CHAPTER XIX

FUNCTIONAL ANALYSIS

The rise and spread of functional analysis in the 20th century had two main causes. On the one hand it became desirable to interpret from a uniform point of view the copious factual material accumulated in the course of the 19th century in various, often hardly connected, branches of mathematics. The fundamental concepts of functional analysis were formed and crystalized under various aspects and for various reasons. Many of the fundamental concepts of functional analysis emerged in a natural fashion in the process of development of the calculus of variations, in problems on oscillations (in the transition from the oscillations of systems with a finite number of degrees of freedom to oscillations of continuous media), in the theory of integral equations, in the theory of differential equations both ordinary and partial (in boundary problems, problems on eigenvalues, etc.) in the development of the theory of functions of a real variable, in operator calculus, in the discussion of problems in the theory of approximation of functions, and others. Functional analysis permitted an understanding of many results in these domains from a single point of view and often promoted the derivation of new ones. In recent decades the preparatory concepts and apparatus were then used in a new branch of theoretical physics—in quantum mechanics.

On the other hand, the investigation of mathematical problems connected with quantum mechanics became a crucial feature in the further development of functional analysis itself: It created, and still creates at the present time, fundamental branches of this development.

Functional analysis has not yet reached its completion by far. On the contrary, undoubtedly in its further development the questions and requirements of contemporary physics will have the same significance for it as classical mechanics had for the rise and development of the differential and integral calculus in the 18th century.

It is impossible here to include in this chapter all, or even only all the fundamental, problems of functional analysis. Many important branches exceed the limitations of this book. Nevertheless, by confining ourselves to certain selected problems, we wish to acquaint the reader with some fundamental concepts of functional analysis and to illustrate as far as possible the connections of which we have spoken here. These problems were analyzed mainly at the beginning of the 20th century on the basis of the classical papers of Hilbert, who was one of the founders of functional analysis. Since then functional analysis has developed very vigorously and has been widely applied in almost all branches of mathematics; in partial differential equations, in the theory of probability, in quantum mechanics, in the quantum theory of fields, etc. Unfortunately these further developments of functional analysis cannot be included in our account. In order to describe them we would have to write a separate large book, and therefore, we restrict ourselves to one of the oldest problems, namely the theory of eigenfunctions.

§1. n-Dimensional Space

In what follows we shall make use of the fundamental concepts on n-dimensional space. Although these concepts have been introduced in the chapters on linear algebra and on abstract spaces, we do not think it superfluous to repeat them in the form in which they will occur here. For scanning through this section it is sufficient that the reader should have a knowledge of the foundations of analytic geometry.

We know that in analytic geometry of three-dimensional space a point is given by a triplet of numbers (f_1, f_2, f_3), which are its coordinates. The distance of this point from the origin of coordinates is equal to $\sqrt{f_1^2+f_2^2+f_3^2}$. If we regard the point as the end of a vector leading to it from the origin of coordinates, then the length of the vector is also equal to $\sqrt{f_1^2+f_2^2+f_3^2}$. The cosine of the angle between nonzero vectors leading from the origin of coordinates to two distinct points $A(f_1, f_2, f_3)$ and $B(g_1, g_2, g_3)$ is defined by the formula

$$\cos\phi = \frac{f_1g_1+f_2g_2+f_3g_3}{\sqrt{f_1^2+f_2^2+f_3^2}\sqrt{g_1^2+g_2^2+g_3^2}}.$$

From trigonometry we know that $|\cos\phi| \leqslant 1$. Therefore we have the inequality

$$\frac{|f_1g_1+f_2g_2+f_3g_3|}{\sqrt{f_1^2+f_2^2+f_3^2}\sqrt{g_1^2+g_2^2+g_3^2}} \leqslant 1,$$

and hence always

$$(f_1g_1 + f_2g_2 + f_3g_3)^2 \leqslant (f_1^2 + f_2^2 + f_3^2)(g_1^2 + g_2^2 + g_3^2). \tag{1}$$

This last inequality has an algebraic character and is true for arbitrary six numbers (f_1, f_2, f_3) and (g_1, g_2, g_3), since any six numbers can be the the coordinates of two points of space. All the same, the inequality (1) was obtained from purely geometric considerations and is closely connected with geometry, and this enables us to give it an easily visualized meaning.

In the analytic formulation of a number of geometric relations, it often turns out that the corresponding facts remain true when the triplet of numbers is replaced by n numbers. For example, our inequality (1) can be generalized to $2n$ numbers $(f_1, f_2, \cdots, f_n)$ and $(g_1, g_2, \cdots, g_n)$. This means that for arbitrary $2n$ numbers $(f_1, f_2, \cdots, f_n)$ and $(g_1, g_2, \cdots, g_n)$ an inequality analogous to (1) is true, namely:

$$(f_1g_1 + f_2g_2 + \cdots + f_ng_n)^2 \leqslant (f_1^2 + f_2^2 + \cdots + f_n^2)(g_1^2 + g_2^2 + \cdots + g_n^2). \tag{1'}$$

This inequality, of which (1) is a special case, can be proved purely analytically.* In a similar way many other relations between triplets of numbers derived in analytic geometry can be generalized to n numbers. This connection of geometry with relations between numbers (numerical relations) for which the cited inequality is an example becomes particularly lucid when the concept of an n-dimensional space is introduced. This concept was introduced in Chapter XVI. We repeat it here briefly.

A collection of n numbers $(f_1, f_2, \cdots, f_n)$ is called a point or *vector* of n-dimensional space (we shall more often use the latter name). The vector $(f_1, f_2, \cdots, f_n)$ will from now on be abbreviated by the single letter f.

Just as in three-dimensional space on addition of vectors their components are added, so we define the sum of the vectors

$$f = \{f_1, f_2, \cdots, f_n\} \quad \text{and} \quad g = \{g_1, g_2, \cdots, g_n\}$$

as the vector $\{f_1 + g_1, f_2 + g_2, \cdots, f_n + g_n\}$ and we denote it by $f + g$.

The product of the vector $f = \{f_1, f_2, \cdots, f_n\}$ by the number λ is the vector $\lambda f = \{\lambda f_1, \lambda f_2, \cdots, \lambda f_n\}$.

The length of the vector $f = \{f_1, f_2, \cdots, f_n\}$, like the length of a vector in three-dimensional space, is defined as $\sqrt{f_1^2 + f_2^2 + \cdots + f_n^2}$.

* See Chapter XVI.

The angle ϕ between the two vectors $f = \{f_1, f_2, \cdots, f_n\}$ and $g = \{g_1, g_2, \cdots, g_n\}$ in n-dimensional space is given by its cosine in exactly the same way as the angle between vectors in three-dimensional space. For it is defined by the formula*

$$\cos\phi = \frac{f_1g_1 + f_2g_2 + \cdots + f_ng_n}{\sqrt{f_1^2 + f_2^2 + \cdots + f_n^2}\sqrt{g_1^2 + g_2^2 + \cdots + g_n^2}}. \tag{2}$$

The scalar product of two vectors is the name for the product of their lengths by the cosine of the angle between them. Thus, if $f = \{f_1, f_2, \cdots, f_n\}$ and $g = \{g_1, g_2, \cdots, g_n\}$, then since the lengths of the vectors are $\sqrt{f_1^2 + f_2^2 + \cdots + f_n^2}$ and $\sqrt{g_1^2 + g_2^2 + \cdots + g_n^2}$, respectively, their scalar product, which is denoted by (f, g), is given by the formula

$$(f, g) = f_1g_1 + f_2g_2 + \cdots + f_ng_n. \tag{3}$$

In particular, the condition of orthogonality (perpendicularity) of two vectors is the equation $\cos\phi = 0$; i.e., $(f, g) = 0$.

By means of the formula (3) the reader can verify that the scalar product in n-dimensional space has the following properties:

1. $(f, g) = (g, f)$.
2. $(\lambda f, g) = \lambda(f, g)$.
3. $(f, g_1 + g_2) = (f, g_1) + (f, g_2)$.
4. $(f, f) \geqslant 0$, and the equality sign holds for $f = 0$ only, i.e., when $f_1 = f_2 = \cdots = f_n = 0$.

The scalar product of a vector f with itself (f, f) is equal to the square of the length of f.

The scalar product is a very convenient tool in studying n-dimensional spaces. We shall not study here the geometry of an n-dimensional space but shall restrict ourselves to a single example.

As our example we choose the theorem of Pythagoras in n-dimensional space: The square of the hypotenuse is equal to the sum of the squares of the sides. For this purpose we give a proof of this theorem in the plane which is easily transferred to the case of an n-dimensional space.

Let f and g be two perpendicular vectors in a plane. We consider the right-angled triangle constructed on f and g (figure 1). The hypotenuse of this triangle is equal in length to the vector $f + g$. Let us write down in vector form the theorem of Pythagoras in our notation. Since the square of the length of a vector is equal to the scalar product of the vector with

---

* The fact that $|\cos\phi| \leqslant 1$ follows from the inequality (1').

itself, Pythagoras' theorem can be written in the language of scalar products as follows:

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

The proof immediately follows from the properties of the scalar product. In fact,

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

and the two middle summands are equal to zero owing to the orthogonality of f and g.

In this proof we have only used the definition of the length of a vector, the perpendicularity of vectors, and the properties of the scalar product. Therefore nothing changes in the proof when we assume that f and g are two orthogonal vectors of an n-dimensional space. And so Pythagoras' theorem is proved for a right-angled triangle in n-dimensional space.

