X. Introduction to Wavelets, 434-544

DOI: 10.3792/euclid/9781429799911-10

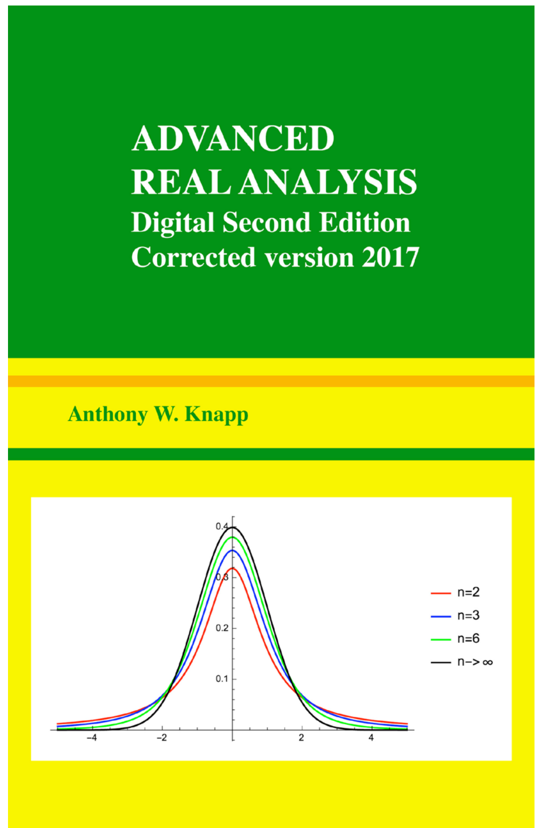
from

Advanced Real Analysis

Digital Second Edition
Corrected version 2017

Anthony W. Knapp

Full Book DOI: 10.3792/euclid/9781429799911
ISBN: 978-1-4297-9991-1



Anthony W. Knapp
81 Upper Sheep Pasture Road
East Setauket, N.Y. 11733–1729, U.S.A.
Email to: aknapp@math.stonybrook.edu
Homepage: www.math.stonybrook.edu/∼aknapp

Title: Advanced Real Analysis
Cover: Normal distribution as a limit of Gosset's t distribution; see page 421.

Mathematics Subject Classification (2010): 46–01, 42–01, 43–01, 35–01, 34–01, 47–01, 58–01, 60A99, 60F05, 28C10, 42C40, 65T60.

First Edition, ISBN-13 978-0-8176-4382-9

Published by Birkhäuser Boston

Digital Second Edition, not to be sold, no ISBN

Published by the Author

CHAPTER X

Introduction to Wavelets

Abstract. This chapter introduces the relatively recent subject of wavelets, which is an outgrowth of Fourier analysis in mathematics and signal processing in engineering. Except in one case, construction of examples of wavelets tends to be difficult. Much of the chapter is devoted to construction of some of the better known examples and lists of their most important properties.

Section 1 defines wavelets and discusses three features of traditional Fourier analysis: the Uncertainty Principle, Gibbs phenomenon, and the Shannon Sampling Theorem. It ends with a brief essay on the need for wavelets in various applications.

Section 2 establishes that the Haar system is an orthonormal basis of $L^2(\mathbb{R})$. The Haar wavelet predates the general theory of wavelets by many decades but provides a prototype for some of the development. The section ends with some discussion of convergence of one-sided Haar expansions for function spaces besides $L^2(\mathbb{R})$.

Section 3 begins the general theory of wavelets, introducing the notion of multiresolution analysis to abstract the construction in Section 2 of the Haar wavelet. The ingredients of a multiresolution analysis are a scaling function, traditionally called φ, and an increasing sequence of closed subspaces V_j of $L^2(\mathbb{R})$ with certain properties. The wavelet that is constructed is traditionally called ψ.

Section 4 introduces the Shannon wavelet, whose construction is immediate from the theory of multiresolution analyses. The new ingredient here, beyond the ideas used for the Haar wavelet, is the careful use of the generating function of the scaling function to obtain a formula for the wavelet.

Section 5 supplements the theory of Section 3 by showing how to build a multiresolution analysis out of a candidate for the scaling function.

Section 6 introduces the Meyer wavelets, each of which is smooth and has Fourier transform of a prescribed order of differentiability. The full theory of Sections 3 and 5 is used in their construction.

Section 7 introduces splines, examines one example, and sees the need for more theory. It develops one further aspect of the general theory, showing how to replace a "Riesz system" with an orthonormal set. It therefore allows one to relax the conditions needed in Section 5 for a function to be a scaling function. In addition, it uses elementary complex analysis to prove a series expansion for $\pi^2/\sin^2 \pi z$ that is needed in Section 8.

Section 8 continues the discussion of the role of splines in the theory of wavelets, introducing the Battle–Lemarié wavelets. As with the Meyer wavelets each is smooth and has Fourier transform of a prescribed order of differentiability.

Section 9 develops the Daubechies wavelets. These have compact support, but except for the first one, neither they nor their Fourier transforms have known formulas in closed form. The construction begins by pinpointing necessary conditions on the generating function.

Section 10 deals with smoothness questions. It contains three results. The first gives an estimate for the decay of the Fourier transform of the Daubechies scaling function of each order. The second deduces a certain amount of differentiability of a scaling function from the estimate in the first result. The third shows in the converse direction that a Daubechies wavelet can never be of class

C^∞. The section concludes with a table summarizing properties of the specific wavelets that have been constructed in Sections 2–9.

Section 11 gives a quick introduction to applications. It discusses the discrete wavelet transform and its use in storage and compression of data, it identifies some applications of wavelets in one and two dimensions, and it makes brief remarks about some of the applications.

1. Introduction

For current purposes a **wavelet** is a function ψ in $L^2(\mathbb{R})$ such that the functions

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$$

form an orthonormal basis of $L^2(\mathbb{R})$ as j and k range through $\mathbb{Z}$. Such wavelets are called "orthogonal" and "one-dimensional" by some authors. We postpone consideration of modifications of this definition to Section 11 at the end of this chapter.

Corresponding to a wavelet ψ is the wavelet expansion

$$f(x) = \sum_{j,k\in\mathbb{Z}} \Big(\int_{\mathbb{R}} f(y)\overline{\psi_{j,k}(y)}\,dy\Big)\psi_{j,k}(x),$$

the series being understood as convergent in $L^2(\mathbb{R})$, independently of the order of the terms. Wavelet expansions allow one to isolate certain hidden features of functions, much as Fourier expansions do, the particular features depending on properties of ψ. Examples of wavelets and their properties will be discussed beginning in Section 2.

To have a practical guide for the theory, it may be helpful to regard a given L^2 function as a "signal," a function of one real variable t that represents time. The values of the function represent a voltage, positive or negative, or perhaps a mechanical analog of such a voltage. Let us digress for the time being to consider the role of traditional Fourier analysis in understanding such a signal. If we consider the function on all of $\mathbb{R}$, the Fourier transform has frequency as its variable. Alternatively we can think of masking the function, looking at only those values of t in an interval. If we take that interval to have length 2π, then we can form the Fourier series of the restricted function, and the variable in the result will be the subscript on the Fourier coefficients, telling what multiple of a fundamental frequency is under consideration. Or if we allow ourselves intervals of a more general length T, then we get the kind of Fourier series appropriate to functions of period T.

A two-dimensional analog of a signal is a representation of a two-dimensional picture. If the picture is in black and white, the value of the function at a point can be the intensity of light at that point. If the picture is in color, the value of the function can be a three-dimensional vector representation of the red, green, and blue values at that point, or by a change of variables, the luminance, the blue chrominance, and the red chrominance values at that point.

Let us begin by mentioning three relevant features of traditional Fourier analysis in this setting. The Fourier transform on $L^2(\mathbb{R})$ is the continuous extension $\mathcal{F}$ from $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ to $L^2(\mathbb{R})$ of the linear function

$$f \mapsto \widehat{f} \qquad \text{given by} \qquad \widehat{f}(\omega) = \int_{\mathbb{R}} f(t) e^{-2\pi i t\omega}\, dt.$$

Then $\mathcal{F}$ is a continuous linear function from $L^2(\mathbb{R})$ onto $L^2(\mathbb{R})$, and it satisfies $\|\mathcal{F}f\|_2 = \|f\|_2$ for all $f \in L^2(\mathbb{R})$. Normally we use $\mathcal{F}$ to refer to all versions of the Fourier transform, completely avoiding using the notation $\widehat{\ }$.

The first feature of traditional Fourier analysis is the **Uncertainty Principle**, which says in effect that the amount of detail detected in the time domain limits the amount of detail detectable in the frequency domain. Since our interest is only in illustrating the principle, it will not be necessary for us to seek maximum generality. Accordingly we formulate this result as follows. Fix a member f other than 0 in the Schwartz space $\mathcal{S}(\mathbb{R})$. Then $tf(t)$ and $\omega\,\mathcal{F}f(\omega)$ are in $\mathcal{S}(\mathbb{R})$, and we can define "mean values" t_0 and ω_0 of t and ω by[1]

$$t_0 = \|f\|_2^{-2} \int_{\mathbb{R}} t|f(t)|^2\, dt \qquad \text{and} \qquad \omega_0 = \|f\|_2^{-2} \int_{\mathbb{R}} \omega\, |\mathcal{F}f(\omega)|^2\, d\omega,$$

as well as variances in t and ω by

$$\sigma_{f,t}^2 = \|f\|_2^{-2} \int_{\mathbb{R}} (t-t_0)^2 |f(t)|^2\, dt, \quad \sigma_{f,\omega}^2 = \|f\|_2^{-2} \int_{\mathbb{R}} (\omega-\omega_0)^2 |\mathcal{F}f(\omega)|^2\, d\omega.$$

Proposition 10.1 (Uncertainty Principle). If f is a nonzero member of $\mathcal{S}(\mathbb{R})$, then

$$\sigma_{f,t}^2 \sigma_{f,\omega}^2 \geq \frac{1}{16\pi^2}.$$

REMARKS. The quantity $\sigma_{f,t}$ is to be regarded as a measure of the time duration of the signal $f(t)$, and $\sigma_{f,\omega}$ is to be regarded as a measure of the frequency dispersion or bandwidth of the signal. The inequality in the proposition is not the only limitation on the time duration of $f(t)$ and the dispersion of $\mathcal{F}f(\omega)$; for example, if f has compact support, then $\mathcal{F}f$ extends to be analytic in the whole complex plane, and the zeros of $\mathcal{F}f$ therefore cannot have a limit point anywhere on the real axis unless $f = 0$.

[1] These are the mean values, in the sense of Chapter IX, of the random variables t and ω relative to the probability distributions $\|f\|_2^{-2}|f(t)|^2\, dt$ and $\|f\|_2^{-2}|\mathcal{F}f(\omega)|^2\, dw$, respectively, and similarly for the variances lower down.

PROOF. We treat first the special case that $t_0 = \omega_0 = 0$. Let f' be the derivative of f. Since $\mathcal{F}f'(\omega) = 2\pi i\omega\,\mathcal{F}f(\omega)$, we have

$$\begin{aligned}\|f\|_2^4\,\sigma_{f,t}^2\sigma_{f,\omega}^2 &= \int_{\mathbb{R}} |tf(t)|^2\,dt \int_{\mathbb{R}} |\omega\,\mathcal{F}f(\omega)|^2\,d\omega \\ &= (2\pi)^{-2} \int_{\mathbb{R}} |tf(t)|^2\,dt \int_{\mathbb{R}} |\mathcal{F}f'(\omega)|^2\,d\omega \\ &= (2\pi)^{-2} \int_{\mathbb{R}} |tf(t)|^2\,dt \int_{\mathbb{R}} |f'(t)|^2\,dt \\ &\geq (2\pi)^{-2} \Big| \int_{\mathbb{R}} \big(tf(t)\overline{f'(t)}\big)\,dt \Big|^2 \\ &= (2\pi)^{-2} \Big| \int_{\mathbb{R}} \tfrac{1}{2}t\big(f(t)\overline{f'(t)} + f'(t)\overline{f(t)}\big)\,dt \Big|^2 \\ &= (2\pi)^{-2} \Big| \int_{\mathbb{R}} \tfrac{1}{2}t\tfrac{d}{dt}\big(f(t)\overline{f(t)}\big)\,dt \Big|^2,\end{aligned}$$

the inequality holding by the Schwarz inequality. In the integral on the right side, we integrate by parts, differentiating $\frac{1}{2}t$ and integrating $\frac{d}{dt}\big(f(t)\overline{f(t)}\big)$. The integrated term is 0 because f is a Schwartz function, and the right side simplifies to

$$= \tfrac{1}{4}(2\pi)^{-2} \Big| \int_{\mathbb{R}} -|f(t)|^2\,dt \Big|^2 = \tfrac{1}{4}(2\pi)^{-2}\|f\|_2^4.$$

Thus

$$\sigma_{f,t}^2\sigma_{f,\omega}^2 \geq \frac{1}{16\pi^2},$$

and the proof is complete under the assumption that $t_0 = \omega_0 = 0$.

For the general case, we consider the function

$$g(t) = e^{-2\pi i\omega_0 t} f(t + t_0),$$

which has $\|g\|_2 = \|f\|_2$ and $\mathcal{F}g(\omega) = e^{2\pi i(\omega+\omega_0)(t_0)}\mathcal{F}f(\omega + \omega_0)$. For g, the mean value of t is

$$\begin{aligned}\mu_{g,t} &= \|f\|_2^{-2} \int_{\mathbb{R}} t|g(t)|^2\,dt \\ &= \|f\|_2^{-2} \int_{\mathbb{R}} t|f(t + t_0)|^2\,dt \\ &= \|f\|_2^{-2} \int_{\mathbb{R}} (t - t_0)|f(t)|^2\,dt = t_0 - t_0 = 0,\end{aligned}$$

and the mean value of ω is

$$\begin{aligned}\mu_{g,\omega} &= \|f\|_2^{-2} \int_{\mathbb{R}} \omega\,|\mathcal{F}g(\omega)|^2\,d\omega \\ &= \|f\|_2^{-2} \int_{\mathbb{R}} \omega\,|\mathcal{F}f(\omega + \omega_0)|^2\,d\omega \\ &= \|f\|_2^{-2} \int_{\mathbb{R}} (\omega - \omega_0)|\mathcal{F}f(\omega)|^2\,d\omega = \omega_0 - \omega_0 = 0.\end{aligned}$$

The special case therefore applies to g. The variance in t for g is

$$\begin{aligned}\sigma_{g,t}^2 = \|g\|_2^{-2} \int_{\mathbb{R}} (t - 0)^2|g(t)|^2\,dt &= \|f\|_2^{-2} \int_{\mathbb{R}} t^2|f(t + t_0)|^2\,dt \\ &= \|f\|_2^{-2} \int_{\mathbb{R}} (t - t_0)^2|f(t)|^2\,dt = \sigma_{f,t}^2,\end{aligned}$$

and similarly the variance $\sigma_{g,\omega}^2$ in ω for g equals $\sigma_{f,\omega}^2$. Hence the conclusion in the special case of g gives the desired inequality for f. □

The second feature of traditional Fourier analysis is **Gibbs phenomenon**. Informally this is the statement that the partial sums of the Fourier series of a nice real-valued function with a jump discontinuity overshoot the expected limit, above and below, by about 9%. The more precise statement is in Proposition 10.2.

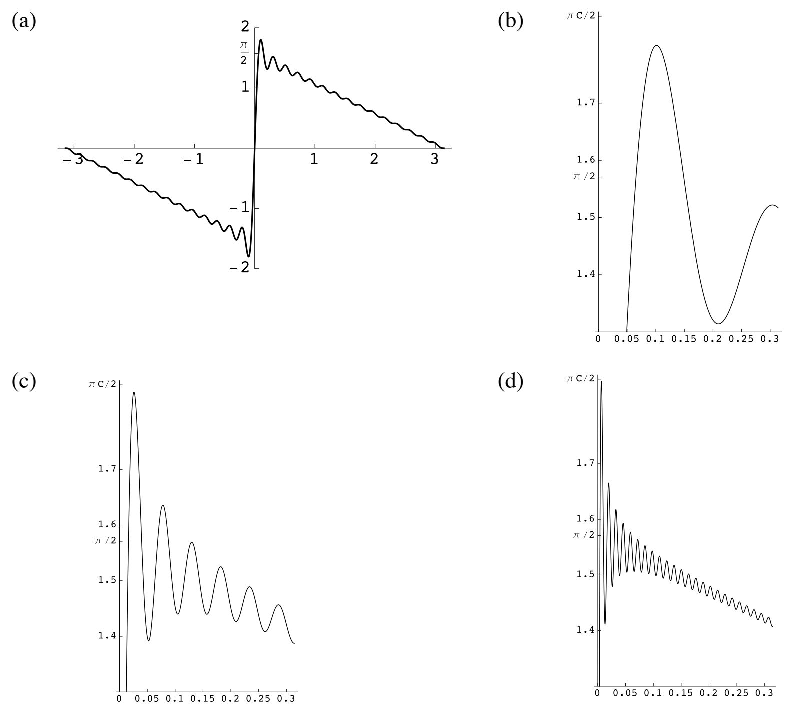


FIGURE 10.1. Gibbs phenomenon for the partial sums of the Fourier series $s_N(f;x)$ when f is continued from $\frac{1}{2}(\pi - x)$ with $0 < x < 2\pi$ so as to have period 2π. (a) Graph of $s_{30}(f;x)$. (b) Detail in graph of $s_{30}(f;x)$, C being ≈ 1.179. (c) Same detail in graph of $s_{120}(f;x)$. (d) Same detail in graph of $s_{480}(f;x)$.

Proposition 10.2 (Gibbs phenomenon). Let f be a real-valued periodic function of period 2π. Suppose that f is of bounded variation and has an isolated jump discontinuity when $t = t_0$; specifically suppose that the jump $J = \lim_{t\downarrow t_0} f(t) - \lim_{t\uparrow t_0} f(t)$ is > 0. If $s_N(f;t)$ is the partial sum of the Fourier series of f at t, then

$$\lim_{N\to\infty} s_N\big(f; t_0 + \tfrac{\pi}{N}\big) - \lim_{N\to\infty} s_N(f; t_0 - \tfrac{\pi}{N}) = CJ, \quad \text{the limits existing,}$$

where $C = \frac{2}{\pi}\int_0^{\pi} (\sin t)/t\, dt \approx 1.179$.

REMARKS. Thus the overshoot above and below is at least by half of .179 times the size of the jump. The overshoot is illustrated in Figure 10.1, which shows a graph of the 30^{th} partial sum, together with enlargements of a certain portion of the graphs of of the 30^{th}, 120^{th}, and 480^{th} partial sums.

PROOF. Let $g(t)$ be the periodic extension of the function that equals $\frac{1}{2}(\pi - t)$ for $0 < t < 2\pi$. This is periodic of period 2π, is of bounded variation, and is continuous except at the multiples of 2π, where it jumps by π. Thus the function h with $h(t) = f(t) - \pi^{-1} J g(t - t_0)$ is of bounded variation and is continuous is an interval about $t = t_0$. By Theorem 6.55 of *Basic*, the Fourier series of h converges uniformly to h in an interval about t_0. Consequently it is enough to consider the function g and the point t_0; in this case the jump is $J = \pi$.

From Section I.10 of *Basic*, the Fourier series of g is $\sum_{n=1}^{\infty} \dfrac{\sin nt}{n}$. If $s_N(g; t)$ denotes the N^{th} partial sum of this series, then

$$s_N(g; \pi/N) = \sum_{n=1}^{N} \frac{\sin(n\pi/N)}{n},$$

which is a Riemann sum for the Riemann integral of the continuous function $(\sin t)/t$ from 0 to π. The partition is equally spaced with mesh π/N, and the function is evaluated at the right-hand endpoint of each interval of the partition. By Theorem 1.35 of *Basic*,

$$\lim_{N\to\infty} s_N(g; \pi/N) = \int_0^{\pi} \tfrac{\sin t}{t}\, dt = \tfrac{\pi}{2} C,$$

with C as in the statement of the proposition. Since the periodic function $g(x)$ is odd, we similarly obtain

$$\lim_{N\to\infty} s_N(g; -\pi/N) = -\tfrac{\pi}{2} C.$$

Therefore

$$\lim_{N\to\infty} s_N(g; \pi/N) - \lim_{N\to\infty} s_N(g; -\pi/N) = \pi C,$$

as asserted. □

The third feature of traditional Fourier analysis concerns "sampling." As a practical matter, signals in engineering cannot be expected to be known exactly. Realistic information about the signal can be obtained only at certain instants of time, possibly instants that are very close to one another.[2] To what extent does

[2] Actually what seems to be the value at an instant of time may really be a kind of average over a very short interval, but we shall ignore this distinction.

the sampled version of the signal determine the signal exactly? The **Shannon Sampling Theorem** gives an answer. It assumes that the signal is "band-limited," i.e., its Fourier transform vanishes outside some interval. This is a reasonable assumption in practice, since each piece of equipment for studying signals has its own limitations in coping with high frequencies.

Proposition 10.3 (Shannon Sampling Theorem). Suppose that f is a member of $L^2(\mathbb{R})$ whose Fourier transform $\mathcal{F}f$ vanishes outside the interval $[-\frac{1}{2}\Omega, \frac{1}{2}\Omega]$. Then f can be taken to be smooth and satisfies

$$f(t) = \sum_{k=-\infty}^{\infty} f\Big(\frac{k}{\Omega}\Big) \frac{\sin(\pi(\Omega t - k))}{\pi(\Omega t - k)},$$

the series being convergent in $L^2(\mathbb{R})$ and also uniformly convergent in t.

PROOF. Until the very end we assume that $\Omega = 1$. Since $\mathcal{F}f$ is in $L^2(\mathbb{R})$ and has compact support, it is in $L^1(\mathbb{R})$, and we can recover f from it almost everywhere by means of the formula

$$f(t) = \int_{\mathbb{R}} e^{2\pi i\omega t}(\mathcal{F}f)(\omega)\,d\omega = \int_{-1/2}^{1/2} e^{2\pi i\omega t}(\mathcal{F}f)(\omega)\,d\omega. \qquad (*)$$

From this formula it follows that f in the variable t is the restriction to $\mathbb{R}$ of an entire function; in particular, f is smooth, and $(*)$ holds for every t. Since $\mathcal{F}f$ is supported on $[-\frac{1}{2}, \frac{1}{2}]$, we can treat it as an L^2 periodic function on $[-\frac{1}{2}, \frac{1}{2}]$ of period 1, and we can expand it in Fourier series as $\mathcal{F}f(\omega) = \sum_{k=-\infty}^{\infty} c_k e^{2\pi ik\omega}$, the series convergent in $L^2([-\frac{1}{2}, \frac{1}{2}])$. Here $c_k = \int_{-1/2}^{1/2}(\mathcal{F}f)(\omega)e^{-2\pi ik\omega}\,dw = f(-k)$, the second equality holding by $(*)$. Thus the Fourier series expansion of $(\mathcal{F}f)(\omega)$ is

$$\mathcal{F}f(\omega) = \sum_{k=-\infty}^{\infty} c_k e^{2\pi ik\omega} = \sum_{k=-\infty}^{\infty} f(-k)e^{2\pi ik\omega} = \sum_{k=-\infty}^{\infty} f(k)e^{-2\pi ik\omega}. \qquad (**)$$

Substituting $(**)$ into $(*)$ gives

$$f(t) = \int_{-1/2}^{1/2} e^{2\pi i\omega t}(\mathcal{F}f)(\omega)\,d\omega = \int_{-1/2}^{1/2} \sum_{k=-\infty}^{\infty} f(k)e^{2\pi i(t-k)\omega}\,d\omega. \qquad (\dagger)$$

If we allow ourselves to interchange sum and integral on the right side and if we assume that t is not an integer, we see that $(\dagger)$ is

$$= \sum_{k=-\infty}^{\infty} f(k) \int_{-1/2}^{1/2} e^{2\pi i(t-k)\omega}\,d\omega$$

$$= \sum_{k=-\infty}^{\infty} f(k)\Big[\frac{e^{\pi i(t-k)} - e^{-\pi i(t-k)}}{2\pi i(t-k)}\Big]$$

$$= \sum_{k=-\infty}^{\infty} f(k)\,\frac{\sin\pi(t-k)}{\pi(t-k)},$$

as required.

Under the assumption that $\Omega = 1$, we are left with justifying the interchange of integral and sum on the right side of (†) and with addressing the restriction $t \notin \mathbb{Z}$. Let $I_E(\omega)$ be the indicator function of $[-\frac{1}{2}, \frac{1}{2}]$, and define c_k as above. Put

$$f_N(t) = \mathcal{F}^{-1}\big(I_E(\omega) \sum_{k=-N}^{N} c_k e^{2\pi i k\omega}\big)(t) = \mathcal{F}^{-1}\big(I_E(\omega) \sum_{k=-N}^{N} f(-k) e^{2\pi i k\omega}\big)(t).$$

Since $I_E(\omega) \sum_{k=-N}^{N} c_k e^{2\pi i k\omega}$ tends to $I_E(\omega)(\mathcal{F}f)(\omega)$ in L^2 and is supported in the fixed bounded set E, it converges also in L^1, and it follows that its inverse Fourier transform $f_N(t)$ converges uniformly to $f(t)$, as well as in $L^2(\mathbb{R})$. The above interchange of limits works for f_N because the sum is a finite sum, and thus we obtain

$$f_N(t) = \sum_{k=-N}^{N} f(k)\, \frac{\sin \pi(t-k)}{\pi(t-k)}$$

for every N. Use of continuity shows that we do not need to assume $t \notin \mathbb{Z}$. Letting N tend to infinity, we obtain the desired identity for $\Omega = 1$ with uniform convergence and convergence in $L^2(\mathbb{R})$.

For general Ω, suppose that f has $\mathcal{F}f(\omega) = 0$ for $|\omega| > \frac{1}{2}\Omega$. Put $g(t) = \Omega^{-1} f(\Omega^{-1}t)$, so that $(\mathcal{F}g)(\omega) = (\mathcal{F}f)(\Omega\omega)$. Then $(\mathcal{F}g)(\omega) = 0$ for $|\omega| > \frac{1}{2}$, and the case proved above applies to g. Then we obtain

$$g(t) = \sum_k g(k)\, \frac{\sin \pi(t-k)}{\pi(t-k)}.$$

Substitution gives

$$\Omega^{-1} f(\Omega^{-1}t) = \Omega^{-1} \sum_k f(\Omega^{-1}k)\, \frac{\sin \pi(t-k)}{\pi(t-k)}.$$

Multiplying both sides by Ω and replacing t by Ωt completes the proof. □

We turn now to wavelets. The subject of wavelets has grown out of many areas, theoretical and applied, and some of the background is quite deep. For example, the dilations by powers of 2 that enter the definition are motivated to an extent by Littlewood–Paley theory, a part of Euclidean Fourier analysis beyond what is discussed in Chapter III of this book. Our approach will be to take the resulting theory as a subject on its own, often without presenting the deep motivation for particular definitions.

Many of the applications of wavelets are to problems of analysis and/or compression of data, possibly with noise reduction. A traditional example of compression, from the days when telephone calls were transmitted over copper wires, is how the voice signals of telephone calls were handled. The telephone company was not content with sending just one signal over a wire. It wanted to send many signals simultaneously. To do so, it filtered each conversation to eliminate frequencies other than those between 300 and 2700 cycles per second and then multiplied the signal by a high-frequency carrier wave to obtain a signal in a high-frequency band of width 2400 cycles per second. Allowing different bands for different conversations allowed it to transmit many conversations at once over the same wires. Some information was lost during this step of compression, including all frequencies outside the band 300 to 2700 cycles per second.

In the setting we are studying, the data can be in analog form, as was the case in the example just given, but let us think of the given data as digital. In the one-dimensional case we then have a digitized signal, perhaps of speech or music or some other kind of sound. In the two-dimensional case, we have an image, perhaps the photographic image produced by a digital camera or an image of a fingerprint. We want to know what is happening in the signal and to process the signal accordingly, perhaps enhancing some parts of it and damping other parts.

If traditional Fourier analysis is to be used, then any compression will likely be achieved by discarding high-frequency information. For example, the JPEG file of a digital photograph[3] is obtained by treating the image as built from squares of size 8 pixels by 8 pixels, doing a discrete Fourier analysis on each piece, and discarding some information. For most photographs this method of compression is perfectly adequate, but interactions occur at the boundary between adjacent 8-by-8 squares. At high compression ratios these interactions can produce a visible effect known as "blocking," and the result can sometimes be disconcerting. On an image like that of a fingerprint, JPEG is distinctly inadequate.

In any of these situations the theory of wavelets is available to handle the analysis and/or processing in a different way. Different wavelets have different advantages, and one may want one kind of wavelet for one situation and another kind for another situation. One study of speech[4] for designing hearing aids classified short intervals of speech into four possible kinds—"voiced, plosive, fricative, and silent segments"—and it proceeded from there. One of its tools was the Daubechies wavelet of order $N = 3$.

Over the next nine sections we shall construct directly certain wavelets and

[3]"JPEG" stands for Joint Photographic Experts Group. The committee has a website, namely `https://jpeg.org`.

[4]B. T. Tan, R. Lang, Heiko Schroder, A. Spray, and P. Dermody, "Applying wavelet analysis to speech segmentation and classification," *Proc. SPIE* (International Society for Optical Engineering) 2242 (1994), Orlando, FL, 750–761.

families of wavelets, extracting properties of each. We construct the first historical example of a wavelet in Section 2 and develop a fairly general method of constructing wavelets in Section 3. Subsequently we shall make refinements to the construction. Applying the method and its refinements, we obtain several important explicit families of wavelets in Sections 4–9. All along, we shall observe certain properties of these families; these properties influence the usefulness of the various families for certain kinds of applications. Section 10 completes the derivation of the properties, and a table listing the names of the families and summarizing the properties appears as Figure 10.18.

In Section 11 we shall take a look at the wavelet transform. Analyzing the nature of the transform and of its calculation gives an inkling of how to take advantage of wavelets in analyzing signals. At the end of the section, we make some remarks about one-dimensional applications and also about two-dimensional wavelets and their use in improving the JPEG algorithm and the compression of fingerprint records. The applications to images involve a new wrinkle in that the sight of asymmetries appears to be more noticeable that one might at first expect. As a result it is desirable for wavelets involved in image processing to be symmetric in a certain sense. This symmetry is impossible for compactly supported orthogonal wavelets, and some relaxation of the definition is warranted.

In the end, analysis using wavelets is one tool in signal processing; Fourier analysis is another. There are also others. One should not think of wavelet analysis as a substitute for these other methods, however. As Y. Meyer put it in his groundbreaking book[5] listed in the Selected References, “But wavelet analysis cannot entirely replace Fourier analysis, indeed, the latter is used in constructing the orthonormal bases of wavelets needed for analysis with wavelet series.”

2. Haar Wavelet

The **Haar system** is the system of functions in $L^2(\mathbb{R})$ defined by

$$\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k) \qquad \text{for } j \text{ and } k \text{ in } \mathbb{Z},$$

where ψ is the function in $L^2(\mathbb{R})$ defined by

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

In this section we shall prove the following result and consider some of its ramifications.

[5] Y. Meyer, *Wavelets and Operators*, p. 1.”

Theorem 10.4. The Haar system is an orthonormal basis of $L^2(\mathbb{R})$.

Therefore the above function ψ is a wavelet for what is called the **Haar system**, and one refers to ψ as the **Haar wavelet**. Any integer translate $\psi_{0,k}$ of $\psi = \psi_{0,0}$ would serve equally well for this purpose. The proof will make use of an auxiliary function φ in $L^2(\mathbb{R})$ called the **scaling function** for the system. The Haar scaling function φ is the indicator function of the interval [0, 1), namely

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

In notation analogous to that for ψ, we introduce the system of functions

$$\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k) \qquad \text{for } j \text{ and } k \text{ in } \mathbb{Z}.$$

Graphs of φ and ψ appear in Figures 10.2a and 10.2b.

The details of the proof of Theorem 10.4 are of some importance for other systems as well as the Haar system, and we shall therefore give the proof in steps and extract some further information from it after it is complete.

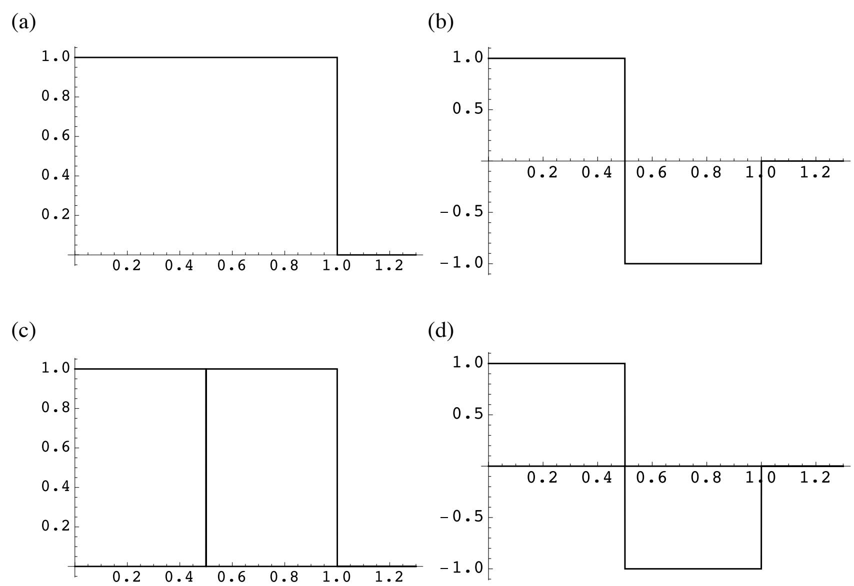


FIGURE 10.2. Graphs of Haar scaling function and wavelet.
(a) $\varphi(x)$. (b) $\psi(x)$. (c) $\varphi(x) = \varphi(2x) + \varphi(2x-1)$.
(d) $\psi(x) = \varphi(2x) - \varphi(2x-1)$.

PROOF OF THEOREM 10.4.

Step 1. We prove that $\{\psi_{j,k}\}$ is an orthonormal system.

The functions being real, we can drop all references to complex conjugation. Using the change of variables $2^j x - k = y$, which has $x = 2^{-j}(y+k)$ and $dx = 2^{-j}\,dy$, we examine the integral

$$\begin{aligned}\int_{\mathbb{R}} \psi_{j,k}(x)\psi_{m,n}(x) &= \int_{\mathbb{R}} 2^{(j+m)/2}\psi(2^j x - k)\psi(2^m x - n)\,dx,\\ &= \int_{\mathbb{R}} 2^{(-j+m)/2}\psi(y)\psi(2^{m-j}(y+k)-n)\,dy.\end{aligned}$$

If $j = m$, this reduces to

$$= \int_{\mathbb{R}} \psi(y)\psi(y+k-n)\,dy;$$

for $k = n$, the integrand is $\psi(y)^2 = \varphi(y)$, and the integral is 1, while for $k \neq n$, the two factors of the integrand are nonzero on disjoint sets and the integral is 0. If $j \neq m$, we may assume by symmetry that $r = m - j$ is > 0. Then it is enough to show that

$$\int_{\mathbb{R}} \psi(y)\psi(2^r y + t)\,dy$$

is 0 for $r > 0$ when t is any integer. Under the change of variables $2^r y + t = u$, $dy = 2^{-r} du$, this integral becomes

$$\begin{aligned}&= \int_0^{1/2} \psi(2^r y + t)\,dy - \int_{1/2}^1 \psi(2^r y + t)\,dy\\ &= 2^{-r}\int_t^{2^{r-1}+t} \psi(u)\,du - 2^{-r}\int_{2^{r-1}+t}^{2^r+t} \psi(u)\,du.\end{aligned}$$

Each of the integrals on the right side is an integral of ψ over an interval of integer length, and any such integral is 0. This completes the proof of orthonormality.

Step 2. If V_0 is the closed subspace of members of $L^2(\mathbb{R})$ that are constant almost everywhere on each interval $[k, k+1)$ with k in $\mathbb{Z}$, then the functions $x \mapsto \varphi(x-k)$ form an orthonormal basis of V_0.

In fact, the function $x \mapsto \varphi(x-k)$ is 1 on $[k, k+1)$ and is 0 otherwise. Its square integral is therefore 1, and the inner product of any two distinct such functions is 0. Thus the set of functions $x \mapsto \varphi(x-k)$ is orthonormal. Since the vector space of finite linear combinations of the functions $x \mapsto \varphi(x+k)$ is dense in V_0, it follows that the set of functions $x \mapsto \varphi(x-k)$ is an orthonormal basis of V_0.

Step 3. We construct some closed subspaces V_j of $L^2(\mathbb{R})$ for $j \geq 1$, we observe for each j that the functions $\varphi_{j,k}$, as k varies, form an orthonormal basis of V_j, and we note some properties of V_j.

We let V_j be the set of functions in $L^2(\mathbb{R})$ that are constant almost everywhere on each interval $[2^{-j}k, 2^{-j}(k+1))$. Then it is plain that for all $j \geq 0$,

$$V_j \subseteq V_{j+1} \tag{$*$}$$

and

$$f(x) \text{ is in } V_j \quad \text{if and only if} \quad f(2x) \text{ is in } V_{j+1}. \tag{$**$}$$

Arguing as in Step 2, we see for each $j \geq 0$ that the set of functions

$$x \mapsto \varphi_{j,k}(x) \quad \text{is an orthonormal basis of } V_j. \tag{$\dagger$}$$

Step 4. We prove that $\bigcup_{j=0}^{\infty} V_j$ is dense in $L^2(\mathbb{R})$.

In fact, suppose $f \in L^2(\mathbb{R})$ is to be approximated, and let $\epsilon > 0$ be given. Corollary 6.4 of *Basic* shows that $C_{\text{com}}(\mathbb{R})$ is dense in $L^2(\mathbb{R})$. Thus we can choose $g \in C_{\text{com}}(\mathbb{R})$ with $\|f-g\|_2 < \epsilon$. Let E be the support of g. Since g is uniformly continuous on the compact set E and hence on all of $\mathbb{R}$, we can find a negative power of 2, say 2^{-l}, small enough so that $|x-y| < 2^{-l}$ implies $|g(x)-g(y)| < \epsilon/|E|^{1/2}$, where $|E|$ is the measure of E. Define a function $h(x)$ to equal $g(n2^{-l})$ for $n2^{-l} \leq x < (n+1)2^{-l}$. Then h is in V_l, and h satisfies an estimate

$$|g(x)-h(x)| = |g(x)-g(n2^{-l})| < \epsilon/|E|^{1/2} \qquad \text{for } x \in [n2^{-l}, (n+1)2^{-l})$$

and hence for all x. Then

$$\|g-h\|_2^2 \leq |E|(\epsilon/|E|^{1/2})^2 \leq \epsilon^2,$$

and hence $\|g-h\|_2 \leq \epsilon$. Therefore $\|f-h\|_2 < 2\epsilon$, and $\bigcup_{j=0}^{\infty} V_j$ is dense in $L^2(\mathbb{R})$.

Step 5. We construct some closed subspaces W_j of $L^2(\mathbb{R})$ for $j \geq 0$, and we examine the properties of W_j.

Let us define W_j for $j \geq 0$ to be the orthogonal complement of V_j in V_{j+1}. For $j \geq 0$, let us check that

$$f(x) \text{ is in } W_j \quad \text{if and only if} \quad f(2x) \text{ is in } W_{j+1}. \tag{$\dagger\dagger$}$$

In fact, if $f(x)$ is in W_j, then it is in V_{j+1}. So $f(2x)$ is in V_{j+2}. If $g(x)$ is in V_{j+1}, then $g(\frac{1}{2}x)$ is in V_j by $(**)$, and it has $\int_{\mathbb{R}} f(2x)g(x)\,dx = 2\int_{\mathbb{R}} f(x)g(\frac{1}{2}x)\,dx = 0$. Hence $f(2x)$ is in the orthogonal complement of V_{j+1} in V_{j+2}, which is W_{j+1}.

In the reverse direction if $f(2x)$ is in W_{j+1}, then $f(2x)$ is in V_{j+2}, and $(**)$ shows that $f(x)$ is in V_{j+1}. Every $g(x)$ in V_{j+1} has $0 = \int_{\mathbb{R}} f(2x)g(x)\,dx =$

$\frac{1}{2}\int_{\mathbb{R}} f(x)g(\frac{1}{2}x)\,dx$. The function $g(\frac{1}{2}x)$ is the most general element of V_j by $(**)$, and thus $f(x)$ is in W_j. This proves $(\dagger\dagger)$.

Let us write $\perp$ for orthogonal direct sum. For $n \geq 0$, we then have

$$\begin{aligned} V_n &= V_{n-1} \perp W_{n-1} = V_{n-2} \perp (W_{n-2} \perp W_{n-1}) \\ &= V_{n-3} \perp (W_{n-3} \perp W_{n-2} \perp W_{n-1}) \\ &= \cdots = V_0 \perp (W_0 \perp W_1 \perp \cdots \perp W_{n-1}). \end{aligned} \tag{$\ddagger$}$$

Step 6. We observe that φ is in V_0, that ψ is in V_1, that the functions φ and ψ satisfy the equations

$$\varphi(x) = \varphi(2x) + \varphi(2x-1)$$

and

$$\psi(x) = \varphi(2x) - \varphi(2x-1),$$

and that ψ is in W_0. These equations are reflected in the graphs appearing in Figures 10.2c and 10.2d.

