N. I. Akhiezer and I. M. Glazman

THEORY OF LINEAR OPERATORS IN HILBERT SPACE

(Teoriia lineinykh operatorov v Gil'bertovom prostranstve, Chapters I to V)

Translated from the Russian by
MERLYND NESTELL
Southern Missionary College

VOLUME I

FREDERICK UNGAR PUBLISHING CO
NEW YORK

Chapter I

HILBERT SPACE

1. Linear Spaces

A set R of elements f, g, h, . . . , (also called points or vectors) forms a *linear space* if

(a) there is an operation, called addition and denoted by the symbol $+$, with respect to which R is an abelian group (the zero element of this group is denoted by 0); $AB = BA$

(b) multiplication of elements of R by (real or complex) numbers $\alpha, \beta, \gamma, \ldots$ is defined such that

$$\alpha(f+g) = \alpha f + \alpha g,$$

$$(\alpha + \beta) f = \alpha f + \beta f,$$

$$\alpha(\beta f) = (\alpha\beta) f,$$

$$1 \cdot f = f, \quad 0 \cdot f = 0.$$

Elements $f_1, f_2, \ldots, f_n$ in R are *linearly independent* if the relation

$$\alpha_1 f_1 + \alpha_2 f_2 + \ldots + \alpha_n f_n = 0 \tag{1}$$

holds only in the trivial case with $\alpha_1 = \alpha_2 = \ldots = \alpha_n = 0$; otherwise $f_1, f_2, \ldots, f_n$ are *linearly dependent*. The left member of equation (1) is called a *linear combination* of the elements $f_1, f_2, \ldots, f_n$. Thus, linear independence of $f_1, f_2, \ldots, f_n$ means that every nontrivial linear combination of these elements is different from zero. If one of the elements $f_1, f_2, \ldots, f_n$ is equal to zero, then these elements are evidently linearly dependent. If, for example, $f_1 = 0$, then we obtain the nontrivial relation (1) by taking

$$\alpha_1 = 1, \alpha_2 = \alpha_3 = \ldots = \alpha_n = 0.$$

A linear space R is *finite dimensional* and, moreover, *n-dimensional* if R contains n linearly independent elements and if any $n + 1$ elements of R are linearly dependent. Finite dimensional linear spaces are studied in linear algebra. If a linear space has arbitrarily many linearly independent elements, then it is *infinite dimensional*.

2. The Scalar Product

A linear space R is *metrizable* if for each pair of elements $f, g \in$ R there is a (real or complex) number (f, g) which satisfies the conditions:[1]

(a) $(g,f) = \overline{(f,g)}$,
(b) $(\alpha_1 f_1 + \alpha_2 f_2, g) = \alpha_1 (f_1, g) + \alpha_2 (f_2, g)$,
(c) $(f,f) \geqq 0$, with equality only for $f = 0$.

The number (f,g) is called the *scalar product*[2] of f and g.

Property (b) expresses the linearity of the scalar product with respect to its first argument. The analogous property with respect to the second argument is

$$(\bar{\text{b}}) \qquad (f, \beta_1 g_1 + \beta_2 g_2) = \bar{\beta}_1 (f, g_1) + \bar{\beta}_2 (f, g_2).$$

Property ($\bar{\text{b}}$) is derived as follows:

$$(f, \beta_1 g_1 + \beta_2 g_2) = \overline{(\beta_1 g_1 + \beta_2 g_2, f)} = \overline{\beta_1 (g_1, f) + \beta_2 (g_2, f)} = \\ = \bar{\beta}_1 (f, g_1) + \bar{\beta}_2 (f, g_2).$$

The positive square root $\sqrt{(f,f)}$ is called the *norm* of the element (vector) f and is denoted by the symbol $\|f\|$. The norm is analogous to the length of a line segment. As with line segments, the norm of a vector is zero if and only if it is the zero vector. In addition, it follows that

$$1^\circ \qquad \|\alpha f\| = |\alpha| \cdot \|f\|.$$

This is shown by using properties (b) and ($\bar{\text{b}}$) of the scalar product:

$$(\alpha f, \alpha f) = \alpha (f, \alpha f) = \alpha \bar{\alpha} (f,f) = |\alpha|^2 (f,f),$$

from which 1° follows.

We shall prove that for any two vectors f and g,

$$2^\circ \qquad |(f,g)| \leqq \|f\| \cdot \|g\|,$$

with equality if and only if f and g are linearly dependent. We call 2° the Cauchy-Bunyakovski inequality[3], because in the two most important particular cases, about which we shall speak below, it was first used by Cauchy and Bunyakovski.

In the proof of 2°, we may assume that $(f,g) \neq 0$. Letting

$$\theta = \frac{(f,g)}{|(f,g)|},$$

we find that for any real λ,

$$0 \leqq (\bar{\theta} f + \lambda g, \bar{\theta} f + \lambda g) = \lambda^2 (g,g) + 2\lambda |(f,g)| + (f,f).$$

[1] A bar over a complex number denotes complex conjugation.
[2] Translator's Note: The phrase *inner product* is also used. Henceforth, we shall call R an *inner product space*.
[3] Translator's Note: This is often called the Schwarz or the Cauchy-Schwarz inequality.

On the right we have a quadratic in λ. For real λ this polynomial is non-negative, which implies that

$$|(f,g)|^2 \leqq (f,f)\cdot(g,g),$$

and this proves 2°. The equality sign will hold only in case the polynomial under consideration has a double root, in other words, only if

$$\bar{\theta} f + \lambda g = 0$$

for some real λ. But this equation implies that the vectors f and g are linearly dependent.

We shall derive one more property of the norm, the inequality,

3° $$\|f+g\| \leqq \|f\| + \|g\|.$$

There is equality if $f=0$ or $g=\lambda f$, $\lambda \geqq 0$. This property is called the *triangle inequality*, by analogy with the inequality for the sides of a triangle in elementary geometry.

In order to prove the triangle inequality, we use the relation

$$\|f+g\|^2 = (f+g, f+g) = (f,f) + (f,g) + (g,f) + (g,g).$$

Hence, by the Cauchy-Bunyakovski inequality

$$\|f+g\|^2 \leqq \|f\|^2 + 2\|f\|\cdot\|g\| + \|g\|^2 = \{\|f\| + \|g\|\}^2$$

which implies that

$$\|f+g\| \leqq \|f\| + \|g\|.$$

For equality, it is necessary that

$$(g,f) = \|f\|\cdot\|g\|.$$

If $f \neq 0$, then, by 2°, it is necessary that

$$g = \lambda f$$

for some λ. From this it is evident that

$$\lambda(f,f) = \|f\|\cdot\|\lambda f\|,$$

whence it also follows that $\lambda \geqq 0$.

An inner product space R becomes a *metric space*, if the distance between two points $f, g \in$ R is defined as

$$D[f,g] = \|f-g\|.$$

It follows from the properties of the norm that the distance function satisfies the usual conditions.[4]

The scalar product yields a definition for the angle between two vectors. However, for what follows, this concept will not be needed. We confine ourselves to the more limited concept of *orthogonality*. Two vectors f and g are *orthogonal* if

$$(f,g) = 0.$$

[4] These conditions are

(a) $D[f,g] = D[g,f] > 0$ (for $f \neq g$),

(b) $D[f,f] = 0$,

(c) $D[f,g] \leqq D[f,h] + D[h,g]$ (triangle inequality).

3. Some Topological Concepts

In the present section we consider some general concepts which are introduced in the study of point sets in an arbitrary metric space. We denote a metric space by E, and speak of the distance $D[f, g]$ between two elements of E. Let us bear in mind that in what will follow we shall consider only the case with $E = R$ and $D[f, g] = \|f - g\|$, i.e., the case with the metric generated by a scalar product.

If f_0 is a fixed element of E, and ρ is a positive number, then the set of all points f for which

$$D[f, f_0] < \rho$$

is called the *sphere* in E with *center* f_0 and *radius* ρ. Such a sphere is a neighborhood, more precisely a ρ-neighborhood of the point f_0.

We say that a sequence of points $f_n \in E$ $(n = 1, 2, 3, \ldots)$ has the *limit point* $f \in E$, and we write

$$f_n \to f \text{ or } \lim_{n\to\infty} f_n = f \tag{1}$$

when

$$\lim_{n\to\infty} D[f_n, f] = 0. \tag{2}$$

It is not difficult to see that (1) implies

$$\lim_{m,\, n\to\infty} D[f_m, f_n] = 0 \tag{3}$$

where m and n tend to infinity independently. In fact, by the triangle inequality,

$$D[f_m, f_n] \leqq D[f_m, f] + D[f_n, f].$$

But the converse is not always correct, i.e., if for the sequence $f_n \in E$ $(n = 1, 2, 3, \ldots)$ relation (3) holds, then there may not exist an element $f \in E$ to which the sequence converges. If (3) is satisfied, then the sequence is called *fundamental*. Thus, a fundamental sequence may or may not converge to an element of the space.

A metric space E is called *complete* if every fundamental sequence in E converges to some element of the space. If a metric space is not complete, then it is possible to complete it by introducing certain new elements. This operation is similar to the introduction of irrational numbers by Cantor's method.

If each neighborhood of $f \in E$ contains infinitely many points of a set M in E, then f is called a limit point of M. If a set contains all its limit points, then it is said to be *closed*. The set consisting of M and its limit points is called the *closure* of M and is denoted by $\overline{M}$.

If the metric space E is the closure of some countable subset of E, then E is said to be *separable*. Thus, in a separable space there exists a countable set N such that, for each point $f \in E$ and each $\varepsilon > 0$, there exists a point $g \in N$ such that

$$D[f, g] < \varepsilon.$$

4. Hilbert Space

A *Hilbert space* H is an infinite dimensional inner product space which is a complete metric space with respect to the metric generated by the inner product. This definition, similar to those in preceding sections, has an axiomatic character. Various concrete linear spaces satisfy the conditions in the definition. Therefore, H is often called an *abstract* Hilbert space, and the concrete spaces mentioned are called *examples* of this abstract space.