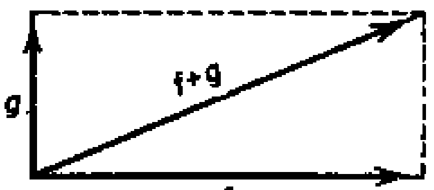

FIG. 1.

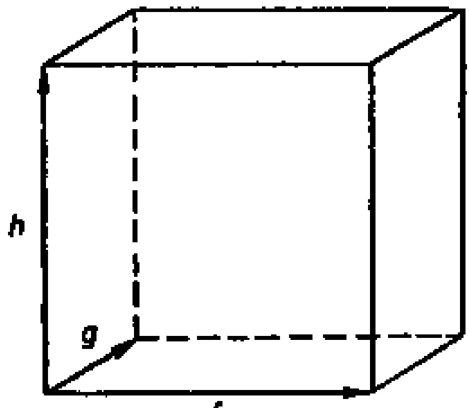

FIG. 2.

If three pairwise orthogonal vectors f, g and h are given in n-dimensional space, then their sum $f+g+h$ is the diagonal of the right-angled parallelepiped constructed from these vectors (figure 2) and we have the equation

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

which signifies that the square of the length of the diagonal of a parallelepiped is equal to the sum of the squares of the lengths of its edges. The proof of this statement, which is entirely analogous to the one given earlier for Pythagoras' theorem, is left to the reader. Similarly, if in an n-dimensional space there are k pairwise orthogonal vectors $f^1, f^2, \cdots, f^k$ then the equation

$$(f^1+f^2+\cdots+f^k, f^1+f^2+\cdots+f^k) = (f^1, f^1) + (f^2, f^2) + \cdots + (f^k, f^k), \qquad (4)$$

which is just as easy to prove, signifies that the square of the length of the diagonal of a "k-dimensional parallelelipiped" in n-dimensional space is also equal to the sum of the squares of the lengths of its edges.

§2. Hilbert Space (Infinite-Dimensional Space)

Connection with n-dimensional space. The introduction of the concept of n-dimensional space turned out to be useful in the study of a number of problems of mathematics and physics. In its turn this concept gave the impetus to a further development of the concept of space and to its application in various domains of mathematics. An important role in the development of linear algebra and of the geometry of n-dimensional spaces was played by problems of small oscillations of elastic systems.

FIG. 3.

Let us consider the following classical example of such a problem (figure 3). Let AB be a flexible string spanned between the points A and B. Let us assume that a weight is attached at a certain point C to the string. If it is moved from its position of equilibrium, it begins to oscillate with a certain frequency ω, which can be computed when we know the tension of the string, the mass m and the position of the weight. The state of the system at every instant is then given by a single number, namely the displacement y_1 of the mass m from the position of equilibrium of the string.

Now let us place n weights on the string AB at the points $C_1, C_2, \cdots, C_n$. The string itself is taken to be weightless. This means that its mass is so small that compared with the masses of the weights it can be neglected. The state of such a system is given by n numbers $y_1, y_2, \cdots, y_n$ equal to the displacements of the weights from the position of equilibrium. The collection of numbers $y_1, y_2, \cdots, y_n$ can be regarded (and this turns out to be useful in many respects) as a vector $(y_1, y_2, \cdots, y_n)$ of an n-dimensional space.

The investigation of the small oscillations that take place under these circumstances turns out to be closely connected with fundamental facts of the geometry of n-dimensional spaces. We can show, for example, that the determination of the frequency of the oscillations of such a system can be reduced to the task of finding the axes of a certain ellipsoid in n-dimensional space.

Now let us consider the problem of the small oscillations of a string spanned between the points A and B. Here we have in mind an idealized string, i.e., an elastic thread having a finite mass distributed continuously along the thread. In particular, by a homogeneous string we understand one whose density is constant.

Since the mass is distributed continuously along the string, the position of the string can no longer be given by a finite set of numbers $y_1, y_2, \cdots, y_n$, and instead the displacement $y(x)$ of every point x of the string has to be given. Thus, the state of the string at each instant is given by a certain function $y(x)$.

The state of a thread with n weights attached at the points with the abscissas $x_1, x_2, \cdots, x_n$, is represented graphically by a broken line with n members (figure 4), so that when the number of weights is increased, then the number of segments of the broken line increases correspondingly. When the number of weights grows without bound and the distance between adjacent weights tends to zero, we obtain in the limit a continuous distribution of mass along the thread, i.e., an idealized string. The broken line that describes the position of the thread with weights then goes over into a curve describing the position of the string (figure 5).

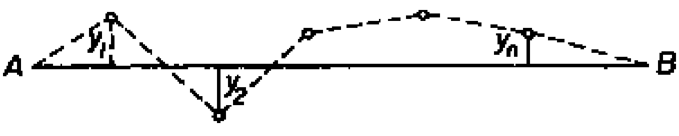

FIG. 4.

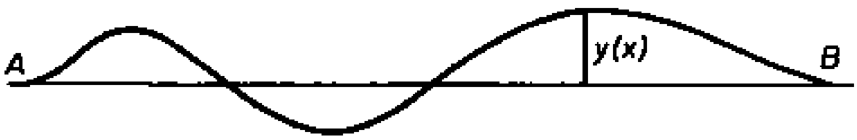

FIG. 5.

So we see that there exists a close connection between the oscillations of a thread with weights and the oscillations of a string. In the first problem the position of the system was given by a point or vector of an n-dimensional space. Therefore it is natural to regard the function $f(x)$ that describes the position of the oscillating string in the second case as a vector or a point of a certain infinite-dimensional space. A whole series of similar problems leads to the same idea of considering a space whose points (vectors) are functions $f(x)$ given on a certain interval.*

* As another such problem let us consider the electrical oscillations set up in a series of connected electrical circuits (figure 6).

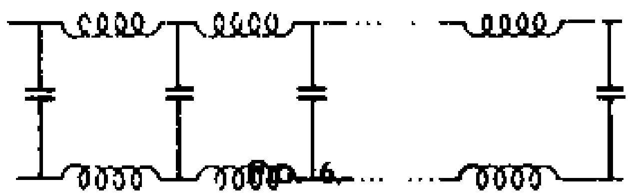

FIG. 6.

This example of oscillation of a string, to which we shall return again in §4, suggests to us how we shall have to introduce the fundamental concepts in an infinite-dimensional space.

Hilbert space. Here we shall discuss one of the most widespread concepts of an infinite-dimensional space of the greatest importance for the applications, namely the concept of the Hilbert space.

A vector of an n-dimensional space is defined as a collection of n numbers f_i, where i ranges from 1 to n. Similarly a vector of an infinite-dimensional space is defined as a function $f(x)$, where x ranges from a to b.

Addition of vectors and multiplication of a vector by a number is defined as addition of the functions and multiplication of the function by a number.

The length of a vector f in an n-dimensional space is defined by the formula

$$\sqrt{\sum_{i=1}^{n} f_i^2}$$

Since for functions the role of the sum is taken by the integral, the length of the vector $f(x)$ of a Hilbert space is given by the formula

$$\sqrt{\int_a^b f^2(x)\,dx}. \qquad (5)$$

The distance between the points f and g in an n-dimensional space is defined as the length of the vector $f - g$, i.e., as

$$\sqrt{\sum_{i=1}^{n} (f_i - g_i)^2}.$$

Similarly the "distance" between the elements $f(t)$ and $g(t)$ in a functional space is equal to

$$\sqrt{\int_a^b [f(t) - g(t)]^2\,dt}.$$

The state of such a series can be expressed by the set of n numbers $u_1, u_2, \cdots, u_n$, where u_i is the voltage on the condensor of the ith circuit of the chain. The collection of the n numbers $(u_1, \cdots, u_n)$ is a vector of an n-dimensional space.

Now let us imagine a two-wire line, i.e., a line consisting of two conductors having finite capacity and inductance, distributed along the line. The electric state of the line is expressed by a certain function $u(x)$, which gives the distribution of the voltage along the line. This function is a vector of the infinite-dimensional space of functions given on the interval (a, b).

The expression $\int_a^b [f(t) - g(t)]^2\,dt$ is called the mean-square deviation of the functions $f(t)$ and $g(t)$. Thus, the mean-square deviation of two elements of Hilbert space is taken to be a measure of their distance.

Let us now proceed to the definition of the angle between vectors. In an n-dimensional space the angle ϕ between the vectors $f = \{f_i\}$ and $g = \{g_i\}$ is defined by the formula

$$\cos\phi = \frac{\sum_{i=1}^n f_i g_i}{\sqrt{\sum_{i=1}^n f_i^2}\sqrt{\sum_{i=1}^n g_i^2}}.$$

In an infinite-dimensional space the sums are replaced by the corresponding integrals and the angle ϕ between the two vectors f and g of Hilbert space is defined by the analogous formula

$$\cos\phi = \frac{\int_a^b f(t)\,g(t)\,dt}{\sqrt{\int_a^b f^2(t)\,dt}\sqrt{\int_a^b g^2(t)\,dt}}. \tag{6}$$

This expression can be regarded as the cosine of a certain angle ϕ, provided the fraction on the right-hand side is an absolute value less than one, i.e., if

$$\left|\int_a^b f(t)\,g(t)\,dt\right| < \sqrt{\int_a^b f^2(t)\,dt}\sqrt{\int_a^b g^2(t)\,dt}. \tag{7}$$

This inequality in fact holds for two arbitrary functions $f(t)$ and $g(t)$. It plays an important role in analysis and is known as the Cauchy-Bunjakovskiĭ inequality. Let us prove it.