The facts that φ is in V_0 and ψ is in V_1 are clear from the definitions. Now

$$\varphi(2x) = \begin{cases} 1 & \text{if } 0 \leq x < \frac{1}{2}, \\ 0 & \text{otherwise,} \end{cases}$$

and

$$\varphi(2x-1) = \begin{cases} 1 & \text{if } \frac{1}{2} \leq x < 1, \\ 0 & \text{otherwise.} \end{cases}$$

Their sum is 1 for $0 \leq x < 1$ and thus equals $\varphi(x)$, while their difference is 1 for $0 \leq x < \frac{1}{2}$ and is -1 for $\frac{1}{2} \leq x < 1$ and thus equals $\psi(x)$. For inner products we have

$$(\varphi_{0,k}, \psi) = \int_{\mathbb{R}} \varphi(x-k)\psi(x)\,dx = \int_k^{k+1} \psi(x)\,dx,$$

and this is 0 for all k; for $k = 0$, it is 0 because ψ has integral 0, and for $k \neq 0$, it is 0 because the set where ψ is not 0 does not meet $[k, k+1)$. Thus ψ is orthogonal to every member of the orthonormal basis of V_0 given in Step 2. Since ψ is known to be in V_1, ψ is in W_0.

Step 7. We show that the functions $x \mapsto \psi(x-k)$ form an orthonormal basis of W_0.

Being nonzero on disjoint sets, they are orthonormal. Let f be any member of W_0. Since W_0 is contained in V_1 and since the functions $\varphi_{1,k}$ form an orthonormal basis of V_1, f is of the form

$$\begin{aligned} f(x) &= \sum_k c_k\varphi_{1,k}(x) = \sum_{k \text{ even}} c_k\varphi_{1,k}(x) + \sum_{k \text{ odd}} c_k\varphi_{1,k}(x) \\ &= \sqrt{2}\sum_l c_{2l}\varphi(2x-2l) + \sqrt{2}\sum_l c_{2l+1}\varphi(2x-2l-1) \\ &= \tfrac{\sqrt{2}}{2}\sum_l (c_{2l}+c_{2l+1})\big(\varphi(2x-2l)+\varphi(2x-2l-1)\big) \\ &\quad + \tfrac{\sqrt{2}}{2}\sum_l (c_{2l}-c_{2l+1})\big(\varphi(2x-2l)-\varphi(2x-2l-1)\big) \\ &= \tfrac{\sqrt{2}}{2}\sum_l (c_{2l}+c_{2l+1})\varphi(x-l) + \tfrac{\sqrt{2}}{2}\sum_l (c_{2l}-c_{2l+1})\psi(x-l), \end{aligned}$$

the last equality holding by Step 6. Since f is orthogonal to V_0, taking the inner product of both sides with any particular $\varphi(x-l)$ shows that the coefficient of $\varphi(x-l)$ is 0. That being true for all l, f is exhibited as in the closed linear span of the functions $\psi(x-l)$.

Step 8. We show that for each $j \geq 0$, the functions $\psi_{j,k}$, as k varies, form an orthonormal basis of W_j.

In fact, this is immediate from (††) and Step 7.

Step 9. The functions $\varphi_{0,k}$ for $k \in \mathbb{Z}$ and the functions $\psi_{j,k}$ for $j \geq 0$ and $k \in \mathbb{Z}$ together form an orthonormal basis of $L^2(\mathbb{R})$. Here the $\varphi_{0,k}$ form an orthonormal basis of V_0, and the other functions form an orthonormal basis of the orthogonal complement of V_0 in $L^2(\mathbb{R})$.

The functions $\varphi_{0,k}$ for $k \in \mathbb{Z}$ and the functions $\psi_{j,k}$ for $0 \leq j \leq n-1$ and $k \in \mathbb{Z}$ together form an orthonormal basis of V_j, by (‡), Step 2, and Step 8. Taking the union on j of the nested spaces V_j and applying Step 4, we find that the functions $\varphi_{0,k}$ for $k \in \mathbb{Z}$ and the functions $\psi_{j,k}$ for $j \geq 0$ and $k \in \mathbb{Z}$ together form an orthonormal basis of $L^2(\mathbb{R})$.

Step 10. We extend the definitions of the spaces V_j and W_j to $j < 0$, and we observe that (∗), (∗∗), (†), (††) extend to be valid for all j, and (‡) has a natural extension to handle the extended definitions.

The space V_j is the set of functions in $L^2(\mathbb{R})$ that are constant almost everywhere on each interval $[2^{-j}k, 2^{-j}(k+1))$, and the space W_j is the orthogonal complement of V_j in V_{j+1}. Then it is immediate that (∗), (∗∗), (†), (††) extend to be valid for all j. The extended form of (‡) is

$$V_n = V_m \perp (W_m \perp W_{m+1} \perp \cdots \perp W_{n-1}) \qquad \text{for } m \leq n, \qquad (‡‡)$$

and this too is immediate.

Step 11. The functions $\varphi_{m,k}$ for $k \in \mathbb{Z}$ and the functions $\psi_{j,k}$ for $j \geq m$ and $k \in \mathbb{Z}$ together form an orthonormal basis of $L^2(\mathbb{R})$.

This is immediate from Step 9, the extended Step 8, and (‡‡).

Step 12. We prove that $\bigcap_{j=-\infty}^{n} V_j = 0$ for every integer n.

In fact, let f be a member of $\bigcap_{j=-\infty}^{n} V_j$. Being in each V_j, f is constant almost everywhere on each interval $[2^{-j}k, 2^{-j}(k+1))$ for $j \leq n$ and $k \in \mathbb{Z}$, say equal to $c_{j,k}$. If $E_{k,j}$ denotes the exceptional subset of $[2^{-j}k, 2^{-j}(k+1))$ of measure 0 where $f(x) \neq c_{j,k}$ and if $E = \bigcup_{j,k} E_{k,j}$, then E is a set of measure 0 and has the following property. For any $x \notin E$, f is constantly equal to $c_{j,k}$ on all intervals $[2^{-j}k, 2^{-j}(k+1))$ containing x. Fix $x_0 \notin E$. We can choose an increasing sequence of intervals $[2^{-j}k, 2^{-j}(k+1))$ containing x_0 and having union $\mathbb{R}$. The constants $c_{j,k}$ for the members of this sequence must match the value $f(x_0)$, and it follows that f is constantly equal to $f(x_0)$ almost everywhere. Since the only constant function in $L^2(\mathbb{R})$ is 0, we conclude that $\bigcap_{j=-\infty}^{n} V_j = 0$.

Step 13. The functions $\psi_{j,k}$ for $j \in \mathbb{Z}$ and $k \in \mathbb{Z}$ together form an orthonormal basis of $L^2(\mathbb{R})$.

From (‡‡) and Step 12 it follows that V_0 is the orthogonal direct sum of the spaces $W_{-1}, W_{-2}, W_{-3}, \dots$. From Step 9 it then follows that $L^2(\mathbb{R})$ is the orthogonal direct sum of the spaces W_j for $j \in \mathbb{Z}$. By Step 8 the functions $\psi_{j,k}$ for $j \in \mathbb{Z}$ and $k \in \mathbb{Z}$ together form an orthonormal basis of $L^2(\mathbb{R})$. $\square$

As was mentioned in Section 1, corresponding to any wavelet ψ is the wavelet expansion

$$f(x) = \sum_{j,k \in \mathbb{Z}} \Big(\int_{\mathbb{R}} f(y) \overline{\psi_{j,k}(y)}\, dy \Big) \psi_{j,k}(x),$$

the series being understood as convergent in $L^2(\mathbb{R})$. This remark applies in particular to the Haar wavelet $\psi(x)$, and the resulting series is called the two-sided **Haar series expansion** of the function f. In a Haar series the complex conjugate sign can be dropped because ψ is real-valued, but we shall include it anyway.

When writing down an expansion in Haar series, however, there is some advantage in taking into account aspects of the proof of Theorem 10.4. The proof was organized around a certain increasing sequence of closed subspaces V_j whose union is dense in $L^2(\mathbb{R})$. Let us see how these spaces include finer and finer detail about f. The way in which the functions $\varphi_{j,k}$ and $\psi_{j,k}$ are defined involves evaluating φ or ψ at $2^j x - k$. It is helpful of this operation as one of *magnification*, affecting the *resolution* of what we can see. Namely it is helpful to think of the operation of passing from x to $2^j x$ as one of magnifying a graph by 2^j or of introducing better resolution so that details at the level of 2^{-j} become visible.

Let us be more specific. Step 9 observed that the functions $\varphi_{0,k}$ for $k \in \mathbb{Z}$ and the functions $\psi_{j,k}$ for $j \geq 0$ and $k \in \mathbb{Z}$ together form an orthonormal basis of $L^2(\mathbb{R})$. The corresponding expansion is

$$f(x) = \sum_{k\in\mathbb{Z}} \big(\int_{\mathbb{R}} f(y)\overline{\varphi_{0,k}(y)}\,dy \big)\varphi_{0,k}(x) + \sum_{j=0}^{\infty} \sum_{k\in\mathbb{Z}} \big(\int_{\mathbb{R}} f(y)\overline{\psi_{j,k}(y)}\,dy \big)\psi_{j,k}(x)$$

and is called a one-sided **Haar series expansion** of the function f. The first term represents the orthogonal projection of f on V_0, and the second term represents the orthogonal projection on the orthogonal complement of V_0.

This one-sided expansion is better than the two-sided expansion at emphasizing that the large positive values of j correspond to what is happening as one passes to the limit.

Step 11 of the proof gives another way of thinking about this process. It says that for any m (think of m as ≥ 0), the functions $\varphi_{m,k}$ and the functions $\psi_{j,k}$ with $j \geq m$ together form an orthonormal basis of $L^2(\mathbb{R})$. Comparing the resulting expansion of f with the display above, we see that $\sum_{k\in\mathbb{Z}} \big(\int_{\mathbb{R}} f(y)\overline{\varphi_{m,k}(y)}\,dy \big)\varphi_{m,k}(x)$ represents a partial sum of the one-sided Haar series expansion. It is in fact the orthogonal projection of f on the subspace V_m. Thus we can think of the orthogonal projection on V_0 as giving a zeroth approximation to f. Then we add the terms $\sum_{k\in\mathbb{Z}} \big(\int_{\mathbb{R}} f(y)\overline{\psi_{0,k}(y)}\,dy \big)\psi_{0,k}(x)$ and obtain the first approximation to f, namely the orthogonal projection on V_1. We continue by adding the terms $\sum_{k\in\mathbb{Z}} \big(\int_{\mathbb{R}} f(y)\overline{\psi_{1,k}(y)\,dy} \big)\psi_{1,k}(x)$ and obtain the second approximation to f, namely the orthogonal projection on V_2. With each step we improve the resolution, obtaining a better approximation to f. The limit of the approximations, taken in the L^2 sense, is f.

The process thus consists in looking at f with an infinite system of finer and finer resolutions. It is for this reason that the term "multiresolution analysis" will be used for this construction starting in the next section.

Notice that what is happening here is quite different from the situation with Fourier series. With Fourier series a new term in a series represents taking into account a new frequency, an enlargement of the frequency domain. With wavelets a new term represents taking into account a higher resolution, thus giving better knowledge of what is happening in the time domain.

Some readers may be helped by a piece of intuition used by people in signal analysis. They think of the scaling function φ as being akin to a low-pass filter and the mother wavelet ψ as being akin to a high-pass filter. It is helpful to think of these filters as working on fineness of detail, however, rather than frequencies. For example, in the construction in Theorem 10.4, φ and its integer translates together yielded the orthogonal projection on the fundamental space V_0, while ψ and its integer translates yielded the orthogonal projection on the detail space consisting of the orthogonal complement of V_0 in V_1. As Barbara Hubbard explains in her book,[6] it was this piece of intuition that finally brought mathematicians and engineers together on the subject of wavelets in 1986.

We conclude this section with a brief discussion of spaces of functions other than L^2 on the real line. We shall return to L^2 from the beginning of the next section. This discussion will enable us to make some comparisons of Haar series and Fourier series. The results with Haar series that we mention are just the beginning. Considerable research has gone into the study of Haar series and other wavelet expansions in connection with spaces other than L^2, but we shall not be delving into it beyond the few remarks that we include here.

To begin with, for any complex-valued function f on $\mathbb{R}$ satisfying suitable conditions, we define

$$(P_m f)(x) = \sum_{k\in\mathbb{Z}} \Big(\int_{\mathbb{R}} f(y)\overline{\varphi_{m,k}(y)}\, dy\Big)\varphi_{m,k}(x).$$

If we are working with L^2 functions, the operator P_m is the orthogonal projection of $L^2(\mathbb{R})$ on V_m. But our interest now will be in other spaces of functions.

The function $\varphi_{m,k}(x)$ is nonzero if and only if $k \le 2^m x < k+1$. Consequently for fixed m and x, there is exactly one value of k for which $\varphi_{m,k}(x)$ is nonzero. Thus the sum defining $(P_m f)(x)$ has only one term. Also the function $\overline{\varphi_{m,k}(y)}$ has compact support, and thus the integral $\int_{\mathbb{R}} f(y)\overline{\varphi_{m,k}(y)}\, dy$ is well defined as soon as f is a measurable function on $\mathbb{R}$ that is locally integrable, i.e., integrable on each compact set. Consequently the expression defining $(P_m f)(x)$ is well defined whenever the function f on $\mathbb{R}$ is locally integrable. Its value is just $2^m \int_{k\le 2^m y<k+1} f(y)\, dy$, where k is the unique integer such that $k \le 2^m x < k+1$. It is a simple matter to use Step 5 and its dilates in Theorem 10.4 to see that

$$(P_{m+1} f)(x) - (P_m f)(x) = \sum_{k\in\mathbb{Z}} \Big(\int_{\mathbb{R}} f(y)\overline{\psi_{m,k}(y)}\, dy\Big)\psi_{m,k}(x).$$

Therefore convergence results about $P_m f$ as m tends to $+\infty$ are in effect convergence results about the one-sided Haar series expansion of f. We summarize as follows.

[6] *The World According to Wavelets*, listed in the Selected References.

Proposition 10.5. If f is a locally integrable complex-valued function on $\mathbb{R}$ and if k is the unique integer such that $k \le 2^m x < k+1$, then the partial sum $(P_m f)(x)$ of the one-sided Haar series expansion is well defined and is given by

$$|I_{m,k}|^{-1} \int_{I_{m,k}} f(y)\,dy,$$

where $I_{m,k} = \{y \in \mathbb{R} \mid k \le 2^m y < k+1\}$ and $|I_{m,k}|$ is its measure, namely 2^{-m}.

The expression for $(P_m f)(x)$ in Lemma 10.5 is the one that arises in the theory of differentiation of integrals, which is discussed in Section VI.6 of *Basic*. Application of the results of that section immediately yields the following corollary.

Corollary 10.6. If f is a locally integrable complex-valued function on $\mathbb{R}$, then the one-sided Haar series of f converges to f almost everywhere.

If f is continuous, the situation is much simpler, since we no longer need the results of Section VI.6 of *Basic*.

Corollary 10.7. If f is a continuous complex-valued function on $\mathbb{R}$, then the one-sided Haar series of f converges to f pointwise. The convergence is uniform on any set on which f is uniformly continuous.

REMARK. Observe that this behavior is much better than what happens for Fourier series: the Fourier series of a continuous function can diverge at a point.

PROOF. For $x \in I_{m,k}$, we have

$$\begin{aligned} \Big||I_{m,k}|^{-1} \int_{I_{m,k}} f(y)\,dy - f(x)\Big| &= |I_{m,k}|^{-1} \Big| \int_{I_{m,k}} \big(f(y) - f(x)\big)\,dy\Big| \\ &\le |I_{m,k}|^{-1} \int_{I_{m,k}} \big|f(y) - f(x)\big|\,dy. \qquad (*) \end{aligned}$$

Let $\epsilon > 0$ be given, and choose the corresponding δ of uniform continuity on E. If m is large enough so that $2^{-m} < \delta$, then $|y - x| \le \delta$ whenever x and y are both in $I_{m,k}$, and hence $|f(y) - f(x)| \le \epsilon$. Consequently $(*)$ is $\le \epsilon$. □

Next we show that one-sided Haar series do not exhibit Gibbs phenomenon in the way that Proposition 9.2 says that Fourier series do.

Lemma 10.8. If f is a bounded locally integrable function of $\mathbb{R}$ and if x in $\mathbb{R}$, then $|(P_m f)(x)| \le \|f\|_{\text{sup}}$.

PROOF. Choose k so that $k \le 2^m x < k+1$. Then

$$|(P_m f)(x)| = 2^m \Big| \int_{k \le 2^m y \le k+1} f(y)\,dy\Big| \le 2^m \|f\|_{\text{sup}} \int_{k \le 2^m y \le k+1} dy = \|f\|_{\text{sup}}.$$

□

Proposition 10.9. Suppose that f is a real-valued function on $\mathbb{R}$ that is continuous everywhere except at the one point x_0, where f has a jump discontinuity. Then the one-side Haar series of f does not exhibit Gibbs phenomenon. Specifically if $\lim_{x\uparrow x_0} f(x) < \lim_{x\downarrow x_0} f(x)$, then along every sequence $\{x_n\}$ decreasing to x_0 and sequence m_n tending to $+\infty$, the inequality

$$\limsup_{n\to\infty}(P_{m_n} f)(x_n) \le \lim_{x\downarrow x_0} f(x)$$

holds.

PROOF. Let $J = \lim_{x\downarrow x_0} f(x) - \lim_{x\uparrow x_0} f(x)$ be the size of the jump discontinuity of f at x_0. The value of f at x_0 itself does not affect the Haar series of f, and we may therefore redefine $f(x_0)$ to equal $\lim_{x\uparrow x_0} f(x)$.

Let $g(x)$ be the function that equals 0 for $x \le x_0$, equals $x_0 + 1 - x$ for $x_0 < x \le x_0 + 1$, and equals 0 for $x \ge x_0 + 1$. This is continuous except at x_0, where it has a jump discontinuity of 1. By Lemma 10.8,

$$|(P_m g)(x_n)| \le \|g\|_{\sup} = 1 \qquad \text{for every } m \text{ and } n. \tag{$*$}$$

Write $f = (f - Jg) + Jg$. The function $f - Jg$ is continuous everywhere, and $(P_m(f - Jg))(x)$ tends to $(f - Jg)(x)$ uniformly for x in a neighborhood of x_0, by Corollary 10.7. Proposition 1.16 of *Basic* therefore shows that

$$\lim_{n\to\infty}(P_{m_n}(f - Jg))(x_n) = (f - Jg)(x_0) = \lim_{x\downarrow x_0}(f - Jg)(x). \tag{$**$}$$

The proposition follows by adding the equality $(**)$ and J times the inequality $(*)$. □

Thus there are senses in which one-sided Haar series are much better behaved than Fourier series. But there are other senses in which Fourier series are the better behaved. One of these is that the existence of derivatives of a function in the subject of Fourier series forces the Fourier coefficients to have at least a certain rate of decrease. The relevant estimate comes from using integration by parts in the formula for the n^{th} Fourier coefficient. Any attempt to imitate this argument for Haar series is doomed because the functions φ and ψ are not even continuous. Consequently a Haar series cannot be expected to have a very small remainder term if terms are discarded from the series. This feature is a drawback of Haar series and is a reason to search for other wavelets for which the series expansions of smooth functions are rapidly convergent.

3. Multiresolution Analysis

In this section we isolate the essential features of the construction of the Haar system in Section 2 and arrive at a construction of wavelets that applies in many situations. A **multiresolution analysis** consists of an increasing system $\{V_j\}_{n=-\infty}^{\infty}$ of closed vector subspaces of $L^2(\mathbb{R})$ and a function φ in $L^2(\mathbb{R})$ such that

(i) $\bigcup_j V_j$ is dense in $L^2(\mathbb{R})$,
(ii) $\bigcap_j V_j = 0$,
(iii) for each $j \in \mathbb{Z}$, $x \mapsto f(x)$ is in V_j if and only if $x \mapsto f(2x)$ is in V_{j+1}, and
(iv) the system of functions $\{x \mapsto \varphi(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis of V_0.

The reason for the cumbersome term "multiresolution analysis" was explained in Section 2 after the definition of one-sided Haar series expansion. The function φ is called the **scaling function** of the multiresolution analysis.

For each integer j, we define functions $\varphi_{j,k}$ in $L^2(\mathbb{R})$ for $k \in \mathbb{Z}$ by

$$\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k).$$

That is all that is needed. The function ψ has not been mentioned!

The main result of the section will be that the existence ψ is built into the definition of multiresolution analysis, and there is little choice for ψ. Before stating such a result as a theorem, let us go through the steps of the proof of Theorem 10.4 to see how much has been captured by this definition in the case of the Haar system.

Indeed, for the Haar system we defined the scaling function φ at the outset, and we defined closed subspaces V_j of $L^2(\mathbb{R})$ in Steps 2, 3, and 10 of the proof of Theorem 10.4. Step 4 established (i) above, Step 12 established (ii), $(**)$ within Steps 3 and 10 established (iii), and Step 2 established (iv). Therefore the Haar system is a multiresolution analysis.

The proof of Theorem 10.4 established some additional properties of the Haar multiresolution analysis. Line (†) of Step 3 observed the functions $\varphi_{j,k}$ form an orthonormal basis of V_j for $j \geq 0$, and (†) in Step 10 said that the same thing is true for $j < 0$. Finally Step 6 established the all-important equation

$$\varphi(x) = \varphi(2x) + \varphi(2x-1)$$

in the Haar case. All the other intermediate conclusions in the proof of Theorem 10.4 related to ψ.

To construct ψ, we make use of the analog of the above equation for a general multiresolution analysis. For the general case φ is in V_0, which is contained in V_1, and the functions $\varphi_{1,k}$ form an orthonormal basis of V_1. Expansion of φ in this orthonormal basis leads us to a series $\varphi = \sum_{k=-\infty}^{\infty} a_k\varphi_{1,k}$ convergent in $L^2(\mathbb{R})$. In this expansion the coefficients are inner products $a_k = (\varphi, \varphi_{1,k})$ satisfying $\sum |a_k|^2 = \|\varphi\|_2^2 = 1$. Substituting the definition of $\varphi_{1,k}$ allows us to rewrite this expansion as

$$\varphi(x) = \sum_{k=-\infty}^{\infty} a_k\sqrt{2}\,\varphi(2x-k).$$

We call this result the **scaling equation**[7] of the multiresolution analysis. We study the scaling equation through its Fourier transform.

Associated to the scaling equation is a useful L^2 periodic function m_0 of period 1 that behaves like a generating function. Namely the Riesz–Fischer Theorem (Theorem 6.51 of *Basic*) shows that there exists a periodic L^2 function m_0 of period 1 such that m_0 is given on [0, 1] by a Fourier series[8] as

$$m_0(y) = \tfrac{1}{\sqrt{2}} \sum_{k=-\infty}^{\infty} a_k e^{-2\pi iky}.$$

This function has

$$\|m_0\|^2_{L^2([0,1])} = \tfrac{1}{2} \sum_{k=-\infty}^{\infty} |a_k|^2 = \tfrac{1}{2}.$$

The Fourier transform of the scaling equation is

$$(\mathcal{F}\varphi)(y) = \sum_{k=-\infty}^{\infty} \tfrac{1}{\sqrt{2}}\, a_k e^{-\pi iky}(\mathcal{F}\varphi)(\tfrac{1}{2}y).$$

By inspection the generating function m_0 allows us to write the Fourier transform of the scaling equation in the tidy form

$$\boxed{(\mathcal{F}\varphi)(y) = m_0(y/2)(\mathcal{F}\varphi)(y/2)}$$

If there is a wavelet ψ corresponding to the multiresolution analysis consisting of $\{V_j\}_{j=-\infty}^{\infty}$ and φ, then it lies in the orthogonal complement W_0 of V_0 in V_1, and the functions $x \mapsto \psi(x-k)$ form an orthonormal basis of W_0. The main result of this section is that all members of W_0 have a certain elegant form, and it shows which functions of that form can be taken as the desired wavelet ψ. Often we shall use a particular choice of this function as the desired wavelet, as we remark after the statement of the theorem.

[7] Some authors treat the factors of $\sqrt{2}$ in a different way, incorporating them into the coefficients. Some authors call this the dilation equation; other authors do not give it a name.

[8] Strictly speaking, this is the Fourier series of $m_0(-y)$ rather than $m_0(y)$, but we follow the convention used in the book by Daubechies and echoed in the book by Pinsky. Regardless of how authors handle the factors of $\sqrt{2}$, the factors have disappeared by the time that one computes the values of $m_0(y)$.

Theorem 10.10. Let $\{V_j\}_{j=-\infty}^{\infty}$ and φ constitute a multiresolution analysis, and let m_0 be the generating function of the scaling equation. Then

(a) the most general member f of the orthogonal complement W_0 of V_0 in V_1 has Fourier transform of the form

$$(\mathcal{F}f)(y) = e^{\pi i y}\nu(y)\overline{m_0(\tfrac{1}{2}y + \tfrac{1}{2})}(\mathcal{F}\varphi)(\tfrac{1}{2}y),$$

where ν is a periodic function of period 1 and with $\|\nu\|^2_{L^2([0,1])} = \|f\|_2^2$, and

(b) a member f of W_0 as in (a) will serve as a wavelet for the multiresolution analysis if and only if $|\nu(y)| = 1$ almost everywhere.

The proof will be preceded by some discussion and three lemmas. Once we have settled on a particular choice of wavelet ψ from Theorem 10.10, we can write down a **wavelet equation** analogous to the scaling equation. The reason is that ψ is in V_1 and the functions $\varphi_{1,k}$ form an orthonormal basis of V_1. Expansion of ψ in this orthonormal basis leads us to a series $\psi = \sum_{k=-\infty}^{\infty} b_k\varphi_{1,k}$ convergent in $L^2(\mathbb{R})$. In this expansion the coefficients are inner products $b_k = (\psi, \varphi_{1,k})$ satisfying $\sum |b_k|^2 = \|\psi\|_2^2 = 1$. Substituting the definition of $\varphi_{1,k}$ allows us to rewrite this wavelet equation as

$$\psi(x) = \sum_{k=-\infty}^{\infty} b_k\sqrt{2}\,\varphi(2x-k).$$

We argue with this equation just as we did with the scaling equation. Associated to the wavelet equation is a useful L^2 periodic function m_1 of period 1 that behaves like a generating function. Namely there exists a periodic L^2 function m_1 of period 1 such that m_1 is given on [0, 1] by a Fourier series as

$$m_1(y) = \frac{1}{\sqrt{2}}\sum_{k=-\infty}^{\infty} b_k e^{-2\pi i k y}.$$

This function has

$$\|m_1\|^2_{L^2([0,1])} = \frac{1}{2}\sum_{k=-\infty}^{\infty}|b_k|^2 = \frac{1}{2}.$$

The Fourier transform of the wavelet equation is

$$(\mathcal{F}\psi)(y) = \sum_{k=-\infty}^{\infty}\frac{1}{\sqrt{2}}b_k e^{-\pi i k y}(\mathcal{F}\varphi)(\tfrac{1}{2}y),$$

and the generating function m_1 allows us to write the Fourier transform of the wavelet equation in the tidy form

$$(\mathcal{F}\psi)(y) = m_1(y/2)(\mathcal{F}\varphi)(y/2).$$

EXAMPLE. In the case of the Haar system, Step 6 of the proof of Theorem 10.4 shows that the scaling equation and the wavelet equation are

$$\varphi(x) = \varphi(2x) + \varphi(2x-1)$$
$$\psi(x) = \varphi(2x) - \varphi(2x-1).$$

In other words, $a_0 = 1/\sqrt{2}$ and $a_1 = 1/\sqrt{2}$, while $b_0 = 1/\sqrt{2}$ and $b_1 = -1/\sqrt{2}$. The functions m_0 and m_1 are given by

$$m_0(y) = \tfrac{1}{2}(1 + e^{-2\pi iy})$$
$$m_1(y) = \tfrac{1}{2}(1 - e^{-2\pi iy}),$$

and we have

$$m_1(y) = e^{2\pi iy}(-e^{-4\pi iy})\overline{m_0(y+\tfrac{1}{2})}.$$

Let us return to the specific choice of a wavelet in the general case. According to Theorem 10.10, we are to specify an L^2 function $\nu(y)$ of period 1 with $|\nu(y)| = 1$ almost everywhere, and then the formula for m_1 in terms of m_0 (after replacement of $y/2$ by y) will be

$$m_1(y) = e^{2\pi iy}\nu(y)\overline{m_0(y+\tfrac{1}{2})}.$$

The seemingly natural choice is to take $\nu(y) = 1$, and we shall make that choice for a while in the proof of Theorem 10.10b, obtaining

$$m_1(y) = e^{2\pi iy}\overline{m_0(y+\tfrac{1}{2})}.$$

That choice leads in the general case to the formula

$$\boxed{b_k = (-1)^{k+1}\,\overline{a_{-k-1}}}$$

for the coefficients of the wavelet equation.

The example of the Haar system shows that other choices for ν will sometimes be appropriate; for it the choice $\nu(y) = -e^{-4\pi iy}$ was what produced the Haar wavelet from the Haar scaling function. The choice $\nu(y) = -e^{-4\pi iy}$ leads to the formulas

$$m_1(y) = -e^{-2\pi iy}\overline{m_0(y+\tfrac{1}{2})}$$

and

$$b_k = (-1)^k\,\overline{a_{-k+1}}.$$

As a general matter, taking $\nu(y)$ to be the product of a power of $e^{2\pi iy}$ and a constant of absolute value one has the effect of moving the wavelet left or right so that it looks better. This will be especially convenient later for the Daubechies wavelets, which are of compact support. Problem 3 at the end of the chapter shows that the only valid choices of $\nu(y)$ in the case of compact support are the products of a power of $e^{2\pi iy}$ and a constant of absolute value one.

Now we come to the three lemmas. The first and the third will be used in the course of proving Theorem 10.10. The second lemma will not be used until Section 7, but the techniques for its proof take some of the mystery out of the proof of Theorem 10.10.

Lemma 10.11. If $h(x)$ is a function in $L^2(\mathbb{R})$, then the set of functions $\{h(x-k)\}_{k=-\infty}^{\infty}$ is orthonormal if and only if

$$\sum_{l=-\infty}^{\infty} |(\mathcal{F}h)(y+l)|^2 = 1$$

almost everywhere.

PROOF. By the Plancherel Theorem (Section VIII.3 of *Basic*) we have

$$\begin{aligned}
\delta_{0k} &\stackrel{?}{=} \int_{\mathbb{R}} h(x)\overline{h(x-k)}\,dx \\
&= \int_{\mathbb{R}} (\mathcal{F}h)(y)\overline{e^{-2\pi iky}(\mathcal{F}h)(y)}\,dy \\
&= \int_{\mathbb{R}} e^{2\pi iky}|(\mathcal{F}h)(y)|^2\,dy \\
&= \sum_{l=-\infty}^{\infty} \int_l^{l+1} e^{2\pi iky}|(\mathcal{F}h)(y)|^2\,dy \\
&= \sum_{l=-\infty}^{\infty} \int_0^1 e^{2\pi ik(y'+l)}|(\mathcal{F}h)(y'+l)|^2\,dy' \\
&= \sum_{l=-\infty}^{\infty} \int_0^1 e^{2\pi iky'}|(\mathcal{F}h)(y'+l)|^2\,dy'. \qquad (*)
\end{aligned}$$

If we were to insert absolute value signs in the integrand of $(*)$, we would obtain the result $\int_{\mathbb{R}} |(\mathcal{F}h)(y')|^2\,dy' = \int_{\mathbb{R}} |h(x)|^2\,dx$ by the Plancherel formula (Theorem 8.6 of *Basic)*, and this is finite. Therefore Fubini's Theorem allows us to interchange sum and integral in $(*)$. Doing the interchange and changing y' back to y, results in the equation

$$\delta_{0k} \stackrel{?}{=} \int_0^1 e^{2\pi iky}\Big(\sum_{l=-\infty}^{\infty} |(\mathcal{F}h)(y+l)|^2\Big)\,dy. \qquad (**)$$

This says that the system is orthonormal if and only if the function in $L^1([0,1])$ given by the almost everywhere convergent series $\sum_{l=-\infty}^{\infty} |(\mathcal{F}h)(y+l)|^2$ has the same Fourier coefficients as the constant function 1. By the uniqueness theorem for Fourier series (Corollary 6.50 of *Basic*), we obtain the result that the system is orthonormal if and only if

$$\sum_{l=-\infty}^{\infty} |(\mathcal{F}h)(y+l)|^2 = 1 \qquad \text{almost everywhere.} \qquad \square$$

Lemma 10.12. Suppose $\varphi(x)$ is a function in $L^2(\mathbb{R})$ such that the set of functions $\{\varphi_{0,k}\}_{k=-\infty}^{\infty} = \{x \mapsto \varphi(x-k)\}_{k=-\infty}^{\infty}$ is orthonormal, and let V be the closure of the linear span of this set of functions. Let ℓ^2 be the set of square integrable doubly infinite sequences $\{c_k\}_{k=-\infty}^{\infty}$, and write $\delta_{\cdot\,k}$ for the member of ℓ^2 that is 1 at k and is 0 otherwise. Parseval's Theorem produces unitary[9] mappings $\alpha : \ell^2 \to V$ and $\beta : L^2([0,1]) \to \ell^2$ such that $\alpha(\delta_{\cdot\,k}) = \varphi_{0,k}$ and $\beta(e^{-2\pi iky}) = \delta_{\cdot\,k}$. Regard members of $L^2([0,1])$ as extended periodically to all of $\mathbb{R}$ with period 1. Then for every $f \in V$,

(a) the product of $(\beta^{-1}\alpha^{-1})(f)$ and $\mathcal{F}\varphi$ equals $\mathcal{F}f$, and
(b) $\|(\beta^{-1}\alpha^{-1})(f)\|_{L^2([0,1])} = \|(\beta^{-1}\alpha^{-1})(f)\cdot(\mathcal{F}\varphi)\|_{L^2(\mathbb{R})}$.

PROOF. In the statement of the lemma, the unitary property of α follows from the assumption that $\{\varphi_{0,k}\}$ is an orthonormal basis of V and from the abstract Parseval Theorem for Hilbert spaces. The unitary property of β follows from the Parseval Theorem for Fourier series and the Riesz–Fischer Theorem.

In (a) the mapping $\beta^{-1}\alpha^{-1}$ acts on V by $\beta^{-1}\alpha^{-1}\big(\sum_k c_k\varphi_{0,k}\big) = \sum_k c_k e^{-2\pi iky}$, the convergence of the sum on the left being in $V \subseteq L^2(\mathbb{R})$ and the convergence of the sum on the right being in $L^2([0,1])$. Since $\mathcal{F}\varphi_{0,k} = e^{-2\pi iky}(\mathcal{F}\varphi)(y) = \beta^{-1}\alpha^{-1}(\varphi_{0,k})(\mathcal{F}\varphi)(y)$, we immediately obtain $\mathcal{F}f = (\beta^{-1}\alpha^{-1})(f)\cdot\mathcal{F}\varphi$ for all members f of the vector-space linear span of $\{\varphi_{0,k}\}$. Also if $f_n \to f$ in $L^2(\mathbb{R})$, we know that $(\beta^{-1}\alpha^{-1})(f_n) \to (\beta^{-1}\alpha^{-1})(f)$ in $L^2([0,1])$. This proves (a).

For the norm equality of (b), let us abbreviate $(\beta^{-1}\alpha^{-1}f)(y)$ as $\mu(y)$. Then

$$\begin{aligned}
\int_{\mathbb{R}} |\mu(y)(\mathcal{F}\varphi)(y)|^2\,dy &= \sum_{k=-\infty}^{\infty}\int_k^{k+1} |\mu(y)|^2|(\mathcal{F}\varphi)(y)|^2\,dy \\
&= \sum_{k=-\infty}^{\infty}\int_0^1 |\mu(y-k)|^2|(\mathcal{F}\varphi)(y-k)|^2\,dy \\
&= \sum_{k=-\infty}^{\infty}\int_0^1 |\mu(y)|^2|(\mathcal{F}\varphi)(y-k)|^2\,dy \quad \text{by periodicity}
\end{aligned}$$

[9]"Unitary" means linear, norm-preserving, and onto.

$$= \int_0^1 |\mu(y)|^2 \sum_{k=-\infty}^{\infty} |(\mathcal{F}\varphi)(y-k)|^2 \, dy$$

$$= \int_0^1 |\mu(y)|^2 \, dy \qquad \text{by Lemma 10.11.}$$

□

Lemma 10.13. If a multiresolution analysis is given with scaling function φ, then the generating function m_0 of the scaling equation satisfies the identity

$$|m_0(y)|^2 + |m_0(y + \tfrac{1}{2})|^2 = 1 \qquad \text{almost everywhere.}$$

REMARKS. This lemma gives a sense in which the generating functions $m_0(y)$ and $m_1(y)$ are complementary. The method of proof will be used several times in what follows.

PROOF. We apply Lemma 10.11 with $h = \varphi$. Substituting for $(\mathcal{F}\varphi)(y+l)$ its value from the Fourier transform of the scaling equation rewrites this equality as

$$\sum_{l=-\infty}^{\infty} |m_0(\tfrac{1}{2}y + \tfrac{1}{2}l)|^2 |(\mathcal{F}\varphi)(\tfrac{1}{2}y + \tfrac{1}{2}l)|^2 = 1 \qquad \text{a.e.}$$

Replacing $y/2$ by y in this relation shows that almost everywhere

$$\sum_{l=-\infty}^{\infty} |m_0(y + \tfrac{1}{2}l)|^2 |(\mathcal{F}\varphi)(y + \tfrac{1}{2}l)|^2 = 1.$$

We separate the even-numbered terms on the left from the odd-numbered terms and use that m_0 is periodic of period 1 to see that

$$\begin{aligned} 1 &= \sum_{l \text{ even}} |m_0(y + \tfrac{1}{2}l)|^2 |(\mathcal{F}\varphi)(y + \tfrac{1}{2}l)|^2 + \sum_{l \text{ odd}} |m_0(y + \tfrac{1}{2}l)|^2 |(\mathcal{F}\varphi)(y + \tfrac{1}{2}l)|^2 \\ &= |m_0(y)|^2 \sum_{l \text{ even}} |(\mathcal{F}\varphi)(y + \tfrac{1}{2}l)|^2 + |m_0(y + \tfrac{1}{2})|^2 \sum_{l \text{ odd}} |(\mathcal{F}\varphi)(y + \tfrac{1}{2}l)|^2 \\ &= |m_0(y)|^2 \sum_{l=-\infty}^{\infty} |(\mathcal{F}\varphi)(y + l)|^2 + |m_0(y + \tfrac{1}{2})|^2 \sum_{l=-\infty}^{\infty} |(\mathcal{F}\varphi)(y + l + \tfrac{1}{2})|^2 \\ &= |m_0(y)|^2 + |m_0(y + \tfrac{1}{2})|^2 \qquad \text{almost everywhere,} \end{aligned}$$

the last step holding by two applications of Lemma 10.11. □

PROOF OF THEOREM 10.10a. Let f be an arbitrary member of the orthogonal complement W_0 of V_0 in V_1. Since the functions $\varphi_{1,k}$ form an orthonormal basis of V_1, we can expand f in this orthonormal basis and obtain a series $f = \sum_{k=-\infty}^{\infty} f_k \varphi_{1,k}$ convergent in $L^2(\mathbb{R})$. In this expansion the coefficients are inner products $f_k = (f, \varphi_{1,k})$ satisfying $\sum_k |f_k|^2 = \|f\|_2^2$. Substituting the definition of $\varphi_{1,k}$ allows us to rewrite this expansion as

$$f(x) = \sum_{k=-\infty}^{\infty} f_k \sqrt{2}\, \varphi(2x - k).$$

Arguing as with φ and the scaling equation, we obtain

$$(\mathcal{F}f)(y) = \sum_{k=-\infty}^{\infty} \tfrac{1}{\sqrt{2}} f_k e^{-\pi i k y} (\mathcal{F}\varphi)(\tfrac{1}{2} y) = m_f(y/2)(\mathcal{F}\varphi)(y/2), \qquad (*)$$

where m_f is the function of period 1 given on [0, 1] by its Fourier series as

$$m_f(y) = \tfrac{1}{\sqrt{2}} \sum_{k=-\infty}^{\infty} f_k e^{-2\pi i k y}.$$

This function satisfies

$$\|m_f\|_{L^2([0,1])}^2 = \tfrac{1}{2} \sum_{k=-\infty}^{\infty} |f_k|^2 = \tfrac{1}{2} \|f\|_2^2. \qquad (**)$$

The condition that f is orthogonal to V_0 means that f is orthogonal to all $\varphi_{0,k}$. By the Plancherel Theorem (Theorem 8.6 of *Basic*) this is the condition that

$$0 = \int_{\mathbb{R}} (\mathcal{F}f)(y) \overline{e^{-2\pi i k y} (\mathcal{F}\varphi)(y)}\, dy \qquad \text{for } k \in \mathbb{Z}.$$

Arguing as with $(*)$ and $(**)$ in the proof of Lemma 10.11, we can rewrite this condition as

$$0 = \int_0^1 e^{2\pi i k y} \Big(\sum_{l=-\infty}^{\infty} (\mathcal{F}f)(y + l) \overline{(\mathcal{F}\varphi)(y + l)} \Big)\, dy.$$

In other words, the function in $L^1([0, 1])$ given almost everywhere by the absolutely convergent series $\sum_{l=-\infty}^{\infty} (\mathcal{F}f)(y+l)\overline{(\mathcal{F}\varphi)(y+l)}$ has all Fourier coefficients 0. From the uniqueness theorem for Fourier series (Corollary 6.50 of *Basic*), we obtain the almost-everywhere equality

$$\sum_{l=-\infty}^{\infty} (\mathcal{F}f)(y + l) \overline{(\mathcal{F}\varphi)(y + l)} = 0.$$

Substituting for $(\mathcal{F}f)(y+l)$ and $(\mathcal{F}\varphi)(y+l)$ the values from $(*)$ and the Fourier transform of the scaling equation allows us to rewrite this equality as

$$\sum_{l=-\infty}^{\infty} m_f(\tfrac{1}{2}y + \tfrac{1}{2}l)\overline{m_0(\tfrac{1}{2}y + \tfrac{1}{2}l)}\,(\mathcal{F}\varphi)(\tfrac{1}{2}y + \tfrac{1}{2}l)\overline{(\mathcal{F}\varphi)(\tfrac{1}{2}y + \tfrac{1}{2}l)} = 0.$$

Replacement of $y/2$ by y in this relation shows that almost everywhere

$$\sum_{l=-\infty}^{\infty} m_f(y + \tfrac{1}{2}l)\overline{m_0(y + \tfrac{1}{2}l)}\,|(\mathcal{F}\varphi)(y + \tfrac{1}{2}l)|^2 = 0.$$

Just as in the proof of Lemma 10.13, we separate the even-numbered terms in the sum from the odd-numbered terms and use that m_0 and m_f are periodic of period 1 to see that

$$m_f(y)\overline{m_0(y)} + m_f(y + \tfrac{1}{2})\overline{m_0(y + \tfrac{1}{2})} = 0 \qquad (\dagger)$$

almost everywhere.

