An Introduction to Wavelets

Lectures delivered at Nicholas Copernicus University, Toruń

March 15th & 22nd, 2000

Richard M. Timoney
School of Mathematics, Trinity College, Dublin 2, Ireland.
Email: richardt@maths.tcd.ie

March 22, 2000

Abstract

We introduce wavelets as a particular way of choosing bases in function spaces.

The concept of a multiresolution analysis provides a setting for constructing certain wavelets and where practical algorithms have been developed. The constructions and the algorithms depend on a sequence of coefficients.

The term 'wavelet' is a relatively new term and most of the ideas are new, but there is a very active interest in using wavelet techniques in applications. We will mention some of these uses.

1 Introduction: A Review of Bases

In this section, we review briefly some concepts of basis and also recall the notion of a Fourier series.

1.1 Vector spaces and algebraic bases

The most familiar context for bases is in linear algebra. If V is a vector space over any field $\mathbb{K}$, then a *basis* for V is a subset $B \subset V$ such that every vector $v \in V$ can be expressed uniquely as a finitely nonzero sum

$$v = \sum_{b \in B} v_b b \qquad (v_b \in \mathbb{K} \forall b \in B, \{b \in B : v_b \neq 0\} \text{ finite})$$

Every vector space has a basis. We will refer to this type of basis as an *algebraic basis* when we need to make a distinction with other types of bases to be introduced below.

Choosing a basis B for an n-dimensional vector space V and an ordering $B = \{b_1, b_2, \ldots, b_n\}$ for the basis is equivalent to choosing a linear isomorphism

$$\begin{array}{rcl} : \mathbb{K}^n & \to & B \\ (\alpha_1, \alpha_2, \ldots, \alpha_n) & \mapsto & \displaystyle\sum_{j=1}^{n} \alpha_j b_j \end{array}$$

The fact that a given vector space will have many different bases (unless it is $\{0\}$), can be viewed as an advantage because it often allows us to choose a basis that adapts to the problem at hand. A simple example is the ability (in good cases) to choose a basis of eigenvectors for a given linear transformation $T \colon V \to V$.

Though bases always exist in theory, there are many infinite-dimensional cases of interest where one cannot write down any basis explicitly.

1.2 Finite-dimensional inner product spaces and orthonormal bases

Notation 1.2.1 *From now on the symbol $\mathbb{K}$ is reserved for a field which can only be either the reals $\mathbb{R}$ or the complex field $\mathbb{C}$.*

When given a finite-dimensional vector space V over $\mathbb{K}$ equipped with an inner product $\langle \cdot, \cdot \rangle \colon V \times V \to \mathbb{K}$, it is very convenient to work with an *orthonormal basis* $B = \{e_1, e_2, \ldots, e_n\}$ (satisfying $\langle e_j, e_k \rangle = 0$ for $j \neq k$ and $\langle e_j, e_j \rangle = 1$ for all j ($1 \leq j, k \leq n$)).

Then every vector can be expressed in terms of the basis in a computable way

$$v = \sum_{j=1}^{n} \langle v, e_j \rangle e_j \qquad (v \in V).$$

(Our notation is that inner products are linear in the first variable, and conjugate linear in the second variable in the complex case.)

An (ordered) orthonormal basis for a finite-dimensional inner product space $(V, \langle \cdot, \cdot \rangle)$ gives us an inner-product preserving linear isomorphism from $\mathbb{K}^n$ with the standard (euclidean) inner product to V.

For hermitian linear operators $T \colon V \to V$ on a finite-dimensional inner product space ($\langle Tv, w \rangle = \langle v, Tw \rangle$ for all $v, w \in V$) we can always find an orthonormal basis of eigenvectors.

1.3 Banach spaces and Schauder bases

Recall that a *Banach space* consists of a vector space X over $\mathbb{K}$ equipped with a *norm* $\| \cdot \|_X$ so that X is *complete* in that norm (every Cauchy sequence in X converges to a limit in X). Convergence in X is taken with respect to the metric (or distance) associated with the norm $d(x_1, x_2) = \|x_1 - x_2\|_X$.

With a norm we can contemplate infinite linear combinations, when we define infinite sums via limits, as in

$$\sum_{n=1}^{\infty} x_n = \lim_{n \to \infty} \sum_{j=1}^{n} x_j$$

Definition 1.3.1 *A* Schauder basis *for a Banach space X is a sequence $(e_n)_{n=1}^{\infty}$ of vectors in X such that every $x \in X$ can be expressed uniquely as an 'infinite linear combination'*

$$x = \sum_{n=1}^{\infty} x_n e_n \qquad (\textit{with } x_n \in \mathbb{K} \forall n)$$

Examples 1.3.2 (i) The classical sequence spaces c_0 and ℓ^p ($1 \le p < \infty$) have as bases the 'standard basis' $(e_n)_{n=1}^{\infty}$, where

$$e_n = (\delta_{nj})_{j=1}^{\infty}$$

($\delta_{nj} = 0$ if $n \neq j$ while $\delta_{nn} = 1$ for all n).

ℓ^p ($1 \leq p < \infty$) is the space of all sequences $(\alpha_j)_{j=1}^\infty$ ($\alpha_j \in \mathbb{K} \forall j$) such that

$$\left\|(\alpha_j)_{j=1}^\infty\right\|_p = \left(\sum_{j=1}^\infty |\alpha_j|^p\right)^{1/p} < \infty.$$

c_0 consists of all sequences $(\alpha_j)_{j=1}^\infty$ such that $\lim_{j\to\infty} \alpha_j = 0$ with the supremum norm

$$\left\|(\alpha_j)_{j=1}^\infty\right\|_\infty = \sup_{1\leq j<\infty} |\alpha_j|$$

(ii) The function spaces $L^p[0,1]$ and $L^p(\mathbb{R}^n)$ are defined (somewhat) similarly to ℓ^p except that they are (almost everywhere equivalence classes of) measurable functions f on the domain which have $|f|^p$ integrable with respect to Lebesgue measure.

Bases are not such a convenient idea for general Banach spaces. Only separable Banach spaces can have a basis and many separable Banach spaces fail to have a basis.

In general the order of summation in the infinite linear combination is important, and this inhibits one from considering a more general kind of basis where uncountably many basis vectors could be allowed. To remove dependence on the order, one can consider unconditional bases, but even fewer Banach spaces have these than have Schauder bases.

As in the vector space case, things are considerably simpler if one assumes that the there is an inner product. However, in the vector space case inner products can always be chosen (not necessarily in a useful and natural way) but in the Banach space case the existence of a compatible inner product is a severe restriction.

1.4 Hilbert spaces and orthonormal bases

Recall that a *Hilbert space* is an inner product space $(H, \langle\cdot,\cdot\rangle_H)$ which is complete (a Banach space) is the associated norm given by $\|x\|_H = \sqrt{\langle x, x\rangle_H}$.

Common examples are $L^2[0,1]$ and $L^2(\mathbb{R}^n)$ with the inner product given by

$$\langle f, g\rangle = \int f\bar{g}$$

(integrals with respect to Lebesgue measure).

Every Hilbert space H has an *orthonormal basis* in a sense that involves convergence. In the Hilbert space case an orthonormal basis may be defined as a maximal orthonormal subset $B \subset H$ — a set consisting of unit norm pairwise orthogonal elements of H with the property that it is not a proper subset of any other such set.

Given any orthonormal basis B for H, we can write every $x \in H$ as

$$x = \sum_{b \in B} \langle x, b \rangle_H b$$

in the sense that there are at most a countable number of nonzero terms in the summation and for any enumeration

$$\{b \in B : \langle x, b \rangle_H \neq 0\} = \{b_1, b_2, \ldots\}$$

of the nonzero terms, we have

$$x = \lim_{n \to \infty} \sum_{j=1}^{n} \langle x, b_j \rangle_H b_j.$$

(No limit is needed if there are only a finite number of nonzero terms.)

