Transform Methods & Signal Processing

lecture 09

Matthew Roughan
<matthew.roughan@adelaide.edu.au>

Discipline of Applied Mathematics
School of Mathematical Sciences
University of Adelaide

September 8, 2010

Wavelets

In previous lectures we saw that the STFT had problems. The Wavelet transform is the way to overcome these problems. One of the nicest aspects of wavelets is that they are so natural: they have been invented several times, each time from a different viewpoint, so we will consider several approaches that naturally result in a Wavelet transform, starting by extending our understanding of the uncertainty principle and Windowed Fourier Transforms.

This lecture introduces wavelets.

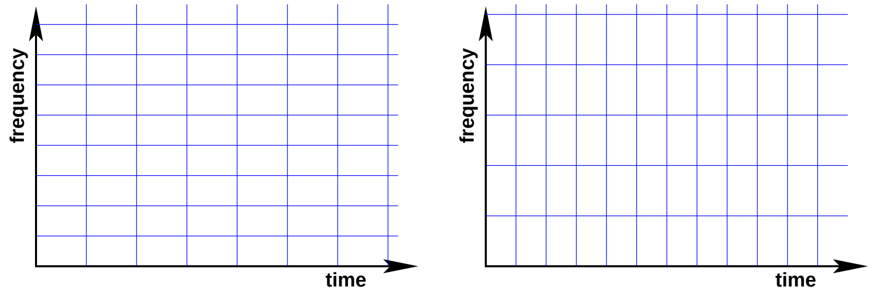
The main reference for this part of the course is Stéphane Mallat's book "A Wavelet Tour of Signal Processing", 2n edition, Academic Press, San Diego, 2001.

Limitations of the STFT

- computational cost $O(nm \log m)$
- time/frequency resolution tradeoff
  - small m better time, worse frequency resolution
- time/frequency resolution tradeoff is fixed
  - higher freq. can change faster than low freq.
  - appropriate resolution for each frequency?
- how can we do better?
  - some improvement might be gained through using better window functions (I have just used rectangular windows above)
  - lets try to get a more theoretical understanding of windows, and uncertainty bounds

Cutting up the time-frequency space

STFT partition of time-frequency



Areas of boxes don't get smaller!

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.3/71

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.4/71

Scaling property of FT

If we scale a function in time, then

$$\mathcal{F}\{f(at)\} = \frac{1}{a}F\left(\frac{s}{a}\right)$$

- Reciprocal scaling in each domain
- Tighter in Time, makes it looser in Fourier domain
- This contributes to uncertainty!!!!
  - in the STFT we use a window function to restrict the support of basis functions
  - tighter support on window function (less uncertainty in the time domain) results in a wider function in the frequency domain (and so more uncertainty there).

Heisenberg's Uncertainty Principle

Heisenberg's inequality is

$$\Delta x \Delta p \geq \frac{h}{2\pi}$$

where Δx and Δp are the unknown errors in position and momentum, respectively. It arises because, when one measures, say the location of a particle, one must bounce a photon on the particle. The impact of the photon changes the momentum of the particle by an unknown amount. One can reduce the energy of the photon to reduce the range of uncertainty in this change in momentum, but only by reducing the photon's frequency, thereby reducing the accuracy of the localization gained through the measurement.

Refresher on properties of the FT from lecture 2:

- Linearity: $af_1(t) + bf_2(t) \to aF_1(s) + bF_2(s)$
- Time shift: $f(t - t_0) \to F(s)e^{-i2\pi s t_0}$
- Time scaling: $f(at) \to \frac{1}{|a|}F\left(\frac{s}{a}\right)$
- Duality: $F(t) \to f(-s)$
- Frequency shift: $f(t)e^{-i2\pi s_0 t} \to F(s - s_0)$
- Convolution: $f_1(t) * f_2(t) \to F_1(s)F_2(s)$
- Differentiation I: $\frac{d^n}{dt^n} f(t) \to (i2\pi s)^n F(s)$
- Differentiation II: $(-i2\pi t)^n f(t) \to \frac{d^n}{ds^n} F(s)$
- Integration: $\int_{-\infty}^{t} f(s)ds \to \frac{1}{i2\pi s}F(s) + \pi F(0)\delta(s)$

Heisenberg's Uncertainty Principle led (in part) to the development of quantum mechanics, one of the most successful physics theories ever. Part of the theory is concerned with the dual nature of sub-atomic objects (electrons, photons, etc.) as both particles and waves. Waves relate this back to our course.

Uncertainty Principle

Given a transient signal $f(t)$, we want to localize this signal in time and frequency. We measure mean location of transient time and frequency by

$$
\begin{aligned}
u &= \frac{1}{\|f\|^2}\int_{-\infty}^{\infty} t|f(t)|^2\,dt \\
\xi &= \frac{1}{\|f\|^2}\int_{-\infty}^{\infty} s|F(s)|^2\,ds
\end{aligned}
$$

Measure uncertainties in time and frequency by variance about the mean, e.g.

$$
\begin{aligned}
\sigma_t^2 &= \frac{1}{\|f\|^2}\int_{-\infty}^{\infty} (t-u)^2|f(t)|^2\,dt \\
\sigma_s^2 &= \frac{1}{\|f\|^2}\int_{-\infty}^{\infty} (s-\xi)^2|F(s)|^2\,ds
\end{aligned}
$$

Uncertainty Principle

Theorem: For a function $f \in L^2$, the temporal and frequency variance satisfy

$$\sigma_t \sigma_s \geq \frac{1}{4\pi}$$

And this is an equality only if there exist $(u,\xi,a,b) \in \mathbb{R}^2 \times \mathbb{C}^2$ such that

$$f(t) = a e^{-b(t-u)^2} e^{i2\pi\xi t}$$

for which

$$
\begin{aligned}
\sigma_t^2 &= \frac{1}{4\pi b^2} \\
\sigma_s^2 &= \frac{b^2}{4\pi}
\end{aligned}
$$

Signals are not usually deltas, i.e., they have some extent in time and frequency. So we can't talk as if they are perfectly localized in either time or frequency. When we talk about location, we are talking about the mean location,. Remember that

$$\|f\|^2 = \int_{-\infty}^{\infty} |f(t)|^2\,dt = \|F\|^2 = \int_{-\infty}^{\infty} |F(t)|^2\,dt.$$

The uncertainty is not a measurement artifact – we can talk about uncertainty of a signal $f(t)$ without any randomness in the measurements. It is simple the fact that the signal is spread out in time (and/or frequency).

In particular, if you had two such signals that overlap, then the degree of overlap determines whether you can resolve them as separate signals. So uncertainty tells us something about resolution (in time and frequency).

Why is it important here? We will be using function as a basis in order to represent our signal. If the functions must satisfy the uncertainty principle, then so too must our representation.

Note we will be concerned with signals for which the above quantities are defined, and finite (i.e. signals that drop to zero "fast enough"). This is fair enough for transient signals.

Uncertainty Principle

Proof: It is sufficient to prove the theorem for f such that $u = \xi = 0$ as we can always perform shifts in time and frequency, e.g. by taking $\exp(i2\pi\xi t) f(t-u)$, to get the general case. In the case $u = \xi = 0$ we get

$$\sigma_t^2 \sigma_s^2 = \frac{1}{\|f\|^4} \int_{-\infty}^{\infty} t^2 |f(t)|^2 \, dt \int_{-\infty}^{\infty} s^2 |F(s)|^2 \, ds$$

Remember $\mathcal{F}\left\{\frac{df}{dt}\right\} = (i2\pi s) F(s)$, so Rayleigh's theorem implies

$$\int_{-\infty}^{\infty} |i2\pi s F(s)|^2 \, ds = 4\pi^2 \int_{-\infty}^{\infty} |f'(t)|^2 \, dt$$

Rayleigh's theorem says

$$\int_{-\infty}^{\infty} |F(s)|^2 \, ds = \int_{-\infty}^{\infty} |f(x)|^2 \, dx$$

Uncertainty Principle

Proof: Hence we can write

$$\sigma_t^2 \sigma_s^2 = \frac{1}{4\pi^2 \|f\|^4} \int_{-\infty}^{\infty} t^2 |f(t)|^2 \, dt \int_{-\infty}^{\infty} |f'(t)|^2 \, dt$$

Schwarz's inequality (for real functions)

$$\int_a^b \psi_1(x)^2 dx \int_a^b \psi_2(x)^2 dx \geq \left[\int_a^b \psi_1(x) \psi_2(x) \, dx \right]^2$$

with equality only if $\psi_2(x) = \alpha \psi_1(x)$ for some constant α.

$$\sigma_t^2 \sigma_s^2 \geq \frac{1}{4\pi^2 \|f\|^4} \left[\int_{-\infty}^{\infty} t f'(t) f(t) \, dt \right]^2$$

For Schwarz's inequality (sometimes called the Cauchy-Schwarz or Buniakowsky inequality) see Gradshteyn and Ryzhik, p.1099, or
`http://mathworld.wolfram.com/SchwarzsInequality.html`

A quick proof-sketch: if we integrate $[\psi_1(x) + t\psi_2(x)]^2$ the terms inside the integral are squared, and so non-negative, so the integral is non-negative, i.e.,

$$\int_a^b [\psi_1(x) + t\psi_2(x)]^2 \, dx \geq 0.$$

Expand the integral into its components and we get

$$\begin{aligned} \int_a^b [\psi_1(x) + t\psi_2(x)]^2 \, dx &= \int_a^b \psi_1^2(x) \, dx + 2t \int_a^b \psi_1(x)\psi_2(x) \, dx + t^2 \int_a^b \psi_2^2(x) \, dx \\ &= A + tB + t^2 C \geq 0 \end{aligned}$$

Now again the integrands of A and C are non-negative so $A, C \geq 0$. So the quadratic curve above has a minimum, which we know is greater than zero. A quadratic curve such as this has zeros if $B^2 - 4AC \geq 0$, so we know that $B^2 \leq 4AC$, and thence Schwarz's inequality, with equality only if $\psi_1(x) + t\psi_2(x) = 0$ for some value of t, for all x.