One of the important examples of H is the space l^2. The construction of the general theory, to which the present book is devoted, was begun for this particular space by Hilbert in connection with his theory of linear integral equations. The elements of the space l^2 are sequences (of real or complex numbers)

$$f = \{x_n\}_1^\infty, \qquad g = \{y_n\}_1^\infty, \ldots,$$

such that

$$\sum_{n=1}^{\infty} |x_n|^2 < \infty, \qquad \sum_{n=1}^{\infty} |y_n|^2 < \infty, \ldots \quad .$$

The numbers $x_1, x_2, x_3, \ldots$, are called *components* of the vector f or *coordinates* of the point f. The zero vector is the vector with all components zero. The addition of vectors is defined by the formula

$$f + g = \{x_n + y_n\}_1^\infty.$$

The relation

$$\sum_{n=1}^{\infty} |x_n + y_n|^2 < \infty$$

follows from the inequality

$$|x + y|^2 \leq 2\,|x|^2 + 2\,|y|^2.$$

The multiplication of a vector f by a number λ is defined by

$$\lambda f = \{\lambda x_n\}_1^\infty.$$

The scalar product in the space l^2 has the form

$$(f, g) = \sum_{n=1}^{\infty} x_n \bar{y}_n.$$

The series on the right converges absolutely because

$$|xy| \leq \tfrac{1}{2}|x|^2 + \tfrac{1}{2}|y|^2.$$

The inequality

$$|(f,g)| \leqq \|f\| \cdot \|g\|$$

now has the form

$$\left|\sum_{n=1}^{\infty} x_n \bar{y}_n\right| \leqq \sqrt{\sum_{n=1}^{\infty} |x_n|^2} \cdot \sqrt{\sum_{n=1}^{\infty} |y_n|^2}$$

and is due to Cauchy.

The space l^2 is separable. A particular countable dense subset of l^2 consists of all vectors with only finitely many nonzero components and with these components rational, i.e., the components are of the form $\xi + i\eta$ where ξ and η are rational numbers.

In addition, the space l^2 is complete. In fact, if the sequence of vectors

$$f^{(k)} = \{x_n^{(k)}\}_{n=1}^{\infty} \qquad (k = 1, 2, 3, \ldots)$$

is fundamental, then each of the sequences of numbers

$$\{x_n^{(k)}\}_{k=1}^{\infty} \qquad (n = 1, 2, 3, \ldots)$$

is fundamental and, hence, converges to some limit x_n ($n = 1, 2, 3, \ldots$). Now, for each $\varepsilon > 0$ there exists an integer N such that for $r > N, s > N$

$$\sqrt{\sum_{n=1}^{\infty} |x_n^{(r)} - x_n^{(s)}|^2} < \varepsilon.$$

Consequently, for every m,

$$\sqrt{\sum_{n=1}^{m} |x_n^{(r)} - x_n^{(s)}|^2} < \varepsilon.$$

Let s tend to infinity to obtain

$$\sqrt{\sum_{n=1}^{m} |x_n^{(r)} - x_n|^2} \leqq \varepsilon.$$

But, because this is true for every m,

$$\sqrt{\sum_{n=1}^{\infty} |x_n^{(r)} - x_n|^2} \leqq \varepsilon.$$

Hence, it follows that

$$f = \{x_n\}_1^{\infty} \in l^2,$$

because
$\sum_m |x_n|^2 \le 2\sum_n |x_n^{(r)} - x_n|^2 + 2\sum_m |x_n^{(r)}|$
$= x_n - x_n^{(r)} + x_n^{(r)}$

and, since $\varepsilon > 0$ is arbitrary,

$$f^{(k)} \to f.$$

Thus, the completeness of the space l^2 is established.

As we demonstrated, the space l^2 is separable. Originally, the requirement of separability was included in the definition of an abstract Hilbert space. However, as time passed it appeared that this requirement was not necessary for a great deal of the theory, and therefore, it is not included in our definition of the space H.

But the requirement of completeness is essential for almost all of our considerations. Therefore, it is included in the definition of H. The appropriate reservation is made in the cases for which this requirement is superfluous.

The space l^2 is infinite dimensional because the *unit vectors*

$$\begin{aligned} e_1 &= \{1, 0, 0, \ldots\}, \\ e_2 &= \{0, 1, 0, \ldots\}, \\ e_3 &= \{0, 0, 1, \ldots\}, \\ &\ldots\ldots\ldots, \end{aligned}$$

are linearly independent. The space l^2 is the infinite dimensional analogue of E_m, the (complex) m-dimensional *Euclidean space*, the elements of which are finite sequences

$$f = \{x_n\}_1^m,$$

and most of the theory which we present consists of generalizations to H of well-known facts concerning E_m.

5. Linear Manifolds and Subspaces

One often considers particular subsets of R (and, in particular, of H). Such a subset L is called a *linear manifold* if the hypothesis $f, g \in$ L implies that $\alpha f + \beta g \in$ L for arbitrary numbers α and β. One of the most common methods of obtaining a linear manifold is the construction of a *linear envelope*. The point of departure is a finite or infinite set M of elements of R. Consider the set L of all finite linear combinations

$$\alpha_1 f_1 + \alpha_2 f_2 + \ldots + \alpha_n f_n$$

of elements $f_1, f_2, \ldots, f_n$ of M. This set L is the smallest linear manifold which contains M. It is called the linear envelope of M or the linear manifold spanned by M. If R is a metric space, then the closure of the linear envelope of a set M is called the *closed linear envelope* of M.

In what follows, closed linear manifolds in H will have a particularly important significance. Each such manifold G is a linear space, metrizable with respect to the scalar product defined in H. Furthermore, G is complete. In fact, every fundamental sequence of elements of G has a limit in H because H is complete, and this limit must belong to G because G is closed. From what has been said, it follows that G itself is a Hilbert space if it contains an infinite number of linearly independent elements; otherwise G is a Euclidean space. Therefore, G is called a *subspace* of the space H.

6. The Distance from a Point to a Subspace

Consider a linear manifold L which is a *proper subset* of H. Choose a point $h \in H$ and let

$$\delta = \inf_{f \in L} \| h - f \|.$$

The question arises as to whether there exists a point $g \in L$ for which

$$\| h - g \| = \delta.$$

In other words, is there a point in L nearest to the point h?[5]

We prove first that there exists at most one point $g \in L$ such that $\delta = \| h - g \|$. Assume that there exist two such points, g' and g''. Since $\frac{1}{2}(g' + g'') \in L$, we have

$$\left\| h - \frac{g' + g''}{2} \right\| \geqq \delta;$$

on the other hand

$$\left\| h - \frac{g' + g''}{2} \right\| \leqq \frac{1}{2} \| h - g' \| + \frac{1}{2} \| h - g'' \| = \delta.$$

Consequently,

$$\left\| h - \frac{g' + g''}{2} \right\| = \delta$$

and therefore

$$\left\| h - \frac{g' + g''}{2} \right\| = \frac{1}{2} \| h - g' \| + \frac{1}{2} \| h - g'' \|.$$

But this is the triangle inequality with the sign of equality. Since

$$h - g' \neq 0$$

we have

$$h - g'' = \lambda (h - g')$$

for some $\lambda \geqq 0$. If $\lambda = 1$, the proof is complete. If $\lambda \neq 1$, then

$$h = \frac{g'' - \lambda g'}{1 - \lambda}$$

so that $h \in L$, which contradicts our assumption. Thus, our assertion is proved.

But, in general, does there exist a point $g \in L$ nearest to the point h? In the most important case, the answer is yes, and the following theorem holds.

THEOREM: *If* G *is a subspace of the space* H *and if*

$$\delta = \inf_{f \in G} \| h - f \|,$$

[5] Translator's Note. The case with $h \in L$ is trivial. Henceforth, the author assumes without saying so that $h \notin L$.

then there exists a vector $g \in G$ (its uniqueness was proved above) for which

$$\|h - g\| = \delta.$$

Proof: According to the definition of the greatest lower bound, there exists an infinite sequence of vectors $\{g_n\}_1^\infty$, in G, for which

$$\lim_{n\to\infty} \| h - g_n \| = \delta.$$

Now,

$$\left\| h - \frac{g_m + g_n}{2} \right\| \leqq \frac{1}{2}\| h - g_m \| + \frac{1}{2}\| h - g_n \|.$$

Therefore

$$\overline{\lim_{m,\, n\to\infty}} \left\| h - \frac{g_m + g_n}{2} \right\| \leqq \delta,$$

and since

$$\left\| h - \frac{g_m + g_n}{2} \right\| \geqq \delta$$

we have

$$\lim_{m,\, n\to\infty} \left\| h - \frac{g_m + g_n}{2} \right\| = \delta.$$

In the easily proved relation,

$$2\|f'\|^2 + 2\|f''\|^2 = \|f' + f''\|^2 + \|f' - f''\|^2,$$

let

$$f' = h - g_m, \qquad f'' = h - g_n$$

to obtain

$$\| g_n - g_m \|^2 = 2\| h - g_m\|^2 + 2\| h - g_n\|^2 - 4\left\| h - \frac{g_m + g_n}{2} \right\|^2.$$

Therefore,

$$\lim_{m,\, n\to\infty} \|g_n - g_m\| = 0.$$

So the sequence of vectors $\{g_n\}_1^\infty$ converges to some vector $g \in G$. It remains to prove that

$$\| h - g \| = \delta.$$

Now

$$\lim_{n\to\infty} \|g - g_n\| = 0, \qquad \lim_{n\to\infty} \|h - g_n\| = \delta$$

and

$$\|h - g\| \leqq \|h - g_n\| + \|g - g_n\|;$$

consequently

$$\| h - g \| \leqq \delta.$$

But, by the hypothesis of the theorem, $\| h - g \| \geqq \delta$. Thus, the theorem is proved.