For almost every y, at least one of $m_0(y)$ and $m_0(y + \frac{1}{2})$ is nonzero, according to Lemma 10.13. If y has the property that both $m_0(y)$ and $m_0(y+\frac{1}{2})$ are nonzero, then we can set

$$\lambda(y) = \frac{m_f(y)}{\overline{m_0(y + \frac{1}{2})}} \qquad \text{and} \qquad \lambda(y + \tfrac{1}{2}) = \frac{m_f(y + \frac{1}{2})}{\overline{m_0(y)}}, \qquad (\dagger\dagger)$$

and we see from $(\dagger)$ that

$$\lambda(y) + \lambda(y + \tfrac{1}{2}) = 0. \qquad (\ddagger)$$

If $m_0(y) = 0$ and $m_0(y + \frac{1}{2}) \neq 0$, then we define $\lambda(y)$ by the first equation in $(\dagger\dagger)$ and $\lambda(y+\frac{1}{2})$ by $(\ddagger)$, while if $m_0(y) \neq 0$ and $m_0(y + \frac{1}{2}) = 0$, then we define $\lambda(y + \frac{1}{2})$ by the second equation in $(\dagger\dagger)$ and $\lambda(y)$ by $(\ddagger)$. The result is that we have a definition almost everywhere of a periodic function λ of period 1 such that $(\ddagger)$ holds almost everywhere and such that

$$m_f(y) = \lambda(y)\overline{m_0(y + \tfrac{1}{2})} \qquad (\ddagger\ddagger)$$

almost everywhere. This equation shows that

$$\begin{aligned}
\|m_f\|^2_{L^2[0,1]} &= \textstyle\int_0^1 |\lambda(y)|^2 |m_0(y + \frac{1}{2})|^2\,dy \\
&= \textstyle\int_0^{1/2} |\lambda(y)|^2 |m_0(y + \frac{1}{2})|^2\,dy + \int_{1/2}^1 |\lambda(y)|^2 |m_0(y + \frac{1}{2})|^2\,dy \\
&= \textstyle\int_0^{1/2} |\lambda(y)|^2 |m_0(y + \frac{1}{2})|^2\,dy + \int_0^{1/2} |\lambda(y)|^2 |m_0(y)|^2\,dy \quad \text{by } (\ddagger) \\
&= \textstyle\int_0^{1/2} |\lambda(y)|^2\,dy \qquad \text{by Lemma 10.13.}
\end{aligned} \qquad (\S)$$

From (‡) it follows that $\nu(2y) = e^{-2\pi i y}\lambda(y)$ with $\nu(y)$ periodic of period 1. We substitute from (‡‡) into (∗) and obtain

$$(\mathcal{F}f)(y) = m_f(\tfrac{1}{2}y)(\mathcal{F}\varphi)(\tfrac{1}{2}y) = \lambda(\tfrac{1}{2}y)\overline{m_0(\tfrac{1}{2}y+\tfrac{1}{2})}(\mathcal{F}\varphi)(\tfrac{1}{2}y).$$

Thus

$$(\mathcal{F}f)(y) = e^{\pi i y}\nu(y)\overline{m_0(\tfrac{1}{2}y+\tfrac{1}{2})}(\mathcal{F}\varphi)(\tfrac{1}{2}y),$$

as required.

Finally we have

$$\int_0^1 |\nu(y)|^2\,dy = 2\int_0^{1/2}|\nu(2y)|^2\,dy = 2\int_0^{1/2}|\lambda(y)|^2\,dy = 2\|m_f\|^2_{L^2[0,1]} = \|f\|_2^2 \tag{§§}$$

by definition of ν and use of (§) and (∗∗). □

PROOF OF THEOREM 10.10b. Taking $\nu = 1$, we define ψ in $L^2(\mathbb{R})$ by

$$(\mathcal{F}\psi)(y) = e^{\pi i y}\overline{m_0(\tfrac{1}{2}y+\tfrac{1}{2})}(\mathcal{F}\varphi)(\tfrac{1}{2}y).$$

We prove that $\{\psi(y-k)\}_{k=-\infty}^{\infty}$ is an orthonormal basis of W_0. To see that it is orthonormal, we use Lemma 10.11, breaking the sum for ψ into even-numbered and odd-numbered terms and applying Lemma 10.11 for φ in each of the terms. We have

$$\begin{aligned}\sum_l |\mathcal{F}\psi(y+l)|^2 &= \sum_l |m_0(\tfrac{1}{2}y+\tfrac{1}{2}l+\tfrac{1}{2})|^2|\mathcal{F}\varphi(\tfrac{1}{2}y+\tfrac{1}{2}l)|^2\\ &= |m_0(\tfrac{1}{2}y+\tfrac{1}{2})|^2\sum_l|\mathcal{F}\varphi(\tfrac{1}{2}y+l)|^2\\ &\qquad + |m_0(\tfrac{1}{2}y)|^2\sum_l|\mathcal{F}\varphi(\tfrac{1}{2}y+\tfrac{1}{2}+l)|^2\\ &= |m_0(\tfrac{1}{2}y+\tfrac{1}{2})|^2 + |m_0(\tfrac{1}{2}y)|^2 \qquad \text{a.e. by Lemma 10.11 for } \varphi\\ &= 1 \qquad \text{a.e. by Lemma 10.13,}\end{aligned}$$

and thus Lemma 10.11 indeed allows us to conclude that $\{\psi(y-k)\}_{k=-\infty}^{\infty}$ is an orthonormal set.

To show completeness when $\nu = 1$, we are to show that any f in W_0 has an expansion in $L^2(\mathbb{R})$ as $f = \sum_k f_k\psi_{0,k}$. This is the question whether equality holds in Bessel's inequality $\|f\|_2^2 \geq \sum_k |f_k|^2$, the coefficient f_k being given by $f_k = \int_{\mathbb{R}} f(x)\overline{\psi(x-k)}\,dx$.

We know from (a) that any f in W_0 is of the form $(\mathcal{F}f)(y) = \nu(y)(\mathcal{F}\psi)(y)$ for a periodic function ν of period 1 with $\|\nu\|^2_{L^2[0,1]} = \|f\|_2^2$. By Parseval's equality for $L^2[0, 1]$, we can write $\nu(x) = \sum_k c_k e^{-2\pi ikx}$ with $\sum_k |c_k|^2 = \|\nu\|^2_{L^2[0,1]}$, and (§§) has shown that $\|\nu\|^2_{L^2[0,1]} = \|f\|_2^2$. Thus completeness will follow if we show that $f_k = c_k$ for all k.

The verification that $f_k = c_k$ is rather similar to the argument in Lemma 10.11. On the one hand, direct computation gives

$$\begin{aligned} f_k &= \int_{\mathbb{R}} f(x)\overline{\psi(x-k)}\,dx \\ &= \int_{\mathbb{R}} (\mathcal{F}f)(y)\overline{(\mathcal{F}\psi_{0,k})(y)} \\ &= \int_{\mathbb{R}} (\mathcal{F}f)(y)\overline{e^{-2\pi iky}(\mathcal{F}\psi)(y)} \\ &= \int_{\mathbb{R}} \nu(y)e^{2\pi iky}|(\mathcal{F}\psi)(y)|^2\,dy \\ &= \sum_{l=-\infty}^{\infty} \int_l^{l+1} \nu(y)e^{2\pi iky}|(\mathcal{F}\psi)(y)|^2\,dy \\ &= \sum_{l=-\infty}^{\infty} \int_0^1 \nu(y)e^{2\pi iky}|(\mathcal{F}\psi)(y+l)|^2\,dy, \end{aligned}$$

and on the other hand, Lemma 10.11 gives

$$\begin{aligned} c_k &= \int_0^1 \nu(y)e^{2\pi iky}\,dy \\ &= \int_0^1 \sum_{l=-\infty}^{\infty} \nu(y)e^{2\pi iky}|(\mathcal{F}\psi)(y+l)|^2\,dy. \end{aligned}$$

Thus the equality $f_k \overset{?}{=} c_k$ comes down to an interchange of the limit and the sum. If we were instead to consider the expression

$$\int_0^1 \sum_{l=-\infty}^{\infty} |\nu(y)||(\mathcal{F}\psi)(y+l)|^2\,dy,$$

the result would be $\int_0^1 |\nu(y)|\,dy \le \|\nu\|^2_{L^2([0,1])} = \|f\|_2^2$, which is finite. Therefore Fubini's Theorem shows the interchange to be justified, and we indeed have $f_k = c_k$ for all k. This proves completeness in W_0 of the orthonormal system $\{\psi_{0,k}\}_{k=-\infty}^{\infty}$.

Now suppose that $\nu(y)$ is any function in $L^2([0, 1])$ periodic of period 1 such that $|\nu(y)| = 1$ almost everywhere. Let f be the function in W_0 that corresponds to ν, i.e., has $(\mathcal{F}f)(y) = \nu(y)(\mathcal{F}\psi)(y)$. Applying both directions of Lemma 10.11 to the identity

$$\sum_{l=-\infty}^{\infty} |\mathcal{F}f(y+l)|^2 = |\nu(y)|^2 \sum_{l=-\infty}^{\infty} |\mathcal{F}\psi(y+l)|^2 = \sum_{l=-\infty}^{\infty} |\mathcal{F}\psi(y+l)|^2$$

we see that $\{x \mapsto f(x-k)\}_{k=-\infty}^{\infty}$ is an orthonormal set in W_0. Suppose that g is given in W_0 and that g arises from a periodic function $\mu(y)$ of period 1 by the formula $(\mathcal{F}g)(y) = \mu(y)(\mathcal{F}\psi)(y)$. Then

$$(\mathcal{F}g)(y) = (\mu(y)\nu(y)^{-1})\nu(y)(\mathcal{F}\psi)(y) = (\mu(y)\nu(y)^{-1})(\mathcal{F}f)(y),$$

and the function $\mu(y)\nu(y)^{-1}$ is in $L^2([0,1])$. The same argument that shows the equality $f_k = c_k$ above shows that g is in the closed linear span of the functions $\{x \mapsto f(x-k)\}_{k=-\infty}^{\infty}$, and therefore $\mathcal{F}^{-1}(\nu(\mathcal{F}\psi))$ is a wavelet, as was asserted.

Finally if $\nu(y)$ is any periodic function of period 1 such that the function h with $(\mathcal{F}h)(y) = \nu(y)(\mathcal{F}\psi)(y)$ is a wavelet, then application of both halves of Lemma 10.11 shows that

$$1 = \sum_{l=-\infty}^{\infty} |\mathcal{F}h(y+l)|^2 = |\nu(y)|^2 \sum_{l=-\infty}^{\infty} |\mathcal{F}\psi(y+l)|^2 = |\nu(y)|^2$$

almost everywhere, and $|\nu(y)| = 1$ almost everywhere. □

4. Shannon Wavelet

We saw in Section 2 how the Haar wavelet could be defined and analyzed immediately from its definition. The advantage of the Haar system, not seen with Fourier series, is that classes of functions defined by their size are handled well by the Haar system. On the other hand, the functions $\psi_{j,k}$ of the Haar system are discontinuous, and wavelet expansions therefore cannot take advantage of smoothness of the function being expanded.

Thus other wavelets are needed with properties better suited for other purposes. The trouble is that the definitions of other wavelets are not so transparent. In this section we introduce the first of several useful examples of wavelets that are constructed directly from a multiresolution analysis with the aid of Theorem 10.10. With most such examples some supplementary argument is needed because the hypotheses of Theorem 10.10 do not directly fit with the example.

The example in this section, called the **Shannon system**, is exceptional in that verification of the hypotheses is fairly straightforward. It has

$$\varphi(x) = \frac{\sin(\pi x)}{\pi x} \qquad \text{for } x \in \mathbb{R}$$

as scaling function and $V_j = \{f \in L^2(\mathbb{R}) \,\big|\, (\mathcal{F}f)(y) = 0 \text{ for } |y| > 2^{j-1}\}$ as the j^{th} closed subspace. As usual, φ is to lie in V_0. The function φ arose in the Shannon Sampling Theorem (Proposition 10.3) when $\Omega = 1$. The proof of that result showed that $(\mathcal{F}\varphi)(y)$ equals the indicator function $I_{[-\frac{1}{2},\frac{1}{2}]}(y)$ almost everywhere on $\mathbb{R}$. Thus indeed φ lies in V_0.

Theorem 10.14. The Shannon system, consisting of the set of closed subspaces $\{V_j\}_{j=-\infty}^{\infty}$ and the function φ, is a multiresolution analysis. The corresponding wavelet may be taken to be

$$\psi(x) = -\frac{2\big(\sin(2\pi x) - \cos(\pi x)\big)}{\pi(2x-1)} \qquad \text{for } x \in \mathbb{R}.$$

Its Fourier transform almost everywhere equals

$$(\mathcal{F}\psi)(y) = e^{-\pi i y}(I_{[-1,-\frac{1}{2}]}(y) + I_{[\frac{1}{2},1]}(y)) \qquad \text{for } y \in \mathbb{R}.$$

REMARK. Graphs of φ and ψ appear in Figure 10.3. A new ingredient with this example, not seen with the Haar system, is that the scaling equation has infinitely many nonzero coefficients. We get at these coefficients through the associated generating function m_0 defined by $(\mathcal{F}\varphi)(2y) = m_0(y)(\mathcal{F}\varphi)(y)$, working with m_0 as a whole rather than with the individual coefficients. The function m_0 is periodic of period 1.

(a)

(b)

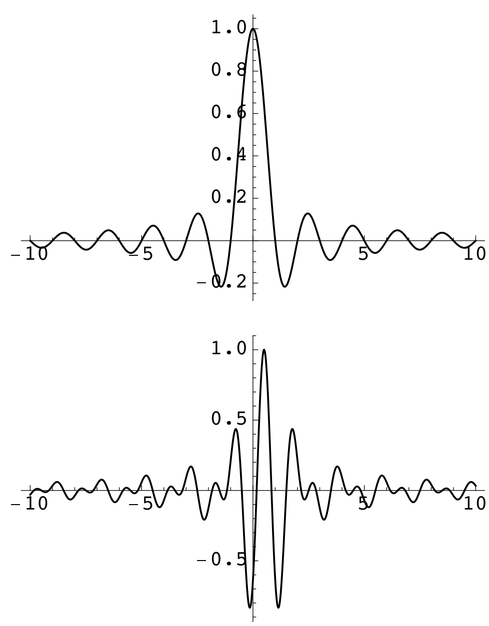


FIGURE 10.3. Graphs of Shannon scaling function and wavelet.
(a) Scaling function. (b) Wavelet.

PROOF. The closed subspaces V_j are nested, and their union is the set of functions in $L^2(\mathbb{R})$ whose Fourier transforms have compact support. This space is dense in $L^2(\mathbb{R})$ by the Plancherel Theorem. The intersection of the spaces V_j is the subspace of functions in $L^2(\mathbb{R})$ whose Fourier transforms are supported at 0, and that subspace is 0. Finally we have $(\mathcal{F}\varphi_{0,k})(y) = e^{-2\pi iky} I_{[-\frac{1}{2},\frac{1}{2}]}(y)$. Since the functions $e^{-2\pi iky}$ form an orthonormal basis of $L^2([0,1])$, the functions $\mathcal{F}\varphi_{0,k}$ form an orthonormal basis of $\mathcal{F}(V_0)$, and the functions $\varphi_{0,k}$, by the Plancherel Theorem, form an orthonormal basis of V_0. Thus the Shannon system is a multiresolution analysis.

To get at ψ, we start from the formula $(\mathcal{F}\varphi)(2y) = m_0(y)(\mathcal{F}\varphi)(y)$, and we find for $|y| \le \frac{1}{2}$ that

$$m_0(y) = \frac{(\mathcal{F}\varphi)(2y)}{\mathcal{F}\varphi(y)} = \frac{I_{[-\frac{1}{2},\frac{1}{2}]}(2y)}{I_{[-\frac{1}{2},\frac{1}{2}]}(y)} = I_{[-\frac{1}{4},\frac{1}{4}]}(y).$$

For $|y| > \frac{1}{2}$, the function is to be extended periodically with period 1. Then we have

$$(\mathcal{F}\psi)(2y) = m_1(y)(\mathcal{F}\varphi)(y).$$

The formula for $m_1(y)$ involves $\overline{m_0(y+\frac{1}{2})}$, which is given by

$$\overline{m_0(y+\tfrac{1}{2})} = \begin{cases} 1 & \text{for } -\frac{3}{4} \le y \le -\frac{1}{4}, \\ 0 & \text{for } -\frac{1}{4} \le y \le \frac{1}{4}, \\ 1 & \text{for } \quad \frac{1}{4} \le y \le \frac{3}{4}, \\ 0 & \text{for } \quad \frac{3}{4} \le y \le \frac{5}{4}. \end{cases}$$

For $|y| \le \frac{3}{4}$, this equals $I_{[-\frac{3}{4},-\frac{1}{4}]}(y) + I_{[\frac{1}{4},\frac{3}{4}]}(y)$, and thus

$$\overline{m_0(y+\tfrac{1}{2})} = I_{[-\frac{1}{2},-\frac{1}{4}]}(y) + I_{[\frac{1}{4},\frac{1}{2}]}(y) \qquad \text{for } |y| \le \tfrac{1}{2}.$$

Theorem 10.10a says that we can take ψ to be f in the formula

$$(\mathcal{F}f)(y) = e^{\pi iy}\nu(y)\overline{m_0(\tfrac{1}{2}y+\tfrac{1}{2})}(\mathcal{F}\varphi)(\tfrac{1}{2}y)$$

if we use $\nu(y) = e^{-2\pi iy}$. (We could as well use any other integer power of $e^{2\pi iy}$ as $\nu(y)$, and the effect will be to translate ψ by an integer. Our choice is arranged to make the graph in Figure 10.3b look pleasing.) In this case

$$(\mathcal{F}\psi)(2y) = e^{-2\pi iy}\overline{m_0(y+\tfrac{1}{2})}(\mathcal{F}\varphi)(y).$$

Since $(\mathcal{F}\varphi)(y) = I_{[-\frac{1}{2},\frac{1}{2}]}(y)$ for all $y \in \mathbb{R}$, we obtain

$$(\mathcal{F}\psi)(2y) = e^{-2\pi i y}\big(I_{[-\frac{1}{2},-\frac{1}{4}]}(y) + I_{[\frac{1}{4},\frac{1}{2}]}(y)\big)$$

and thus
$$(\mathcal{F}\psi)(y) = e^{-\pi i y}\big(I_{[-1,-\frac{1}{2}]}(y) + I_{[\frac{1}{2},1]}(y)\big),$$

as required. Forming the inverse Fourier transform of this function by direct computation yields

$$\psi(x) = \frac{2\big(\sin(2\pi(x-\frac{1}{2})) - \sin(\pi(x-\frac{1}{2}))\big)}{\pi(2x-1)} = -\frac{2\big(\sin(2\pi x) - \cos(\pi x)\big)}{\pi(2x-1)}. \quad \square$$

The wavelet ψ produced by Theorem 10.14 is called the **Shannon wavelet**. Since the Shannon wavelet is smooth, we have gotten around one defect of the Haar wavelet. But we have introduced other defects: the functions φ and ψ in the Shannon system fail to be in $L^1(\mathbb{R})$, and their Fourier transforms are discontinuous.

The examples in later sections will have more favorable smoothness properties, but each will have other drawbacks.

5. Construction of a Wavelet from a Scaling Function

We now aim for the Meyer wavelets, whose Fourier transforms have compact support and a specified degree of smoothness. Correspondingly the wavelets themselves are all of class C^∞ and have a specified degree of decrease at infinity. As is true with many examples, the spaces V_j in the relevant multiresolution analysis are hard to pin down without referring to the candidate for the scaling function φ. By contrast, in the Haar system the members of the spaces V_j were L^2 functions that were constant on certain kinds of intervals, and in the Shannon system the members of the spaces V_j were L^2 functions whose Fourier transforms vanished on certain sets.

However, when we are allowed to refer to φ, property (iv) of a multiresolution analysis shows that V_0 is always the closed subspace for which $\{x \mapsto \varphi(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis, and the other spaces V_j can be defined in terms of V_0 by the dilation property (iii). Thus the question arises how to tell whether a function φ in $L^2(\mathbb{R})$ for which $\{x \mapsto \varphi(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal set is the scaling function of a multiresolution analysis. The sequence $\{V_j\}$ will be increasing as soon as $V_0 \subseteq V_1$, i.e., as soon as φ satisfies a scaling equation. Evidently what is needed is to have reasonable sufficient conditions for the first two properties of a multiresolution analysis to be satisfied, namely

(i) $\bigcup_j V_j$ is dense in $L^2(\mathbb{R})$,
(ii) $\bigcap_j V_j = 0$.

We give such conditions in this section.

Proposition 10.15. Suppose that φ is a member of $L^2(\mathbb{R})$ such that

$$\{x \mapsto \varphi(x-k)\}_{k=-\infty}^{\infty}$$

is an orthonormal set. Let V_0 be the closed linear span of this orthonormal set, and define dilated spaces V_j by $V_j = \{f \in L^2(\mathbb{R}) \mid x \mapsto f(2^{-j}x) \text{ is in } V_0\}$. Then $\bigcap_{j=-\infty}^{\infty} V_j = 0$.

REMARK. In other words, property (ii) is automatic; no additional hypothesis on φ is needed. By contrast in Proposition 10.17 we shall impose an additional hypothesis on φ to ensure that property (i) holds.

PROOF. Define $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$. The set $\{\varphi_{j,k}\}_{k=-\infty}^{\infty}$ is an orthonormal basis of V_j, and thus any f in V_j satisfies Parseval's equality,

$$\|f\|_2^2 = \sum_{j=-\infty}^{\infty} |(f, \varphi_{j,k})|^2.$$

Let f be in $\bigcap_{j=-\infty}^{\infty} V_j$. We are to prove that $f = 0$. Let $\epsilon > 0$ be given, and choose $g \in C_{\text{com}}(\mathbb{R})$ with $\|f - g\|_2 \le \varepsilon$. Let M be large enough so that the interval $[-M, M]$ contains the support of g, and suppose that $j < 0$ is large enough so that $2^{-|j|}M < \frac{1}{2}$. If P_j denotes the orthogonal projection of $L^2(\mathbb{R})$ on V_j, then

$$\|f - P_j g\|_2 = \|P_j(f - g)\|_2 \le \|f - g\|_2 \le \varepsilon,$$

and hence

$$\|f\|_2 \le \varepsilon + \|P_j g\|_2. \tag{$*$}$$

Also we have

$$\begin{aligned}
\|P_j g\|_2^2 &= \sum_{k=-\infty}^{\infty} |(g, \varphi_{j,k})|^2 \\
&\le 2^{-|j|} \sum_{k=-\infty}^{\infty} \Big(\int_{x\in\mathbb{R}} |g(x)||\varphi(2^{-|j|}x - k)|\,dx \Big)^2 \\
&= 2^{-|j|} \sum_{k=-\infty}^{\infty} \Big(\int_{|x|\le M} |g(x)||\varphi(2^{-|j|}x - k)|\,dx \Big)^2 \\
&\le 2^{-|j|}\|g\|_{\text{sup}}^2 \sum_{k=-\infty}^{\infty} \Big(\int_{|x|\le M} |\varphi(2^{-|j|}x - k)|\,dx \Big)^2 \\
&\le 2^{-|j|}\|g\|_{\text{sup}}^2\, 2M \sum_{k=-\infty}^{\infty} \int_{|x|\le M} |\varphi(2^{-|j|}x - k)|^2\,dx \\
&\qquad\qquad \text{by the Schwarz inequality} \\
&= \|g\|_{\text{sup}}^2\, 2M \sum_{k=-\infty}^{\infty} \int_{|x'|\le 2^{-|j|}M} |\varphi(x' - k)|^2\,dx' \\
&= \|g\|_{\text{sup}}^2\, 2M \sum_{k=-\infty}^{\infty} \int_{|x+k|\le 2^{-|j|}M} |\varphi(x)|^2\,dx.
\end{aligned}$$

Since $2^{-|j|}M < \frac{1}{2}$, the sets of integration on the right side are disjoint, and we can write the right side as

$$= \|g\|_{\sup}^2 \, 2M \int_{E_j} |\varphi(x)|^2 \, dx,$$

where E_j is the set of reals at a distance of $\leq 2^{-|j|}M$ from $\mathbb{Z}$. As $|j|$ tends to infinity, these sets decrease to $\mathbb{Z}$, which has measure 0. In view of Corollary 5.3 of *Basic*, the complete additivity of the finite measure $|\varphi(x)|^2 \, dx$ implies that this expression tends to 0. By $(*)$, $\|f\|_2 \leq \varepsilon$. Since ε is arbitrary, f is the zero element of $L^2(\mathbb{R})$. □

Lemma 10.16. Suppose that φ is a member of $L^2(\mathbb{R})$ such that

$$\{x \mapsto \varphi(x-k)\}_{k=-\infty}^{\infty}$$

is an orthonormal set. Let V_0 be the closed linear span of this orthonormal set, and define dilated spaces V_j by $V_j = \{f \in L^2(\mathbb{R}) \mid x \mapsto f(2^{-j}x) \text{ is in } V_0\}$. For $j > 0$, let P_j be the orthogonal projection of $L^2(\mathbb{R})$ on V_j. If f is a member of $L^2(\mathbb{R})$ whose Fourier transform $\mathcal{F}f$ is bounded and is supported in the interval $[-M, M]$, then

$$\|P_j f\|_2^2 = \int_{-M}^{M} |(\mathcal{F}f)(y)|^2 \, |(\mathcal{F}\varphi)(2^{-j}y)|^2 \, dy$$

when j is large enough so that $2^{j-1} > M$.

PROOF. With $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$, we have

$$\begin{aligned}
\|P_j f\|_2^2 &= \sum_{k\in\mathbb{Z}} |(f, \varphi_{j,k})|^2 \\
&= \sum_{k\in\mathbb{Z}} \Big| \textstyle\int_{-M}^{M} (\mathcal{F}f)(y)\overline{(\mathcal{F}\varphi_{j,k})(y)} \, dy \Big|^2 \qquad \text{by the Plancherel Theorem} \\
&= \sum_{k\in\mathbb{Z}} \Big| \textstyle\int_{-M}^{M} (\mathcal{F}f)(y) e^{-2\pi i k 2^{-j} y} 2^{-j/2} \overline{(\mathcal{F}\varphi)(2^{-j}y)} \, dy \Big|^2 \\
&= \sum_{k\in\mathbb{Z}} 2^{-j} \Big| \textstyle\int_{-2^{j-1}}^{2^{j-1}} (\mathcal{F}f)(y) e^{-2\pi i k 2^{-j} y} \overline{(\mathcal{F}\varphi)(2^{-j}y)} \, dy \Big|^2, \qquad (*)
\end{aligned}$$

the last equality holding since $2^{j-1} > M$. The integral in the k^{th} term on the right is the k^{th} Fourier coefficient of an L^2 function on $[-2^{j-1}, 2^{j-1}]$, specifically of

$$(\mathcal{F}f)(y) \, \overline{(\mathcal{F}\varphi)(2^{-j}y)},$$

and Parseval's Theorem shows that the sum over $k \in \mathbb{Z}$ of the absolute values squared of these coefficients equals the norm squared of the function in the space $L^2([-2^{j-1}, 2^{j-1}], dx)$. Thus $(*)$ equals

$$\int_{-2^{j-1}}^{2^{j-1}} |(\mathcal{F}f)(y)\,\overline{(\mathcal{F}\varphi)(2^{-j}y)}|^2\,dy. \qquad \square$$

Proposition 10.17. Suppose that φ is a member of $L^2(\mathbb{R})$ such that

$$\{x \mapsto \varphi(x-k)\}_{k=-\infty}^{\infty}$$

is an orthonormal set. Let V_0 be the closed linear span of this orthonormal set, and define dilated spaces V_j by $V_j = \{f \in L^2(\mathbb{R}) \mid x \mapsto f(2^{-j}x) \text{ is in } V_0\}$. Under the additional hypothesis that $\mathcal{F}\varphi$ is a bounded function that is continuous and nonzero at $y=0$, the vector subspace $\sum_{j=-\infty}^{\infty} V_j$ generated by all the V_j is dense in $L^2(\mathbb{R})$.

REMARK. We shall use this result only when $V_0 \subseteq V_1$, in which case $\{V_j\}$ is an increasing sequence and $\sum_{j=-\infty}^{\infty} V_j = \bigcup_{j=-\infty}^{\infty} V_j$.

PROOF. Let P_j be the orthogonal projection on V_j. Arguing by contradiction, suppose that f is a nonzero element in the orthogonal complement of $\bigcup_{j=-\infty}^{\infty} V_j$. Certainly $P_j f = 0$ for all j. Let $\epsilon > 0$ be given, and assume that ϵ is small enough so that

$$\epsilon(1 + |(\mathcal{F}\varphi)(0)|) < |(\mathcal{F}\varphi)(0)|\|f\|_2. \qquad (*)$$

This is possible since $(\mathcal{F}\varphi)(0)$ and $\|f\|_2$ are nonzero. Choose a function g in $L^2(\mathbb{R})$ such that $\mathcal{F}g$ has compact support and $\|f-g\|_2 \le \epsilon$. Then $\|P_j g\|_2 = \|P_j(g-f)\|_2 \le \|g-f\|_2 \le \epsilon$.

Say that the support of $\mathcal{F}g$ is contained in the interval $[-M, M]$. Referring to Lemma 10.16, we then have

$$\epsilon^2 \ge \|P_j g\|_2^2 = \int_{-M}^{M} |(\mathcal{F}g)(y)|^2\,|\mathcal{F}\varphi(2^{-j}y)|^2\,dy. \qquad (**)$$

Since $\mathcal{F}\varphi$ is bounded and $\mathcal{F}g$ is square integrable and $\mathcal{F}\varphi$ is continuous at 0, the Dominated Convergence Theorem applies and shows that the right side of $(**)$ tends to

$$\int_{-M}^{M} |(\mathcal{F}g)(y)||(\mathcal{F}\varphi)(0)|^2\,dy$$

as j tends to $+\infty$. Thus we obtain

$$\epsilon^2 \ge |(\mathcal{F}\varphi)(0)|^2\|\mathcal{F}g\|_2^2 = |(\mathcal{F}\varphi)(0)|^2\|g\|_2^2 \ge |(\mathcal{F}\varphi)(0)|^2(\|f\|_2 - \epsilon)^2,$$

an inequality that contradicts $(*)$. $\square$

6. Meyer Wavelets

Now we can proceed with the construction of the Meyer wavelets. The definition of the Meyer multiresolution analysis involves the choice of a continuous function ν on $\mathbb{R}$ that is of class C^m or perhaps C^∞ and that is 0 for $y \le 0$ and is 1 for $y \ge 1$. It is assumed that ν satisfies

$$\nu(y) + \nu(1-y) = 1 \qquad \text{for } y \in \mathbb{R}.$$

This function will be incorporated into the definition of the Fourier transform of the scaling function φ.

There are no further assumptions. However, the interest in the Meyer wavelets is normally in the effect of the order of continuous differentiability on what happens, and at least when the order of continuous differentiability is finite, the convention is to choose the function ν of class C^m as computationally simple as possible so that it meets the above conditions.[10] With this understanding for each finite $m \ge 0$, there is a unique such ν whose polynomial part has lowest degree.[11] That degree turns out to be $2m+1$, and the polynomial is

$$x^{m+1} \sum_{k=0}^{m} \binom{m+k}{k} (1-x)^k.$$

The convention is to take ν to be that polynomial on [0, 1], extended by 0 for $x \le 0$ and extended by 1 for $x \ge 1$. We shall call the resulting multiresolution analysis the "Meyer multiresolution analysis" of **index** m. Figure 10.4 shows that graph of $\nu(x)$ for $m = 3$. The table in Figure 10.5 lists the polynomial explicitly for a few cases.

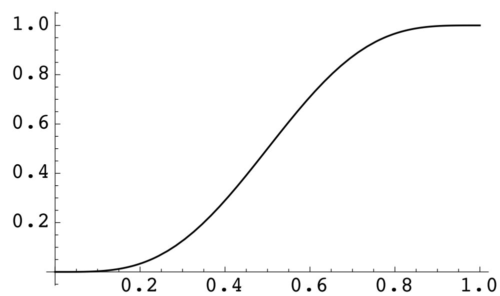


FIGURE 10.4. Graph of Meyer polynomial function when $m = 3$.

[10] Specifically the conditions are that ν is 0 for $y \le 0$, is 1 for $y \ge 1$, is a polynomial function between 0 and 1, satisfies $\nu(y) + \nu(1-y) = 1$, and is in the class C^m on $\mathbb{R}$.

[11] See Problems 13–17 at the end of the chapter.

m	Polynomial $\nu(x)$ for $0 \le x \le 1$
0	x
1	$x^2(3 - 2x)$
2	$x^3(10 - 15x + 6x^2)$
3	$x^4(35 - 84x + 70x^2 - 20x^3)$
4	$x^5(126 - 420x + 540x^2 - 315x^3 + 70x^4)$
5	$x^6(462 - 1980x + 3465x^2 - 3080x^3 + 1386x^4 - 252x^5)$
6	$x^7(1716 - 9009x + 20020x^2 - 24024x^3 + 16380x^4 - 6006x^5 + 924x^6)$

FIGURE 10.5. Table of values of the polynomial part of $\nu(x)$ used in defining the Meyer wavelet of index m.

With ν in hand, we define the Meyer scaling function φ through its Fourier transform $\mathcal{F}\varphi$ by

$$(\mathcal{F}\varphi)(y) = \begin{cases} 1 & \text{for } |y| \le \frac{1}{3}, \\ \cos\left(\frac{\pi}{2}\nu(3|y| - 1)\right) & \text{for } \frac{1}{3} \le |y| \le \frac{2}{3}, \\ 0 & \text{for } \frac{2}{3} \le |y|. \end{cases}$$

A plot of $\mathcal{F}\varphi$ for the case with $m = 3$ in the above table appears in Figure 10.6.

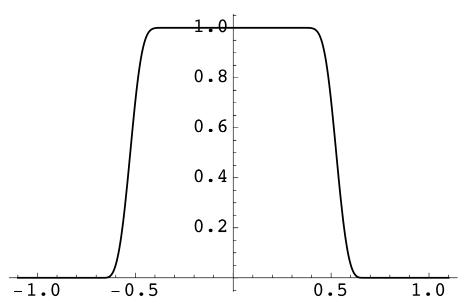


FIGURE 10.6. Graph of $\mathcal{F}\varphi$ for the Meyer wavelet when $m = 3$.

It will be seen in Theorem 10.18 below that $\{x \mapsto \varphi(x - k)\}_{k=-\infty}^{\infty}$ is an orthonormal set, hence that φ is a candidate for a scaling function. Then we let V_0 be the closed linear span of this orthonormal set and define dilated spaces V_j as usual by $V_j = \{f \in L^2(\mathbb{R}) \mid x \mapsto f(2^{-j}x) \text{ is in } V_0\}$. The spaces V_j will be seen to be nested. The **Meyer system corresponding to** ν consists of φ and the nested sequence of spaces $\{V_j\}_{j=-\infty}^{\infty}$.

Theorem 10.18. Let $\nu : \mathbb{R} \to [0, 1]$ be a continuous function such that

(i) $\nu(t) = 0$ for $t \le 0$,

(ii) $\nu(t) = 1$ for $t \geq 1$,
(iii) $\nu(t) + \nu(1-t) = 1$ for all $t \in \mathbb{R}$.

Then the Meyer system corresponding to ν is a multiresolution analysis. The corresponding wavelet ψ may be taken to have Fourier transform given by

$$(\mathcal{F}\psi)(y) = e^{\pi i y}\big((\mathcal{F}\varphi)(y+1) + (\mathcal{F}\varphi)(y-1)\big)(\mathcal{F}\varphi)(\tfrac{1}{2}y).$$

REMARKS. It is instructive to see how $e^{-\pi i y}(\mathcal{F}\psi)$ relates to $\mathcal{F}\varphi$. For this purpose, compare the graph of $\mathcal{F}\varphi$ in Figure 10.6 with that of $e^{-\pi i y}(\mathcal{F}\psi)$ in Figure 10.7 below. Figure 10.8 after the proof shows graphs of φ and ψ.

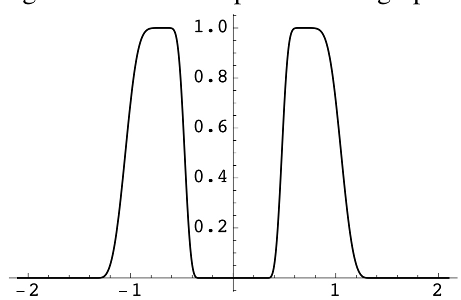


FIGURE 10.7. Graph of $e^{-\pi i y}\mathcal{F}\psi$ for the Meyer wavelet when $m = 3$.

PROOF. To see that φ and the spaces V_j form a multiresolution analysis, let us first compute

$$\textstyle\sum_l |(\mathcal{F}\varphi)(y+l)|^2. \qquad (*)$$

Expression $(*)$ is manifestly periodic of period 1, and we may assume that $0 \leq y \leq 1$. For $0 \leq y \leq \frac{1}{3}$, the only nonzero term is for $l = 0$, and it is 1. For $\frac{2}{3} \leq y \leq 1$, the only nonzero term is for $l = -1$, and it is 1. For $\frac{1}{3} \leq y \leq \frac{2}{3}$, the only terms that contribute are for $l = 0$ and $l = -1$, and the sum of the contributions is

$$\cos^2(\tfrac{\pi}{2}\nu(3y-1)) + \cos^2(\tfrac{\pi}{2}\nu(2-3y)). \qquad (**)$$

The arguments of ν in $(**)$ have sum 1, and therefore the sum of their values is 1, since we are assuming that $\nu(t) + \nu(1-t) = 1$ for all real t. Thus $(**)$ is of the form $\cos^2\frac{\pi}{2}u + \cos^2\frac{\pi}{2}(1-u) = \cos^2 u + \sin^2 u = 1$. In other words, $(*)$ equals 1 for all y. By Lemma 10.11, $\{x \mapsto \varphi(x-k)\}_{k=-\infty}^{\infty}$ is an orthonormal set.