Moreover, we have a convenient representation of the norm on H in terms of the basis coefficients

$$\|x\|_H = \sqrt{\sum_{b \in B} |\langle x, b \rangle_H|^2}$$

1.5 Fourier series

From now on, we will typically consider $\mathbb{C}$-valued function spaces.

Example 1.5.1 *For $H = L^2[0,1]$ there is a very simple orthonormal basis that is so frequently used that it is almost the standard basis. It is $\{e_n : n \in \mathbb{Z}\}$ where*

$$e_n(t) = \exp(2\pi i n t)$$

From the general theory, we know that every $f \in L^2[0,1]$ can be expressed as a sum

$$f = \sum_{n \in \mathbb{Z}} \langle f, e_n \rangle e_n$$

with convergence of the sum in L^2-norm. This is normally known as the *Fourier series* of the function f and we often write

$$\hat{f}(n) = \langle f, e_n \rangle = \int_0^1 f(t) \exp(-2\pi i n t)\, dt.$$

It is also common to write

$$S_n f(t) = \sum_{j=-n}^{n} \hat{f}(j) \exp(2\pi i j t)$$

for certain (symmetrical) partial sums of the series and the general Hilbert space theory tells us that for all $f \in L^2[0,1]$ we have

$$\lim_{n \to \infty} \|f - S_n f\|_2 = 0$$

but it is a much deeper result due to Carleson that we also have almost everywhere convergence of $S_n f(t)$ to $f(t)$.

For $f \in L^p[0,1]$ and $1 < p < \infty$ one also knows that

$$\lim_{n \to \infty} \|f - S_n f\|_p = 0$$

so that the exponentials e_n also form a Schauder basis for $L^p[0,1]$.

They do *not* form a basis for $L^1[0,1]$ or for $C[0,1] =$ the continuous functions $f\colon [0,1] \to \mathbb{K}$ (with the supremum norm). What we can say however is that for $f \in L^1[0,1]$, there is enough information in the Fourier coefficients $(\hat{f}(n))_{n \in \mathbb{Z}}$ to completely determine f, but it is difficult to determine whether a given sequence $(a_n)_{n \in \mathbb{Z}}$ is the sequence of Fourier coefficients of some (unknown) function $f \in L^1[0,1]$.

Remark 1.5.2 *Why do we use Fourier series?*

One may justify the choice of the Fourier series example of an orthonormal basis for $L^2[0,1]$ on the basis that it has proved its value over time, but one may also argue that the complex exponentials $e_n(t) = \exp(2\pi i n t)$ are 'eigenvectors' of the differentiation operator $\frac{d}{dt}$.

Perhaps there is a slight problem because the operator is not globally defined $: L^2[0,1] \to L^2[0,1]$ and there are other eigenvectors $e_\lambda(t) = \exp(2\pi i \lambda t)$ with $\lambda \in \mathbb{C}$.

However, one may argue that the right context is to deal with is that of periodic functions (having 1 as period). For this, we extend all functions $f \in L^2[0,1]$ periodically to $\mathbb{R}$ by discarding the value at 1 and extending from $[0,1)$ to $\mathbb{R}$ using period 1 ($f(x) \stackrel{def}{=} f(x - [x])$ with $[x]$ the greatest integer $\leq x$). At least the terms of the Fourier series are naturally periodic in this way and the e_n are the 1-periodic eigenvectors of differentiation.

In the context of periodic functions, one can argue that we are actually dealing with functions on the unit circle of the complex plane $\mathbb{T} = \{\exp(2\pi i t) : t \in$

$\mathbb{R}\} = \{\exp(2\pi it) : t \in [0,1)\}$ and in this case the complex exponentials e_n may be viewed as the irreducible unitary representations of $\mathbb{T}$.

There is a generalisation of the Fourier theory to $L^2(G)$, for G a locally compact abelian group. We define the L^2 space with respect to Haar measure on G (normalised to give G measure 1 in the case G is compact) and then we have a 'Fourier series' representation of every $f \in L^2(G)$ where the series is indexed by the set $\hat{G}$ of irreducible unitary representations of G in place of $\mathbb{Z}$.

2 Fourier Transform

We will continue in this section to consider $\mathbb{C}$-valued function spaces. However, at some places it will be convenient to assume that a given $f \in L^2$ to be represented is actually $\mathbb{R}$-valued. Most of the remarks where this assumption is invoked can be adapted by linearity or other means to the case of general complex valued f.

We progress to consider Fourier series where the period is not 1 and from there, by a limiting argument, we arrive at the Fourier transform on $\mathbb{R}$. We then deal with some of the limitations of the Fourier transform as a preparation for motivating the notion of a wavelet later.

2.1 Fourier series on general intervals

We can transfer the theory of Fourier series on the unit interval $[0,1]$ to any other interval $[a,b]$ ($a < b \in \mathbb{R}$) by a simple change of variables. We have an isometric (and inner product preserving) map

$$\begin{aligned} : L^2[0,1] &\quad\to\quad L^2[a,b] \\ f &\quad\mapsto\quad \left(s \mapsto \frac{1}{\sqrt{b-a}} f\left(\frac{s-a}{b-a}\right)\right) \end{aligned}$$

and so a way to transfer Fourier series to $L^2[a,b]$.

Specialising to the case $[a,b] = [-T,T]$ we have that for each $g \in L^2[-T,T]$,

$$g(s) = \sum_{n\in\mathbb{Z}} \hat{g}(n) \left(\frac{(-1)^n}{\sqrt{2T}} \exp\left(\frac{2\pi ins}{2T}\right)\right)$$

with

$$\hat{g}(n) = \frac{(-1)^n}{\sqrt{2T}} \int_{-T}^{T} g(s) \exp\left(\frac{-2\pi ins}{2T}\right)\, ds.$$

2.2 The Fourier transform

Combining these last two together, we can say

$$g(s) = \sum_{n\in\mathbb{Z}} \frac{1}{2T}\left(\int_{-T}^{T} g(t)\exp\left(\frac{-2\pi int}{2T}\right)\,dt\right)\exp\left(\frac{2\pi ins}{2T}\right)$$

for $g \in L^2[-T,T]$.

If we take $g \in L^2(\mathbb{R})$ with compact support then we have

$$g(s) = \sum_{n\in\mathbb{Z}} \frac{1}{2T}\left(\int_{-\infty}^{\infty} g(t)\exp\left(\frac{-2\pi int}{2T}\right)\,dt\right)\exp\left(\frac{2\pi ins}{2T}\right)$$

for all T large.

If we now define a function $\mathcal{F}(g)$ by

$$\mathcal{F}(g)(\xi) = \int_{-\infty}^{\infty} g(t)\exp(-2\pi i\xi t)\,dt$$

we have

$$g(s) = \sum_{n\in\mathbb{Z}} \frac{1}{2T}\mathcal{F}(g)\left(\frac{n}{2T}\right)\exp\left(\frac{2\pi ins}{2T}\right)$$

By treating this summation as a Riemann sum for an integral and taking the limit as $T\to\infty$ we can justify

$$g(s) = \int_{-\infty}^{\infty} \mathcal{F}(g)(\xi)\exp(2\pi i\xi s)\,d\xi \tag{1}$$

for $g \in L^2(\mathbb{R})$ compactly supported. In fact, the formula (1) holds for all $g \in L^2(\mathbb{R})$.

The map $\mathcal{F}$ is called the *Fourier transform* on $\mathbb{R}$ and it can be proved (Parseval's theorem) that $\mathcal{F}\colon L^2(\mathbb{R}) \to L^2(\mathbb{R})$ is an isometric isomorphism. The formula (1) is the *Fourier inversion* formula, which exhibits the inverse transform as being almost of the same form as $\mathcal{F}$ (in fact it is the adjoint of $\mathcal{F}$).