Uncertainty Principle

Proof: When $\psi_1(x)$ and $\psi_1(x)$ are complex, a more appropriate form of Schwarz's inequality is (from Bracewell, p.176) gives

$$4\int_a^b |\psi_1(x)|^2 dx \int_a^b |\psi_2(x)|^2 dx \geq \left[\int_a^b (\psi_1^*(x)\psi_2(x) + \psi_1(x)\psi_2^*(x))\, dx\right]^2$$

So

$$\begin{aligned} \sigma_t^2 \sigma_s^2 &\geq \frac{1}{16\pi^2 \|f\|^4}\left[\int_{-\infty}^{\infty} t\left(f'(t)f^*(t) + f^{*\prime}(t)f(t)\right) dt\right]^2 \\ &\geq \frac{1}{16\pi^2 \|f\|^4}\left[\int_{-\infty}^{\infty} t\frac{d}{dt}\left(f(t)f^*(t)\right) dt\right]^2 \\ &\geq \frac{1}{16\pi^2 \|f\|^4}\left[\left[t\,|f(t)|^2\right]_{-\infty}^{\infty} + \int_{-\infty}^{\infty} |f(t)|^2\, dt\right]^2 \end{aligned}$$

Uncertainty Principle

Proof: The Theorem holds for all $f \in L^2(\mathbb{R})$, but we are mainly interested in **transient** signals

- **transient** signals go to zero at some point
- lets have a fairly weak definition $\lim_{|t|\to\infty} \sqrt{t}\, f(t) = 0$
- in this case, the first term in the integration by parts is zero, so

$$\begin{aligned} \sigma_t^2 \sigma_s^2 &\geq \frac{1}{16\pi^2 \|f\|^4}\left[\int_{-\infty}^{\infty} |f(t)|^2\, dt\right]^2 \\ &\geq \frac{1}{16\pi^2 \|f\|^4}\left[\|f\|^2\right]^2 \\ &\geq \frac{1}{16\pi^2} \end{aligned}$$

The last step is the result of integration by parts.

Uncertainty Principle

Proof: To obtain an equality, note that Schwarz's inequality requires $\psi_2(x) = \alpha\psi_1(x)$ for some constant α, which in this case implies that

$$f'(t) = -2bt f(t)$$

which is true only for

$$f(t) = ae^{-bt^2}$$

This is the result for $(u,\xi) = (0,0)$. We perform a frequency and time translation to freq. ξ and time u to get

$$f(t) = ae^{-b(t-u)^2} e^{i2\pi\xi t}$$

$\square$

Gabor function

Definition: A Gabor function

$$f_{a,b,u,\xi}(t) = ae^{-b\pi(t-u)^2} e^{i2\pi\xi t}$$

It has FT

$$F_{a,b,u,\xi}(s) = \frac{a}{\sqrt{b}} e^{-\pi(s-\xi)^2/b} e^{-i2\pi su}$$

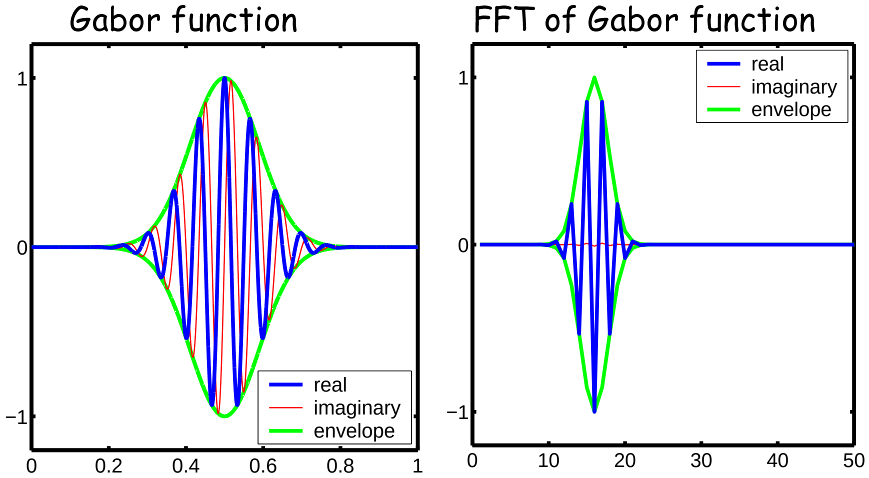
Mean position and frequency are u and ξ, and the uncertainty in location is

$$\begin{aligned}
\sigma_t^2 &= \frac{1}{\|f\|^2} \int_{-\infty}^{\infty} (t-u)^2 |f(t)|^2 \, dt = \frac{1}{b} \int_{-\infty}^{\infty} t^2 e^{-2b\pi t^2} \, dt = \frac{1}{4\pi b^2} \\
\sigma_s^2 &= \frac{1}{\|f\|^2} \int_{-\infty}^{\infty} (s-\xi)^2 |F(s)|^2 \, ds = \frac{1}{b} \int_{-\infty}^{\infty} t^2 e^{-2b\pi t^2} \, dt = \frac{b^2}{4\pi}
\end{aligned}$$

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.13/71

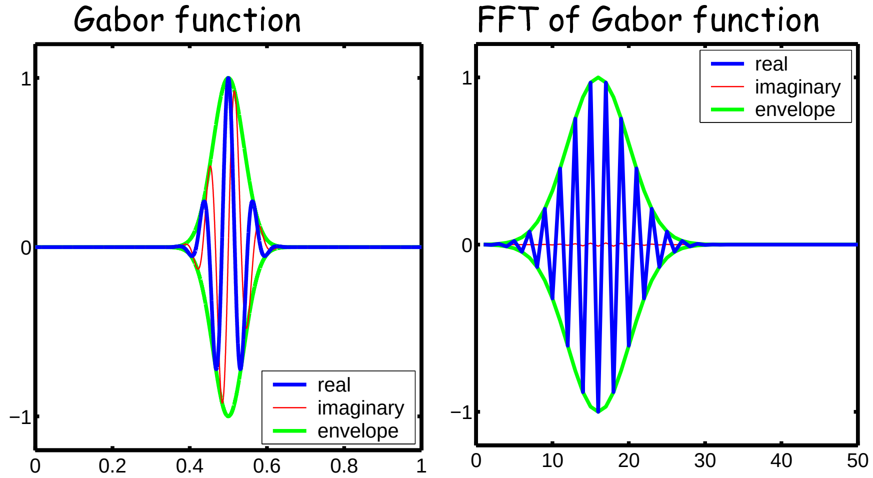
Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.14/71

Gabor function



Gabor function
= Gaussian window applied to a complex sinusoid

Gabor function

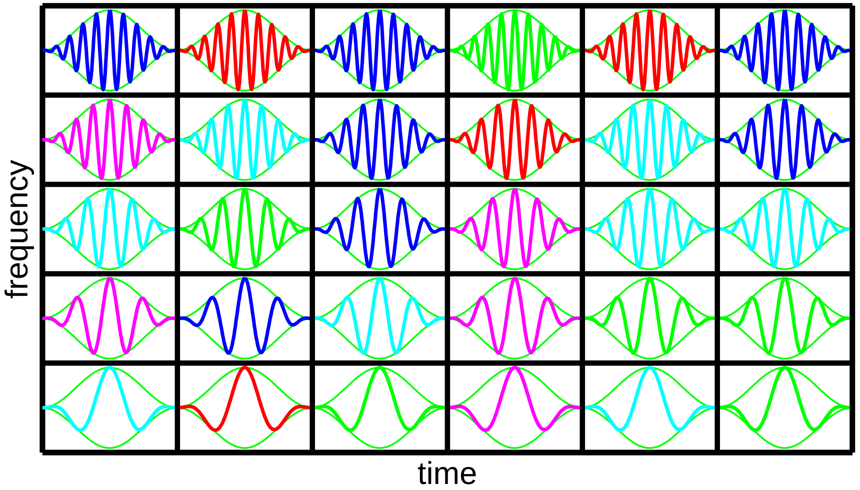


If we make the time-domain function narrower
the frequency domain function gets wider


```
function s = chirp(t, t_0, f_0, T, beta)
% chirp generates a Gaussian chirp signal
%    a special case is a Gabor function
%
% file:       chirp.m, (c) Matthew Roughan, Thu Aug 12 2004
% author:   Matthew Roughan
% email:    matthew.roughan@adelaide.edu.au
%
% see Bracewell, p. 135, 502
% and Mallat, p. 71, and 100
%
%    inputs:
%        t:      time points for chirp samples
%        t_0:    mid-pulse time        (t_0=u for a Gabor function)
%        f_0:    mid-pulse frequency   (f_0=xi for a Gabor function)
%        T:      Gaussian window width (T=1/b for a Gabor function)
%        2*beta: chirp sweep rate      (beta = 0 for a Gabor function)
%

s = exp(-pi*(t-t_0).^2/T) .* exp(i*2*pi*(f_0 * (t-t_0) + beta*(t-t_0).^2));
```

Cutting up the time-frequency space

Basis-like functions for a STFT with a window function



An alternative

Remember that we can scale window functions to change the resolution in time and frequency.