As at the beginning of Section 5, we define V_0 to be the closed linear span of $\{x \mapsto \varphi(x-k)\}_{k\in\mathbb{Z}}$, and we put $V_j = \{f \mid x \mapsto f(2^{-j}x) \text{ is in } V_0\}$. We need to show that $V_0 \subseteq V_1$, i.e., that φ satisfies a scaling equation. We define a periodic function μ_0 of period 1 by

$$\mu_0(y) = \sum_{l=-\infty}^{\infty} (\mathcal{F}\varphi)(2y+2l). \qquad (\dagger)$$

In a neighborhood of any fixed y, this is a finite sum because $\mathcal{F}\varphi$ has compact support. Thus $\mu_0(y)$ is meaningful, and on $[0,1]$, the function μ_0 is square integrable. Let check that $\mu_0(y)$ satisfies

$$\mu_0(y/2)(\mathcal{F}\varphi)(y/2) = (\mathcal{F}\varphi)(y) \qquad \text{for all } y \in \mathbb{R}. \tag{††}$$

To prove (††), we observe for each $l \neq 0$ that the contribution to (††) given by $(\mathcal{F}\varphi)(y+2l)(\mathcal{F}\varphi)(y/2)$ is 0. In fact, for it to be nonzero requires that $|y+2l| < \frac{2}{3}$ (and hence $|y| > \frac{4}{3}$), as well as $|y/2| < \frac{2}{3}$, and there are no values of y where these conditions are met. Thus proving (††) comes down to proving the equality

$$(\mathcal{F}\varphi)(y) = (\mathcal{F}\varphi)(y)(\mathcal{F}\varphi)(y/2) \qquad \text{for all } y \in \mathbb{R}. \tag{‡}$$

There are two cases. One case is that $|y| \geq \frac{2}{3}$, and then $(\mathcal{F}\varphi)(y) = 0$ and both sides are equal. The other case is that $|y| \leq \frac{2}{3}$, and then $|y/2| \leq \frac{1}{3}$, so that $(\mathcal{F}\varphi)(y/2) = 1$; in this case each side of (‡) equals $(\mathcal{F}\varphi)(y)$, and the two sides are equal. Thus (‡) holds in both cases, and in particular, (††) is proved.

Since μ is periodic of period 1, we can write $\mu_0(y) = \frac{1}{\sqrt{2}}\sum_k c_k e^{-2\pi i k y}$. If we substitute this expression into (††) and apply $\mathcal{F}^{-1}$ to (††), we see that φ satisfies the scaling equation $\varphi(x) = \sum_k c_k \sqrt{2}\varphi(2x-k)$. That is, we can conclude from (†) and (††) together that $V_0 \subseteq V_1$ and that the generating function for φ is given by[12]

$$m_0(y) = \sum_{l=-\infty}^{\infty} (\mathcal{F}\varphi)(2y+2l).$$

Thus we are in a position to apply Propositions 10.15 and 10.17 to see that we have a multiresolution analysis. Proposition 10.15 says that condition (ii) is always satisfied when the V_j's are defined as above from the candidate for a scaling function. Proposition 10.17 says that condition (i) is satisfied if $\mathcal{F}\varphi$ is bounded, is continuous at 0, and has $(\mathcal{F}\varphi)(0) \neq 0$; all of these hypotheses can be seen by inspection of the definition of $(\mathcal{F}\varphi)(y)$ before the statement of the theorem, and thus indeed we have a multiresolution analysis.

Theorem 10.10 applies, and we are left with computing ψ. The theorem tells us what functions we can take as ψ. One of them is given by

$$(\mathcal{F}\psi)(y) = e^{\pi i y}\overline{m_0(\tfrac{1}{2}y + \tfrac{1}{2})}(\mathcal{F}\varphi)(\tfrac{1}{2}y).$$

Since $\mathcal{F}\varphi$ is real-valued, we can drop the complex conjugation and obtain

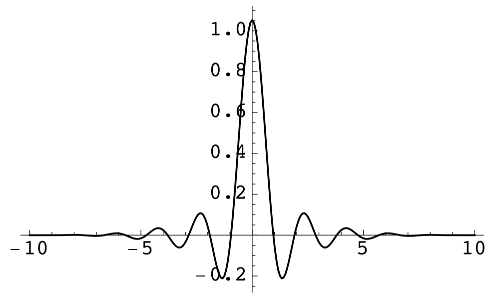
$$(\mathcal{F}\psi)(y) = e^{\pi i y}\sum_{l=-\infty}^{\infty} (\mathcal{F}\varphi)(y+2l+1)(\mathcal{F}\varphi)(\tfrac{1}{2}y).$$

[12]Even with an interpretation of L^2 convergence in the formula defining μ_0, we make no assertion that this kind of formula yields m_0 beyond the context of Meyer wavelets.

For the l^{th} term to be nonzero, we must have $|y/2| < \frac{2}{3}$ and $|y + 2l + 1| < \frac{2}{3}$, and thus $l = 0$ or $l = -1$. Thus

$$(\mathcal{F}\psi)(y) = e^{\pi i y}\big((\mathcal{F}\varphi)(y+1) + (\mathcal{F}\varphi)(y-1)\big)(\mathcal{F}\varphi)(\tfrac{1}{2}y). \qquad \square$$

(a)



(b)

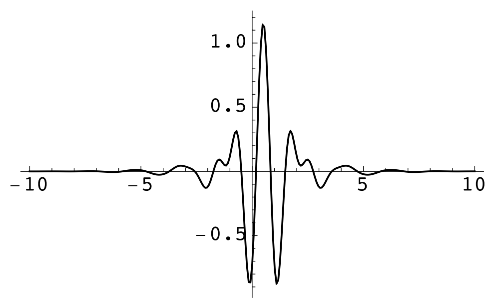


FIGURE 10.8. Graphs of Meyer scaling function and wavelet when $m = 3$.
(a) Scaling function. (b) Wavelet.

7. Splines

A "spline" for our purposes is a piecewise-polynomial function on an interval $[a, b]$ with certain smoothness conditions at the edges of the pieces. Specifically let

$$a = x_0 < x_1 < \cdots < x_m = b$$

be a partition of $[a, b]$. If $n < m$, a **spline of degree** n relative to this partition[13] is a function $\sigma(x)$ such that

(i) for each i, the restriction of $\sigma(x)$ to the interval $[x_{i-1}, x_i]$ is a real-valued polynomial p_i of degree at most n,

(ii) $\sigma(x)$ is of class C^{n-1} on $[a, b]$.

[13]The situation in which some points of the partition are repeated will not be of interest to us and will not be addressed.

The points x_i of the partition are called the **knots** of the spline.[14]

We shall be interested only in splines with equally spaced knots, for example at the integer points in the interval $[a, b]$, and we shall be especially interested in splines that can be extended to members of $C^{n-1}(\mathbb{R})$ that are zero on $(-\infty, a]$ and $[b, +\infty)$. These we call the **splines of compact support** (with knots in $\mathbb{Z}$).

The case of degree 0 is something we have seen before. In the case of degree 0, the splines of compact support with knots in $\mathbb{Z}$ are the functions on $\mathbb{R}$ that are constant on every interval between consecutive integers and that vanish outside a compact subset of $\mathbb{R}$. These are the finite linear combinations of integer translates of the Haar scaling function,[15] which we shall call ${}^0\gamma$ in this section and the next. The closure of this vector subspace of $L^2(\mathbb{R})$ is the subspace V_0 from which the Haar multiresolution analysis is built.

We wish to generalize the Haar construction to splines of higher degree. For $n > 0$, the spline function $\sigma(x)$ has to be continuous at the knots, and so do its derivatives up through order $n-1$. Being a polynomial on each interior interval, $\sigma(x)$ has right and left limits at each knot, and the same thing is true of each derivative. The difference between these limits can be nonzero only for the n^{th} derivative, all derivatives higher than n giving 0, and

$$\lim_{x \downarrow x_i} p_{i+1}^{(n)}(x) - \lim_{x \uparrow x_i} p_i^{(n)}(x) = c_i$$

is the jump of the n^{th} derivative. Then it follows that

$$p_{i+1}(x) - p_i(x) - \frac{c_i}{n!}(x - x_i)^n$$

has all derivatives 0 in a neighborhood of x_i and must be the zero polynomial:

$$\boxed{p_{i+1}(x) = p_i(x) + \frac{c_i}{n!}(x - x_i)^n.}$$

This formula gives us a handle on the dimension of the vector space of splines of degree n on $[a, b]$. Again the knots are understood to be at the integer points.

For now, let us restrict attention to the case of degree 1. We introduce a spline ${}^1\gamma$ of degree 1 by the definition

$${}^1\gamma(x) = \begin{cases} x & \text{for } 0 \le x \le 1, \\ 2 - x & \text{for } 1 \le x \le 2, \\ 0 & \text{otherwise.} \end{cases}$$

[14]Under our assumption that their are no repetitions among the points of the spline, the knots are all "simple."

[15]The values of the function at integer points are not relevant as long as we are studying $L^2(\mathbb{R})$.

See Figure 10.9. The integer translates of ${}^1\gamma$, given by ${}^1\gamma_{0,k}(x) = {}^1\gamma(x-k)$, will also be of interest. Each of them is a spline of compact support. Comparison of their supports shows that they are linearly independent.

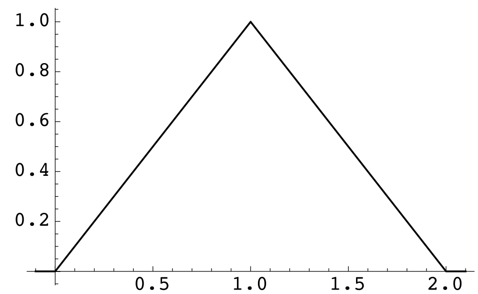


FIGURE 10.9. Graph of the B-spline ${}^1\gamma$.

The functions ${}^1\gamma_{0,k}$ are called ***B*-splines** of degree 1, the "B" being short for basis. The following result justifies this terminology.

Proposition 10.19. The B-splines of degree 1 form a vector-space basis of the vector space of all splines of degree 1 of compact support with knots in $\mathbb{Z}$.

PROOF. We need to prove spanning. Suppose that the spline is supported in $[M, N]$ with $M < N$. We shall subtract a multiple of a B-spline to reduce the support. Thus suppose the support of a spline $s(x)$ is contained in $[M, N]$. Let us say that the derivative $s'(x)$ has a jump by c at $x = M$, c possibly being 0. In this case the boxed jump formula shows that $s(x) = c(x - M)$ for $M \le x \le M + 1$. Consequently $s(x) - c\,{}^1\gamma_{0,M}(x)$ is a spline of degree 1 supported in $[M + 1, N]$. The support has been reduced, and in finitely many steps we end up with the difference of $s(x)$ and a linear combination of B-splines exhibited as supported in the interval $[N - 1, N]$. A spline of degree 1 that vanishes at each end has to be the 0 spline, and thus the difference of $s(x)$ and a linear combination of B-splines is 0. In other words, $s(x)$ equals a linear combination of B-splines. □

Let V_0 be the closure in $L^2(\mathbb{R})$ of the vector spaces of all splines of degree 1 of compact support with knots in $\mathbb{Z}$. It begins to look as if V_0 and ${}^1\gamma$ might yield a multiresolution analysis generalizing the Haar multiresolution analysis—until we stop to realize that the B-splines ${}^1\gamma_{0,k}$ are not orthogonal to one another. In fact, two of them are orthogonal if and only if their supports have no nontrivial interval in common. Nevertheless we plunge ahead by writing down a scaling equation and by computing the Fourier transform of ${}^1\gamma$.

The scaling equation for ${}^1\gamma$ is

$$ {}^1\gamma(x) = \tfrac{1}{2}\,{}^1\gamma(2x) + {}^1\gamma(2x-1) + \tfrac{1}{2}\,{}^1\gamma(2x-2); $$

we can check this equality directly when x is an integer multiple of $\frac{1}{2}$, and then the equality follows everywhere.

The computation of the Fourier transform of ${}^1\gamma$ can be carried out directly, using integration by parts suitably, but we prefer the following alternative argument, which deduces it as Corollary 10.22 from the Fourier transform of the indicator function $I_{[0,1]}$ and an identity concerning convolutions.

Proposition 10.20. The Fourier transform of ${}^0\gamma = I_{[0,1]}$ is given by

$$(\mathcal{F}\,{}^0\gamma)(y) = \frac{e^{-2\pi iy} - 1}{-2\pi iy} = e^{-\pi iy}\Big(\frac{\sin \pi y}{\pi y}\Big).$$

PROOF. The quantity in question equals $\int_0^1 e^{-2\pi ixy}\,dx$. □

Proposition 10.21. The B-spline ${}^1\gamma$ is the convolution of the indicator function ${}^0\gamma = I_{[0,1]}$ with itself:

$${}^1\gamma(x) = \int_{\mathbb{R}} I_{[0,1]}(x-t) I_{[0,1]}(t)\,dt = \int_0^1 I_{[0,1]}(x-t)\,dt.$$

PROOF. The integral in question is 0 if $x \le 0$ or $x \ge 2$. For $0 \le x \le 2$, the change of variables $u = x - t$ shows that it equals

$$\textstyle\int_{x-1}^{x} I_{[0,1]}(u)\,du = \int_{\max\{0,x-1\}}^{\min\{1,x\}} 1\,du = \min\{1, x\} - \max\{0, x-1\};$$

this equals x if $0 \le x \le 1$ or else equals $1 - (x-1) = 2 - x$ if $1 \le x \le 2$. □

Corollary 10.22. The Fourier transform of the B-spline ${}^1\gamma$ is given by

$$(\mathcal{F}\,{}^1\gamma)(y) = e^{-2\pi iy}\Big(\frac{\sin \pi y}{\pi y}\Big)^2.$$

PROOF. The Fourier transform of a convolution is the product of the Fourier transforms, by Proposition 8.1c of *Basic*. Thus the result follows by combining Propositions 10.20 and 10.21. □

The Fourier transform of the scaling equation of ${}^1\gamma$ is

$$(\mathcal{F}^1\gamma)(y) = \tfrac{1}{2}(\tfrac{1}{2}(\mathcal{F}^1\gamma)(y/2) + 1(\tfrac{1}{2}e^{-\pi iy}(\mathcal{F}^1\gamma)(y/2)) + \tfrac{1}{2}(\tfrac{1}{2}e^{-2\pi iy}(\mathcal{F}^1\gamma)(y/2)).$$

We can check this equation explicitly by using Corollary 10.22 and thereby verify the original scaling equation itself. The question is whether

$$\begin{aligned} e^{-2\pi iy}\Big(\frac{\sin \pi y}{\pi y}\Big)^2 &\overset{?}{=} \tfrac{1}{2}(\mathcal{F}^1\gamma)(y/2)(\tfrac{1}{2} + e^{-\pi iy} + \tfrac{1}{2}e^{-2\pi iy}) \\ &= \tfrac{1}{2}(\mathcal{F}^1\gamma)(y/2)e^{-\pi iy}(1 + \cos \pi y) \end{aligned}$$

$$= (\mathcal{F}^1\gamma)(y/2)e^{-\pi i y}\cos^2 \pi y/2$$
$$= e^{-\pi i y}e^{-\pi i y}\Big(\frac{\sin^2 \pi y/2}{\pi^2 y^2/4}\Big)\cos^2 \pi y/2,$$

and this comes down to the identity $\sin 2\theta = 2\sin\theta\cos\theta$.

Motivated by Lemma 10.11, we now calculate

$$\Delta(y) = \sum_{l=-\infty}^{\infty} |(\mathcal{F}^1\gamma)(y+l)|^2$$

We know that the sum cannot be almost everywhere equal to 1, since $\{{}^1\gamma_{0,k}\}$ is not an orthonormal set. However, the value of the sum will contain useful information for us. To make the calculation, we need to use some elementary complex analysis. We shall take as our starting point the results of Appendix B of *Basic*.

Proposition 10.23. For complex z,

$$\frac{\pi^2}{\sin^2 \pi z} = \sum_{n=-\infty}^{\infty} \frac{1}{(z-n)^2},$$

the series being uniformly convergent on compact sets that contain no integers.

PROOF. The series is convergent for $z \notin \mathbb{Z}$ by comparison with the series $\sum_{n=1}^{\infty}(1/n^2)$. If the terms corresponding to integers in a fixed closed disk about 0 are excluded, then the same comparison shows that the convergence is uniform on the disk. By Problem 55 in Appendix B of *Basic*, the sum of the series is meromorphic in $\mathbb{C}$ with poles of order 2 at each integer. The function $\pi^2/\sin^2(\pi z)$ has the same property, and therefore the difference

$$g(z) = \tfrac{\pi^2}{\sin^2 \pi z} - \sum_{n=-\infty}^{\infty} \tfrac{1}{(z-n)^2} \qquad (*)$$

is analytic in all of $\mathbb{C}$. We show that $g(z)$ is the zero function.

Each term on the right side of $(*)$ is periodic of period 1, and thus the same thing is true of $g(z)$. By Liouville's Theorem it is enough to prove that $|g(x+iy)|$ tends to 0 uniformly for $|x| \le 1$ as $y \to \infty$. Consider the first term on the right side of $(*)$. Direct calculation gives

$$|\sin z|^2 = \tfrac{1}{4}(e^{i(x+iy)} - e^{-i(x+iy)})\overline{(e^{i(x+iy)} - e^{-i(x+iy)})}$$
$$= \tfrac{1}{4}(e^{-2y} - e^{-2ix} - e^{2ix} + e^{2y}) = \cosh^2 y - \cos^2 x,$$

and the reciprocal of this tends to 0 uniformly for all x as $y \to \infty$. This takes care of the first term of the right side of $(*)$. For the second term we have

$$\left|\tfrac{1}{(z-n)^2}\right| = \tfrac{1}{(x-n)^2+y^2}.$$

This has the required behavior for the terms with $|n| \le 2$. For the terms with $|n| \ge 3$, we have $\frac{1}{(x-n)^2+y^2} \le \frac{1}{(|n|-1)^2+y^2}$. Given $\epsilon > 0$ with ϵ equal to the reciprocal of an integer, suppose that $|y| \ge \epsilon^{-1}$. Then

$$\sum_{|n|=2}^{1/\epsilon} \tfrac{1}{(|n|-1)^2+y^2} \le \sum_{|n|=2}^{1/\epsilon} \tfrac{1}{y^2} \le 2(1/\epsilon)y^{-2} \le 2\epsilon$$

and

$$\sum_{|n|=1/\epsilon}^{\infty} \tfrac{1}{(|n|-1)^2+y^2} \le \sum_{|n|=1/\epsilon}^{\infty} \tfrac{1}{(|n|-1)^2} \le \sum_{|n|=1/\epsilon}^{\infty} \left(\tfrac{1}{|n|-2} - \tfrac{1}{|n|-1}\right) = \tfrac{2}{\epsilon^{-1}-2},$$

which is $\le 6\epsilon$ as long as $\epsilon < \frac{1}{3}$. This takes care of the second term on the right side of $(*)$. □

Corollary 10.24. For every integer $m \ge 0$,

$$\left(\frac{d}{dz}\right)^m \left(\frac{\pi^2}{\sin^2 \pi z}\right) = (m+1)! \sum_{n=-\infty}^{\infty} \frac{1}{(z-n)^{m+2}}$$

PROOF. This formula results from successive term-by-term differentiation of the result of Proposition 10.23. To justify the interchange of derivative and sum, we make repeated use the standard technique of Appendix B of *Basic*, writing the complex derivative as a suitable line integral and using Fubini's Theorem to interchange the integral and the infinite sum. □

Returning to the calculation of $\sum_{l=-\infty}^{\infty} |(\mathcal{F}^1\gamma)(y+l)|^2$, we combine Corollary 10.22 with the formula of Corollary 10.24 when $m = 2$. Then

$$\begin{aligned}
\Delta(y) = \sum_{l=-\infty}^{\infty} |(\mathcal{F}^1\gamma)(y+l)|^2 &= \sum_{l=-\infty}^{\infty} \left|\frac{\sin \pi(y+l)}{\pi(y+l)}\right|^4 \\
&= \frac{\sin^4(\pi y)}{\pi^4} \sum_{l=-\infty}^{\infty} \frac{1}{(y+l)^4} \\
&= \frac{\sin^4(\pi y)}{6\pi^4} \frac{d^2}{dy^2}\left(\frac{\pi^2}{\sin^2 \pi y}\right) \\
&= \tfrac{1}{3}(\sin^2 \pi y + 3\cos^2 \pi y) \\
&= \tfrac{1}{3}(1 + 2\cos^2 \pi y).
\end{aligned}$$

In other words, the periodic function $\Delta(y) = \sum_{l=-\infty}^{\infty} |(\mathcal{F}\,{}^1\gamma)(y+l)|^2$ of period 1 is bounded above and below by the positive constants 1 and $\frac{1}{3}$. In terminology used by many authors, this means that the integer translates of $\mathcal{F}\,{}^1\gamma$ form a "Riesz system." In practical terms it means that we can use a general method to construct an orthonormal system out of $\mathcal{F}\,{}^1\gamma$. We proceed as follows.

We define a member φ of $L^2(\mathbb{R})$ by

$$(\mathcal{F}\varphi)(y) = \Delta(y)^{-1/2}(\mathcal{F}\,{}^1\gamma)(y).$$

By Lemma 10.11, the set of functions $\{x \mapsto \varphi(x-k)\}_{k=-\infty}^{\infty}$ is orthonormal. What is not apparent is that they form an orthonormal basis of the same space V_0 and that we thereby obtain a multiresolution analysis and an associated wavelet. The relevant tool is Lemma 10.12, which we have not used explicitly so far.

Theorem 10.25. Let V_j be the closure in $L^2(\mathbb{R})$ of the space of all splines of degree 1 of compact support with knots in $2^{-j}\mathbb{Z}$, let ${}^1\gamma$ be the B-spline of degree 1 defined above and having Fourier transform as in Corollary 10.22, and let φ be the member of $L^2(\mathbb{R})$ defined by

$$(\mathcal{F}\varphi)(y) = \Delta(y)^{-1/2}(\mathcal{F}\,{}^1\gamma)(y).$$

Then

(a) the integer translates of φ are in V_0 and form an orthonormal basis of it,
(b) $\{V_j\}_{j=-\infty}^{\infty}$ and φ constitute a multiresolution analysis,
(c) the corresponding wavelet ψ may be taken to have Fourier transform $(\mathcal{F}\psi)(y)$ equal to

$$(\sin^2 \tfrac{1}{2}\pi y)\Big(\frac{1+2\sin^2 \frac{1}{2}\pi y}{1+2\cos^2 \pi y}\Big)^{1/2}\Delta(\tfrac{1}{2}y)^{-1/2}(\mathcal{F}\,{}^1\gamma)(\tfrac{1}{2}y),$$

(d) the Fourier series $\Delta(y)^{-1/2} = \sum_k d_k e^{-2\pi iky}$ has the property that $\varphi(x)$ is given by $\varphi(x) = \sum_k d_k\,{}^1\gamma(x-k)$, the series being locally a finite sum,
(e) the Fourier series $\sum_k u_k e^{-\pi iky}$ of period 2 of the function

$$U(y) = (\sin^2 \tfrac{1}{2}\pi y)\Big(\frac{1+2\sin^2 \frac{1}{2}\pi y}{1+2\cos^2 \pi y}\Big)^{1/2}\Delta(\tfrac{1}{2}y)^{-1/2}$$

has the property that $\psi(x)$ is given by $\psi(x) = 2\sum_k u_k\,{}^1\gamma(2x-k)$, the series being locally a finite sum,
(f) $\varphi(x)$ and $\psi(x)$ are splines of degree 1 with infinite support, the knots of φ being in $\mathbb{Z}$ and the knots of ψ being in $\frac{1}{2}\mathbb{Z}$,
(g) $\varphi(x)$ and $\psi(x)$ decay exponentially fast as $|x|$ tends to infinity.

REMARKS. The wavelet obtained from Theorem 10.25 is called the **Battle–Lemarié wavelet** of degree[16] 1. Another name for it is the **Franklin wavelet**. Following Daubechies, we shall refer to the device of using $\Delta(y)$ the way we are using it here as the **orthonormalization trick**. The formulas for φ in (d) and for ψ in (e) allow one to understand φ and ψ without applying $\mathcal{F}^{-1}$ to $\mathcal{F}\varphi$ and $\mathcal{F}\psi$. Graphs of φ and ψ appear in Figure 10.10. To make the graphs, one computes enough Fourier coefficients d_k and u_k numerically to be able to use partial sums of the series in (d) and (e) as good approximations to φ and ψ.

PROOF. For (a), put $\varphi_{0,k}(x) = \varphi(x-k)$. We saw above that the set of functions $\{\varphi_{0,k}\}_{k=-\infty}^{\infty}$ is orthonormal. Let V be the closure of its linear span. We shall use Lemma 10.12 with this V and this φ. Let α and β be as in the statement of that lemma. Let $c = \{c_k\}_{k=-\infty}^{\infty}$ be the sequence of coefficients of the expansion of the function $\Delta(y)^{1/2} = \big(\frac{1}{3}(1+2\cos^2 \pi y)\big)^{1/2}$ in series as $\sum_{k=-\infty}^{\infty} c_k e^{-2\pi i k y}$. If f is the member of V given by $f = \alpha(c) = \sum_k c_k \varphi_{0,k}$, then $\beta^{-1}\alpha^{-1}(f) = \Delta^{1/2}$, and Lemma 10.12a shows that $\mathcal{F}f = \Delta^{1/2}\cdot(\mathcal{F}\varphi)$, with the dot indicating pointwise product. The right side equals $\mathcal{F}\,{}^1\gamma$, and therefore $f = {}^1\gamma$, i.e.,

$$(\mathcal{F}\,{}^1\gamma)(y) = \Delta(y)^{1/2}(\mathcal{F}\varphi)(y). \qquad (*)$$

Consequently ${}^1\gamma$ is in V, and the closure V_0 of the space of splines of degree 1 of compact support with knots in $\mathbb{Z}$ is contained in V.

We shall show that equality holds: $V = V_0$. Since $\Delta^{1/2}$ is smooth and is bounded above and below by positive constants, $\Delta^{-1/2}$ is a smooth periodic function on [0, 1]. Expand $\Delta(y)^{-1/2}$ in series as $\sum_{k=-\infty}^{\infty} d_k e^{-2\pi i k y}$, and let p_n be its n^{th} partial sum. By Proposition 1.56 of *Basic*, $\{p_n\}$ converges uniformly on [0, 1]. The limit is $\Delta^{-1/2}$ by Fejér's Theorem (Theorem 1.59 of *Basic*).

We know that the function $\mathcal{F}^{-1}\big(e^{-2\pi i k y}(\mathcal{F}\,{}^1\gamma)(y)\big) = \big(x \mapsto {}^1\gamma(x-k)\big)$ is in V_0 for each k, and hence so is the linear combination $\mathcal{F}^{-1}\big(p_n(y)(\mathcal{F}\,{}^1\gamma)(y)\big) = \mathcal{F}^{-1}\big(p_n(y)\Delta(y)^{1/2}(\mathcal{F}\varphi)(y)\big)$ for each n. The product $p_n\Delta^{1/2}$ converges uniformly to $\Delta^{-1/2}\Delta^{1/2} = 1$, and uniform convergence implies convergence in $L^2[0,1]$. Lemma 10.12 then allows us to conclude that $p_n(y)\Delta(y)^{1/2}(\mathcal{F}\varphi)(y)$ converges to $(\mathcal{F}\varphi)(y)$ in $L^2(\mathbb{R})$. Applying $\mathcal{F}^{-1}$, we see that there is a sequence in V_0 converging to φ. Since V_0 is closed, φ is in V_0, and it follows that $V = V_0$.

[16]Terminology varies. Some authors use the word "order" in place of "degree." Some authors shift the indices by 1, saying that the case here is of order 2; for such authors the Haar wavelet becomes the Battle-Lemarié wavelet of order 1.

For (b), we have just seen that the closed subspace of $L^2(\mathbb{R})$ generated by the integer translates of φ is V_0, and we know for this explicit V_0 that the dilates V_j form an increasing sequence. Propositions 10.15 and 10.17 show that φ is a scaling function, since $\mathcal{F}\varphi$ is continuous, is bounded, and has $(\mathcal{F}\varphi)(0) \neq 0$. Thus $\{V_j\}_{j=-\infty}^{\infty}$ and φ constitute a multiresolution analysis.

For (c), we are to compute ψ. The Fourier transform of the scaling equation for φ gives us $(\mathcal{F}\varphi)(y) = m_0(y/2)(\mathcal{F}\varphi)(y/2)$, from which we have

$$\begin{aligned} m_0(y/2) &= \frac{(\mathcal{F}\varphi)(y)}{(\mathcal{F}\varphi)(y/2)} \\ &= \frac{\big(\frac{1}{3}(1+2\cos^2 \pi y)\big)^{-1/2}(\mathcal{F}\,{}^1\gamma)(y)}{\big(\frac{1}{3}(1+2\cos^2 \pi y/2)\big)^{-1/2}(\mathcal{F}\,{}^1\gamma)(y/2)} \\ &= \frac{\big(\frac{1}{3}(1+2\cos^2 \pi y/2)\big)^{1/2} e^{-2\pi i y}(\pi y)^{-2}\sin^2 \pi y)}{\big(\frac{1}{3}(1+2\cos^2 \pi y)\big)^{1/2} e^{-\pi i y}(\pi y/2)^{-2}\sin^2 \pi y/2)} \\ &= e^{-\pi i y}\Big(\frac{1+2\cos^2 \pi y/2}{1+2\cos^2 \pi y}\Big)^{1/2}\cos^2 \pi y/2. \end{aligned}$$

From Theorem 10.10 we may take ψ to have

$$\begin{aligned} (\mathcal{F}\psi)(y) &= e^{\pi i y}\nu(y)\overline{m_0(\tfrac{1}{2}y+\tfrac{1}{2})}(\mathcal{F}\varphi)(\tfrac{1}{2}y) \\ &= -e^{2\pi i y}\nu(y)\sin^2\tfrac{1}{2}\pi y\Big(\frac{1+2\sin^2\frac{1}{2}\pi y}{1+2\cos^2 \pi y}\Big)^{1/2}(\mathcal{F}\varphi)(\tfrac{1}{2}y) \\ &= -e^{2\pi i y}\nu(y)\sin^2\tfrac{1}{2}\pi y\Big(\frac{1+2\sin^2\frac{1}{2}\pi y}{1+2\cos^2 \pi y}\Big)^{1/2}\Delta(\tfrac{1}{2}y)^{-1/2}(\mathcal{F}\,{}^1\gamma)(\tfrac{1}{2}y) \end{aligned}$$

with ν periodic of period 1. If we take $\nu(y) = -e^{-2\pi i y}$, then the asserted formula follows.

For (d), we observe that $\mathcal{F}^{-1}(e^{-2\pi i k y}(\mathcal{F}\,{}^1\gamma)(y)) = \gamma(y-k)$ and hence that $\mathcal{F}^{-1}\big(\sum_{k=-n}^{n} d_k e^{-2\pi i k y}(\mathcal{F}\,{}^1\gamma)(y)\big) = \sum_{k=-n}^{n} d_k\,{}^1\gamma(x-k)$. That is,

$$\mathcal{F}^{-1}(p_n\Delta^{1/2}\mathcal{F}\varphi) = \sum_{k=-n}^{n} d_k\,{}^1\gamma(x-k). \qquad (**)$$

We saw in the proof of (a) that $p_n\Delta^{1/2}\mathcal{F}\varphi$ converges in $L^2(\mathbb{R})$ to $\mathcal{F}\varphi$. By continuity of $\mathcal{F}^{-1}$ on $L^2(\mathbb{R})$, the left side of $(**)$ converges to φ in $L^2(\mathbb{R})$. Therefore

$$\varphi(x) = \lim_n \sum_{k=-n}^{n} d_k\,{}^1\gamma(x-k), \qquad (\dagger)$$

the limit taken in $L^2(\mathbb{R})$. This proves the equality in (d). Since ${}^1\gamma$ has compact support, the series in (†) is locally the finite sum of continuous functions. Therefore the equality (†) is meaningful pointwise, as well as in $L^2(\mathbb{R})$, if we redefine φ on a set of measure 0 so that it coincides with the pointwise sum.

For (e), let ${}^1\gamma_{1,k}(x) = \sqrt{2}\,{}^1\gamma(2x-k)$. Changes of variables show that

$$(\mathcal{F}\,{}^1\gamma_{1,k})(y) = \tfrac{1}{\sqrt{2}}e^{-\pi iky}(\mathcal{F}\,{}^1\gamma)(\tfrac{1}{2}y). \tag{††}$$

Without any justification of interchanges of limits, the argument for (e) is that

$$\begin{aligned}(\mathcal{F}\psi)(y) &= U(y)(\mathcal{F}\,{}^1\gamma)(\tfrac{1}{2}y)\\ &= \sum_{k=-\infty}^{\infty} u_k e^{-\pi iky}(\mathcal{F}\,{}^1\gamma)(\tfrac{1}{2}y)\\ &= \sqrt{2}\sum_{k=-\infty}^{\infty} u_k(\mathcal{F}\,{}^1\gamma_{1,k})(y)\\ &= \mathcal{F}(\sqrt{2}\sum_{k=-\infty}^{\infty} u_k\,{}^1\gamma_{1,k})(y).\end{aligned} \tag{‡}$$

Application of $\mathcal{F}^{-1}$ then gives

$$\psi(x) = \sqrt{2}\sum_{k=-\infty}^{\infty} u_k\,{}^1\gamma_{1,k}(x) = 2\sum_{k=-\infty}^{\infty} u_k\,{}^1\gamma(2x-k). \tag{‡‡}$$

We shall interpret (‡‡) as an equality in $L^2(\mathbb{R})$, with convergence in the L^2 sense. With this interpretation the right side is locally a finite sum of continuous functions and hence is continuous. Then the L^2 function ψ can be adjusted on a set of measure 0 so as to agree with this continuous function, and the result is that (‡‡) is also a pointwise equality, the sum being uniformly convergent on each compact subset of $\mathbb{R}$.

Thus we are to prove (‡‡) in the L^2 sense. Since $\mathcal{F}$ is a unitary operator from $L^2(\mathbb{R})$ onto itself, (‡‡) is immediate from (‡), and we are to justify the steps of (‡). The first equality of (‡) is merely a restatement of conclusion (c) of the theorem, and the third equality follows from (††). For the second equality of (‡), we have substituted the series expansion of $U(y)$, but we have left ambiguous how to interpret this expression in terms of convergence. Is the convergence to be that of periodic functions in $L^2([0,2])$, or is it to be convergence in $L^2(\mathbb{R})$ of the whole expression? If we were working with φ instead of ${}^1\gamma$, the two would come to the same thing, by Lemma 10.12b, but in our situation we have no comparable result. Finally the fourth equality of (‡) requires the justification of a clear-cut interchange of limits.

We can handle both difficulties at once if we show that

$$\lim_{n\to\infty}\Big[(\mathcal{F}\psi)(y)-\sum_{|k|\le n}u_k e^{-\pi iky}(\mathcal{F}\,{}^1\gamma)(\tfrac{1}{2}y)\Big]=0,$$

the convergence being in $L^2(\mathbb{R})$. The computation is

$$\begin{aligned}\Big\|\sum_{|k|>n}u_k e^{-i\pi ky}(\mathcal{F}\,{}^1\gamma)(\tfrac{1}{2}y)\Big\|_{L^2(\mathbb{R})} &= \Big\|\sum_{|k|>n}u_k e^{-i\pi ky}\Delta(\tfrac{1}{2}y)^{1/2}(\mathcal{F}\varphi)(\tfrac{1}{2}y)\Big\|_{L^2(\mathbb{R})}\\ &= \Big\|\big(\sum_{|k|>n}u_k e^{-i\pi ky}\big)\Delta(\tfrac{1}{2}y)^{1/2}\Big\|_{L^2([0,2])}\end{aligned}$$

by Lemma 10.12b applied to the space V_1 and the function $\varphi_{1,0}$, with the understanding that the measure on $L^2([0,2])$ is normalized so as to have total mass 1. The right side is

$$\le\big(\sup_{0\le y\le 2}\Delta(\tfrac{1}{2}y)^{1/2}\big)\Big\|\sum_{|k|>n}u_k e^{-i\pi ky}\Big\|_{L^2([0,2])},$$

and this tends to 0 as n tends to infinity, since $\Delta(\frac{1}{2}y)^{1/2}$ is bounded and $\sum_k |u_k|^2$ is finite. This completes the proof of (e).

For (f), the fact that each series is locally finite implies that we can differentiate term by term away from the knots. The second derivatives are zero, while the functions themselves are continuous. Thus $\varphi(x)$ and $\psi(x)$ are splines of degree 1. Their support is infinite since ${}^1\gamma$ has compact support and since arbitrarily large translates of it are involved.

For (g), the idea is that the coefficients d_k and u_k decrease geometrically fast, while ${}^1\gamma$ has compact support. The reason for the geometric decrease of the two sequences of Fourier coefficients is that the functions in question are the restrictions to the unit circle of analytic functions in a neighborhood within $\mathbb{C}$ of the unit circle. By Theorem B.47 of Appendix B of *Basic*, these analytic functions each have Laurent series expansions valid within an open annulus that contains the unit circle, and the usual estimates on Taylor coefficients imply that the Fourier coefficients decrease geometrically fast. □

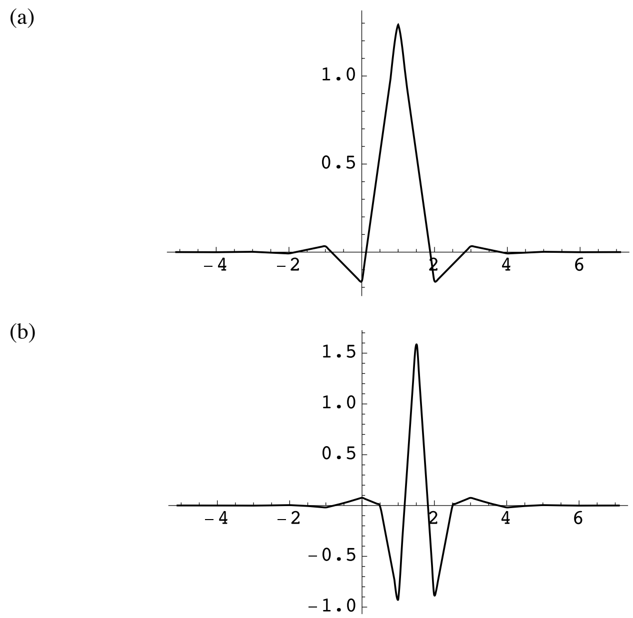


FIGURE 10.10. Graphs of Battle–Lemarié scaling function and wavelet of degree 1. (a) Scaling function. (b) Wavelet.

8. Battle–Lemarié Wavelets

In Section 7 we saw how splines of degree 1 of compact support with knots in $\mathbb{Z}$ lead to a multiresolution analysis and a wavelet. The wavelet is known as the Battle–Lemarié wavelet of degree 1. In more detail the B-splines of degree 1 with knots in $\mathbb{Z}$ are exactly the integer translates of one of them, which we called ${}^1\gamma$, and they form a vector space basis of the space of all splines of degree 1 of compact support with knots in $\mathbb{Z}$. The closure of this vector space is the space V_0 of the analysis. Unfortunately the B-splines are not orthogonal to one another, and an additional step was necessary to fit everything into our standard set-up.

In this section we shall consider the case of higher degree. The principles will all be the same, but the technical details are more complicated. Before we begin, let us pause to realize that this theory also works in degree 0. In this case the splines in question are functions that are constant between consecutive integers and have compact support, and the B-splines are the integer translates of the Haar scaling function. When we follow through the construction, we arrive at the Haar wavelet. The complication from nonorthogonal B-splines does not arise.

Now we pass to degree m with $m \geq 1$. The first step is to exhibit a nonzero spline of degree m with compact support and with knots in $\mathbb{Z}$. Specifically we seek an analog of the B-spline ${}^1\gamma$. When $m = 2$, an analog is

$$
{}^2\gamma(x) = \begin{cases} \frac{1}{2}x^2 & \text{for } 0 \leq x \leq 1, \\ \frac{3}{4} - (x - \frac{3}{2})^2 & \text{for } 1 \leq x \leq 2, \\ \frac{1}{2}(x-3)^2 & \text{for } 2 \leq x \leq 3, \\ 0 & \text{otherwise.} \end{cases}
$$

One readily checks that ${}^2\gamma$ and its first derivative are continuous at $x = 0, 1, 2, 3$ and that ${}^2\gamma$ is supported in $[0, 2]$. Thus $\gamma^{(2)}$ is a nonzero spline of degree 2 of compact support with knots in $\mathbb{Z}$.

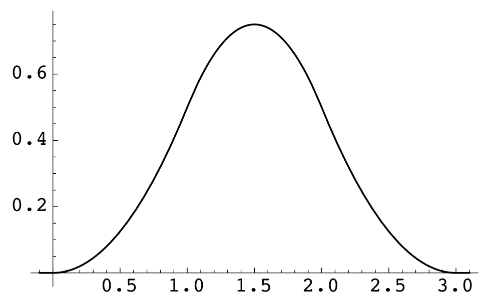


FIGURE 10.11. Graph of the B-spline ${}^2\gamma$.

