MGA Tutorial, September 08, 2004

Construction of Wavelets

Jan-Olov Strömberg

Department of Mathematics
Royal Institute of Technology (KTH)
Stockholm, Sweden

Department of Numerical Analysis and Computer Science (NADA)
Royal Institute of Technology (KTH)
Stockholm, Sweden

Outline of talk

- Introduction .
- Building blocks:Standard sampling, Fourier and Wavelets
- Some more Fourier Analysis. Time - Frequency plane
- More about building blocks: Orthonormal bases.
- The Continuous Wavelet Transform
- Discrete wavelet transform.
  - Desired properties of the wavelets.
  - Different approaches in their construction.
  - Multi-scale-analysis and bi-orthogonal bases, scaling equation
  - Lowpass, Highpass filter, Wavelet filter-tree.
- Wavelets basis in dimension 2.

- Wavelet packets filter-tree and wavelet packets library.
- Cost functions, Best basis and adaptiveness.
- Interpolets and sparse sampling in high dimension.

Introduction

Wavelets= small packet of waves

Theory of wavelet offspring from:

- Mathematics: Fourier analysis / Harmonic analysis
- Signal processing: Quadratic mirror filter

Wavelet theory 1985-

Example data on a Music CD: Use blackboard

Expansion of functions in trigonometric series by

Jean Baptiste Joseph Fourier (1768 - 1830) around 1807

. . . . . . but

use of approximation by trigonometric functions was used earlier by

Leonard Euler(1707-1783)

. . . . . . but

even earlier by

Daniel Bernuolli(1700 - 1783). “He showed that the movements of strings of musical instruments are composed of and infinite number of harmonic vibrations all superimposed on the string.” (late 1720th)

Building blocks of a signal

- Sampling of a signal:
  representation in standard basis
- Frequency description of the signal
  Fourier basis
- Wavelets an compromise between those two extremes.

Time - frequency plane

$$|f(t)|^2/\|f\|^2$$

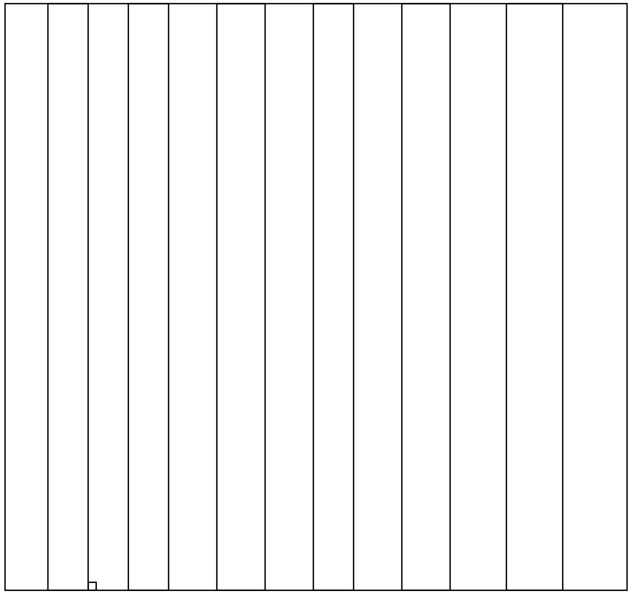
is the density distribution of function in time.

$$|\hat{f}(\omega)|^2/\|\hat{f}\|^2$$

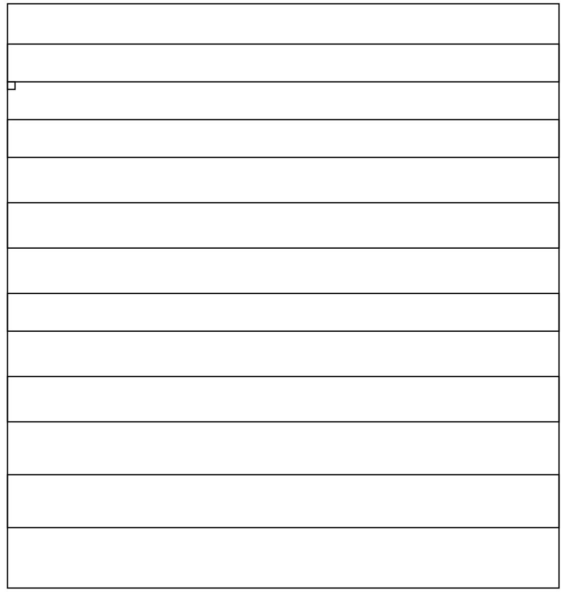
is the density distribution of function in frequency.

We will look at the product as a density distribution in the Time-Frequency plane.