7. Projection of a Vector on a Subspace

Let G be a subspace of H. By the preceding section, to each element $h \in \mathrm{H}$ there corresponds a unique element $g \in \mathrm{G}$ such that

$$\|h - g\| = \inf_{g' \in \mathrm{G}} \|h - g'\|. \tag{1}$$

Considering h and g as points, we say that g is the point of the subspace G nearest the point h. If the elements g and h are considered as vectors, then it is said that g is the particular vector of G which deviates least from h. Now, using (1), we show that the vector $h-g$ is orthogonal to the subspace G; i.e., orthogonal to every vector $g' \in \mathrm{G}$.

For the proof, we assume that the vector $h-g$ is not orthogonal to every vector $g' \in \mathrm{G}$. Let

$$(h - g, g_0) = \sigma \neq 0 \qquad (g_0 \in \mathrm{G}).$$

We define the vector

$$g^* = g + \frac{\sigma}{(g_0, g_0)} g_0 \in \mathrm{G}.$$

Then

$$\begin{aligned} \|h - g^*\|^2 &= \left(h - g - \frac{\sigma}{(g_0, g_0)} g_0, h - g - \frac{\sigma}{(g_0, g_0)} g_0\right) = \\ &= \|h - g\|^2 - \frac{\bar{\sigma}}{(g_0, g_0)}(h - g, g_0) - \frac{\sigma}{(g_0, g_0)}(g_0, h-g) + \frac{|\sigma|^2}{(g_0, g_0)} = \\ &= \|h - g\|^2 - \frac{|\sigma|^2}{(g_0, g_0)}, \end{aligned}$$

so that

$$\|h - g^*\| < \|h - g\|,$$

which contradicts (1).

From the proof it follows that h has a representation of the form[6]

$$h = g + f,$$

where $g \in \mathrm{G}$ and f is orthogonal to G (in symbols, $f \perp \mathrm{G}$). It follows easily that

$$\|h\|^2 = \|g\|^2 + \|f\|^2.$$

It is natural to call g the *component* of h in the subspace G or the *projection* of h on G.[7]

---

[6] Translator's Note: It is easy to prove that this representation is unique. The author uses the uniqueness later.

[7] This contrasts with the situation in analytic geometry, where a projection is a number and a component is a vector. Here projection and component are equivalent terms.

We denote by F the set of all vectors f orthogonal to the subspace G. We show that F is closed, so that F is a subspace. In fact, let $f_n \in$ F ($n = 1, 2, 3, \ldots$) and $f_n \to f$. Then $(f_n, g) = 0$ and

$$(f, g) = (f - f_n, g).$$

In absolute value, the right member does not exceed

$$\|f - f_n\| \cdot \|g\|,$$

which converges to zero as $n \to \infty$. Hence, $(f, g) = 0$, so that $f \in$ F and the manifold F is closed.

The subspace F is called the *orthogonal complement* of G and is expressed by

$$\text{(2)} \qquad \mathrm{F} = \mathrm{H} \ominus \mathrm{G}.$$

It is easy to see that

$$\text{(2')} \qquad \mathrm{G} = \mathrm{H} \ominus \mathrm{F}.$$

Both relations (2) and (2′) are expressed by the equation

$$\mathrm{H} = \mathrm{G} \oplus \mathrm{F},$$

because H is the so-called direct sum of the subspaces F and G (in the given case, the *orthogonal sum*).

In general, a set $\mathrm{M} \subset \mathrm{H}$ is called the *direct sum* of a finite number of linear manifolds $\mathrm{M}_k \subset \mathrm{H}$ ($k = 1, 2, 3, \ldots, n$) and is expressed by

$$\mathrm{M} = \mathrm{M}_1 \oplus \mathrm{M}_2 \oplus \ldots \oplus \mathrm{M}_n$$

if each element $g \in$ M is represented uniquely in the form

$$g = g_1 + g_2 + \ldots + g_n$$

where $g_k \in \mathrm{M}_k$ ($k = 1, 2, 3 \ldots, n$). It is evident that M is also a linear manifold.

It will be necessary for us to consider direct sums of an infinite number of linear manifolds only in cases for which the manifolds are pairwise orthogonal subspaces of the given space. This is done as follows.

DEFINITION: *Let* $\{\mathrm{H}_a\}$ *be a countable or uncountable class of pairwise orthogonal subspaces of* H. *Their orthogonal sum*

$$\sum_a \oplus \mathrm{H}_a$$

is defined as the closed linear envelope of the set of all finite sums of the form

$$\mathrm{H}_{a'} \oplus \mathrm{H}_{a''} \oplus \ldots.$$

Often it is necessary to determine the projection of a vector on a finite dimensional subspace. We consider this question in some detail. Let G be an n-dimensional subspace and let

$$(3) \qquad g_1, g_2, \ldots, g_n$$

be n linearly independent elements of G. Since any $n+1$ elements of G are linearly dependent, each vector $g' \in$ G can be represented (uniquely) in the form

$$g' = \lambda_1 g_1 + \lambda_2 g_2 + \ldots + \lambda_n g_n.$$

In other words, G is the linear envelope of the set of vectors (3).

We choose an arbitrary vector $h \in$ H and denote by g its projection on G. The vector g has a unique representation,

$$g = \alpha_1 g_1 + \alpha_2 g_2 + \ldots + \alpha_n g_n.$$

According to the definition of a projection, the difference $h - g = f$ must be orthogonal to the subspace G, i.e., f is orthogonal to each of the vectors $g_1, g_2, \ldots, g_n$. Therefore,

$$(4) \quad (f, g_k) \equiv (h, g_k) - \alpha_1(g_1, g_k) - \alpha_2(g_2, g_k) - \ldots - \alpha_n(g_n, g_k) = 0$$
$$(k = 1, 2, 3, \ldots, n).$$

This is a system of n linear equations in the unknowns $\alpha_1, \alpha_2, \ldots, \alpha_n$. We have shown that it has a unique solution for each vector h. Therefore, the determinant of this system is different from zero[8]. This determinant

$$\Gamma(g_1, g_2, \ldots, g_n) = \begin{vmatrix} (g_1, g_1) & (g_2, g_1) & \ldots & (g_n, g_1) \\ (g_1, g_2) & (g_2, g_2) & \ldots & (g_n, g_2) \\ \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots \\ (g_1, g_n) & (g_2, g_n) & \ldots & (g_n, g_n) \end{vmatrix}$$

is called the *Gram determinant* of the vectors $g_1, g_2, \ldots, g_n$. It is easy to see that if the vectors $g_1, g_2, \ldots, g_n$ are linearly dependent, then the Gram determinant is equal to zero. Hence, for the linear independence of the vectors it is necessary and sufficient that their Gram determinant be different from zero.

We proceed to determine the number

$$\delta = \min_{g' \in G} \| h - g' \|.$$

We shall express δ by means of the Gram determinant. As above let g be the projection of h on G and let $f = h - g$. Then $\delta = \|f\| = \|h - g\|$ and

$$\delta^2 = (f, f) = (f, h),$$

[8] Translator's Note: In (4), suppose that $(h, g_k) = 0$ for $k = 1, 2, 3, \ldots, n$. In other words, let $h \perp$ G. Then $\|g\|^2 = \|\alpha_1 g_1 + \ldots + \alpha_n g_n\|^2 = \sum_{k=1}^{n} \bar{\alpha}_k[(\alpha_1 g_1, g_k) + \ldots + (\alpha_n g_n, g_k)] = \sum_{k=1}^{n} \bar{\alpha}_k(h, g_k) = 0$ so that $g = \alpha_1 g_1 + \alpha_2 g_2 + \ldots + \alpha_n g_n = 0$. Since $g_1, \ldots, g_n$ are linearly independent, $\alpha_1 = \alpha_2 = \ldots \alpha_n = 0$. This proves that the homogeneous system $\alpha_1(g_1, g_k) + \ldots + \alpha_n(g_n, g_k) = 0$ $(k = 1, 2, 3, \ldots, n)$ has only the trivial solution $\alpha_1 = \ldots = \alpha_n = 0$. Consequently, the determinant of the system is not zero.

since $(f,g)=0$. Let $g=\alpha_1 g_1+\alpha_2 g_2+\dots+\alpha_n g_n$, where the g_k are as in equation (3), to obtain

(5) $$\delta^2=(h,h)-\alpha_1(g_1,h)-\alpha_2(g_2,h)-\dots-\alpha_n(g_n,h).$$

The determination of δ^2 is reduced to the elimination of the quantities α_i from equations (4) and (5). This elimination yields

$$\begin{vmatrix} (h,h)-\delta^2 & (g_1,h) & (g_2,h) & \dots & (g_n,h) \\ (h,g_1) & (g_1,g_1) & (g_2,g_1) & \dots & (g_n,g_1) \\ \dots & \dots & \dots & \dots & \dots \\ (h,g_n) & (g_1,g_n) & (g_2,g_n) & \dots & (g_n,g_n) \end{vmatrix}=0.$$

Hence,

(6) $$\delta^2=\frac{\Gamma(h,g_1,g_2,\dots,g_n)}{\Gamma(g_1,g_2,\dots,g_n)}.$$

This is the formula we wished to obtain.

Since $\Gamma(g_1)=(g_1,g_1)>0$ (for $g_1\neq 0$), it follows from formula (6) that the Gram determinant of linearly independent vectors is always positive. This fact can be regarded as a generalization of the Cauchy-Bunyakovski inequality, which asserts that

$$\Gamma(g_1,g_2)>0$$

for linearly independent vectors g_1 and g_2.

8. Orthogonalization of a Sequence of Vectors

Two sets M and N of vectors in H are said to be *equivalent* if their linear envelopes coincide. Therefore, the sets M and N are equivalent if and only if each element of one of these sets is a linear combination of a finite number of vectors belonging to the other set.

If the elements of the set M are pairwise orthogonal vectors, and if each of the vectors is *normalized*, i.e., if each has norm equal to one, then the set M is called an *orthonormal system*. If, in addition, the set M is countable, then it is also called an *orthonormal sequence*.