In the Fourier series case, functions on $[0,T)$, or periodic functions with period T, are exhibited as superpositions of exponentials $t \mapsto \exp(2\pi nt/T)$ (with periods a multiple of T). For the infinite line, we no longer have this granularity of the periods and we need almost all possible periods $1/\xi$ for the Fourier inversion formula. We can argue that the continuous range of periods used implies that the summation in Fourier series becomes an integral in the case of $\mathbb{R}$. But we can still think of every $f \in L^2(\mathbb{R})$ as being given by a 'superposition' of exponentials.

We can view these exponentials as being all the bounded eigenfunctions of the derivative operator on $\mathbb{R}$ and this gives a clue to important applications of the Fourier transform in differential equations.

We can also view the Fourier transform from the group-theoretical point of view. Then the dual group of $\mathbb{R}$ (that is, the space of irreducible unitary representations) is again $\mathbb{R}$ — if we identify $\xi \in \mathbb{R}$ with the one dimensional representation of $\mathbb{R}$ given by its matrix as $t \mapsto \exp(2\pi i \xi t)$.

For future reference note that $t \mapsto \exp(2\pi i \xi t)$ is periodic with period $1/|\xi|$. We can say it repeats $|\xi|$ times when t increases by 1, and this justifies saying that it has *frequency* $|\xi|$ (measured in cycles or repetitions per unit of t).

2.3 Shannon-Nyquist sampling

For applications to digital communication, it is important to consider sending a signal over a channel which is limited in frequency range. One can communicate over such a channel only signals $f(t)$ with the property that the Fourier transform $\mathcal{F}f$ is supported inside the range of the channel. By a simple phase change of the signal (multiplying by a suitable complex exponential $\exp(2\pi i \xi_0 t)$), we may assume that the channel can carry frequencies in the range $[-W/2, W/2]$.

Now, if we take $g = \mathcal{F}f$, we can reconstruct g from its Fourier series

$$g(\xi) = \sum_{n \in \mathbb{Z}} \hat{g}(n) \left(\frac{1}{\sqrt{W}} \exp\left(\frac{2\pi i n \xi}{W} \right) \right) \qquad (|\xi| < W)$$

where[1]

$$\begin{aligned} \hat{g}(n) &= \frac{1}{\sqrt{W}} \int_{-W/2}^{W/2} g(s) \exp\left(\frac{-2\pi i n s}{W} \right) ds \\ &= \frac{1}{\sqrt{W}} \int_{-W/2}^{W/2} \mathcal{F}(f)(s) \exp\left(\frac{-2\pi i n s}{W} \right) ds \\ &= \frac{1}{\sqrt{W}} \int_{-\infty}^{\infty} \mathcal{F}(f)(s) \exp\left(\frac{-2\pi i n s}{W} \right) ds \end{aligned}$$

(using the fact that $\mathcal{F}(f)(s) = 0$ for $|s| > W/2$). This should be $(1/\sqrt{W})f(-n/W)$ by the Fourier inversion formula. However the Fourier inversion formula is a formula in $L^2(\mathbb{R})$ and so cannot be applied pointwise in general. It turns out (see below) that band-limited functions are automatically continuous, and then

[1] Strictly speaking we should have some factors $(-1)^n$ in $\hat{g}(n)$ and also on the exponentials above, but we have canceled these signs.

we can justify this pointwise application of the inversion formula. We can show $\hat{g}(n) = (1/\sqrt{W})f(-n/W)$. Thus the Fourier series formula says

$$\mathcal{F}(f)(\xi) = \sum_{n\in\mathbb{Z}} f\left(\frac{-n}{W}\right)\left(\frac{1}{W}\exp\left(\frac{2\pi i n\xi}{W}\right)\right) \qquad (|\xi| < W)$$

and using the Fourier inversion formula (plus the assumption that f is band-limited) we can show the following result.

Theorem 2.3.1 *If $f \in L^2(\mathbb{R})$ is band-limited so that $\mathcal{F}(f)$ is supported in $[-W/2, W/2]$, then f is completely determined by its values $f(n/W)$ ($n \in \mathbb{Z}$) and in fact*

$$f(t) = \sum_{n=-\infty}^{\infty} f(n/W)\frac{\sin(\pi(n - Wt))}{\pi(n - Wt)}$$

This can be interpreted to mean that if a channel is band-limited to a frequency band of total width W (or limited by $|\xi| < W/2$), then we cannot transmit a continuous function along the channel, but only one value every time interval $1/W$.

2.4 Compact support

We can be tempted from looking at band-limited functions (compact support of the Fourier transform) to consider compactly supported functions, but we cannot have both compact support for f and $\mathcal{F}(f)$ simultaneously. This is a consequence of the following theorem.

Theorem 2.4.1 (Paley-Wiener) *If $f \in L^2(\mathbb{R})$ has compact support then its Fourier transform $\mathcal{F}(f)(\xi)$ ($\xi \in \mathbb{R}$) extends to be an analytic function $\zeta \mapsto \mathcal{F}(f)(\zeta) \colon \mathbb{C} \to \mathbb{C}$ (or an entire function).*

In fact this entire function must be of exponential type, that is it must satisfy

$$|\mathcal{F}(f)(\zeta)| \leq Ae^{B|\zeta|} \qquad (\zeta \in \mathbb{C})$$

for some constants $A, B \geq 0$.

Moreover, the (restrictions to $\mathbb{R}$ of) entire functions of exponential type are exactly the Fourier transforms of compactly supported functions in $L^2(\mathbb{R})$.

A similar result applies to the inverse Fourier transform, so that the band-limited functions are those that are restrictions to $\mathbb{R}$ of entire functions of exponential type.

As an entire function cannot be zero on any interval of positive length, unless it is identically zero, it follows that f and $\mathcal{F}(f)$ cannot both be compactly supported.

A quantitative form of this fact can be stated for functions which are somewhat localised in time and have Fourier transforms that are somewhat localised as well.

For $f \in L^2(\mathbb{R})$ normalised to have unit norm ($\|f\|_2 = 1$ or $\int_{\mathbb{R}} |f(t)|^2 \, dt = 1$) we can treat $|f(t)|^2$ as a probability density function on $\mathbb{R}$. Then we can try to compute the mean μ and variance σ^2 of this probability distribution.

$$\mu = \int_{\mathbb{R}} t|f(t)|^2 \, dt, \qquad \sigma^2 = \int_{\mathbb{R}} (t-\mu)^2 |f(t)|^2 \, dt$$

We will certainly be able to do this for compactly supported $f \in L^2(\mathbb{R})$, but we will also succeed for more general (normalised) $f \in L^2(\mathbb{R})$.

As $\|\mathcal{F}(f)\|_2 = \|f\|_2 = 1$, we can also contemplate the mean $\hat{\mu}$ and variance $(\hat{\sigma})^2$ of $|\mathcal{F}(f)|^2$.

Theorem 2.4.2 (Heisenberg uncertainty principle) *If we have $f \in L^2(\mathbb{R})$ and also $f'(t), f''(t), tf(t), t^2 f(t) \in L^2(\mathbb{R})$ and $\|f\|_2 = 1$, then*

$$\sigma\hat{\sigma} \geq \frac{1}{4\pi}$$

We omit the proof of this, though it requires only some simple properties of the Fourier transform, the Cauchy-Schwarz inequality and integration by parts. See [1, Appendix F] for a proof.

2.5 Windowed Fourier transforms

A drawback of both Fourier series and Fourier transforms is that they destroy local information. Both allow reconstruction of functions in L^2 (and L^p for certain p, though the reconstruction is not possible by exactly the same elegant inversion formula). But it is part of the nature of the Fourier transform that the whole transform is needed to recover the function.