Let $f(x)$ and $g(x)$ be two functions, not identically equal to zero, given on the interval (a, b). We choose arbitrary numbers λ and μ and form the expresson

$$\int_a^b [\lambda f(x) - \mu g(x)]^2\,dx.$$

Since the function $[\lambda f(x) - \mu g(x)]^2$ under the integral sign is nonnegative, we have the following inequality

$$\int_a^b [\lambda f(x) - \mu g(x)]^2\,dx \geqslant 0;$$

i.e.,

$$2\lambda\mu\int_a^b f(x)\,g(x)\,dx \leqslant \lambda^2\int_a^b f^2(x)\,dx + \mu^2\int_a^b g^2(x)\,dx.$$

For brevity we introduce the notation

$$\left|\int_a^b f(x)\,g(x)\,dx\right| = C,\ \int_a^b f^2(x)\,dx = A,\ \int_a^b g^2(x)\,dx = B. \tag{8}$$

In this notation the inequality can be rewritten as follows:*

$$2\lambda\mu C \leqslant \lambda^2 A + \mu^2 B. \tag{9}$$

This inequality is valid for arbitrary values of λ and μ; in particular we may set

$$\lambda = \sqrt{\frac{C}{A}}, \mu = \sqrt{\frac{C}{B}}. \tag{10}$$

Substituting these values of λ and μ in (9), we obtain

$$\frac{C}{\sqrt{AB}} \leqslant 1.$$

When we replace A, B and C by their expressions in (8), we finally obtain the Cauchy-Bunjakovskiĭ inequality.

In geometry the scalar product of vectors is defined as the product of their lengths by the cosine of the angle between them. The lengths of the vectors f and g in our case are equal to

$$\sqrt{\int_a^b f^2(x)\,dx} \quad \text{and} \quad \sqrt{\int_a^b g^2(x)\,dx},$$

and the cosine of the angle between them is defined by the formula

$$\cos\phi = \frac{\int_a^b f(x)\,g(x)\,dx}{\sqrt{\int_a^b f^2(x)\,dx}\sqrt{\int_a^b g^2(x)\,dx}}.$$

When we multiply out these expressions, we arrive at the following formula for the scalar product of two vectors of Hilbert space:

$$(f, g) = \int_a^b f(x)\,g(x)\,dx. \tag{11}$$

From this formula it is clear that the scalar product of the vector f with itself its the square of its length.

If the scalar product of the nonzero vectors f and g is equal to zero, it means that $\cos\phi = 0$, i.e., that the angle ϕ ascribed to them by our definition is 90°. Therefore functions f and g for which

$$(f, g) = \int_a^b f(x)\,g(x)\,dx = 0,$$

are called orthogonal.

Pythagoras' theorem (see §1) holds in Hilbert space as in an n-dimen-

* For C we have to take the modulus of the integral because of the arbitrary sign of λ or μ.

sional space. Let $f_1(x), f_2(x), \cdots, f_N(x)$ be N pairwise orthogonal functions

$$f(x) = f_1(x) + f_2(x) + \cdots + f_N(x)$$

and their sum. Then the square of the length of f is equal to the sum of the squares of the lengths of $f_1, f_2, \cdots, f_N$.

Since the lengths of vectors in Hilbert space are given by means of integrals, Pythagoras' theorem in this case is expressed by the formula

$$\int_a^b f^2(x)\,dx = \int_a^b f_1^2(x)\,dx + \int_a^b f_2^2(x)\,dx + \cdots + \int_a^b f_N^2(x)\,dx. \tag{12}$$

The proof of this theorem does not differ in any respect from the one given previously (§1) for the same theorem in n-dimensional space.

So far we have not made precise what functions are to be regarded as vectors in Hilbert space. For such functions we have to take all those for which $\int_a^b f^2(x)\,dx$ has a meaning. It might appear natural to confine ourselves to continuous functions for which $\int_a^b f^2(x)\,dx$ always exists. However, the theory of Hilbert space becomes more complete and natural if $\int_a^b f^2(x)\,dx$ is interpreted in a generalized sense, namely as a Lebesgue integral (see Chapter XV).

This extension of the concept of integrals (and correspondingly of the class of functions to be discussed) is necessary for functional analysis in the same way as a strict theory of the real numbers is necessary for the foundation of the differential and integral calculus. Thus, the generalization of the ordinary concept of an integral that was created at the beginning of the 20th century in connection with the development of the theory of functions of a real variable turned out to be quite essential for functional analysis and the branches of mathematics connected with it.

§3. Expansion by Orthogonal Systems of Functions

Definition and examples of orthogonal systems of functions. If in a plane two arbitrary mutually perpendicular vectors e_1 and e_2 of unit length are chosen (figure 7), then every vector of the same plane can be decomposed in the directions of these two vectors, i.e., can be represented in the form

$$f = a_1 e_1 + a_2 e_2,$$

where a_1 and a_2 are the numbers equal to the projections of the vector f in the direction of the axis of e_1 and e_2. Since

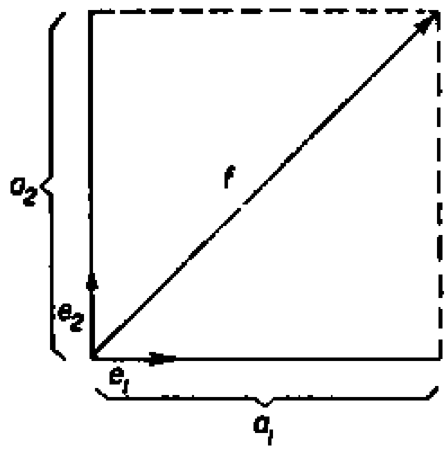

FIG. 7.

the projection of f on an axis is equal to the product of the length of f by the cosine of the angle between f and the axis, we can write, remembering the definition of the scalar product,

$$a_1 = (f, e_1),$$
$$a_2 = (f, e_2).$$

Similarly if in a three-dimensional space any three mutually perpendicular vectors e_1, e_2, e_3 of unit length are chosen, then every vector f in this space can be written in the form

$$f = a_1e_1 + a_2e_2 + a_3e_3,$$

where

$$a_k = (f, e_k) \ (k = 1, 2, 3).$$

In Hilbert space we can also consider systems of pairwise orthogonal vectors of the space, i.e., functions $\phi_1(x), \phi_2(x), \cdots, \phi_n(x), \cdots$.

Such systems of functions are called orthogonal and play an important role in analysis. They occur in very diverse problems of mathematical physics, integral equations, approximate computations, the theory of functions of a real variable, etc. The ordering and unification of the concepts relating to such systems formed one of the motivations that led at the beginning of the 20th century to the creation of the general concept of a Hilbert space.

Let us give a precise definition. A system of functions

$$\phi_1(x), \phi_2(x), \cdots, \phi_n(x), \cdots$$

is called *orthogonal* if any two functions of the system are orthogonal, i.e., if

$$\int_a^b \phi_i(x)\,\phi_k(x)\,dx = 0 \quad \text{for} \quad i \neq k. \tag{13}$$

In three-dimensional space we required that the vectors of the system should be of unit length. Recalling the definition of length of a vector we see that in the case of Hilbert space this requirement can be written as follows:

$$\int_a^b \phi_k^2(x)\,dx = 1. \tag{14}$$

A system of functions satisfying the conditions (13) and (14) is called orthonormal.

Let us give examples of such systems of functions.

1. On the interval $(-\pi, \pi)$ we consider the sequence of functions

$$1, \cos x, \sin x, \cos 2x, \sin 2x, \cdots, \cos nx, \sin nx, \cdots.$$

Any two functions of this sequence are orthogonal to each other. This can be verified by the simple computation of the corresponding integrals. The square of the length of a vector in Hilbert space is the integral of the square of the function. Thus, the squares of the lengths of the vectors of the sequence

$$1, \cos x, \sin x, \cos 2x, \sin 2x, \cdots, \cos nx, \sin nx, \cdots$$

are the integrals

$$\int_{-\pi}^{\pi} dx = 2\pi, \int_{-\pi}^{\pi} \cos^2 nx\, dx = \pi, \int_{-\pi}^{\pi} \sin^2 nx\, dx = \pi,$$

i.e., the vectors of our sequence are orthogonal, but not normalized. The length of the first vector of the sequence is equal to $\sqrt{2\pi}$, and all the others are of length $\sqrt{\pi}$. When we divide every vector by its length, we obtain the orthonormal system of trigonometric functions

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \cdots, \frac{\cos nx}{\sqrt{\pi}}, \frac{\sin nx}{\sqrt{\pi}}, \cdots.$$

This system is historically one of the first and most important examples of orthogonal systems. It appeared in the works of Euler, D. Bernoulli, and d'Alembert in connection with problems on the oscillations of strings. The study of it plays an essential role in the development of the whole of analysis.*

The appearance of the orthogonal system of trigonometrical functions in connection with problems on oscillations of strings is not accidental. Every problem on small oscillations of a medium leads to a certain system of orthogonal functions that describe the so-called characteristic oscillations of the given system (see §4). For example, in connection with problems on the oscillations of a sphere there appear the so-called spherical functions, in connection with problems on the oscillations of a circular membrane or a cylinder there appear the so-called cylinder functions, etc.