Suppose given a finite or infinite sequence of independent vectors

(1) $$g_1,g_2,\dots,g_n,\dots .$$

We show how to construct an equivalent orthonormal sequence of vectors

(2) $$e_1,e_2,\dots,e_n,\dots .$$

For the first vector, we take

$$e_1=\frac{g_1}{\|g_1\|},$$

the norm of which is equal to one. The vectors e_1 and g_1 generate the same (one dimensional) subspace E_1. The vector e_2 is constructed in two steps. First, we subtract from the vector g_2 its projection on E_1 to get

$$h_2 = g_2 - (g_2, e_1)e_1,$$

which is orthogonal to the subspace E_1. Since the vectors (1) are linearly independent, g_2 does not belong to E_1, so that $h_2 \neq 0$. Now let

$$e_2 = \frac{h_2}{\| h_2 \|}.$$

The vectors e_1 and e_2 generate the same (two dimensional) subspace E_2 as do the vectors g_1 and g_2. We now construct the vector e_3. First, we subtract from g_3 its projection on E_2 to get

$$h_3 = g_3 - (g_3, e_1)e_1 - (g_3, e_2)e_2,$$

which is different from zero and orthogonal to the subspace E_2, i.e., h_3 is orthogonal to each of the vectors e_1 and e_2. Next we let

$$e_3 = \frac{h_3}{\| h_3 \|}.$$

We continue in the same way. If the vectors

$$e_1, e_2, \ldots, e_n$$

have been constructed, then we let

$$h_{n+1} = g_{n+1} - \sum_{k=1}^{n} (g_{n+1}, e_k)\, e_k,$$

and

$$e_{n+1} = \frac{h_{n+1}}{\| h_{n+1} \|}.$$

The method described is called *orthogonalization.*[9]

In the solution of many problems concerning manifolds generated by a given sequence of vectors, preliminary orthogonalization of the sequence turns out to be very useful. We illustrate this in the problem considered in the preceding paragraph. That problem concerned the determination of the distance from a point $h \in \mathrm{H}$ to a linear manifold G, which was the closed linear envelope of the given sequence (1). We shall show how elegantly this problem is solved if the system (1) is orthogonalized beforehand.

[9] Often, in particular cases, one does not bother to normalize the system (2) of pairwise orthogonal elements which is equivalent to the system (1). The transition from (1) to such an orthogonal system likewise is called orthogonalization. (See Section 11 below.)

Assume given the orthogonal sequence (2) and a vector $h \in \mathrm{H}$. For each integer n, the vector h can be expressed in the form

$$h = \sum_{k=1}^{n} (h, e_k)\, e_k + f_n$$

where the vector f_n is orthogonal to each of the vectors $e_1, e_2, \ldots, e_n$. The vector

(3) $$s_n = \sum_{k=1}^{n} (h, e_k)\, e_k$$

belongs to the set of vectors

(4) $$\lambda_1 e_1 + \lambda_2 e_2 + \ldots + \lambda_n e_n$$

and, of these vectors, s_n is nearest to the vector h. The distance from s_n to h is

(5) $$\delta_n = \min_{\lambda_k} \| h - \lambda_1 e_1 - \lambda_2 e_2 - \ldots - \lambda_n e_n \| = \| f_n \| = \sqrt{\|h\|^2 - \sum_{k=1}^{n} |(h, e_k)|^2}.$$

This is the distance from the point h to the linear envelope G_n of the set consisting of the first n vectors of the sequence (2). If instead of the linear combination of the nth order (4), we wish to find the linear combination of the $(n + 1)$th order,

(4′) $$\mu_1 e_1 + \mu_2 e_2 + \ldots + \mu_{n+1} e_{n+1},$$

which is nearest to the vector h, then we must take the vector

$$s_{n+1} = \sum_{k=1}^{n+1} (h, e_k)\, e_k.$$

Thus, we do not change the coefficients in the linear combination (3). Rather, we merely add one more term,

$$(h, e_{n+1}) e_{n+1}$$

to the right member of (3).

These considerations show that, being given the infinite orthonormal sequence (2), it is appropriate to associate with each vector $h \in \mathrm{H}$ the *infinite series*

(6) $$\sum_{k=1}^{\infty} (h, e_k) e_k.$$

Equation (5) yields the important inequality

(7) $$\sum_{k=1}^{\infty} |(h, e_k)|^2 \leqq \| h \|^2.$$

The convergence of the series

$$\sum_{k=1}^{\infty} |(h, e_k)|^2$$

implies that

$$\|s_n - s_m\|^2 = \sum_{k=m+1}^{n} |(h, e_k)|^2 \qquad (m < n)$$

converges to zero as $m, n \to \infty$, i.e., the series (6) converges in H.[10] We see that the square of the distance from the point h to the manifold G is

$$\|h\|^2 - \sum_{k=1}^{\infty} |(h, e_k)|^2$$

and that the vector h belongs to the manifold G if and only if there is equality in formula (7).

We shall say that a system of vectors is *closed* in H if its linear envelope is dense in H. From our considerations, the orthonormal system (2) is closed in H if and only if

$$\|h\|^2 = \sum_{k=1}^{\infty} |(h, e_k)|^2 \tag{8}$$

for each $h \in$ H. Following V. A. Steklov,[11] we call this equality the *closure relation.*[12] We show next that if the Parseval relation holds for each vector $h \in$ H, then for any pair of vectors $g, h \in$ H, the general Parseval relation

$$(g, h) = \sum_{k=1}^{\infty} (g, e_k)(e_k, h) \tag{9}$$

holds. In fact, we have the Parseval relation for each vector $g + \lambda h$:

$$\|g + \lambda h\|^2 = \sum_{k=1}^{\infty} |(g + \lambda h, e_k)|^2,$$

which yields

$$(g, g) + \lambda(h, g) + \bar{\lambda}(g, h) + |\lambda|^2 (h, h) =$$

$$= \sum_{k=1}^{\infty} \{|(g, e_k)|^2 + \lambda(h, e_k)(e_k, g) + \bar{\lambda}(g, e_k)(e_k, h) + |\lambda|^2 |(h, e_k)|^2\}$$

and

$$\lambda(h, g) + \bar{\lambda}(g, h) = \lambda \sum_{k=1}^{\infty} (h, e_k)(e_k, g) + \bar{\lambda} \sum_{k=1}^{\infty} (g, e_k)(e_k, h).$$

Since λ is arbitrary, equation (9) follows.

---

[10] We remark that convergence in H of any series $f_1 + f_2 + \dots$, where $(f_i, f_k) = 0$ $(i \neq k)$ is equivalent to the convergence of the numerical series $\|f_1\|^2 + \|f_2\|^2 + \|f_3\|^2 + \dots$.

[11] V. A. Steklov showed for the first time the important significance of the closure relation in various problems of analysis and mathematical physics. Before the work of V. A. Steklov, the relation under consideration was known only for systems of trigonometric functions (the so-called Parseval-Liapunov equation).

[12] Translator's Note: Following rather common English practice, we shall refer to (8) henceforth as the *Parseval relation.* Some authors refer to it as the completeness relation.

9. Complete Orthonormal Systems

The vectors of an orthonormal system cannot be linearly dependent. Therefore, in n-dimensional Euclidean space each orthonormal system of vectors contains at most n vectors.

We say that an orthonormal system M is *complete* in H if M is not contained in a larger orthonormal system in H, i.e., if there is no nonzero vector in H which is orthogonal to every vector of the system M. In Euclidean n-space any orthonormal system of n vectors is complete. In Hilbert space a complete orthonormal system contains an infinite number of elements, and there arises the problem of the cardinality of such systems. This problem is solved easily for separable spaces. We begin with them.

THEOREM 1: *If the space* H *is separable, then every orthonormal system of vectors in* H *consists of a finite or countable number of elements.*

Proof: Let

(1) $$f_1, f_2, f_3, \ldots$$

be a sequence of vectors which is dense in H, and let M be an orthonormal system of vectors. We proceed to show that M can be enumerated. Let e and e' be distinct vectors in M. From (1) choose vectors f_k and $f_{k'}$ such that

(2) $$\|e - f_k\| < \tfrac{1}{2}\sqrt{2}$$

and similarly for e' and k'. We show that $k' \neq k$. In fact,

$$\|e - e'\|^2 = \|e\|^2 + \|e'\|^2 = 2$$

so that

$$\sqrt{2} = \|e - e'\| \leqq \|e - f_k\| + \|e' - f_k\| < \tfrac{1}{2}\sqrt{2} + \|e' - f_k\|.$$

Therefore,

$$\|e' - f_k\| > \tfrac{1}{2}\sqrt{2}$$

so that $f_{k'} \neq f_k$ and $k \neq k'$. Thus, we can associate with each vector of M a different integer k. This proves that the set M is finite or countable.

The existence of a nondenumerable orthonormal system of vectors in H implies that the space is not separable. An important example of this kind will be considered later.

THEOREM 2: *An infinite orthonormal sequence*

(3) $$e_1, e_2, e_3, \ldots$$

is complete in H *if and only if the sequence is closed in* H.

