Math 6610

- Classic Tool for periodic processes.
  
  Fourier Series
  
  $f$ 2π periodic
  
  $f(x) \approx F_n(x) = \frac{a_0}{2} + \sum_{k=1}^{n} a_k \cos kx + \sum_{k=1}^{n} b_k \sin kx$

- The Fourier coefficients $a_k$ and $b_k$ are determined by the requirement
  
  $\min_{n} \int_{-\pi}^{\pi} (f(x) - F_n(x))^2 \, dx$

- This is least squares approximation with respect to the inner product
  
  $(f, g) = \int_{-\pi}^{\pi} f(x) g(x) \, dx$

- It turns out (easy exercise) that the basis functions $\cos kx$ and $\sin kx$ are orthogonal with respect to $\langle \cdot, \cdot \rangle$ and the coefficients are
  
  $a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos kx \, dx \quad k = 0, 1, 2, \ldots$

  $b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin kx \, dx \quad k = 1, 2, \ldots$
- Drawback of Fourier Series: action in a small interval of time affects all Fourier coefficients.
- e.g., square wave
  \[
  \frac{1}{k}
  \]
  coefficients decline like \( \frac{1}{k} \)
- also: Gibbs Phenomenon
- This can be overcome in many ways, including various modifications of Fourier Series
- big subject, but we'll skip it
- Another, more recent, approach is based on \underline{wavelets}
- particularly in signal processing and computer graphics.
- Good first introduction:
  Ingrid Daubechies, Ten Lectures on Wavelets, SIAM publication, 1992,
- we'll just look at one example, the Haar wavelet

- you want to look for:
  - use of many time scales (multiresolution analysis)
  - local support of basis functions
  - orthogonality.

- The "mother wavelet"

\[ \psi(x) = \begin{cases} 
1 & 0 \leq x < \frac{1}{2} \\
-1 & \frac{1}{2} \leq x < 1 \\
0 & \text{else}
\end{cases} \]

- Note local support.

- This corresponds to the basic sine function

inner product \( (f, g) = \int f(x)g(x)dx \)
- we'll use translations and dilations of \( \psi \)

\[
\psi_{m,n}(x) = 2^{-m/2} \psi(2^{-m}x - n)
\]

\( m, n \) integer

- get a bi-infinite family

- support of \( \psi_{m,0} \) is \([0, 2^m]\)

" " \( \psi_{m,n} \) is \([n2^m, (n+1)2^m]\)

- Basic facts

\[
\int \psi_{m,n}^2(x) \, dx = 1 \quad \int \psi_{m,n} = 0
\]

\[
\int \psi_{m,n}^* (x) \psi_{m^*,n^*}(x) \, dx = 0 \quad \text{unless} \quad m=m^* \text{ and } n=n^*
\]

- most surprising

Any \( L^2 \) function \( f \) \((\int f(x)^2 \, dx < \infty)\) can be arbitrarily well approximated by a linear combination of the \( \psi_{m,n} \)

- clearly, changing \( f \) at a point affects only terms whose support includes that point.
we only use these terms whose scale is relevant to the current application \((m \text{ not too small})\)

- Orthornormality: 
  \[ S = S^\omega \]
  \[ S \psi_{mn} \psi_{mn}^* = 0 \quad \text{if} \quad n \neq n^* \]
  since the supports of \(\psi_{mn}\) and \(\psi_{mn}^*\) do not overlap.

suppose now we have two functions \(\psi_{mn}\) and \(\psi_{m*n*}\) and \(m^* + n^*\)

- assume \(m < m^* \Rightarrow\) The support of \(\psi_{m*n*}\) lies entirely in a region where \(\psi_{mn}\) is constant

\[ \Rightarrow S \psi_{mn} \psi_{m*n*}^* = 0 \]
How can we approximate a function \( f \) (which may be everywhere positive, for example) by a linear combination of functions whose average is zero?

- Most amazing!

- Suppose \( \int f^2 < \infty \)

- \( f \) can be arbitrarily well approximated (in the norm \( \| f \| = \| f^2 \|^{1/2} \)) by a piecewise constant function.

