$\|\phi_\alpha\| \neq 0$ for every α, then $\{\phi_\alpha/\|\phi_\alpha\|\}$ is orthonormal. Henceforth, we will assume that $\|\phi_\alpha\| \neq 0$ for all α for an orthogonal system $\{\phi_\alpha\}$. This implies that no element is zero and that no two elements are equal.

(8.21) Theorem *Any orthogonal system $\{\phi_\alpha\}$ in L^2 is countable.*

Proof. We may assume that $\{\phi_\alpha\}$ is orthonormal. Then for $\alpha \neq \beta$, we have

$$\|\phi_\alpha - \phi_\beta\|^2 = \int (\phi_\alpha - \phi_\beta)(\bar{\phi}_\alpha - \bar{\phi}_\beta) = \|\phi_\alpha\|^2 + \|\phi_\beta\|^2 = 2,$$

so that $\|\phi_\alpha - \phi_\beta\| = \sqrt{2}$. Since L^2 is separable, it follows that $\{\phi_\alpha\}$ must be countable.

A collection $\psi_1, \ldots, \psi_N$ is said to be *linearly independent* if $\sum_{k=1}^N a_k\psi_k(\mathbf{x}) = 0$ (a.e.) implies that every a_k is zero. An infinite collection of functions is called *linearly independent* if each finite subcollection is. No function in a linearly independent set can be zero a.e.

(8.22) Theorem *If $\{\psi_k\}$ is orthogonal, it is linearly independent.*

Proof. Suppose that $a_1\psi_{k_1} + \cdots + a_N\psi_{k_N} = 0$. Multiplying both sides by $\bar{\psi}_{k_1}$ and integrating, we obtain by orthogonality that $a_1 = 0$. Similarly, $a_2 = \cdots = a_N = 0$.

The converse of (8.22) is not true. However, the next result shows that if $\{\psi_k\}$ is linearly independent, then the system formed from suitable linear combinations of its elements is orthogonal.

(8.23) Theorem (*Gram-Schmidt Process*) *If $\{\psi_k\}$ is linearly independent, then the system $\{\phi_k\}$ defined by*

$$\begin{aligned} \phi_1 &= \psi_1 \\ \phi_2 &= a_{21}\psi_1 + \psi_2 \\ &\vdots \\ \phi_k &= a_{k1}\psi_1 + \cdots + a_{k,k-1}\psi_{k-1} + \psi_k \\ &\vdots \end{aligned}$$

is orthogonal for proper selection of the a_{ij}.

Proof. Having $\phi_1 = \psi_1$, we proceed by induction, assuming that $\phi_1, \ldots, \phi_{k-1}$ have been chosen as required. We will determine constants $b_{k1}, \ldots, b_{k,k-1}$ so that the function ϕ_k defined by

$$\phi_k = b_{k1}\phi_1 + \cdots + b_{k,k-1}\phi_{k-1} + \psi_k$$

is orthogonal to $\phi_1, \ldots, \phi_{k-1}$. If $j < k$,

$$\langle\phi_k,\phi_j\rangle = b_{kj}\langle\phi_j,\phi_j\rangle + \langle\psi_k,\phi_j\rangle$$

by orthogonality. Since $\langle\phi_j,\phi_j\rangle \neq 0$, b_{kj} can be chosen so that $\langle\phi_k,\phi_j\rangle = 0$, $j < k$. Since each ϕ_j with $j < k$ is a linear combination of $\psi_1, \ldots, \psi_j$, the theorem follows.

When the ϕ_k are selected by the Gram-Schmidt process, we shall say that they are *generated* from the ψ_k. Note that the triangular character of the matrix in (8.23) means that each ψ_k can also be written as a linear combination of the ϕ_j, $j \leq k$.

We call an orthogonal system $\{\phi_k\}$ *complete* if the only function which is orthogonal to every ϕ_k is zero; that is, $\{\phi_k\}$ is complete if $\langle f,\phi_k\rangle = 0$ for all k implies that $f = 0$ a.e. Thus, a complete orthogonal system is one which is maximal in the sense that it is not properly contained in any larger orthogonal system.

The *span* of a set of functions $\{\psi_k\}$ is the collection of all finite linear combinations of the ψ_k. In speaking of the span of $\{\psi_k\}$, we may always assume that $\{\psi_k\}$ is orthogonal by discarding any dependent functions and applying the Gram-Schmidt process to the resulting linearly independent set.

A set $\{\psi_k\}$ is called a *basis* for L^2 if its span is dense in L^2; that is, $\{\psi_k\}$ is a basis if given $f \in L^2$ and $\varepsilon > 0$, there exist N and $\{a_k\}$ such that $\|f - \sum_{k=1}^N a_k\psi_k\| < \varepsilon$. The a_k can always be chosen with rational real and imaginary parts. Any countable dense set in L^2 is of course a basis. It follows that L^2 has an orthogonal basis.

(8.24) Theorem *Any orthogonal basis in L^2 is complete. In particular, there exists a complete orthonormal basis for L^2.*

Proof. Let $\{\psi_k\}$ be an orthogonal basis for L^2. We may assume that $\{\psi_k\}$ is orthonormal. To show that it is complete, let $\langle f,\psi_k\rangle = 0$ for all k. Then $\langle f,f\rangle = \langle f, f - \sum_{k=1}^N a_k\psi_k\rangle$ for any N and a_k. By Schwarz's inequality, $|\langle f,f\rangle| \leq \|f\|\|f - \sum_{k=1}^N a_k\psi_k\|$, and so, since the term on the right can be chosen arbitrarily small, $\langle f,f\rangle = 0$. Therefore, $f = 0$ a.e., which completes the proof.

6. Fourier Series; Parseval's Formula

Let $\{\phi_k\}$ be any orthonormal system for L^2. If $f \in L^2$, the numbers defined by

$$c_k = c_k(f) = \langle f,\phi_k\rangle = \int_E f\bar{\phi}_k$$

are called the *Fourier coefficients* of f with respect to $\{\phi_k\}$. The series $\sum_k c_k\phi_k$ is called the *Fourier series* of f with respect to $\{\phi_k\}$, and denoted $S[f] = \sum_k c_k\phi_k$. We also write

$$f \sim \sum_k c_k\phi_k.$$

The first question we ask is how well $S[f]$, or more precisely, the sequence