Proof: (A) Let the system (3) be closed in H. Then, for each vector of H, the Parseval relation holds. Assume that the system (3) is not complete, and denote by h a nonzero vector which is orthogonal to each of the vectors (3). Thus, $(h, e_k) = 0$ $(k = 1, 2, 3, \ldots)$ and the Parseval relation for h reduces to the contradiction

$$0 \neq \|h\|^2 = 0.$$

(B) We suppose now that the system (3) is complete. We choose an arbitrary vector $h \in \mathrm{H}$ and consider the sequence of vectors

$$s_n = \sum_{k=1}^{n} (h, e_k)\, e_k \qquad (n = 1, 2, 3, \ldots).$$

By the preceding section, $\{s_n\}$ is fundamental, which implies that it converges to some vector g. Then

$$(4) \qquad (g, e_k) = \lim_{n\to\infty} (s_n, e_k) = (h, e_k) \qquad (k = 1, 2, 3, \ldots)$$

and g belongs to the closed linear envelope of the sequence (3). Consequently, the Parseval relation is valid for g:

$$(5) \qquad \| g \|^2 = \sum_{k=1}^{\infty} |(g, e_k)|^2 = \sum_{k=1}^{\infty} |(h, e_k)|^2.$$

It follows from (4) that the vector $g-h$ is orthogonal to each vector of the sequence (3). The assumption that this sequence is complete implies that $g-h = 0$, so that $g = h$, and (5) takes the form

$$\| h \|^2 = \sum_{k=1}^{\infty} |(h, e_k)|^2.$$

We have shown that for an arbitrary vector $h \in \mathrm{H}$, the Parseval relation holds. Thus, it is proved that (3) is closed in H.

THEOREM 3: *The space* H *contains a complete orthonormal sequence if and only if it is separable.*

Proof: (A) We assume that the space H is separable, and let N denote a countable set of vectors which is dense in H. Deleting from the sequence N any vector which is a linear combination of the preceding ones, and orthogonalizing the resulting sequence, we obtain an orthonormal sequence M. This sequence is complete. For, let the vector $h \in \mathrm{H}$ be orthogonal to each element of the sequence M. Then h is orthogonal to each vector of N. For each $\varepsilon > 0$ there exists a vector $f \in \mathrm{N}$ such that

$$\| h - f \| < \varepsilon$$

which implies that

$$\| h \|^2 = (h, h) = (h - f, h) \leqq \| h - f \| \cdot \| h \| < \varepsilon \| h \|$$

and

$$\| h \| < \varepsilon.$$

Since $\varepsilon > 0$ is arbitrary, $h = 0$ and the orthonormal sequence M is complete in H.

(B) We assume that (3) is a complete orthonormal sequence in H. Let N be the set of all linear combinations of the form

$$\gamma_1^{(n)} e_1 + \gamma_2^{(n)} e_2 + \ldots + \gamma_n^{(n)} e_n \qquad (n = 1, 2, 3, \ldots),$$

where $\gamma_k^{(r)} = \alpha_k^{(r)} + i\beta_k^{(r)}$, and $\alpha_k^{(r)}$, $\beta_k^{(r)}$ are rational numbers. The set N is countable. For each $h \in$ H and each $\varepsilon > 0$ there exists an integer n such that

$$\left\| h - \sum_{k=1}^{n} (h, e_k)\, e_k \right\| < \frac{\varepsilon}{2}\,.$$

It is possible to approximate the complex numbers (h, e_k), $(k = 1, 2, 3, \ldots, n)$, by numbers of the form $\gamma_k^{(n)}$ such that

$$\left\| \sum_{k=1}^{n} \{(h, e_k) - \gamma_k^{(n)}\}\, e_k \right\| < \frac{\varepsilon}{2}\,.$$

Thus, there exists a vector,

$$f = \sum_{k=1}^{n} \gamma_k^{(n)} e_k,$$

in N, for which

$$\| h - f \| < \varepsilon,$$

and this implies that H is separable.

The question of the cardinality of a complete orthonormal system in a separable space now can be answered completely: every complete orthonormal system in a separable space is necessarily an infinite sequence --- a so-called *orthonormal basis* of the space.

Now we consider arbitrary Hilbert spaces. First, we remark that whatever the cardinality of an orthonormal system M, each vector f has no more than a countable set of nonzero projections on the elements of the system M. This follows from the fact that for any sequence of elements $e', e'', e''', \ldots$ of M the inequality

$$|(f, e')|^2 + |(f, e'')|^2 + |(f, e''')|^2 + \ldots \leqq \|f\|^2$$

holds, which shows that it is possible to enumerate the set of all nonzero numbers (f, e) with $e \in$ M. Further, we have

THEOREM 4: *Any two complete orthonormal systems in a Hilbert space have the same cardinal number.*

Proof: Let M and N be two orthonormal systems, each complete in H, with cardinalities $\mathfrak{m}$ and $\mathfrak{n}$, respectively. Choose $e \in$ M. At least one of the scalar products (e, f), $f \in$ N, is different from zero because otherwise it would be possible to extend the orthonormal system N by appending to it the vector e. On the other hand, by the remark of the previous paragraph, there exists no more than a countable set of elements $f \in$ N for which $(e, f) \neq 0$. We denote these elements by

$$f_1, f_2, \ldots, f_n \qquad (1 \leqq n \leqq \infty). \tag{6}$$

We define a function φ with domain M such that $\varphi(e)$ is the set of vectors, (6), for which $(e, f) \neq 0$. The function φ is at least single valued and at most countably valued. This function maps M onto a set of countable subsets of N. Each $f^* \in$ N satisfies $f^* \in \varphi(e^*)$ for some $e^* \in$ M, since for each $f^* \in$ N there exists an element $e^* \in$ M which is not orthogonal to the element f^*. From what has been said it follows that

$$\mathfrak{m} \geqq \mathfrak{n}.$$

Reversing the roles of the systems M and N, we get

$$\mathfrak{n} \geqq \mathfrak{m}.$$

And so

$$\mathfrak{m} = \mathfrak{n},$$

and the theorem is proved.[13] The following definition is based on this theorem.

DEFINITION: *The dimension of a Hilbert space* H *is the cardinality of a complete orthonormal system in* H.

It is not necessary to make a separate definition of the dimension of a subspace G $\subset$ H. The dimension of an arbitrary linear manifold L $\subset$ H is defined as the dimension of the corresponding subspace $\bar{L}$.

If two Hilbert spaces H and H′ have the same dimension, then they are *isomorphic* in the sense that there exists a one-to-one correspondence between H and H′ having the following property. If the elements $f, g \in$ H correspond to the elements $f', g' \in$ H′, respectively, then (1) $\alpha f + \beta g$ corresponds to $\alpha f' + \beta g'$ and (2) $(f, g)_H = (f', g')_{H'}$. In fact, since the spaces H and H′ have the same dimension they possess complete orthonormal systems of identical cardinalities. We choose any one-to-one correspondence between the elements of these two orthonormal systems and extend this correspondence to the linear envelopes of the orthogonal systems under consideration in such a way that condition (1) is satisfied. Then condition (2) is automatically satisfied, which permits one to get, by passage to the limit, the required correspondence for all the elements of the spaces H and H′.

From the proof it follows that any separable space is isomorphic to the space l^2. It is evident that two Hilbert spaces of different dimensions are not isomorphic. Therefore, two abstract Hilbert spaces (similarly, two abstract Euclidean spaces) differ from each other only in their dimensions.

[13] The actual construction of a complete orthonormal system in a non-separable space requires transfinite induction.

10. The Space L^2

Let (a, b) denote a finite or infinite interval[14] on the real axis. We denote by $L^2(a, b)$ (or simply by L^2) the set of all complex valued Lebesgue measurable functions f defined on (a, b) such that $|f|^2$ is Lebesgue integrable on (a, b). We do not regard as distinct elements of L^2 a pair of functions which differ only on a set of measure zero.

It follows by means of the inequality

$$|\alpha + \beta|^2 \leqq 2|\alpha|^2 + 2|\beta|^2$$

that $f + g \in L^2$ whenever $f, g \in L^2$. Furthermore, for each complex number λ and each $f \in L^2$, it follows that $\lambda f \in L^2$. Thus, L^2 is a linear space and the zero element is a function which is equal to zero almost everywhere in (a, b). In this linear space the scalar product is defined by the formula

$$(f, g) = \int_a^b f(t)\overline{g(t)}\,dt.$$

The existence of the integral on the right side is a consequence of the inequality

$$|\alpha\beta| \leqq \tfrac{1}{2}|\alpha|^2 + \tfrac{1}{2}|\beta|^2.$$

In the present case, the inequality

$$|(f, g)| \leqq \|f\| \cdot \|g\|$$

has the form

$$\left| \int_a^b f(t)\overline{g(t)}\,dt \right| \leqq \sqrt{\int_a^b |f(t)|^2\,dt}\ \sqrt{\int_a^b |g(t)|^2\,dt}\,.$$

This inequality was obtained by Bunyakovski for Riemann integrals.

Now we show that L^2 is complete, from which it will follow that L^2 is a Hilbert space. Let the sequence of functions $f_n \in L^2$ $(n = 1, 2, 3, \ldots)$ be fundamental, i.e., let

$$\lim_{m, n \to \infty} \int_a^b |f_n(t) - f_m(t)|^2\,dt = 0.$$

Then there exists an infinite sequence of integers

$$k_1 < k_2 < k_3 < \ldots < k_r < \ldots,$$

[14] Instead of an interval, the domain of definition of the function could be any measurable set (of finite or infinite measure) on the real axis, on the plane, or in Euclidean n-space.

for which

$$\int_a^b |f_{k_{r+1}}(t) - f_{k_r}(t)|^2 dt < \frac{1}{8^r} \qquad (r = 1, 2, 3, \ldots).$$