It is a little hard to see how to generalize the above formula, and a different notation will make matters a little more transparent. Let us introduce the name $(\,\cdot\,)_+$ for the function with $t_+ = \max\{0, t\}$. This is a spline of degree 1, and its only knot is in $\mathbb{Z}$; but it does not have compact support. The function $(\,\cdot\,)_+^m$ with $t_+^m = (\max\{0, t\})^m$ is a spline of degree m. It similarly has its only knot in $\mathbb{Z}$, and it too fails to have compact support. The integer translates of this function are of the form $(\,\cdot\, - k)_+^m$ with $(t - k)_+^m = (\max\{0, t - k\})^m$. We define

$$
{}^m\gamma(t) = \frac{1}{m!} \sum_{k=0}^{m+1} (-1)^k \binom{m+1}{k} (t - k)_+^m.
$$

This formula agrees with the concrete formula for ${}^2\gamma(t)$ given above. Also it agrees with the formula ${}^0\gamma = I_{[0,1]}$ except at integer points.

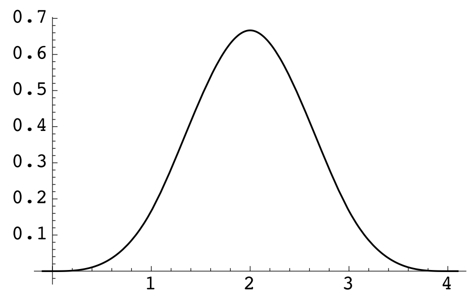


FIGURE 10.12. Graph of the B-spline ${}^3\gamma$.

The function ${}^m\gamma$ is a spline of degree m with knots in $\mathbb{Z}$. It is nonzero because if $0 < t < 1$, then ${}^m\gamma(t) = (m!)^{-1}t^m \neq 0$. It is certainly 0 when $t \leq 0$. We shall prove that it is 0 when $t \geq m+1$. To do so, we make use of the first half of the following combinatorial lemma.

Lemma 10.26.

(a) For any function $g : \mathbb{R} \to \mathbb{R}$, write $(\delta g)(x) = g(x) - g(x-1)$. Then for any $m \geq 1$, $(\delta^m g)(x) = \sum_{k=0}^{m} (-1)^k \binom{m}{k} g(x-k)$.

(b) For any function $h : \mathbb{R} \to \mathbb{R}$, write $(\varepsilon h)(x) = h(x+1) - h(x)$. Then for any $m \geq 1$, $(\varepsilon^m h)(x) = \sum_{k=0}^{m} (-1)^{m-k} \binom{m}{k} h(x+k)$.

PROOF. In (a) the argument is for all g by induction on m. The case $m = 1$ is the definition of δg. Assume inductively that equality holds for $m-1 \geq 1$. Then

$$\begin{aligned}
\delta^m g(x) &= (\delta^{m-1}(\delta g))(x) \\
&= \sum_{k=0}^{m-1} (-1)^k \binom{m-1}{k} (\delta g)(x-k) \\
&= \sum_{k=0}^{m-1} (-1)^k \binom{m-1}{k} g(x-k) - \sum_{k=0}^{m-1} (-1)^k \binom{m-1}{k} g(x-k-1) \\
&= \sum_{k=0}^{m-1} (-1)^k \binom{m-1}{k} g(x-k) + \sum_{l=1}^{m} (-1)^l \binom{m-1}{l-1} g(x-l) \\
&= \sum_{k=0}^{m} (-1)^k \binom{m}{k} g(x-k).
\end{aligned}$$

and the induction is complete. Part (b) follows by applying (a) to the function $g(x) = h(-x)$. $\square$

Returning to ${}^m\gamma$, we use Lemma 10.26 to show that ${}^m\gamma(t) = 0$ for $t \geq m+1$. If $t \geq m+1$, then

$$ {}^m\gamma(t) = \sum_{k=0}^{m+1} (-1)^k \binom{m+1}{k} (t-k)^m $$

Because of Lemma 10.26, we recognize the polynomial on the right side as $\delta^{m+1}(t^m)$. The operation of δ reduces the degree of a polynomial by 1, and hence $\delta^{m+1}(t^m) = 0$. Thus we see that ${}^m\gamma(t) = 0$ for $t \geq m+1$.

This proves that ${}^m\gamma$ is supported in the interval $[0, m+1]$. Translating by an integer, we conclude that the vector space of splines of degree m with knots in $\mathbb{Z}$ and with support in $[N, N+m]$ is not 0 for each integer N.

Let us put that result aside for a moment and verify an integral formula for ${}^m\gamma(x)$. In turn the integral formula will allow us to see that the integer translates of ${}^m\gamma(x)$ form a vector-space basis of the space of all splines of degree m of compact support with knots in $\mathbb{Z}$. Let us write

$$ J_0 = I_{[0,1]}, $$

where $I_{[0,1]}$ is the indicator function of the interval $[0, 1]$, and let us further define inductively

$$ J_m = J_{m-1} * I_{[0,1]} \qquad \text{for } m \geq 1. $$

We have defined J_0 to equal ${}^0\gamma = I_{[0,1]}$, and we saw in Proposition 10.21 that J_1 equals ${}^1\gamma$. We are going to extend this result to higher degrees.

Lemma 10.27.

(a) $J_l'(x) = \delta J_{l-1}(x) = J_{l-1}(x) - J_{l-1}(x-1)$ for $l \geq 1$ as long as $x \notin \mathbb{Z}$ if $l = 1$,

(b) $\int_{\mathbb{R}} f'(x) J_l(x)\,dx = \int_{\mathbb{R}} (\varepsilon f)(x) J_{l-1}(x)\,dx$ for $l \geq 1$ if f is of class C^1.

PROOF. For (a), we have $J_l(x) = (J_{l-1} * I_{[0,1]})(x) = \int_0^1 J_{l-1}(x-t)\,dt$. Differentiation under the integral sign gives

$$ J_l'(x) = \int_0^1 J_{l-1}'(x-t)\,dt = -[J_{l-1}(x-t)]_{t=0}^{t=1} = J_{l-1}(x) - J_{l-1}(x-1). $$

For (b), we first suppose that $l \geq 1$. Then the function $f J_l$ is of class C^1 and has compact support. Therefore $0 = \int_{\mathbb{R}} \frac{d}{dx}(f J_l)\,dx = \int_{\mathbb{R}} f J_l'\,dx + \int_{\mathbb{R}} f' J_l\,dx$. Substituting from (a), we obtain

$$ \begin{aligned} \int_{\mathbb{R}} f'(x) J_l(x)\,dx &= -\int_{\mathbb{R}} f(x) J_l'(x)\,dx \\ &= -\int_{\mathbb{R}} f(x) J_{l-1}(x)\,dx + \int_{\mathbb{R}} f(x) J_{l-1}(x-1)\,dx \\ &= -\int_{\mathbb{R}} f(x) J_{l-1}(x)\,dx + \int_{\mathbb{R}} f(x+1) J_{l-1}(x)\,dx \\ &= \int_{\mathbb{R}} (\varepsilon f)(x) J_{l-1}(x)\,dx. \end{aligned} $$

Now suppose that $l = 1$. In this case the function $J_1(x)$ fails to be C^1 at the points $x = 0, 1, 2$. We compute $\int_{\mathbb{R}} \frac{d}{dx}(f J_1)\,dx$ as $\int_{-1}^{0} + \int_0^1 + \int_1^2 + \int_2^3$ and see that it equals $[f J_1]_{x=-1}^{0} + [f J_1]_{x=0}^{1} + [f J_1]_{x=1}^{2} + [f J_1]_{x=2}^{3}$. Because of the continuity of J_1 at $x = 0, 1, 2$, we still get cancellation and we still have $\int_{\mathbb{R}} \frac{d}{dx}(f J_1)\,dx = 0$. Thus the above argument extends to be valid for $l = 1$. □

Proposition 10.28. For $m \geq 1$,

(a) ${}^m\gamma(x) = J_m(x)$ for $m \geq 0$ as long as $x \notin \mathbb{Z}$ if $m = 0$, and
(b) $\int_{\mathbb{R}} {}^m\gamma(x)\,dx = 1$ for $m \geq 0$.

PROOF. For (a), we shall prove equality for each point x_0. To handle x_0, we work with the function $g(t) = \dfrac{(-1)^m}{(m-1)!}(x_0 - t)_+^{m-1}$. For $k \leq m-2$, its k^{th} derivative is the continuous function

$$g^{(k)}(t) = \frac{(-1)^{m-k}}{(m-k-1)!}(x_0 - t)_+^{m-1-k}.$$

For $k = m-1$, the k^{th} derivative for $t \neq x_0$ is

$$g^{(m-1)}(t) = \begin{cases} 1 & \text{for } t > x_0 \\ 0 & \text{for } t < x_0, \end{cases}$$

and it does not exist at $t = x_0$.

We have

$$\begin{aligned} J_m(x_0) &= \int_{\mathbb{R}} J_{m-1}(x_0 - t) I_{[0,1]}(t) \\ &= \int_0^1 J_{m-1}(x_0 - t)\,dt \\ &= \int_{x_0-1}^{x_0} J_{m-1}(s)\,ds \qquad\qquad \text{under } s = x_0 - t \\ &= \int_{t > x_0 - 1} J_{m-1}(t)\,dt - \int_{t > x_0} J_{m-1}(t)\,dt \\ &= \int_{\mathbb{R}} \big(g^{(m-1)}(t+1) - g^{(m-1)}(t)\big) J_{m-1}\,dt \\ &= \int_{\mathbb{R}} (\varepsilon g^{(m-1)})(t) J_{m-1}(t)\,dt. \end{aligned}$$

From here we can apply Lemma 10.27b recursively. The above expression is

$$\begin{aligned} &= \int_{\mathbb{R}} (\varepsilon^2 g^{(m-2)})(t) J_{m-2}(t)\,dt \\ &= \cdots = \int_{\mathbb{R}} (\varepsilon^{m-1} g^{(1)})(t) J_1(t) \\ &= \int_{\mathbb{R}} (\varepsilon^m g)(t) J_0(t)\,dt \\ &= \int_0^1 (\varepsilon^m g)(t)\,dt \end{aligned}$$

$$= \sum_{k=0}^{m} \frac{(-1)^{m-k}}{(m-1)!}\binom{m}{k} \int_0^1 (x_0 - t - k)_+^{m-1}\,dt$$
$$= \sum_{k=0}^{m} \frac{(-1)^{m-k}}{(m-1)!}\binom{m}{k} \int_{x_0-k-1}^{x_0-k} s_+^{m-1}\,ds$$
$$= \sum_{k=0}^{m} \frac{(-1)^{m-k}}{m!}\binom{m}{k}\big((x_0-k)_+^m - (x_0-k-1)_+^m\big)$$
$$= \sum_{k=0}^{m} \frac{(-1)^{m-k}}{m!}\binom{m}{k}(x_0-k)_+^m + \sum_{l=1}^{m+1} \frac{(-1)^{m-l}}{m!}\binom{m}{l-1}(x_0-l)_+^m$$
$$= \sum_{k=0}^{m} \frac{(-1)^{m-k}}{m!}\binom{m+1}{k}(x_0-k)_+^m$$
$$= {}^m\gamma(x_0).$$

This completes the proof of (a).

For (b), we have $\int_{\mathbb{R}} {}^m\gamma(x)\,dx = \int_{\mathbb{R}} J_m(x)\,dx$. The integral over $\mathbb{R}$ of the convolution of two functions is the product of their integrals. Thus $\int_{\mathbb{R}} J_m(x)\,dx$ is the product of $m+1$ factors of 1. □

Corollary 10.29. The vector space of splines of degree m with compact support contained in $[0, m]$ and with knots in $\mathbb{Z}$ is 0.

PROOF. We proceed by induction on m. For $m = 1$, a spline of degree 1 with support in $[0, 1]$ and knots in $\mathbb{Z}$ is given for $0 \le x \le 1$ by a linear function $s(x)$, and the support condition implies that $s(0) = s(1) = 0$. Then we must have $s = 0$.

Assume the result for degree $m - 1 \ge 1$. If $s(x)$ is a spline of degree m with knots in $\mathbb{Z}$ and with support in $[0, m]$, then the derivative $s'(x)$ is a spline of degree $m - 1$ with knots in $\mathbb{Z}$ and with support in $[0, m]$. Arguing by contradiction, we may assume that $s(x)$ is not the 0 spline, so that $s'(x)$ is not the 0 spline. Since ${}^{m-1}\gamma$ is a spline of degree $m-1$ with knots in $\mathbb{Z}$ and with support in $[0, m]$, the boxed jump formula in Section 7 shows that $s'(x) - c\,{}^{m-1}\gamma$ for a suitable constant $c \neq 0$ is a spline of degree $m-1$ with knots in $\mathbb{Z}$ and with support in $[1, m]$. By inductive hypothesis the difference must be 0. Thus $s'(x)$ is the nonzero constant multiple c of ${}^{m-1}\gamma$. However, any spline of compact support that is a derivative has integral over $\mathbb{R}$ equal to 0, and Proposition 10.28b shows that ${}^{m-1}\gamma$ has nonzero integral. We conclude that $s' = 0$, and we have arrived at a contradiction. □

We write ${}^m\gamma_{0,k}(x) = {}^m\gamma(x-k)$ for the integer translates of the spline ${}^m\gamma$. Each of them is a spline of degree m of compact support with knots in $\mathbb{Z}$.

Comparison of their supports shows that they are linearly independent. The functions ${}^m\gamma_{0,k}$ are called **B-splines** of degree m, the "B" short for basis. The following result justifies this terminology.

Corollary 10.30. The B-splines of degree m, i.e., the integer translates of ${}^m\gamma$, form a vector-space basis of the space of splines of degree m with compact support and with knots in $\mathbb{Z}$.

PROOF. We need to prove spanning. Suppose that a given spline $s(x)$ of degree m with knots in $\mathbb{Z}$ is supported in $[M, N]$ with $M \leq N$. Corollary 10.29 shows that $M + m + 1 \leq N$ unless s is the 0 spline. We shall subtract a multiple of a B-spline to reduce the support. Since $s(x)$ vanishes for $x < M$ and since $s(x)$ is of class C^{m-1}, $s^{(k)}(M) = 0$ for $k \neq m$. Let us say that $s^{(k)}(x)$ has a jump by c at $x = M$, c possibly being 0. In this case the boxed jump formula shows that $s(x) = (m!)^{-1}c(x-M)^m$ for $M \leq x \leq M+1$. Consequently $s(x) - c\,{}^m\gamma_{0,M}(x)$ is a spline of degree m, and its support is contained in $[M+1, \max\{M+m+1, N\}]$, which is contained in $[M+1, N]$ since $M + m + 1 \leq N$. The support has been reduced, and in finitely many steps we end up with the difference of $s(x)$ and a linear combination of B-splines exhibited as supported in the interval $[N-m, N]$. By Corollary 10.29 this difference is 0. In other words, $s(x)$ equals a linear combination of B-splines. □

Corollary 10.31. The Fourier transform of the B-spline ${}^m\gamma$ is given by

$$(\mathcal{F}\,{}^m\gamma)(y) = e^{-(m+1)\pi i y}\Big(\frac{\sin \pi y}{\pi y}\Big)^{m+1}.$$

PROOF. The Fourier transform of a convolution is the product of the Fourier transforms, by Proposition 8.1c of *Basic*. Thus the result follows by combining Propositions 10.20 and 10.28a. □

In our effort to generalize Theorem 10.25 to higher order splines, we are up to the stage of testing the B-splines for the extent to which they are orthonormal. Two B-splines with overlapping support are certainly not orthogonal, but we calculate the sum

$$\Delta_{2m}(y) = \sum_{l=-\infty}^{\infty} |(\mathcal{F}\,{}^m\gamma)(y+l)|^2$$

anyway. As before, this is a periodic function of period 1. Once again we combine Corollary 10.22 with the appropriate case of Corollary 10.24. Then

$$\Delta_{2m}(y) = \sum_{l=-\infty}^{\infty} |(\mathcal{F}^m \gamma)(y+l)|^2 = \sum_{l=-\infty}^{\infty} \left| \frac{\sin \pi(y+l)}{\pi(y+l)} \right|^{2m+2}$$
$$= \frac{\sin^{2m+2}(\pi y)}{\pi^{2m+2}} \sum_{l=-\infty}^{\infty} \frac{1}{(y+l)^{2m+2}}$$
$$= \frac{\sin^{2m+2}(\pi y)}{(2m+1)!\,\pi^{2m+2}} \frac{d^{2m}}{dy^{2m}}\Big(\frac{\pi^2}{\sin^2 \pi y}\Big).$$

The values of $\Delta_{2m}(y)$ for $1 \le m \le 4$ appear as a table in Figure 10.13. We shall be ready to state a generalization of Theorem 10.25 once we prove the following lemma.

Lemma 10.32. For each $m \ge 1$, the function

$$\Delta_{2m}(y) = \frac{\sin^{2m+2}(\pi y)}{(2m+1)!\,\pi^{2m+2}} \frac{d^{2m}}{dy^{2m}}\Big(\frac{\pi^2}{\sin^2 \pi y}\Big)$$

is a polynomial in $\sin \pi y$ and $\cos \pi y$ with $P_{2m}(y) > 0$ for all real y.

PROOF. Let $q(y)$ be any polynomial in $\sin \pi y$ and $\cos \pi y$. The formula

$$\frac{d}{dy}\Big(\frac{q(y)}{\sin^k \pi y}\Big) = \frac{q'(y)}{\sin^k \pi y} - \frac{kq(y)\cos \pi y}{\sin^{k+1} \pi y} = \frac{q'(y)\sin \pi y - kq(y)\cos \pi y}{\sin^{k+1} \pi y}$$

shows that the derivative of a quotient of a polynomial in $\sin \pi y$ and $\cos \pi y$ by a power of $\sin \pi y$ is the quotient of a polynomial in $\sin \pi y$ and $\cos \pi y$ by one higher power of $\sin \pi y$. Consequently the expression $\Delta_{2m}(y)$ in the statement of the lemma is indeed a polynomial in $\sin \pi y$ and $\cos \pi y$.

The computation before the lemma shows that

$$(2m+1)!\,\pi^{2m}\Delta_{2m}(y) = (2m+1)!\,\pi^{2m} \sum_{l=-\infty}^{\infty} \left| \frac{\sin \pi(y+l)}{\pi(y+l)} \right|^{2m+2}.$$

The expression on the right side can vanish only if $\sin \pi(y+l) = 0$ for all l, hence only if y is an integer. By periodicity we have only to examine $y = 0$. There the quotient $\sin(\pi y)\big/(\pi y)$ is not 0. So $\Delta_{2m}(y)$ is positive for all y. □

m	$\Delta_{2m}(y)$
1	$\frac{1}{3}(2 + 2\cos 2\pi y)$
2	$\frac{1}{60}(33 + 26\cos 2\pi y + \cos 4\pi y)$
3	$\frac{1}{2520}(1208 + 1191\cos 2\pi y + 120\cos 4\pi y + \cos 6\pi y)$
4	$\frac{1}{2835}(62 + 1072\cos^2 \pi y + 1452\cos^4 \pi y + 247\cos^6 \pi y + 2\cos^8 \pi y)$

FIGURE 10.13. Values of $\Delta_{2m}(y)$ for small m.

Theorem 10.33. For $m \geq 1$, define

$$\Delta_{2m}(y) = \sum_{l=-\infty}^{\infty} |(\mathcal{F}\,{}^m\gamma)(y+l)|^2.$$

Let V_j be the closure in $L^2(\mathbb{R})$ of the space of all splines of degree m of compact support with knots in $2^{-j}\mathbb{Z}$, let ${}^m\gamma$ be the B-spline of degree m defined above and having Fourier transform as in Corollary 10.22, and let φ be the member of $L^2(\mathbb{R})$ defined by

$$(\mathcal{F}\varphi)(y) = \Delta_{2m}(y)^{-1/2}(\mathcal{F}\,{}^m\gamma)(y).$$

Then

(a) the integer translates of φ are in V_0 and form an orthonormal basis of it,
(b) $\{V_j\}_{j=-\infty}^{\infty}$ and φ constitute a multiresolution analysis,
(c) the corresponding wavelet ψ may be taken to have Fourier transform $(\mathcal{F}\psi)(y)$ equal to

$$s(y)(\sin^{m+1}\tfrac{1}{2}\pi y)\Big(\frac{\Delta_{2m}(\frac{1}{2}y+\frac{1}{2})}{\Delta_{2m}(y+1)}\Big)^{1/2}\Delta_{2m}(\tfrac{1}{2}y)^{-1/2}(\mathcal{F}\,{}^1\gamma)(\tfrac{1}{2}y),$$

where $s(y)$ is 1 if m is odd and is $e^{\pi i y}$ if m is even,
(d) the Fourier series $\Delta_{2m}(y)^{-1/2} = \sum_k d_k e^{-2\pi i k y}$ has the property that $\varphi(x)$ is given by $\varphi(x) = \sum_k d_k{}^m\gamma(x-k)$, the series being locally a finite sum,
(e) the Fourier series $\sum_k u_k e^{-\pi i k y}$ of period 2 of the function

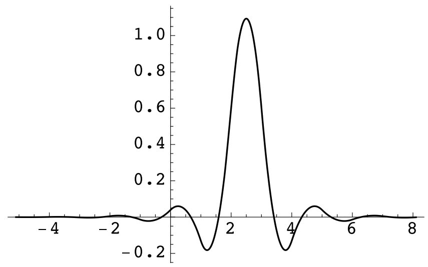
$$U(y) = s(y)(\sin^{m+1}\tfrac{1}{2}\pi y)\Big(\frac{\Delta_{2m}(\frac{1}{2}y+\frac{1}{2})}{\Delta_{2m}(y+1)}\Big)^{1/2}\Delta_{2m}(\tfrac{1}{2}y)^{-1/2}$$

has the property that $\psi(x)$ is given by $\psi(x) = 2\sum_k u_k{}^m\gamma(2x-k)$, the series being locally a finite sum,
(f) $\varphi(x)$ and $\psi(x)$ are splines of degree m with infinite support, the knots of φ being in $\mathbb{Z}$ and the knots of ψ being in $\frac{1}{2}\mathbb{Z}$,
(g) $\varphi(x)$ and $\psi(x)$ decay exponentially fast as $|x|$ tends to infinity.

REMARK. The wavelet obtained from Theorem 10.33 is called the **Battle–Lemarié wavelet** of degree[17] m. See Figures 10.14 and 10.15 for graphs of φ and ψ in the cases $m = 2$ and $m = 3$.

[17] As was mentioned in connection with Theorem 10.25, terminology varies. Some authors use the word "order" in place of "degree." Some authors shift the indices by 1, saying that the case here is of order $m + 1$.

(a)



(b)

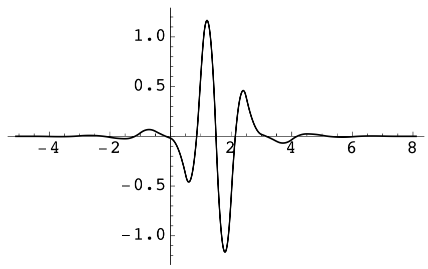
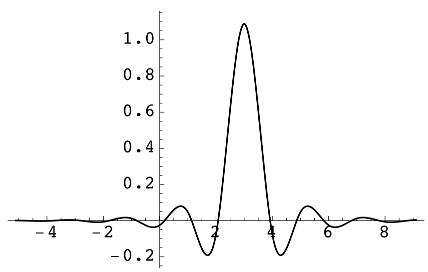


FIGURE 10.14. Graphs of Battle–Lemarié scaling function and wavelet of degree 2. (a) Scaling function. (b) Wavelet.

(a)



(b)

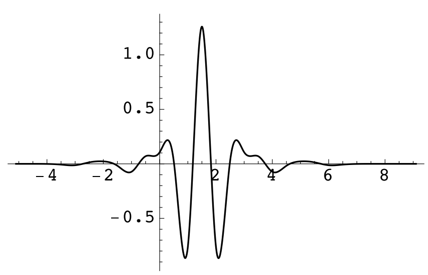


FIGURE 10.15. Graphs of Battle–Lemarié scaling function and wavelet of degree 3. (a) Scaling function. (b) Wavelet.

PROOF. Lemma 10.32 shows that $P_{2m}(y)$ is a polynomial in $\sin \pi y$ and $\cos \pi y$ that is everywhere positive. Being a continuous periodic function that is everywhere positive, $\Delta_{2m}(y)$ is bounded above and below by positive constants. We can therefore use the orthonormalization trick introduced in the proof of Theorem 10.25 to show that the integer translates of φ form an orthonormal set.

For (a), we define $\varphi_{0,k}(x) = \varphi(x-k)$, and let V be the closure of its linear span. We use Lemma 10.12 to argue as in Theorem 10.25a that $V = V_0$.

For (b) we use Propositions 10.15 and 10.17 just as in the proof of Theorem 10.25b to complete the argument that $\{V_j\}_{j=-\infty}^{\infty}$ and φ constitute a multiresolution analysis.

For (c) we have to compute ψ. The Fourier transform of the scaling equation gives us $(\mathcal{F}\varphi)(y) = m_0(y/2)(\mathcal{F}\varphi)(y/2)$, from which we have

$$\begin{aligned} m_0(y/2) &= \frac{(\mathcal{F}\varphi)(y)}{(\mathcal{F}\varphi)(y/2)} \\ &= \frac{\Delta_{2m}(y)^{-1/2}(\mathcal{F}^1\gamma)(y)}{\Delta_{2m}(y/2)^{-1/2}(\mathcal{F}^1\gamma)(y/2)} \\ &= \frac{\Delta_{2m}(y)^{-1/2}e^{-(m+1)\pi i y}(\pi y)^{-(m+1)}\sin^{m+1}\pi y}{\Delta_{2m}(y/2)^{-1/2}e^{-\frac{1}{2}(m+1)\pi i y}(\pi y/2)^{-(m+1)}\sin^{m+1}\pi y/2} \\ &= e^{-\frac{1}{2}(m+1)\pi i y}\Big(\frac{\Delta_{2m}(\frac{1}{2}y)}{\Delta_{2m}(y)}\Big)^{1/2}\cos^{m+1}\tfrac{1}{2}\pi y. \end{aligned}$$

From Theorem 10.10 we may take

$$\begin{aligned} (\mathcal{F}\psi)(y) &= e^{\pi i y}\nu(\tfrac{1}{2}y)\overline{m_0(\tfrac{1}{2}y+\tfrac{1}{2})}(\mathcal{F}\varphi)(\tfrac{1}{2}y) \\ &= e^{\pi i y}\nu(\tfrac{1}{2}y)e^{\frac{1}{2}(m+1)\pi i y}e^{\frac{1}{4}\pi i(m+1)} \\ &\quad\times(\sin^{m+1}\tfrac{1}{2}\pi y)\Big(\frac{\Delta_{2m}(\frac{1}{2}y+\frac{1}{2})}{\Delta_{2m}(y+1)}\Big)^{1/2}(\mathcal{F}\varphi)(\tfrac{1}{2}y) \\ &= \nu(\tfrac{1}{2}y)e^{\frac{1}{2}(m+3)\pi i y}e^{\frac{1}{4}(m+1)\pi i} \\ &\quad\times(\sin^{m+1}\tfrac{1}{2}\pi y)\Big(\frac{\Delta_{2m}(\frac{1}{2}y+\frac{1}{2})}{\Delta_{2m}(y+1)}\Big)^{1/2}\Delta_{2m}(\tfrac{1}{2}y)^{-1/2}(\mathcal{F}^m\gamma)(\tfrac{1}{2}y). \end{aligned}$$

with ν periodic of period 1. If we take

$$\nu(y) = \begin{cases} e^{-\frac{1}{4}(m+1)\pi i}e^{-(m+3)\pi i y} & \text{if } m \text{ is odd} \\ e^{-\frac{1}{4}(m+1)\pi i}e^{-(m+2)\pi i y} & \text{if } m \text{ is even,} \end{cases}$$

then

$$\nu(\tfrac{1}{2}y)e^{\frac{1}{2}(m+3)\pi iy}e^{\frac{1}{4}(m+1)\pi i} = \begin{cases} 1 & \text{if } m \text{ is odd} \\ e^{\pi iy} & \text{if } m \text{ is even,} \end{cases}$$

and (c) follows.

The proofs of (d) through (g) are straightforward adaptations of the corresponding arguments within the proof of Theorem 10.25. □

9. Daubechies Wavelets

Recall that the Haar wavelet has the desirable property of having compact support and the undesirable property of being discontinuous. Studying the Haar wavelet in the context of the general theory of wavelets leads one naturally to the question whether there are continuous wavelets of compact support. The existence of the Daubechies wavelets, as established in this section, gives an affirmative answer to this question.

Indeed, the compactness of their supports is such a helpful property that the Daubechies wavelets are the ones used in many applications. Although the general method used in their construction allows for many alternative formulations, their usual definitions include just one wavelet for each positive integer N, which is called the **order**. The order of a Daubechies wavelet is one more than the degree of a certain polynomial used in the definition of the scaling function, and it is half the number of nonzero coefficients of the generating function[18] m_0. It is related to the order of differentiability of the wavelet in a complicated way that will be discussed in part in the next section. The Daubechies wavelet of order 1 is the Haar wavelet. All Daubechies wavelets of higher order are continuous. Unlike with the Meyer wavelets, there is no Daubechies wavelet of infinite order; we shall see in the next section that no compactly supported wavelet of class C^∞ exists.

No closed form is known for the definition of the Daubechies wavelets of order greater than 1. They are constructed within the context of a multiresolution analysis by a limiting process.

The starting point for the construction is the scaling equation, specifically the generating function m_0 that appears in the scaling equation. Recall that the scaling equation is of the form

$$\varphi(x) = \sum_{k=-\infty}^{\infty} a_k \sqrt{2}\,\varphi(2x-k)$$

[18] The function m_0 is sometimes called the **scaling filter** or **low-pass filter**, but its role as a filter will not be part of our discussion until Section 11.

for constants $a_k = \int_{\mathbb{R}} \sqrt{2}\,\varphi(x)\overline{\varphi(2x-k)}\,dx$ with $\sum |a_k|^2 = 1$ and that the associated generating function is given by

$$m_0(y) = \tfrac{1}{\sqrt{2}} \sum_{k=-\infty}^{\infty} a_k e^{-2\pi i k y}.$$

If φ has compact support, then a_k can be nonzero for only finitely many k, and hence $m_0(y)$ is a trigonometric polynomial.[19] Following the Daubechies approach, we wish to arrange that *the scaling function φ is real-valued and has compact support*. The inner product formula for the coefficients a_k then shows that a_k is real for all k.

The wavelet is given by a similar equation, the wavelet equation, with coefficients b_k related to the a_k's. As a consequence only finitely many b_k's will be nonzero, and thus we see that having φ of compact support forces the wavelet ψ to have compact support.

Before beginning the construction, let us peek ahead and see that the compactly supported scaling function φ must have $|(\mathcal{F}\varphi)(0)| = 1$. (In particular, φ must have nonzero integral over $\mathbb{R}$.) Since φ is to be a scaling function for a multiresolution analysis, the integer translates of φ are to form an orthonormal set. The closed linear span of the integer translates will be the space V_0, and we form spaces V_j as usual. The intersection of the V_j's is 0 automatically, according to Proposition 10.15. But when the V_j are increasing, it is not automatic that their union is dense in $L^2(\mathbb{R})$. Proposition 10.17 gives a sufficient condition, namely that $\mathcal{F}\varphi$ is a bounded function that is continuous and nonzero at $y = 0$.

Since φ is in $L^2(\mathbb{R})$ and has compact support, it has to be in $L^1(\mathbb{R})$ and its Fourier transform $\mathcal{F}\varphi$ has to be bounded and continuous. But what about the condition $(\mathcal{F}\varphi)(0) \neq 0$? Let us see that the union of the V_j's will actually fail to be dense in $L^2(\mathbb{R})$ unless $|(\mathcal{F}\varphi)(0)| = 1$. In fact, let f be any nonzero member of $L^2(\mathbb{R})$ whose Fourier transform is bounded and is compactly supported, say in the interval $[-M, M]$. If P_j is the orthogonal projection on V_j, then Lemma 10.16 gives

$$\|P_j f\|_2^2 = \int_{-M}^{M} |(\mathcal{F}f)(y)|^2\, |(\mathcal{F}\varphi)(2^{-j}y)|^2\, dy$$

when j is large enough so that $2^{j-1} > M$. The integrand is bounded by a multiple of $|(\mathcal{F}f)(y)|^2$, and dominated convergence gives us

$$\begin{aligned}\lim_{j\to\infty} \|P_j f\|_2^2 &= \int_{-M}^{M} |(\mathcal{F}f)|^2 |(\mathcal{F}\varphi)(0)|^2\, dy \\ &= |(\mathcal{F}\varphi)(0)|^2 \|\mathcal{F}f\|_2^2 = |(\mathcal{F}\varphi)(0)|^2 \|f\|_2^2.\end{aligned}$$

[19] A **trigonometric polynomial** is defined for current purposes to be a polynomial in $e^{2\pi i y}$ and $e^{-2\pi i y}$.

If $\bigcup_j V_j$ is dense in $L^2(\mathbb{R})$, then the left side equals $\|f\|_2^2$, which we are assuming is nonzero. Then it follows that $|(\mathcal{F}\varphi)(0)| = 1$. Hence a φ with compact support cannot yield a multiresolution analysis unless $|(\mathcal{F}\varphi)(0)| = 1$.

We are ready to construct the multiresolution analysis. For each step we shall derive necessary conditions and then either show that those conditions are sufficient or add some additional conditions to make them sufficient. These are the two steps:

(1) find the possible trigonometric polynomials that might serve as m_0,
(2) find how to determine φ from m_0, and find conditions on φ so that φ is the scaling function for a multiresolution analysis.

Step 1. In this step we look for m_0 in the form of a trigonometric polynomial. Let us write down conditions that m_0 must satisfy. A first condition is the identity

$$|m_0(y)|^2 + |m_0(y + \tfrac{1}{2})|^2 = 1$$

given by Lemma 10.13. We do not need to add the words "almost everywhere" because m_0 as a trigonometric polynomial is automatically continuous. We have just seen that we must have $(\mathcal{F}\varphi)(0) \neq 0$. From the Fourier transform of the scaling equation, namely

$$(\mathcal{F}\varphi)(y) = m_0(y/2)(\mathcal{F}\varphi)(y/2),$$

we see that $m_0(0) = 1$. Since $|m_0(0)|^2 + |m_0(\frac{1}{2})|^2 = 1$, we see that $m_0(\frac{1}{2}) = 0$. Then we can write

$$m_0(y) = \Big(\frac{1 + e^{-2\pi i y}}{2}\Big)^N \mathcal{L}(y)$$

for some trigonometric polynomial $\mathcal{L}(y)$, where $N \geq 1$ is the order of the zero of $m_0(y)$ at $y = \frac{1}{2}$. It will be observed in the proof below that since $m_0(y)$ has real coefficients, so does $\mathcal{L}(y)$. In Propositions 10.34 and 10.36 below, we shall determine all the functions m_0 that satisfy all conditions so far.

Proposition 10.34. Let a trigonometric polynomial $m_0(y)$ with real coefficients be given in the form

$$m_0(y) = \Big(\frac{1 + e^{-2\pi i y}}{2}\Big)^N \mathcal{L}(y)$$

with $N \geq 1$. If m_0 satisfies the equation $|m_0(y)|^2 + |m_0(y + \frac{1}{2})|^2 = 1$, then the trigonometric polynomial $L(y) = |\mathcal{L}(y)|^2$ can be written as

$$L(y) = P(\sin^2 \pi y)$$

for some ordinary polynomial P of the form

$$P(w) = P_N(w) + w^N R(\tfrac{1}{2} - w)$$

with

$$P_N(w) = \sum_{k=0}^{N-1} \binom{N-1+k}{k} w^k$$

and with R an odd ordinary polynomial such that $P(w) \geq 0$ for $0 \leq w \leq 1$. Conversely if the trigonometric polynomial $L(y) = |\mathcal{L}(y)|^2$ with real coefficients can be written as

$$L(y) = P(\sin^2 \pi y)$$

for some ordinary polynomial P of the form

$$P(w) = P_N(w) + w^N R(\tfrac{1}{2} - w)$$

with

$$P_N(w) = \sum_{k=0}^{N-1} \binom{N-1+k}{k} w^k$$

and with R an odd ordinary polynomial such that $P(w) \geq 0$ for $0 \leq w \leq 1$, then $m_0(y) = \left(\dfrac{1 + e^{-2\pi i y}}{2}\right)^N \mathcal{L}(y)$ satisfies the equation $|m_0(y)|^2 + |m_0(y + \frac{1}{2})|^2 = 1$.

REMARKS. What Proposition 10.34 says briefly is that a trigonometric polynomial $m_0(y) = \left(\frac{1}{2}(1 + e^{-2\pi i y})\right)^N \mathcal{L}(y)$ satisfying the functional equation leads us to a certain kind of polynomial P and that that kind of P leads us back to the integer N and the modulus squared $|\mathcal{L}(y)|^2$ of the function that appears in the formula for $m_0(y)$. To get all the way back to $m_0(y)$, rather than merely back to $|m_0(y)|^2$, we need a way of passing from $|\mathcal{L}(y)|^2$ to $\mathcal{L}(y)$. That step is postponed to Proposition 10.36.

To prove Proposition 10.34, we need a lemma.

Lemma 10.35. Let $P_N(w)$ be the polynomial $\sum_{k=0}^{N-1} \binom{2N-1}{k} w^k (1-w)^{N-k-1}$. Then P_N has the properties

(a) $P_N(w)(1-w)^N + w^N P_N(1-w) = 1$,

(b) $P_N(w) = \sum_{k=0}^{N-1} \binom{N-1+k}{k} w^k$,

(c) $0 \leq P_N(y) \leq P_N(1) = \binom{2N-1}{N}$ for $0 \leq y \leq 1$.

N	Polynomial $P_N(w)$
1	1
2	$1+2w$
3	$1+3w+6w^2$
4	$1+4w+10w^2+20w^3$
5	$1+5w+15w^2+35w^3+70w^4$
6	$1+6w+21w^2+56w^3+126w^4+252w^5$
7	$1+7w+28w^2+84w^3+210w^4+462w^5+924w^6$

FIGURE 10.16. Table of values of the polynomial $P_N(w)$ used in defining the Daubechies wavelet of order N.

PROOF. The Binomial Theorem gives

$$\begin{aligned}
1 &= (w+(1-w))^{2N-1} \\
&= \sum_{k=0}^{2N-1} \binom{2N-1}{k} w^k (1-w)^{2N-1-k} \\
&= (1-w)^N \sum_{k=0}^{N-1} \binom{2N-1}{k} w^k (1-w)^{N-1-k} \\
&\quad + w^N \sum_{k=N}^{2N-1} \binom{2N-1}{k} w^{k-N} (1-w)^{2N-1-k} \\
&= (1-w)^N P_N(w) + w^N P_N(1-w).
\end{aligned}$$

This proves (a). For (b) we observe that $Q_N(w) = \sum_{k=0}^{N-1} \binom{N-1+k}{k} w^k$ equals the sum of the terms of the Taylor series of $(1-w)^{-N}$ about $w=0$ through order $N-1$. From the result of (a),

$$(1-w)^{-N} = P_N(w) + w^N (1-w)^{-N} P_N(1-w).$$

On the other hand, $P_N(w)$ on the right side is a polynomial of degree $N-1$, and the other term on the right side is the product of w^N and an analytic function for $|w| < 1$. Thus $P_N(y) = Q_N(y)$. This proves (b). For (c), we use the result of (b) to see that $0 \le P_N(y) = \sum_{k=0}^{N-1} \binom{N-1+k}{k} y^k \le \sum_{k=0}^{N-1} \binom{N-1+k}{k} = P_N(1)$, and $P_N(1) = \binom{2N-1}{N}$ from the definition of $P_N(w)$. This proves (c). □

PROOF OF NECESSITY IN PROPOSITION 10.34. If a trigonometric polynomial $m(y) = \sum c_k e^{-2\pi iky}$ with real coefficients is divisible by $1 + e^{-2\pi iy}$, say as $m(y) = q(y)(1 + e^{-2\pi iy})$ with $q(y) = \sum d_k e^{-2\pi iky}$ equal to a trigonometric polynomial, then we see that $c_k = d_k + d_{k-1}$ for all k. Since $d_{k-1} = 0$ for some k, it follows that we can recursively determine the coefficients d_k and see that they are real. Iterating this result, we see that the trigonometric polynomial $\mathcal{L}(y)$ in the statement of the proposition has real coefficients.

The function $L(y) = |\mathcal{L}(y)|^2$ is a trigonometric polynomial with real coefficients, say $L(y) = \sum A_k e^{-2\pi iky}$, and it is real valued. Taking its complex conjugate, we see that $A_k = A_{-k}$ for each k, hence that $L(y)$ is a finite linear combination of the functions $\cos 2\pi ky$ with real coefficients. Each $\cos 2\pi ky$ is a polynomial function of $\cos 2\pi y$, and thus $L(y)$ is a polynomial in $\cos 2\pi y$ with real coefficients, say $L(y) = Q(\cos 2\pi y)$.