- higher frequencies can change more quickly
- why not change frequency resolution to match the frequency?
- just have to make the window width a function of frequency
- e.g. for the Gabor functions $f(t) = ae^{-b\pi(t-u)^2} e^{i2\pi\xi t}$ make the window frequency dependent by making b a function of ξ
  - higher frequencies make the window narrower
  - so for larger ξ we want smaller b.

Note that the colors in the plot are only there to help distinguish the different functions.

Note that the functions are not really a basis, because we have not shown that they are linearly independent, or that all possible functions can be represented. It is perhaps better to think of the functions as atoms which combined form a Dictionary, which we can use to describe other functions.

Wavelets

Wavelet are the natural result of this idea.

- start with a function we call the **Mother Wavelet**
  - e.g. a rectangular pulse, or a Gabor function
  - denote by $\psi(t)$
  - require $\psi \in L^2$, $\|\psi\| = 1$ and $\int_{-\infty}^{\infty} \psi(t)\, dt = 0$
- construct a set of atomic functions $\psi_{u,s}$ (atoms) from this function by
  - dilation (stretching and shrinking by s)
  - translation (shifting in time by u)

$$\psi_{u,s}(t) = \frac{1}{\sqrt{s}} \psi\left(\frac{t-u}{s}\right)$$

- e.g. could generate any Gabor function this way

Definition: Atoms

Time-frequency **atoms** $\{\phi_\gamma\}$, underly many transforms

- $\phi_\gamma \in L^2$
- $\|\phi_\gamma\| = 1$
- Transform $F(\gamma) = \langle f(t), \phi_\gamma(t) \rangle$

For example the STFT

$$\phi_\gamma(t) = g_{\xi,u}(t) = e^{-i2\pi\xi t} g(t-u)$$

where $g(t)$ is the (suitably normalized) window function.

Continuous wavelet transform

Wavelet Transform (analysis)

$$\mathcal{W}\{f(t)\} = W_f(u,s) = \langle f, \psi_{u,s} \rangle = \int_{-\infty}^{\infty} f(t) \frac{1}{\sqrt{s}} \psi^* \left(\frac{t-u}{s} \right) dt$$

Wavelet Reconstruction (synthesis), choose a complete, orthogonal set of wavelets $\{\psi_{j,n}\}$, then

$$f = \sum_j \sum_n \langle f, \psi_{n,j} \rangle \, \psi_{n,j}$$

Similar to the generalized Fourier transform.

Wavelets

There are many possible Mother Wavelets

- Haar
- Daubechies
- Mexican hat
- Gabor
- ...

Each has slightly different properties – much the same as when we considered window functions.

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.21/71

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.22/71

Wavelet Example

Mexican Hat wavelets are given by the second derivative of a Gaussian function, e.g.

$$\psi(t) = \frac{2}{\pi^{1/4}\sqrt{3\sigma}} \left(\frac{t^2}{\sigma^2} - 1 \right) \exp\left(\frac{-t^2}{2\sigma^2} \right)$$

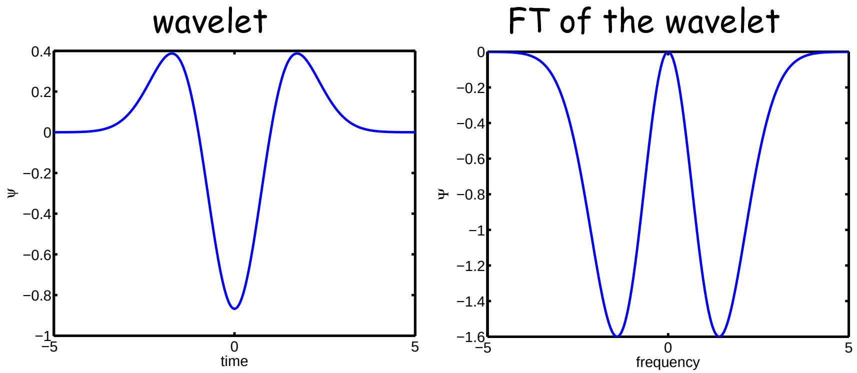
Its FT is

$$\Psi(\omega) = \frac{-\sqrt{8}\sigma^{5/2}\pi^{1/4}}{\sqrt{3}} \omega^2 \exp\left(\frac{-\sigma^2\omega^2}{2} \right)$$

where $\omega = 2\pi s$ is frequency in radians per time unit.

Wavelet Example

Mexican Hat wavelets



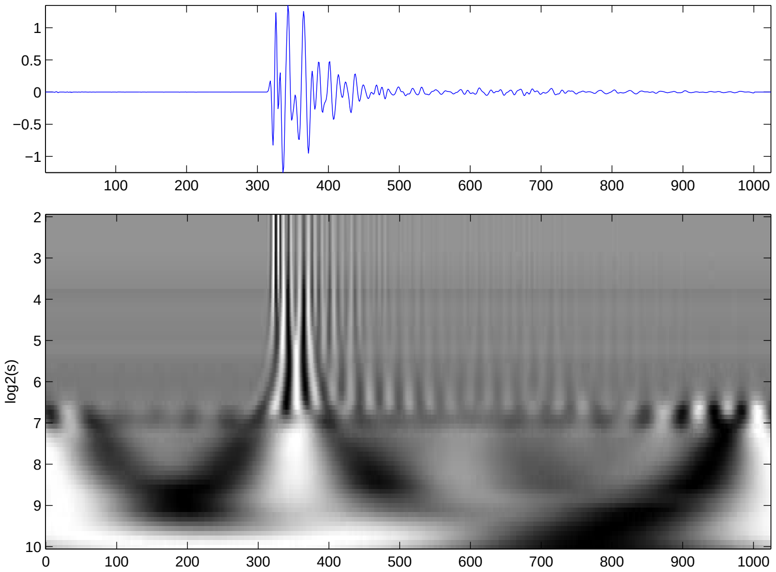
Once again, we would generate all other wavelets via a translation and dilation of this mother wavelet.

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.23/71

```
% file:       wavelets_mex_hat.m, (c) Matthew Roughan, Wed Aug  3 2005
%                 Mexican hat wavelets

t = -5:0.01:5;
w = -5:0.01:5;
sigma = 1;
psi = (2/(sqrt(sqrt(pi)*3*sigma))) * ...
      ((t.^2/sigma^2)-1) .* exp(-t.^2/(2*sigma^2));
Psi = -( sqrt(8)*sigma^(5/2)*pi^(1/4)/sqrt(3) ) * w.^2 .* ...
      exp(-sigma^2*w.^2/2);


figure(1)
hold off
plot(t, psi, 'linewidth', 3);
hold on
set(gca, 'linewidth', 3, 'fontsize', 18);
xlabel('time');
ylabel('\psi');
print('-depsc', 'Plots/wavelets_mexican_hat_psi.eps');

figure(2)
hold off
plot(t, Psi, 'linewidth', 3);
hold on
set(gca, 'linewidth', 3, 'fontsize', 18);
xlabel('frequency');
ylabel('\Psi');
print('-depsc', 'Plots/wavelets_mexican_hat_Psi.eps');
```

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.24/71

Example wavelet transform



Wavelet Basis

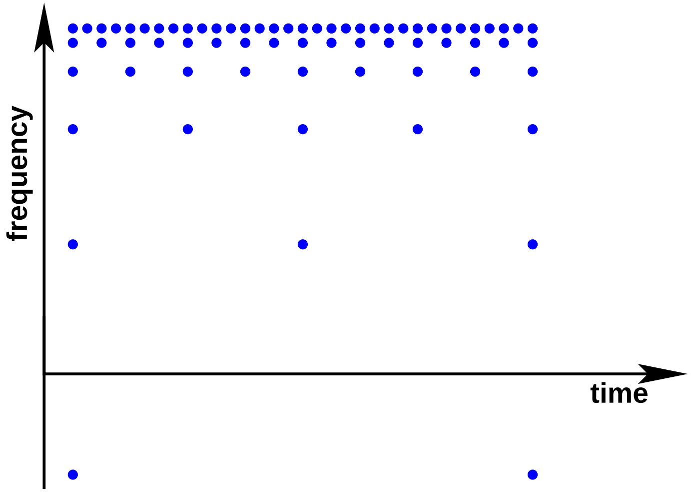
We don't need to consider all possible wavelet translations and dilations:

- We can think of the wavelet transform as a generalized FT
- So we want to find an orthogonal basis
- Also want time resolution tuned to frequency
- Choose a set of wavelets such that we get this
- Choose points on the **dyadic grid**

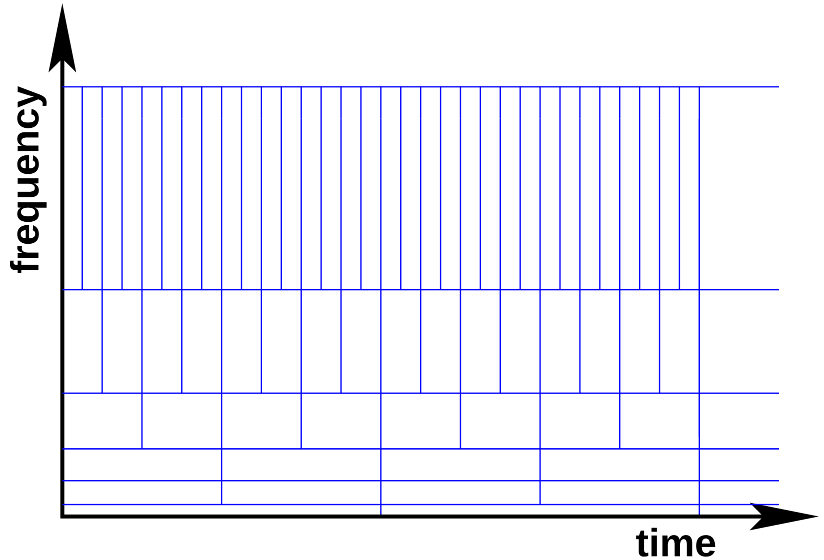
The figure is a reproduction of a figure from Mallat, p.81, using WaveLab `http://www-stat.stanford.edu/~wavelab/`, and in particular the tool WT-Browser. The signal is transformed using a large range of possible dilations and translations of the Mexican Hat wavelet.

Dyadic grid

Higher frequencies change more rapidly than low frequencies and so need to be sampled at a higher rate.



Wavelet Partition

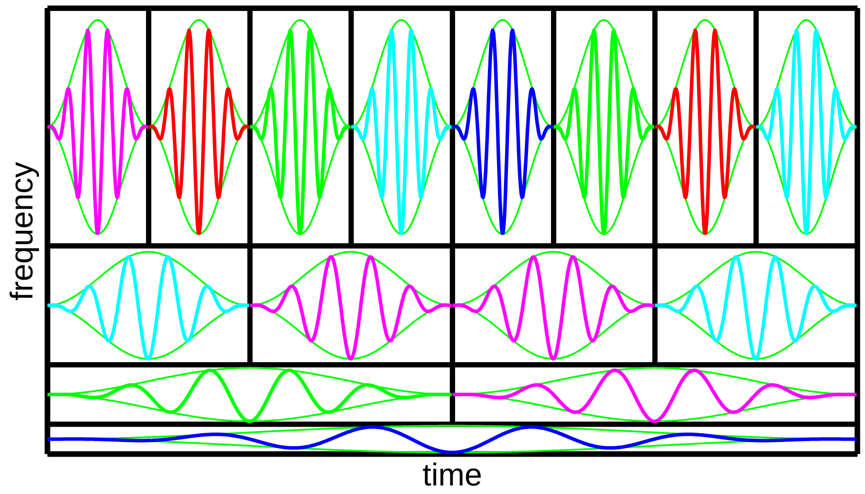


Blue dots indicate sample points within the time-frequency space.

Note that for high-frequencies we have lower frequency resolution, but better spacial resolution. Even so, the area of the rectangles is still constant.

Cutting up the time-frequency space

Basis functions for a wavelet(-like) transform



Note that the colors in the plot are only there to help distinguish the different functions.

This time, if we choose the mother-wavelet and sample points correctly, we can derive a set of basis functions for the space (though we haven't shown this yet).

Wavelet transforms

- Continuous Wavelet Transform (CWT) is the transform onto the whole space (u, s).
- Discrete Wavelet Transform (DWT) is the continuous transform, onto the discrete space given by the dyadic grid.
  - wavelet basis on dyadic grid defined by

$$\begin{aligned} s &= 2^j \\ u &= 2^j n \end{aligned}$$

where n and j are integers. So we get the basis

$$\psi_{n,j}(t) = \frac{1}{\sqrt{2^j}} \psi\left(\frac{t}{2^j} - n\right)$$

Note that the Discrete wavelet transform is still a continuous transform (it involves an integral over $\mathbb{R}$, but it maps to a discrete set of basis functions (indexed by (j, n) on the dyadic grid).

Scalogram

Take the power in each wavelet coefficient, e.g.

$$|W_f(u,s)|^2$$

and call this the **scalogram**

- analogous to periodogram (power of Fourier transform)
- analogous to spectrogram (power of STFT)

Time-Frequency Measurement

We can perform transform in either time or frequency domain

$$\mathcal{W}\{f\} = W_f(u,s) = \int_{-\infty}^{\infty} f(t)\psi_{u,s}^*(t)\,dt = \int_{-\infty}^{\infty} F(r)\Psi_{u,s}^*(r)\,dr$$

where $\Psi_{u,s}^*(r) = \mathcal{F}\left\{\psi_{u,s}^*(t)\right\}$

Note that

$$\Psi_{u,s}(r) = e^{-i2\pi ur}\sqrt{s}\Psi(sr)$$

using the scaling and time-translation properties.

$$\int_{-\infty}^{\infty} f(t)\psi_{u,s}^*(t)\,dt = \int_{-\infty}^{\infty} F(r)\Psi_{u,s}^*(r)\,dr$$

due to Plancheral's theorem (see Lecture 7).

Time-Frequency resolution

Time-frequency resolution of a wavelet

$$\mathcal{W}\{f(t)\} = W_f(u,s) = \langle f, \psi_{u,s}\rangle = \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{s}}\psi^*\left(\frac{t-u}{s}\right)\,dt$$

Suppose WLOG that ψ is centered at 0, which implies $\psi_{u,s}$ is centered at u, then

$$\int_{-\infty}^{\infty}(t-u)^2|\psi_{u,s}|^2\,dt = \int_{-\infty}^{\infty} t^2|\psi_{0,s}|^2\,dt = s^2\int_{-\infty}^{\infty} t^2|\psi(t)|^2\,dt = s^2\sigma_t^2$$

So the energy spread of a wavelet atom $\psi_{u,s}$ is a "box" $s\sigma_t$ wide in time.

- σ_t depends on the particular mother wavelet

Time-Frequency resolution

The FT of a wavelet is

$$\Psi_{u,s}(r) = e^{-i2\pi ur}\sqrt{s}\Psi(sr)$$

The center frequency is therefore η_ψ/s, where η_ψ is the center frequency of the mother wavelet.

- hence we call s the **scale**, and note that is it proportional to one over the frequency.
- the center frequency of the mother wavelet is given by

$$\eta_\psi = \int_{-\infty}^{\infty}\omega|\Psi(\omega)|^2\,d\omega$$

For instance we already know $\sigma_t = \frac{1}{2\sqrt{\pi b}}$ for the Gabor wavelet.

Time-Frequency resolution

The energy spread of the wavelet about the central frequency η_ψ / s is

$$\frac{1}{2\pi}\int_0^\infty \left(\omega - \frac{\eta}{s}\right)^2 |\Psi_{u,s}(\omega)|\, d\omega = \frac{\sigma_\omega^2}{s^2}$$

where

$$\sigma_\omega^2 = \frac{1}{2\pi}\int_0^\infty (\omega-\eta)^2 |\Psi(\omega)|\, d\omega$$

So the energy spread of a wavelet atom $\psi_{u,s}$ is a "box"

- $s\sigma_t$ wide in time (wider for lower frequencies)
- σ_ω / s in frequency (finer for lower freq.)

MultiResolution Approximation and Wavelets

Wavelets were independently invented from several different viewpoints. In this section we start by considering how we can approximate functions at different levels of detail, and by doing so come up again with the notion of wavelets.

The previous section presents wavelets from one point of view: as a better way of doing a STFT. Generating a set of atomic functions by scaling and translation is a very general approach, and by sampling these atomic function appropriately we create a representation in the time-frequency domain that adapts it resolution to the correct point in the plane.

However, wavelets were independently invented from several different viewpoints, and there is another one that provide a great deal of insight into wavelets, and in particular the "scaling function". We tackle this in this section.

MultiResolution Analysis

- a noted, we call s scale
- time-resolution at a particular scale s is fixed
- at different scales, the time resolution is proportional to the scale
- like observing the data at multiple scales
- hence the name **multiresolution analysis**
  - we can take this concept further by considering multiresolution approximation

Approximation

Definition: An approximation of a function $f \in L^2$ in subspace $\mathbf{V}$ is defined as the orthogonal projection of f onto $\mathbf{V}$ (e.g. the projection $\hat{f} \in \mathbf{V}$ that minimizes $\|f - \hat{f}\|$).