2. We can give an example of an orthogonal system of functions in

---

* See Chapter XII, §1.

which every function is a polynomial. Such an example is the sequence of Legendre polynomials

$$P_n(x) = \frac{1}{2^n n!} \frac{d^n(x^2 - 1)^n}{dx^n},$$

i.e., $P_n(x)$ is (apart from a constant factor) the nth derivative of $(x^2 - 1)^n$. Let us write down the first few polynomials of this sequence:

$$P_0(x) = 1;$$
$$P_1(x) = x;$$
$$P_2(x) = \tfrac{1}{2}(3x^2 - 1);$$
$$P_3(x) = \tfrac{1}{2}(5x^3 - 3x).$$

Obviously $P_n(x)$ is a polynomial of degree n. We leave it to the reader to convince himself that these polynomials are an orthogonal sequence on the interval $(-1, 1)$.

The general theory of orthogonal polynomials (the so-called orthogonal polynomials with weights) was developed in the second half of the 19th century by the famous Russian mathematician P. L. Čebyšev.

Expansion by orthogonal systems of functions. Just as in three-dimensional space every vector can be represented in the form of a linear combination of three pairwise orthogonal vectors e_1, e_2, e_3 of unit length

$$f = a_1 e_1 + a_2 e_2 + a_3 e_3,$$

so in a functional space there arises the problem of the decomposition of an arbitrary function f in a series with respect to an orthonormal system of functions, i.e., of the representation of f in the form

$$f(x) = a_1\phi_1(x) + a_2\phi_2(x) + \cdots + a_n\phi_n(x) + \cdots. \quad (15)$$

Here the convergence of the series (15) to the function f has to be understood in the sense of the distance between elements in Hilbert space. This means that the mean-square deviation of the partial sum of the series

$$S_n(t) = \sum_{k=1}^{n} a_k \phi_k(t)$$

from the function $f(t)$ tends to zero for $n \to \infty$; i.e.,

$$\lim_{n\to\infty} \int_a^b [f(t) - S_n(t)]^2 \, dt = 0. \quad (16)$$

This convergence is usually called "convergence in mean."

Expansions in various systems of orthogonal functions often occur in analysis and are an important method for the solution of problems of mathematical physics. For example, if the orthogonal system is the system of trigonometric functions on the interval $(-\pi, \pi)$

$$1, \cos x, \sin x, \cos 2x, \sin 2x, \cdots, \cos nx, \sin nx, \cdots,$$

then this expansion is the classical expansion of a function in a trigonometric series*

$$f(x) = a_0 + a_1 \cos x + b_1 \sin x + a_2 \cos 2x + b_2 \sin 2x + \cdots.$$

Let us assume that an expansion (15) is possible for every function f of a Hilbert space and let us find its coefficients a_n. For this purpose we multiply both sides of the equation scalarly by one and the same function ϕ_m of our system. We obtain the equation

$$(f, \phi_m) = a_1 (\phi_1, \phi_m) + a_2(\phi_2, \phi_m) + \cdots + a_m(\phi_m, \phi_m) + a_{m+1}(\phi_{m+1}, \phi_m) + \cdots,$$

in virtue of the fact that $(\phi_m, \phi_n) = 0$ for $m \neq n$ and $(\phi_m, \phi_m) = 1$, this determines the value of the coefficient a_m

$$a_m = (f, \phi_m) \quad (m = 1, 2, \cdots).$$

We see that, as in ordinary three-dimensional space (see the beginning of this section), the coefficients a_m are equal to the projections of the vector f in the direction of the vectors ϕ_k.

Recalling the definition of the scalar product we see that the coefficients of the expansion of $f(x)$ by the normal orthogonal system of functions $\phi_1(x), \phi_2(x), \cdots, \phi_n(x), \cdots$

$$f(x) = a_1\phi_1(x) + a_2\phi_2(x) + \cdots + a_n\phi_n(x) + \cdots \tag{17}$$

are determined by the formulas

$$a_m = \int_a^b f(t)\, \phi_m(t)\, dt. \tag{18}$$

As an example let us consider the normal orthogonal trigonometric system of functions mentioned previously:

$$\frac{1}{\sqrt{2\pi}}, \frac{\cos x}{\sqrt{\pi}}, \frac{\sin x}{\sqrt{\pi}}, \frac{\cos 2x}{\sqrt{\pi}}, \frac{\sin 2x}{\sqrt{\pi}}, \cdots.$$

* Such a decomposition often occurs in various problems of physics in the decomposition of an oscillation into its harmonic constituents. See Chapter VI, §5.

Then

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} a_n \cos nx + b_n \sin nx,$$

where

$$a_0 = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x)\, dx, \quad a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos nx\, dx,$$

$$b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin nx\, dx.$$

So we have obtained the formula for the computation of the coefficients of the expansion of a function in trigonometric series, assuming of course that this expansion is possible.*

We have established the form of the coefficients of the expansion (18) of the function $f(x)$ by an orthogonal system of functions under the assumptions that this expansion holds. However, an infinite orthogonal system of functions $\phi_1, \phi_2, \cdots, \phi_n, \cdots$ may turn out to be insufficient for every function of a Hilbert space to have such an expansion. For such an expansion to be possible, the system of orthogonal functions must satisfy an additional condition, namely the so-called condition of completeness.

An orthogonal system of functions is called *complete* if it is impossible to add to it even one function, not identically equal to zero, that is orthogonal to all the functions of the system.

It is easy to give an example of an incomplete orthogonal system. For this purpose we choose an arbitrary orthogonal system, for example that of the trigonometric functions, and remove one of the functions of the system, for example $\cos x$. The remaining infinite system of functions

$$1, \sin x, \cos 2x, \sin 2x, \cdots, \cos nx, \sin nx, \cdots$$

is orthogonal as before, but of course it is not complete, since the function $\cos x$ which we have excluded is orthogonal to all the functions of the system.

If a system of functions is incomplete, then not every function of a Hilbert space can be expanded by it. For if we attempt to expand by such a system a nonzero function $f_0(x)$ that is orthogonal to all the functions of the system, then by (18) all the coefficients turn out to be zero, whereas the function $f_0(x)$ is not equal to zero.

The following theorem holds: If a complete orthonormal system of

* On trigonometric series see also Chapter XII, §7.

functions in a Hilbert space $\phi_1(x), \phi_2(x), \cdots, \phi_n(x), \cdots$, is given, then every function $f(x)$ can be expanded in a series by functions of this system*

$$f(x) = a_1\phi_1(x) + a_2\phi_2(x) + \cdots + a_n\phi_n(x) + \cdots.$$

Here the coefficients a_n of the expansion are equal to the projections of the vectors f on the elements of the normal orthogonal system

$$a_n = (f, \phi_n) = \int_a^b f(x)\,\phi_n(x)\,dx.$$

Pythagoras' theorem in Hilbert space, which was established in §2, enables us to find an interesting relation between the coefficients a_k and the function $f(x)$. We denote by $r_n(x)$ the difference between $f(x)$ and the sum of the first n terms of its series; i.e.,

$$r_n(x) = f(x) - [a_1\phi_1(x) + \cdots + a_n\phi_n(x)].$$

The function $r_n(x)$ is orthogonal to $\phi_1(x), \phi_2(x), \cdots, \phi_n(x)$. Let us verify for example that it is orthogonal to $\phi_1(x)$, i.e., that $\int_a^b r_n(x)\,\phi_1(x)\,dx = 0$. We have

$$\int_a^b r_n(x)\,\phi_1(x)\,dx = \int_a^b [f(x) - a_1\phi_1(x) - a_2\phi_2(x) - \cdots - a_n\phi_n(x)]\,\phi_1(x)\,dx$$
$$= \int_a^b f(x)\,\phi_1(x)\,dx - a_1\int_a^b \phi_1^2(x)\,dx.\dagger$$

Since $a_1 = \int_a^b f(x)\,\phi_1(x)\,dx$, and $\int_a^b \phi_1^2(x)\,dx = 1$, it follows from this that $\int_a^b r_n(x)\phi_1(x)\,dx = 0$.

Thus, in the equation

$$f(x) = a_1\phi_1(x) + a_2\phi_2(x) + \cdots + a_n\phi_n(x) + r_n(x) \tag{19}$$

the individual terms on the right-hand side are orthogonal to each other. Hence, by Pythagoras' theorem as formulated in §1, the square of the length of $f(x)$ is equal to the sum of the squares of the lengths of the summands of the right-hand side in (19); i.e.,

$$\int_a^b f^2(x)\,dx = \int_a^b [a_1\phi_1(x)]^2\,dx + \cdots + \int_a^b [a_n\phi_n(x)]^2\,dx + \int_a^b r_n^2(x)\,dx.$$

* This series is related to its sum in the sense defined in formula (16).

† The remaining integrals are equal to zero, because the functions $\phi_k(x)$ are orthogonal to each other.