A simple example is provided by the characteristic functions of intervals $[0, a)$. By a simple computation the Fourier transform is

$$\mathcal{F}(\chi_{[0,a)})(\xi) = \frac{1 - \exp(-2\pi i a \xi)}{2\pi i \xi}$$

and we can see that a small shift in a affects the Fourier transform at each ξ.

In particular a relatively small localised change in the function requires a re-calculation of the Fourier transform/series and the whole transform is likely to be altered by such a change. Similarly a small local change on the frequency side (a small change of the Fourier transform) will normally affect the whole function.

One may analyse a signal or a sound wave such as a piece of music played over a certain time by taking its Fourier transform. First we must consider the signal (or sound wave) to extend over infinite time (perhaps by extending it with zero backwards in time to $-\infty$ and forwards in time to $+\infty$). One may consider (and this is typically done) the absolute value of the Fourier transform of the signal at frequency $\pm\xi$ to be a measure of the amount of the total amount of frequency $|\xi|$ present in the signal. (Assuming that the signal is a real one, we should combine $\mathcal{F}(f)(\xi)\exp(2\pi i\xi t)+\mathcal{F}(f)(-\xi)\exp(-2\pi i\xi t)$ to get $A(\xi)\cos(2\pi\xi(t-t_0))$ with amplitude $A(\xi)=2|\mathcal{F}(f)(\xi)|$ and *phase* t_0.)

For example in the case $f(t)=\chi_{[0,a)}(t)$, we get

$$A(\xi)=2\frac{\sin(\pi a\xi)}{\pi\xi}, \qquad t_0=\frac{a}{2}-\frac{1}{4\xi}.$$

The phase can be thought of loosely as representing a time shift in frequency ξ, but it is not easy to interpret it directly as related to a beginning time as one must consider the cancellation between all the terms to recover the signal. Thus the Fourier transform of the sound wave from a piece of music must contain all the notes as the original sound wave can be recovered from it. However, there is no simple way to detect from the Fourier transform which notes were played at which times.

In the L^2 setting (or even L^1) we cannot handle a sound wave that consists of a single note played with constant intensity for all time, but if we considered a suitable generalisation (distributions) where such signals could be handled, then the Fourier transformed signal would be concentrated at that one frequency. In this way we can justify the interpretation of the amplitude as the 'amount' of a given frequency present in a signal.

In an effort to also get a hold of local information, the windowed Fourier transform takes the Fourier transform of many localised versions of the original signal. We take a particular fixed function g of compact support, such as $g(t)=\chi_{[-1,1]}$ or a smoother version. Then we translate g by arbitrary amounts to get $t\mapsto g(t-a)$ and take the Fourier transform of $f(t)g(t-a)$. We get a transform for each translation amount a

$$\mathcal{WF}(f)(a,\xi)=\mathcal{F}(f(t)g(t-a))(\xi)=\int_{\mathbb{R}} f(t)g(t-a)\exp\left(-2\pi i\xi t\right)dt$$

The definition of $\mathcal{WF}$ depends on the choice of g and g is known as the *window function*. One may find strange effects caused by discontinuities or lack of smoothness of g and so choosing a smooth compactly supported window function is desirable as a rule.

The resulting $\mathcal{WF}$ does give local information about the function f, at the expense of introducing a 'position' parameter a. For a fixed a we must be able to find from $\mathcal{WF}(a, \cdot)$ all the information about f in the support of $t \mapsto g(t-a)$ (which is the support of g translated by a).

However, the information is still not very easy to decode. We can say that $\mathcal{WF}(f)(a, \xi)$ represents the amount of frequency ξ present in the graph of f intersection the window around a, though there will be an effect from the shape of the graph of g also.

We can argue that $\mathcal{WF}$ must be limited in its ability to localise because of the Heisenberg uncertainty principle. Applying the window to get $f(t)g(t-a)$ certainly gives us a windowed signal with relatively small standard deviation σ, but we do not always get good control on $\hat{\sigma}$, the standard deviation on the transform side. At the best, we could generate window functions where $\hat{\sigma}$ would be proportional to the reciprocal of the width of the window.

One might argue that for a relatively high frequency ξ where the corresponding period $1/\xi$ is smaller than the window, we are in a reasonably good position to detect whether there is a frequency ξ component inside the window. For relatively small frequency ξ, and large period, we do not really have a big enough window on the graph of f to say that we are detecting any variation in the graph at that frequency.

Of course, when we take the Fourier transform, we do see something at all frequencies, including small frequencies, but we are really detecting features caused by the zero-extension of the graph of $f(t)g(t-a)$ past the window.

At high frequencies, small periods, we have room for several oscillations inside the window and we are now back to a similar position to that we had with the Fourier transform $\mathcal{F}(f)(\xi)$ where we are not able to say where within the window the high frequency change in f happens.

If we managed to optimise $\hat{\sigma}$ as roughly proportional to the reciprocal of the width of the window, we would be able to do quite a good job of pin-pointing the position of reasonably low frequency components of f, but for higher frequencies we could not expect to pin down the location except within a multiple of the period, a high multiple in the case of high frequencies.

3 Wavelets via Frames

We continue the ideas introduced in studying the windowed Fourier transform to get one approach to wavelets, perhaps the most general approach. Later we will take a more restrictive setting where there are better computational algorithms available.

3.1 Continuous wavelet transform

If we take the discussion above a step further, it suggests that if we want to know where we can find a component of frequency ξ in the graph of f, we should not expect to be able to answer unless we can consider a section (or window) of the graph of length comparable to the period $1/\xi$.

We then take a window such as

$$\psi(t) = \begin{cases} \sin(2\pi t) & -1/2 < t < 1/2 \\ 0 & t \geq 1/2 \text{ or } t \leq -1/2 \end{cases}$$

but we could take any ψ. Normally we assume that ψ is compactly supported, $\psi \in L^2(\mathbb{R})$ and $\int_{\mathbb{R}} \psi(t)\,dt = 0$. We should think of the graph of ψ as a single 'cycle' of a more or less periodic wave, but we will often want additional properties such as smoothness of ψ which will make this only approximately correct.

Instead of translating the wavelet around as we did for windows, we also scale the wavelet to give stretched versions of the original wavelet ψ with the same basic shape but at different scale or frequency.

We define

$$\psi_{a,b}(t) = \frac{1}{\sqrt{a}}\psi\left(\frac{t-b}{a}\right) \qquad (a > 0, b \in \mathbb{R})$$

and this will give us a single 'wavelet' with support stretched to a of its previous length, frequency $1/a$ times the original frequency in some sense. (The factor $1/\sqrt{a}$ is not really essential, but it is there to preserve L^2 norms — $\|\psi_{a,b}\|_2 = \|\psi\|_2$.)

Instead of imposing a further frequency on this, we just consider its inner product with a given f.

Definition 3.1.1 *The* continuous wavelet transform *of $f \in L^2(\mathbb{R})$ corresponding to a choice of 'wavelet' ψ is*

$$W(f)(a,b) = \langle f, \psi_{a,b} \rangle = \int_{\mathbb{R}} f(t) \frac{1}{\sqrt{a}} \overline{\psi\left(\frac{t-b}{a}\right)}\,dt \qquad (a > 0, b \in \mathbb{R}).$$

The key to the continuous wavelet transform is that we can choose ψ so that $W(f)(a,b)$ contains enough information to reconstruct the function f. The simplest example where this is possible is called the *Haar Wavelet*

$$\psi(t) = \begin{cases} 1 & 0 \leq t < 1/2 \\ -1 & 1/2 \leq t < 1 \\ 0 & t \geq 1 \text{ or } t < 0 \end{cases}$$

Theorem 3.1.2 (Inverse wavelet transform) *If $\psi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ is real-valued and satisfies the* admissibility condition

$$c_\psi = \int_0^\infty \frac{|\mathcal{F}(\psi)(\xi)|^2}{\xi} \, d\xi < \infty$$

then for $f \in L^2(\mathbb{R})$

$$\|f\|_2 = \sqrt{c_\psi} \left(\int_{(a,b)\in(0,\infty)\times\mathbb{R}} |W(f)(a,b)|^2 \, \frac{da}{a^2} \, db \right)^{1/2}$$

For $f \in L^p(\mathbb{R})$ $(1 < p < \infty)$

$$f(t) = \frac{1}{c_\psi} \int_{(a,b)\in(0,\infty)\times\mathbb{R}} W(f)(a,b)\psi_{a,b}(t) \, \frac{da}{a^2} \, db$$

(if the integral is interpreted in a distributional sense).