If an orthonormal basis $\{\phi_\gamma\}$ for $\mathbf{V}$ exists, then the projection into the space is given by

$$\hat{f} = \sum_\gamma \langle f, \phi_\gamma \rangle \phi_\gamma$$

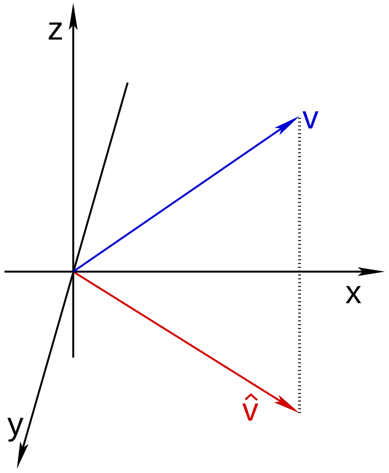
Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.37/71

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.38/71

Projection

Simple example of projection:

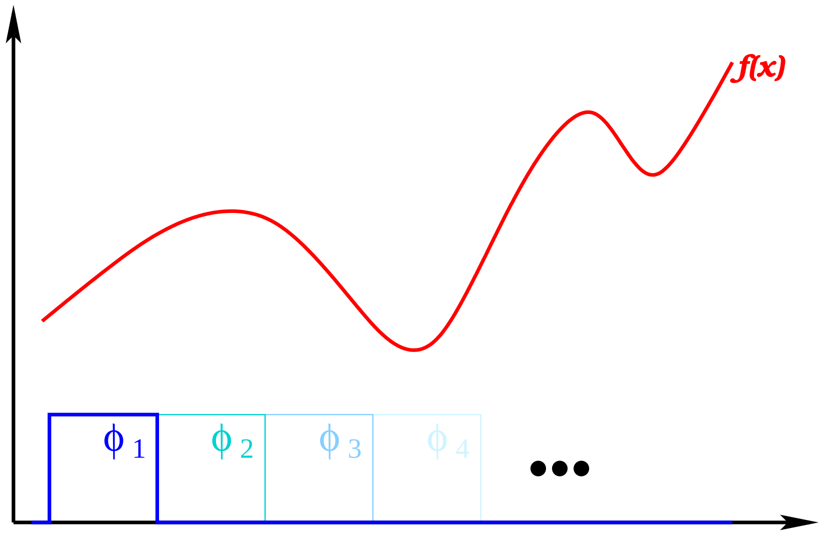
- projecting an (x,y,z) vector into the $x-y$ plane.
- vector $v \in \mathbb{R}^3$ is projected to $\hat{v} \in \mathbb{R}^2$
- take $(1,0,0)$ and $(0,1,0)$ as the basis vectors of the $x-y$ plane.
- inner product is just vector dot product



$$
\begin{aligned}
\hat{v} &= [v.(1,0,0)](1,0,0) + [v.(0,1,0)](0,1,0) \\
&= (v_1, v_2, 0)
\end{aligned}
$$

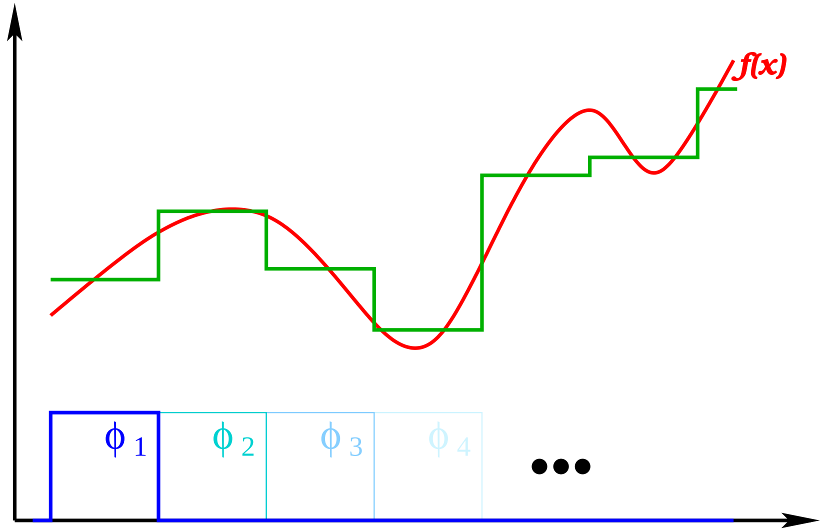
This example is from first year maths, but the idea of projection is much more general, and in our case we want to apply it to function spaces.

Approximation



The ϕ_i are our basis functions. They are simple rectangular pulses, translated along the x-axis. $f(x)$ is the function we wish to approximate.

Approximation



The approximation in terms of the rectangular pulses is obvious. The approximation can only be made of a linear combination of these rectangular functions. The functions do not overlap, so the resulting approximation will be a piecewise constant curve. Assume that we use the standard L^2 inner product

$$\langle f, g \rangle = \int_{-\infty}^{\infty} f(x) g(x)\, dx$$

If the basis functions are one unit wide, then

$$\langle f, \phi_i \rangle = \int_{-\infty}^{\infty} f(x)\phi_i(x)\, dx = \int_i^{i+1} f(x)\, dx$$

So the inner product is the average value of the function over the interval, $[i, i+1]$, which we will denote $bar\, f_i$ and the corresponding approximation is

$$\hat{f}(x) = \sum_i \bar{f}_i \phi_i(x)$$

It should be obvious that this function is piecewise constant, and its value on each interval $[i, i+1]$, is the mean of the function on that interval $\bar{f}_i$.

MultiResolution Approximation (MRA)

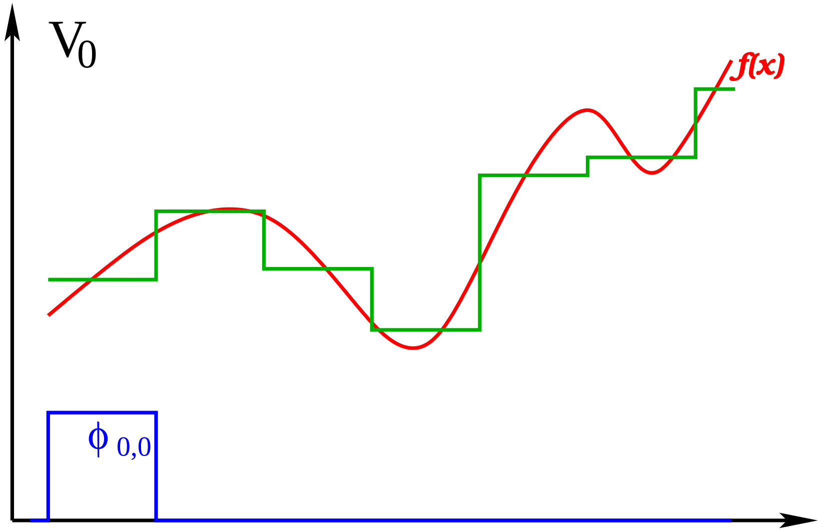
A sequence $\{\mathbf{V}_j\}_{j\in\mathbb{Z}}$ of closed subspaces of $L^2(\mathbb{R})$ is called a MultiResolution Approximation (MRA) if

1. $\mathbf{V}_{j+1} \subset \mathbf{V}_j$ for all $j \in \mathbb{Z}$
2. $f(t) \in \mathbf{V}_j \Leftrightarrow f(t - 2^j k) \in \mathbf{V}_j$ for all $j, k \in \mathbb{Z}$
3. $f(t) \in \mathbf{V}_j \Leftrightarrow f(t/2) \in \mathbf{V}_{j+1}$ for all $j, k \in \mathbb{Z}$
4. $\lim_{j\to\infty} \mathbf{V}_j = \{0\}$
5. $\lim_{j\to-\infty} \mathbf{V}_j = L^2(\mathbb{R})$
6. $\exists \theta$ such that $\{\theta(t-n)\}_{n\in\mathbb{Z}}$ is a Riesz basis of $\mathbf{V}_0$.

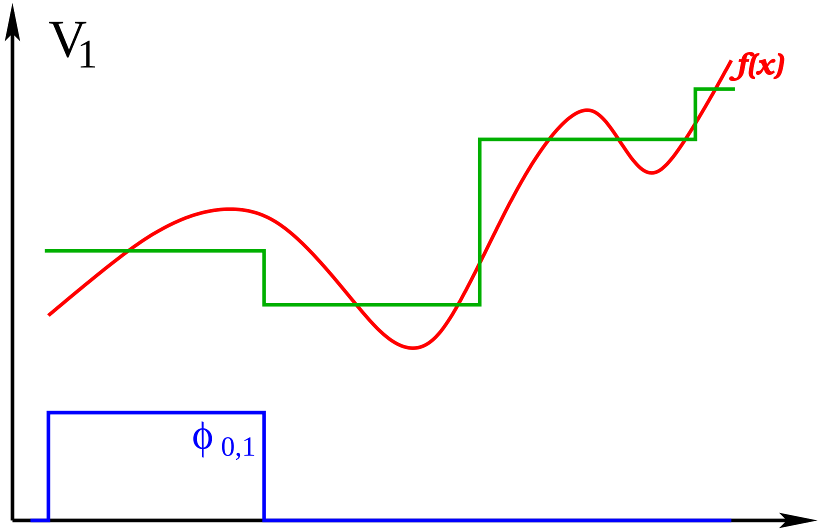
We can think of $\mathbf{V}_j$ grouping together the approximations at **scale** 2^j. Sometimes call j the **octave** (through analogy to music).