Since the system of functions $\phi_1, \phi_2, \cdots, \phi_n$ is normalized [equation (14)], we have

$$\int_a^b f^2(x)\,dx = a_1^2 + a_2^2 + \cdots + a_n^2 + \int_a^b r_n^2(x)\,dx. \tag{20}$$

The series $\sum_{k=1}^{\infty} a_k\phi_k(x)$ converges in mean. This means that

$$\int_a^b [f(x) - a_1\phi_1(x) - \cdots - a_n\phi_n(x)]^2\,dx \to 0,$$

i.e., that

$$\int_a^b r_n^2(x)\,dx \to 0.$$

But then we obtain from the formula (20) the equation

$$\sum_{k=1}^{\infty} a_k^2 = \int_a^b f^2(x)\,dx,^* \tag{21}$$

which states that the integral of the square of a function is equal to the sum of the squares of the coefficients of its expansion by a closed orthogonal system of functions. If the condition (21) holds for an arbitrary function of the Hilbert space, it is called the condition of completeness.

We wish to draw attention to the following important question. Which numbers a_k can be the coefficients of the expansion of a function in Hilbert space? The equation (21) asserts that for this purpose the series $\sum_{k=1}^{\infty} a_k^2$ must converge. Now it turns out that this condition is also sufficient; i.e., a sequence of numbers a_k is the sequence of coefficients of the expansion by an orthogonal system of functions in Hilbert space if and only if the series $\sum_{k=1}^{\infty} a_k^2$ converges.

We remark that this fundamental theorem holds if Hilbert space is interpreted as the collection of all functions with integrable square in the sense of Lebesgue (see §2). If we were to confine ourselves in Hilbert space, for example, to the continuous functions, then the solution of the problem as to which numbers a_k can be the coefficients of an expansion would become unnecessarily complicated.

The arguments given here are only one of the reasons that have led to the use of an integral in a generalized (Lebesgue) sense in the definition of Hilbert space.

---

* Geometrically, this means that the square of the length of a vector in Hilbert space is equal to the sum of the squares of its projections onto a complete system of mutually orthogonal directions.

§4. Integral Equations

In this section the reader will become acquainted with one of the most important and, historically, one of the first branches of functional analysis, namely the theory of integral equations, which has also played an essential role in the subsequent development of functional analysis. Quite apart from internal requirements of mathematics [for example, boundary problems for partial differential equations (Chapter VI)], various problems of physics were of great importance in the development of the theory of integral equations. Side by side with differential equations, the integral equations are, in the 20th century, one of the most important means of the mathematical investigation of various problems of physics. In this section we shall give a certain amount of information concerning the theory of integral equations. The facts we shall explain here are closely connected and have essentially sprung up (directly or indirectly) in connection with the study of small oscillations of elastic systems.

The problem of small oscillations of elastic systems. We return to the problem of small oscillations discussed in §2. Let us find equations that describe such oscillations. For the sake of simplicity we assume that we are dealing with the oscillation of a linear elastic system. As examples of such systems we can take, say, a string of length l (figure 8) or an elastic rod (figure 9). We shall assume that in the position of equilibrium our elastic system is situated along the segment Ol of the x-axis. We apply a unit force at the point x. Under the action of this force all the points of the system receive a certain displacement. The displacement arising at the point y (figure 8) is denoted by $k(x, y)$.

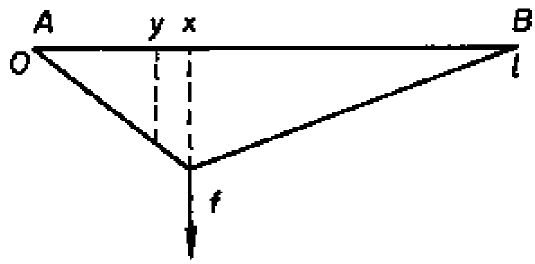

FIG. 8.

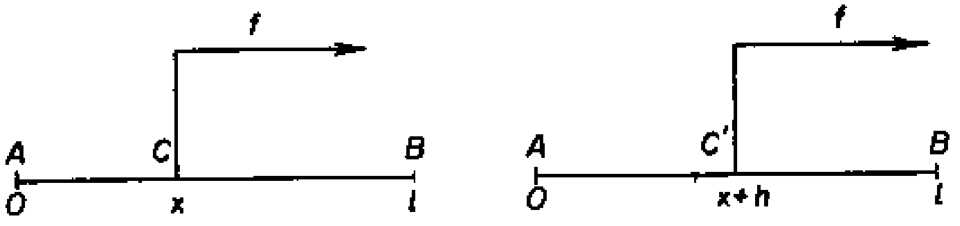

FIG. 9.

The function $k(x, y)$ is a function of two points: the point x at which the force is applied, and the point y at which we measure the displacement. It is called the influence function (Green's function).

From the law of conservation of energy, we can deduce an important property of the Green's function $k(x, y)$, namely the so-called reciprocity law: The displacement arising at the point y under the action of a force applied at the point x is equal to the displacement arising at the point x under the action of the same force applied at the point y. In other words, this means that

$$k(x, y) = k(y, x). \tag{22}$$

Let us find, for example, the Green's function for the longitudinal oscillations of an elastic rod (in figure 8 we have illustrated other transverse displacements). We consider a rod AB of length fixed at the ends (figure 9). At the point C we apply a force f acting in the direction of B. Under the action of this force the rod is deformed and the point C is shifted into the position C'. We denote the magnitude of the shift of C by h. Let us find the value of h. By means of h we can then find the shift at an arbitrary point y. For this purpose we shall make use of Hooke's law, which states that the force is proportional to the relative extension (i.e., to the ratio of the amount of displacement to the length). A similar relation holds for compressions.

Under the action of the force f the part AC of the rod is stretched. We denote the reaction arising here by T_1. At the same time the part CB of the rod is compressed, giving rise to a reaction T_2. By Hooke's law

$$T_1 = \kappa \frac{h}{x}, \quad T_2 = \kappa \frac{h}{l - x},$$

where κ is the coefficient of proportionality that characterizes the elastic properties of the rod. The position of equilibrium of the forces acting at the point C gives us

$$f = \kappa \frac{h}{x} + \kappa \frac{h}{l - x}, \quad \text{i.e.,} \quad f = \frac{\kappa l h}{x(l - x)}.$$

Hence

$$h = \frac{f}{\kappa l} x(l - x).$$

In order to find the displacement arising at a certain point y on the segment AC, i.e., for $y < x$, we note that it follows from Hooke's law that under an extension of the rod the relative extension (i.e., the ratio of the displacement of the point to its distance from the fixed end) does not depend on the position of the point. We denote the displacement of the point y by k.

Then by comparing the relative displacements at the points x and y we obtain

$$\frac{k}{y} = \frac{h}{x};$$

hence

$$k = h\frac{y}{x} = \frac{f}{\kappa l}y(l - x) \quad \text{for} \quad y < x.$$

Similarly, if the point lies on the segment CB ($y > x$), we obtain

$$k = h\frac{l - y}{l - x} = \frac{f}{\kappa l}x(l - y).$$

Bearing in mind that the Green's function $k(x, y)$ is the displacement at the point y under the action of a unit force applied at the point x, we obtain that on the longitudinal oscillations of an elastic rod the Green's function has the form

$$k(x, y) = \begin{cases} \dfrac{1}{\kappa l}y(l - x) & \text{for} \quad y < x, \\ \dfrac{1}{\kappa l}x(l - y) & \text{for} \quad y > x. \end{cases}$$

In a more or less similar way we could have found the Green's function for a string. If the tension of the string is T and the length l, then under the action of a unit force applied at the point x the string assumes the form illustrated in figure 8, and the displacement $k(x, y)$ at the point y is given by the formula

$$k(x, y) = \begin{cases} \dfrac{1}{Tl}x(l - y), & \text{for} \quad x < y, \\ \dfrac{1}{Tl}y(l - x), & \text{for} \quad x > y, \end{cases}$$

which coincides with the Green's function for the rod which we have derived.

In terms of the Green's function we can express the displacement of the system from its position of equilibrium provided that it is acted upon by a continuously distributed force of density $f(y)$. Since on an interval of length Δy there acts a force $f(y)\,\Delta y$, which we can regard approximately as concentrated at the point y, under the action of this force at the point x there arises a displacement $k(x, y)f(y)\,\Delta y$. The displacement undeı the action of the whole load is approximately equal to the sum

$$\sum k(x, y)f(y)\,\Delta y.$$

Passing to the limit for $\Delta y \to 0$ we obtain that the displacement $u(x)$ at the point x under the action of the force $f(y)$ distributed along the system is given by the formula

$$u(x) = \int_a^b k(x, y) f(y)\, dy. \tag{23}$$

Let us assume that our elastic system is not subject to the action of external forces. If it is displaced from its position of equilibrium, it then begins to move. These motions are called the free oscillations of the system.

Now let us write down in terms of the Green's function $k(x, y)$ the equation that the free oscillations of the elastic system in question have to obey. For this purpose we denote by $u(x, t)$ the displacement from the position of equilibrium at the point x and the instant of time t. Then the acceleration of x at the time t is equal to $\partial^2 u(x, t)/\partial t^2$.

If ρ is the linear density of the field, i.e., $\rho\, dy$ the mass of the element of length dy, then we obtain by a fundamental law of mechanics the equation of motion by replacing in (23) the force $f(y)\, dy$ by the product of the mass and the acceleration $[\partial^2 u(y, t)/\partial t^2]\, \rho\, dy$ taken with the opposite sign.