We will not try to prove this result.

Notice that the Haar wavelet has Fourier transform

$$\mathcal{F}(\psi)(\xi) = \frac{(1 - \exp(-\pi i \xi))^2}{2\pi i \xi}$$

and it is admissible.

It is easy to see that a compactly supported C^∞ function ψ with $\int_{\mathbb{R}} \psi(t) \, dt = 0$ is 'admissible' because the Fourier transform of any such function will be analytic, have a zero at $\xi = 0$ and decay faster than any power of $1/|\xi|$ as $|\xi| \to \infty$.

3.2 Discretisation of the CWT

In the case of the Haar wavelet at least, it is easy to see that there is a great redundancy of information in $W(f)(a,b)$. Among the $\psi_{a,b}$, those with $a = 1$ contain an orthonormal set in $L^2(\mathbb{R})$:

$$\{\psi_{1,m} : m \in \mathbb{Z}\} = \{\chi_{[m,m+1/2)} - \chi_{[m+1/2,m+1)} : m \in \mathbb{Z}\}$$

This is clearly not an orthonormal basis because no function can be expressed in terms of these unless it is constant on the intervals with endpoints at adjacent half integers.

However, the orignal work of Haar showed that

$$\{\psi_{2^n,2^n m} : n, m \in \mathbb{Z}\}$$

does form an orthonormal basis for $L^2(\mathbb{R})$. Thus we can express every $f \in L^2(\mathbb{R})$ as

$$f = \sum_{n,m\in\mathbb{Z}} \langle f, \psi_{2^n,2^n m}\rangle\, \psi_{2^n,2^n m} = \sum_{n,m\in\mathbb{Z}} W(f)(2^n, 2^n m)\psi_{2^n,2^n m}$$

from the general theory of orthonormal bases in Hilbert spaces.

This has been known and used by various authors since Haar invented it, but it has not had so many applications because the basis functions are not continuous and not well suited to many applications.

What is new in the wavelet context is that we can use other admissible ψ which have good behaviour in many ways. We can even find much more regular (for example continuous or continuously differentiable) ψ so that

$$\{\psi_{2^n,2^n m} : n, m \in \mathbb{Z}\}$$

forms an orthonormal basis for $L^2(\mathbb{R})$.

For some purposes, we might be happy with less, a discrete set of points (a_n, b_m) which are sufficient to recover f from $W(f)(a_n, b_m)$.

3.3 Frames

There are a number of ways to generalise the concept of an orthonormal basis in a separable Hilbert space.

One generalisation, known as a *Riesz basis*, allows for a sequence of vectors $x_n \in H$ with the property that they form a Schauder basis for H and so that for any convergent infinite linear combination $\sum_n \alpha_n x_n$ ($\alpha_n \in \mathbb{K}$ for all n) we have

$$A \left\| \sum_n \alpha_n x_n \right\|_H^2 \leq \sum_n |\alpha_n|^2 \leq B \left\| \sum_n \alpha_n x_n \right\|_H^2$$

for some fixed $A, B > 0$.

Every Riesz basis for H can be got by applying an invertible continuous linear map $T\colon H \to H$ to some orthonormal basis of H and so we can view a Riesz basis as an orthonormal basis which has been somewhat distorted.

The more general concept of a frame was first introduced by Duffin and Schaeffer in 1952.

Definition 3.3.1 *A sequence $(x_n)_{n=1}^\infty$ of elements of a Hilbert space H is called a* frame *for H if there are constants $A, B > 0$ so that*

$$A\|x\|_H^2 \leq \sum_{n=1}^{\infty} |\langle x, x_n\rangle|^2 \leq B\|x\|_H^2$$

holds for all $x \in H$.

Examples 3.3.2 (i) A *wavelet frame* for $L^2(\mathbb{R})$ is a frame of the type

$$\{a^{-n/2}\psi(a^{-n}t - mb)\}_{n,m\in\mathbb{Z}}$$

for a fixed $\psi \in L^2(\mathbb{R})$ and some $a > 0$, $b \neq 0$.

(ii) Relatively uninteresting examples of frames can be generated from orthonormal bases $e_1, e_2, \ldots$ by listing each basis vector a finite number of times in the frame. For example $e_1, e_1, e_2, e_2, e_3, e_3, \ldots$ (where each is repeated twice) is a frame with $A = B = 2$.

A frame with $A = B$ is called a *tight* frame and a frame which ceases to be a frame on removal of one vector from the sequence is called an *exact* frame. The frame generated from an orthonormal basis by

$$\frac{1}{\sqrt{2}}e_1, \frac{1}{\sqrt{2}}e_1, e_2, e_3, e_4, \ldots$$

is tight with $A = B = 1$ but is not exact.

To make a connection between frames and bases, we introduce the coefficient operator $U\colon H \to \ell^2$ associated with a frame $x_1, x_2, \ldots$ in H given by

$$Ux = (\langle x, x_n\rangle)_{n=1}^{\infty} \in \ell^2$$

What we know is that

$$A\|x\|_H^2 \leq \|Ux\|_2^2 = \langle Ux, Ux\rangle = \langle U^*Ux, x\rangle_H \leq B\|x\|_H^2$$

and this implies that $U^*U : H \to H$ is a positive bounded linear operator on H (in the sense of positive definiteness that $\langle U^*Ux, x\rangle_H \geq 0$ for all $x \in H$). Also U^*U satisfies the inequalities[2]

$$A\,\mathrm{id}_H \leq U^*U \leq B\,\mathrm{id}_H$$

in the sense that the differences $U^*U - A\,\mathrm{id}_H$ and $B\,\mathrm{id}_H - U^*U$ are positive operators. Moreover we can compute

$$U^*Ux = \sum_{n=1}^{\infty}\langle x, x_n\rangle x_n$$

We call U^*U the *frame operator* and denote it by S.

[2] id_H denotes the identity operator $: H \to H$

The frame operator S is invertible and if $x \in H$ we can show

$$x = \sum_{n=1}^{\infty} \langle S^{-1}x, x_n \rangle x_n = \sum_{n=1}^{\infty} \langle x, x_n \rangle S^{-1} x_n$$

With a frame we do not have to have linear independence in general, but we can show an optimality property for the ℓ^2 norm of the coefficients in the above canonical expansion of x in terms of x_n (with coefficients $\langle S^{-1}x, x_n \rangle$). If $x = \sum_{n=1}^{\infty} \alpha_n x_n$ for some scalars α_n, then

$$\sum_{n=1}^{\infty} |\alpha_n|^2 = \sum_{n=1}^{\infty} |\langle S^{-1}x, x_n \rangle|^2 + \sum_{n=1}^{\infty} |\langle S^{-1}x, x_n \rangle - \alpha_n|^2 \geq \sum_{n=1}^{\infty} |\langle S^{-1}x, x_n \rangle|^2$$

If we have a tight frame ($A = B$) then $S = A\,\mathrm{id}_H$ and

$$x = \frac{1}{A} \sum_{n=1}^{\infty} \langle x, x_n \rangle x_n$$

By scaling the frame, we can arrive at one where $A = 1$. Every $x \in H$ has a representation in terms of such a frame which has many of the properties of an expansion in terms of an orthonormal basis, except uniqueness.