1. The approximation at octave j has all the information needed for the approximation at octave $j+1$, so anything we can represent in V_{j+1} will be possible to represent in V_j.
2. The approximation at octave j can be translated by an integer multiple of 2^j, and it will still be a valid approximation at octave j
3. Dilating a function in V_j by 2 puts it into a coarser resolution V_{j+1}.
4. When octave goes to ∞, we lose all details, and the only possible approximation is the zero function.
5. When octave goes to $-\infty$, we can represent any function in L^2, i.e. we can obtain an arbitrarily good level of detail in our approximations.
6. See appendices for definition of Riesz basis. We need a basis to make projection simple.

MRA example



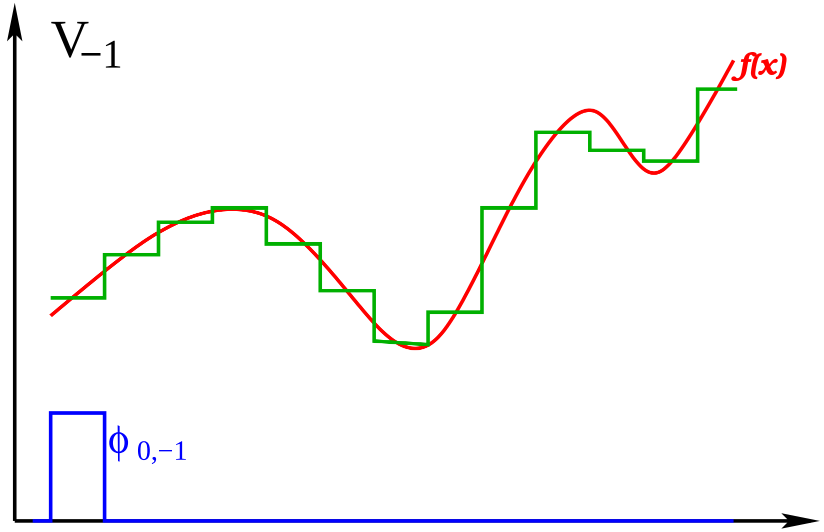
MRA example



Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.43/71

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.44/71

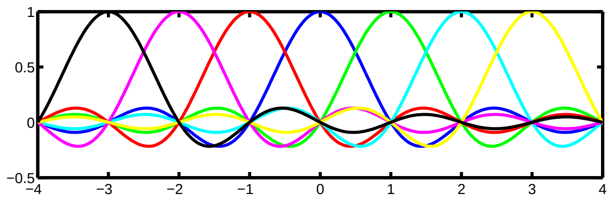
MRA example



MRA examples

Examples:

- **piecewise constant:** see above.
- **Shannon approximation:** using frequency band-limited functions (which hence must have infinite support in the time domain). Orthonormal basis $\mathrm{sinc}(t-n)$.



- **Spline approximation:**

MRA and scaling functions

From the Riesz basis $\exists\theta$ for the MRA, we can derive an orthonormal basis $\{\phi_{n,j}(t)\}_{n\in\mathbb{Z}}$ for $\mathbf{V}_j$. The functions ϕ are called **scaling functions**, and can be derived from a mother scaling function as with wavelets, e.g.

$$\phi_{n,j}(t) = \frac{1}{\sqrt{2^j}}\phi\left(\frac{t}{2^j} - n\right)$$

The approximation of a function $f \in L^2(\mathbb{R})$ is given by

$$\hat{f}_j(t) = \sum_{n\in\mathbb{Z}} \langle f, \phi_{n,j}\rangle \phi_{n,j}(t)$$

where

$$\langle f, \phi_{n,j}\rangle = \int_{-\infty}^{\infty} f(t)\phi_{n,j}(t)\, dt = \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{2^j}}\phi\left(\frac{t}{2^j} - n\right) dt = \left[f * \bar{\phi}_j\right](n)$$

The form of approximation

$$\hat{f}_j(t) = \sum_{n\in\mathbb{Z}} \langle f, \phi_{n,j}\rangle \phi_{n,j}(t)$$

is just the standard projection operation.

$\bar{\phi}_j$ is the time reversed version of ϕ_j, i.e.

$$\bar{\phi}_j(t) = \phi_j(-t)$$

So that

$$\begin{aligned} \langle f, \phi_{n,j}\rangle &= \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{2^j}}\phi\left(\frac{t}{2^j} - n\right) dt \\ &= \int_{-\infty}^{\infty} f(t)\frac{1}{\sqrt{2^j}}\bar{\phi}\left(n - \frac{t}{2^j}\right) dt \end{aligned}$$

which is just a standard convolution of $f(t)$ with the function $\bar{\phi}_j = \frac{1}{\sqrt{2^j}}\bar{\phi}\left(\frac{t}{2^j}\right)$, sampled at the points n.

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.47/71

The Approximation

The approximation of a function $f \in L^2(\mathbb{R})$ is given by

$$\hat{f}_j(t) = \sum_{n\in\mathbb{Z}} \langle f, \phi_{n,j}\rangle \phi_{n,j}(t) = \sum_{n\in\mathbb{Z}} a_j(n)\phi_{n,j}(t)$$

where $a_j(n) = \langle f, \phi_{n,j}\rangle = \left[f * \bar{\phi}_j\right](n)$

- frequency response of the approximation coefficients $a_j(n)$ depends on the frequency response of the scaling function
- scaling function typically a low-pass, so this becomes a low-frequency approximation.
- larger scale gives a coarse approx, so lower-freq.
- consistent with scaling law (as we dilate scaling function, the filter pass-band is reduced)

We can write approximation coefficients

$$a_j(n) = \langle f, \phi_{n,j}\rangle = \left[f * \bar{\phi}_j\right](n)$$

where $*$ is a generalization of the convolution operation, and $\bar{\phi}_j$ is the time reversed version of ϕ_j.

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.48/71

Relationship to wavelets

The approximation of a function $\hat{f}_j \in V_j$ into V_{j+1} is

$$\hat{f}_{j+1}(t) = \sum_{n\in\mathbb{Z}} \langle \hat{f}_j, \phi_{n,j} \rangle \phi_{n,j}(t)$$

- when we approximate a function $f \in V_j$ with a coarser approximation $f \in V_{j+1}$ we lose detail
- prefer a decomposition of V_j into an orthogonal sum of V_{j+1} and W_{j+1}
  - W_{j+1} are the bits we lost in the approximation
  - should be able to recombine V_{j+1} and W_{j+1} to get back to $f \in V_{j+1}$
- natural to associate W_{j+1} somehow with the wavelet

Some rough notes (I am not being very precise here, its just to give you some idea)

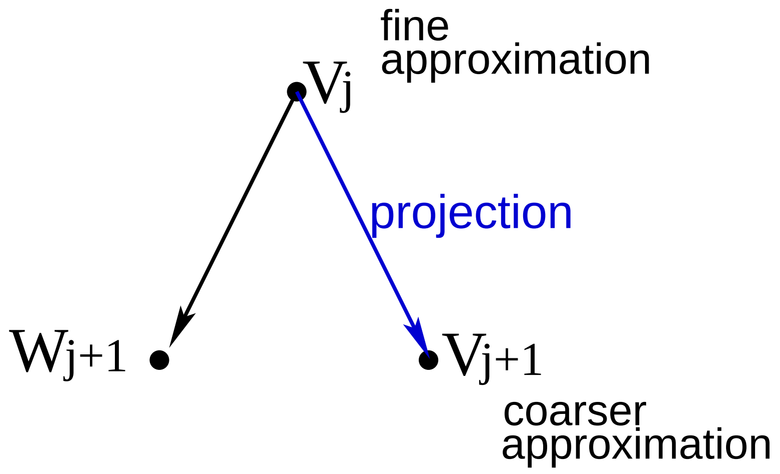
The direct sum of two subspaces (e.g. V_{j+1} and W_{j+1}) is often denoted $V_{j+1} \oplus W_{j+1}$, and implies that $V_{j+1} \cap W_{j+1} = \{0\}$, i.e., the intersection of the two sets is the zero element.

Assume that we have $V_j = V_{j+1} \oplus W_{j+1}$, and we have an (positive definite) inner product defined on V_j, then the orthogonal compliment of V_{j+1} is

$$V_{j+1}^{\perp} = \{ v \in V_j \,|\, \langle v, u \rangle = 0, \forall u \in V_{j+1} \}$$

Given V_j and its orthogonal compliment $W_{j+1} = V_{j+1}^{\perp}$ the space $V_j = V_{j+1} \oplus W_{j+1}$.

Relationship to wavelets



Projections from V_j into V_{j+1} and W_{j+1} work by

$$\begin{aligned}
\hat{f}_{j+1} &= \sum_{n\in\mathbb{Z}} \langle \hat{f}_j, \phi_{n,j+1} \rangle \phi_{n,j+1} \\
&= \sum_{n\in\mathbb{Z}} a_{n,j+1} \phi_{n,j+1} \\
\dot{f}_{j+1} &= \sum_{n\in\mathbb{Z}} \langle \hat{f}_j, \psi_{n,j+1} \rangle \psi_{n,j+1} \\
&= \sum_{n\in\mathbb{Z}} d_{n,j+1} \psi_{n,j+1}
\end{aligned}$$

where $\{\phi_{n,j+1}\}_{n\in\mathbb{Z}}$ and $\{\psi_{n,j+1}\}_{n\in\mathbb{Z}}$ are the respective bases for V_{j+1} and W_{j+1}, and $\hat{f}_{j+1}$ and $\dot{f}_{j+1}$ are the projections into these spaces.