Consider $|m_0(y)|^2$. This has

$$\begin{aligned} |m_0(y)|^2 &= \Big(\frac{1+e^{-2\pi iy}}{2}\Big)^N \Big(\frac{1+e^{2\pi iy}}{2}\Big)^N L(y) \\ &= \big(\cos^2 \pi y\big)^N Q(\cos 2\pi y). \end{aligned} \tag{$*$}$$

Since $\sin^2 \pi y = \frac{1}{2}(1 - \cos 2\pi y)$, we can rewrite $Q(\cos 2\pi y)$ as $P(\sin^2 \pi y)$ for a polynomial P with real coefficients. Define a function $w = w(y)$ by $w(y) = \sin^2 \pi y$. Then $1 - w(y) = \cos^2 \pi y$, and $(*)$ becomes

$$|m_0(y)|^2 = (1 - w(y))^N P(w(y)). \tag{$**$}$$

To work with the corresponding formula for $|m_0(y + \frac{1}{2})|^2$, we need to compute $w(y+\frac{1}{2})$. Observe that $\sin \pi(y+\frac{1}{2}) = \sin \pi y \cos \pi/2 + \cos \pi y \sin \pi/2 = \cos \pi y$. Thus $w(y + \frac{1}{2}) = \sin^2 \pi(y + \frac{1}{2}) = \cos^2 \pi y = 1 - \sin^2 \pi y = 1 - w(y)$, and

$$|m_0(y + \tfrac{1}{2})|^2 = (1 - w(y + \tfrac{1}{2}))^N P(w(y + \tfrac{1}{2})) = (w(y))^N P(1 - w(y)). \tag{$\dagger$}$$

Adding $(**)$ and $(\dagger)$ gives

$$1 = |m_0(y)|^2 + |m_0(y + \tfrac{1}{2})|^2 = (1 - w)^N P(w) + w^N P(1 - w). \tag{$\dagger\dagger$}$$

In view of Lemma 10.35a, we have $1 = (1 - w)^N P_N(w) + w^N P_N(1 - w)$. Subtracting $(\dagger\dagger)$ and this equation gives

$$(1 - w)^N \big(P(w) - P_N(w)\big) + w^N \big(P(1 - w) - P_N(1 - w)\big) = 0, \tag{$\ddagger$}$$

from which it follows that w^N divides $P(w) - P_N(w)$.

Let write $P(w) - P_N(w) = w^N R(\frac{1}{2} - w)$ for a polynomial R. Equation $(\ddagger)$ shows that

$$w^N (1 - w)^N R(\tfrac{1}{2} - w) + w^N (1 - w)^N R(\tfrac{1}{2} - (1 - w)) = 0.$$

Hence $R(-w) = -R(w)$, and R is odd. $\square$

PROOF OF SUFFICIENCY IN PROPOSITION 10.34. If $L(y) = |\mathcal{L}(y)|^2$ can be written as $L(y) = P(\sin^2 \pi y)$ with P of the form

$$P(w) = P_N(w) + w^N R(\tfrac{1}{2} - w)$$

and R odd, then we use Lemma 10.35a to write

$$P_N(w)(1-w)^N + w^N P_N(1-w) = 1. \qquad (\ddagger\ddagger)$$

The oddness of R ensures that

$$w^N R(\tfrac{1}{2} - w)(1-w)^N + w^N(1-w)^N R(\tfrac{1}{2} - (1-w)) = 0. \qquad (\S)$$

Adding (‡‡) and (§) gives

$$P(w)(1-w)^N + w^N P(1-w) = 1.$$

Substituting $w = \sin^2 \pi y$ shows that

$$P(\sin^2 \pi y)(\cos^2 y)^N + (\sin^2 y)^N P(\cos^2 \pi y) = 1. \qquad (\S\S)$$

We saw in (∗∗) that $|m_0(y)|^2 = (1-w)^N P(w)$ and in (†) that $|m_0(y+\frac{1}{2})|^2 = w^N P(1-w)$. Substituting these two relations into (§§) gives

$$|m_0(y)|^2 + |m_0(y + \tfrac{1}{2})|^2 = 1,$$

as required. □

Proposition 10.36 (Fejér and F. Riesz).[20] If $t(y) = \sum_{k=-M}^{M} c_k e^{-2\pi i k y}$ is a trigonometric polynomial that is everywhere ≥ 0, then there exists a trigonometric polynomial $\mathcal{L}(y) = \sum_{k=-M}^{M} d_k e^{-2\pi i k y}$ such that $t(y) = |\mathcal{L}(y)|^2$. If all the c_k are real, then all the d_k can be taken to be real.

REMARKS.

(1) Proposition 10.36 adds to Proposition 10.34 a final step in the inverse direction. Instead of merely passing from a certain kind of polynomial P to the integer N and $|\mathcal{L}(y)|^2$, we can now pass back from P to N and $\mathcal{L}(y)$ itself, therefore to m_0.

(2) No uniqueness result is asserted in Proposition 10.36. In the construction of $\mathcal{L}(y)$ in the proof below, we do not need to take all the roots α_j inside the unit disk and none of the roots $\overline{\alpha_j}^{\,-1}$ outside the unit disk. We merely have to select one root from each pair $\alpha_j, \overline{\alpha_j}^{\,-1}$. If there are r pairs of roots, the total number of possibilities for $\mathcal{L}(y)$ is 2^r.

[20] The original version appears on page 117 of the book by Riesz and Sz.-Nagy.

PROOF. We may assume that $t(y)$ is not identically 0. Since $t(y)$ is real valued, we have an equality

$$\sum_{k=-M}^{M} c_{-k} e^{2\pi i k y} = \sum_{k=-M}^{M} c_k e^{-2\pi i k y} = t(y) = \overline{t(y)} = \sum_{k=-M}^{M} \overline{c_k} e^{2\pi i k y},$$

and we must have equality term by term. Therefore $c_{-k} = \overline{c_k}$ for all k. We may assume that M is as small as possible, and then $c_{-M} \neq 0$. Define an ordinary polynomial by

$$Q(z) = c_{-M} + \cdots + c_M z^{2M} = \overline{c_M} + \cdots + \overline{c_{-M}} z^{2M}.$$

Then

$$t(y) = e^{2\pi i M y} Q(e^{-2\pi i y}) \qquad (*)$$

and

$$z^{2M} Q(z^{-1}) = \overline{Q(\bar{z})}. \qquad (**)$$

From $(**)$ it follows that if $Q(z_0) = 0$ with z_0 outside the closed unit disk, then $Q(1/\overline{z_0}) = 0$ with $1/\overline{z_0}$ inside the unit disk, and vice versa. Furthermore, the multiplicities match. Observe that $Q(z)$ cannot vanish at $z = 0$ because $c_{-M} \neq 0$.

Suppose that $Q(z)$ has a zero at $e^{-2mi\theta_0}$ of order m. Then $\theta \mapsto Q(e^{-2\pi i\theta})$ has a zero at $\theta = \theta_0$, and this zero is of the same order m because the exponential function has a locally defined inverse about every point of its image. Similarly $\theta \mapsto e^{2\pi i M\theta} Q(e^{-2\pi i\theta})$ has a zero of order m at $\theta = \theta_0$. This function is real-valued, matching $t(\theta)$ for $\theta \in \mathbb{R}$, and it equals the sum of a multiple of $(\theta - \theta_0)^m$ and a small error term for $|\theta - \theta_0|$ small. Since t takes on no negative values, m has to be even.

All that being so, let $\{\alpha_j\}$ be the zeros of $Q(z)$ inside the open unit disk, repeated according to their multiplicities, and let $\{\beta_k\}$ be the distinct zeros of $Q(z)$ on the unit circle, with β_k having multiplicity $2m_k$. Since a polynomial is determined up to a multiplicative constant by its roots and their multiplicities, we have

$$Q(z) = c\Big(\prod_j (z - \alpha_j)(z - (\overline{\alpha_j})^{-1})\Big)\Big(\prod_k (z - \beta_k)^{2m_k}\Big)$$

and

$$t(y) = ce^{2\pi i M y}\Big(\prod_j (e^{-2\pi i y} - \alpha_j)(e^{-2\pi i y} - (\overline{\alpha_j})^{-1})\Big)\Big(\prod_k (e^{-2\pi i y} - \beta_k)^{2m_k}\Big).$$

Put

$$r(y) = \Big(\prod_j (e^{-2\pi i y} - \alpha_j)\Big)\Big(\prod_k (e^{-2\pi i y} - \beta_k)^{m_k}\Big).$$

Then

$$|r(y)|^2 = \Big(\prod_j |e^{-2\pi iy} - \alpha_j|^2\Big)\Big(\prod_k |e^{-2\pi iy} - \beta_k|^{2m_k}\Big)$$

Now

$$\begin{aligned}|e^{-2\pi iy} - \alpha_j|^2 &= (e^{-2\pi iy} - \alpha_j)(e^{2\pi iy} - \overline{\alpha_j})\\ &= -e^{2\pi iy}\overline{\alpha_j}(e^{-2\pi iy} - \alpha_j)(e^{-2\pi iy} - (\overline{\alpha_j}\,)^{-1})\end{aligned}$$

and

$$|e^{-2\pi iy} - \beta_k|^2 = (e^{-2\pi iy} - \beta_k)(e^{2\pi iy} - \overline{\beta_k}\,) = -\overline{\beta_k}e^{2\pi iy}(e^{-2\pi iy} - \beta_k)^2.$$

Comparing, we see that there is some integer p with

$$|r(y)|^2 = c'e^{2\pi ipy}t(y) \qquad \text{and with } c' \text{ in } \mathbb{C}.$$

Since $|r(y)|^2$ and $t(y)$ are nonnegative and not identically 0, we see that c' is positive and that $p = 0$. Putting $\mathcal{L}(y) = r(y)\sqrt{c'}^{\,-1}$, we obtain a trigonometric polynomial whose absolute value squared equals $t(y)$.

If all the c_k are real, then $Q(z)$ has real coefficients and its roots come in complex conjugate pairs. If α_j used in the definition of $r(y)$, then we use $\overline{\alpha_j}$ also, with the same multiplicity. Then the result is that $r(y)$ has real coefficients, and so does $\mathcal{L}(y)$. □

Step 2. In this step we determine φ from m_0. We start from the Fourier transform of the scaling equation,

$$(\mathcal{F}\varphi)(y) = m_0(y/2)(\mathcal{F}\varphi)(y/2).$$

We know from earlier that $m_0(0) = 1$, that $m_0(\frac{1}{2}) = 0$, and that m_0 is a trigonometric polynomial periodic of period 1. If we iterate the formula, we obtain

$$(\mathcal{F}\varphi)(y) = \big(\prod_{j=1}^{n} m_0(2^{-j}y)\big)(\mathcal{F}\varphi)(2^{-n}y).$$

Proposition 10.37. If φ is a compactly supported scaling function, then the infinite product

$$\prod_{j=1}^{n} m_0(2^{-j}y)$$

converges (to a limit that is nonzero at points where all the factors are nonzero), the convergence being uniform on compact sets in $\mathbb{R}$, and

$$(\mathcal{F}\varphi)(y) = \big(\prod_{j=1}^{\infty} m_0(2^{-j}y)\big)(\mathcal{F}\varphi)(0)$$

with $|(\mathcal{F}\varphi)(0)| = 1$.

PROOF. Fix $M > 0$. We prove uniform convergence for $|y| \le M$. By Taylor's Theorem, $|m_0(y) - 1| \le C|y|$ for $|y| \le M$, say. Then also $|m_0(2^{-j}y) - 1| \le C2^{-j}M$. Since $\sum_{j=1}^{\infty} C2^{-j}M$ converges, the given infinite product converges as asserted.[21] Therefore

$$(\mathcal{F}\varphi)(y) = \Big(\prod_{j=1}^{\infty} m_0(2^{-j}y)\Big)\Big(\lim_{n\to\infty}(\mathcal{F}\varphi)(2^{-n}y)\Big).$$

The second factor on the right equals $(\mathcal{F}\varphi)(0)$ since the Fourier transform of the L^1 function φ is continuous, and we have seen that $|(\mathcal{F}\varphi)(0)| = 1$. □

Now let us consider the converse direction, to pass from m_0 to φ. We need to remember that we are seeking a real-valued φ. Then $(\mathcal{F}\varphi)(0)$, being the integral of φ, must be real, and so we must have $(\mathcal{F}\varphi)(0) = \pm 1$. This minus sign is harmless, and we might as well aim for $(\mathcal{F}\varphi)(0) = 1$ and define $\mathcal{F}\varphi$ by the product formula in the proposition:

$$(\mathcal{F}\varphi)(y) = \prod_{j=1}^{\infty} m_0(2^{-j}y).$$

Sadly if we start from m_0 produced from Step 1 and we define φ this way, the result is not necessarily a scaling function: the set of its integer translates can fail to be orthonormal. We need an extra assumption. A necessary and sufficient condition is known, but we shall not give it. Instead we give a sufficient condition that we can easily verify in the examples of interest.

Proposition 10.38. Suppose that $m_0(y)$ is a trigonometric polynomial such that

(i) $m_0(0) = 1$,
(ii) $|m_0(y)|^2 + |m_0(y + \frac{1}{2})|^2 = 1$, and
(iii) $m_0(y)$ is nonzero for $|y| \le \frac{1}{4}$.

Define

$$h(y) = \prod_{j=1}^{\infty} m_0(2^{-j}y).$$

Then the infinite product converges uniformly on compact sets of $\mathbb{R}$, and the limit function h is a smooth function in $L^2(\mathbb{R})$ with $h(0) = 1$. Moreover the inverse Fourier transform $\varphi = \mathcal{F}^{-1}h$ has compact support, has the property that its integer translates are orthonormal, and satisfies the scaling equation $(\mathcal{F}\varphi)(y) = m_0(y/2)(\mathcal{F}\varphi)(y/2)$. Consequently φ is the scaling function for a multiresolution analysis. If $m_0(y)$ has real coefficients, then φ is real-valued.

[21] For a proof of the standard facts relating convergence of infinite series with convergence of infinite products, see the author's *Elliptic Curves*, page 195. For the uniformity of the convergence, we are incorporating the Weierstrass M test into the relevant theorem.

PROOF. For the moment let us regard the variable y as complex. Observe that $m_0(y)$ is a trigonometric polynomial and is therefore entire. We check that the product defining $h(y)$ converges uniformly on compact subsets of $\mathbb{C}$, and hence $h(y)$ is an entire function.[22] Fix $K > 0$. From assumption (i), $m_0(0) = 1$, and therefore $|m_0(y) - 1| \le A|y|$ for all complex y with $|y| \le K$. Then also $|m_0(2^{-j}y) - 1| \le A2^{-j}K$. Since $\sum_{j=1}^{\infty} A2^{-j}K$ converges, the given infinite product $h(y) = \prod_{j=1}^{\infty} m_0(2^{-j}y)$ converges uniformly[23] on the compact subset of $\mathbb{C}$ where $|y| \le K$. The number K being arbitrary, $h(y)$ is an entire function of the complex variable y.

We are going to estimate the size of $h(y)$ as a function of the complex variable y. We begin by estimating the size of the entire function $m_0(y)$. Write m_0 out in the form of a trigonometric polynomial as $m_0(y) = \frac{1}{\sqrt{2}} \sum_{k=-M}^{M} a_k e^{-2\pi i k y}$. Since $m_0(0) = 1$, we have

$$\sum_{k=-M}^{M} \tfrac{1}{\sqrt{2}} a_k e^{-2\pi i k y} - 1 = \sum_{k=-M}^{M} \tfrac{1}{\sqrt{2}} a_k (e^{-2\pi k y} - 1)$$

and thus

$$|m_0(y) - 1| \le \sum_{k=-M}^{M} \tfrac{1}{\sqrt{2}} |a_k| |e^{2\pi i k y} - 1|.$$

Put $C = \sum_{k=-M}^{M} \frac{1}{\sqrt{2}} |a_k|$. Let us record that $\sum_{k=-M}^{M} \frac{1}{\sqrt{2}} a_k = 1$ implies

$$C = \sum_{k=-M}^{M} \tfrac{1}{\sqrt{2}} |a_k| \ge 1. \qquad (*)$$

By comparing power series term by term, we have

$$|e^{2\pi i k y} - 1| \le e^{2\pi k |y|} - 1 \le e^{2\pi M |y|} - 1,$$

Consequently

$$|m_0(y) - 1| \le \Big(\sum_{k=-M}^{M} \tfrac{1}{\sqrt{2}} |a_k| \Big) (e^{2\pi M |y|} - 1) = C(e^{2\pi M |y|} - 1).$$

[22] A sequence of analytic functions converges to an analytic function if the convergence is uniform on compact sets, according to Problem 55 in Appendix B of *Basic*.

[23] This is the same argument as in the proof of Proposition 10.37 except that we are now allowing the variable to be complex.

Since $C \geq 1$ by $(*)$, we have

$$C(e^{2\pi M|y|}-1)=\sum_{n=1}^{\infty}\frac{C(2\pi M|y|)^n}{n!}\leq\sum_{n=1}^{\infty}\frac{(2\pi CM|y|)^n}{n!}=e^{2\pi CM|y|}-1.$$

Thus

$$|m_0(y)|\leq 1+|m_0(y)-1|\leq e^{2\pi CM|y|}. \qquad (**)$$

The product of the first J factors defining $h(y)$ is $\prod_{j=1}^{J} m_0(2^{-j}y)$, and its absolute value, according to $(**)$, is

$$\prod_{j=1}^{J}|m_0(2^{-j}y)|\leq\prod_{j=1}^{J}e^{2\pi CM|2^{-j}y|}=\exp\big(2\pi CM|y|\sum_{j=1}^{J}2^{-j}\big)\leq e^{2\pi CM|y|}.$$

Letting J tend to infinity, we obtain

$$|h(y)|\leq e^{2\pi CM|y|}. \qquad (\dagger)$$

This estimate, which we shall return to a little later, is summarized by saying that the entire function $h(y)$ is of **exponential type**.

For the reminder of the proof, y can be regarded as a real variable. Let us write the inner product of f and g in $L^2(\mathbb{R})$ as (f, g), and let us define a translation operator $T_k : L^2(\mathbb{R}) \to L^2(\mathbb{R})$ for $k \in \mathbb{Z}$ by $T_k f(x) = f(x-k)$. Each operator T_k is unitary. Put

$$\mathcal{C}=\{f\in L^2(\mathbb{R})\bigm| \|f\|_2=1 \text{ and } (f,T_kf)=0 \text{ for all } k\neq 0 \text{ in } \mathbb{Z}\}.$$

We shall prove that $\mathcal{C}$ is a closed set. In fact, suppose that $\{f_n\}$ is a sequence in $\mathcal{C}$ convergent to f in $L^2(\mathbb{R})$. Since the norm is continuous, $\|f\|_2 = 1$. For nonzero k in $\mathbb{Z}$, the equality

$$(f,T_kf)-(f_n,T_kf_n)=(f-f_n,T_kf)+(f_n,T_kf-T_kf_n)$$

together with the triangle inequality, the Schwarz inequality, and the fact that T_k is unitary give

$$\begin{aligned}|(f,T_kf)-(f_n,T_kf_n)|&\leq\|f-f_n\|_2\|T_kf\|_2+\|f_n\|_2\|T_k(f-f_n)\|_2\\&=\|f\|_2\|f-f_n\|_2+\|f_n\|_2\|f-f_n\|_2\\&=2\|f-f_n\|_2.\end{aligned}$$

The right side has limit 0. Since $k \neq 0$ and f_n is in $\mathcal{C}$, the left side reduces to $|(f, T_k f)|$, and consequently $(f, T_k f) = 0$. Thus $\mathcal{C}$ is a closed set.

Let φ_0 be the L^2 function with $\mathcal{F}\varphi_0 = I_{[-\frac{1}{2},\frac{1}{2}]}$, i.e., the scaling function of the Shannon wavelet in Section 4. It was shown in the proof of Theorem 10.14 that the integer translates of φ_0 are orthonormal. Hence φ_0 is in $\mathcal{C}$. Since m_0 has to be bounded as a function of a real variable, we can define L^2 functions φ_n inductively for $n \geq 1$ by

$$(\mathcal{F}\varphi_n)(y) = (\mathcal{F}\varphi_0)(2^{-n}y)\, m_0(2^{-1}y) \cdots m_0(2^{-n}y). \qquad (\dagger\dagger)$$

We shall prove inductively that each φ_n is in $\mathcal{C}$. The base case $n = 0$ of the induction having been settled, assume inductively that φ_{n-1} is in $\mathcal{C}$. Since $(\mathcal{F}\varphi_n)(y) = m_0(y/2)(\mathcal{F}\varphi_{n-1})(y/2)$, we have

$$\begin{aligned}
\sum_{l\in\mathbb{Z}} |(\mathcal{F}\varphi_n)(y+l)|^2 &= \sum_{l \text{ even}} |m_0(\tfrac{1}{2}(y+l))|^2 |(\mathcal{F}\varphi_{n-1})(\tfrac{1}{2}(y+l))|^2 \\
&\qquad + \sum_{l \text{ odd}} |m_0(\tfrac{1}{2}(y+l))|^2 |(\mathcal{F}\varphi_{n-1})(\tfrac{1}{2}(y+l))|^2 \\
&= |m_0(\tfrac{1}{2}y)|^2 \sum_{l\in\mathbb{Z}} |(\mathcal{F}\varphi_{n-1})(\tfrac{1}{2}y+l)|^2 \\
&\qquad + |m_0(\tfrac{1}{2}y+\tfrac{1}{2})|^2 \sum_{l\in\mathbb{Z}} |(\mathcal{F}\varphi_{n-1})(\tfrac{1}{2}(y+1)+l)|^2 \\
&= |m_0(\tfrac{1}{2}y)|^2 + |m_0(\tfrac{1}{2}y+\tfrac{1}{2})|^2,
\end{aligned}$$

the last equality holding by Lemma 10.11 and the inductive hypothesis. The last line of the above expression equals 1 by assumption (ii), and it follows from another application of Lemma 10.11 that φ_n is in $\mathcal{C}$. This completes the proof by induction that φ_n is in $\mathcal{C}$.

We know that the partial products $\prod_{j=1}^{J} m_0(2^{-j}y)$ converge uniformly on compact sets to what we have defined to be $h(y)$. Since $(\mathcal{F}\varphi_0)$ is continuous at 0 and has $(\mathcal{F}\varphi_0)(0) = 1$, a glance at $(\dagger\dagger)$ shows that

$$\lim_n (\mathcal{F}\varphi_n)(y) = h(y)$$

uniformly on compact sets. Applying Fatou's Lemma[24] to the absolute values squared, we obtain

$$\int_{\mathbb{R}} |h(y)|^2\, dy \leq \liminf \int_{\mathbb{R}} |(\mathcal{F}\varphi_n)(y)|^2\, dy.$$

The right side equals 1 since φ_n is in $\mathcal{C}$, and therefore the restriction of the entire function h to $\mathbb{R}$ is in $L^2(\mathbb{R})$ with $\|h\|_2 \leq 1$. We let $\varphi = \mathcal{F}^{-1}h$, so that φ is in $L^2(\mathbb{R})$ with $\|\varphi\|_2 \leq 1$.

[24] Theorem 5.29 of *Basic*.

We know that $\lim(\mathcal{F}\varphi_n) = \mathcal{F}\varphi$ uniformly on compact sets, and we are going to prove that the convergence takes place in $L^2(\mathbb{R})$. Once we have done so, the Plancherel Theorem will yield $\lim \varphi_n = \varphi$ in $L^2(\mathbb{R})$. Since $\mathcal{C}$ is closed and each φ_n is in $\mathcal{C}$, we will then have proved that the integer translates of φ are orthonormal, and we will be on our way toward obtaining a multiresolution analysis.

To prove that $\mathcal{F}\varphi_n$ converges to $\mathcal{F}\varphi$ in $L^2(\mathbb{R})$, we are going to use dominated convergence, showing that for all n, $|(\mathcal{F}\varphi_n)(y)|$ is $\le C_1|(\mathcal{F}\varphi)(y)|$ for some constant C_1 and all real y. By assumption (iii), $|m_0(y)|$ is nonzero for real y with $|y| \le \frac{1}{4}$. The function $m_0(y)$ is continuous (being a trigonometric polynomial), and thus there is a positive number c such that $|m_0(y)| \ge c > 0$ for real y with $|y| \le \frac{1}{4}$. By the uniform convergence of the partial products of $h(y)$ on compact sets, we can choose an integer J large enough so that $\prod_{j>J} |m_0(y2^{-j})| \ge \frac{1}{2}$ whenever $|y| \le \frac{1}{2}$ and y is real. Then we have

$$|(\mathcal{F}\varphi)(y)| = \Big(\prod_{j=1}^{J} |m_0(y2^{-j})|\Big)\Big(\prod_{j>J} |m_0(y2^{-j})|\Big) \ge \tfrac{1}{2}\prod_{j=1}^{J} |m_0(y2^{-j})| \ge \tfrac{1}{2}c^J$$

for $|y| \le \frac{1}{2}$. In other words, $\inf_{|y|\le 1/2} |(\mathcal{F}\varphi)(y)| \ge c_1 > 0$ for some positive constant c_1. The definition of $\mathcal{F}\varphi_n$ shows that $(\mathcal{F}\varphi_n)(y) = 0$ for $|y| > 2^{n-1}$ and hence that $|(\mathcal{F}\varphi_n)(y)| \le c_1^{-1}|(\mathcal{F}\varphi)(y)|$ there. Meanwhile for $|y| \le 2^{n-1}$, we have $(\mathcal{F}\varphi)(y) = (\mathcal{F}\varphi_n)(y)(\mathcal{F}\varphi)(2^{-n}y)$ and hence also

$$|(\mathcal{F}\varphi_n)(y)| = \left|\frac{(\mathcal{F}\varphi)(y)}{(\mathcal{F}\varphi)(2^{-n}y)}\right| \le c_1^{-1}|(\mathcal{F}\varphi)(y)|.$$

Thus

$$|(\mathcal{F}\varphi_n)(y)| \le c_1^{-1}|(\mathcal{F}\varphi)(y)| \qquad \text{for all } y.$$

Consequently $|(\mathcal{F}\varphi_n)(y) - (\mathcal{F}\varphi)(y)|^2$ is dominated for all y by the multiple $(c_1^{-1} + 1)^2$ of $|(\mathcal{F}\varphi)(y)|^2$, which in turn has been shown to be integrable. By dominated convergence we therefore have

$$\lim \int_{\mathbb{R}} |(\mathcal{F}\varphi_n)(y) - (\mathcal{F}\varphi)(y)|^2\, dy = \int_{\mathbb{R}} \lim |(\mathcal{F}\varphi_n)(y) - (\mathcal{F}\varphi)(y)|^2\, dy = 0.$$

Thus $\mathcal{F}\varphi_n$ converges to $\mathcal{F}\varphi$ in $L^2(\mathbb{R})$, as asserted.

Let us recapitulate. We formed $h(y)$ as an infinite product of dilates of the trigonometric polynomial $m_0(y)$, exhibited it as a limit of functions $\mathcal{F}\varphi_n$ uniformly on compact sets, and deduced from Fatou's Lemma that h is in $L^2(\mathbb{R})$. We defined $\varphi = \mathcal{F}^{-1}h$ as a member of $L^2(\mathbb{R})$, and we saw that φ is in $\mathcal{C}$, i.e., its integer translates form an orthonormal set. The definitions made $h(y) = m_0(y/2)h(y/2)$, i.e.,

$$(\mathcal{F}\varphi)(y) = m_0(y/2)(\mathcal{F}\varphi)(y/2),$$

and thus φ satisfies a scaling equation. We now define $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$ and let V_j be the closure of the linear span of $\{\varphi_{j,k}\}_{k=-\infty}^{\infty}$. The sequence $\{V_j\}$ is increasing, since φ satisfies a scaling equation. The claim is that $\{V_j\}$ and φ together form a multiresolution analysis in the sense of Section 3. Condition (iv) in the definition is satisfied, and condition (iii) is built into the definition. Proposition 10.15 shows that condition (ii) is automatic, and Proposition 10.17 says that condition (i) holds if $\mathcal{F}\varphi$ is a bounded function that is continuous and nonzero at $y = 0$. But $\mathcal{F}\varphi$ is just h, and we saw that h is the restriction to $\mathbb{R}$ of an entire function with $h(0) = 1$. Thus indeed (i) holds. Consequently $\{V_j\}$ and φ together form a multiresolution analysis.

Next we are to show that φ has compact support. We saw that the L^2 function h extends to an entire function of exponential type. By the classical Paley–Wiener Theorem,[25] the L^2 function $\varphi = \mathcal{F}^{-1}h$ has compact support.

Finally we are to show that if $m_0(y)$ has real coefficients, then φ is real-valued. Since $m_0(y)$ has real coefficients, m_0 satisfies $\overline{m_0(y)} = m_0(-y)$. Referring to the formulas, we then see that $h(y) = \overline{h(-y)}$ and $\overline{\varphi(x)} = \varphi(x)$. □

Now we are in a position to collect all our results and produce the wavelets of compact support.

Theorem 10.39. Fix an integer $N \geq 1$, and define

$$P_N(w) = \sum_{k=0}^{N-1} \binom{N-1+k}{k} w^k.$$

Let $\mathcal{L}(y)$ be any of the trigonometric polynomials produced by Proposition 10.36 such that $|\mathcal{L}(y)|^2 = P_N(\sin^2 \pi y)$, and define

$$m_0(y) = \Big(\frac{1 + e^{-2\pi i y}}{2}\Big)^N \mathcal{L}(y).$$

Let $h(y)$ be the L^2 function

$$h(y) = \prod_{j=1}^{\infty} m_0(2^{-j}y),$$

and define $\varphi = \mathcal{F}^{-1}h$. Then φ has compact support and is the scaling function of a multiresolution analysis whose wavelet can be taken to have Fourier transform given by

$$(\mathcal{F}\psi)(y) = e^{-i\pi y}\overline{m_0(\tfrac{1}{2}y + \tfrac{1}{2})}h(y/2).$$

[25]The classical Paley–Wiener Theorem says that a function h in $L^2(\mathbb{R})$ is of the form $h = \mathcal{F}\varphi$ for a function φ in $L^2(\mathbb{R})$ of compact support if and only if h (after adjustment on a set of measure 0) can be extended to all of $\mathbb{C}$ as an entire function of exponential type. This theorem is not included in Chapter VIII of *Basic*, but instead the sufficiency appears as Theorem 10.41 at the end of the present section. The necessity is much easier but is not needed.

REMARKS.

(1) Proposition 10.36 produces a number of trigonometric polynomials $\mathcal{L}(y)$ with $|\mathcal{L}(y)|^2 = P_N(\sin^2 \pi y)$, corresponding to the subsets of roots α_j of a certain polynomial inside the unit disk. The **Daubechies wavelet** of order N arises by choosing the subset to consist of *all* the roots within the unit disk. All other choices of subsets of roots, however, lead to wavelets. The subsets that are closed under complex conjugation lead to φ and ψ real-valued.

(2) Theorem 10.39 contains no assertion of smoothness or even continuity for the Daubechies wavelets. This issue is postponed until the next section, in which we shall see that the Daubechies wavelets of order ≥ 2 are continuous and that they become progressively more differentiable as N increases. However, they are never infinitely differentiable. Graphs of the scaling functions and wavelets appear in Figures 10.17a and 10.17b, except that the wavelets have been translated by integers to make them better centered.[26] The cases of small N look as if they were drawn with a shaky hand, and this fact reflects the low order of differentiability in these cases.

(3) The Daubechies wavelet of order 1 is just the Haar wavelet. In fact, we have $P_1(w) = 1$. The polynomial $\mathcal{L}(y)$ is to have $|\mathcal{L}(y)|^2 = 1$ and thus can be taken to be 1. Then $m_0(y) = \frac{1}{2}(1 + e^{-2\pi i y})$, in agreement with the formula in the example in Section 3. To compute $h(y) = \prod_{j=1}^{\infty} m_0(2^{-j}y)$, we write $m_0(y) = e^{-\pi i y}\cos \pi y$. The product of the exponentials $e^{-\pi i 2^{-j} y}$ comes out to be $e^{-\pi i y}$, and thus $h(y) = e^{-\pi i y}\prod_{j=1}^{\infty}\cos(2^{-j}\pi y)$. By making repeated use of the identity $\sin 2\theta = 2\sin\theta\cos\theta$, we obtain $2^J \sin(2^{-J}\pi y)\prod_{j=1}^{J}\cos(2^{-j}\pi y) =$ $= 2^{J-1}\sin(2^{-J+1}\pi y)\prod_{j=1}^{J-1}\cos(2^{-j}\pi y) = \cdots = \sin \pi y$. Therefore

$$\prod_{j=1}^{J}\cos(2^{-j}\pi y) = \frac{\sin \pi y}{2^J \sin(2^{-J}\pi y)}.$$

Using $\lim_{\theta\to 0}\theta^{-1}\sin\theta = 1$ and passing to the limit, we obtain $\prod_{j=1}^{\infty}\cos(2^{-j}\pi y) = (\sin \pi y)\big/(\pi y)$. Hence

$$h(y) = e^{-i\pi y}\,\frac{\sin \pi y}{\pi y}.$$

The right side is the Fourier transform of the indicator function $I_{[0,1]}$, and thus $\varphi = I_{[0,1]}$, as asserted.

[26] As the proof notes, the definition of $\mathcal{F}\psi$ involves an arbitrary periodic function ν of absolute value 1, and the figures use a power of $\nu(y) = e^{2\pi i y}$ to achieve this translation.

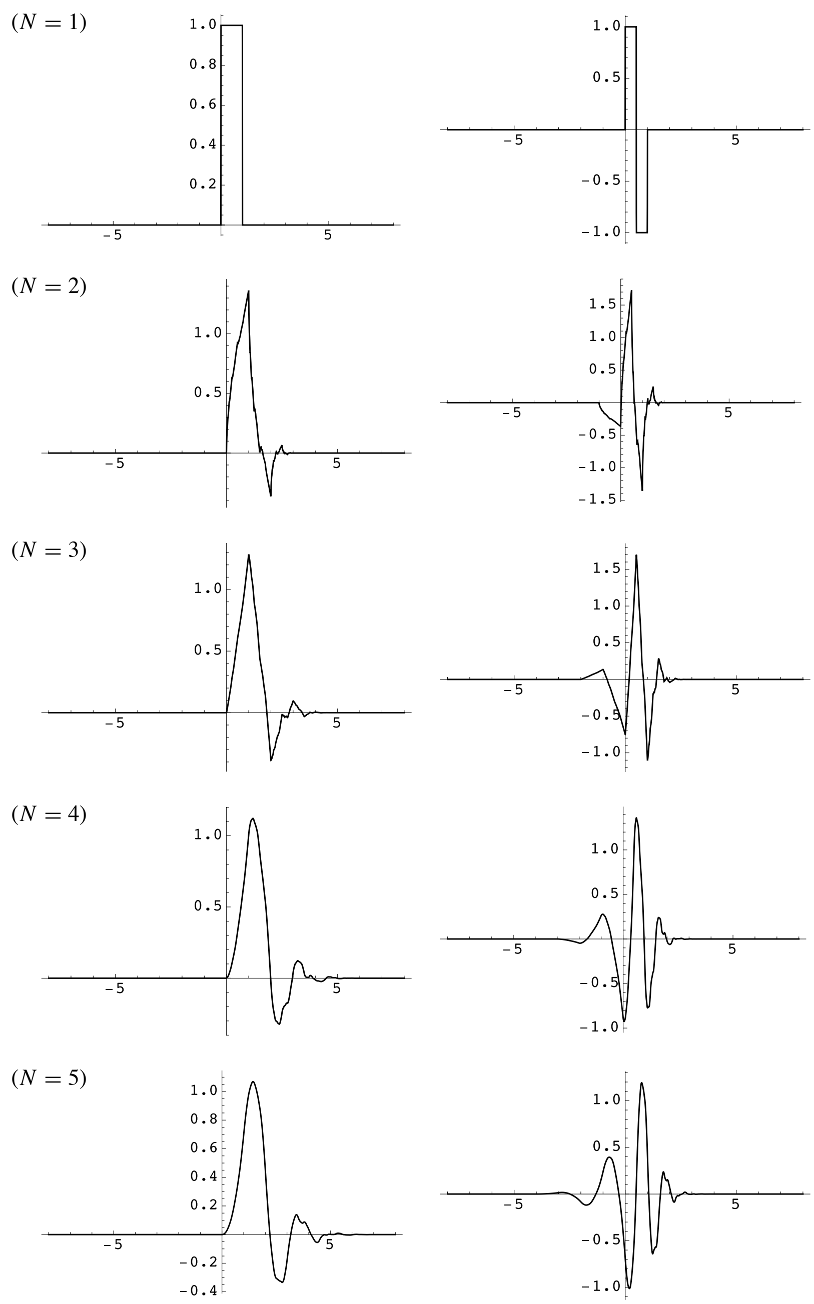


FIGURE 10.17a. Daubechies scaling function (left) and wavelet (right) of order N for $N = 1, 2, 3, 4, 5$.

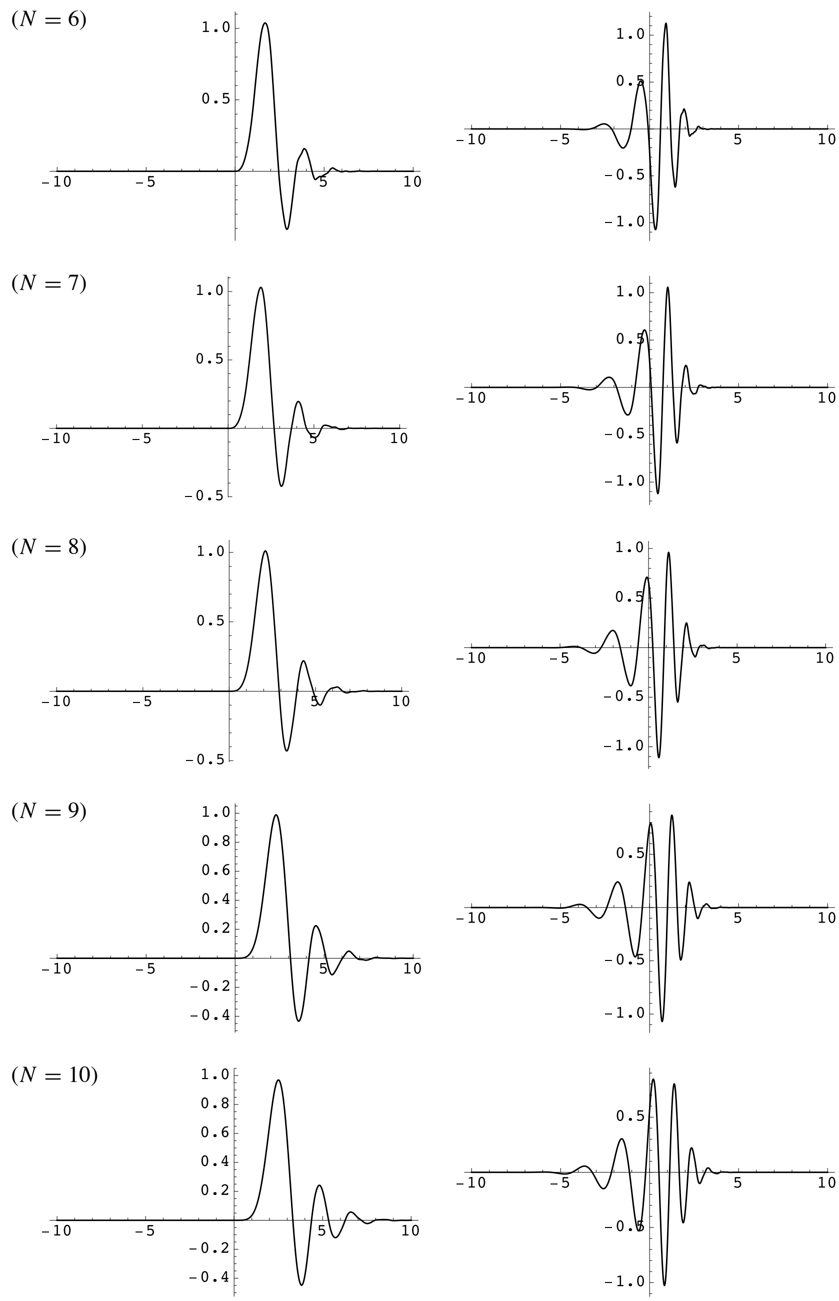


FIGURE 10.17b. Daubechies scaling function (left) and wavelet (right) of order N for $N = 6, 7, 8, 9, 10$.