For non-tight frames, the necessity to compute S^{-1} to arrive at a concrete representation for $x \in H$ is an obstacle to the practical use of such frames.

Proposition 3.3.3 *Suppose $\psi \in L^1(\mathbb{R})$ has the property that $\mathcal{F}(\psi)(\xi)$ has no zeros for $1 < |\xi| < k$ for some $k > 2$. Then the set*

$$\{2^{n/2}\psi(2^n t - 2^n m) : n, m \in \mathbb{Z}\}$$

is a frame in $L^2(\mathbb{R})$.

4 Wavelets via Multiresolution Analysis

In this section we take a more restrictive approach to wavelets, which leads to a more effective approach in practice than the frame-based approach.

4.1 Multiresolution analysis on R

Looking at the Haar wavelet example, we have the discrete version where every $f \in L^2(\mathbb{R})$ can be expressed

$$f = \sum_{n,m \in \mathbb{Z}} \langle f, \psi_{2^n,2^n m} \rangle \, \psi_{2^n,2^n m}$$

$$\psi_{2^n,2^n m}(t) = 2^{n/2} \left(\chi_{[2^{-n}m, 2^{-n}(m+1/2)} - \chi_{[2^{-n}(m+1/2), 2^{-n}(m+1)} \right)$$

For a fixed n we see functions constant on intervals of length 2^{-n-1} (and all the functions have integral 0). Recall that we can have n positive and negative so that we get short as well as long intervals.

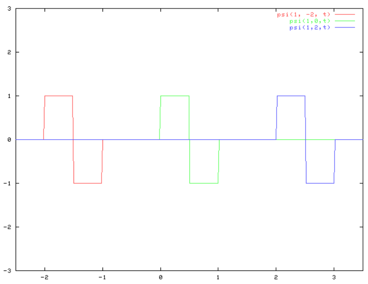
If we take

$$W_n = \text{span}\{\psi_{2^{-n},2^{-n}m} : m \in \mathbb{Z}\} \qquad V_n = \text{span}\{\psi_{2^{-k},2^{-k}m} : k, m \in \mathbb{Z}, k < n\}$$

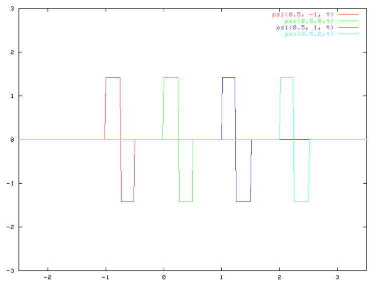
(we take the closed linear spans), then we can deduce from orthonormality and the fact the the $\psi_{2^n,2^n m}$ form a basis that

$$\begin{aligned}
\{0\} = \bigcap_{n \in \mathbb{Z}} V_n \quad &\subset \quad \cdots \subset V_{-1} \subset V_0 \subset V_1 \subset V_2 \subset \cdots \\
&\subset \quad \bigcup_{n \in \mathbb{Z}} V_n \text{ dense in } L^2(\mathbb{R}) \qquad (2) \\
f(t) \in V_n \quad &\iff \quad f(2t) \in V_{n+1} \qquad (3) \\
V_n \quad &= \quad V_{n-1} \oplus W_n \\
& \qquad \text{(orthogonal direct sum)} \qquad (4) \\
V_n \quad &= \quad \text{span} \left\{ \begin{array}{l} \text{functions constant on intervals} \\ \text{of length } 2^{-n} \text{ starting at } 0 \end{array} \right\} \\
W_n \quad &= \quad \text{span} \left\{ \begin{array}{l} \text{functions constant on intervals} \\ \text{of length } 2^{-n-1} \text{ starting at } 0 \\ \text{and which average } 0 \text{ on the intervals} \\ \text{of length } 2^{-n} \end{array} \right\}
\end{aligned}$$

See Figure 1 for a graphical view of this. In Figure 2, we try to show what happens when we build up linear combinations of $\psi_{2^n,2^n m}$ starting with $n = 1$



A selection of graphs of $\psi_{1,m}$, $m = -2, 0, 2$



A selection of graphs of $\psi_{1/2,b}$, $b = -1, 0, 1, 2$

Figure 1: Haar wavelets at different scales

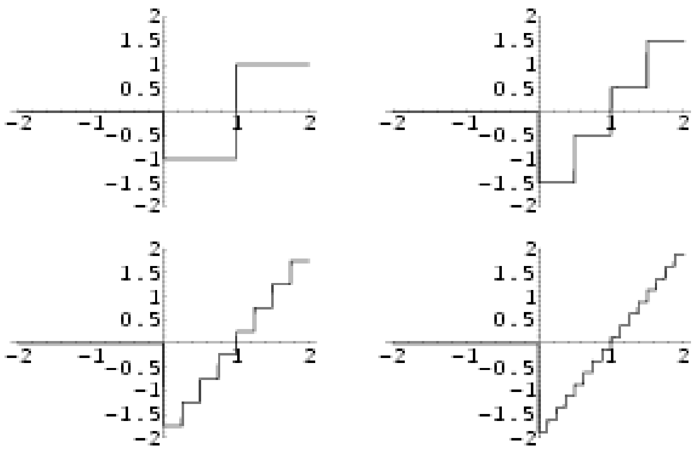


Figure 2: Combining Haar wavelets at finer scales

(constant on intervals of length 1) and then adding some more terms (with smaller coefficients) with $n = 0, -1, -2$. The additional terms affect the details of the graph. In fact, we show successive approximations to $2(x-1)\chi_{[0,1)}(x)$, first using $\psi_{2,0}$, then adding a combination of $\psi_{1,m}$, next adding a combination of $\psi_{1/2,(1/2)m}$, etc.

We take the point of view that V_n is what one can see if one is restricted to taking averages over (dyadic) blocks of length $1/2^n$ (blocks $[k/2^n, (k+1)/2^n)$).

A sequence V_n of subspaces on $L^2(\mathbb{R})$ with the properties (2) and (3) is called a *multiresolution analysis* of $L^2(\mathbb{R})$.

In the example at hand, we have one additional key property, that all the V_n

are spanned by the integer translates of one function. For example

$$V_0 \text{ is spanned by the translates } \phi(x-m) \qquad (m \in \mathbb{Z}) \tag{5}$$

of the function $\phi(t) = \chi_{[0,1)}(t)$. In fact these form an orthonormal basis for V_0 (which is even more convenient). Then property (3) implies that $\phi_{2^{-n},2^{-n}m}(t) = 2^{n/2}\phi(2^n t - m)$ $(m \in \mathbb{Z})$ form an orthonormal spanning sequence for V_n (all $n \in \mathbb{Z}$).

Moreover, the orthogonal complement W_{n-1} of V_{n-1} inside V_n is also spanned by the translates of one function, $\{\psi_{2^{-n+1},2^{-n+1}m} : m \in \mathbb{Z}\}$.

It turns out that, in general, we can deduce that W_n is spanned by the translates of one function based on the fact that (2), (3), (4) and (5) hold. We have

$$\phi(t) \in V_1 = \text{span}\{\phi_{1/2,m/2}(t) = \sqrt{2}\phi(2t-m) : m \in \mathbb{Z}\}$$

and since the $\sqrt{2}\phi(2t-m)$ are orthonormal, this means we must be able to express

$$\phi = \sum_{m\in\mathbb{Z}} \langle \phi, \phi_{1/2,m/2}\rangle \phi_{1/2,m/2} = \sum_{m\in\mathbb{Z}} c_m \phi_{1/2,m/2}.$$

Indeed, in the particular Haar case we were considering, we have

$$\phi(t) = \phi(2t) + \phi(2t-1) = \frac{1}{\sqrt{2}}\phi_{1/2,0}(t) + \frac{1}{\sqrt{2}}\phi_{1/2,1/2}(t)$$

and notice that we also have

$$\psi(t) = \frac{1}{\sqrt{2}}\phi_{1/2,0}(t) - \frac{1}{\sqrt{2}}\phi_{1/2,1/2}(t)$$

It turns out in general that something similar happens.