From this inequality it follows that the set of points of the interval (a, b) for which

$$|f_{k_{r+1}}(t) - f_{k_r}(t)| \geqq \frac{1}{2^r}$$

has measure less than $\frac{1}{2^r}$. For $s = 1, 2, 3, \ldots$, let I_s denote the set of points $t \in (a, b)$ such that

$$|f_{k_{s+1}}(t) - f_{k_s}(t)| < \frac{1}{2^s},$$

$$|f_{k_{s+2}}(t) - f_{k_{s+1}}(t)| < \frac{1}{2^{s+1}},$$

$$\cdots\cdots\cdots\cdots\cdots\cdots$$

The complement of I_s with respect to the interval $I = (a, b)$ has measure

$$m(I - I_s) < \sum_{n=s}^{\infty} \frac{1}{2^n} = \frac{1}{2^{s-1}}.$$

Since $I_n \subset I_{n+1} \subset \ldots$, $\lim\limits_{n\to\infty} I_n = I^*$ exists and

$$m(I - I^*) = 0.$$

The sequence $\{f_{k_r}(t)\}_{r=1}^{\infty}$ converges uniformly on each set I_s. This follows from the inequality

$$|f_{k_n}(t) - f_{k_m}(t)| \leqq \sum_{r=m}^{n-1} |f_{k_{r+1}}(t) - f_{k_r}(t)| < \sum_{r=m}^{n-1} \frac{1}{2^r} < \frac{1}{2^{m-1}} \quad (n > m > s)$$

which is valid for $t \in I_s$. Consequently, $\{f_{k_r}(t)\}_{r=1}^{\infty}$ converges on I^* (i.e., almost everywhere in I). Let

$$f(t) = \begin{cases} \lim\limits_{r\to\infty} f_{k_r}(t) & (t \in I^*), \\ 0 & (t \in I - I^*). \end{cases}$$

Since $\{f_n\}$ was assumed to be a fundamental sequence, for each $\varepsilon > 0$ there exists an integer $N(\varepsilon)$ such that

$$\int_{I_s(a)} |f_m(t) - f_{k_r}(t)|^2 dt \leqq \|f_m - f_{k_r}\|^2 < \varepsilon$$

for $m, k_r > N(\varepsilon)$, where $I_s(\alpha) = I_s$ if the interval (a, b) is finite and $I_s(\alpha) = I_s \cap (-\alpha, \alpha)$ otherwise. In $I_s(\alpha)$, convergence of the sequence $\{f_{k_r}(t)\}_{r=1}^{\infty}$ is uniform. Therefore, in the integral, passage to the limit is legitimate and we obtain

$$\int_{I_s(\alpha)} |f_m(t) - f(t)|^2 dt \leqq \varepsilon \qquad (m > N(\varepsilon)).$$

It follows that

$$\int_{I_s} |f_m(t) - f(t)|^2 dt \leqq \varepsilon,$$

where s is arbitrary. Hence

$$\int_a^b |f_m(t) - f(t)|^2 dt \leqq \varepsilon.$$

This implies that $f_m - f \in L^2$, so that $f \in L^2$. Since $\varepsilon > 0$ is arbitrary, we have proved also that

$$\lim_{m \to \infty} \int_a^b |f_m(t) - f(t)|^2 dt = 0 .$$

In the process of the proof, we have obtained the following fact: if the sequence $f_n \in L^2$ $(n = 1, 2, 3, \ldots)$ converges to f in norm, i.e., if $\|f_n - f\| \to 0$ $(n \to \infty)$, then there exists a subsequence $\{f_{k_r}(t)\}_{r=1}^{\infty}$ which converges to $f(t)$ almost everywhere. Furthermore, if a proper set of arbitrarily small measure is removed from the interval (a, b), then in the remaining set the subsequence $\{f_{k_r}(t)\}_{r=1}^{\infty}$ converges to $f(t)$ uniformly.

We remark that it is possible to consider the space $L^2(a, b)$ as a subspace of $L^2(a_1, b_1)$ if $a_1 \leqq a < b \leqq b_1$ and, in particular, as a subspace of $L^2(-\infty, \infty)$. For this, it is necessary to extend each function $f \in L^2(a, b)$ beyond the limits of the interval (a, b) by defining $f(t)$ to be zero for t outside of (a, b).

Convergence in the metric space L^2 is called *convergence in the mean* and is denoted by

$$f(t) = \underset{n \to \infty}{\mathrm{l.i.m.}} f_n(t),$$

if

$$\lim_{n \to \infty} \int_a^b |f(t) - f_n(t)|^2 dt = 0$$

(l. i. m.) is an abbreviation for limes in medio, i.e. *limit in the mean*).

11. Complete Orthonormal Systems in L^2

In the present paragraph, we show that there exist complete orthonormal sequences in $L^2(a, b)$, where a and b are finite or infinite. Hence, by Theorem 2, Section 9, it will follow that the space L^2 is separable. It would be possible to prove this latter fact immediately. In fact, using the definition of the Lebesgue integral, it is not difficult to prove that the linear envelope of the set of functions f such that $f \equiv 1$ on some finite interval and $f \equiv 0$ outside, is complete in L^2. Hence, the separability of L^2 follows.

A. We begin with the space $L^2(0, 2\pi)$. In this space, the functions

$$\frac{1}{\sqrt{2\pi}} e^{ikt} \qquad (\pm k = 0, 1, 2, \ldots)$$

form an orthonormal system. We wish to show that this *trigonometric system* is complete. Assume there exists $f \in L^2(0, 2\pi)$ such that

$$\int_0^{2\pi} |f(t)|\, dt \neq 0$$

and

$$\int_0^{2\pi} f(t) e^{-ikt} dt = 0 \qquad (\pm k = 1, 2, 3, \ldots). \tag{1}$$

It follows by means of integration by parts that the function

$$F(t) = \int_0^t f(t)\, dt$$

satisfies the equations

$$\int_0^{2\pi} \{F(t) - C\} e^{-ikt}\, dt = 0 \qquad (\pm k = 1, 2, 3, \ldots) \tag{2}$$

for any constant C. We specify this constant so that equation (2) holds also for $k = 0$. Since the function

$$\Phi(t) = F(t) - C$$

is continuous, the well-known theorem of Weierstrass applies: for each $\varepsilon > 0$ there exists a trigonometric sum

$$\sigma(t) = \sum_{k=-n}^{n} A_k e^{ikt}$$

such that

$$|\Phi(t) - \sigma(t)| < \varepsilon.$$

Therefore, using relation (2), we obtain

$$\int_0^{2\pi} |\Phi(t)|^2 dt = \int_0^{2\pi} \overline{\Phi(t)}\{\Phi(t) - \sigma(t)\}\, dt \leqq$$

$$\leqq \varepsilon \int_0^{2\pi} |\Phi(t)|\, dt \leqq \varepsilon\sqrt{2\pi} \sqrt{\int_0^{2\pi} |\Phi(t)|^2 dt}$$

whence

$$\int_0^{2\pi} |\Phi(t)|^2 dt \leqq 2\pi\varepsilon^2.$$

Since $\varepsilon > 0$ is arbitrary, this implies that $\Phi(t) = 0$, so that $F(t) = C$ and $f(t) = 0$ almost everywhere. Thus, the completeness of the trigonometric system is proved.

B. We consider now the space $L^2(a, b)$, where (a, b) is an arbitrary finite interval. The orthogonalization of the sequence of functions

$$1, t, t^2, \ldots$$

yields the sequence of polynomials

$$C_k \frac{d^k\{(t-a)(t-b)\}^k}{dt^k} \qquad (k = 0, 1, 2, \ldots),$$

where C_k are certain positive constants. These are the well-known *Legendre polynomials.* They are usually considered for $a = -1$, $b = 1$. The completeness of this orthonormal system may be proved in the same way as the completeness of the trigonometric system.

C. We consider the space $L^2(-\infty, \infty)$. The orthogonalization of the system

$$e^{-\frac{t^2}{2}}, te^{-\frac{t^2}{2}}, t^2e^{-\frac{t^2}{2}}, \ldots$$

yields the sequence of *Tchebysheff-Hermite functions,*

$$\varphi_k(t) = (-1)^k e^{\frac{t^2}{2}} \frac{d^k e^{-t^2}}{dt^k} = H_k(t) e^{-\frac{t^2}{2}} \qquad (k = 0, 1, 2, \ldots),$$

where $H_k(t)$ is the so-called *Tchebysheff-Hermite polynomial* of degree k. The Tchebysheff-Hermite functions satisfy the relations

$$\int_{-\infty}^{\infty} \varphi_k(t)\varphi_m(t)\, dt = \begin{cases} 0 & (k \neq m), \\ 2^m m! \sqrt{\pi} & (k = m), \end{cases}$$

so that they are pairwise orthogonal but not normalized. We prove next that the sequence of Tchebysheff-Hermite functions is complete. Assume there exists a nonzero function $f \in L^2(-\infty, \infty)$ such that

$$\int_{-\infty}^{\infty} f(t)\,\varphi_k(t)\,dt = 0 \qquad (k = 0, 1, 2, \ldots)$$

or, equivalently, such that

$$\int_{-\infty}^{\infty} f(t)\,e^{-\frac{t^2}{2}} t^k\,dt = 0 \qquad (k = 0, 1, 2, \ldots). \tag{3}$$

We introduce the function

$$F(z) = \int_{-\infty}^{\infty} f(t)\,e^{-\frac{t^2}{2}} e^{itz}\,dt$$

which, it is evident, exists for every complex z. The function $F(z)$ has the finite derivative

$$F'(z) = \int_{-\infty}^{\infty} f(t)\,e^{-\frac{t^2}{2}} e^{itz}\,it\,dt.$$

Since this equation holds everywhere in the complex plane, $F(z)$ is an entire function. But, by (3)

$$F^{(k)}(0) \equiv \int_{-\infty}^{\infty} f(t)\,e^{-\frac{t^2}{2}} (it)^k\,dt = 0 \qquad (k = 0, 1, 2, \ldots)$$

so that $F(z)$ is identically zero. Therefore,

$$\int_{-\infty}^{\infty} f(t)\,e^{-\frac{t^2}{2}} e^{itx}\,dt = 0 \qquad (-\infty < x < \infty).$$

Multiplying this equality by e^{-ixy}, where y is real, and integrating with respect to x from $-\omega$ to ω, we get

$$\int_{-\infty}^{\infty} f(t)\,e^{-\frac{t^2}{2}} \frac{\sin \omega\,(t-y)}{t-y}\,dt = 0,$$

which is valid for every real y and ω. Hence, as is proved in analysis courses, it follows that $f(t) = 0$ almost everywhere, and this contradicts our original assumption.