(4) The Daubechies wavelet of order 2 is sufficiently complicated that no closed form is known for it. However, it is sufficiently simple that we can compute $m_0(y)$ exactly. For it, we have $P_2(w) = 1 + 2w$. Thus $P_2(\sin^2 \pi y) = 1 + 2\sin^2 \pi y = |\mathcal{L}(y)|^2$. Proposition 10.36 tells us how to find the possibilities for $\mathcal{L}(y)$. In the proposition the function t has $t(y) = 1 + 2\sin^2 \pi y = 2 - \cos 2\pi y = -\frac{1}{2}e^{-2\pi iy} + 2 - \frac{1}{2}e^{2\pi iy}$. In the notation of the proof of that proposition, we have $M = 1$, $Q(y) = -\frac{1}{2} + 2y - \frac{1}{2}y^2$, and $z^2 Q(z^{-1}) = \overline{Q(\bar{z})}$. The zeros of $Q(z)$ occur at $2 \pm \sqrt{3}$, and we are to use the one inside the unit disk, namely $2 - \sqrt{3}$, in defining $\mathcal{L}(y)$. The proof of the proposition says to use $r(y) = e^{-2\pi iy} - (2 - \sqrt{3})$ as a first approximation to $\mathcal{L}(y)$. Then $|r(y)|^2 = (8 - 4\sqrt{3}) - 2(2 - \sqrt{3}) \cos 2\pi y = c'(2 - \cos 2\pi y) = c' t(y)$, where $c' = 4 - 2\sqrt{3}$. Since $(c')^{-1/2} = \frac{1}{2}(1 + \sqrt{3})$, the proof of the proposition says that $\mathcal{L}(y) = \frac{1}{2}(1 + \sqrt{3}) r(y)$ has $|\mathcal{L}(y)|^2 = t(y)$. In other words, $\mathcal{L}(y) = \frac{1}{2}(1 + \sqrt{3})(e^{-2\pi iy} - (2 - \sqrt{3}))$. Then $m_0(y)$ is given by $= \frac{1}{4}(1 + e^{-2\pi iy})^2 \mathcal{L}(y)$, i.e.,

$$m_0(y) = \tfrac{1}{2}(1 + \sqrt{3})(e^{-2\pi iy} - (2 - \sqrt{3}))\tfrac{1}{4}(1 + e^{-2\pi iy})^2.$$

Observe that the expansion of $m_0(y)$ in terms of exponentials has exactly four nonzero coefficients—those for 1, $e^{-2\pi iy}$, $e^{-4\pi iy}$, and $e^{-6\pi iy}$.

(5) The Daubechies wavelet of order N, when expanded in terms of exponentials, has exactly $2N$ nonzero coefficients—those for $1, e^{-2\pi ix}, \ldots, e^{-2\pi i(2N-1)x}$. Decimal approximations of these coefficients for $2 \le N \le 10$ appear in a table[27] on page 195 of the award-winning book by Daubechies, *Ten Lectures on Wavelets*.

PROOF OF THEOREM 10.39. Proposition 10.34 shows that when defined as in the statement of the theorem, $m_0(y)$ satisfies $|m_0(y)|^2 + |m_0(y + \frac{1}{2})|^2 = 1$, as well as $m_0(0) = 1$. Also

$$\begin{aligned} |m_0(y)|^2 &= 2^{-2N} |1 + e^{-2\pi iy}|^{2N} P_N(\sin^2 \pi y) \\ &= (\cos^2 \pi y)^N \sum_{k=0}^{N-1} \binom{2N-1}{k} (\sin^2 \pi y)^k (\cos^2 \pi y)^{N-k} \\ &\ge (\cos^2 \pi y)^{2N}, \end{aligned}$$

and this is > 0 for $|y| \le \frac{1}{4}$. Therefore the hypotheses of Proposition 10.38 are satisfied, and we can conclude that

$$h(y) = \prod_{j=1}^{\infty} m_0(2^{-j} y)$$

[27] The coefficients in the table differ from those in the present book by a factor of $\sqrt{2}$. One can verify this fact for $N = 2$ by using the coefficients obtained for $N = 2$ in the previous remark.

defines an entire function for $y \in \mathbb{C}$ with $h(0) = 1$ whose restriction to $\mathbb{R}$ is in $L^2(\mathbb{R})$. Moreover, the proposition says that $\varphi = \mathcal{F}^{-1}h$ is compactly supported and is the scaling function for a multiresolution analysis. According to Theorem 10.10, the corresponding wavelet is any function of the form

$$(\mathcal{F}f)(y) = e^{\pi i y}\nu(y)\overline{m_0(\tfrac{1}{2}y + \tfrac{1}{2})}(\mathcal{F}\varphi)(y/2)$$

with ν periodic of period 1 and with $|\nu(y)| = 1$ everywhere. If we take into account that $\mathcal{F}\varphi = h$ and if we put $\nu(y) = e^{-2\pi i y}$, then the resulting formula is

$$(\mathcal{F}\psi)(y) = e^{-i\pi y}\overline{m_0(\tfrac{1}{2}y + \tfrac{1}{2})}h(y/2). \qquad \square$$

Before turning to smoothness questions, we insert as Theorem 10.41 the hard direction of the **Paley–Wiener Theorem**. This result was used at a critical point in the proof of Proposition 10.38. In the course of proving Theorem 10.41, we shall make use of the **Phragmén–Lindelöf Theorem**, which is a variant of the Three Lines Theorem (*Basic*, Lemma 9.19B, p. 471).

Lemma 10.40 (Phragmén–Lindelöf Theorem). Let f be a function that is analytic in an open neighborhood of the closure $\overline{D}$ of an open sector D of $\mathbb{C}$ with angular opening $< \pi$. Suppose that $|f(z)| \le C$ on the boundary of D and that f satisfies a growth estimate $|f(z)| \le C'e^{M|z|}$ throughout $\overline{D}$. Then $|f(z)| \le C$ throughout $\overline{D}$.

PROOF. Without loss of generality, we may assume that D is positioned symmetrically about the positive x axis with vertex at the origin and with angle ψ on each side of the axis. Since $\psi < \frac{\pi}{2}$ by assumption, we can choose and fix $\alpha > 1$ with $\alpha\psi < \frac{\pi}{2}$. For each positive number ε, define $g_\varepsilon(z) = f(z)e^{-\varepsilon z^\alpha}$, where z^α refers to the principal value. For $z = re^{i\theta}$ in $\overline{D}$ with $|\theta| \le \psi$, we have

$$|g_\varepsilon(z)| \le C'e^{M|z|}|e^{-\varepsilon(re^{i\theta})^\alpha}| = C'e^{Mr}e^{-\varepsilon r^\alpha \cos(\alpha\theta)} \le C'e^{Mr}e^{-\varepsilon r^\alpha \cos(\alpha\psi)}.$$

Since $\alpha > 1$, this tends to 0 as r tends to infinity, and we can regard g_ε as continuous on the compact subset $\overline{D} \cup \{\infty\}$ of the extended plane. Then $|g_\varepsilon(z)|$ has an absolute maximum at some point z_ε of $\overline{D} \cup \{\infty\}$. The point z_ε cannot be ∞, and the maximum principle for analytic functions then forces z_ε to be on the boundary of $\overline{D}$. Hence $|g_\varepsilon(z_\varepsilon)| \le C$. Consequently $|g_\varepsilon(z)| \le C$ and $|f(z)| \le C|e^{-\varepsilon z^\alpha}|$ everywhere on $\overline{D}$. Passing to the limit as ε decreases to 0, we see that $|f(z)| \le C$ everywhere on $\overline{D}$. $\square$

Theorem 10.41 (Paley–Wiener Theorem). If h is an L^2 function on $\mathbb{R}$ that extends to an entire function on $\mathbb{C}$ of exponential type, then the Fourier transform of h has compact support.

REMARK. In our application within the proof of Proposition 10.38, we needed $\varphi = \mathcal{F}^{-1}h$ to be of compact support, and the assertion here is about $\mathcal{F}h$. However, the question of compact support comes to the same thing for $\mathcal{F}h$ and $\mathcal{F}^{-1}h$, since $(\mathcal{F}^{-1}h)(\xi) = (\mathcal{F}h)(-\xi)$.

PROOF. We regard h as an entire function on $\mathbb{C}$ with $|h(z)| \leq Ae^{2\pi M|z|}$ for all z. Allusions to the Fourier transform are to the restriction of h to $\mathbb{R}$. For most of the proof, we shall assume that h is bounded on $\mathbb{R}$, say with $|h(x)| \leq B$. Toward the end of the proof, we shall show how to drop this assumption.

We shall prove that $(\mathcal{F}h)(\xi) = 0$ a.e. for $|\xi| > M$. We handle $\xi > M$ and $\xi < -M$ separately. First suppose that $\xi > M$. For $\varepsilon > 0$, we introduce $h_\varepsilon(z) = h(z)/(1 + i\varepsilon z)^2$. This is analytic in a neighborhood of the lower half plane $\operatorname{Im} z \leq 0$, and its restriction to $\mathbb{R}$ is in $L^1(\mathbb{R})$, being the product of two functions in $L^2(\mathbb{R})$. We shall show that its Fourier transform $\widehat{h_\varepsilon}(\xi)$ vanishes for $\xi > M$. We are going to move the contour of integration in the integral $\widehat{h_\varepsilon}(\xi) = \int_{-\infty}^{\infty} h_\varepsilon(x)e^{-2\pi i\xi x}\,dx$.

Let $b > 0$ be arbitrary. Form the rectangle in $\mathbb{C}$ whose top side extends from $-R$ to R on the real axis and whose bottom side extends from $-R - ib$ to $-R + ib$. Orient the rectangle clockwise. The total integral over the rectangle of $h_\varepsilon(z)e^{-2\pi i\xi z}$ is 0 by the Cauchy Integral Theorem, and we are going to show that in the limit $R \to \infty$ and then $b \to \infty$, the other three sides each contribute 0. Then the conclusion will be that the contribution from the top side is 0, i.e., that $\widehat{h_\varepsilon}(\xi) = 0$.

To estimate the contributions from the vertical sides and bottom of the rectangle, we shall apply Lemma 10.40 (the Phragmén–Lindelöf Theorem) twice to the analytic function $h(z)e^{-2\pi iMz}$ in the lower half plane. Consider the quadrant with $\operatorname{Re} z \geq 0$ and $\operatorname{Im} z \leq 0$. On the positive real axis we have

$$|h(x)e^{-2\pi iMx}| = |h(x)| \leq B,$$

while on the negative imaginary axis $\{-iy \text{ with } y \geq 0\}$ we have

$$|h(-iy)e^{-2\pi iM(-iy)}| \leq Ae^{2\pi M|-iy|}e^{-2\pi My} = A.$$

On the whole quadrant where $\operatorname{Re} z \geq 0$ and $\operatorname{Im} z \leq 0$, we have

$$|h(z)||e^{-2\pi iMz}| = |h(z)|e^{2\pi M|\operatorname{Im} z|} \leq Ae^{2\pi M|z|}e^{2\pi M|\operatorname{Im} z|} \leq Ae^{4\pi M|z|}.$$

Thus the lemma applies with $C = \max\{A, B\}$ and gives the bound

$$|h(z)e^{-2\pi iMz}| \leq C \qquad (*)$$

everywhere on the quadrant. We can argue similarly with the quadrant for which $\operatorname{Re} z \le 0$ and $\operatorname{Im} z \le 0$. The estimates are the same, and Lemma 10.40 yields the bound $(*)$ whenever $\operatorname{Re} z \le 0$ and $\operatorname{Im} z \le 0$. Consequently

$$|h(x+iy)| \le C|e^{2\pi i M(x+iy)}| = Ce^{2\pi M|y|} \qquad (**)$$

whenever $y \le 0$.

To estimate the contributions from the vertical sides, where $|x| = R$, we apply $(**)$ and make use of the bound

$$|h_\varepsilon(z)| \le |h(z)|(1+\varepsilon^2|z|^2)^{-1} \le Ce^{2\pi M|y|}(1+\varepsilon^2 R^2)^{-1} \le C\varepsilon^{-2}R^{-2}e^{2\pi Mb}.$$

The integral over the right vertical side is

$$\le \int_0^{-b} |h_\varepsilon(R+it)||e^{-2\pi i\xi(R+it)}|\,dt \le C\varepsilon^{-2}R^{-2}e^{2\pi Mb}\int_0^{-b} e^{2\pi\xi t}\,dt.$$

Since we are holding b fixed as we let R tend to infinity, this term tends to 0. Similarly the integral over the left vertical side tends to 0.

After the passage to the limit in R and the application of the Cauchy Integral Theorem, we see that $\widehat{h_\varepsilon}(\xi)$ equals the integral over the bottom side of the rectangle, i.e., that

$$\widehat{h_\varepsilon}(\xi) = \int_{-\infty}^{\infty} h_\varepsilon(x-ib)e^{-2\pi i\xi(x-ib)}\,dx.$$

On the right side we use the bound $(**)$ and the estimate

$$|h_\varepsilon(z)| = |h(z)|(1+\varepsilon^2|z|^2)^{-1} \le |h(z)|(1+\varepsilon^2x^2)^{-1}$$

to see that

$$|\widehat{h_\varepsilon}(\xi)| \le \int_{-\infty}^{\infty}(1+\varepsilon^2x^2)^{-1}Ce^{2\pi Mb}e^{-2\pi\xi b}\,dx = ce^{-2\pi b(\xi-M)}.$$

As b tends to infinity, the right side tends to 0 because $\xi > M$. Thus $\widehat{h_\varepsilon}(\xi) = 0$.

Still with $\xi > M$, we shall let ε tend to 0. Let g be any function in $C^\infty_{\text{com}}(\mathbb{R})$ with support in the interval $(M, +\infty)$. Then the multiplication formula gives $0 = \int_{\mathbb{R}} \widehat{h_\varepsilon} g\,dx = \int_{\mathbb{R}} h_\varepsilon \widehat{g}\,dx$, i.e.,

$$0 = \int_{\mathbb{R}} \frac{h(x)\widehat{g}(x)}{(1+i\varepsilon x)^2}\,dx$$

for every $\varepsilon > 0$. Here h and $\widehat{g}$ are in $L^2(\mathbb{R})$ and also $|(1+i\varepsilon x)^{-2}| \le 1$. Thus we have dominated convergence as ε tends to 0, and we obtain

$$0 = \int_{\mathbb{R}} h(x)\widehat{g}(x)\,dx = \int_{\mathbb{R}} (\mathcal{F}h)(\xi)g(\xi)\,d\xi.$$

Since g is any smooth function supported in $(M,+\infty)$, we see that $(\mathcal{F}h)(\xi)=0$ a.e. for $\xi>M$.

In similar fashion we argue for $\xi<-M$ by using approximating functions $h_\varepsilon(z)=h(z)/(1-i\varepsilon z)^2$ and working in the upper half plane. Lemma 10.40 is to be applied to the function $h(z)e^{2\pi iMz}$ in the two quadrants of the upper half plane. The new version of estimate $(**)$ is that $|h(x+iy)|\le Ce^{2\pi M|y|}$ whenever $y\ge 0$. The result is that $(\mathcal{F}h)(\xi)=0$ a.e. for $\xi<-M$.

This completes the proof under the assumption that h is bounded. To see that we can drop this assumption, let $u\ge 0$ be in $C_{\text{com}}(\mathbb{R})$, and form the convolution $h_u(z)=\int_{\mathbb{R}}h(z-t)u(t)\,dt$. The function h_u is entire, and it has

$$\begin{aligned}|h_u(z)|&=\Big|\int_{\mathbb{R}}h(z-t)u(t)\,dt\Big|\le\int_{\mathbb{R}}|h(z-t)|u(t)\,dt\\&\le A\int_{\mathbb{R}}e^{2\pi M|z-t|}u(t)\,dt\le Ae^{2\pi M|z|}\int_{\mathbb{R}}e^{2\pi M|t|}u(t)\,dt=A'e^{2\pi M|z|}.\end{aligned}$$

The argument in the special case applies to h_u and shows that $\mathcal{F}h_u$ vanishes a.e. for $|\xi|>M$. If we replace $u(x)$ by $u_\varepsilon(x)=\varepsilon^{-1}u(\varepsilon^{-1}x)$ and let ε decrease to 0, then $\lim h_{u_\varepsilon}=h$ in $L^2(\mathbb{R})$, and so $\lim\mathcal{F}h_{u_\varepsilon}=\mathcal{F}h$ in $L^2(\mathbb{R})$. Since $\mathcal{F}h_{u_\varepsilon}$ vanishes a.e. for $|\xi|>M$, so does $\mathcal{F}h$. □

10. Smoothness Questions

The Daubechies wavelet of order N that was constructed in Theorem 10.39 has compact support, but we have not yet proved that it and its scaling function are continuous if $N\ge 2$. (For $N=1$, the resulting wavelet is the Haar wavelet, which is certainly not continuous.)

We shall get at this continuity in this section. As a general principle, the more rapidly the Fourier transform of a function f decays at infinity, the smoother that f is. An explanation in simple terms is that under certain technical assumptions given in Proposition 8.1 of *Basic*, the Fourier transform of $\frac{d}{dx}f(x)$ is $2\pi iyf(y)$. Thus the starting point for our study is an investigation of the rapidity of decay at infinity of $\mathcal{F}\varphi$, φ being the scaling function of the Daubechies wavelet of order N.

Lemma 10.42. The binomial coefficient $\binom{2n}{n}$ satisfies $\binom{2n}{n}\le\dfrac{4^n}{\sqrt{\pi n}}$ for all $n\ge 1$.

REMARK. In fact, the ratio of the two sides of the inequality tends to 1, as is easily seen by examining the proof of the lemma closely.

PROOF. Because of the orthogonality of the exponentials on $[-\pi,\pi]$, the 0^{th} Fourier coefficient of $\cos^{2n} x$ is

$$\begin{aligned}\frac{1}{2\pi}\int_{-\pi}^{\pi}\cos^{2n}x\,dx &= \frac{2^{-2n}}{2\pi}\int_{-\pi}^{\pi}(e^{ix}+e^{-ix})^{2n}\,dx\\ &= \frac{2^{-2n}}{2\pi}\int_{-\pi}^{\pi}\sum_{k=0}^{2n}\binom{2n}{k}e^{ikx}e^{i(2n-k)x}\,dx = 2^{-2n}\binom{2n}{n}.\end{aligned}$$

Since $\cos x$ is even,

$$\binom{2n}{n} = \frac{4^n}{\pi}\int_{-\pi/2}^{\pi/2}\cos^{2n}x\,dx. \qquad (*)$$

Consider the function $f(x) = \log\cos x + \frac{1}{2}x^2$ on $[-\pi/2, \pi/2]$. It has $f(0) = f'(0) = 0$ and $f''(x) = -\sec^2 x + 1$. Therefore it has an absolute maximum at $x = 0$, and we obtain the inequality

$$\log\cos x \le -\tfrac{1}{2}x^2 \qquad \text{for } -\tfrac{\pi}{2} \le x \le \tfrac{\pi}{2}.$$

Exponentiating and raising both sides to the $(2n)^{\text{th}}$ power gives

$$\cos^{2n}x \le e^{-nx^2} \qquad \text{for } -\tfrac{\pi}{2} \le x \le \tfrac{\pi}{2}.$$

Thus $(*)$ is

$$\le \frac{4^n}{\pi}\int_{-\pi/2}^{\pi/2}e^{-nx^2}\,dx \le \frac{4^n}{\pi}\int_{-\infty}^{\infty}e^{-nx^2}\,dx = \frac{4^n}{\pi}\sqrt{\frac{\pi}{n}}\int_{-\infty}^{\infty}e^{-\pi y^2}\,dy = \frac{4^n}{\sqrt{\pi n}}. \qquad \square$$

Proposition 10.43. For every $N \ge 1$ and every integer $j \ge 1$, the scaling function φ of the Daubechies wavelet of order N satisfies an estimate

$$|(\mathcal{F}\varphi)(y)|^2 \le \Big(\frac{2^{N-1}}{\sqrt{\pi N}}\Big)(\sqrt{4\pi N}\,)^{-j} \qquad \text{for } 2^{j-1} \le |y| \le 2^j.$$

Consequently

$$\int_{2^{j-1}\le|y|\le 2^j}|(\mathcal{F}\varphi)(y)|\,dy \le \Big(\frac{2^{N-1}}{\sqrt{\pi N}}\Big)^{1/2}(\pi N/4)^{-j/4}.$$

PROOF. For the moment we allow y to be an arbitrary real number. Theorem 10.39 says that

$$(\mathcal{F}\varphi)(y) = \prod_{j=1}^{\infty}m_0(2^{-j}y),$$

where $m_0(y) = e^{-\pi i N y}(\cos \pi y)^N \mathcal{L}(y)$ and $|\mathcal{L}(y)|^2 = P_N(\sin^2 \pi y)$. Moreover, Lemma 10.35c gives $0 \le P_N(\sin^2 \pi y) \le \binom{2N-1}{N}$. Thus

$$\begin{aligned}
|(\mathcal{F}\varphi)(y)|^2 &= \prod_{k=1}^{\infty} |m_0(2^{-k} y)|^2 \\
&\le \prod_{k=1}^{j+1} |m_0(2^{-k} y)|^2 && \text{since } |m_0(y)| \le 1 \\
&= \prod_{k=1}^{j+1} (\cos 2^{-k}\pi y)^{2N} P_N(\sin^2 2^{-k}\pi y) \\
&\le \prod_{k=1}^{j+1} \Big((\cos 2^{-k}\pi y)^{2N} \tbinom{2N-1}{N}\Big) && \text{by Lemma 10.35c} \\
&= \Big(\prod_{k=1}^{j+1} \cos 2^{-k}\pi y\Big)^{2N} \tbinom{2N-1}{N}^{j+1} \\
&= \Big(\prod_{k=1}^{j+1} \cos 2^{-k}\pi y\Big)^{2N} \Big(\tfrac{N}{2N}\tbinom{2N}{N}\Big)^{j+1} \\
&\le \Big(\prod_{k=1}^{j+1} \cos 2^{-k}\pi y\Big)^{2N} \Big(\tfrac{1}{2}\tfrac{4^N}{\sqrt{\pi N}}\Big)^{j+1} && \text{by Lemma 10.42.} \qquad (*)
\end{aligned}$$

To handle the product of powers of cosine, we use the same trick as in Remark 3 with Theorem 10.39, namely repeated application of the identity $\sin 2\theta = 2\sin\theta\cos\theta$:

$$\begin{aligned}
2^{j+1}\sin(2^{-(j+1)}\pi y)\prod_{k=1}^{j+1}\cos(2^{-k}\pi y) &= 2^j \sin(2^{-j}\pi y)\prod_{k=1}^{j}\cos(2^{-k}\pi y) \\
&= \cdots = \sin \pi y.
\end{aligned}$$

Thus

$$\prod_{k=1}^{j+1} \cos 2^{-k}\pi y = \frac{\sin \pi y}{2^{j+1}\sin 2^{-(j+1)}\pi y},$$

and $(*)$ is

$$\le \Big(\frac{\sin \pi y}{2^{j+1}\sin 2^{-(j+1)}\pi y}\Big)^{2N}\Big(\tfrac{1}{2}\tfrac{4^N}{\sqrt{\pi N}}\Big)^{j+1}. \qquad (**)$$

We now bring in the constraint $2^{j-1} \le |y| \le 2^j$ in order to give an upper bound for the first factor. In the numerator we use $|\sin \pi y| \le 1$. For the denominator, the absolute value of the argument of the sine lies between $\pi/4$ and $\pi/2$. Thus $|\sin 2^{-(j+1)}\pi y| \ge \sin \pi/4 = 2^{-1/2}$, and $(**)$ is

$$\le 4^{-jN} 2^{-(1/2)(2N)}\Big(\tfrac{1}{2}\tfrac{4^N}{\sqrt{\pi N}}\Big)^{j+1} = 2^{-N}\Big(\tfrac{1}{2}\tfrac{4^N}{\sqrt{\pi N}}\Big)(2\sqrt{\pi N}\,)^{-j} = \Big(\tfrac{2^{N-1}}{\sqrt{\pi N}}\Big)(2\sqrt{\pi N}\,)^{-j}.$$

This proves the upper bound for $|(\mathcal{F}\varphi)(y)|^2$.

An upper bound for the integral is the product of the square root of this bound by the measure 2^j of the set of integration. The result follows. □

As soon as $N\pi > 4$, namely as soon as $N \geq 2$, the sum over $j \geq 1$ of the integrals in Proposition 10.43 is finite, and it follows that $\mathcal{F}\varphi$ is integrable. Consequently the function $\varphi = \mathcal{F}^{-1}\mathcal{F}\varphi$, after adjustment on a set of measure 0, is a bounded continuous function vanishing at infinity. We already knew that it had compact support (and hence vanishes at infinity), and the new fact is that it is continuous. Since in the case of compact support the wavelet is a finite linear combination of translates of dilates of the scaling function, the wavelet itself is continuous. Let us state that result as a corollary.

Corollary 10.44. For every $N \geq 2$, the scaling function φ of the Daubechies wavelet of order N is continuous, and so is the wavelet ψ itself.

The pointwise estimate in Proposition 10.43 gives more, however. One way of proceeding is to observe as a further corollary of Proposition 10.41 that $|(\mathcal{F}\varphi)(y)|^2(1+|y|^2)^s$ is integrable for specific values of s. In the terminology of Problems 8–12 for Chapter III, f is in certain spaces $H^s(\mathbb{R})$ of "Bessel potentials." We elaborate on this approach in Problems 18–19 at the end of the present chapter. But for now we shall proceed somewhat differently, so as to avoid invoking results from Chapter III.

If $0 < \alpha < 1$, we say that a continuous function f on $\mathbb{R}$ satisfies a **Hölder condition** with exponent α if there is a constant C such that $|f(x+h) - f(x)| \leq C|h|^\alpha$ whenever x and h are in $\mathbb{R}$ with $|h| \leq 1$.

Proposition 10.45. Fix $\alpha > 0$. Suppose that

(i) f is a compactly supported member of $L^2(\mathbb{R})$,
(ii) $\mathcal{F}f$ is integrable and f has been adjusted on a set of measure 0 so as to be continuous, and
(iii) there is a constant C with

$$\int_{2^{j-1}\leq|y|\leq 2^j} |(\mathcal{F}f)(y)|\,dy \leq C2^{-\alpha j} \quad \text{for every integer } j \geq 0.$$

If $0 < \alpha < 1$, then f satisfies a Hölder condition with exponent α. If the inequality $n < \alpha < n+1$ holds for some positive integer n, then f is of class C^n, and its n^{th} derivative $f^{(n)}$ is in $L^2(\mathbb{R})$ and satisfies a Hölder condition with exponent $\alpha - n$.

REMARK. In our case we define α so that $2^{-\alpha} = (\pi N/4)^{-1/4}$, and Proposition 10.43 establishes the estimate in the hypothesis of Proposition 10.45. In other words, α is any number $\leq \frac{1}{4}\log_2(\pi N/4)$.

PROOF. First suppose that $0 < \alpha < 1$. Under the hypotheses on f, we can write f pointwise as the inverse Fourier transform of $\mathcal{F}f$, and then

$$f(x+h) - f(x) = \int_{\mathbb{R}} e^{2\pi ixy}(e^{2\pi ihy} - 1)(\mathcal{F}f)(y)\,dy,$$

the integrand being integrable. We are to estimate

$$|f(x+h) - f(x)| \le \int_{\mathbb{R}} |e^{2\pi ihy} - 1||(\mathcal{F}f)(y)|\,dy \qquad (*)$$

With $|h| \le 1$, we define the index $j = j(h) \ge 0$ so that $2^{-(j+1)} \le |h| < 2^{-j}$, and then we break the region of integration in $(*)$ into three sets S_1, S_2, and S_3 as follows: $S_1 = S_1(h)$ where $0 \le |y| < \frac{1}{2}$, $S_2 = S_2(h)$ where $\frac{1}{2} \le |y| < 2^{j(h)}$, and $S_3 = S_3(h)$ where $2^{j(h)} \le |y|$.

For the integral over S_1, we use the inequality $|e^{i\theta} - 1| \le |\theta|$, which is valid for all real θ, to write

$$\int_{S_1} |e^{2\pi ihy} - 1|\,|(\mathcal{F}f)(y)|\,dy \le 2\pi|h| \int_{S_1} |y|\,|(\mathcal{F}f)(y)|\,dy \le \pi|h|^\alpha \|\mathcal{F}f\|_1.$$

For the integral over S_2, we have similarly

$$\begin{aligned}
\int_{S_2} |e^{2\pi ihy} - 1|\,|(\mathcal{F}f)(y)|\,dy &\le 2\pi|h| \int_{S_2} |y|\,|(\mathcal{F}f)(y)|\,dy \\
&= 2\pi|h| \sum_{k=0}^{j(h)} \int_{2^{k-1} \le |y| < 2^k} |y|\,|(\mathcal{F}f)(y)|\,dy \\
&\le 2\pi|h| \sum_{k=0}^{j(h)} 2^k \int_{2^{k-1} \le |y| < 2^k} |(\mathcal{F}f)(y)|\,dy \\
&\le 2\pi|h| \sum_{k=0}^{j(h)} 2^k C 2^{-\alpha k} \\
&= 2\pi C|h| \sum_{k=0}^{j(h)} 2^{(1-\alpha)k} \\
&\le 2\pi C|h|(2^{1-\alpha} - 1)^{-1} 2^{(1-\alpha)(j(h)+1)}.
\end{aligned}$$

Since $2^{-(j(h)+1)} \le |h| \le 2^{-j(h)}$, we have $|h| = |h|^\alpha |h|^{1-\alpha} \le |h|^\alpha 2^{-(1-\alpha)j(h)}$, and thus

$$\int_{S_2} |e^{2\pi ihy} - 1|\,|(\mathcal{F}f)(y)|\,dy \le 2\pi C 2^{1-\alpha}(2^{1-\alpha} - 1)^{-1}|h|^\alpha.$$

For the integral over S_3, we have

$$\begin{aligned}\int_{S_3} |e^{2\pi ihy} - 1|\,|(\mathcal{F}f)(y)|\,dy &\le 2\int_{S_3} |(\mathcal{F}f)(y)|\,dy \\ &= 2\sum_{k=j(h)+1}^{\infty} \int_{2^{k-1}\le|y|\le 2^k} |(\mathcal{F}f)(y)|\,dy \\ &\le 2\sum_{k=j(h)+1}^{\infty} C2^{-\alpha k} \\ &= 2C(1-2^{-\alpha})^{-1}2^{-\alpha(j(h)+1)} \\ &\le 2C(1-2^{-\alpha})^{-1}|h|^{\alpha}.\end{aligned}$$

Combining the estimates for S_1, S_2, and S_3, we obtain

$$|f(x+h) - f(x)| \le C'|h|^{\alpha},$$

as required. This completes the proof for $0 < \alpha < 1$.

Now suppose that $\alpha > 1$. Proceeding inductively, we shall show that f is of class C^1 and that f' satisfies the hypotheses of the proposition for the index $\alpha - 1$. Then the result will follow by induction.

First of all, we have

$$\textstyle\int_{2^{j-1}\le|y|\le 2^j} |y|\,|(\mathcal{F}f)(y)|\,dy \le \int_{2^{j-1}\le|y|\le 2^j} 2^j|(\mathcal{F}f)(y)|\,dy \le C2^{-(\alpha-1)j},$$

and it follows that $|y|\,|(\mathcal{F}f)(y)|$ is integrable.

To prove[28] that f is of class C^1, we make use of the inequality

$$|h^{-1}(e^{2\pi iyh} - 1)| \le 2\pi|y| \qquad (**)$$

valid for $|h| \le 1$; this inequality follows from the inequality $|e^{i\theta} - 1| \le |\theta|$ that was used for estimating S_1. We have

$$\frac{f(x+h) - f(x)}{h} = \int_{\mathbb{R}} (\mathcal{F}f)(y)e^{2\pi iyx}\Big[\frac{e^{2\pi iyh} - 1}{h}\Big]\,dy.$$

The expression in brackets is $\le 2\pi|y|$ in absolute value by $(**)$, and we have seen that $y(\mathcal{F}f)(y)$ is integrable. Therefore we have dominated convergence, and we conclude that

$$f'(x) = \int_{\mathbb{R}} (\mathcal{F}f)(y)(2\pi iy)e^{2\pi iyx}\,dy,$$

[28] This argument is a version of the proof of Proposition 8.1g of *Basic*, but it is easier to write out the details of the argument than to show that the C^1 property follows from Proposition 8.1g.

the derivative existing. Except for an extra minus sign in the exponential, this formula exhibits f' as the Fourier transform of the integrable function $2\pi i y(\mathcal{F}f)(y)$, and f' is therefore continuous. Since f has compact support, so does f'. The Fourier inversion formula is applicable, and we obtain

$$(\mathcal{F}f')(y) = (2\pi i y)(\mathcal{F}f)(y).$$

We conclude that f' satisfies the hypotheses of the proposition for the index $\alpha - 1$, and the proof is complete. □

Corollary 10.46. Let n be a positive integer. If N is large enough so that $\frac{1}{4}\log_2(\pi N/4) > n$, then the scaling function of the Daubechies wavelet of order N is of class C^n, and the same thing is true of the wavelet itself.

PROOF. This follows immediately by combining Proposition 10.43 and Proposition 10.45. □

Qualitatively Corollary 10.46 says that if n is given, then the scaling function is of class C^n if N is sufficiently large. The growth of n as a function of N is logarithmic. For example, the corollary asks that N be at least 21 before it guarantees that the scaling function is of class C^1. Quantitatively this result does not appear to be very sharp if one takes into account the appearance of the curves in Figures 10.17a and 10.17b. In fact, Daubechies with much more work shows in Chapter 7 of her book *Ten Lectures on Wavelets* that asymptotically for large N, the scaling function for the Daubechies wavelet of order N is of class $C^{\mu n}$ with $\mu = \frac{3}{4}\frac{\log 3}{\log 2} - 1 \approx .1887$.

Anyway, Daubechies wavelets can have as many derivatives as we like. But it turns out that no such wavelet can have infinitely many. This fact will be a corollary of the following proposition.

Proposition 10.47. Suppose that ψ is a compactly supported continuous function on $\mathbb{R}$ such that the set of functions $\{2^{j/2}\psi(2^j x - k)\}_{j,k\in\mathbb{Z}}$ is an orthogonal set. If ψ is of class C^m, then $\int_{\mathbb{R}} x^l \psi(x)\,dx = 0$ for $0 \le l \le m$.

PROOF. We write $\psi^{(0)} = \psi, \psi^{(1)}, \dots, \psi^{(m)}$ for the successive derivatives, and we introduce successive integrals inductively by the formula

$$\psi^{(-l)}(x) = \int_{-\infty}^{x} \psi^{(-l+1)}(t)\,dt \qquad \text{for } 1 \le l \le m.$$

We prove by induction on l for $0 \le l \le m$ that $\int_{\mathbb{R}} x^l \psi(x)\,dx = 0$.

The case $l = 0$ makes use of the assumed orthogonality. Let $a = k/2^j$ be an arbitrary diadic rational number. We do not necessarily assume that k is an odd integer. Then

$$0 = \int_{\mathbb{R}} \psi(x)\overline{\psi(2^j x - k)}\,dx = \int_{\mathbb{R}} \psi(x)\overline{\psi(2^j(x - a))}\,dx,$$

and the change of variables $x = a + x'2^{-j}$ yields

$$0 = \int_{\mathbb{R}} \psi(a + x'2^{-j})\overline{\psi(x')}\,dx'. \tag{$*$}$$

In this expression we can let j tend to $+\infty$, since we have not assumed that k is odd. Then dominated convergence yields the equality $0 = \psi(a)\int_{\mathbb{R}} \psi(x')\,dx'$. Either $\int_{\mathbb{R}} \psi(x)\,dx = 0$ or $\psi(a) = 0$ for all a, in which case $\int_{\mathbb{R}} \psi(x)\,dx = 0$ by the assumed continuity of ψ. This completes the argument for $l = 0$. It also shows that $\psi^{(-1)}(x)$ has compact support.

Assume inductively that $0 \le l < m$ and that $\int_{\mathbb{R}} x^l \psi(x)\,dx = 0$ and that $\psi^{(0)}, \dots, \psi^{(-l-1)}$ have compact support. Integrating ($*$) by parts l times is allowable since with each integration, both functions in the integrand have compact support, and we obtain

$$0 = \int_{\mathbb{R}} \psi^{(l+1)}(a + x2^{-j})\overline{\psi^{(-l-1)}(x)}\,dx.$$

Letting j tend to $+\infty$ as above, we obtain $\psi^{(l+1)}(a)\int_{\mathbb{R}} \overline{\psi^{(-l-1)}(x)}\,dx = 0$ for all a. We cannot have $\psi^{(l+1)}(a) = 0$ for all a, since the continuity of $\psi^{(l+1)}$ would force ψ to be a polynomial and that is not the case. Thus we obtain

$$\int_{\mathbb{R}} \psi^{(-l-1)}(x)\,dx = 0. \tag{$**$}$$

This shows that $\psi^{(-l-2)}$ has compact support and completes one part of the inductive step.

We now integrate ($**$) by parts, differentiating the integrated ψ factor and integrating a complementary power of x. After one such integration we obtain $\int_{\mathbb{R}} x\psi^{(-l)}(x)\,dx = 0$. After a total of l such integrations by parts, we have $\int_{\mathbb{R}} x^l\psi^{(-1)}(x)\,dx = 0$. Finally we integrate by parts once more, obtaining $\int_{\mathbb{R}} x^{l+1}\psi(x)\,dx = 0$. This completes the induction and the proof. □

Corollary 10.48. If ψ is a compactly supported continuous function on $\mathbb{R}$ such that the set of functions $\{2^{j/2}\psi(2^j x - k)\}_{j,k\in\mathbb{Z}}$ is an orthogonal set, then ψ cannot be of class C^∞ unless $\psi = 0$.

PROOF. If ψ is a compactly supported C^∞ function on $\mathbb{R}$ such that the set of functions $\{2^{j/2}\psi(2^j x - k)\}_{j,k\in\mathbb{Z}}$ is an orthogonal set, then $\int_{\mathbb{R}} x^m\psi(x)\,dx = 0$ for all $m \ge 0$, according to Proposition 10.47, and therefore $\int_{\mathbb{R}} p(x)\psi(x)\,dx = 0$ for every polynomial p. By the Weierstrass Approximation Theorem (Theorem 1.52 of *Basic*), choose a sequence $\{p_n\}$ of polynomials converging uniformly to $\overline{\psi}$ on the support of ψ. Then $\{p_n\psi\}$ tends uniformly to $|\psi|^2$ on the support of ψ, and it follows that $\int_{\mathbb{R}} |\psi(x)|^2\,dx = 0$. Therefore $\psi = 0$. □

It is time to take stock of where we are. We have constructed a number of families of wavelets, often by applying some general theory, and we have seen that each of these families has certain properties, some desirable and some undesirable. We summarize this information in the table in Figure 10.18. In the table, ψ has the same decay and smoothness qualitatively as φ does. The expression "jump" points to the presence of a jump discontinuity, while "compact supp." means that the indicated function has compact support and "Schwartz" means that the function is in the Schwartz class $\mathcal{S}$. A function on $\mathbb{R}$ is **real analytic** if it is the restriction to $\mathbb{R}$ of an analytic function on an open subset of $\mathbb{C}$ that contains $\mathbb{R}$.

Family	φ decay	$\mathcal{F}\varphi$ decay	φ smoothness	$\mathcal{F}\varphi$ smoothness
Haar	compact supp.	$\leq c/\lvert y\rvert$	jump	real analytic
Shannon	$\leq C/\lvert x\rvert$	compact supp.	real analytic	jump
Meyer, index m	$\leq C\lvert x\rvert^{m+2}$	compact supp.	real analytic	C^m
Meyer, index ∞	Schwartz	compact supp.	real analytic	C^∞
Battle–LeMarié, deg. 1	exponential	$\leq c/\lvert y\rvert^2$	C^0	real analytic
Battle–LeMarié, deg. m	exponential	$\leq c\lvert y\rvert^{m+1}$	C^{m-1}	real analytic
Daubechies, order 1	compact supp.	$\leq c/\lvert y\rvert$	jump	real analytic
Daubechies, order N	compact supp.	$\leq c/\lvert y\rvert^{n(N)}$	$C^{n'(N)}$	real analytic

FIGURE 10.18. Summary of properties of constructed wavelets.

11. A Quick Introduction to Applications

If it were not for the usefulness of wavelets in applications, the subject of wavelets might have remained something known only to experts in one corner of mathematics. But there are by now applications in many areas of science and engineering, even in areas of social science, and the subject cannot be ignored by a well educated mathematician. Not only are there applications, but also the applications have in many cases driven the theory. In this section we shall list some areas where wavelets have been useful, and we shall give a few details for some of them.

As we often saw throughout the theoretical development, the hypotheses on wavelets that we were accustomed to were not exactly the hypotheses that were needed for new steps. This phenomenon persists with applications, as we shall see. Let us distinguish two situations, corresponding to one and two dimensions.

