Chapter 7

Fourier Series

What makes Hilbert spaces so powerful in many applications is the possibility of expressing a problem in terms of a suitable orthonormal basis. In this chapter, we study *Fourier series*, which correspond to the expansion of periodic functions with respect to an orthonormal basis of trigonometric functions. We explore a variety of applications of Fourier series, and introduce an important related class of orthonormal bases, called *wavelets*.

7.1 The Fourier basis

A function $f : \mathbb{R} \to \mathbb{C}$ is 2π-periodic if

$$f(x + 2\pi) = f(x) \qquad \text{for all } x \in \mathbb{R}.$$

The choice of 2π for the period is simply for convenience; different periods may be reduced to this case by rescaling the independent variable. A 2π-periodic function on $\mathbb{R}$ may be identified with a function on the circle, or one-dimensional torus, $\mathbb{T} = \mathbb{R}/(2\pi\mathbb{Z})$, which we define by identifying points in $\mathbb{R}$ that differ by $2\pi n$ for some $n \in \mathbb{Z}$. We could instead represent a 2π-periodic function $f : \mathbb{R} \to \mathbb{C}$ by a function on a closed interval $f : [a, a + 2\pi] \to \mathbb{C}$ such that $f(a) = f(a + 2\pi)$, but the choice of a here is arbitrary, and it is clearer to think of the function as defined on the circle, rather than an interval.

The space $C(\mathbb{T})$ is the space of continuous functions from $\mathbb{T}$ to $\mathbb{C}$, and $L^2(\mathbb{T})$ is the completion of $C(\mathbb{T})$ with respect to the L^2-norm,

$$\|f\| = \left(\int_{\mathbb{T}} |f(x)|^2 \, dx \right)^{1/2}.$$

Here, the integral over $\mathbb{T}$ is an integral with respect to x taken over any interval of length 2π. An element $f \in L^2(\mathbb{T})$ can be interpreted concretely as an equivalence class of Lebesgue measurable, square integrable functions from $\mathbb{T}$ to $\mathbb{C}$ with respect to the equivalence relation of almost-everywhere equality. The space $L^2(\mathbb{T})$ is a

Hilbert space with the inner product

$$\langle f, g\rangle = \int_{\mathbb{T}} \overline{f(x)} g(x)\, dx.$$

The Fourier basis elements are the functions

$$e_n(x) = \frac{1}{\sqrt{2\pi}} e^{inx}. \tag{7.1}$$

Our first objective is to prove that $\{e_n \mid n \in \mathbb{Z}\}$ is an orthonormal basis of $L^2(\mathbb{T})$. The orthonormality of the functions e_n is a simple computation:

$$\begin{aligned} \langle e_m, e_n\rangle &= \int_{\mathbb{T}} \overline{\frac{1}{\sqrt{2\pi}} e^{imx}} \frac{1}{\sqrt{2\pi}} e^{inx}\, dx \\ &= \frac{1}{2\pi} \int_0^{2\pi} e^{i(n-m)x}\, dx \\ &= \begin{cases} 1 & \text{if } m = n, \\ 0 & \text{if } m \neq n. \end{cases} \end{aligned}$$

Thus, the main result we have to prove is the completeness of $\{e_n \mid n \in \mathbb{Z}\}$. We denote the set of all finite linear combinations of the e_n by $\mathcal{P}$. Functions in $\mathcal{P}$ are called *trigonometric polynomials*. We will prove that any continuous function on $\mathbb{T}$ can be approximated uniformly by trigonometric polynomials, a result which is closely related to the Weierstrass approximation theorem in Theorem 2.9. Since uniform convergence on $\mathbb{T}$ implies L^2-convergence, and continuous functions are dense in $L^2(\mathbb{T})$, it follows that the trigonometric polynomials are dense in $L^2(\mathbb{T})$, so $\{e_n\}$ is a basis.

The idea behind the completeness proof is to obtain a trigonometric polynomial approximation of a continuous function f by taking the convolution of f with an approximate identity that is a trigonometric polynomial. Convolutions and approximate identities are useful in many other contexts, so we begin by describing them.

The *convolution* of two continuous functions $f, g : \mathbb{T} \to \mathbb{C}$ is the continuous function $f * g : \mathbb{T} \to \mathbb{C}$ defined by the integral

$$(f * g)(x) = \int_{\mathbb{T}} f(x-y) g(y)\, dy. \tag{7.2}$$

By changing variables $y \to x - y$, we may also write

$$(f * g)(x) = \int_{\mathbb{T}} f(y) g(x-y)\, dy,$$

so that $f * g = g * f$.

Definition 7.1 A family of functions $\{\varphi_n \in C(\mathbb{T}) \mid n \in \mathbb{N}\}$ is an *approximate identity* if:

$$\text{(a)} \quad \varphi_n(x) \geq 0; \tag{7.3}$$

$$\text{(b)} \quad \int_{\mathbb{T}} \varphi_n(x)\, dx = 1 \qquad \text{for every } n \in \mathbb{N}; \tag{7.4}$$

$$\text{(c)} \quad \lim_{n\to\infty} \int_{\delta \leq |x| \leq \pi} \varphi_n(x)\, dx = 0 \qquad \text{for every } \delta > 0. \tag{7.5}$$

In (7.5), we identify $\mathbb{T}$ with the interval $[-\pi, \pi]$.

Thus, each function φ_n has unit area under its graph, and the area concentrates closer to the origin as n increases. For large n, the convolution of a function f with φ_n therefore gives a local average of f.

Theorem 7.2 If $\{\varphi_n \in C(\mathbb{T}) \mid n \in \mathbb{N}\}$ is an approximate identity and $f \in C(\mathbb{T})$, then $\varphi_n * f$ converges uniformly to f as $n \to \infty$.

Proof. From (7.4), we have

$$f(x) = \int_{\mathbb{T}} \varphi_n(y) f(x)\, dy.$$

We also have that

$$(\varphi_n * f)(x) = \int_{\mathbb{T}} \varphi_n(y) f(x-y)\, dy.$$

We may therefore write

$$(\varphi_n * f)(x) - f(x) = \int_{\mathbb{T}} \varphi_n(y) \left[f(x-y) - f(x)\right] dy. \tag{7.6}$$

To show that the integral on the right-hand side of this equation is small when n is large, we consider the integrand separately for y close to zero and y bounded away from zero. The contribution to the integral from values of y close to zero is small because f is continuous, and the contribution to the integral from values of y bounded away from zero is small because the integral of φ_n is small.

More precisely, suppose that $\epsilon > 0$. Since f is continuous on the compact set $\mathbb{T}$, it is bounded and uniformly continuous. Therefore, there is an M such that $|f(x)| \leq M$ for all $x \in \mathbb{T}$, and there is a $\delta > 0$ such that $|f(x) - f(y)| \leq \epsilon$ whenever $|x - y| < \delta$. Then, estimating the integral in (7.6), we obtain that

$$\begin{aligned} |(\varphi_n * f)(x) - f(x)| &\leq \int_{-\pi}^{\pi} \varphi_n(y) \left|f(x-y) - f(x)\right| dy \\ &\leq \int_{|y|<\delta} \varphi_n(y) \left|f(x-y) - f(x)\right| dy \end{aligned}$$

$$
\begin{aligned}
& + \int_{|y|\geq\delta} \varphi_n(y) \left| f(x-y) - f(x) \right| \, dy \\
\leq \quad & \epsilon \int_{|y|<\delta} \varphi_n(y) \, dy + \int_{|y|\geq\delta} \varphi_n(y) \left[|f(x-y)| + |f(x)| \right] \, dy \\
\leq \quad & \epsilon + 2M \int_{|y|\geq\delta} \varphi_n(y) \, dy.
\end{aligned}
$$

Taking the sup of this inequality over x, the lim sup as $n \to \infty$, and using (7.5), we find that

$$\limsup_{n\to\infty} \|\varphi_n * f - f\|_\infty \leq \epsilon.$$

Since $\epsilon > 0$ is arbitrary, it follows that $\varphi_n * f \to f$ uniformly in $C(\mathbb{T})$. □

Theorem 7.3 The trigonometric polynomials are dense in $C(\mathbb{T})$ with respect to the uniform norm.

Proof. For each $n \in \mathbb{N}$, we define the function $\varphi_n \geq 0$ by

$$\varphi_n(x) = c_n \left(1 + \cos x\right)^n. \tag{7.7}$$

We choose the constant c_n so that

$$\int_{\mathbb{T}} \varphi_n(x) \, dx = 1. \tag{7.8}$$

Since $1 + \cos x$ has a strict maximum at $x = 0$, the graph of φ_n is sharply peaked at $x = 0$ for large n, and the area under the graph concentrates near $x = 0$. In particular, $\{\varphi_n\}$ satisfies (7.5) (see Exercise 7.1). It follows that $\{\varphi_n\}$ is an approximate identity, and hence $\varphi_n * f$ converges uniformly to f from Theorem 7.2.

To complete the proof, we show that $\varphi_n * f$ is a trigonometric polynomial for any continuous function f. First, φ_n is a trigonometric polynomial; in fact,

$$\varphi_n(x) = \sum_{k=-n}^{n} a_{nk} e^{ikx}, \quad \text{where} \quad a_{nk} = 2^{-n} c_n \binom{2n}{n+k}.$$

Therefore,

$$
\begin{aligned}
\varphi_n * f(x) &= \int_{\mathbb{T}} \sum_{k=-n}^{n} a_{nk} e^{ik(x-y)} f(y) \, dy \\
&= \sum_{k=-n}^{n} a_{nk} e^{ikx} \int_{\mathbb{T}} e^{-iky} f(y) \, dy \\
&= \sum_{k=-n}^{n} b_k e^{ikx},
\end{aligned}
$$

where

$$b_k = a_{nk} \int_{\mathbb{T}} e^{-iky} f(y)\, dy.$$

Thus, $\varphi_n * f$ is a trigonometric polynomial. □

From the completeness of the Fourier basis, it follows that any function $f \in L^2(\mathbb{T})$ may be expanded in a Fourier series as

$$f(x) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \widehat{f}_n e^{inx},$$

where the equality means convergence of the partial sums to f in the L^2-norm, or

$$\lim_{N\to\infty} \int_{\mathbb{T}} \left| \frac{1}{\sqrt{2\pi}} \sum_{n=-N}^{N} \widehat{f}_n e^{inx} - f(x) \right|^2 dx = 0.$$

From orthonormality, the *Fourier coefficients* $\widehat{f}_n \in \mathbb{C}$ of f are given by $\widehat{f}_n = \langle e_n, f\rangle$, or

$$\widehat{f}_n = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{T}} f(x) e^{-inx}\, dx.$$

Moreover, Parseval's identity implies that

$$\int_{\mathbb{T}} \overline{f(x)} g(x)\, dx = \sum_{n=-\infty}^{\infty} \overline{\widehat{f}_n} \widehat{g}_n.$$

In particular, the L^2-norm of a function can be computed either in terms of the function or its Fourier coefficients, since

$$\int_{\mathbb{T}} |f(x)|^2\, dx = \sum_{n=-\infty}^{\infty} \left|\widehat{f}_n\right|^2. \tag{7.9}$$

Thus, the periodic Fourier transform $\mathcal{F} : L^2(\mathbb{T}) \to \ell^2(\mathbb{Z})$ that maps a function to its sequence of Fourier coefficients, by

$$\mathcal{F}f = \left(\widehat{f}_n\right)_{n=-\infty}^{\infty},$$

is a Hilbert space isomorphism between $L^2(\mathbb{T})$ and $\ell^2(\mathbb{Z})$. The projection theorem, Theorem 6.13, implies that the partial sum

$$f_N(x) = \frac{1}{\sqrt{2\pi}} \sum_{n=-N}^{N} \widehat{f}_n e^{inx}$$

is the best approximation of f by a trigonometric polynomial of degree N in the sense of the L^2-norm.

An important property of the Fourier transform is that it maps the convolution of two functions to the pointwise product of their Fourier coefficients. The convolution of two L^2-functions may either be defined by (7.2), where the integral is a Lebesgue integral, or by a density argument using continuous functions, as in the following proposition.