Theorem 4.1.1 *Suppose $(V_n)_{n\in\mathbb{Z}}$ is a sequence of subspaces of $L^2(\mathbb{R})$ which form a multiresolution analysis (that is, they satisfy (2) and (3))) and suppose that (5) holds for a function ϕ with orthonormal translates $\phi(t-m)$ $(m \in \mathbb{Z})$.*

(i) Then we must have

$$\sum_{m\in\mathbb{Z}} |c_m|^2 = 1 \qquad \text{for } c_m = \langle \phi, \phi_{1/2,m/2}\rangle$$

and

$$\sum_{m\in\mathbb{Z}} c_m \overline{c_{m-2\ell}} = 0 \text{ for } 0 \neq \ell \in \mathbb{Z} \tag{6}$$

(ii) *If we define*

$$\psi = \sum_{m\in\mathbb{Z}} (-1)^m \overline{c_{1-m}} \phi_{1/2,m/2}$$

then $\psi \in V_1$ and it is orthogonal to V_0.

Moreover the translates $\psi_{1,m}$ ($m \in \mathbb{Z}$) form an orthonormal basis for the orthogonal complement of V_0 in V_1 and the whole collection of translated and scaled functions

$$\{\psi_{2^n,2^n m} : n, m \in \mathbb{Z}\}$$

based on ψ is an orthonormal basis for $L^2(\mathbb{R})$.

(iii) *If we assume that $\phi \in L^1(\mathbb{R}) \cap L^2(\mathbb{R})$, $\int_{\mathbb{R}} \phi(t)\, dt \neq 0$ and $\{m : c_m \neq 0\}$ is finite, then we have*

$$\sum_{m\in\mathbb{Z}} c_m = \sqrt{2} \tag{7}$$

$$\mathcal{F}(\phi)(\xi) = P(\xi/2)\mathcal{F}(\phi)(\xi/2) \tag{8}$$

$$\text{where } P(\xi) = \sum_{m\in\mathbb{Z}} c_m e^{2\pi i m\xi}$$

$$\mathcal{F}(\phi)(\xi) = \mathcal{F}(\phi)(0)\prod_{k=1}^{\infty} P(\xi/2^k) \tag{9}$$

Proof.

(i) The first part follows from orthonormality of $\phi(t-m)$ and the second by expanding $\langle \phi(t), \phi(t-\ell)\rangle = 0$ in terms of $\phi = \sum_{m\in\mathbb{Z}} c_m \phi_{1/2,m/2}$.

(ii) We can express

$$\psi(t-\ell) = \sum_{m\in\mathbb{Z}} (-1)^m \overline{c_{1-(m+2\ell)}} \phi_{1/2,m/2}(t)$$

and so expand

$$\langle \phi(t), \psi(t-\ell)\rangle = \sum_{m\in\mathbb{Z}} (-1)^m c_m c_{1-(m+2\ell)} = 0$$

(since the m and $1-m-2\ell$ terms have opposite signs).

The orthonormality of the $\psi_{1,m}$ follows by (i).

To show that they span the orthogonal complement of V_0 in V_1, suppose $f \in V_1$ is perpendicular to V_0 and to all $\psi_{1,m}$. We can write $f = \sum_{m\in\mathbb{Z}} f_m \phi_{1/2,m/2} = \sum_{m\in\mathbb{Z}} \langle f, \phi_{1/2,m/2}\rangle \phi_{1/2,m/2}$ and let us denote the sequence of coefficients of

f by $F = (f_m)_{m\in\mathbb{Z}}$. Suppose we also denote by $\Phi_\ell = (c_{m-2\ell})_{m\in\mathbb{Z}}$ the coefficients of $\phi(t-\ell) = \phi_{1,\ell}$ and by $\Psi_k = \left((-1)^m \overline{c_{1-(m+2k)}}\right)_{m\in\mathbb{Z}}$ the coefficients of $\psi(t-k) = \psi_{1,k}(t)$ in the same basis for V_0.

Consider the matrix M that has as its columns

$$(\ldots, \Phi_1, \Psi_1, \Phi_0, \Psi_0, \Phi_{-1}, \Psi_{-1}, \Phi_{-2}, \ldots).$$

These columns are orthonormal by the earlier parts of the proof. The matrix M represents a linear operator on ℓ^2 (where we take $\mathbb{Z}$ as the index set for the sequence space) that transforms the standard basis to the orthonormal sequence given by the columns. When we compute M^*M we get the identity matrix. But $M^*F = 0$ by assumption and so what we want is MM^* to be the identity.

If we break M into 2×2 blocks M_{rs} ($r, s \in \mathbb{Z}$) we get

$$M_{rs} = \begin{pmatrix} c_{2r+2s} & \overline{c_{2s-2r+1}} \\ c_{2r+2s+1} & -\overline{c_{2s-2r}} \end{pmatrix}$$

and if we work out MM^* we get the matrix made up of the blocks

$$\begin{aligned}
&\sum_j M_{r,j}(M_{s,j})^* \\
&\quad = \sum_j \begin{pmatrix} c_{2r+2j} & \overline{c_{2j-2r+1}} \\ c_{2r+2j+1} & -\overline{c_{2j-2r}} \end{pmatrix} \begin{pmatrix} \overline{c_{2s+2j}} & \overline{c_{2s+2j+1}} \\ c_{2j-2s+1} & -c_{2j-2s} \end{pmatrix} \\
&\quad = \begin{pmatrix} \sum_j (c_{2r+2j}\overline{c_{2s+2j}} + \overline{c_{2j-2r+1}}c_{2j-2s+1}) & \sum_j (c_{2r+2j}\overline{c_{2s+2j+1}} - \overline{c_{2j-2r+1}}c_{2j-2s}) \\ \sum_j (c_{2r+2j+1}\overline{c_{2s+2j}} - \overline{c_{2j-2r}}c_{2j-2s+1}) & \sum_j (c_{2r+2j+1}\overline{c_{2s+2j+1}} + \overline{c_{2j-2r}}c_{2j-2s}) \end{pmatrix} \\
&\quad = \begin{pmatrix} \delta_{rs} & 0 \\ 0 & \delta_{rs} \end{pmatrix}
\end{aligned}$$

(The off-diagonal sums rearrange to 0 and the orthogonality relations (6) show that the diagonal entries are 0 unless $r = s$, when they are 1 by (i).) Hence MM^* is the identity, $F = MM^*F = 0$ and so $f = 0$. Thus the collection $\{\phi_{1,m} : m \in \mathbb{Z}\} \cup \{\psi_{1,m} : m \in \mathbb{Z}\}$ is a maximal orthonormal subset (an orthonormal basis) of V_0.

Since $\psi_{1,\ell}$ is orthogonal to ϕ, it follows that $\langle \psi_{1,\ell}, \phi_{1,m}\rangle = \langle \psi_{1,\ell-m}, \phi\rangle = 0$ for all m. As $\phi_{2,0} = \sum_{m\in\mathbb{Z}} c_m \phi_{1,m}$ it follows that $\langle \psi_{1,\ell}, \phi_{2,0}\rangle = 0$.

Iterating these ideas, we can show that $\psi_{2^n,2^n m}$ are all orthonormal. Since $\bigcap_{n\in\mathbb{Z}} V_n = \{0\}$ we can show that any $f \in V_k$ which is orthogonal to all

$\psi_{2^n,2^n m}$ $(n > -k)$ must be zero and $\{\psi_{2^n,2^n m} : n > -k\}$ is an orthonormal basis for V_n. Using density of $\bigcup_{n\in\mathbb{Z}} V_n$ in $L^2(\mathbb{R})$ we can show that $\psi_{2^n,2^n m}$ form an orthonormal basis of $L^2(\mathbb{R})$.