Thus, the equation of the free oscillations has the form

$$u(x, t) = -\int_a^b k(x, y) \frac{\partial^2 u(y, t)}{\partial t^2} \rho\, dy.$$

An important role in the theory of oscillations is played by the so-called harmonic oscillations of the elastic system, i.e., the motions for which

$$u(x, t) = u(x) \sin \omega t.$$

They are characterized by the fact that every fixed point performs harmonic oscillations (moves according to a sinusoidal law) with a certain frequency ω, and that this frequency is one and the same for all the points x.

Later on we shall see that every free oscillation is composed of harmonic oscillations.

We set

$$u(x, t) = u(x) \sin \omega t$$

in the equation of the free oscillations and cancel $\sin \omega t$. Then we obtain the following equation to determine the function $u(x)$

$$u(x) = \rho\omega^2 \int_a^b k(x, y)\, u(y)\, dy. \tag{24}$$

Such an equation is called a homogeneous integral equation for the function $u(x)$.

Obviously the equation (24) has for every ω the uninteresting solution $u(x) \equiv 0$, which corresponds to the state of rest. Those values of ω for which there exist other solutions of the equation (24), different from zero, are called the eigenfrequencies of the system.

Since nonzero solutions do not exist for every value of ω, the system can perform free oscillations only with definite frequencies. The smallest of these is called the fundamental tone of the system, and the remaining ones are overtones.

Now it turns out that for every system there exists an infinite sequence of eigenfrequencies, the so-called frequency spectrum

$$\omega_1, \omega_2, \cdots, \omega_n, \cdots$$

The nonzero solution $u_n(x)$ of the equation (24) corresponding to the the eigenfrequency ω_n gives us the form of the corresponding characteristic oscillation.

For example, if the elastic system is a string stretched between the points O and l and fastened at these points, then the possible frequencies of the characteristic oscillations of the system are equal to

$$a\frac{\pi}{l}, 2a\frac{\pi}{l}, 3a\frac{\pi}{l}, \cdots, na\frac{\pi}{l}, \cdots,$$

where a is a coefficient depending on the density and the tension of the

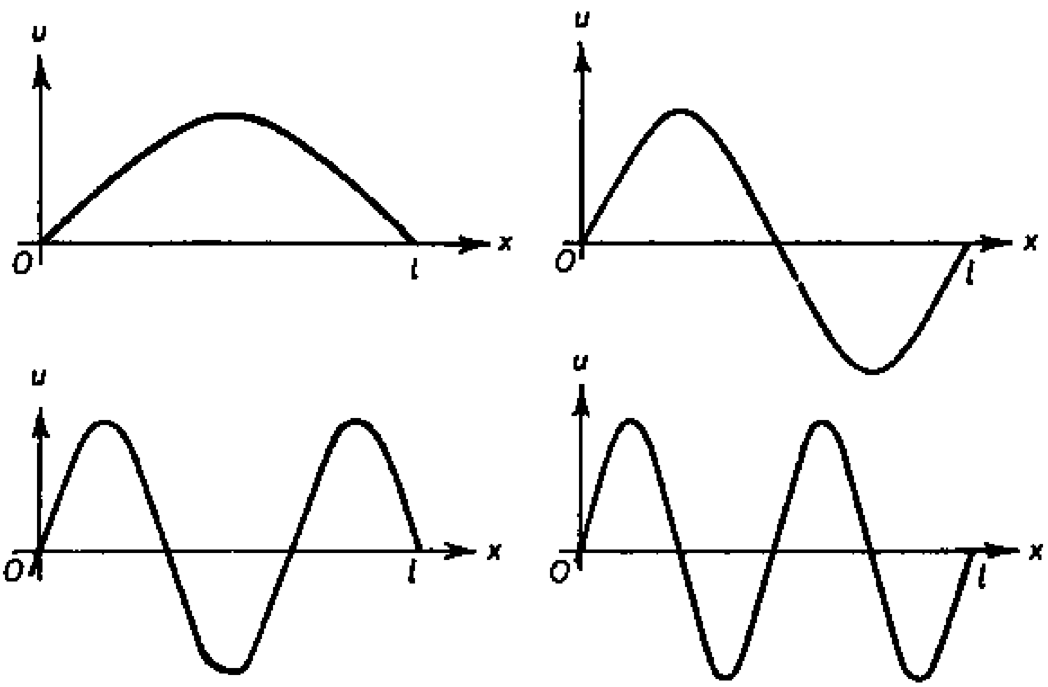

FIG. 10.

string, namely, $a = \sqrt{T/\rho}$. The fundamental tone is here $\omega_1 = a(\pi/l)$, and the overtones are $\omega_2 = 2\omega_1$, $\omega_3 = 3\omega_1$, $\cdots$, $\omega_n = n\omega_1$. The form of the corresponding harmonic oscillations is given by the equation

$$u_n(x) = \sin \frac{n\pi}{l} x$$

and are illustrated for $n = 1, 2, 3, 4$ in figure 10.

So far we have discussed free oscillations of elastic systems. Now if an exterior harmonic force acts on the elastic system during the motion, then, in determining the harmonic oscillations under the action of this force, we arrive at the function $u(x)$ at the so-called inhomogeneous integral equation

$$u(x) = \rho\omega^2 \int_a^b k(x, y)\, u(y)\, dy + h(x). \tag{25}$$

Properties of integral equations. Previously we have become acquainted with examples of integral equations

$$f(x) = \lambda \int_a^b k(x, y) f(y)\, dy \tag{26}$$

and

$$f(x) = \lambda \int_a^b k(x, y) f(y)\, dy + h(x), \tag{27}$$

the first of which was obtained in the solution of the problem on the free oscillations of an elastic system, and the second in the discussion of forced oscillations, i.e., oscillations under the action of external forces.

The unknown function in these equations is $f(x)$. The given function $k(x, y)$ is called the *kernel* of the integral equation. The equation (27) is called an *inhomogeneous linear integral equation*, and the equation (26) is *homogeneous*. It is obtained from the inhomogeneous one by setting $h(x) = 0$.

It is clear that the homogeneous equation always has the zero solution, i.e., the solution $f(x) = 0$. A close connection exists between the solutions, of the inhomogeneous and the homogeneous integral equations. By way of example we mention the following theorem: If the homogeneous integral equation has only the zero solution, then the corresponding inhomogeneous equation is soluble for every function $h(x)$.

If for a certain value λ a homogeneous equation has the solution $f(x)$, not identically equal to zero, then this value λ is called an *eigenvalue* and the corresponding solution $f(x)$ an *eigenfunction*. We have seen earlier that when an integral equation describes the free oscillations of an elastic system, then the eigenvalues are closely connected with the frequencies of the oscillations of the system (namely $\lambda = \rho\omega^2$). The eigenfunctions then give the form of the corresponding harmonic oscillations.

In the problems on oscillations it followed from the law of conservation of energy that

$$k(x, y) = k(y, x). \tag{28}$$

A kernel satisfying the condition (28) is called *symmetric*.

The eigenfunctions and eigenvalues of an equation with a symmetric kernel have a number of important properties. One can prove that such an equation always has a sequence of real eigenvalues

$$\lambda_1, \lambda_2, \cdots, \lambda_n, \cdots.$$

To every eigenvalue there correspond one or several eigenfunctions. Here eigenfunctions corresponding to distinct eigenvalues are always orthogonal to each other.*

Thus, for every integral equation with a symmetric kernel the system of eigenfunctions is an orthogonal system of functions. There arises the question of when this system is complete, i.e., when can every function of the Hilbert space be expanded in a series by a system of eigenfunctions of the integral equation. In particular, if the equation

$$\int_a^b k(x, y) f(y)\, dy = 0 \tag{29}$$

is satisfied for $f(y) \equiv 0$ only, then the system of eigenfunctions of the integral equation

$$\lambda \int_a^b k(x, y) f(y)\, dy = f(x)$$

is a complete orthogonal system.†

Thus, every function $f(x)$ with integrable square can in this case be expanded in a series by eigenfunctions. By discussing various types of integral equations, we obtain a general and powerful method of proving

* The latter statement will be proved in the next section.

† In the case when $k(x, y)$ is the Green's function of an elastic system, the equation (29) assumes a simple physical meaning. In fact [see formula (23)] we have seen that under the action of a force $f(y)$ distributed along the system the displacement of the system from the position of equilibrium is expressed by the formula $u(x) = \int_b^a k(x, y) f(y)\, dy$. Thus, the condition (29) signifies that every nonzero force takes the system out of its position of equilibrium.

that various important orthogonal systems are closed, i.e., that the functions are expandable in series by orthogonal functions. By this method we can prove the completeness of the system of trigonometric functions, of cylinder functions, spherical functions, and many other important systems of functions.

The fact that an arbitrary function can be expanded in a series by eigenfunctions means in the case of oscillations that every oscillation can be decomposed into a sum of harmonic oscillations. Such a decomposition yields a method that is widely applicable in solving problems on oscillations in various domains of mechanics and physics (oscillations of elastic bodies, acoustic oscillations, electromagnetic waves, etc.).

The development of the theory of linear integral equations gave the impetus to the creation of the general theory of linear operators of which the theory of linear integral equations forms an organic part. In the last few decades the general methods of the theory of linear operators have vigorously contributed to the further development of the theory of integral equations.

§5. Linear Operators and Further Developments of Functional Analysis

In the preceding section we have seen that problems on the oscillations of an elastic system lead to the search for the eigenvalues and eigenfunctions of integral equations. Let us note that these problems can also be reduced to the investigation of the eigenvalues and eigenfunctions of linear differential equations.* Many other physical problems also lead to the task of computing the eigenvalues and eigenfunctions of linear differential or integral equations.