In one dimension the objective is to analyze some function f in $L^2(\mathbb{R})$. It will be convenient to think of the domain variable as representing time and f as being some kind of signal. The orthogonal wavelets that we have studied are in many practical cases a suitable tool for the analysis of f. We fix a wavelet ψ that comes from a scaling function φ, and just as we did with the Haar wavelet, we introduce the corresponding **discrete wavelet transform**, which carries f to the system $\{(f, \psi_{j,k})\}_{j,k\in\mathbb{Z}}$ of inner products of f with the members of the orthonormal basis $\{\psi_{j,k}\}$. We can think of these inner products as the wavelet coefficients. As with the Haar wavelet, we are really interested in a one-sided **wavelet expansion** of the form

$$f(x) = \sum_{k\in\mathbb{Z}} \big(\textstyle\int_{\mathbb{R}} f(y)\overline{\varphi_{0,k}(y)}\,dy\big)\varphi_{0,k}(x) + \sum_{j=0}^{\infty}\sum_{k\in\mathbb{Z}} \big(\textstyle\int_{\mathbb{R}} f(y)\overline{\psi_{j,k}(y)}\,dy\big)\psi_{j,k}(x).$$

In effect this expansion groups all the contributions from the $\psi_{j,k}$ with $j < 0$ into the first term. This expansion represents a process of looking at f with an infinite system of finer and finer resolutions. The first term gives the result of using an initial resolution corresponding to $j = 0$. New terms in the approximating expansion represent taking into account higher and higher resolutions, thus providing better and better knowledge of what is happening in the time domain. In short, the new terms represent the detail in the signal. Changing k represents changing the center of the interval of time on which we are concentrating; increasing j represents increasing the resolution. Experts in signal processing think of the process as one of passing the signal through a sequence of filters. Barbara Hubbard explains matters in the following way in her book:[29]

> Mathematicians classify functions in all kinds of ways. The viewpoint of signal processors is simpler: to them, a function is either a signal to be analyzed or a filter with which to analyze a function. A classical filter is an electric circuit with one wire that carries a signal in and another wire that carries a signal out. But a filter can also be a function (or, if it is digital, a sequence of numbers). The effect of a filter, whether physical or abstract, is easier to understand in Fourier space: the Fourier transform of the signal is multiplied by the Fourier transform of the filter, letting certain frequencies pass through while blocking others.
>
> If, for example, the Fourier transform of the filter is almost 1 near zero, and almost 0 everywhere else . . . , the signal's low frequencies will survive this multiplication by 1, but the high frequencies will be essentially eliminated. This is a low-pass filter. The result in "physical" space is to smooth the signal: the small variations given by the high frequencies disappear, leaving the general tendency. . . .
>
> Mallat [the person who formally introduced multiresolution analyses] realized that one can incorporate wavelets into a system that uses a cascade of filters to decompose a signal. Each resolution has its own pair of filters: a low-pass filter associated with the scaling

[29] *The World According to Wavelets*, second edition, pp. 166–167.

> function, giving an overall picture of the signal, and a high-pass filter associated with the wavelet, letting through only the high frequencies associated with the details. The two filters complement each other; that which one blocks, the other lets through. (Of course "high" or "low" frequencies are relative. The low frequencies encoded by the low-pass filter at a fine resolution may be higher than the high frequencies encoded by the high-pass filter at a coarse resolution.)

In quantitative terms the process described by Hubbard is as follows: For each index j, let P_j be the orthogonal projection of $L^2(\mathbb{R})$ on V_j, and let Q_j be the orthogonal projection on W_j. Then we have $P_J f = P_{J-1} f + Q_{J-1} f$ and $P_{J-1} f = P_{J-2} f + Q_{J-2} f$ and so on, with the result that

$$P_J f = P_0 f + Q_{J-1} f + Q_{J-2} f + \cdots + Q_1 f + Q_0 f.$$

We shall come back to this decomposition shortly when we discuss the discrete wavelet transform in more detail.

For many purposes one wants to chop off the expansion at some point by dropping small terms. This is already what we did above in passing from f to $P_J f$. The same kind of chopping would be the simplest thing to do if one were trying to compress the signal or counteract the effect of noise. Meyer[30] says the following about this process:

> The most astonishing result we obtain will be the remarkable fact that "full" wavelet series (those having plenty of non-zero coefficients) represent really pathological functions, whereas "normal" functions have "sparse" or "lacunary" wavelet series. On the other hand, Fourier series of the usual functions are "full," whereas lacunary Fourier series represent pathological functions.
>
> This phenomenon has a simple explanation. Analysis by wavelets is a local Fourier analysis which takes place at every scale. It has the advantage of being concentrated near the singular support of the function analyzed. In other words, away from the singular support, the function analyzed is infinitely differentiable and the corresponding wavelet coefficients are negligible.

The book by Burris, Gopinath, and Guo[31] has some interesting graphs giving an example of how this works in practice. The authors have created a one-dimensional signal out of the Houston skyline, and they show how the signal is

[30] *Wavelets and Operators*, p. 113. Meyer makes this statement in a specific context, which is a little different from ours, but that distinction will not deter us.

[31] See the Selected References.

decomposed when the discrete wavelet transform is applied using the Daubechies wavelet[32] of order 4.

Other applications use this and other wavelets. As a rule, authors tend not to explain how they came to use one particular wavelet rather than another. The wavelets in use tend to be the Meyer wavelets with index ≤ 3, the Battle–Lamarié wavelets of degree ≤ 3, and the Daubechies wavelets of order ≤ 6. Occasionally an author will use some other kind of wavelet for some particular purpose. Some of the applications that have been discussed in print are the following:

(a) Automatic analysis of an electrocardiogram. This is discussed in the book by Louis, Maass, and Rieder, pp. 232ff. The objective is to automate the obvious diagnoses of irregular heartbeats, leaving a more careful analysis to a human being. The questions are whether the rhythm of the cardiac valve is in synchronization with that of the heart muscle and whether the heart muscle relaxes between beats. The Daubechies wavelet of order 2 is used for this purpose.

(b) Speech storage. Speech takes a great deal of computer memory to store. Imagine that one wanted to store the content of all telephone calls worldwide that might be of national security interest. One wants a way of compressing the speech without making it unintelligible.

(c) Music storage. Music takes a great deal of space to store but for different reasons from speech. One has to handle a greater range of frequencies without losing subtle techniques of the performers.

(d) Speech recognition. The idea is to use wavelet analysis to identify phonemes from a signal representing speech.

(e) Hearing aids for understanding speech. Modern hearing aids are capable of doing a complicated nonlinear analysis and synthesis of speech, amplifying part of a signal and suppressing another part, in order to make speech more recognizable to a patient using a hearing aid. Simple amplification is not enough. The usual difficulty that a patient has in hearing is an inability to distinguish among certain consonants, and the processing of the signal is supposed to emphasize those features that the patient needs in order to recognize what is being said, all the while eliminating background noise. Smoothing of the signal is definitely not what is called for; smoothed speech sounds largely like vowels, and the difficulty in recognizing consonants is aggravated if speech is smoothed.[33]

(f) Applications in economics. Wavelet analysis can reveal relationships between economic variables and indicate how those relationships evolve with time.

---

[32] This is in Section 2.7 of the book. There are two standard ways of naming the Daubechies wavelets—by the order N and by the number of nonzero coefficients $2N$ in the scaling equation.

[33] There is a problem with being too specific about how particular hearing aids work in that a certain amount of the information is proprietary and therefore unpublished and unavailable.

(g) Applications in finance. Economic variables affect markets according to various time scales, short term and medium term, for example. Some are leading indicators, and some are lag indicators. In principle wavelets decompose time series data into different scales and can reveal relationships that are not obvious in the aggregate data.

Let us come back to the discrete wavelet transform. Calculating a one-sided wavelet expansion as at the beginning of this section looks as if it involves a great many integrals. However, these integrals are related to one another by the scaling equation and the wavelet equation, and matters are not so complicated. As indicated with the Hubbard quotation, we are to think of the goal as forming $P_J f$ with J chosen large enough so that $\|f - P_J f\|_2$ is small.[34] Write $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$ for each integer j and k. For fixed j, these functions form an orthonormal basis of V_j, and we write $c_{j,k}$ for the coefficient in the expansion of f in this basis:

$$c_{j,k} = (f, \varphi_{j,k}) = \int_{\mathbb{R}} f\,\overline{\varphi_{j,k}}\,dx.$$

We write $\mathbf{c}_j$ for the square-summable sequence $\{c_{j,k}\}_{k\in\mathbb{Z}}$.

To be systematic, we shall introduce an algorithm of "decomposition" of $\mathbf{c}_j$ and an algorithm of "reconstruction" of $\mathbf{c}_j$. Each of these makes use of the two functions $m_0(y)$ and $m_1(y)$ that we have carried along ever since Theorem 10.10. The function $m_0(y)$ was introduced just before the statement of that theorem in terms of the scaling equation

$$\varphi(x) = \sum_{k=-\infty}^{\infty} a_k\sqrt{2}\,\varphi(2x-k), \quad \text{where } a_k = (\varphi, \varphi_{1,k}),$$

the definition being

$$m_0(y) = \tfrac{1}{\sqrt{2}} \sum_{k=-\infty}^{\infty} a_k e^{-2\pi i k y}.$$

The function $m_1(y) = e^{2\pi i y}\nu(y)\overline{m_0(y+\frac{1}{2})}$ was introduced just after the statement of the theorem and involved the wavelet equation

$$\psi(x) = \sum_{k=-\infty}^{\infty} b_k\sqrt{2}\,\varphi(2x-k), \quad \text{where } b_k = (\psi, \varphi_{1,k}).$$

Its formula was

$$m_1(y) = \tfrac{1}{\sqrt{2}} \sum_{k=-\infty}^{\infty} b_k e^{-2\pi i k y}.$$

In this formula the coefficients b_k are related to the coefficients a_k. The exact relationship depends on the choice of the function $\nu(y)$, but we saw, for example, that if $\nu(y) = 1$, then $b_k = (-1)^{k+1}\,\overline{a_{-k-1}}$.

[34] From a theoretical standpoint this norm can be estimated with the aid of Lemma 10.16.

The above formulas relate matters in V_1 to those in V_0. Let us see the relationship between V_j and V_{j-1}. We have

$$\begin{aligned}\varphi_{j-1,k}(x) &= 2^{(j-1)/2}\varphi(2^{j-1}x-k) = 2^{(j-1)/2}\varphi(\tfrac{1}{2}(2^jx-2k)) \\ &= 2^{(j-1)/2}\sum_{l=-\infty}^{\infty} a_l\sqrt{2}\,\varphi(2^jx-2k-l) \\ &= \sum_{l=-\infty}^{\infty} a_l 2^{j/2}\varphi(2^jx-2k-l) \\ &= \sum_{l=-\infty}^{\infty} a_l\varphi_{j,2k+l}(x)\end{aligned}$$

Similarly

$$\psi_{j-1,k}(x) = \sum_{l=-\infty}^{\infty} b_l\varphi_{j,2k+l}(x).$$

For the **decomposition algorithm**, we substitute the relationship between $\varphi_{j-1,k}$ and the $\varphi_{j,2k+l}$ into the definition of the coefficient $c_{j-1,k}$. Then we obtain

$$c_{j-1,k} = (f,\varphi_{j-1,k}) = \Big(f,\sum_{l=-\infty}^{\infty} a_l\varphi_{j,2k+l}\Big) = \sum_{l=-\infty}^{\infty}\overline{a_l}c_{j,2k+l} = \sum_{l=-\infty}^{\infty}\overline{a_{l-2k}}c_{j,l}$$

If we define coefficients $d_{j,k}$ by $d_{j,k} = (f,\psi_{j,k})$, then we similarly have

$$d_{j-1,k} = (f,\psi_{j-1,k}) = \Big(f,\sum_{l=-\infty}^{\infty} b_l\varphi_{j,2k+l}\Big) = \sum_{l=-\infty}^{\infty}\overline{b_l}c_{j,2k+l} = \sum_{l=-\infty}^{\infty}\overline{b_{l-2k}}c_{j,l}.$$

Then one step of the decomposition algorithm is the passage from the system of coefficients $\{c_{j,k}\}_{k\in\mathbb{Z}}$ to the systems $\{c_{j-1,k}\}_{k\in\mathbb{Z}}$ and $\{d_{j-1,k}\}_{k\in\mathbb{Z}}$ by the *same* operation. To make this step more transparent, let us extend our definition of $\mathbf{c}_j$ above by giving names to all of our various square-summable sequences:

$$\mathbf{a} = \{a_l\}_{l\in\mathbb{Z}} \quad\text{and}\quad \mathbf{b} = \{b_l\}_{l\in\mathbb{Z}},$$

$$\mathbf{a}^{\mathrm{tr}} = \{a_{-l}\}_{l\in\mathbb{Z}} \quad\text{and}\quad \mathbf{b}^{\mathrm{tr}} = \{b_{-l}\}_{l\in\mathbb{Z}},$$

and with j fixed,

$$\mathbf{c}_j = \{c_{j,k}\}_{k\in\mathbb{Z}} \qquad\text{and}\qquad \mathbf{d}_j = \{d_{j,k}\}_{k\in\mathbb{Z}}.$$

Then $\mathbf{c}_{j-1}$ is obtained from $\mathbf{c}_j$ by convolving with the sequence $\overline{\mathbf{a}}^{\mathrm{tr}}$ and retaining only the even-numbered entries. Similarly $\mathbf{d}_{j-1}$ is obtained from $\mathbf{c}_j$ by convolving with the sequence $\overline{\mathbf{b}}^{\text{tr}}$ and retaining only the even-numbered entries. Iterating this step, starting at $\mathbf{c}_J$, allows us to pass from the sequence $\mathbf{c}_J$ to the system of sequences $\mathbf{d}_{J-1}, \mathbf{d}_{J-2}, \dots, \mathbf{d}_0, \mathbf{c}_0$.

The **reconstruction algorithm** will pass from the system of sequences $\mathbf{d}_{J-1}$, $\mathbf{d}_{J-2}, \dots, \mathbf{d}_0, \mathbf{c}_0$ to the sequence $\mathbf{c}_J$. To understand matters, it is enough to carry out one step of the algorithm, namely to pass from $\mathbf{c}_{j-1}$ and $\mathbf{d}_{j-1}$ to $\mathbf{c}_j$. The relevant formula in terms of projections is

$$P_j f = P_{j-1} f + Q_{j-1} f,$$

which we write out as

$$\sum_{k\in\mathbb{Z}} (f, \varphi_{j,k})\varphi_{j,k} = \sum_{k\in\mathbb{Z}} (f, \varphi_{j-1,k})\varphi_{j-1,k} + \sum_{k\in\mathbb{Z}} (f, \psi_{j-1,k})\psi_{j-1,k},$$

hence as

$$\begin{aligned}\sum_{k\in\mathbb{Z}} c_{j,k}\varphi_{j,k} &= \sum_{k\in\mathbb{Z}} c_{j-1,k}\varphi_{j-1,k} + \sum_{k\in\mathbb{Z}} d_{j-1,k}\psi_{j-1,k} \\ &= \sum_{k\in\mathbb{Z}} c_{j-1,k}\Big(\sum_{l\in\mathbb{Z}} a_l\varphi_{j,2k+l}\Big) + \sum_{k\in\mathbb{Z}} d_{j-1,k}\Big(\sum_{l\in\mathbb{Z}} b_l\varphi_{j,2k+l}\Big) \\ &= \sum_{l\in\mathbb{Z}}\sum_{k\in\mathbb{Z}} (c_{j-1,k}a_{l-2k} + d_{j-1,k}b_{l-2k})\varphi_{j,l}.\end{aligned}$$

Since the functions $\varphi_{j,l}$ are an orthonormal basis of V_j, we obtain

$$c_{j,l} = \sum_{k\in\mathbb{Z}} (c_{j-1,k}a_{l-2k} + d_{j-1,k}b_{l-2k}),$$

which is a formula for recovering $\mathbf{c}_j$ from $\mathbf{c}_{j-1}$ and $\mathbf{d}_{j-1}$.

Both sums in the expression for $c_{j,k}$ can be viewed as convolutions, but they are more subtle than the ones in the decomposition algorithm. To see them as convolutions, define

$$\tilde{c}_{j-1,n} = \begin{cases} c_{j-1,n/2} & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd,} \end{cases}$$

and let $\tilde{\mathbf{c}}_{j-1} = \{\tilde{c}_{j-1,n}\}_{n\in\mathbb{Z}}$. Define $\tilde{d}_{j-1,n}$ and $\tilde{\mathbf{d}}_{j-1}$ similarly. Then $\mathbf{c}_j$ is the sum of two terms; the first is the convolution of $\tilde{\mathbf{c}}_{j-1}$ and $\mathbf{a}$, and the second is the convolution of $\tilde{\mathbf{d}}_{j-1}$ and $\mathbf{b}$.

We conclude this discussion of the one-dimensional case with some general remarks.

In the language of signal processing, the decomposition algorithm and the reconstruction algorithm are the analysis and synthesis steps of a scheme of "sub-band filtering with exact reconstruction." The point is to do some compression or other processing between the decomposition and reconstruction steps. The decomposition step draws attention to terms that are candidates for dropping without the loss of too much information, and then one drops those terms. The reconstruction step is applied to the terms that remain. Here the theory has built this idea from signal processing into a mathematical transform that isolates features of a signal that are important.

In practice a given signal will often be given by discrete pulses obtained by sampling. To make the discrete pulses into a function on $\mathbb{R}$, one defines the signal to be constant or perhaps linear over each interval between sampling points. In this situation the Shannon Sampling Theorem (Proposition 10.3) indicates that there is an upper limit to how much resolution will contain useful information. We omit the details.

Let us turn our attention to two dimensions. In two dimensions the objective is to analyze some square integrable function f on $\mathbb{R}^2$. Let us think of this function as representing a visual image. If the image is in black-and-white, then the function is scalar valued. If it is in color, then it is vector-valued. The usual thing is to make the vectors be three-dimensional. Initially the three dimensions for color can be regarded as representing the intensity of three colors; red, green, and blue are one choice. We return to this matter in a moment.

We shall concentrate on two situations:

(i) compression and storage of fingerprints, and
(ii) traditional JPEG compression of images from digital cameras and the wavelet version, JPEG 2000.

In terms of multiresolution analysis, proceeding in the setting of $\mathbb{R}^2$ is not fundamentally more complicated than proceeding in the setting of $\mathbb{R}^1$. The idea is that one can use a kind of completed tensor product to convert one-variable data into two-variable data. Thus for example, if φ and $\{V_j\}$ define a multiresolution analysis is one dimension, then the set of functions $\{x \mapsto \varphi(x-k)\}_{k\in\mathbb{Z}}$ is an orthonormal basis of V_0. For the two-dimensional setting the corresponding orthonormal basis is $\{(x, y) \mapsto \varphi(x-k)\varphi(y-l)\}_{k,l\in\mathbb{Z}}$, and one calls the space $V_0 \widetilde{\otimes} V_0$. The spaces $V_j \widetilde{\otimes} V_j$ are obtained as usual by dilation. The details will not be of concern to us.

What is of more significance is that for some reason, the human brain seems especially sensitive to visual asymmetries. When one goes through the process of decomposition, compression, and reconstruction for a visual image by the process that was just described, the use of a real-valued wavelet ψ that lacks symmetry about an axis turns out to be noticeable. As it happens, all Daubechies wavelets other than the Haar wavelet fail to be symmetric or antisymmetric about any vertical axis, and in fact the same thing is true in all real-valued compactly supported cases except in the Haar case.[35] The effect is that one has to enlarge the theory that we have presented to allow wavelets that are not orthogonal. It is enough to use "biorthogonal wavelets," which use two scaling functions that are related in a certain way. This expansion of the theory will take us too far afield, and thus we omit it.

However, we can mention the name of the best known family of biorthogonal wavelets, namely the Cohen–Daubechies–Feauveau family. The two members of the family that are most used have indices 5/3 and 9/7, and for current purposes we need not know what these indices refer to; the one with indices 5/3 is for compression with no loss of information, and the one with indices 9/7 is for compression that allows loss of information.

Now let us come to the two situations mentioned above. First let us consider fingerprints.[36] As of 1995 the database of fingerprints maintained by the Federal Bureau of Investigation contained more than 200 million records, and there was a need for computerizing these records, which were all in the form of inked impressions on cards, one card per person. Taking a full uncompressed 8-bit gray-scale digital image at a resolution of 500 dots per inch would have required too much computer storage, and a system for compressing digital images was sought that would be accurate enough to make the distinctions that one could traditionally make by hand if given enough time. A group set up to investigate the situation found that the JPEG method used in digital cameras unavoidably adds faint horizontal and vertical lines to the images in certain places if the compression ratio is at least 10 to 1, and that feature was unacceptable. After trying a number of other possibilities, the group settled on some guidelines in 1997. In principle these guidelines could be met by a number of wavelets, and initially the one put in use was the Cohen–Daubechies–Feauveau 9/7 wavelet. That choice seemed to produce reliable images even after a process that involved 15 to 1 compression.

Finally we consider JPEG and JPEG 2000 for the compression and storage of photographic and other images.[37] For each we treat black-and-white images first and then color images.

Of the two systems, JPEG is the conceptually simpler one. It uses classical Fourier analysis rather than wavelets, but it still has steps of decomposition, compression, and reconstruction. It works independently on 8-by-8 blocks of pixels, doing an analysis on each one. The transform at the decomposition stage is a version of Fourier series that avoids complex numbers. Within the classical treatment of Fourier series, a Fourier cosine expansion is obtained for functions on

[35]The result in question is Theorem 8.1.4 of the book by Daubechies in the Selected References.

[36]The article by Brislawn listed in the Selected References is a readable expository account.

[37]The two items by Austin listed in the Selected References are readable expository accounts. See the article by Rabbabi and Joshi for more detail about JPEG 2000.

$[0, \pi]$ by extending the functions to be even functions on $[-\pi, \pi]$ that are periodic of period 2π; if the Fourier series is written out with cosines and sines in place of complex exponentials, the sine terms drop out, and the desired series results. In our situation our function is defined at 8 points, and we think of extending it to be defined on 16 points so as to be even. In defining the Fourier cosine expansion, there is a choice how the 16 points are to be distributed in the given interval, and the particular choice that is made in the case of JPEG is what is sometimes called the discrete cosine transform of type II. If the given interval were $[0, \pi]$, doubled to $[-\pi, \pi]$, this particular choice of transform would visualize $[0, \pi]$ as divided into 8 equal pieces, and then the midpoint of each piece would be used. Thus the points for evaluation of a function h would be the points $\pi(2x + 1)/16$ with $0 \leq x \leq 7$, and functions on these points would be be expanded as linear combinations of the functions $\cos[\pi(2x + 1)u/16]$ with $0 \leq u \leq 7$. For our application we abbreviate the cumbersome $h(\pi(2x + 1)/16)$ as $f(x)$, and we expand $f(x)$ in terms of the functions $\cos[\pi(2x + 1)u/16]$ with $0 \leq u \leq 7$.

Using the formula $\cos A \cos B = \frac{1}{2}(\cos(A + B) + \cos(A - B))$, we readily check that the eight functions of x given by $\cos[\pi(2x + 1)u/16]$ with $0 \leq u \leq 7$ are orthogonal in the sense that

$$\sum_{x=0}^{7} \cos[\pi(2x + 1)u/16] \cos[\pi(2x + 1)u'/16] = 0 \qquad \text{for } u \neq u'.$$

Hence they are linearly independent and form a basis of an 8-dimensional vector space. As to their normalization, $\sum_{x=0}^{7} \cos^2[\pi(2x + 1)u/16]$ equals 8 if $u = 0$ and equals 4 if $1 \leq u \leq 7$. Thus the system of functions

$$\left\{\tfrac{1}{2}C(u) \cos[\pi(2x + 1)u/16]\right\}_{u=0}^{7}$$

is an orthonormal basis of our space of functions of x if $C(u)$ is defined to be 1 for $1 \leq u \leq 7$ and to be $1/\sqrt{2}$ for $u = 0$. The **discrete cosine transform** is the passage from f to the system of coefficients F of f is this basis. Specifically in terms of f we let

$$F(u) = \tfrac{1}{2}C(u) \sum_{x=0}^{7} f(x) \cos[\pi(2x + 1)u/16],$$

and we recover f from F by the formula

$$f(x) = \tfrac{1}{2} \sum_{u=0}^{7} C(u)F(u) \cos[\pi(2x + 1)u/16].$$

For the application to JPEG, the 8-by-8 block of pixels in a photographic image is a function $f(x, y)$ of two variables x and y, and we do a discrete cosine transform in each variable. The result is the function $F(u, v)$ given by

$$F(u, v) = \tfrac{1}{4}C(u)C(v) \sum_{x=0}^{7} \sum_{y=0}^{7} f(x, y) \cos[\pi(2x+1)u/16] \cos[\pi(2y+1)v/16],$$

and we recover $f(x, y)$ by the inversion formula

$$f(x, y) = \tfrac{1}{4} \sum_{u=0}^{7} \sum_{v=0}^{7} C(u)C(v)F(u, v) \cos[\pi(2x+1)u/16] \cos[\pi(2y+1)v/16].$$

The above formula for $F(u, v)$ takes care of the decomposition algorithm except for the question of its fast implementation, which will not concern us. The next step is the compression algorithm. A naive approach would be to chop off higher frequencies, but there are two problems. One is that there are not many frequencies in this analysis, and the other is that such a chopping process would introduce Gibbs phenomenon. Instead, one goes through a process that involves linear combinations of frequencies.

This consists of a step of "quantization" or rounding off, followed by an invertible packaging step that will not concern us. Part of the input to the JPEG process is a quality parameter q with $1 \le q \le 100$. This parameter is converted to another number α by the formula

$$\alpha = \begin{cases} \frac{50}{q} & \text{if } 1 \le q \le 50, \\ 2 - \frac{50}{q} & \text{if } 50 \le q \le 100. \end{cases}$$

Data on human perception leads to the definition of a matrix Q with rows and columns numbered from 0 to 7 and given by[38]

$$Q = \begin{pmatrix} 16 & 11 & 10 & 16 & 24 & 40 & 51 & 61 \\ 12 & 12 & 14 & 19 & 26 & 58 & 60 & 55 \\ 14 & 13 & 16 & 24 & 40 & 57 & 69 & 56 \\ 14 & 17 & 22 & 29 & 51 & 87 & 80 & 62 \\ 18 & 22 & 37 & 56 & 68 & 109 & 103 & 77 \\ 24 & 35 & 55 & 64 & 81 & 104 & 113 & 92 \\ 49 & 64 & 78 & 87 & 103 & 121 & 120 & 101 \\ 72 & 92 & 95 & 98 & 112 & 100 & 103 & 99 \end{pmatrix}.$$

In the quantizing step, we replace the 64 real numbers $F(u, v)$ by the 64 integers $\operatorname{round}\big(F(u, v)/(\alpha Q(u, v))\big)$, where "round" refers to the nearest integer. The

[38]The exact matrix appears in the JPEG standard as a matter of information, not as a requirement. The article by Wallace in the Selected References includes an example of what happens to some given data when this matrix is used and α equals 1.

effect of this step is to replace all of the numbers $F(u, v)$ by integers, most of which are 0. Instead of recording each zero, the algorithm records the number of zeros. It is here that space is saved. After the invertible packaging step, the data can be saved.

For the reconstruction algorithm one reverses the steps—unpackaging, unquantizing (i.e., multiplication of entries by $\alpha Q(u, v)$), and inversion of the discrete cosine transform.

For color images the main difference with the black-and-white case is that the functions in question have values that are 3-component vectors rather than real numbers. The three components could correspond to the red, green, and blue components, but it is customary to use certain linear combinations of these instead. These are called luminance Y, blue chrominance C_b, and red chrominance C_r. The relationship is given by[39]

$$\begin{pmatrix} Y \\ C_b \\ C_r \end{pmatrix} = \begin{pmatrix} 0.29900 & 0.58700 & 0.11400 \\ -0.16874 & -0.33126 & 0.50000 \\ 0.50000 & -0.41869 & -0.08131 \end{pmatrix} \begin{pmatrix} R \\ G \\ B \end{pmatrix}.$$

Then one goes through the same steps as above, using suitable matrices Q for each component. The matrix Q for the luminance is the one given above, but the matrices Q for the chrominance components are different.

The ingredients for JPEG 2000 are based on those for JPEG but are more complicated. The distinction between black-and-white as involving scalar-valued functions and color as involving vector-valued functions persists, and in the latter case one still uses luminance, blue chrominance, and red chrominance. The processing involves decomposition, compression, and reconstruction just as before, but there is also a step of preprocessing. As before, we ignore the packaging step, which is more complicated than before.

The preprocessing step consists first of a decision about how to partition the input image into rectangular and nonoverlapping tiles (except possibly for the tiles at the image borders). These can be as large as the original image itself or as small as a single pixel. The various tiles are processed independently. There are other aspects to the preprocessing, but we ignore them.

For the decomposition a discrete wavelet transform is used, the same transform in each variable. As was mentioned earlier, orthogonal wavelets are not suitable in the wavelet transform for a visual image. Instead, biorthogonal wavelets are used. These have different scaling functions and wavelets for the decomposition and reconstruction stages. Thinking in terms of the wavelet transform as given by filters, one speaks of using a pair of filter banks (one for decomposition and one for reconstruction) rather than just one. The two that are mentioned in the article by Rabbani and Joshi are the Cohen–Daubechies–Feauveau 5/3 biorthogonal

[39]An adjustment is made to force the chrominance components to be ≥ 0, but we omit this part.

wavelet[40] and the Cohen–Daubechies–Feauveau $9/7$ biorthogonal wavelet. To describe either of these, one has only to give the coefficients of the two generating functions. They are integers in the first case and are merely real numbers in the second case. The coefficients are written down explicitly in the Rabbani–Joshi article.

For the compression stage the idea is to drop low-pass terms beyond a certain threshold. This is done with the aid of a "budget" for how many bits are to be allowed and an algorithm for comparing the distortion that results from the various choices.

BIBLIOGRAPHICAL REMARKS ABOUT CHAPTER X. In Section 1, Proposition 10.1 is taken from Chapter 3 of Debnath's *Wavelet Transforms*, and Proposition 10.3 is taken from from Chapter 2 of Daubechies's *Ten Lectures on Wavelets*. Sections 2 through 10 are adapted from material in the following chapters of books listed in the Selected References:

Section 2: Pinsky, Chapter 6.
Section 3: Daubechies, Chapter 5.
Section 4: Daubechies, Chapter 5; Hernandez–Weiss, Chapter 2.
Section 5: Daubechies, Chapter 5; Pinsky, Chapter 6.
Section 6: Daubechies, Chapters 4 and 5.
Section 7: Ahlfors, Chapter 5; Chui, Chapters 1 and 4; Daubechies, Chapter 5; Meyer, Chapter 2.
Section 8: Chui, Chapter 4.
Section 9: Daubechies, Chapter 6; Debnath, Chapter 7; Pinsky, Chapter 6.
Section 10: Daubechies, Chapter 7; Pinsky, Chapter 6.

The proof of Lemma 10.42 in Section 10 is from the note by N. D. Elkies listed in the Selected References. The proofs of the Paley–Wiener Theorem and its lemma in Section 10 are based on the Junior Paper of L. B. Pierce. Section 11 draws on the above books in the Selected References, as well as the expository articles by Austin and Brislawn and the books by Burris et al. and by Louis et al.

12. Problems

1. Find all Schwartz functions f on $\mathbb{R}$ for which equality holds in the inequality of the Uncertainty Principle (Proposition 10.1). Assume that the mean values t_0 and ω_0 are 0.
2. Referring to the proof of the Shannon Sampling Theorem (Proposition 10.3), give an example of a nonzero continuous function f in $L^2(\mathbb{R})$ such that $f(k) = 0$ for every integer k and such that $\mathcal{F}f$ has compact support.

[40] Rabbani and Joshi refer to the the Cohen–Daubechies–Feauveau $5/3$ biorthogonal wavelet as the "LeGall (5,3) spline."

3. Theorem 10.10b limits the wavelets ψ that can correspond to a given scaling function φ. The form of the most general such ψ allows for a periodic function $\nu(y)$ of period 1 with $|\nu(y)| = 1$ almost everywhere.
   (a) Prove that if φ has compact support, then $\nu(y)$ is a trigonometric polynomial.
   (b) Prove that if $\nu(y)$ is a trigonometric polynomial such that $|\nu(y)| = 1$ almost everywhere, then $\nu(y) = ce^{2\pi iny}$ for some integer n and some constant c with $|c| = 1$.

4. (a) Show that if $\{V_j\}$ and φ form a multiresolution analysis, then $\varphi^\#$ with $\varphi^\#(x) = \varphi(-x)$ is a scaling function for a suitable sequence of spaces $\{V_j^\#\}$, and identify the corresponding spaces $\{V_j^\#\}$. Why does it follow that $V_j^\# = V_j$ if $\varphi^\#$ is an integer translate of φ?
   (b) Show that the scaling function $\varphi^\#$ is an integer translate of φ in the cases of the Haar wavelet, the Shannon wavelet, the Meyer wavelets, and the Battle–Lemarié wavelets.
   (c) In the case of the Daubechies wavelets of order ≥ 2, the function $\varphi^\#$ is not an integer translate of φ. Nevertheless, show that $\varphi^\#$ arises in some way from the same kind of construction.

5. Let K be the interval $\big[-\frac{2}{3}, \frac{1}{3}\big)$, and let φ be the indicator function of K. Prove that φ is the scaling function of a multiresolution analysis.

6. In connection with Proposition 10.38, define $m_0(y) = \frac{1}{2}(1 + e^{-4\pi iy})$.
   (a) Define $h(y) = \prod_{j=1}^{\infty} m_0(2^{-j}y)$, and check that $h(y) = (1 - e^{-4\pi iy})/(4\pi iy) = (\mathcal{F} I_{[0,2]})(y)$.
   (b) Verify that the integer translates of $\varphi = I_{[0,2]}$ do not form an orthonormal sequence. Conclude that hypothesis (iii) of Proposition 10.38 cannot be weakened to the point of allowing this particular trigonometric polynomial but still deducing that $\mathcal{F}^{-1}h$ is a scaling function.

Problems 7–12 concern the Haar system of Section 2. Let φ be the indicator function of $[0, 1)$, and define $\varphi_{j,k}(x) = 2^{j/2}\varphi(2^j x - k)$ as usual. The orthogonal projection of $L^2(\mathbb{R})$ on the closed linear span V_m of the subset $\{\varphi_{m,k}\}_{k\in\mathbb{Z}}$ of $L^2(\mathbb{R})$ is

$$(P_m f)(x) = \sum_{k\in\mathbb{Z}} \big(\textstyle\int_{\mathbb{R}} f(y)\overline{\varphi_{m,k}(y)}\, dy \big) \varphi_{m,k}(x).$$

7. Why is $P_m f$ meaningful for all f in $L^1(\mathbb{R})$? For f in $L^1(\mathbb{R})$, why is $P_m f$ convergent to f in $L^1(\mathbb{R})$ as m tends to $+\infty$?

8. Give an example of an L^1 function for which $P_m f$ does not tend to 0 in $L^1(\mathbb{R})$ as m tends to $-\infty$.

9. Suppose that $f(t) = 1$ for $0 \le t < 1/3$ and that $f(t) = 0$ elsewhere. Show that $\liminf_{m\to+\infty} P_m f(1/3) < \limsup_{m\to+\infty} P_m f(1/3)$, so that the Haar series of f diverges at the point $t = 1/3$. (By contrast the Fourier series of a function of bounded variation converges at every point.)

10. The Haar system of wavelets consists of the functions $\psi_{j,k}(x) = 2^{j/2}\psi(2^j x - k)$ on $\mathbb{R}$ with j and k in $\mathbb{Z}$. Here $\psi(x)$ equals 1 for $0 \le x < \frac{1}{2}$, equals -1 for $\frac{1}{2} \le x < 1$, and equals 0 otherwise. Prove that the nonzero restrictions to $[0, 1)$ of φ and the functions $\psi_{j,k}$ with $j \ge 0$ together form an orthonormal basis of $L^2([0, 1])$.

11. The Haar scaling function φ satisfies $\varphi(x) = \varphi(2x) + \varphi(2x - 1)$ almost everywhere. By working with Fourier transforms, show that any function f in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ with $f(x) = f(2x) + f(2x - 1)$ almost everywhere equals a multiple of $I_{[0,1]}(x)$ almost everywhere.

12. Let $\Phi(x) = I_{[-\frac{1}{2},\frac{1}{2})}(x)$. Using the orthogonality of the functions $x \mapsto \Phi(2x-k)$ for k in $\mathbb{Z}$, show that Φ cannot satisfy an equation

$$\Phi(x) = \sum_{k\in\mathbb{Z}} a_k \Phi(2x - k) \qquad \text{a.e.}$$

for complex constants a_k such that $\sum |a_k|^2 < \infty$.

Problems 13–17 concern the polynomial used in the construction of the Meyer wavelets in Section 6. Fix an integer $m \ge 1$. Construction of the Meyer wavelet of index m makes use of a C^m function ν on $\mathbb{R}$ such that $\nu(x) = 0$ for $x \le 0$, $\nu(x) = 1$ for $x \ge 1$, $\nu(y) + \nu(1 - y) = 1$ everywhere, and $\nu(x)$ equals a polynomial $P(x)$ for $0 \le x \le 1$. Define a polynomial $P(x)$ for current purposes to be "admissible" if it is divisible by x^{m+1}, has $P(1) = 1$, and has $P^{(k)}(1) = 0$ for $1 \le k \le m$.

13. Show that if P is usable as the polynomial in the definition of the Meyer wavelet of index m, then $P(x)$ is admissible.

14. Show that an admissible polynomial exists if and only if there exists an admissible polynomial of degree $\le 2m + 1$, and in this case the admissible polynomial of degree $\le 2m + 1$ is unique.

15. Show that an admissible polynomial P of the least possible degree necessarily satisfies $P(x) + P(1-x) = 1$ for all x. Deduce that such a polynomial is usable as the polynomial in the definition of the Meyer wavelet of index m.

16. This problem establishes a certain formula for the alternating sum of products of two binomial coefficients. To do so, it combines a technique in the proof of Lemma 10.35 with the technique used in proving the "Vandermonde convolution" formula $\sum_{j=0}^{n} \binom{n}{j}\binom{m}{k-j} = \binom{n+m}{k}$ for binomial coefficients, namely of recognizing the two sides of the formula as the coefficient of x^k on the two sides of the equation $(1 + x)^n(1 + x)^m = (1 + x)^{n+m}$.

(a) With m as above and with $p \leq m$, prove the formula

$$(1-z)^{m+1} = \sum_{q=0}^{p} (-1)^{p-q} \binom{m+1}{p-q} z^{p-q} + [z^{p+1}],$$

where $[z^{p+1}]$ is an analytic function in a disk about $z = 0$ that is divisible by z^{p+1}.

(b) With m and p as above, prove the formula

$$(1-z)^{-(m+1)} = \sum_{k=0}^{p} \binom{m+k}{k} z^k + [z^{p+1}],$$

where $[z^{p+1}]$ is another analytic function in a disk about $z = 0$ that is divisible by z^{p+1}.

(c) Taking the product of the results of (a) and (b), prove that

$$\sum_{q=0}^{p} (-1)^{p-q} \binom{m+1}{p-q} \binom{m+q}{q}$$

equals 1 if $p = 0$ and equals 0 if $0 < p \leq m$.

17. Let D be the differentiation operator on polynomials in one variable. If f and g are polynomials in x, the Leibniz rule says that

$$D^n(fg) = \sum_{k=0}^{n} \binom{n}{k} (D^{n-k} f)(D^k g).$$

(a) With m as above and $p \leq m$, apply the Leibniz rule to compute D^p of the polynomial

$$P(x) = x^{m+1} \sum_{k=0}^{m} \binom{m+k}{k} (1-x)^k.$$

Then evaluate at $x = 1$, and obtain the identity

$$(D^p P)(1) = p! \sum_{q=0}^{p} (-1)^q \binom{m+1}{p-q} \binom{m+q}{q}.$$

(b) Combine the above results to prove that the polynomial P is admissible for index m, and conclude that $P(x)$ is a polynomial of degree $2m+1$ that is usable as the polynomial in the definition of the Meyer wavelet of index m.

Problems 18–19 refer to the spaces $H^s(\mathbb{R})$ of Bessel potentials studied in Problems 8–12 for Chapter III. The idea is to show that the Daubechies wavelet of order N lies in a suitable space $H^s(\mathbb{R})$ with s equal to a real number depending on N. Then one can conclude that the Daubechies scaling function and wavelet of order N automatically have whatever smoothness is forced on all functions in the space $H^s(\mathbb{R})$.

18. Using the estimate

$$|(\mathcal{F}\varphi)(y)|^2 \le \Big(\frac{2^{N-1}}{\sqrt{\pi N}}\Big)(\sqrt{4\pi N}\,)^{-j} \qquad \text{for } 2^{j-1} \le |y| \le 2^j$$

from Proposition 10.43, show that the Daubechies scaling function φ and wavelet ψ of order N lie in $H^s(\mathbb{R})$ if $s < \frac{1}{4}\log_2(\pi N)$.

19. Deduce Corollary 10.46 from the previous problem in combination with Problem 12c for Chapter III, showing that if m is an integer ≥ 0 with $m < \frac{1}{4}\log_2(\pi N/4)$, then the Daubechies scaling function φ and wavelet ψ of order N are of class C^m.