(iii) Integrating both sides of $\phi = \sum_{m\in\mathbb{Z}} c_m \phi_{1/2,m/2}$ gives (7). Taking Fourier transforms of both sides gives (8). Iterating (8) gives

$$\mathcal{F}(\phi)(\xi) = \mathcal{F}(\phi)(\xi/2^n) \prod_{k=1}^{n} P(\xi/2^k)$$

and letting $n \to \infty$ gives (9).

4.2 Daubechies wavelets

By Theorem 4.1.1, we can find wavelets if we can find finitely nonzero sequences of coefficients c_m that satisfy the conditions of part (i) of that theorem and (7). There is another step required, to show that the infinite product in (9) converges to the Fourier transform of a function in $L^1(\mathbb{R}) \cap L^2(\mathbb{R})$ with nonzero integral. This step usually works because of results which we will not explain and the resulting ϕ (and ψ) will be compactly supported in most cases.

The function ψ is the basic wavelet (sometimes called the 'mother wavelet') resulting from the multiresolution analysis and the ϕ is a generating function that is sometimes called the 'father function', but also known as the *scaling function*.

Examples 4.2.1 (i) If we allow only two non-zero coefficients c_0 and c_1, then the conditions we have force us to have the Haar wavelet situation.

Allowing a possible nonzero c_2 as well as c_0 and c_1 still reduces to the Haar wavelet situation.

(ii) Daubechies showed that there are a family of choices of coefficients $c_0, c_1, \ldots, c_{2p-1}$ $(p \geq 1)$ that lead to continuous compactly-supported ϕ (and continuous compactly-supported wavelets). As p increases the corresponding wavelet becomes more differentiable.

For $p = 2$ she exhibited

$$c_0 = \frac{1+\sqrt{3}}{4\sqrt{2}}, c_1 = \frac{3+\sqrt{3}}{4\sqrt{2}}, c_2 = \frac{3-\sqrt{3}}{4\sqrt{2}}, c_3 = \frac{1-\sqrt{3}}{4\sqrt{2}}$$

4.3 Higher dimensions

Generalisations to $\mathbb{R}^d$ often require the use of $2^d - 1$ wavelets and not just one. So 3 wavelets for $\mathbb{R}^2$, for example.

The simplest approach is to take a wavelet ψ arising from a multiresolution of $L^2(\mathbb{R})$ and its associated scaling function ϕ. Then we consider the wavelet basis of $L^2(\mathbb{R}^2)$ given by

$$\{\psi^{ws}_{2^n,2^n m} : n \in \mathbb{Z}, m \in \mathbb{Z}^2\} \cup \{\psi^{sw}_{2^n,2^n m}\} \cup \{\psi^{ww}_{2^n,2^n m}\}$$

where

$$\psi^{ws}(x,y) = \psi(x)\phi(y), \quad \psi^{sw}(x,y) = \phi(x)\psi(y), \quad \psi^{ww}(x,y) = \psi(x)\psi(y)$$

are 'tensor products' of the wavelet ψ in one variable with the scaling function ϕ or with ψ in the other variable.

This corresponds to a multiresolution analysis of $L^2(\mathbb{R}^2)$ which respects the dilations $x \mapsto 2x$ and the translations $x \mapsto x + b$ with $b \in \mathbb{Z}^2$.

One can replace the dilation $x \mapsto 2x$ by $x \mapsto Ax$ where A is a matrix with integer entries which is expansive ($\|A^n x\| \to \infty$ as $n \to \infty$ for each $0 \neq x \in \mathbb{R}^2$). The number of wavelets required is then $|\det(A)| - 1$ and this can be 1, for example if

$$A = \begin{pmatrix} 1 & -1 \\ 1 & 1 \end{pmatrix}$$

Although only 1 wavelet may be needed, there are some A where there is no scaling function to generate a multiresolution analysis and there do not seem to be effective algorithms with such approaches.

5 Applications and Concluding Remarks

There have been a vast range of practical applications of wavelets.

There are effective algorithms for dealing with orthonormal wavelets that arise from a multiresolution analysis. All the calculations are done in terms of the coefficients c_m.

Practical applications normally rely on a finite-dimensional version of the wavelet basis, similar to the way the Discrete Fourier Transform (DFT) works. There is a Fast Wavelet Transform algorithm to rival (or even surpass) the Fast Fourier Transform (FFT) algorithm for computing the DFT.

Wavelets are applied in removing high-frequency noise from signals and in trying to identify strong features of signals. Strong features (such as steep slopes or discontinuities) will normally result in significant magnitudes for the wavelet coefficients at that point for many different scales. Similarly high frequency noise will normally only produce significant wavelet coefficients at one scale at each position.

The most spectacular uses are in image analysis and the big success story of wavelets with respect to images was its adoption by the FBI as a standard for compressing digital fingerprint images. An image (in one colour or grayscale) may be regarded as a function of a position in the planar image, with the value of the function being the brightness or intensity of the image at that point. Using 2-dimensional wavelets, one can express the function by its wavelet coefficients.

To compress the image, one discards some of the smaller coefficients and stores or transmits only the 'important' ones. If this is done in a clever way and if one uses a cleverly selected set of wavelet coefficients (c_m), then high compression rates can be obtained without much loss of quality.

Apart from the fingerprints, there is a new JPEG2000 standard that is intended to supplant the JPEG version now used for storing many images on computers. (JPEG is based on an algorithm that uses the Discrete Cosine Transform, a variant of the DFT).

On a more mathematical front there are orthonormal wavelet bases that form unconditional bases for Banach spaces of functions such as $L^p(\mathbb{R})$ and $L^p[0,1]$ ($1 < p < \infty$). See [6]. For the Fourier series case, we get only a conditional basis for $L^p[0,1]$ ($1 < p < \infty, p \neq 2$).

Wavelets have certain apparent disadvantages compared to Fourier series and transforms. The wavelets are not eigenfunctions for differentiation and they do not behave especially well with respect to convolutions. However, some results remain valid that seem to suggest that differentiation is 'almost diagonal' when expressed in a wavelet basis. See the characterisations of Sobolev spaces in terms of coefficients with respect to a wavelet basis ([6]).

The number of potential applications of wavelets in signal processing, image analysis and compression, sound wave analysis, numerical solution of differential equations and noise reduction is very large and there is a vast literature on wavelets. Not all potential applications seem to have succeeded, possibly because the right approach or the right wavelet has not been found. Another explanation is that wavelet-like techniques were already standard techniques in some fields like recognition of shapes in digital images, even before the term 'wavelet' was identified around 1988.

Current research on applications of wavelets is very active and, on the theoretical front, there are various more complex variants of wavelets being investigated. Also the n-dimensional versions of wavelets (associated with matrix dilations) seem to pose some unresolved problems.

References

[1] Burke Hubbard, B., The world according to Wavelets: The stroy of a Mathematical Technique in the Making (2nd ed.), A. K. Peters, Natick, Massachusetts (1998).

[2] Chui, C., Wavelets: a mathematical tool for signal processing, SIAM, Philadelphia, PA, 1997.

[3] Strang, G. and Nguyen, T., Wavelets and filter banks, Wellesley-Cambridge Press, Wellesley, MA, 1996.

[4] G. Strang, *Wavelet transforms versus Fourier transforms*, Bull. Amer. Math. Soc. **28** (1993) 288–305.

[5] B. Vidaković and P. Müller, *Wavelets for kids*, `ftp://ftp.isds.duke.edu/pub/Users/brani/papers/wav4kidsA.ps.Z`

[6] Wojtaszczyk, P., A mathematical introduction to wavelets. London Mathematical Society Student Texts, 37, Cambridge University Press, Cambridge, 1997.

Contents