D. In the space $L^2(0, \infty)$, we have the orthonormal system of *Tchebysheff-Laguerre functions*

$$\psi_k(t) = \frac{e^{-\frac{t}{2}} L_k(t)}{k!} \qquad (k = 0, 1, 2, \ldots),$$

where the $L_k(t)$ are the *Tchebysheff-Laguerre polynomials* which are defined by the formulas

$$L_k(t) = e^t \frac{d^k}{dt^k}(t^k e^{-t}) \qquad (k = 0, 1, 2, \ldots).$$

The completeness of this system can be proved by an argument based on the completeness of the Tchebysheff-Hermite system. We leave this to the reader.

12. The Space L^2_σ

Let a non-decreasing function of bounded variation $\sigma(t)\,(-\infty < t < \infty)$ be given. We assume that it is left-continuous:

$$\sigma(t-0) = \sigma(t).$$

Such a function is often called a distribution function. With the aid of the function $\sigma(t)$ it is possible to construct a measure analogous to the Lebesgue measure but differing from it in that the length $b-a$ of the interval $[a, b]$, $(a \leqq b)$[15] is replaced by the *σ-length* $\sigma(b+0)-\sigma(a)$. Some points may have σ-length different from zero (points of jumps of $\sigma(t)$) and some proper intervals may have σ-length equal to zero (intervals of constancy of $\sigma(t)$). The measure determined by the σ-length is called the *σ-measure*; the *σ-measurable functions* and the corresponding Lebesgue-Stieltjes integral are constructed from it.

We consider the linear space of all σ-measurable functions f for which the Lebesgue-Stieltjes integral

$$\int_{-\infty}^{\infty} |f(t)|^2 d\sigma(t)$$

exists, and metrize it by means of the metric generated by the scalar product

$$(f, g) = \int_{-\infty}^{\infty} f(t)\overline{g(t)}\, d\sigma(t).$$

This linear space is complete, so that it is a Hilbert space. It is denoted by L^2_σ. Special significance is possessed by *characteristic functions.* A characteristic function is equal to one in a certain finite interval of positive σ-length and is equal to zero outside of that interval. We do not exclude improper intervals here. The linear envelope of the set of all characteristic functions is dense in L^2_σ. Using this fact, it is easy to prove that L^2_σ is separable.

The case of a distribution function $\sigma(t)$ for which each integral

$$s_k = \int_{-\infty}^{\infty} t^k\, d\sigma(t) \qquad (k = 0, 1, 2, \ldots)$$

[15] For $a = b$, we get an *improper* interval, i.e., a point.

exists is of great interest. Because of mechanical considerations these integrals are called *moments* of the distribution function $\sigma(t)$. If $\sigma(t)$ has only a finite number of points of increase, then the Stieltjes integral becomes a finite sum. We do not consider this uninteresting case.

Orthogonalizing the sequence,

$$1, t, t^2, \ldots,$$

in L^2_σ, we get a sequence of polynomials $\{P_k(t)\}_{k=1}^{\infty}$ (where $P_k(t)$ is a polynomial of degree k), which satisfy the relations

$$\int_{\infty}^{\infty} P_k(t) P_m(t)\, d\sigma(t) = \begin{cases} 0 & (k \neq m), \\ 1 & (k = m). \end{cases}$$

This polynomial sequence is said to be orthonormal with respect to the distribution function $\sigma(t)$. If the function $\sigma(t)$ is absolutely continuous and if

$$\sigma'(t) = w(t)$$

then the orthonormality relations can be written in the form

$$\int_{-\infty}^{\infty} P_k(t) P_m(t) w(t)\, dt = \begin{cases} 0 & (k \neq m), \\ 1 & (k = m). \end{cases}$$

In this case, it is said that the polynomials $P_k(t)$ are orthonormal with respect to the *weight* $w(t)$. For instance, the Tchebysheff-Hermite polynomials are orthogonal (but not normalized) with respect to the weight e^{-t^2} $(-\infty < t < \infty)$.

If the interval of orthogonality is finite (i.e., if $\sigma(t)$ is constant for $t < a$ and for $t > b$), then the orthogonal polynomials $P_k(t)$ form a complete system. This is proved with the aid of the theorem of Weierstrass in exactly the same way as in the proof of the completeness in $L^2(a, b)$ of the sequence of Legendre polynomials.

If the interval of orthogonality is infinite, then the system of orthogonal polynomials $P_k(t)$ may fail to be complete. We restrict ourselves to one example, which is due to Hamburger.[16] Consider the interval $(0, \infty)$. The orthogonal polynomials on this interval with respect to the weight

$$w(t) = e^{-\frac{\pi\sqrt{t}}{\ln^2 t + \pi^2}}$$

is an incomplete system because the function

$$g(t) = e^{\frac{\ln t}{\ln^2 t + \pi^2}} \sin \frac{\sqrt{t}\ln t + \pi}{\ln^2 t + \pi^2}$$

[16] H. Hamburger [1].

satisfies the relations

$$\int_0^\infty t^k g(t) w(t)\, dt = 0 \qquad (k = 0, 1, 2, \ldots).$$

13. The Space of Almost Periodic Functions

We consider the set of all functions of the form $e^{i\lambda t}$ $(-\infty < t < \infty)$, where the parameter λ is real. We denote by L the linear envelope of this set, i.e., the collection of all "polynomials" of the form

$$\sum_k A_k e^{i\lambda_k t}.$$

Adding to L the limits of sequences of functions of L which are uniformly convergent on the entire real axis, we get a certain set B of continuous functions. As H. Bohr proved, a continuous function $f(t)$ defined on the real axis belongs to the collection B if and only if it is *almost periodic*, i.e., if for each $\varepsilon > 0$ there exists a real number $l = l(\varepsilon)$ such that in every interval of length l there is at least one number τ for which

$$|f(t+\tau) - f(t)| < \varepsilon \qquad (-\infty < t < \infty).$$

We can metrize the linear system L, by defining the scalar product of two polynomials

$$f(t) = \sum_{r=1}^{m} A_r e^{i\lambda_r t},$$

$$g(t) = \sum_{s=1}^{n} B_s e^{i\mu_s t}$$

as

$$(f, g) = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} f(t)\overline{g(t)}\, dt =$$

$$= \lim_{T\to\infty} \sum_{r,s=1}^{m,n} A_r \bar{B}_s \frac{1}{2T} \int_{-T}^{T} e^{it(\lambda_r - \mu_s)}\, dt = \sum_{r,s=1}^{m,n} \delta(\lambda_r, \mu_s) A_r \bar{B}_s$$

where

$$\delta(\lambda, \mu) = \begin{cases} 0 & (\lambda \neq \mu), \\ 1 & (\lambda = \mu). \end{cases}$$

When L is closed by means of the metric generated by this scalar product, we get a certain complete Hilbert space B^2 which contains B as a linear manifold. The space B^2 is not separable. This follows from the fact that in B^2 there exists a continuum of orthonormal vectors $e^{i\lambda t}$ $(-\infty < \lambda < \infty)$, whereas (see paragraph 9) every orthonormal system in a separable space contains a finite or countable number of vectors.

Chapter II

LINEAR FUNCTIONALS AND BOUNDED LINEAR OPERATORS

14. Point Functions

Two kinds of point functions are considered in elementary treatments of 3- and n-dimensional spaces: *scalar functions*, the values of which are (real or complex) numbers, and *vector functions*, which relate the points of a space to other points of the same or another space. In the present book we shall study point functions in Hilbert space. In correspondence with the indicated division of the functions of elementary analysis into scalar functions and vector functions, we introduce in H so-called *functionals* and *operators*. The appropriate definitions follow.

Let D denote a subset of the space H. A function Φ which relates to each point $f \in \mathrm{D}$ a definite complex number $\Phi(f)$ is called a *functional* in the space H with *domain* D. A function T which relates to each element $f \in \mathrm{D}$ a particular element $Tf = g \in \mathrm{H}$ is called an operator[1] in the space H with *domain* D. The set Δ, consisting of all $g = Tf$, where f runs through D, is called the *range* of T. Sometimes we shall denote the domain of a functional Φ by D_Φ and, correspondingly, the domain and range of an operator T by D_T and Δ_T respectively.

The identity operator, i.e., the operator which maps each vector into itself, we shall denote by E. The operator which maps every vector into zero, we shall denote by O.

If the operator T maps each pair of different elements of D into a pair of different elements of Δ, then T has an inverse operator, which maps the elements of Δ into elements of D. The inverse operator is denoted by the symbol T^{-1}, and $T^{-1}g = f$ if and only if $f = Tg$. Moreover,

$$\mathrm{D}_{T^{-1}} = \Delta_T, \qquad \Delta_{T^{-1}} = \mathrm{D}_T.$$

We shall consider two functionals (or operators) to be equivalent if their domains coincide and if for each element of their common domain, the values of these functionals (or operators) coincide.

[1] Sometimes it is necessary to consider also functions which map elements of the space H into elements of some other Hilbert space. These functions are also called operators. They will not be encountered often in this book.

If the domain D_T of the operator T contains the domain D_S of the operator S, i.e., if $D_S \subset D_T$, and if

$$Tf = Sf$$

for each $f \in D_S$, then T is called an *extension* of S and we write

$$S \subset T.$$

The concept of an extension of a functional is defined analogously.

Motivated by the notion of a continuous function, we make the following definition of the *continuity of an operator:* T is continuous at a point $f_0 \in D_T$ if

$$\lim_{f \to f_0} Tf = Tf_0 \qquad (f \in D_T).$$

An equivalent condition is that for each $\varepsilon > 0$, there exists $\delta = \delta(\varepsilon) > 0$ such that if f satisfies the inequality,

$$\|f - f_0\| < \delta, \quad f \in D_T,$$

then

$$\|Tf - Tf_0\| < \varepsilon.$$

The *continuity of a functional* is defined analogously.