Let us give one more example. In modern radio technology the so-called wave guides are widely used for the transmission of electromagnetic oscillations of high frequencies, i.e., hollow metallic tubes in which electromagnetic waves are propagated. It is known that in a wave guide only electromagnetic oscillations of not too large a wave length can be propagated. The search for the critical wave length amounts to a problem on the eigenvalues of a certain differential equation.

Problems on eigenvalues occur, moreover, in linear algebra, in the theory of ordinary differential equations, in questions of stability, etc.

So it became necessary to discuss all these related problems from one single point of view. This common point of view is the general theory of linear operators. Many problems on eigenfunctions and eigenvalues in various concrete cases came to be fully understood only in the light of

* See Chapter VI, §5.

the general theory of operators. Thus, in this and a number of other directions the general theory of operators turned out to be a very fruitful research tool in those domains of mathematics in which it is applicable.

In the subsequent development of the theory of operators, quantum mechanics played a very important role, since it makes extensive use of the methods of the theory of operators. The fundamental mathematical apparatus of quantum mechanics is the theory of the so-called self-adjoint operators. The formulation of mathematical problems arising in quantum mechanics was and still is a powerful stimulus for the further development of functional analysis.

The operator point of view on differential and integral equations turned out to be extremely useful also for the development of practical methods for approximate solutions of such equations.

Fundamental concepts of the theory of operators. Let us now proceed to an explanation of the fundamental definitions and facts in the theory of operators.

In analysis we have come across the concept of a function. In its simplest form this was a relation that associates with every number x (the value of the independent variable) a number y (the value of the function). In the further development of analysis it became necessary to consider relations of a more general type.

Such more general relations are discussed, for example, in the calculus of variations (Chapter VIII), where we associated with every function a number. If with every function a certain number is associated, then we say that we are given a functional. As an example of a functional we can take the association between an arbitrary function $y = f(x)$ $(a \leqslant x \leqslant b)$ and the arc length of the curve represented by it. We obtain another example of a functional if we associate with every function $y = f(x)$ $(a \leqslant x \leqslant b)$ its definite integral $\int_a^b f(x)\,dx$.

If we regard $f(x)$ as a point of an infinite-dimensional space, then a functional is simply a function of the points of the infinite-dimensional space. From this point of view the problems of the calculus of variations concern the search for maxima and minima of functions of the points of an infinite-dimensional space.

In order to define what we mean by a continuous functional it is necessary to define first what we mean by proximity of two points of an infinite-dimensional space. In §2 we gave the distance between two functions $f(x)$ and $g(x)$ (points of an infinite-dimensional space) as

$$\sqrt{\int_a^b [f(x) - g(x)]^2\,dx}.$$

This method of assigning a distance in infinite-dimensional space is often used, but of course it is not the only possible one. In other problems other methods of giving the distance between functions may turn out to be better. We may point, for example, to the problem of the theory of approximation of functions (see Chapter XII, §3), where the distance between functions, which characterizes the measure of proximity of the two functions $f(x)$ and $g(x)$, is given, for example, by the formula

$$\max | f(x) - g(x) |.$$

Other methods of giving a distance between functions are used in the investigation of functionals in the calculus of variations. Distinct methods of giving the distance between functions lead us to distinct infinite-dimensional spaces.

Thus, various infinite-dimensional (functional) spaces differ from each other by their set of functions and by the definition of distance between them. For example, if we take the set of all functions with integrable square and define distance as

$$\sqrt{\int_a^b [f(x) - g(x)]^2 \, dx},$$

then we arrive at the Hilbert space that was introduced in §2; but if we take the set of all continuous functions and define distance as $\max | f(x) - g(x) |$, then we obtain the so-called space (C).

In the discussion of integral equations we come across expressions of the form

$$g(x) = \int_a^b k(x, y) f(y) \, dy.$$

For a given kernel $k(x, y)$ this equation indicates a rule by which every function $f(x)$ is set in correspondence with another function $g(x)$.

This kind of a correspondence that relates with one function f another function g is called an *operator*.

We shall say that we are given a linear operator A in a Hilbert space if we have a rule by which we associate with every function f another function g. The correspondence need not be given for all the functions of the Hilbert space. In that case the set of those functions f for which there exists the function $g = Af$ is called the *domain of definition* of the operator A (similar to the domain of definition of a function in ordinary analysis). The correspondence itself is usually denoted as follows:

$$g = Af. \qquad (30)$$

The linearity of the operator means that the sum of the functions f_1 and f_2 is associated with the sum of Af_1 and Af_2, and the product of f and a number λ with the function λAf; i.e.,

$$A(f_1 + f_2) = Af_1 + Af_2 \tag{31}$$

and

$$A(\lambda f) = \lambda Af. \tag{32}$$

Occasionally continuity is also postulated for linear operators; i.e., it is required that the convergence of a sequence of functions f_n to a function f should imply that the sequence Af_n should converge to Af.

Let us give examples of linear operators.

1. Let us associate with every function $f(x)$ the function $g(x) = \int_a^x f(t)\, dt$, i.e., the indefinite integral of f. The linearity of this operator follows from the ordinary properties of the integral, i.e., from the fact that the integral of the sum is equal to the sum of the integrals and that a constant factor can be taken out of the integral sign.

2. Let us associate with every differentiable function $f(x)$ its derivative $f'(x)$. This operator is usually denoted by the letter D; i.e.,

$$f'(x) = D\, f(x).$$

Observe that this operator is not defined for all the functions of the Hilbert space but only for those that have a derivative belonging to the Hilbert space. These functions form, as we have said previously, the domain of definition of this operator.

3. The examples 1 and 2 were examples of linear operators in an infinite-dimensional space. But examples of linear operators in finite-dimensional spaces have occurred in other chapters of this book. Thus, in Chapter III affine transformations were investigated. If an affine transformation of a plane of space leaves the origin of coordinates fixed, then it is an example of a linear operator in a two-dimensional, or three-dimensional, space. The linear transformations of an n-dimensional space introduced in Chapter XVI now appear as linear operators in n-dimensional space.

4. In the integral equations, we have already met a very important and widely applicable class of linear operators in a functional space, namely the so-called integral operators. Let us choose a certain definite function $k(x, y)$. Then the formula

$$g(x) = \int_a^b k(x, y) f(y)\, dy$$

associates with every function f a certain function g. Symbolically we can write this transformation as follows:

$$g = Af.$$

The operator A in this case is called an integral operator. We could mention many other important examples of integral operators.

In §4 we spoke of the inhomogeneous integral equation

$$f(x) = \lambda \int_a^b k(x, y) f(y)\, dy + h(x).$$

In the notation of the theory of operators this equation can be rewritten as follows

$$f = \lambda Af + h, \tag{33}$$

where λ is a given number, h a given function (a vector of an infinite-dimensional space), and f the required function. In the same notation the homogeneous equation can be written as follows:

$$f = \lambda Af. \tag{34}$$

The classical theorems on integral equations, such as, for example, the theorem formulated in §4 on the connection between the solvability of the inhomogeneous and the corresponding homogeneous integral equation, are not true for every operator equation. However, one can indicate certain general conditions to be imposed on the operator A under which these theorems are true.

These conditions are stated in topological terms and express that the operator A should carry the unit sphere (i.e., the set of vectors whose length does not exceed 1) into a compact set.

Eigenvalues and eigenvectors of operators. The problem of eigenvalues and eigenfunctions of an integral equation to which we were led by problems on oscillations can be formulated as follows: to find the values λ for which there exist a nonzero function f satisfying the equation

$$f(x) = \lambda \int_a^b k(x, y) f(y)\, dy.$$

As before, this equation can be written as follows:

$$f = \lambda Af$$

or

$$Af = \frac{1}{\lambda} f. \tag{35}$$

Now we shall understand by A an arbitrary linear operator. Then a vector f satisfying the equation (35) is called an eigenvector of the operator A, and the number $1/\lambda$ the corresponding eigenvalue.

Since the vector $(1/\lambda)f$ coincides in direction with the vector f (differs from f only by a numerical factor), the problem of finding eigenvectors can also be stated as the problem of finding nonzero vectors f that do not change direction under the transformation A.

This way of looking at the eigenvalues enables us to unify the problem of eigenvalues of integral equations (if A is an integral operator), differential equations (if A is a differential operator), and the problem of eigenvalues in linear algebra (if A is a linear transformation in finite-dimensional space; see Chapter VI and Chapter XVI). In the case of three-dimensional space this problem arises in the search for the so-called principal axes of an ellipsoid.

In the case of integral equations a number of important properties of the eigenfunctions and eigenvalues (for example the reality of the eigenvalues, the orthogonality of the eigenfunctions, etc.) are consequences of the symmetry of the kernel, i.e., of the equation $k(x, y) = k(y, x)$.

For an arbitrary linear operator A in a Hilbert space the analogue of of this property is the so-called self-adjointness of the operator.

The condition for an operator A to be self-adjoint in the general case is that for any two elements f_1 and f_2 the equation

$$(Af_1, f_2) = (f_1, Af_2)$$

holds, where (Af_1, f_2) denotes the scalar product of the vector Af_1 and the vector f_2.