Proposition 7.4 If $f, g \in L^2(\mathbb{T})$, then $f * g \in C(\mathbb{T})$ and

$$\|f * g\|_\infty \leq \|f\|_2 \|g\|_2. \tag{7.10}$$

Proof. If $f, g \in C(\mathbb{T})$, then application of the Cauchy-Schwarz inequality to (7.2) implies that

$$|f * g(x)| \leq \|f\|_2 \|g\|_2.$$

Taking the supremum of this equation with respect to x, we get (7.10). If $f, g \in L^2(\mathbb{T})$, then there are sequences (f_k) and (g_k) of continuous functions such that $\|f - f_k\|_2 \to 0$ and $\|g - g_k\|_2 \to 0$ as $k \to \infty$. The convolutions $f_k * g_k$ are continuous functions. Moreover, they form a Cauchy sequence with respect to the sup-norm since, from (7.10),

$$\begin{aligned}
\|f_j * g_j - f_k * g_k\|_\infty &\leq \|(f_j - f_k) * g_j\|_\infty + \|f_k * (g_j - g_k)\|_\infty \\
&\leq \|f_j - f_k\|_2 \|g_j\|_2 + \|f_k\|_2 \|g_j - g_k\|_2 \\
&\leq M \left(\|f_j - f_k\|_2 + \|g_j - g_k\|_2\right).
\end{aligned}$$

Here, we use the fact that $\|f_j\|_2 \leq M$ and $\|g_k\|_2 \leq M$ for some constant M because the sequences converge in $L^2(\mathbb{T})$. By the completeness of $C(\mathbb{T})$, the sequence $(f_k * g_k)$ converges uniformly to a continuous function $f * g$. This limit is independent of the sequences used to approximate f and g, and it satisfies (7.10). □

The inequality (7.10) is a special case of *Young's inequality* for convolutions (see Theorem 12.58).

Theorem 7.5 (Convolution) If $f, g \in L^2(\mathbb{T})$, then

$$\widehat{(f * g)}_n = \sqrt{2\pi}\, \widehat{f}_n \widehat{g}_n. \tag{7.11}$$

Proof. Because of the density of $C(\mathbb{T})$ in $L^2(\mathbb{T})$, and the continuity of the Fourier transform and the convolution with respect to L^2-convergence, it is sufficient to prove (7.11) for continuous functions f, g. In that case, we may exchange the order of integration in the following computation:

$$\begin{aligned}
\widehat{(f * g)}_n &= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{T}} f * g(x) e^{-inx}\, dx \\
&= \frac{1}{\sqrt{2\pi}} \int_{\mathbb{T}} \left(\int_{\mathbb{T}} f(x - y) g(y)\, dy \right) e^{-inx}\, dx
\end{aligned}$$

$$
\begin{aligned}
&= \int_{\mathbb{T}} \frac{1}{\sqrt{2\pi}} \left(\int_{\mathbb{T}} f(x-y) e^{-in(x-y)}\, dx \right) g(y) e^{-iny}\, dy \\
&= \widehat{f}_n \int_{\mathbb{T}} g(y) e^{-iny}\, dy \\
&= \sqrt{2\pi}\, \widehat{f}_n \widehat{g}_n.
\end{aligned}
$$

This proves the theorem. □

Alternatively, we may prove Theorem 7.5 directly for $f, g \in L^1(\mathbb{T})$. The exchange in the order of integration is justified by Fubini's theorem, Theorem 12.41.

The L^2-convergence of Fourier series is particularly simple. It is nevertheless interesting to ask about other types of convergence. For example, the Fourier series of a function $f \in L^2(\mathbb{T})$ also converges pointwise a.e. to f. This result was proved by Carleson, only as recently as 1966. An analysis of the pointwise convergence of Fourier series is very subtle, and the proof is beyond the scope of this book. For smooth functions, such as continuously differentiable functions, the convergence of the partial sums is uniform, as we will show in Section 7.2 below.

The behavior of the partial sums near a point of discontinuity of a piecewise smooth function is interesting. The sums do not converge uniformly; instead the partial sums oscillate in an interval that contains the point of discontinuity. The width of the interval where the oscillations occur shrinks to zero as $N \to 0$, but the size of the oscillations does not — in fact, for large N, the magnitude of the oscillations is approximately 9%of the jump in f at the jump discontinuity. This behavior is called the *Gibbs phenomenon*. As a result, care is required when one uses Fourier series to represent discontinuous functions.

It is often convenient to modify the orthonormal basis $\{e_n(x)\}$ in (7.1) slightly. First, if we use the non-normalized orthogonal basis $\{e^{inx}\}$, then the Fourier expansion of $f \in L^2(\mathbb{T})$ is

$$f(x) = \sum_{n=-\infty}^{\infty} \widehat{f}_n e^{inx},$$

where

$$\widehat{f}_n = \frac{1}{2\pi} \int_{\mathbb{T}} f(x) e^{-inx}\, dx.$$

Second, the real-valued functions

$$\{1, \cos nx, \sin nx \mid n = 1, 2, 3, \ldots\}$$

also form an orthogonal basis, since

$$\cos nx = \frac{e^{inx} + e^{-inx}}{2}, \qquad \sin nx = \frac{e^{inx} - e^{-inx}}{2i}.$$

The corresponding Fourier expansion of $f \in L^2(\mathbb{T})$ is

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

where

$$a_0 = \frac{1}{\pi}\int_{\mathbb{T}} f(x)\,dx, \quad a_n = \frac{1}{\pi}\int_{\mathbb{T}} f(x)\cos nx\,dx, \quad b_n = \frac{1}{\pi}\int_{\mathbb{T}} f(x)\sin nx\,dx.$$

This basis has the advantage that a real-valued function has real Fourier coefficients a_n, b_n. A second useful property of this basis is that its elements are even or odd. A function f is *even* if $f(-x) = f(x)$ for all x, and *odd* if $f(-x) = -f(x)$ for all x. Even functions f have a *Fourier cosine expansion* of the form

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

while odd functions f have a *Fourier sine expansion* of the form

$$f(x) = \sum_{n=1}^{\infty} b_n \sin(nx).$$

If a function is defined on the interval $[0, \pi]$, then we may extend it to an even or an odd 2π periodic function on $\mathbb{R}$. The original function may therefore be represented by a Fourier cosine or sine expansion on $[0, \pi]$ (see Exercise 7.3). The quality of the approximation of a function by the partial sums of a Fourier series sometimes depends significantly on the basis used for the expansion. This is illustrated in Figure 7.1.

Fourier series of multiply periodic functions are defined in an entirely analogous way. A function $f : \mathbb{R}^d \to \mathbb{C}$ is 2π-periodic in each variable if

$$f(x_1, x_2, \ldots, x_i + 2\pi, \ldots, x_d) = f(x_1, x_2, \ldots, x_i, \ldots, x_d) \qquad \text{for } i = 1, \ldots, d.$$

We may regard a multiply periodic function as a function on the d-dimensional torus $\mathbb{T}^d = \mathbb{R}^d/(2\pi\mathbb{Z})^d$, which is the Cartesian product of d circles. An orthonormal basis of $L^2(\mathbb{T}^d)$ consists of the functions

$$e_{\mathbf{n}}(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}} e^{i\mathbf{n}\cdot\mathbf{x}},$$

where $\mathbf{x} = (x_1, \ldots, x_d) \in \mathbb{T}^d$, $\mathbf{n} = (n_1, \ldots, n_d) \in \mathbb{Z}^d$, and

$$\mathbf{n} \cdot \mathbf{x} = n_1 x_1 + \cdots + n_d x_d.$$

The Fourier series expansion of a function $f \in L^2(\mathbb{T}^d)$ is

$$f(\mathbf{x}) = \frac{1}{(2\pi)^{d/2}} \sum_{\mathbf{n}\in\mathbb{Z}^d} \widehat{f}_{\mathbf{n}} e^{i\mathbf{n}\cdot\mathbf{x}},$$

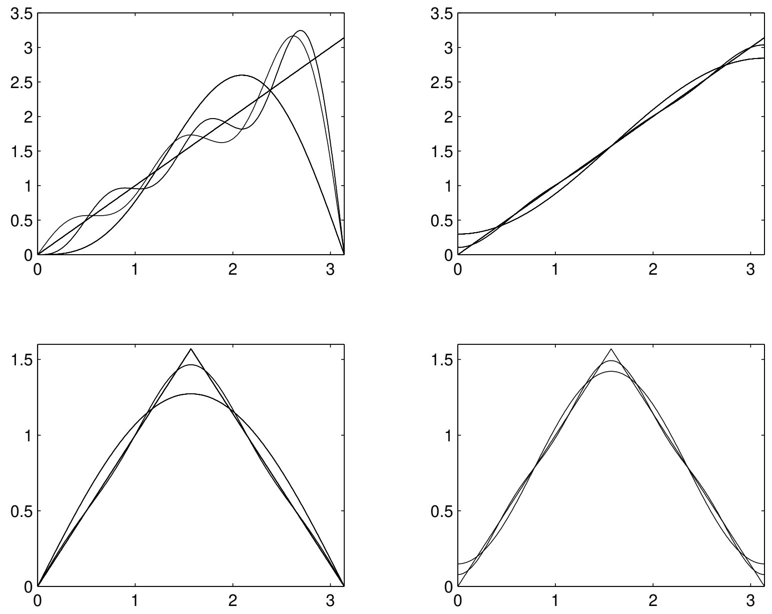

Fig. 7.1 Fourier sine (left) and cosine (right) series for two piecewise linear functions. Note the difference in the quality of the approximations.

where the series converges unconditionally with respect to the L^2-norm, and

$$\widehat{f}_{\mathbf{n}} = \frac{1}{(2\pi)^{d/2}} \int_{\mathbb{T}^d} f(\mathbf{x}) e^{-i\mathbf{n}\cdot\mathbf{x}}\, d\mathbf{x}.$$

7.2 Fourier series of differentiable functions

There is an important connection between the smoothness of a function and the rate of decay of its Fourier coefficients: the smoother a function (that is, the more times it is differentiable), the faster its Fourier coefficients decay. Heuristically, a smooth function contains a small amount of high frequency components.

If $f \in C^1(\mathbb{T})$ is continuously differentiable, then we can relate the Fourier coefficients of f' to those of f using an integration by parts:

$$\begin{aligned} \widehat{f'}_n &= \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} e^{-inx} f'(x)\, dx \\ &= \frac{1}{\sqrt{2\pi}} [f(2\pi) - f(0)] - \frac{1}{\sqrt{2\pi}} \int_0^{2\pi} (-in) e^{-inx} f(x)\, dx \\ &= in\widehat{f}_n. \end{aligned} \tag{7.12}$$

Thus, differentiation of a function corresponds to multiplication of its Fourier coefficients by in. It follows by induction that if $f \in C^k(\mathbb{T})$, then

$$\widehat{f_n^{(k)}} = (in)^k \hat{f}_n.$$

Equation (7.12) may be used to define the notion of the derivative of a function whose derivative is square integrable, but need not be continuous. Such a derivative is called a *weak derivative*. The space of functions in L^2 whose weak derivatives are in L^2 is denoted by H^1, and is an example of a Sobolev space.

Definition 7.6 The *Sobolev space* $H^1(\mathbb{T})$ consists of all functions

$$f(x) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} \hat{f}_n e^{inx} \in L^2(\mathbb{T})$$

such that

$$\sum_{n=-\infty}^{\infty} n^2 \left|\hat{f}_n\right|^2 < \infty.$$

The weak L^2-derivative $f' \in L^2(\mathbb{T})$ of $f \in H^1(\mathbb{T})$ is defined by the L^2-convergent Fourier series

$$f'(x) = \frac{1}{\sqrt{2\pi}} \sum_{n=-\infty}^{\infty} in\hat{f}_n e^{inx}.$$

The space $H^1(\mathbb{T})$ is a Hilbert space with respect to the inner product

$$\langle f, g\rangle_{H^1} = \int_{\mathbb{T}} \left\{\overline{f(x)}g(x) + \overline{f'(x)}g'(x)\right\}\, dx.$$

By Parseval's theorem, the H^1-inner product of two functions may be written in terms of their Fourier coefficients as

$$\langle f, g\rangle_{H^1} = \sum_{n=-\infty}^{\infty} \left(1+n^2\right) \overline{\hat{f}_n}\hat{g}_n.$$

Convergence with respect to the associated H^1-norm corresponds to mean-square convergence of functions and their derivatives.

A continuously differentiable function belongs to $H^1(\mathbb{T})$ and its weak derivative is equal to the usual pointwise derivative. It follows from the density of $C(\mathbb{T})$ in $L^2(\mathbb{T})$ that $H^1(\mathbb{T})$ is the completion of the space $C^1(\mathbb{T})$ of continuously differentiable functions (or the space of trigonometric polynomials) with respect to the H^1-norm.