- So suppose \( f \) is constant on each

\[
[l 2^{-j}, (l+1)2^{-j}]
\]

and its support is

\[
[-2^{j}, 2^{j}]
\]

- Example: Floating point arithmetic, computer graphics.

\( j \) and \( j_0 \) can be arbitrarily large, but finite.
we know construct a sequence of functions as follows:

\[ f^0 = f \]

\[ f^0 = f' + (f^0 - f) = f' + \delta' \]

where \( f' \) is obtained from \( f^0 \) by averaging over groups of 2 neighbouring intervals.

\[ \delta' \]

\[ \delta' \text{ is a linear combination of Haar wavelets!} \]

\[ \delta' \text{ is piecewise constant over the same intervals as } f^0. \]

\[ f^0 = f' + \sum_{k=-\infty}^{\infty} c_{-j_0+k} e^{i\pi(j_0 + k)x} \]
- Naturally, apply the same trick to \( f' \), etc.

- Eventually we get

\[
f^0 = f^{x_0 + x_1} + \sum_{m=-J_0+1}^{J_0} \sum_{l} \gamma_{m,l} \psi_{m,l}
\]

- \( f^{x_0 + x_1} \) consists of two constant pieces.

- However, we can continue, just widen the support!

\[
f^0 = f^{x_0 + x_1 + k} + \sum_{m=-J_0+1}^{J_0} \sum_{l} \psi_{m,l}
\]

- Thus

\[
f - \sum \sum \gamma_{m,l} \psi_{m,l} = f^{x_0 + x_1 + k}
\]

- What about \( \| f^{x_0 + x_1 + k} \| = \sqrt{\sum (f^{x_0 + x_1 + k})^2} \)
- when we average a constant piece with zero we halve the function value and double the support.
- The integral over the square is thereby halved.
- Every time we extend the domain we divide by 2

\[
\begin{align*}
S A^2 &= 5 A^2 \\
S \left( A / 2 \right)^2 &= 2S \left( \frac{A^2}{4} \right) = \frac{5A^2}{2}
\end{align*}
\]
function. This redundancy can be exploited (it is, for instance, possible to compute the wavelet transform only approximately, while still obtaining reconstruction of \( f \) with good precision), or eliminated to reduce the transform to its bare essentials (such as in the image compression work of Mallat and Zhong (1992)). It is in this discrete form that the wavelet transform is closest to the "\( \psi \)-transform" of Frazier and Jawerth (1988).

The choice of the wavelet \( \psi \) used in the continuous wavelet transform or in frames of discretely labelled families of wavelets is essentially only restricted by the requirement that \( C_\psi \), as defined by (1.3.2), is finite. For practical reasons, one usually chooses \( \psi \) so that it is well concentrated in both the time and the frequency domain, but this still leaves a lot of freedom. In the next section we will see how giving up most of this freedom allows us to build orthonormal bases of wavelets.

1.3.3. Orthonormal wavelet bases: Multiresolution analysis. For some very special choices of \( \psi \) and \( a_0, b_0 \), the \( \psi_{m,n} \) constitute an orthonormal basis for \( L^2(\mathbb{R}) \). In particular, if we choose \( a_0 = 2, b_0 = 1,2 \) then there exist \( \psi, \) with good time-frequency localization properties, such that the

\[
\psi_{m,n}(x) = 2^{-m/2} \psi(2^{-m}x - n)
\]

constitute an orthonormal basis for \( L^2(\mathbb{R}) \). (For the time being, and until Chapter 10, we restrict ourselves to \( a_0 = 2 \).) The oldest example of a function \( \psi \) for which the \( \psi_{m,n} \) defined by (1.3.4) constitute an orthonormal basis for \( L^2(\mathbb{R}) \) is the Haar function,

\[
\psi(x) = \begin{cases} 
1 & 0 \leq x < \frac{1}{2} \\
-1 & \frac{1}{2} \leq x < 1 \\
0 & \text{otherwise}
\end{cases}
\]

The Haar basis has been known since Haar (1910). Note that the Haar function does not have good time-frequency localization: its Fourier transform \( \hat{\psi}(\xi) \) decays like \( |\xi|^{-1} \) for \( \xi \to \infty \). Nevertheless we will use it here for illustration purposes. What follows is a proof that the Haar family does indeed constitute an orthonormal basis. This proof is different from the one in most textbooks; in fact, it will use multiresolution analysis as a tool.