In problems of mechanics the condition of self-adjointness of an operator is usually a consequence of the law of conservation of energy. Therefore it is satisfied for operators connected with, say, oscillations for which there is no loss (dissipation) of energy.

The majority of operators that occur in quantum mechanics are also self-adjoint.

Let us verify that an integral operator with a symmetric kernel $k(x, y)$ is self-adjoint. In fact, in this case Af_1 is the function $\int_a^b k(x, y) f_1(y)\, dy$. Therefore the scalar product (Af_1, f_2), which is equal to the integral of the product of this function with f_2, is given by the formula

$$(Af_1, f_2) = \int_a^b \int_a^b k(x, y) f_1(y) f_2(x)\, dy\, dx.$$

Similarly

$$(f_1, Af_2) = \int_a^b \int_a^b k(x, y) f_2(y) f_1(x)\, dy\, dx.$$

The equation $(Af_1, f_2) = (f_1, Af_2)$ is an immediate consequence of the symmetry of the kernel $k(x, y)$.

Arbitrary self-adjoint operators have a number of important properties that are useful in the applications of these operators to the solution of a variety of problems. Indeed, the eigenvalues of a self-adjoint linear operator are always real and the eigenfunctions corresponding to distinct eigenvalues are orthogonal to each other.

Let us prove, for example, the last statement. Let λ_1 and λ_2 be two distinct eigenvalues of the operator A, and f_1 and f_2 eigenvectors corresponding to them. This means that

$$\begin{aligned} Af_1 &= \lambda_1 f_1, \\ Af_2 &= \lambda_2 f_2. \end{aligned} \tag{36}$$

We form the scalar product of the first equation (36) by f_2, and of the second by f_1. Then we have

$$\begin{aligned} (Af_1, f_2) &= \lambda_1(f_1, f_2), \\ (Af_2, f_1) &= \lambda_2(f_2, f_1). \end{aligned} \tag{37}$$

Since the operator A is self-adjoint, we have $(Af_1, f_2) = (Af_2, f_1)$. When we subtract the second equation (37) from the first, we obtain

$$0 = (\lambda_1 - \lambda_2)(f_1, f_2).$$

Since $\lambda_1 \neq \lambda_2$, we have $(f_1, f_2) = 0$, i.e., the eigenvectors f_1 and f_2 are orthogonal.

The investigation of self-adjoint operators has brought clarity into many concrete problems and questions connected with the theory of eigenvalues. Let us dwell in more detail on one of them, namely on the problem of the expansion by eigenefunctions in the case of a continuous spectrum.

In order to explain what a continuous spectrum means, let us turn again to the classical example of the oscillation of a string. Earlier we have shown that for a string of length l the characteristic frequencies of oscillations can assume the sequence of values.

$$a\frac{\pi}{l}, 2a\frac{\pi}{l}, \cdots, na\frac{\pi}{l}, \cdots.$$

Let us plot the points of this sequence on the numerical axis $O\lambda$. When we increase the length of the string l, the distance between any two adjacent points of the sequence will decrease, and they will fill the numerical axis more densely. In the limit, when $l \to \infty$, i.e., for an infinite string, the the eigenfrequencies fill the whole numerical semiaxis $\lambda \geqslant 0$. In this case we say that the system has a continuous spectrum.

We have already said that for a string of length l the expansion in a series by eigenfunctions is an expansion in a series by sines and cosines of $n(\pi/l)x$; i.e., in a trigonometric series

$$f(x) = \frac{a_0}{2} + \sum a_n \cos n \frac{\pi}{l} x + b_n \sin n \frac{\pi}{l} x.$$

For the case of an infinite string we can again show that a more or less arbitrary function can be expanded by sines and cosines. However, since the eigenfrequencies are now distributed continuously along the numerical line, this is not an expansion in a series, but in a so-called Fourier integral

$$f(x) = \int_{-\infty}^{+\infty} [A(\lambda) \cos \lambda x + B(\lambda) \sin \lambda x] \, d\lambda.$$

The expansion in a Fourier integral was already well known and widely used in the 19th century in the solutions of various problems of mathematical physics.

However, in more general cases with a continuous spectrum* many problems referring to an expansion of functions by eigenfunctions were not properly clarified. Only the creation of the general theory of self-adjoint operators brought the necessary clarity to these problems.

Let us mention still another set of classical problems that have been solved on the basis of the general theory of operators. The discussion of oscillations involving dissipation (scattering) of energy belongs to such problems.

In this case we can again look for free oscillations of the system in the form $u(x)\phi(t)$. However, in contrast to the case of oscillations without dissipation of energy, the function $\phi(t)$ is not simply $\cos \omega t$, but has the form $e^{-kt} \cos \omega t$, where $k > 0$. Thus, the corresponding solution has the form $u(x)e^{-kt} \cos \omega t$. In this case every point x again performs oscillations (with frequency ω), however the oscillations are damped because for $t \to \infty$ the amplitude of these oscillations containing the factor e^{-kt} tends to zero.

It is convenient to write the characteristic oscillations of the system in the complex form $u(x)e^{-i\lambda t}$, where in the absence of friction the number λ is real and in the presence of friction λ is complex.

---

* As examples we can take the oscillations of an inhomogeneous elastic medium and also many problems of quantum mechanics.

The problem on the oscillations of a system with dissipation of energy again leads to a problem on eigenvalues, but this time not for self-adjoint operators. A characteristic feature here is the presence of complex eigenvalues indicative of the damping of the free oscillations.

Using a method of the theory of operators in conjuntion with methods of the theory of analytic functions M. V. Keldyš investigated this class of problems in 1950–1951 and proved for it the completeness of the system of eigenfunctions.

Connection of functional analysis with other branches of mathematics and quantum mechanics. We have already mentioned that the creation of quantum mechanics gave a decisive impetus to the development of functional analysis. Just as the rise of the differential and integral calculus in the 18th century was dictated by the requirements of mechanics and classical physics, so the development of functional analysis was, and still is, the result of the vigorous influence of contemporary physics, principally of quantum mechanics. The fundamental mathematical apparatus of quantum mechanics consists of the branches of mathematics relating essentially to functional analysis. We can only briefly indicate the connections existing here, because an explanation of the foundations of quantum mechanics exceeds the framework of this book.

In quantum mechanics the state of the system is given in its mathematical description by a vector of Hilbert space. Such quantities as energy, impulse, and moment of momentum are investigated by means of self-adjoint operators. For example, the possible energy levels of an electron in an atom are computed as eigenvalues of the energy operator. The differences of these eigenvalues give the frequencies of the emitted quantum of light and thus define the structure of the radiation spectrum of the given substance. The corresponding states of the electron are here described as eigenfunctions of the energy operator.

The solution of problems of quantum mechanics often requires the computation of eigenvalues of various (usually differential) operators. In some complicated cases the precise solution of these problems turns out to be practically impossible. For an approximate solution of these problems the so-called perturbation theory is widely used, which enables us to find from the known eigenvalues and functions of a certain self-adjoint operator A the eigenvalues of an operator A_1 slightly different from it. We mention that the perturbation theory has not yet received a full mathematical foundation, which is an interesting and important mathematical problem.

Independently of the approximate determination of eigenvalues, we can often say a good deal about a given problem by means of qualitative investigation. This investigation proceeds in problems of quantum mechanics on the basis of the symmetries existing in the given case. As examples of such symmetries we can take the properties of symmetry of crystals, spherical symmetry in an atom, symmetry with respect to rotation, and others. Since the symmetries form a group (see Chapter XX), the group methods (the so-called representation theory of groups) enables us to answer a number of problems without computation. As examples we may mention the classification of atomic spectra, nuclear transformations, and other problems. Thus, quantum mechanics makes extensive use of the mathematical apparatus of the theory of self-adjoint operators. At the same time the continued contemporary development of quantum mechanics leads to a further development of the theory of operators by placing new problems before this theory.

The influence of quantum mechanics and also the internal mathematical developments of functional analysis have had the effect that in recent years algebraic problems and methods have played a significant role in functional analysis. This intensification of algebraic tendencies in contemporary analysis can well be compared with the growth of the value of algebraic methods in contemporary theoretical physics in comparison with the methods of physics of the 19th century.

In conclusion, we wish to emphasize once more that functional analysis is one of the rapidly developing branches of contemporary mathematics. Its connections and applications in contemporary physics, differential equations, approximate computations, and its use of general methods developed in algebra, topology, the theory of functions of a real variable, etc., make functional analysis one of the focal points of contemporary mathematics.

Suggested Reading

N. Dunford and J. T. Schwartz, *Linear operators. I. General theory*, Interscience, New York, 1958.

I. M. Gel'fand and Z. Ja. Šapiro, *Representations of the rotation group of three-dimensional space and their applications*, Amer. Math. Soc. Translations Series 2, vol. 2, 1956, 207-316.

A. N. Kolmogorov and S. V. Fomin, *Elements of the theory of functions and functional analysis*. Vol. 1, *Metric and normed spaces*. Vol. 2, *Measure, Lebesgue integrals and Hilbert space*. Graylock, New York, 1957/1961.

L. D. Landau and E. M. Lifšic, *Course of theoretical physics*. Vol. 3, *Quantum mechanics*, Pergamon, New York, 1958/1960.

F. Riesz and B. Sz.-Nagy, *Functional analysis*, Frederick Ungar, New York, 1955.

A. E. Taylor, *Introduction to functional analysis*, Wiley, New York, 1958.