Relationship to wavelets

Properties imposed by the relationship

1. $W_{j+1} \subset V_j$, so the basis vectors of W_{j+1} must be $\in V_j$.
   - we want the basis of W_{j+1} to be wavelets, so

$$\psi_{j+1} \in W_{j+1} \subset V_j$$

   - hence we can represent ψ_{j+1} in terms of ψ_j, i.e.,

$$\psi_{0,j+1}(t) = \sum_n a_j(n)\phi_{n,j}(t)$$

2. V_j is an orthogonal sum of V_{j+1} and W_{j+1}, so

$$\langle \phi_{0,j+1}(t), \psi_{n,j+1}(t) \rangle = 0$$

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.51/71

Relationship to wavelets

Take the properties above (for $j = 0$), and work out relationships between mother wavelet, and mother scaling function. First take the property that

$$\psi_{0,j+1}(t) = \sum_n a_j(n)\phi_{n,j}(t)$$

for $j = 0$

$$\psi_{0,1}(t) = \sum_n a_1(n)\phi_{n,0}(t) \quad (1)$$

$$\psi(t/2)/\sqrt{2} = \sum_n a_1(n)\phi(t-n) \quad (2)$$

$$\psi(t) = \sum_n a(n)\phi(2t-n) \quad (3)$$

(2) derives from the scaling relationships.

$$\psi_{n,j}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t}{2^j} - n\right)$$

$$\phi_{n,j}(t) = \frac{1}{\sqrt{2^j}}\phi\left(\frac{t}{2^j} - n\right)$$

(3) Note that $\hat{a}(n) = \sqrt{2}a_1(n)$, and we have substituted $t \to 2t$

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.52/71

Relationship to wavelets

Combining the first and second properties (from p.51)

$$\psi(t) = \sum_n a(n)\phi(2t-n)$$

$$\langle \psi(t), \phi(t-n) \rangle = \int_{-\infty}^{\infty} \psi(t)\phi(t-n)\,dt = 0$$

we get

$$\int_{-\infty}^{\infty} \sum_k a(k)\phi(2t-k)\phi(t-n)\,dt = \sum_k a(k) \int_{-\infty}^{\infty} \phi(2t-k)\phi(t-n)\,dt = 0$$

which defines possible values for $a(k)$

Example: Haar wavelets

Piecewise constant approximation: so take

$$\phi(t) = \begin{cases} 1 & \text{if } 0 \leq t \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

Basis functions for approximations are rectangular pulses.

$$\begin{aligned} \sum_k a(k) \int_{-\infty}^{\infty} \phi(2t-k)\phi(t-n)\,dt &= 0 \\ \sum_k a(k) \int_{n}^{n+1} \phi(2t-k)\,dt &= 0 \end{aligned}$$

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.53/71

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.54/71

Example: Haar wavelets

Now, $\phi(2t-k)$ is only positive in the interval $[n, n+1]$ for $k = 2n$ or $2n+1$

$$\begin{aligned} \sum_k a(k) \int_n^{n+1} \phi(2t-k)\, dt &= 0 \\ a(2n) + a(2n+1) &= 0 \end{aligned}$$

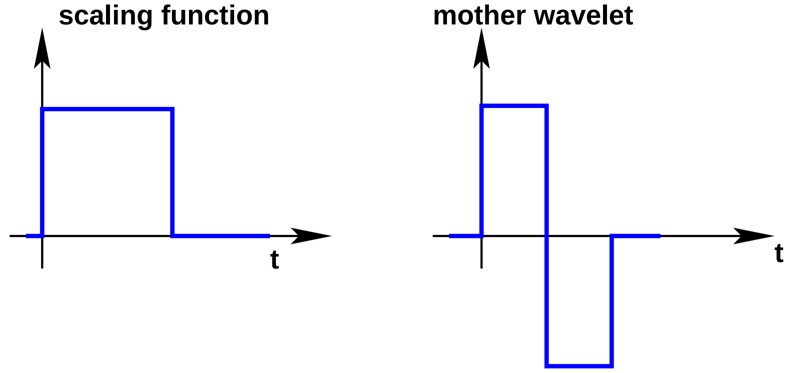
because in both cases the integral is 1.

The function with minimal support that satisfies this relationship has $a(0) = 1$ and $a(1) = -1$ and all other $a(k) = 0$, so

$$\psi(t) = \phi(2t) - \phi(2t-1)$$

Haar wavelets

Scaling and wavelet functions for the Haar transform shown below



Approximations are piecewise constant curves.

Remember

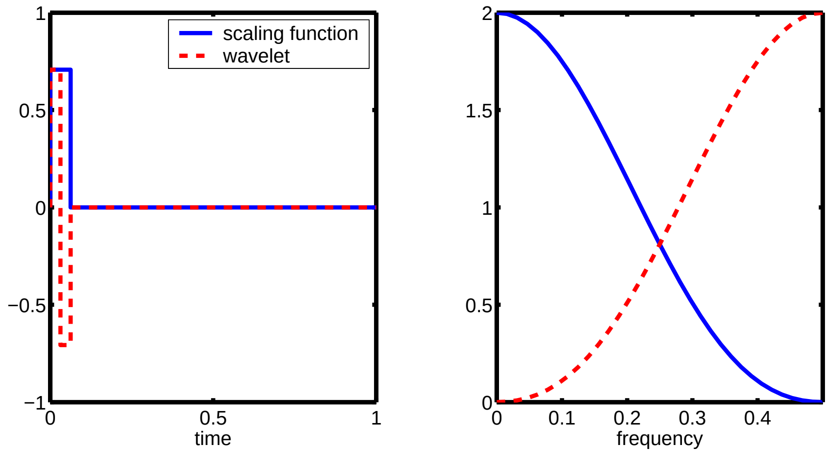
$$\psi(t) = \sum_n a(n)\phi(2t-n)$$

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.55/71

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.56/71

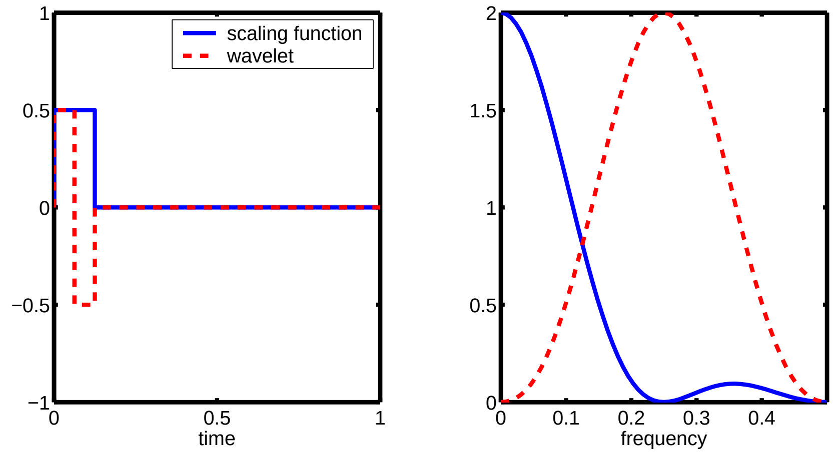
Haar wavelets: freq. representation

At scale $j=0$, scale by 2^0 ($\psi_{0,j}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t}{2^j}\right)$)



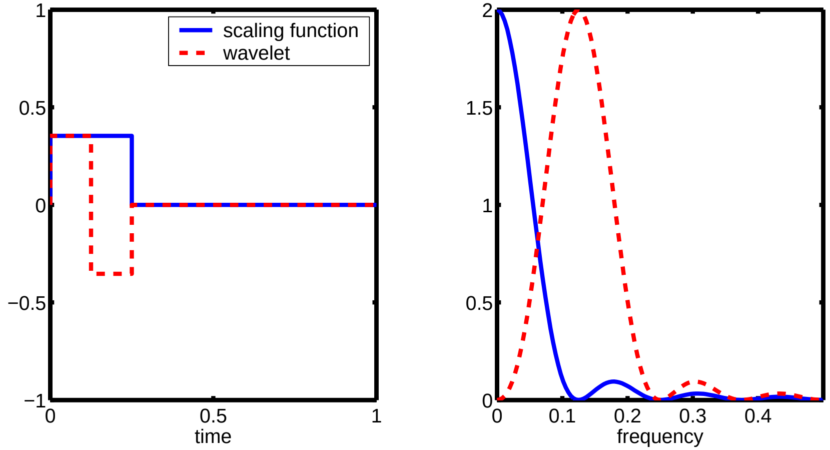
Haar wavelets: freq. representation

At scale $j=1$, scale by 2^1 ($\psi_{0,j}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t}{2^j}\right)$)