In order to prove that the \( \psi_{m,n}(x) \) constitute an orthonormal basis, we need to establish that

1. the \( \psi_{m,n} \) are orthonormal;
2. any \( L^2 \)-function \( f \) can be approximated, up to arbitrarily small precision, by a finite linear combination of the \( \psi_{m,n} \).

Orthonormality is easy to establish. Since \( \text{support}(\psi_{m,n}) = [2^m n, 2^m(n+1)] \), it follows that two Haar wavelets of the same scale (same value of \( m \)) never overlap, so that \( \langle \psi_{m,n}, \psi_{m,n'} \rangle = \delta_{m,m'} \). Overlapping supports are possible if the two wavelets have different sizes, as in Figure 1.5. It is easy to check, however, that if \( m < m' \), then support \( (\psi_{m,n}) \) lies wholly within a region where \( \psi_{m',n'} \) is
possible to coming reconstruc-
tion to its bare hong (1992)). It is "φ-transform"
transform or in nly restricted by practical reasons,
the time and the next section we

analysis. For an orthonormal
then there exist ψ,

(1.3.4)
and until Chap-
function ψ for
basis for $L^2(\mathbb{R})$ is

at the Haar func-
er transform $ψ(ξ)$
are for illustration
indeed constitute
most textbooks; in

ly small precision,

$= [2^m n, 2^m(n+1)]$, value of m) never are possible if the
to check, however,
on where $ψ_{m,n'}$ is

constant (as on the figure). It follows that the inner product of $ψ_{m,n}$ and $ψ_{m',n'}$
is then proportional to the integral of $ψ$ itself, which is zero.

![Figure 1.5](image)

**Fig. 1.5.** Two Haar wavelets; the support of the "narrower" wavelet is completely contained in an interval where the "wider" wavelet is constant.

We concentrate now on how well an arbitrary function $f$ can be approximated by linear combinations of Haar wavelets. Any $f$ in $L^2(\mathbb{R})$ can be arbitrarily well approximated by a function with compact support which is piecewise constant on the $[2^{-j}(ℓ+1)-2^{-j}ℓ]$ (it suffices to take the support and $j$ large enough). We can therefore restrict ourselves to such piecewise constant functions only: assume $f$ to be supported on $[-2^{-j}ℓ, 2^{-j}ℓ]$, and to be piecewise constant on the $[2^{-j}ℓ, (ℓ+1)2^{-j}ℓ]$, where $J_1$ and $J_0$ can both be arbitrarily large (see Figure 1.6). Let us denote the constant value of $f^0 = f$ on $[2^{-j}ℓ, (ℓ+1)2^{-j}ℓ]$ by $f_1$. We now represent $f^0$ as a sum of two pieces, $f^0 = f^1 + δ^1$, where $f^1$ is an approximation to $f^0$ which is piecewise constant over intervals twice as large as originally, i.e., $f^1[2^{-j}ℓ, (ℓ+1)2^{-j}ℓ] = constant = f_1$. The values $f_1$ are given by the averages of the two corresponding constant values for $f^0$, $f_1 = \frac{1}{2}(f^0_{2ℓ} + f^0_{2ℓ+1})$ (see Figure 1.6). The function $δ^1$ is piecewise constant with the same stepwidth as $f^0$; one immediately has

$$δ_{2ℓ} = f^0_{2ℓ} - f_1 = \frac{1}{2}(f^0_{2ℓ} - f^0_{2ℓ+1})$$

and

$$δ_{2ℓ+1} = f^0_{2ℓ+1} - f_1 = \frac{1}{2}(f^0_{2ℓ+1} - f^0_{2ℓ}) = δ_{2ℓ}.$$  It follows that $δ^1$ is a linear combination of scaled and translated Haar functions:

$$δ^1 = \sum_{ℓ=2^{-j}J_0-1}^{2J_1+J_0-1} δ_{2ℓ}ψ(2^{-j}ℓ - ℓ).$$  

We have therefore written $f$ as

$$f = f^0 = f^1 + \sum_{ℓ} c_{-J_0+1, ℓ} ψ_{-J_0+1, ℓ},$$
where \( f^1 \) is of the same type as \( f^0 \), but with stepwidth twice as large. We can apply the same trick to \( f^1 \), so that

\[
f^1 = f^2 + \sum_{\ell} c_{-J_0+2,\ell} \psi_{-J_0+2,\ell},
\]

with \( f^2 \) still supported on \([-2^{J_1}, 2^{J_1}]\), but piecewise constant on the even larger intervals \([k2^{-J_0+2}, (k+1)2^{-J_0+2}]\). We can keep going like this, until we have

\[
f = f^{J_0+J_1} + \sum_{m=-J_0+1}^{J_1} \sum_{\ell} c_{m,\ell} \psi_{m,\ell}.
\]

Here \( f^{J_0+J_1} \) consists of two constant pieces (see Figure 1.7), with \( f^{J_0+J_1}|_{[0,2^{J_1}]} \equiv f_0^{J_0+J_1} \) equal to the average of \( f \) over \([0,2^{J_1}]\), and \( f^{J_0+J_1}|_{[-2^{J_1},0]} \equiv f_{-1}^{J_0+J_1} \) the average of \( f \) over \([-2^{J_1},0]\).

Even though we have "filled out" the whole support of \( f \), we can still keep going with our averaging trick: nothing stops us from widening our horizon from
THE WHAT, WHY, AND HOW OF WAVELETS

2^J_1 to 2^J_1+1, and writing \( f^{J_1+J_2} = f^{J_1+J_2+1} + \delta^{J_1+J_2+1} \), where

\[ f^{J_1+J_2+1}|_{[0,2^{J_1+1}]} = \frac{1}{2} f^{J_1+J_2}, \quad f^{J_1+J_2+1}|_{[-2^{J_1+1},0]} = \frac{1}{2} f^{J_1+J_2} \]

and

\[ \delta^{J_1+J_2} = \frac{1}{2} f_0^{J_1+J_2} \psi(2^{-J_1-1}x) - \frac{1}{2} f_{-1}^{J_1+J_2} \psi(2^{-J_1-1}x + 1) \]

(see Figure 1.7). This can again be repeated, leading to

\[ f = f^{J_0+J_1+K} + \sum_{m=-J_0+1}^{J_1+K} \sum_{\ell} c_{m,\ell} \psi_{m,\ell} \]

where support \( (f^{J_0+J_1+K}) = [-2^{J_1+K}, 2^{J_1+K}] \), and

\[ f^{J_0+J_1+K}|_{[0,2^{J_1+K}]} = 2^{-K} f_0^{J_0+J_1}, f^{J_0+J_1+K}|_{[-2^{J_1+K},0]} = 2^{-K} f_{-1}^{J_0+J_1}. \]

It follows immediately that

\[ \left\| f - \sum_{m=-J_0+1}^{J_1+K} \sum_{\ell} c_{m,\ell} \psi_{m,\ell} \right\|_{L^2}^2 = \left\| f^{J_0+J_1+K} \right\|_{L^2}^2 \]

\[ = 2^{-K/2} \cdot 2^{J_1/2} \left[ \left\| f_0^{J_0+J_1} \right\|^2 + \left\| f_{-1}^{J_0+J_1} \right\|^2 \right]^{1/2}, \]

which can be made arbitrarily small by taking sufficiently large \( K \). As claimed, \( f \) can therefore be approximated to arbitrary precision by a finite linear combination of Haar wavelets!