If the element f_0 does not belong to D_T but $\lim Tf = g_0$ exists as $f \to f_0$ with $f \in D_T$, then the operator T can be defined for f_0, by letting $Tf_0 = g_0$. Proceeding in exactly the same way with all such elements f_0, we arrive at the so-called *extension by continuity of the operator T*. This extension is uniquely defined for each operator T. The *extension by continuity of a functional* is defined analogously.

We shall introduce below notation relating to so-called linear functionals and linear operators, which are the basic objects of our study.

Here we consider only the operator analogue of a "function of a function". Let S and T be two operators such that the range of T intersects the domain of S (i.e., let $\Delta_T \cap D_S \neq 0$). In this case we define the *product* ST of the operators S and T as the operator such that

$$STf = S(Tf)$$

for each element f in its domain, which is defined as the set of all $f \in D_T$ for which $Tf \in D_S$. The product TS is defined analogously whenever the set $\Delta_S \cap D_T$ is non-empty. It is clear that ST and TS are not generally equivalent because their domains may be different, and, moreover, even if g is an element belonging to both domains, it is possible that

$$STg \neq TSg.$$

Since it is difficult to give a reasonable completely general definition of the commutativity of two operators, we restrict ourselves here to the case in which at least one of the two operators S and T is defined everywhere in H. Let $D_S = H$. Then S and T are commutative if

$$ST \subset TS$$

i.e., if $f \in D_T$ implies both $Sf \in D_T$ and

$$STf = TSf.$$

In particular, if both operators are defined everywhere in H, then S and T are commutative if and only if $ST = TS$.

15. Linear Functionals

A functional Φ is said to be *linear* if:

(a) its domain D is a linear manifold and

$$\Phi(\alpha f + \beta g) = \alpha \Phi(f) + \beta \Phi(g)$$

for $f, g \in D$ and any complex numbers α and β;

(b) the inequality

$$\sup_{f \in D, \|f\| \leqq 1} |\Phi(f)| < \infty$$

is satisfied.

The left member of this inequality is called the *norm* of the functional Φ and is denoted by the symbol $\|\Phi\|_D$ or, if $D = H$, simply by $\|\Phi\|$.

If $f \in D$ and $f \neq 0$ then, by the definition of the norm of a functional,

$$\left|\Phi\left(\frac{f}{\|f\|}\right)\right| \leqq \|\Phi\|_D.$$

Hence, for $f \in D$,

(1) $$|\Phi(f)| \leqq \|\Phi\|_D \cdot \|f\|.$$

Relation (1) shows that the linear functional Φ is continuous. In fact, by (1),

$$|\Phi(f) - \Phi(f_0)| = |\Phi(f - f_0)| \leqq \|\Phi\|_D \cdot \|f - f_0\|$$

for $f, f_0 \in D$.

From (1) it also follows that, if $f \in D$ and $\|f\| \leqq 1$, then

$$|\Phi(f)| \leqq \|\Phi\|_D,$$

with strict inequality if $\|f\| < 1$. Therefore, the norm $\|\Phi\|_D$ can be defined by

(2) $$\sup_{f \in D, \|f\| = 1} |\Phi(f)| = \|\Phi\|_D,$$

or, equivalently, by[2]

(2′) $$\sup_{f \in D} \frac{|\Phi(f)|}{\|f\|} = \|\Phi\|_D.$$

[2] Translator's Note: The condition $f \neq 0$ should be included in (2′). Similar conditions should be added in many other relations which occur below. This task is left to the reader.

If Φ and Ψ are two linear functionals with domains of definition D_Φ and D_Ψ, then $\alpha\,\Phi + \beta\,\Psi$, where α and β are constants, is also a linear functional, with the intersection $D_\Phi \cap D_\Psi$ of the domains D_Φ and D_Ψ as domain (of course, only the case when $D_\Phi \cap D_\Psi$ contains elements different from $f = 0$ offers much interest).

If the functional Φ satisfies the condition (a) listed above, which states that Φ is *homogeneous and additive*, and if Φ is continuous at any one point $f_0 \in D$, then Φ also satisfies condition (b) above, i.e., Φ is bounded and, therefore, Φ is a linear functional. In fact, if Φ is continuous at f_0 then for each $\delta > 0$ there exists $\varepsilon > 0$ such that

$$|\Phi(h) - \Phi(f_0)| < \delta$$

for $\|h - f_0\| \leqq \varepsilon$ and $h \in D$. For each $f \in D$ such that $f \neq 0$,

$$\Phi(f) = \frac{\|f\|}{\varepsilon}\,\Phi\left(\frac{\varepsilon f}{\|f\|}\right) = \frac{\|f\|}{\varepsilon}\left\{\Phi\left(\frac{\varepsilon f}{\|f\|} + f_0\right) - \Phi(f_0)\right\}.$$

Since the vector $\dfrac{\varepsilon f}{\|f\|} + f_0 = h$ satisfies the relation $\|h - f_0\| = \varepsilon$ we have, for $f \in D$,

$$|\Phi(f)| < \frac{\delta}{\varepsilon}\|f\|;$$

in other words, for $f \in D$ and $f \neq 0$,

$$\frac{|\Phi(f)|}{\|f\|} < \frac{\delta}{\varepsilon}.$$

This proves that Φ is bounded.

If the linear manifold D on which the linear functional Φ is defined is not closed, then it is possible to extend Φ by continuity to the closure of D. This extension, as is easy to see, leads to a unique linear functional with the same norm as the initial functional.

16. The Theorem of F. Riesz

The following theorem of F. Riesz provides a representation for each linear functional in H.

THEOREM: *Each linear functional Φ in the Hilbert space* H *can be expressed in the form*

$$\Phi(h) = (h, f),$$

where f is an element of H *which is uniquely determined by the functional Φ; furthermore,*

$$\|\Phi\| = \|f\|.$$

Proof: We denote by G the set of all elements $g \in H$ for which

$$\Phi(g) = 0.$$

By the linearity of the functional Φ, the set G is a linear manifold. Furthermore, G is closed, so that G is a subspace. In fact, if $g_n \in G$, $n \geqq 1$, and $g_n \to g$ then, by the continuity of Φ,

$$\Phi(g) = \lim_{n\to\infty} \Phi(g_n),$$

so that $\Phi(g)=0$ and $g \in G$. If $G=H$, then the functional Φ is equal to zero everywhere, and the theorem of Riesz is proved by taking $f=0$. We now suppose that $G \neq H$. Then there exists a nonzero element $f_0 \in H \ominus G$. We consider elements of the form

$$\Phi(h)f_0 - \Phi(f_0)\,h,$$

where h runs through H. These elements belong to G because

$$\Phi\,[\Phi(h)f_0 - \Phi(f_0)\,h] = \Phi(h)\,\Phi(f_0) - \Phi(f_0)\,\Phi(h) = 0.$$

Since $f_0 \in H \ominus G$,

$$(\Phi(h)f_0 - \Phi(f_0)\,h, f_0) = 0$$

and

$$\Phi(h) = \left(h, \frac{\overline{\Phi(f_0)}}{(f_0,f_0)} f_0\right).$$

If we set

$$f = \frac{\overline{\Phi(f_0)}}{(f_0,f_0)} f_0,$$

then it follows from the equality just obtained that

$$\Phi(h) = (h,f).$$

This is the required representation of the functional Φ.

We now prove that f is unique. Assuming the contrary, we have the equation

$$(h,f') = (h,f''),$$

for $h \in H$, where $f' \neq f''$. But this is impossible since the substitution of $h = f' - f''$ yields the contradiction,

$$\|f' - f''\|^2 = 0.$$

It remains to be proved that

$$\|\Phi\| = \|f\|.$$

It follows from the equation

$$\Phi(h) = (h,f)$$

that

$$\|\Phi(h)\| \leqq \|f\| \cdot \|h\|$$

which yields

$$\|\Phi\| \leqq \|f\|.$$

On the other hand, taking $h = f$, we get

$$\Phi(f) = \|f\|^2,$$

whence it follows that

$$\| \Phi \| \geqq \|f\|.$$

Thus the theorem of F. Riesz is proved.

We consider now a linear functional Ψ with domain D_Ψ closed in H. Then D_Ψ is a subspace of H and the theorem of F. Riesz asserts the existence of a unique element $g \in D_\Psi$ such that

$$\Psi(h) = (h, g) \qquad (h \in D_\Psi) \tag{3}$$

and

$$\| \Psi \|_{D_\Psi} = \| g \|.$$

By means of (3), the linear functional Ψ may be extended to the whole space H without increasing the norm.[3] Any other extension of the linear functional Ψ to the whole space H increases the norm of the functional. In fact, if Φ is any extension of Ψ to the whole space, then

$$\Phi(h) = (h, f)$$

and

$$\| \Phi \| = \|f\|.$$

For $h \in D_\Psi$,

$$(h, g) = (h, f)$$

so that $f - g \perp D_\Psi$. Because $g \in D_\Psi$,

$$\|f\|^2 = \|g\|^2 + \|f - g\|^2$$

which implies that

$$\| \Phi \| \geqq \| \Psi \|_{D_\Psi},$$

where there is strict inequality if $f \neq g$.

17. A Criterion for the Closure in H of a Given System of Vectors

According to the definition in Section 8, a system M of vectors is closed in H if it is possible to approximate each $h \in H$ to any degree of accuracy by means of a linear combination of vectors belonging to M.

THEOREM: *In order that the system* M *be closed in* H, *it is necessary and sufficient that a linear functional* Φ *in* H *which vanishes for all* $g \in M$, *be identically equal to zero.*

Proof: The necessity is an immediate consequence of the continuity of the linear functional. In order to prove the sufficiency, let us assume that the system is not closed. Then there exists $\delta > 0$ and a vector $h_0 \in H$ for which

$$\inf_{n, \alpha_i} \| h_0 - \alpha_1 g_1 - \alpha_2 g_2 - \ldots - \alpha_n g_n \| = \delta > 0 \qquad (g_i \in M).$$

[3] Since any linear functional can be extended to the whole space without increasing the norm, one usually considers a linear functional as being defined on the whole space when the domain is not specified.