If $f, g \in H^1(\mathbb{T})$, then the definition of f' and Parseval's theorem imply that

$$\langle f', g\rangle_{L^2} = \sum_{n=-\infty}^{\infty} \overline{in\hat{f}_n}\hat{g}_n = -\sum_{n=-\infty}^{\infty} \overline{\hat{f}_n} in\hat{g}_n = -\langle f, g'\rangle_{L^2}.$$

After replacing f by $\overline{f}$, we see that weak derivatives satisfy integration by parts:

$$\int_{\mathbb{T}} f'g\,dx = -\int_{\mathbb{T}} fg'\,dx. \tag{7.13}$$

There are no boundary terms in the integration by parts formula for periodic functions. We can use (7.13) to give an equivalent definition of the weak L^2-derivative of a function in terms of its integral against a smooth *test function*. If $f \in H^1(\mathbb{T})$, then the linear functional $F : C^1(\mathbb{T}) \subset L^2(\mathbb{T}) \to \mathbb{C}$ defined by

$$F(\varphi) = -\int_{\mathbb{T}} f\varphi'\,dx, \qquad \varphi \in C^1(\mathbb{T}),$$

is bounded. Conversely, if F is bounded for a given function $f \in L^2(\mathbb{T})$, then, since $C^1(\mathbb{T})$ is dense in $H^1(\mathbb{T})$, the bounded linear transformation theorem, Theorm 5.19, implies that F extends to a unique bounded linear functional on $L^2(\mathbb{T})$. The Riesz representation theorem (see Theorem 8.12 below) therefore implies that there is a unique function $f' \in L^2(\mathbb{T})$ such that $F(\varphi) = \langle \overline{f}', \varphi\rangle$ holds for all $\varphi \in C^1(\mathbb{T})$. This leads to the following alternative definition of a weak L^2-derivative.

Definition 7.7 A function $f \in L^2(\mathbb{T})$ belongs to $H^1(\mathbb{T})$ if there is a constant M such that

$$\left|\int_{\mathbb{T}} f\varphi'\,dx\right| \le M\|\varphi\|_{L^2} \qquad \text{for all } \varphi \in C^1(\mathbb{T}).$$

If $f \in H^1(\mathbb{T})$, then the *weak derivative* f' of f is the unique element of $L^2(\mathbb{T})$ such that

$$\int_{\mathbb{T}} f'\varphi\,dx = -\int_{\mathbb{T}} f\varphi'\,dx \qquad \text{for all } \varphi \in C^1(\mathbb{T}).$$

More generally, for any $k \ge 0$, we define the Sobolev space

$$H^k(\mathbb{T}) = \left\{ f \in L^2(\mathbb{T}) \;\middle|\; f(x) = \sum_{n=-\infty}^{\infty} c_n e^{inx}, \quad \sum_{n=-\infty}^{\infty} |n|^{2k}|c_n|^2 < \infty \right\}.$$

If k is a natural number, the space H^k consists of functions with k square-integrable weak derivatives, but the Fourier series definition makes sense even when k is not a natural number.

Lemma 7.8 Suppose that $f \in H^k(\mathbb{T})$ for $k > 1/2$. Let

$$S_N(x) = \frac{1}{\sqrt{2\pi}} \sum_{n=-N}^{N} \widehat{f}_n e^{inx} \tag{7.14}$$

be the Nth partial sum of the Fourier series of f, and define

$$\left\| f^{(k)} \right\| = \left(\sum_{n=-\infty}^{\infty} |n|^{2k} \left|\widehat{f}_n\right|^2 \right)^{1/2}.$$

Then there is a constant C_k, independent of f, such that

$$\|S_N - f\|_\infty \le \frac{C_k}{N^{k-1/2}} \left\| f^{(k)} \right\|,$$

and (S_N) converges uniformly to f as $N \to \infty$.

Proof. Since $S_N \in C(\mathbb{T})$ for every $N \in \mathbb{N}$ and $C(\mathbb{T})$ is complete with respect to the supremum norm, it is sufficient to prove that for all $M > N$,

$$\|S_N - S_M\|_\infty \le \frac{C_k}{N^{k-1/2}} \left\| f^{(k)} \right\|.$$

This equation follows from (7.12) and the Cauchy-Schwarz inequality:

$$\begin{aligned}
\|S_N - S_M\|_\infty &\le \frac{1}{\sqrt{2\pi}} \sum_{N<|n|\le M} \left| \widehat{f}_n \right| \\
&= \frac{1}{\sqrt{2\pi}} \sum_{N<|n|\le M} |n|^k \left| \widehat{f}_n \right| \frac{1}{|n|^k} \\
&\le \frac{1}{\sqrt{2\pi}} \left[\sum_{N<|n|\le M} |n|^{2k} \left| \widehat{f}_n \right|^2 \right]^{1/2} \left[\sum_{N<|n|\le M} \frac{1}{|n|^{2k}} \right]^{1/2} \\
&\le \frac{1}{\sqrt{2\pi}} \left\| f^{(k)} \right\| \left[\int_N^\infty \frac{dr}{r^{2k}} \right]^{1/2} \\
&\le \frac{1}{\sqrt{2\pi(2k-1)}} \frac{1}{N^{k-1/2}} \left\| f^{(k)} \right\|,
\end{aligned}$$

which proves the result with

$$C_k = \frac{1}{\sqrt{2\pi(2k-1)}}.$$

□

A corollary of this lemma is a special case of the *Sobolev embedding theorem*, which implies, in particular, that if a function on $\mathbb{T}$ has a square-integrable weak derivative, then it is continuous.

Theorem 7.9 (Sobolev embedding) If $f \in H^k(\mathbb{T})$ for $k > 1/2$, then $f \in C(\mathbb{T})$.

Proof. From Lemma 7.8, the partial sums of the Fourier series of f converge uniformly, so the limit is continuous. □

More generally, if $f \in H^k(\mathbb{T})$, then the Fourier series for the derivatives $f^{(j)}$ converge uniformly when $k > j + 1/2$, so $f \in C^\ell(\mathbb{T})$, where ℓ is the greatest integer strictly less than $k - 1/2$. For functions of several variables, one finds that $f \in H^k(\mathbb{T}^d)$ is continuous when $k > d/2$, and j-times continuously differentiable when $k > j + d/2$ (see Exercise 7.5). Roughly speaking, there is a "loss" of slightly more than one-half a derivative per space dimension in passing from L^2 derivatives to continuous derivatives.

7.3 The heat equation

Fourier series are an essential tool for the study of a wide variety of problems in engineering and science. In this section, we use Fourier series to solve the heat, or diffusion, equation which models the flow of heat in a conducting body. This was the original problem that led Jean-Baptiste Fourier to develop the series expansion named after him, although similar ideas had been suggested earlier by Daniel Bernoulli. The same equation also describes the diffusion of a dye or pollutant in a fluid.

We consider a thin ring made of a heat conducting material. In a one-dimensional approximation, we can represent the ring by a circle. We choose units of space and time so that the length of the ring is 2π and the thermal conductivity of the material is equal to one. The temperature $u(x,t)$ at time $t \geq 0$ and position $x \in \mathbb{T}$ along the ring satisfies the *heat* or *diffusion equation*,

$$\begin{aligned} u_t &= u_{xx}, \\ u(x,0) &= f(x), \end{aligned} \tag{7.15}$$

where $f : \mathbb{T} \to \mathbb{R}$ is a given function describing the initial temperature in the ring.

If $u(x,t)$ is a smooth solution of the heat equation, then, multiplying the equation by $2u$ and rearranging the result, we get that

$$\left(u^2\right)_t = \left(2uu_x\right)_x - 2u_x^2.$$

Integration of this equation over $\mathbb{T}$, and use of the periodicity of u, implies that

$$\frac{d}{dt}\int_{\mathbb{T}} u^2(x,t)\,dx = -2\int_{\mathbb{T}} |u_x(x,t)|^2\,dx \leq 0.$$

Therefore, $\|u(\cdot,t)\| \leq \|f(\cdot)\|$, where $\|\cdot\|$ denotes the L^2-norm in x, so it is reasonable to look for solutions $u(\cdot,t)$ that belong to $L^2(\mathbb{T})$ for all $t \geq 0$.

To make the notion of solutions that belong to L^2 more precise, let us first suppose that the initial data f is a trigonometric polynomial,

$$f(x) = \sum_{n=-N}^{N} f_n e^{inx}. \tag{7.16}$$

We look for a solution

$$u(x,t) = \sum_{n=-N}^{N} u_n(t) e^{inx}, \tag{7.17}$$

that is also a trigonometric polynomial, with coefficients $u_n(t)$ that are continuously differentiable functions of t. Using (7.17) in the heat equation, computing the t and x derivatives, and equating Fourier coefficients, we find that $u_n(t)$ satisfies

$$\dot{u}_n + n^2 u_n = 0, \tag{7.18}$$

$$u_n(0) = f_n,$$

where the dot denotes a derivative with respect to t. Thus, the PDE (7.15) reduces to a decoupled system of ODEs. The solutions of (7.18) are

$$u_n(t) = f_n e^{-n^2 t}.$$

Therefore, the solution of the heat equation with the initial data (7.16) is given by

$$u(x,t) = \sum_{n=-N}^{N} f_n e^{-n^2 t} e^{inx}. \tag{7.19}$$

We may write this solution more abstractly as

$$u(\cdot, t) = T(t) f(\cdot),$$

where $T(t) : \mathcal{P} \to \mathcal{P}$ is the linear operator on the space $\mathcal{P}$ of trigonometric polynomials defined by

$$T(t)\left[\sum_{n=-N}^{N} f_n e^{inx}\right] = \sum_{n=-N}^{N} f_n e^{-n^2 t} e^{inx}.$$

Parseval's theorem (7.9) implies that, for $t \geq 0$,

$$\|T(t)f\|^2 = \sum_{n=-N}^{N} |f_n|^2 e^{-2n^2 t} \leq \sum_{n=-N}^{N} |f_n|^2 = \|f\|^2.$$

Thus, the solution operator $T(t)$ is bounded with respect to the L^2-operator norm when $t \geq 0$. The operator $T(t)$ is unbounded when $t < 0$.

By the bounded linear transformation theorem, Theorem 5.19, there is a unique bounded extension of $T(t)$ from $\mathcal{P}$ to $L^2(\mathbb{T})$, which we still denote by $T(t)$. Explicitly, if

$$f(x) = \sum_{n=-\infty}^{\infty} f_n e^{inx} \in L^2(\mathbb{T})$$

then $u(\cdot, t) = T(t) f \in L^2(\mathbb{T})$ is given by

$$u(x,t) = \sum_{n=-\infty}^{\infty} f_n e^{-n^2 t} e^{inx}. \tag{7.20}$$

We may regard this equation as defining the exponential

$$T(t) = e^{t\partial^2/\partial x^2}$$

of the unbounded operator $A = \partial^2/\partial x^2$ with periodic boundary conditions. Rather than consider in detail when this Fourier series converges to a continuously differentiable, or *classical*, solution of the heat equation that satisfies the initial condition pointwise, we will simply say that the function u obtained in this way is a *weak solution* of the heat equation (7.15). In Chapter 11, we will see that this point of view corresponds to interpreting the derivatives in (7.15) in a distributional sense.

The operators $T(t)$ have the following properties:

(a) $T(0) = I$;
(b) $T(s)T(t) = T(s+t)$ for $s, t \geq 0$;
(c) $T(t)f \to f$ as $t \to 0^+$ for each $f \in L^2(\mathbb{T})$.

In particular, $T(t)$ converges strongly, but not uniformly, to I as $t \to 0^+$. We say that $\{T(t) \mid t \geq 0\}$ is a C_0*-semigroup*, in contrast with the uniformly continuous group of operators with a bounded generator defined in Theorem 5.49.

The action of the solution operator $T(t)$ on a function is given by multiplication of the function's nth Fourier coefficient by $e^{-n^2 t}$. The convolution theorem, Theorem 7.5, implies that for $t > 0$ the operator has the spatial representation $T(t)f = g^t * f$ of convolution with a function g^t, called the *Green's function*, where

$$g^t(x) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} e^{inx - n^2 t}. \tag{7.21}$$

Using the Poisson summation formula in (11.43), we can write this series as an infinite, periodic sum of Gaussians,

$$g^t(x) = \frac{1}{\sqrt{4\pi t}} \sum_{n=-\infty}^{\infty} e^{-(x-2\pi n)^2/4t}.$$

We can immediately read off from (7.20) several important qualitative properties of the heat equation. The first is the *smoothing property*. For every $t > 0$, we have $u(\cdot, t) \in C^\infty(\mathbb{T})$, because the Fourier coefficients decay exponentially quickly as $n \to \infty$. This holds even if the initial condition has a discontinuity, as illustrated in Figure 7.2 for the case of a step function. In more detail, we have

$$\sum_{n=-\infty}^{\infty} |n|^{2k} \left| f_n e^{-n^2 t} \right|^2 \leq \max_{n \in \mathbb{N}} \left\{ n^{2k} e^{-n^2 t} \right\} \sum_m |f_m|^2 < \infty$$

for each $k \geq 0$, so $u(\cdot, t) \in H^k(\mathbb{T})$ for every $k \in \mathbb{N}$. The Sobolev embedding theorem in Theorem 7.9 implies that $u(\cdot, t) \in C^{k-1}(\mathbb{T})$ for every $k \in \mathbb{N}$, and therefore $u(x, t)$ has continuous partial derivatives with respect to x of all orders. It then follows from the heat equation that u has continuous partial derivatives with respect to t of all orders for $t > 0$.