The argument we just saw has implicitly used a "multiresolution" approach: we have written successive coarser and coarser approximations to \( f \) (the \( f^j \),...
averaging \( f \) over larger and larger intervals), and at every step we have written the difference between the approximation with resolution \( 2^{j-1} \), and the next coarser level, with resolution \( 2^j \), as a linear combination of the \( \psi_{j,k} \). In fact, we have introduced a ladder of spaces \( (V_j)_{j \in \mathbb{Z}} \) representing the successive resolution levels: in this particular case, \( V_j = \{ f \in L^2(\mathbb{R}) ; f \text{ piecewise constant on the } [2^j k, 2^j (k+1)] , k \in \mathbb{Z} \} \). These spaces have the following properties:

\[
\begin{align*}
(1) & \quad \cdots \subset V_2 \subset V_1 \subset V_0 \subset V_{-1} \subset V_{-2} \subset \cdots ; \\
(2) & \quad \bigcap_{j \in \mathbb{Z}} V_j = \{ 0 \}, \quad \bigcup_{j \in \mathbb{Z}} V_j = L^2(\mathbb{R}); \\
(3) & \quad f \in V_j \iff f(2^j \cdot) \in V_0; \\
(4) & \quad f \in V_0 \rightarrow f(\cdot - n) \in V_0 \text{ for all } n \in \mathbb{Z}.
\end{align*}
\]

Property 3 expresses that all the spaces are scaled versions of one space (the "multiresolution" aspect). In the Haar example we found then that there exists a function \( \psi \) so that

\[
\text{Proj}_{V_{j-1}} f = \text{Proj}_{V_j} f + \sum_{k \in \mathbb{Z}} (f, \psi_{j,k}) \psi_{j,k} . \tag{1.3.5}
\]

The beauty of the multiresolution approach is that whenever a ladder of spaces \( V_j \) satisfies the four properties above, together with

\[
(5) \quad \exists \phi \in V_0 \text{ so that the } \phi_{0,n}(x) = \phi(x - n) \text{ constitute an orthonormal basis for } V_0,
\]

then there exists \( \psi \) so that (1.3.5) holds. (In the Haar example above, we can take \( \phi(x) = 1 \) if \( 0 \leq x < 1, \phi(x) = 0 \) otherwise.) The \( \psi_{j,k} \) constitute automatically an orthonormal basis. It turns out that there are many examples of such "multiresolution analysis ladders," corresponding to many examples of orthonormal wavelet bases. There exists an explicit recipe for the construction of \( \psi \): since \( \phi \in V_0 \subset V_{-1} \), and the \( \phi_{-1,n}(x) = \sqrt{2} \phi(2x - n) \) constitute an orthonormal basis for \( V_{-1} \) (by (3) and (5) above), there exist \( \alpha_n = \sqrt{2} \langle \phi, \phi_{-1,n} \rangle \) so that \( \phi(x) = \sum_n \alpha_n \phi(2x - n) \). It then suffices to take \( \psi(x) = \sum_n (-1)^n \alpha_{-n+1} \phi(2x - n) \). The function \( \phi \) is called a scaling function of the multiresolution analysis. The correspondence multiresolution analysis \( \rightarrow \) orthonormal basis of wavelets will be explained in detail in Chapter 5, and further explored in subsequent chapters. This multiresolution approach is also linked with subband filtering, as explained in §5.6 (Chapter 5).

Figure 1.8 shows some examples of pairs of functions \( \phi, \psi \) corresponding to different multiresolution analyses which we will encounter in later chapters. The Meyer wavelets (Chapters 4 and 5) have compactly supported Fourier transform; \( \phi \) and \( \psi \) themselves are infinitely supported; they are shown in Figure 1.8a. The Battle-Lemarié wavelets (Chapter 5) are spline functions (linear in Figure 1.8b, cubic in Figure 1.8c), with knots at \( \mathbb{Z} \) for \( \phi \), at \( \frac{1}{2}\mathbb{Z} \) for \( \psi \). Both \( \phi \) and \( \psi \) have infinite support, and exponential decay; their numerical decay is faster than for the Meyer wavelets (for comparison, the horizontal scale is the same in (a),
Rewritten:

The next fact, we re write the resolution of the spaces of orthonormal above, we consti- tute many many ex- pressions for the (2x - k) and exist to take action of this order linked ordering to- wards. The ansiform; 8. The are 1.8b, 1.8a. The 1.8b, 1.8a have on the contrary than in (a),

\[ \psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k), \quad j, k \in \mathbb{Z}, \]

constitutes an orthonormal basis of \( L^2(\mathbb{R}) \). The figure plots \( \psi \) (the associated scaling function) and \( \psi \) for different constructions which we will encounter in later chapters. (a) The Meyer wavelets; (b) and (c) Battle-Lemarié wavelets; (d) the Haar wavelet; (e) the next member of the family of compactly supported wavelets, \( 2\psi \); (f) another compactly supported wavelet, with less asymmetry.