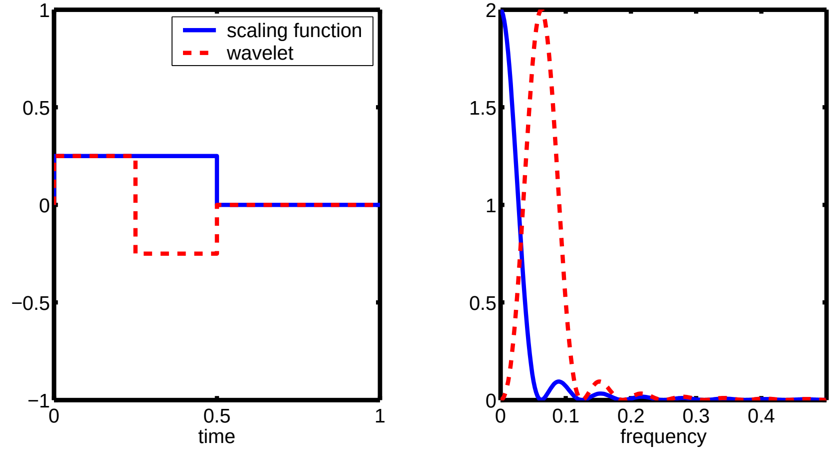
Haar wavelets: freq. representation

At scale $j = 2$, scale by 2^2 ($\psi_{0,j}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t}{2^j}\right)$)



Haar wavelets: freq. representation

At scale $j = 3$, scale by 2^3 ($\psi_{0,j}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t}{2^j}\right)$)

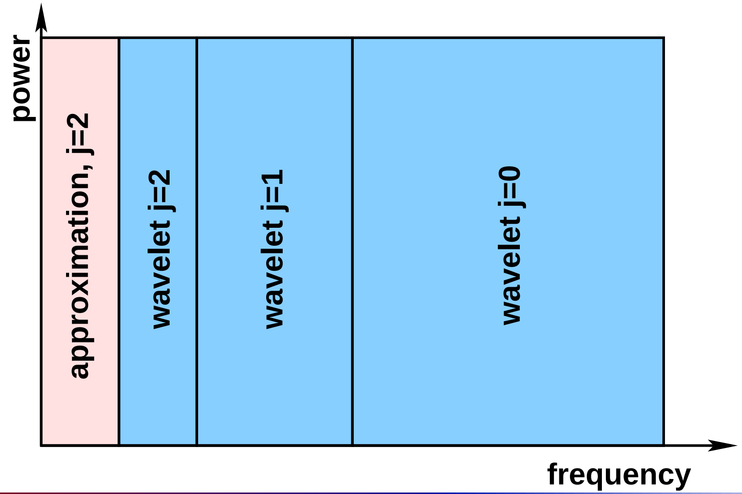


Haar wavelets: freq. representation

- scaling function is a low-pass
  - approximations are low-freq. approximations
  - larger scale, low-frequency stop-band
- wavelet function is a band-pass
  - together with scaling they break up a block of the frequency spectrum

Subband coding

The idea (looking across frequencies or scales) is that the transform breaks frequency spectrum into bands.



Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.61/71

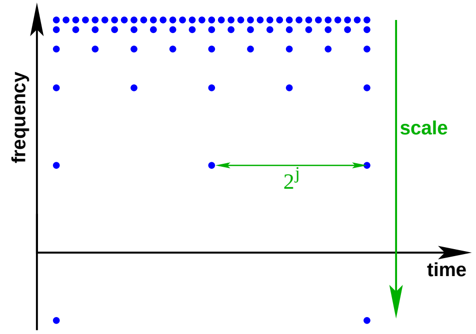
Sudband coding is yet another approach to derive a wavelet transform. We derive the subband characteristics of (Haar) wavelets here, rather than using it to derive wavelets, but we could have started with subband coding as our goal, and derived a wavelet transform.

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.62/71

MRA and wavelets

Take mother wavelet $\psi(t)$, with orthogonal discrete wavelet basis on the dyadic grid

$$\psi_{n,j}(t) = \frac{1}{\sqrt{2^j}}\psi\left(\frac{t}{2^j} - n\right)$$



Form closed subspaces

$$W_j = \mathrm{Sp}\{\psi_{n,j} | n \in \mathbb{Z}\}$$

As noted earlier,

$$V_j = \oplus_{i=j}^{\infty} W_i$$

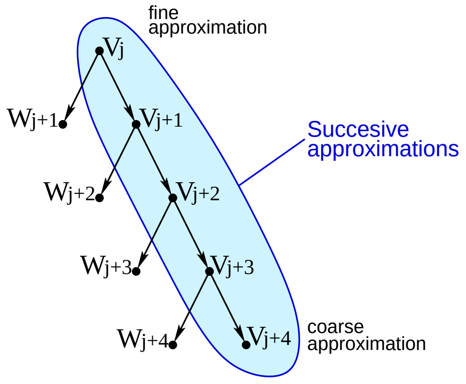
is a MRA and the scaling function ϕ was also given earlier, and $V_{j-1} = V_j \oplus W_j$ so an orthogonal projection into V_{j-1} can be decomposed into projections into V_j and W_j.

Successive decompositions

We can iteratively decompose approximation V_j into a wavelet part (the **details**) and a coarser scale **approximation** $V_{j-1} = V_j \oplus W_j$ using the projection operation

Form $f_{j-1} \in V_{j-1}$ by

$$\begin{aligned}
\hat{f}_{j+1} &= \sum_{n\in\mathbb{Z}} \langle \hat{f}_j, \phi_{n,j+1} \rangle \phi_{n,j+1} \\
&= \sum_{n\in\mathbb{Z}} a_{n,j+1} \phi_{n,j+1} \\
\dot{f}_{j+1} &= \sum_{n\in\mathbb{Z}} \langle \hat{f}_j, \psi_{n,j+1} \rangle \psi_{n,j+1} \\
&= \sum_{n\in\mathbb{Z}} d_{n,j+1} \psi_{n,j+1}
\end{aligned}$$



Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.63/71

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.64/71

MRA and wavelets

$$\begin{aligned}\hat{f}_j &= \hat{f}_{j+1} + \dot{f}_{j+1} \\ &= \sum_{n\in\mathbb{Z}} a_{n,j+1}\phi_{n,j+1} + \sum_{n\in\mathbb{Z}} d_{n,j+1}\psi_{n,j+1}\end{aligned}$$

- $\hat{f}_{j+1}$ is a coarser scale approximation of f
- it loses some "detail"
- **details** are captured in the wavelet component $\dot{f}_{j+1}$
- often call the coefficients
  - $a_{n,j}$ the approximation
  - $d_{n,j}$ the details
- As $j \to -\infty$ the approximation $\hat{f}_j \to f$

The Scaling Function

The above representation requires wavelet coefficients for $s = -\infty, \ldots, \infty$ and $u = -\infty, \ldots, \infty$. We can still manage if we have coefficients $\langle f, \psi_{u,s}\rangle$ for $s < s_0$, by using a **scaling function** $\phi(t)$.

- can be thought of as a low frequency (high scale) approximation of the signal
- form scaling functions $\phi_{u,s}(t)$ by the same dilations and translation used to form wavelets
- scaling function $\phi(t)$ brings in info from scales $s > 1$, so it is the aggregation of wavelets above this scale

$$|\Phi(\omega)|^2 = \int_1^\infty |\Psi(s\omega)|^2 \frac{1}{s}\, ds = \int_\omega^\infty |\Psi(\xi)|^2 \frac{1}{\xi}\, d\xi$$

The coefficients $a_{n,j}$ are often called the approximation, but remember the real approximating function is a linear combination of the basis functions, i.e.

$$\hat{f}_j = \sum_{n\in\mathbb{Z}} a_{n,j}\phi_{n,j}$$

The Scaling Function

- DWT representation

$$f = \sum_{j=j_0}^{\infty} \sum_{n=-\infty}^{\infty} \langle f, \psi_{n,j} \rangle \psi_{n,j} + \sum_{n=-\infty}^{\infty} \langle f, \phi_{n,j_0} \rangle \phi_{n,j_0}$$

Wavelet Properties

Potential wavelet properties

- finite support
- vanishing moments
- orthogonal/ bi-orthogonal
- complex(analytic) or real
- redundant (framelets)

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.67/71

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.68/71

Applications

- edge (and anomaly) detection
- motion detection
- denoising
- compression (JPEG 2000)

To do these, we will need to

- perform wavelet transforms on discrete data.
- make the algorithms efficient (as with FFT)

Appendices

Transform Methods & Signal Processing (APP MTH 4043): lecture 09 – p.69/71

Riesz basis

A family of elements $\{e_n\}_{n\in\mathbb{Z}}$ from a Hilbert space $\mathbf{H}$ is said to be a Riesz basis of $\mathbf{H}$ if it is linearly independent and there exists $A > 0$ and $B > 0$ such that for any $f \in \mathbf{H}$ one can find λ_n with

$$f(t) = \sum_{n=-\infty}^{\infty} \lambda_n e_n$$

which satisfies

$$\frac{1}{B}\|f\|^2 \leq \sum_{n=-\infty}^{\infty} |\lambda_n|^2 \leq \frac{1}{A}\|f\|^2$$

If $A = B$ the frame is said to be tight.