The second property is *irreversibility*. A solution may not exist for $t < 0$, even if the "final data" f is C^∞. For example, if

$$f(x) = \sum_{n=-\infty}^{\infty} e^{-|n|} e^{inx} \in C^\infty(\mathbb{T}),$$

then the Fourier series solution (7.20) diverges for any $t < 0$. Equivalently, letting $t \to -t$, we see that the initial value problem for the *backwards heat equation*,

$$u_t = -u_{xx}, \qquad u(x,0) = f(x),$$

may not have a solution for $t > 0$, and is said to be *ill-posed*. This ill-posedness reflects the impossibility of determining the temperature distribution that led to a given observed temperature distribution because of the rapid damping of temperature variations that fluctuate rapidly in space.

The third property is the *exponential decay* of solutions to an equilibrium state as $t \to +\infty$. It follows from the heat equation that the mean temperature,

$$\langle u \rangle = \frac{1}{2\pi} \int_{\mathbb{T}} f(x)\, dx,$$

is independent of time since

$$\frac{d}{dt} \int_{\mathbb{T}} u(x,t)\, dx = \int_{\mathbb{T}} u_t(x,t)\, dx = \int_{\mathbb{T}} u_{xx}(x,t)\, dx = 0.$$

The solution $u(x,t)$ converges exponentially quickly to its mean value, because

$$\begin{aligned} \sup_{x\in\mathbb{T}} \left| \sum_{n=-\infty}^{\infty} f_n e^{-n^2 t} e^{inx} - \langle u \rangle \right| &\leq \sum_{n\neq 0} |f_n| e^{-n^2 t} \\ &\leq \left[\sum_{n\neq 0} |f_n|^2 \right]^{1/2} \left[2 \sum_{n=1}^{\infty} e^{-2n^2 t} \right]^{1/2} \\ &\leq C \|f\|_{L^2} e^{-t}, \end{aligned}$$

where C is a suitable constant. The exponential decay is a consequence of a *spectral gap* between the lowest eigenvalue, zero, of the operator $\partial^2/\partial x^2$ on $\mathbb{T}$ and the rest of its spectrum.

Heat diffusion on a ring leads to periodic boundary conditions in x. Other types of problems may be analyzed in an analogous way. An interesting example is the modeling of seasonal temperature variations in the earth as a function of depth. If we neglect daily fluctuations, a reasonable assumption is that the surface temperature of the earth is a periodic function of time with period equal to one year, and that the temperature at a depth x below the surface is also a periodic function of time. We further require that the temperature be bounded at large depths.

We choose a time unit so that $1\,\text{year} = 2\pi$, and a length unit so that the thermal conductivity of the earth, assumed constant, is equal to one. The temperature $u(x,t)$ then satisfies the following problem in $x > 0$, $t > 0$:

$$\begin{aligned} &u_t = u_{xx}, \\ &u(0,t) = f(t), \end{aligned}$$

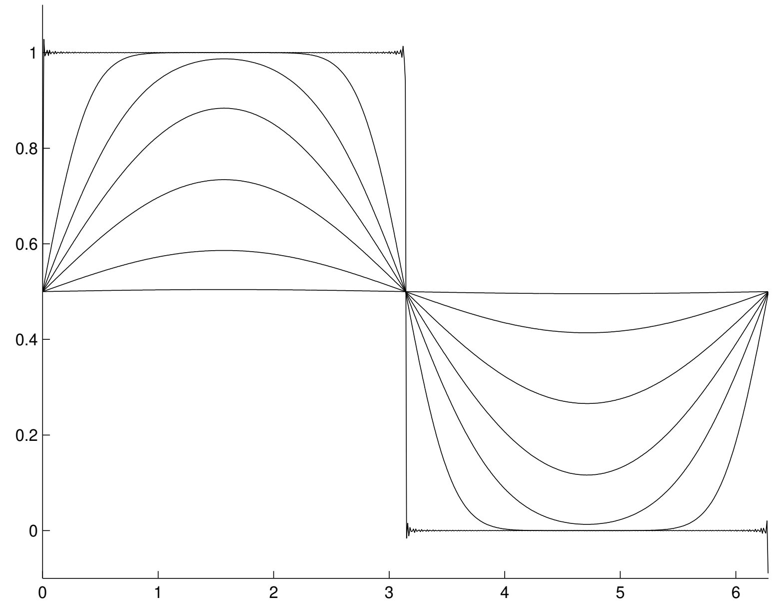

Fig. 7.2 The time evolution of the temperature distribution on a ring. The initial distribution is a step function. A truncated Fourier series is used to approximate the step function, and the Gibbs phenomenon can be seen near the points of discontinuity. The final distribution is uniform.

$$
\begin{aligned}
&u(\cdot, t) \in L^\infty([0, \infty)), \\
&u(x, t) = u(x, t + 2\pi).
\end{aligned}
$$

Here, $f(t)$ is a given real-valued, 2π-periodic function that describes the seasonal temperature variations at the earth's surface.

We expand the temperature $u(x, t)$ at depth x in a Fourier series in t,

$$u(x, t) = \sum_{n=-\infty}^{\infty} u_n(x) e^{int}.$$

Use of this expansion in the heat equation implies that the coefficients $u_n(x)$ satisfy

$$
\begin{aligned}
&-u_n'' + inu_n = 0, \qquad (7.22)\\
&u_n(0) = f_n, \\
&u_n \in L^\infty,
\end{aligned}
$$

where the prime denotes a derivative with respect to x, and f_n is the nth Fourier coefficient of f,

$$f_n = \frac{1}{2\pi} \int_{\mathbb{T}} f(t) e^{-int}\, dt.$$

The solution of (7.22) is

$$u_n(x) = \begin{cases} f_n \exp(\pm\sqrt{n}(1+i)x/\sqrt{2}) & \text{if } n > 0 \\ f_0 & \text{if } n = 0 \\ f_n \exp(\pm\sqrt{|n|}(1-i)x/\sqrt{2}) & \text{if } n < 0 \end{cases}.$$

The solutions with the plus sign in the exponent are excluded because they are unbounded as $x \to \infty$. The solution for $u(x,t)$ is therefore

$$\begin{aligned} u(x,t) &= f_0 + \sum_{n=1}^{\infty} f_n e^{-|n/2|^{1/2}x} e^{i(nt-|n/2|^{1/2}x)} \\ &\quad + \sum_{n=-\infty}^{-1} f_n e^{-|n/2|^{1/2}x} e^{i(nt+|n/2|^{1/2}x)}. \end{aligned} \tag{7.23}$$

For example, suppose that the surface temperature is given by a simple harmonic function

$$u(0,t) = a + b\sin t.$$

Then (7.23) may be written as

$$u(x,t) = a + b\exp\left(-\frac{x}{\sqrt{2}}\right)\sin\left(t - \frac{x}{\sqrt{2}}\right).$$

See Figure 7.3 for a graph of this solution. The exponential damping factor in front of the sine function describes a reduction in the magnitude of the variations in the earth's temperature below the surface. The argument of the sine function indicates that there is a depth-dependent phase shift in the temperature variations. At a depth $x = \sqrt{2}\pi$, the variations are reduced by a factor of $e^{-\pi} \approx 0.04$, and are opposite in phase to the surface temperature. For realistic numerical values of the thermal conductivity of the soil, this happens at a depth of about 13 feet. Thus, 13 feet below the surface the maximum temperature is reached in winter and minimum in summer! At this depth, the difference between winter and summer temperatures is reduced by a factor of about 25, as compared with the temperature difference at the surface. This reduction explains the usefulness of wine cellars, since it is important to store wine at a cool, uniform temperature.

7.4 Other partial differential equations

Fourier series may be used to study periodic solutions of any linear, constant coefficient partial differential equation. In this section, we consider a number of examples, including the wave equation and Laplace's equation, the two other classical linear partial differential equations of applied mathematics, in addition to the heat equation. The Fourier series may be interpreted either as classical solutions if they

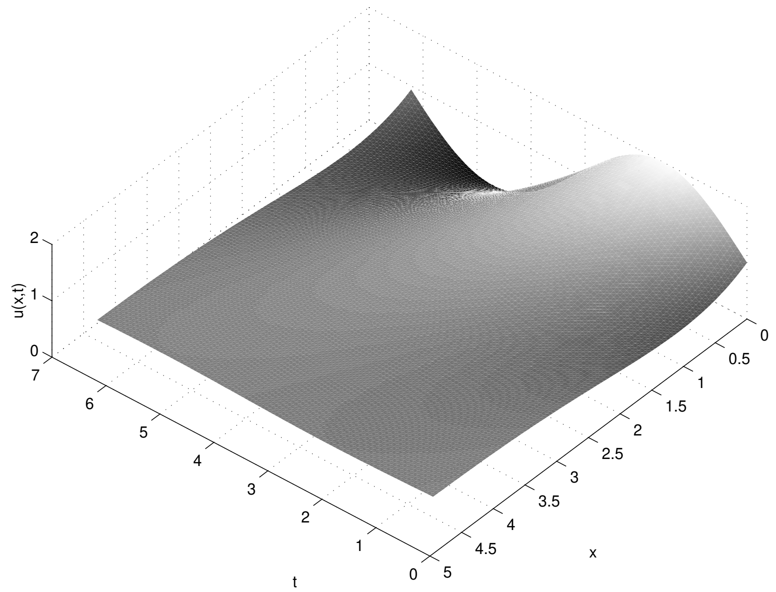

Fig. 7.3 The temperature of the earth as a function of time and depth. The time unit is one year divided by 2π. The unit of depth is roughly 1 meter. The temperature unit is arbitrary. The maximum and minimum temperature at the surface ($x = 0$) represent the maximum and minimum mean soil temperatures attained during summer and winter respectively.

converge sufficiently quickly to have continuous derivatives, or as weak solutions if they do not.

The one-dimensional *wave equation* is

$$u_{tt} = c^2 u_{xx}. \tag{7.24}$$

This equation describes the propagation of waves with a constant speed c, such as waves on an elastic string, sound waves, or light waves in a vacuum. The wave equation (7.24) is second order in time t, so we expect that two initial conditions are required to specify a unique solution. The initial value problem for wave propagation on a circle is

$$\begin{aligned} u_{tt} - c^2 u_{xx} &= 0, \\ u(x, 0) &= f(x), \\ u_t(x, 0) &= g(x), \end{aligned}$$

where $f, g \in L^2(\mathbb{T})$ are given functions.

The separated solutions of the wave equation (7.24), proportional to e^{inx}, are

$$u(x,t) = \left(ae^{inct} + be^{-inct}\right) e^{inx}$$

for $n \neq 0$, and

$$u(x,t) = a + bt$$

for $n = 0$. Superposing these solutions, the general solution is of the form

$$u(x,t) = a_0 + b_0 t + \sum_{n \neq 0} \left\{ a_n e^{in(x+ct)} + b_n e^{in(x-ct)} \right\}. \tag{7.25}$$

The constants a_n and b_n can be determined from the initial conditions as

$$a_0 = f_0, \quad b_0 = g_0, \quad a_n = \frac{1}{2}\left(f_n - \frac{i}{nc} g_n \right), \quad b_n = \frac{1}{2}\left(f_n + \frac{i}{nc} g_n \right),$$

where f_n and g_n are given by

$$f_n = \frac{1}{2\pi} \int_{\mathbb{T}} f(x) e^{-inx}, \quad g_n = \frac{1}{2\pi} \int_{\mathbb{T}} g(x) e^{-inx}.$$

In contrast with the heat equation, the solution exists for both $t > 0$ and $t < 0$, there is no smoothing of the initial data, and the solution does not converge to a stationary solution as $t \to \infty$.