**Fig. 1.8.** Some examples of orthonormal wavelet bases. For every \( \psi \) in this figure, the family \( \psi_{j,k}(x) = 2^{-j/2}\psi(2^{-j}x - k), \) \( j, k \in \mathbb{Z}, \) constitutes an orthonormal basis of \( L^2(\mathbb{R}) \). The figure plots \( \psi \) (the associated scaling function) and \( \psi \) for different constructions which we will encounter in later chapters. (a) The Meyer wavelets; (b) and (c) Battle-Lemarié wavelets; (d) the Haar wavelet; (e) the next member of the family of compactly supported wavelets, \( 2\psi \); (f) another compactly supported wavelet, with less asymmetry.
(b), and (c) of Figure 1.8). The Haar wavelet, in Figure 1.8d, has been known since 1910. It can be viewed as the smallest degree Battle–Lemarié wavelet ($\psi_{\text{Haar}} = \psi_{BL,0}$) or also as the first of a family of compactly supported wavelets constructed in Chapter 6, $\psi_{\text{Haar}} = 1\psi$. Figure 1.8e plots the next member of the family of compactly supported wavelets $N\psi$; $2\phi$ and $2\psi$ both have support width 3, and are continuous. In this family of $N\psi$ (constructed in §6.4), the regularity increases linearly with the support width (Chapter 7). Finally, Figure 1.8f shows another compactly supported wavelet, with support width 11, and less asymmetry (see Chapter 8).

Notes.

1. There exist other techniques for time-frequency localization than the windowed Fourier transform. A well-known example is the Wigner distribution. (See, e.g., Boashash (1990) for a good review on the use of the Wigner distribution for signal analysis.) The advantage of the Wigner distribution is that, unlike the windowed Fourier transform or the wavelet transform, it does not introduce a reference function (such as the window function, or the wavelet) against which the signal has to be integrated. The disadvantage is that the signal enters in the Wigner distribution in a quadratic rather than linear way, which is the cause of many interference phenomena. These may be useful in some applications, especially for, e.g., signals which have a very short time duration (an example is Janse and Kaiser (1983); Boashash (1990) contains references to many more examples); for signals which last for a longer time, they make the Wigner distribution less attractive. Flandrin (1989) shows how the absolute values of both the windowed Fourier transform and the wavelet transform of a function can also be obtained by “smoothing” its Wigner distribution in an appropriate way; the phase information is lost in this process however, and reconstruction is not possible any more.

2. The restriction $b_0 = 1$, corresponding to (1.3.4), is not very serious: if (1.3.4) provides an orthonormal basis, then so do the $\hat{\psi}_{m,n}(x) = 2^{-m/2} \psi(2^{-m}x - nb_0)$, with $\hat{\psi}(x) = |b_0|^{-1/2}\psi(b_0^{-1}x)$, where $b_0 \neq 0$ is arbitrary. The choice $a_0 = 2$ cannot be modified by scaling, and in fact $a_0$ cannot be chosen arbitrarily. The general construction of orthonormal bases we will expose here can be made to work for all rational choices for $a_0 > 1$, as shown in Auscher (1989), but the choice $a_0 = 2$ is the simplest. Different choices for $a_0$ correspond of course to different $\psi$. Although the constructive method for orthonormal wavelet bases, called multiresolution analysis, can work only if $a_0$ is rational, it is an open question whether there exist orthonormal wavelet bases (necessarily not associated with a multiresolution analysis), with good time-frequency localization, and with irrational $a_0$. 