The two-dimensional *Laplace equation* is

$$u_{xx} + u_{yy} = 0. \tag{7.26}$$

We will use Fourier series to solve a boundary value problem for Laplace's equation in the unit disc

$$\Omega = \left\{ (x,y) \mid x^2 + y^2 < 1 \right\}.$$

The Dirichlet problem consists of (7.26) in Ω with the boundary condition

$$u = f \qquad \text{on } \partial\Omega, \tag{7.27}$$

where $f : \mathbb{T} \to \mathbb{R}$ is a given function. In polar coordinates (r, θ) we may write (7.26)–(7.27) for $u(r,\theta)$ as

$$\begin{aligned} &\frac{1}{r}\left(r u_r \right)_r + \frac{1}{r^2} u_{\theta\theta} = 0 \qquad \text{in } r < 1, \\ &u(1,\theta) = f(\theta). \end{aligned}$$

The Laplace equation in polar coordinates has the separated solutions

$$u(r,\theta) = \left(a r^n + b r^{-n} \right) e^{in\theta} \qquad \text{for } n \in \mathbb{Z}.$$

The general solution of Laplace's equation that is bounded inside the unit disc is therefore

$$u(r,\theta) = \sum_{n=-\infty}^{\infty} a_n r^{|n|} e^{in\theta}. \tag{7.28}$$

The boundary condition implies that

$$a_n = \frac{1}{2\pi} \int_{\mathbb{T}} f(\theta) e^{-in\theta}\, d\theta.$$

Using the convolution theorem, we may write (7.28) in $r < 1$ as

$$u(r, \theta) = (g^r * f)(\theta),$$

where $g^r : \mathbb{T} \to \mathbb{R}$ is the *Poisson kernel*,

$$g^r(\theta) = \frac{1}{2\pi} \sum_{n=-\infty}^{\infty} r^{|n|} e^{in\theta}.$$

The geometric series for $n > 0$ and $n < 0$ may be summed to give

$$g^r(\theta) = \frac{1}{2\pi} \frac{1 - r^2}{1 - 2r\cos\theta + r^2}.$$

The series in (7.28) converges to an infinitely differentiable — in fact, analytic — function in $r < 1$ for any $f \in L^2(\mathbb{T})$, so the Laplace equation smoothes the boundary data.

In 1895 Korteweg and de Vries introduced a nonlinear PDE to describe water waves in shallow channels:

$$u_t = uu_x + u_{xxx}. \tag{7.29}$$

This *KdV equation* has exact localized traveling wave solutions called solitary waves, or solitons. A remarkable fact is that, in spite of its nonlinearity, the KdV equation can be solved exactly by the inverse scattering method introduced by Gardner, Greene, Kruskal, and Miura in 1967. This method depends on a surprising connection between the nonlinear KdV equation and a spectral problem for an associated linear operator (see Exercise 9.15). We will not discuss the inverse scattering method here, but we will use Fourier analysis to describe the *dispersive* property of the KdV equation.

If u is sufficiently small, then we do not expect the nonlinear term uu_x to influence the solution significantly, so we omit it in a first approximation. We therefore consider the linearized KdV equation,

$$u_t = u_{xxx}. \tag{7.30}$$

The general solution that is a 2π-periodic function of x is

$$u(x, t) = \sum_{n=-\infty}^{\infty} a_n e^{in(x - n^2 t)}.$$

Notice that the speed of propagation of e^{inx} depends on n, that is, on the wavelength. Since the components in the wave with different wavelengths propagate at

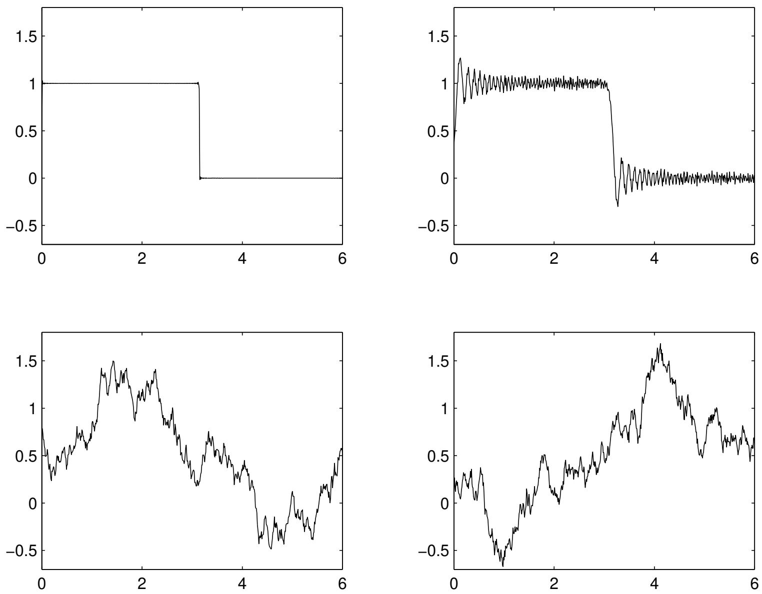

Fig. 7.4 The effect of dispersion is illustrated here with the solution (7.30) of the linearized KdV equation on a ring for times $t = 0, e^{-10}, e^{-2}$, and $t = e$. The initial condition is a step function.

different speeds, a wave generally spreads out or disperses; hence the name *dispersive* waves. In particular, a wave front does not maintain its shape while propagating. See Figure 7.4 for an illustration. Contrast this with the solution of the wave equation (7.25), where different Fourier components propagate at the same speed. The wave equation is said to be *nondispersive.* Another example of a dispersive wave equation, the Schrödinger equation from quantum mechanics, is discussed in Exercise 7.13.

7.5 More applications of Fourier series

The use of Fourier series is not restricted to differential equations. In this section, we consider two other applications.

The first is a solution of the isoperimetric problem, which states that of all closed curves of a given length, a circle encloses the maximum area. This result can also be stated as an inequality: for any closed curve of length L enclosing an area A, we have

$$4\pi A \leq L^2, \tag{7.31}$$

with equality if and only if the curve is a circle. Equation (7.31) is called the *isoperimetric inequality.* There are many different proofs of this result; the one we give, using Fourier series, is due to Hurwitz.

In order to state and prove a precise result, we reformulate the problem analytically. Without loss of generality, we consider curves whose lengths are normalized to 2π, and that are parametrized by arclength, s, positively oriented in the counterclockwise direction. We may represent such a smooth, closed curve Γ in the plane $\mathbb{R}^2$ by

$$(x, y) = (f(s), g(s)), \tag{7.32}$$

where $f, g : \mathbb{T} \to \mathbb{R}$ are continuously differentiable functions such that

$$\dot{f}(s)^2 + \dot{g}(s)^2 = 1. \tag{7.33}$$

Here, the dot denotes a derivative with respect to s.

Green's theorem states that if Ω is a region in the plane with a smooth, positively oriented boundary $\partial\Omega$ and $u, v : \overline{\Omega} \to \mathbb{R}$ are continuously differentiable functions, then

$$\int_\Omega \{u_x + v_y\}\, dxdy = \int_{\partial\Omega} \{udy - vdx\}.$$

If Γ does not intersect itself, then the use of Green's theorem with $u = x/2$ and $v = y/2$ implies that the area A enclosed by Γ is given by

$$A = \frac{1}{2}\int_\mathbb{T} \left\{f(s)\dot{g}(s) - g(s)\dot{f}(s)\right\} ds. \tag{7.34}$$

The expressions in (7.33) and (7.34) make sense for general functions $f, g \in H^1(\mathbb{T})$. Thus, an analytical formulation of the isoperimetric problem is to find functions $f, g \in H^1(\mathbb{T})$ that maximize the area functional A in (7.34) subject to the constraint (7.33).

Theorem 7.10 Suppose that a curve Γ is given by $x = f(s)$, $y = g(s)$, where $f, g \in H^1(\mathbb{T})$ are real-valued functions that satisfy (7.33), and the area A enclosed by Γ is given by (7.34). Then $A \le \pi$, with equality if and only if Γ is a circle.

Proof. We Fourier expand f and g as

$$f(s) = \frac{1}{\sqrt{2\pi}}\sum_{n=-\infty}^{\infty} \widehat{f}_n e^{ins}, \qquad g(s) = \frac{1}{\sqrt{2\pi}}\sum_{n=-\infty}^{\infty} \widehat{g}_n e^{ins}. \tag{7.35}$$

Since f and g are real valued, we have $\widehat{f}_{-n} = \overline{\widehat{f}_n}$ and $\widehat{g}_{-n} = \overline{\widehat{g}_n}$ for all n. Integration of (7.33) over $\mathbb{T}$ gives

$$2\pi = \int_\mathbb{T} \left\{\dot{f}(s)^2 + \dot{g}(s)^2\right\} ds.$$

From Parseval's theorem, this equation implies that

$$2\pi = \sum_{n=-\infty}^{\infty} n^2 \left\{\left|\widehat{f}_n\right|^2 + \left|\widehat{g}_n\right|^2\right\}, \tag{7.36}$$

and equation (7.34) implies that

$$2A = \sum_{n=-\infty}^{\infty} in \left\{ \overline{\widehat{f}_n} \widehat{g}_n - \widehat{f}_n \overline{\widehat{g}_n} \right\}.$$

Subtracting these series and rearranging the result, we find that

$$2\pi - 2A = \frac{1}{2} \sum_{n \neq 0} \left\{ \left| n\widehat{f}_n - i\widehat{g}_n \right|^2 + \left| n\widehat{g}_n + i\widehat{f}_n \right|^2 + \left(n^2 - 1\right) \left(\left| \widehat{f}_n \right|^2 + \left| \widehat{g}_n \right|^2 \right) \right\}.$$

Since the terms in the series on the right hand side of this equation are nonnegative, it follows that $A \leq \pi$. Moreover, we have equality if and only if $\widehat{f}_n = \widehat{g}_n = 0$ for $n \geq 2$, and $\widehat{f}_1 = i\widehat{g}_1$. Equation (7.36) implies that $|\widehat{f}_1| = \sqrt{\pi/2}$, so that

$$\widehat{f}_1 = \sqrt{\frac{\pi}{2}} e^{i\delta}, \qquad \widehat{g}_1 = -i\sqrt{\frac{\pi}{2}} e^{i\delta},$$

for some $\delta \in \mathbb{R}$. Writing $\widehat{f}_0 = \sqrt{2\pi} x_0$ and $\widehat{g}_0 = \sqrt{2\pi} y_0$, where $x_0, y_0 \in \mathbb{R}$, we find from (7.35) that

$$f(s) = x_0 + \cos(s + \delta), \qquad g(s) = y_0 + \sin(s + \delta).$$

Thus, if $A = \pi$, the curve $x = f(s)$, $y = g(s)$ is a circle. □

Our final application is an *ergodic theorem* for one of the simplest dynamical systems one can imagine, namely, rotations of the circle. We will prove another ergodic theorem for more general dynamical systems later on, in Theorem 8.37.

Let $\gamma \in \mathbb{R}$. We define a map $F_\gamma : \mathbb{T} \to \mathbb{T}$ on the circle $\mathbb{T}$ by

$$F_\gamma(x) = x + 2\pi\gamma. \tag{7.37}$$

This map is called the *circle map* or the *rotation map*. For every $x_0 \in \mathbb{T}$, the iterated application of F_γ generates a sequence of points $(x_n)_{n=0}^{\infty}$, where $x_n = F_\gamma^n(x_0)$. The set $\{x_n\}$ is called the *orbit* or *trajectory* of x_0 under F_γ. If γ is rational, then these points eventually repeat, and each orbit contains finitely many distinct points. If γ is irrational, then $x_m \neq x_n$ for $m \neq n$, and there are infinitely many points in each orbit (see Figure 7.5).

If $f : \mathbb{T} \to \mathbb{C}$ is a continuous function on $\mathbb{T}$, we define two averages of f, a *time average*

$$\langle f \rangle_{\mathrm{t}}(x_0) = \lim_{N \to \infty} \frac{1}{N+1} \sum_{n=0}^{N} f(x_n),$$

and a *phase-space average*,

$$\langle f \rangle_{\mathrm{ph}} = \frac{1}{2\pi} \int_{\mathbb{T}} f(x)\, dx.$$

This phase-space average may be regarded as a probabilistic average with respect to a uniform probability measure on $\mathbb{T}$. The following ergodic theorem, proved by Weyl in 1916, states that time averages and phase-space averages are equal when γ is irrational. This result is false when γ is rational.

Theorem 7.11 (Weyl ergodic) If γ is irrational, then

$$\langle f\rangle_{\mathrm{t}}(x_0) = \langle f\rangle_{\mathrm{ph}} \tag{7.38}$$

for all $f \in C(\mathbb{T})$ and all $x_0 \in \mathbb{T}$.

Proof. First, we show that (7.38) holds for the functions e^{imx} for each $m \in \mathbb{Z}$. If $m = 0$, then both averages are equal to 1. If $m \neq 0$, then $\langle e^{imx}\rangle_{\mathrm{ph}} = 0$, and the time average may be explicitly computed as follows:

$$\begin{aligned}
\langle e^{imx}\rangle_{\mathrm{t}} &= \lim_{N\to\infty} \frac{1}{N+1}\sum_{n=0}^{N} e^{im(x_0+2\pi n\gamma)} \\
&= \lim_{N\to\infty} \frac{1}{N+1} e^{imx_0}\sum_{n=0}^{N} [e^{2\pi im\gamma}]^n \\
&= \lim_{N\to\infty} \frac{1}{N+1} e^{imx_0}\left(\frac{1-[e^{2\pi im\gamma}]^{N+1}}{1-e^{2\pi im\gamma}}\right) \\
&= 0,
\end{aligned}$$

where we use the fact that $e^{2\pi im\gamma} \neq 1$ for irrational γ. Since both averages are linear in f, it follows that (7.38) holds for all trigonometric polynomials.

The trigonometric polynomials are dense in $C(\mathbb{T})$. Therefore, if $f \in C(\mathbb{T})$ and $\epsilon > 0$, then there is a trigonometric polynomial p such that $\|f - p\|_\infty \leq \epsilon$, and

$$\left|\frac{1}{N+1}\sum_{n=0}^{N} f(x_n) - \frac{1}{2\pi}\int_0^{2\pi} f(x)\,dx\right| \leq 2\epsilon + \left|\frac{1}{N+1}\sum_{n=0}^{N} p(x_n) - \frac{1}{2\pi}\int_0^{2\pi} p(x)\,dx\right|.$$

Taking the lim sup of this equation as $N \to \infty$, we obtain that

$$\limsup_{N\to\infty}\left|\frac{1}{N+1}\sum_{n=0}^{N} f(x_n) - \langle f\rangle_{\mathrm{ph}}\right| \leq 2\epsilon.$$

Since $\epsilon > 0$ is arbitrary, this proves (7.38) for all $f \in C(\mathbb{T})$ and all $x_0 \in \mathbb{T}$. □

A consequence of this ergodic theorem is the following result, which says that the points in an orbit $\{x_n \mid n \geq 0\}$ are uniformly distributed on the circle.

Corollary 7.12 Suppose that γ is irrational and I is an interval in $\mathbb{T}$ of length λ. Then

$$\lim_{N\to\infty} \frac{\#\{n \mid 0 \leq n \leq N,\, x_n \in I\}}{N+1} = \frac{\lambda}{2\pi}, \tag{7.39}$$

where $\#S$ denotes the number of points in the set S.

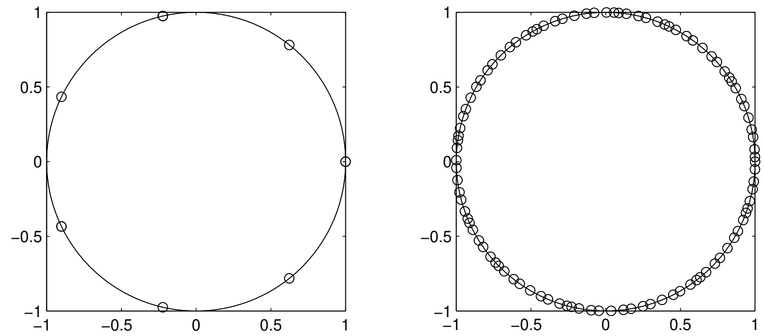

Fig. 7.5 The repeated images of the origin under the circle map, $F_\gamma^n(0)$, for $1 \le n \le 100$. On the left, $\gamma = 2/7$ is rational. On the right, $\gamma = (\sqrt{5}-1)/2$ is the golden ratio, which is irrational.

Proof. Let χ_I be the characteristic function of the interval I. Then (7.39) is equivalent to the statement that

$$\langle \chi_I \rangle_{\mathrm{t}} = \langle \chi_I \rangle_{\mathrm{ph}}. \tag{7.40}$$

This equation does not follow directly from Theorem 7.11 because χ_I is not continuous. We therefore approximate χ_I by continuous functions. We choose sequences (f_k) and (g_k) of nonnegative, continuous functions such that $f_k \le \chi_I \le g_k$ and

$$\int_{\mathbb{T}} f_k(x)\,dx \to \int_{\mathbb{T}} \chi_I(x)\,dx, \quad \int_{\mathbb{T}} g_k(x)\,dx \to \int_{\mathbb{T}} \chi_I(x)\,dx \qquad \text{as } k \to \infty.$$

We leave it to the reader to construct such sequences. Since $f_k \le \chi_I \le g_k$,

$$\frac{1}{N+1}\sum_{n=0}^{N} f_k(x_n) \le \frac{1}{N+1}\sum_{n=0}^{N} \chi_I(x_n) \le \frac{1}{N+1}\sum_{n=0}^{N} g_k(x_n).$$

Taking the limit as $N \to \infty$ of this equation, and applying Theorem 7.11 to the functions f_k and g_k, we obtain that

$$\begin{aligned} \frac{1}{2\pi}\int_{\mathbb{T}} f_k(x)\,dx &\le \liminf_{N\to\infty} \frac{1}{N+1}\sum_{n=0}^{N} \chi_I(x_n) \\ &\le \limsup_{N\to\infty} \frac{1}{N+1}\sum_{n=0}^{N} \chi_I(x_n) \le \frac{1}{2\pi}\int_{\mathbb{T}} g_k(x)\,dx. \end{aligned}$$

Letting $k \to \infty$, we find that

$$\begin{aligned} \frac{1}{2\pi}\int_{\mathbb{T}} \chi_I(x)\,dx &\le \liminf_{N\to\infty} \frac{1}{N+1}\sum_{n=0}^{N} \chi_I(x_n) \\ &\le \limsup_{N\to\infty} \frac{1}{N+1}\sum_{n=0}^{N} \chi_I(x_n) \end{aligned}$$

$$\leq \frac{1}{2\pi}\int_{\mathbb{T}} \chi_I(x)\, dx.$$

It follows that the limit defining the time average of χ_I exists and satisfies equation (7.40) for all $x_0 \in \mathbb{T}$. □

Theorem 7.11 actually holds for every $f \in L^1(\mathbb{T})$, except possibly for a set of initial points x_0 in $\mathbb{T}$ with zero Lebesgue measure. The proof, however, requires additional results from measure theory.

One application of the ergodic theorem is to the numerical integration of functions by the *Monte Carlo method*, in which one approximates the phase average, or integral, of f by a time average. This method is not required in the simple case of functions defined on a circle, but it is useful for the numerical integration of functions that depend on a large number of independent variables, where standard numerical integration formulae may become prohibitively expensive.

7.6 Wavelets

In this section, we introduce a special class of orthonormal bases of $L^2([0,1])$ and $L^2(\mathbb{R})$, called *wavelets*. These bases have proved to be very useful in signal analysis and data compression. With this application in mind, we will refer to the independent variable as a "time" variable. Wavelet bases in several independent variables are equally useful in image compression and many other applications.

Fourier expansions provide an efficient representation of stationary functions whose properties are invariant under translations in time. They are not as efficient in representing other types of functions, such as transient functions that vanish on most of their domain, or functions which vary much more rapidly at some times than at others. In the case of periodic functions, Parseval's identity in Theorem 6.26,

$$\|f\|^2 = \sum_n \left|\hat{f}_n\right|^2,$$

suggests that a large number of terms in a Fourier series expansion of f is needed if the quantity $\|f\|$ is distributed over a large number of coefficients $\hat{f}_n$ which are not too small. For example, from Lemma 7.8, this happens when f is discontinuous, so that its Fourier coefficients decay slowly as $n \to \infty$. Signals with sharp, or almost discontinuous, transitions and transient signals supported on a relatively small portion of the relevant time interval, such as the short beeps transmitted by a modem, are very common.

It is often useful to compress a signal before transmission or storage. To represent a function $f(t)$ accurately on the interval $0 \leq t \leq 1$ by storing a finite number of values $f(n\Delta t)$, where $n = 0, \ldots, N$ with $N = 1/\Delta t$, we need to choose Δt small enough that all rapid transitions can be reconstructed from this list of values. If the changes are rapid, then Δt has to be small and N has to be large. Suppose, however, that we have an orthonormal basis of $L^2([0,1])$ with the property that a finite linear combination of basis elements with M terms, where M is much smaller than N, yields a good approximation of the function f. Then we can store or transmit the function with M instead of N numbers without significant loss of information. Roughly speaking, we would then have compressed the data with a compression ratio of $N : M$. One reason for the use of wavelets in representing signals, or images, is that they allow for large compression ratios. There are many different kinds of wavelets, but all of them share the property that they describe a function at a sequence of different time, or length, scales. This allows us to represent a function efficiently by using wavelets whose local rate of variation is adapted to that of the function. We begin by describing a simple example, the *Haar wavelets.*

We define the *Haar scaling function* $\varphi \in L^2(\mathbb{R})$ by

$$\varphi(x) = \begin{cases} 1 & \text{if } 0 \le x < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{7.41}$$

The function φ is the characteristic function of the interval $[0,1)$, and is often referred to as a "box" function because of the shape of its graph. The basic *Haar wavelet*, or *mother wavelet*, $\psi \in L^2(\mathbb{R})$ is given by

$$\psi(x) = \begin{cases} 1 & \text{if } 0 \le x < 1/2, \\ -1 & \text{if } 1/2 \le x < 1, \\ 0 & \text{otherwise.} \end{cases} \tag{7.42}$$

These functions satisfy the scaling relations

$$\varphi(x) = \varphi(2x) + \varphi(2x-1), \tag{7.43}$$

$$\psi(x) = \varphi(2x) - \varphi(2x-1). \tag{7.44}$$

For $n, k \in \mathbb{Z}$, we define scaled translates $\varphi_{n,k}, \psi_{n,k} \in L^2(\mathbb{R})$ of φ, ψ by

$$\varphi_{n,k}(x) = 2^{n/2}\varphi\left(2^n x - k\right), \quad \psi_{n,k}(x) = 2^{n/2}\psi\left(2^n x - k\right). \tag{7.45}$$

First, consider the Hilbert space $L^2([0,1])$ with its usual inner product. For $n = 0, 1, 2, \ldots$, let V_n be the finite-dimensional subspace

$$V_n = \{f \mid f \text{ is constant on } [k/2^n, (k+1)/2^n) \text{ for } k = 0, \ldots, 2^n - 1\}. \tag{7.46}$$

Elements of V_n are step functions that are constant on intervals of length 2^{-n}. The value of $f \in V_n$ at the right endpoint $x = 1$ is irrelevant, since functions in $L^2([0,1])$ that are equal a.e. are equivalent. Clearly, we have $V_n \subset V_{n+1}$.

The function $\varphi_{n,k}$ is the characteristic function of the interval $[k2^{-n}, (k+1)2^{-n})$. The set

$$A_n = \{\varphi_{n,k} \mid 0 \le k \le 2^n - 1\}$$

is therefore a basis of V_n for each $n \ge 0$. Since $A_{n+1} \supset A_n$, the sets A_n are not disjoint, and we cannot form a basis of $\bigcup_{n\in\mathbb{N}} V_n$ by taking their union. Instead, for

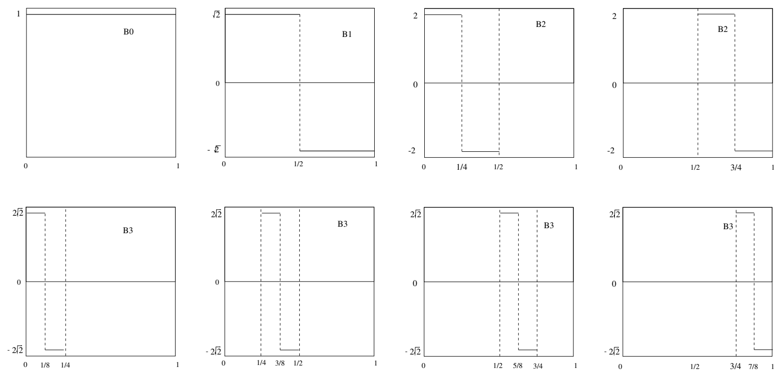

Fig. 7.6 Some members of the sets of functions B_n defined in (7.47).

$n = 0, 1, 2, \ldots$, we define subsets B_n of V_n by

$$B_0 = \{\varphi_{0,0}\}, \quad B_{n+1} = \{\psi_{n,k} \mid k = 0, 1, \ldots, 2^n - 1\}. \tag{7.47}$$

The subsets B_m and B_n are disjoint for $n \neq m$. The union of these sets, $B = \bigcup_{n=0}^{\infty} B_n$, or

$$B = \{\varphi_{0,0}\} \cup \{\psi_{n,k} \mid n = 0, 1, 2, \ldots, k = 0, 1, \ldots, 2^n - 1\}, \tag{7.48}$$

is called the *Haar wavelet basis* of $L^2([0,1])$. Some of these basis functions are illustrated in Figure 7.6.

Solving (7.43)–(7.44) for $\varphi(2x)$ and $\varphi(2x-1)$, we get

$$\varphi(2x) = \frac{1}{2}(\varphi(x) + \psi(x)),$$
$$\varphi(2x-1) = \frac{1}{2}(\varphi(x) - \psi(x)).$$

It follows by induction from dyadic dilations $x \mapsto 2x$ of these equations that $\varphi_{n,k}$ is a linear combination of $\varphi_{0,0}$ and $\psi_{m,k}$ with $m < n$. Hence the linear span of B contains bases A_n of V_n for every $n \in \mathbb{N}$. Using this fact, we can prove that B is a basis of $L^2([0,1])$.

Lemma 7.13 The set B in (7.48) is an orthonormal basis of $L^2([0,1])$.

Proof. It follows from Exercise 7.17 that B is an orthonormal set, so we just have to show that it is complete. Suppose that $f \in C([0,1])$ and $\epsilon > 0$. By Theorem 1.67, f is uniformly continuous, so there is an n such that $|f(x) - f(y)| \leq \epsilon$ for all $x, y \in [0,1]$ with $|x - y| \leq 2^{-n}$. We define the step function approximation $g \in V_n$ of f by

$$g(x) = \sum_{k=0}^{2^n - 1} f\left(k2^{-n}\right) \varphi_{n,k}(x).$$

Then g is in the linear span of $\bigcup_{m=0}^{n-1} B_m$, and

$$\sup_{0\le x<1} |f(x) - g(x)| \le \epsilon. \tag{7.49}$$

Thus any $f \in C([0,1])$ is the uniform limit of finite linear combinations of functions in B. Since the continuous functions are dense in $L^2([0,1])$, and the sup-norm is stronger than the L^2-norm, the orthonormal set B is complete in $L^2([0,1])$. □

We define the Haar wavelet basis B of $L^2(\mathbb{R})$ in a similar way, as

$$B = \{\psi_{n,k} \mid n \in \mathbb{Z}, k \in \mathbb{Z}\}, \tag{7.50}$$

where $\psi_{n,k}$ is defined in (7.45). This basis includes wavelets supported on intervals of arbitrarily large length, when n is large and negative, as well as on intervals of arbitrarily small length, when n is large and positive. The wavelet basis of $L^2(\mathbb{R})$ does not include a scaling function φ, in contrast with the wavelet basis (7.48) of $L^2([0,1])$.

Lemma 7.14 The set B in (7.50) is an orthonormal basis of $L^2(\mathbb{R})$.

Proof. The set B is orthonormal, so we just have to show that it is complete. Suppose that $f \in L^2(\mathbb{R})$ is orthogonal to B. Then f is orthogonal to all wavelets $\psi_{n,k}$ that are supported on any compact interval $[-2^N, 2^N]$. Since we can transform the interval $[-2^N, 2^N]$ to $[0,1]$ by a translation $x \mapsto x + 2^N$ and a dyadic dilation $x \mapsto 2^{-(N+1)}x$, and the basis B is invariant under such translations and dilations, it follows from Lemma 7.13 that f is constant on every compact interval $[-2^N, 2^N]$. Therefore f is constant on $\mathbb{R}$. Since the nonzero constant functions do not belong to $L^2(\mathbb{R})$, we conclude that $f = 0$, so B is complete. □

The Haar wavelets are very simple, compactly supported, orthonormal, step functions that take only three different values. Each wavelet is obtained by dilation and translation of a single basic wavelet ψ, derived from a scaling function φ. These properties make the Haar wavelets especially suitable for the representation of localized functions, as well as functions that vary on different lengthscales at different locations, and functions with a self-similar, fractal structure.

A drawback of the Haar wavelets is that they are discontinuous, so the partial sums approximating a continuous function are also discontinuous. It is often desirable to have continuous approximations of continuous functions, and C^p approximations of C^p functions. This is one motivation for the introduction of other wavelet bases. For definiteness, we consider wavelet bases of $L^2(\mathbb{R})$. The following *axioms of multiresolution analysis* capture the essential properties of the Haar wavelet basis that we want to generalize.

Definition 7.15 (Multiresolution analysis) A family $\{V_n \mid n \in \mathbb{Z}\}$ of closed linear subspaces of $L^2(\mathbb{R})$ and a function $\varphi \in L^2(\mathbb{R})$ are called a *multiresolution* *analysis* of $L^2(\mathbb{R})$ if the following properties hold:

(a) $f(x) \in V_n$ if and only if $f(2x) \in V_{n+1}$ for all $n \in \mathbb{Z}$ (scaling); (7.51)

(b) $V_n \subset V_{n+1}$ for all $n \in \mathbb{Z}$ (inclusion); (7.52)

(c) $\overline{\bigcup_{n\in\mathbb{Z}} V_n} = L^2(\mathbb{R})$ (density); (7.53)

(d) $\bigcap_{n\in\mathbb{Z}} V_n = \{0\}$ (maximality); (7.54)

(e) there is a function $\varphi \in L^2(\mathbb{R}) \cap L^1(\mathbb{R})$ such that $\{\varphi(x-k) \mid k \in \mathbb{Z}\}$ is an orthonormal basis of V_0 (basis). (7.55)

The five properties required in this definition are not independent. One can prove that (d) follows from (a), (b), and (e), and that, under the assumption that (a), (b), and (e) hold, (c) is equivalent to the property that $\widehat{\varphi}(0) \neq 0$, where $\widehat{\varphi}$ is the Fourier transform of φ. For brevity, we do not prove these statements here.

The spaces V_n defined in (7.46) and the function φ defined in (7.41) satisfy these axioms. We call φ the *scaling function* of the multiresolution analysis. We will explain how to obtain an orthonormal wavelet basis of $L^2(\mathbb{R})$ from this structure. When the scaling function φ is the box function, defined in (7.41), this procedure will reproduce the Haar wavelets, but other scaling functions lead to different orthonormal wavelet bases.

From (7.51) and (7.55) it follows that

$$A_n = \{2^{n/2}\varphi(2^n x - k) \mid k \in \mathbb{Z}\}$$

is an orthonormal basis of V_n for each $n \in \mathbb{Z}$. Since $V_n \subset V_{n+1}$, each function in A_n is a linear combination of functions of A_{n+1}, so the sets A_n are not linearly independent. To obtain linearly independent sets of functions, we define closed linear subspaces W_n of V_{n+1} by

$$V_{n+1} = V_n \oplus W_n.$$

The subspaces W_n are called *wavelet subspaces.* From their definition and the inclusion property (7.52), we see that W_m and W_n are orthogonal subspaces for $m \neq n$. Moreover, from Exercise 7.16, properties (7.52)–(7.54) imply that

$$\bigoplus_{n\in\mathbb{Z}} W_n = L^2(\mathbb{R}). \tag{7.56}$$

Now suppose that we have a function $\psi \in L^2(\mathbb{R})$, called a *wavelet*, such that

$$\{\psi(x-k) \mid k \in \mathbb{Z}\} \text{ is an orthonormal basis of } W_0.$$

Equation (7.58) below shows how the wavelet ψ is obtained from the scaling function φ. It then follows from the scaling axiom (7.51) that for each $n \in \mathbb{Z}$ the set

$$B_n = \left\{2^{n/2}\psi(2^n x - k) \mid k \in \mathbb{Z}\right\}$$

is an orthonormal basis of W_n so, from (7.56), their union

$$B = \left\{ 2^{n/2}\psi(2^n x - k) \mid n, k \in \mathbb{Z} \right\}$$

is an orthonormal basis of $L^2(\mathbb{R})$.

The axioms of multiresolution analysis impose severe restrictions on the scaling function. Translates of the scaling function must be orthogonal, and the function must be a linear combination of scaled translates of itself, meaning that there are constants c_k such that

$$\varphi(x) = \sum_{k\in\mathbb{Z}} c_k \varphi(2x - k). \tag{7.57}$$

For example, the Haar scaling function satisfies (7.43), so in that case $c_0 = c_1 = 1$ and $c_k = 0$ otherwise. For simplicity, we assume that $c_k \in \mathbb{R}$ and all but finitely many of the coefficients c_k are zero.

The basic wavelet ψ belongs to V_1 and is orthogonal to V_0. The following function satisfies these conditions:

$$\psi(x) = \sum_{k\in\mathbb{Z}} (-1)^k c_{1-k} \varphi(2x - k). \tag{7.58}$$

For example, in the case of the Haar wavelets, this equation gives (7.44). The function ψ in (7.58) clearly belongs to V_1, since it is a linear combination of the orthonormal basis elements $2^{1/2}\varphi(2x - k)$ of V_1. Moreover, for $j \in \mathbb{Z}$, we find from (7.57) and (7.58) that

$$\langle \varphi(x - j), \psi(x) \rangle = \frac{1}{2} \sum_{k\in\mathbb{Z}} (-1)^k c_{k-2j} c_{1-k}.$$

This sum is zero for every $j \in \mathbb{Z}$, since the change of summation variable from $k-2j$ to $1-k$ implies that

$$\sum_{k\in\mathbb{Z}} (-1)^k c_{k-2j} c_{1-k} = \sum_{k\in\mathbb{Z}} (-1)^{2j+1-k} c_{1-k} c_{k-2j} = -\sum_{k\in\mathbb{Z}} (-1)^k c_{k-2j} c_{1-k}.$$

Hence ψ is orthogonal to V_0. The translates $\{\psi(x - k) \mid k \in \mathbb{Z}\}$ form a basis of W_0, but we omit a proof of this fact here.

Next, we derive restrictions on the coefficients c_k in the scaling equation (7.57). We assume that the integral of φ is nonzero. It can, in fact, be shown that this is necessarily the case. By rescaling φ and x, we may assume without loss of generality that

$$\int_{\mathbb{R}} \varphi(x)\, dx = 1, \qquad \int_{\mathbb{R}} |\varphi(x)|^2\, dx = 1.$$

Changing variables $x \to 2x$, we see that

$$\int_{\mathbb{R}} \varphi(2x)\, dx = \frac{1}{2}, \qquad \int_{\mathbb{R}} |\varphi(2x)|^2\, dx = \frac{1}{2}.$$

Integration of (7.57) over $\mathbb{R}$ therefore implies that

$$\sum_{k \in \mathbb{Z}} c_k = 2. \tag{7.59}$$

Since $\{\varphi(2x-k) | k \in \mathbb{Z}\}$ is an orthogonal set, an application of Parseval's identity to (7.57) implies that

$$\sum_{k \in \mathbb{Z}} c_k^2 = 2. \tag{7.60}$$

The orthogonality of $\varphi(x-j)$ and $\varphi(x)$ for $j \neq 0$, together with (7.57), further imply that

$$\sum_{k \in \mathbb{Z}} c_{k-2j} c_k = 0 \qquad \text{for } j \in \mathbb{Z} \text{ and } j \neq 0. \tag{7.61}$$

Finally, it is often useful to require that several moments of the wavelet ψ vanish, meaning that

$$\int_{\mathbb{R}} x^m \psi(x)\, dx = 0 \qquad \text{for } m = 0, 1, \ldots, p-1. \tag{7.62}$$

The scaling coefficients c_k and the wavelet ψ must therefore satisfy (7.59)–(7.61). For example, the Haar wavelet coefficients $c_0 = 1$, $c_1 = 1$, and $c_k = 0$ for $k \neq 0, 1$ satisfy these conditions, and (7.62) with $p = 1$, but there are many other possible choices of the scaling coefficients.

One interesting choice, that satisfies (7.62) with $p = 2$, is due to Daubechies:

$$c_0 = \frac{1}{4}(1+\sqrt{3}), \qquad c_1 = \frac{1}{4}(3+\sqrt{3}),$$
$$c_2 = \frac{1}{4}(3-\sqrt{3}), \qquad c_3 = \frac{1}{4}(1-\sqrt{3}),$$

and $c_k = 0$ otherwise. We call the corresponding wavelet the D_4 wavelet. We can find the scaling function φ by regarding (7.57) as a fixed point equation and solving it iteratively, starting with the box function, for example, as an initial guess:

$$\varphi_{n+1}(x) = \sum_{k \in \mathbb{Z}} a_k \varphi_n(2x-k), \qquad n \geq 0,$$
$$\varphi_0(x) = \chi_{[0,1)}(x).$$

It is possible to show that φ_n converges to a continuous function φ whose support is the interval $[0, 3]$. There is no explicit analytical expression for φ, which is shown in Figure 7.7. As suggested by this figure, the D_4 scaling and wavelet functions φ and ψ are Hölder continuous (see Definition 12.72) but not differentiable.

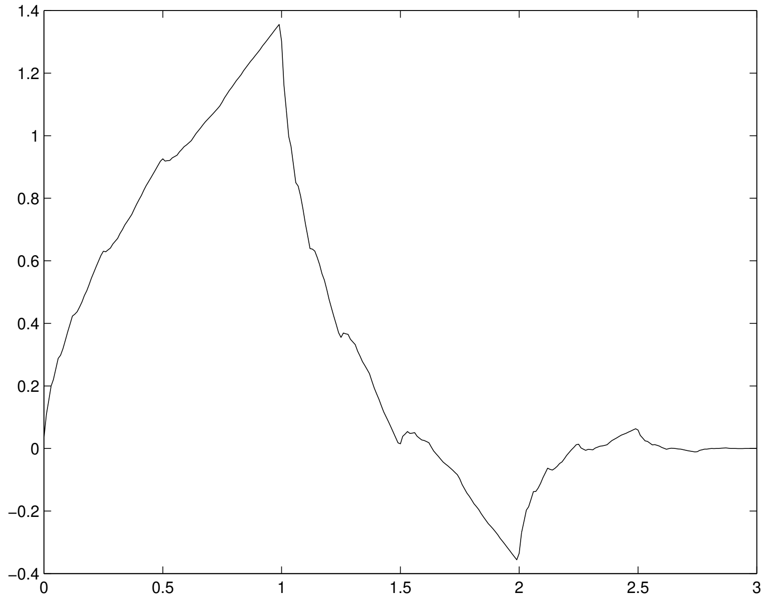

Fig. 7.7 The scaling function φ for the D_4 wavelets.

7.7 References

Beals [3] gives an elegant discussion of Fourier series, Hilbert spaces, and distributions. Rauch [44] discusses Fourier solutions of linear constant coefficient PDEs in more detail. See Whitham [56] for more on dispersive and nondispersive waves. Dym and McKean [10] contains a discussion of the Gibbs phenomenon, a proof of the isoperimetric inequality, and much more besides. Körner [29] is a wide-ranging introduction to the theory and applications of Fourier methods. In particular, it has a discussion of the Monte Carlo integration techniques mentioned in Section 7.5. There are many accounts of wavelets: for example, see Mallet [35]. Some algorithms for the numerical implementation of wavelets are described in [43].

7.8 Exercises

Exercise 7.1 Let φ_n be the function defined in (7.7).

(a) Prove (7.5).
(b) Prove that if the set $\mathcal{P}$ of trigonometric polynomials is dense in the space of periodic continuous functions on $\mathbb{T}$ with the uniform norm, then $\mathcal{P}$ is dense in the space of all continuous functions on $\mathbb{T}$ with the L^2-norm.
(c) Is $\mathcal{P}$ dense in the space of all continuous functions on $[0, 2\pi]$ with the uniform norm?

Exercise 7.2 Suppose that $f : \mathbb{T} \to \mathbb{C}$ is a continuous function, and

$$S_N = \frac{1}{\sqrt{2\pi}} \sum_{n=-N}^{N} \widehat{f}_n e^{inx}$$

is the Nth partial sum of its Fourier series.

(a) Show that $S_N = D_N * f$, where D_N is the *Dirichlet kernel*

$$D_N(x) = \frac{1}{2\pi} \frac{\sin\left[(N + 1/2)\, x\right]}{\sin(x/2)}.$$

(b) Let T_N be the mean of the first $N + 1$ partial sums,

$$T_N = \frac{1}{N+1} \left\{S_0 + S_1 + \ldots + S_N\right\}.$$

Show that $T_N = F_N * f$, where F_N is the *Fejér kernel*

$$F_N(x) = \frac{1}{2\pi(N+1)} \left(\frac{\sin\left[(N+1)x/2\right]}{\sin(x/2)} \right)^2.$$

(c) Which of the families (D_N) and (F_N) are approximate identities as $N \to \infty$? What can you say about the uniform convergence of the partial sums S_N and the averaged partial sums T_N to f?

Exercise 7.3 Prove that the sets $\{e_n \mid n \geq 1\}$ defined by

$$e_n(x) = \sqrt{\frac{2}{\pi}} \sin nx,$$

and $\{f_n \mid n \geq 0\}$ defined by

$$f_0(x) = \sqrt{\frac{1}{\pi}}, \quad f_n(x) = \sqrt{\frac{2}{\pi}} \cos nx \quad \text{for } n \geq 1,$$

are both orthonormal bases of $L^2([0, \pi])$.

Exercise 7.4 Let $T, S \in L^2(\mathbb{T})$ be the triangular and square wave, respectively, defined by

$$T(x) = |x|, \quad \text{if } |x| \leq \pi, \qquad S(x) = \begin{cases} 1 & \text{if } 0 < x < \pi, \\ -1 & \text{if } -\pi < x < 0. \end{cases}$$

(a) Compute the Fourier series of T and S.
(b) Show that $T \in H^1(\mathbb{T})$ and $T' = S$.
(c) Show that $S \notin H^1(\mathbb{T})$.

Exercise 7.5 Consider $f : \mathbb{T}^d \to \mathbb{C}$ defined by

$$f(\mathbf{x}) = \sum_{\mathbf{n}\in\mathbb{Z}^d} a_{\mathbf{n}} e^{i\mathbf{n}\cdot\mathbf{x}},$$

where $\mathbf{x} = (x_1, x_2, \ldots, x_d)$, $\mathbf{n} = (n_1, n_2, \ldots, n_d)$, and $\mathbf{n}\cdot\mathbf{x} = n_1x_1+n_2x_2+\ldots+n_dx_d$. Prove that if

$$\sum_{\mathbf{n}\in\mathbb{Z}^d} |\mathbf{n}|^{2k} |a_{\mathbf{n}}|^2 < \infty$$

for some $k > d/2$, then f is continuous.

Exercise 7.6 Suppose that $f \in H^1([a, b])$ and $f(a) = f(b) = 0$. Prove the *Poincaré inequality*

$$\int_a^b |f(x)|^2 \, dx \leq \frac{(b-a)^2}{\pi^2} \int_a^b |f'(x)|^2 \, dx.$$

Exercise 7.7 Solve the following initial-boundary value problem for the heat equation,

$$\begin{aligned} &u_t = u_{xx}, \\ &u(0,t) = 0, \quad u(L,t) = 0 \qquad \text{for } t > 0, \\ &u(x,0) = f(x) \qquad\qquad\quad \text{for } 0 \leq x \leq L. \end{aligned}$$

Exercise 7.8 Find sufficient conditions on the coefficients a_n and b_n in the solution (7.25) of the wave equation so that $u(x,t)$ is a twice continuously differentiable function of x for all $t \in \mathbb{R}$.

Exercise 7.9 Suppose that $u(x,t)$ is a smooth solution of the one-dimensional wave equation,

$$u_{tt} - c^2 u_{xx} = 0.$$

Prove that

$$\left(u_t^2 + c^2 u_x^2\right)_t - \left(2c^2 u_t u_x\right)_x = 0.$$

If $u(0,t) = u(1,t) = 0$ for all t, deduce that

$$\int_0^1 |u_t(x,t)|^2 + c^2 |u_x(x,t)|^2 \, dx = \text{constant}.$$

Exercise 7.10 Show that

$$u(x,t) = f(x+ct) + g(x-ct)$$

is a solution of the one-dimensional wave equation,

$$u_{tt} - c^2 u_{xx} = 0,$$

for arbitrary functions f and g, This solution is called *d'Alembert's solution.*

Exercise 7.11 Let $\Omega = \{(r,\theta) \mid r < 1\}$ be the unit disc in the plane, where (r,θ) are polar coordinates. The boundary of Ω is the unit circle $\mathbb{T}$. Let $u(r,\theta)$ be a solution of Laplace's equation in Ω,

$$\frac{1}{r}\left(ru_r\right)_r + \frac{1}{r^2}u_{\theta\theta} = 0 \qquad r < 1,$$

and define $f, g \in L^2(\mathbb{T})$ by $f(\theta) = u_\theta(1,\theta)$, $g(\theta) = u_r(1,\theta)$. Show that $g = \mathbb{H}f$ where $\mathbb{H}$ is the periodic Hilbert transform, defined in Example 8.32.

Exercise 7.12 Show that there is initial data $f \in C^\infty(\mathbb{T})$ for which the initial value problem for Laplace's equation,

$$\begin{aligned} &u_{tt} + u_{xx} = 0, \\ &u(x,0) = f(x), \\ &u_t(x,0) = 0, \end{aligned}$$

has no solution with $u(\cdot,t) \in L^2(\mathbb{T})$ in any interval $|t| < \delta$, where $\delta > 0$. (The initial-value problem for Laplace's equation is therefore ill-posed.)

Exercise 7.13 Use Fourier series to solve the following initial-boundary value problem for the Schrödinger equation (6.14), that describes a quantum mechanical particle in a box:

$$iu_t = -u_{xx} \tag{7.63}$$

$$u(0,t) = u(1,t) = 0 \quad \text{for all } t, \tag{7.64}$$

$$u(x,0) = f(x). \tag{7.65}$$

Derive the following two conservation laws from your Fourier series solution and directly from the PDE:

$$\frac{d}{dt}\int_0^1 |u(x,t)|^2 \, dx = 0, \qquad \frac{d}{dt}\int_0^1 |u_x|^2 \, dx = 0.$$

Exercise 7.14 Consider the logistic map

$$x_{n+1} = 4\mu x_n(1 - x_n),$$

where $x_n \in [0,1]$, and $\mu = 1$. Show that the solutions may be written as $x_n = \sin^2\theta_n$ where $\theta^n \in \mathbb{T}$, and

$$\theta_{n+1} = 2\theta_n.$$

What can you say about the orbits of the logistic map, the existence of an invariant measure, and the validity of an ergodic theorem?

Exercise 7.15 Consider the dynamical system on $\mathbb{T}$,

$$x_{n+1} = \alpha x_n \pmod 1,$$

where $\alpha = (1+\sqrt{5})/2$ is the golden ratio. Show that the orbit with initial value $x_0 = 1$ is not equidistributed on the circle, meaning that it does not satisfy (7.39).
HINT: Show that

$$u_n = \left(\frac{1+\sqrt{5}}{2}\right)^n + \left(\frac{1-\sqrt{5}}{2}\right)^n$$

satisfies the difference equation

$$u_{n+1} = u_n + u_{n-1}$$

and hence is an integer for every $n \in \mathbb{N}$. Then use the fact that

$$\left(\frac{1-\sqrt{5}}{2}\right)^n \to 0 \quad \text{as } n \to \infty.$$

Exercise 7.16 If $\{V_n \mid n \in \mathbb{Z}\}$ is a family of closed subspaces of $L^2(\mathbb{R})$ that satisfies the axioms of multiresolution analysis, and $V_{n+1} = V_n \oplus W_n$, prove that

$$L^2(\mathbb{R}) = \bigoplus_{n\in\mathbb{Z}} W_n.$$

(See Exercise 6.5 for the definition of an infinite direct sum.)

Exercise 7.17 Let B_n and V_n be as defined in (7.46) and (7.47). Prove that $\bigcup_{n=0}^N B_n$ is an orthonormal basis of V_N.
HINT. Prove that the set is orthonormal and count its elements.

Exercise 7.18 Suppose that $B = \{e_n(x)\}_{n=1}^{\infty}$ is an orthonormal basis for $L^2([0,1])$. Prove the following:

(a) For any $a \in \mathbb{R}$, the set $B_a = \{e_n(x-a)\}_{n=1}^{\infty}$ is an orthonormal basis for $L^2([a, a+1])$.
(b) For any $c > 0$, the set $B^c = \{\sqrt{c}e_n(cx)\}_{n=1}^{\infty}$ is an orthonormal basis for $L^2([0, c^{-1}])$.
(c) With the convention that functions are extended to a larger domain than their original domain by setting them equal to 0, prove that $B \cup B_1$ is an orthonormal basis for $L^2([0,2])$.
(d) Prove that $\bigcup_{k\in\mathbb{Z}} B_k$ is an orthonormal basis for $L^2(\mathbb{R})$.