Standard sampling and Fourier representation in TF- plane



Standard sampling



Fourier representation

Heisenberg uncertainty principle:

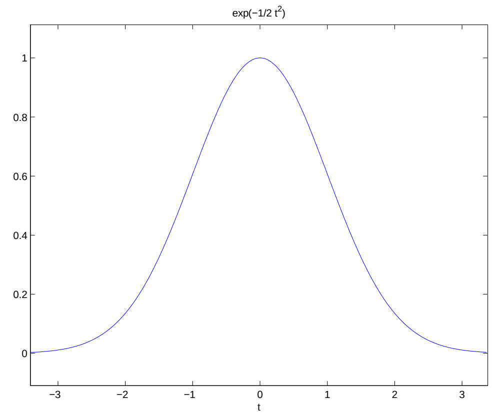
(Assume f is normalized: $\|f\| = 1$.)

Var(function time)$*$Var(function frequency) $\geq$ Constant

Minimum for Gaussian function (Normal distribution)

Function:

$$f(t) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}.$$



(continued)

Fourier transform:

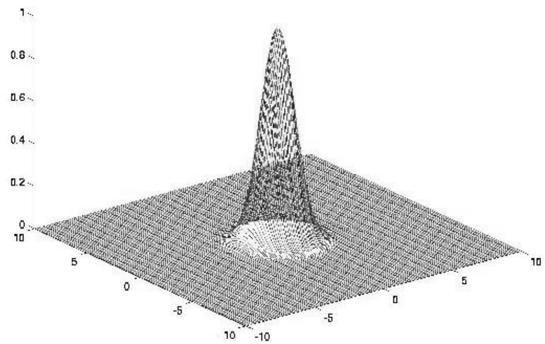
$$f(\omega) = \mathrm{e}^{-\omega^2/2}.$$

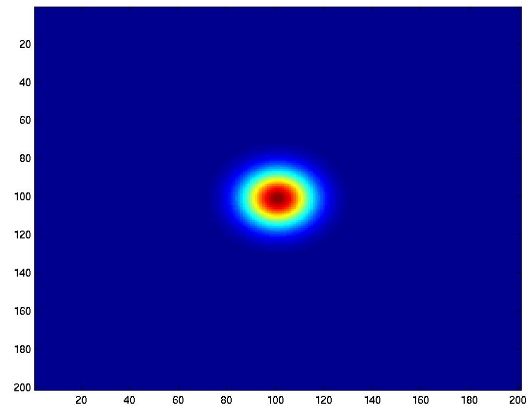
The Gaussian function in time-frequency plane

Function $f(t) = \mathrm{e}^{-t^2/2}$

Distribution function on time-frequency plane:

$$\rho_f(t,\omega) = \frac{1}{2\pi}\mathrm{e}^{-(t^2+\omega^2)}$$

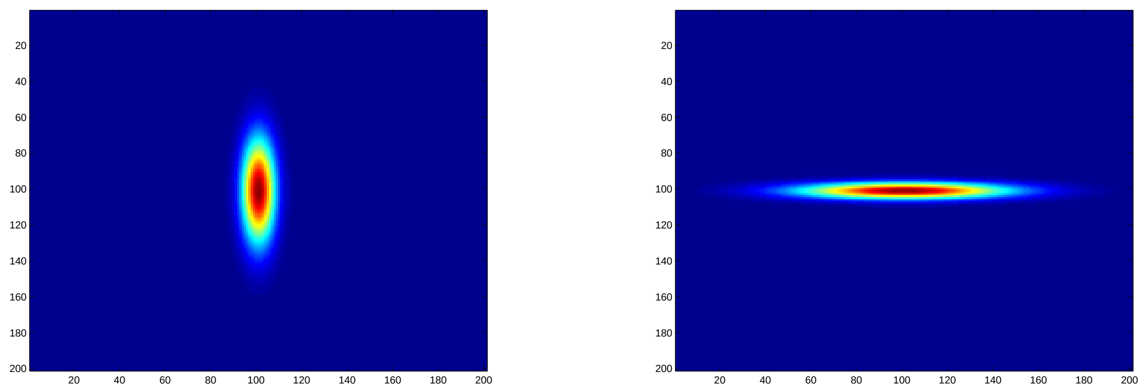




Level curves are circles

Changing scale of the Gaussian function in TF-plane

Function $f(t) = \mathrm{e}^{-\frac{t^2}{2a^2}}$, distribution $\rho_f(t,\omega) = \frac{1}{2\pi}\mathrm{e}^{-(\frac{t^2}{a^2}+a^2\omega^2)}$



Level curves are ellipses.

Next few slides:

Some basic definitions and notations:

Orthonormal basis

Orthonormal family of functions

- Functions are uncorrelated: for any two different functions φ_n and φ_m in the family ($n \neq m$):

$$(\varphi_n, \varphi_m) = \int \varphi_n(t)\overline{\varphi_m(t)}dt = 0.$$

- Function are normalized: any function φ_n in the family has norm equal to 1:

$$\|\varphi_n\|^2 = (\varphi_n, \varphi_n) = \int \varphi_n(t)\overline{\varphi_n(t)}dt = 1.$$

Orthonormal basis

- An orthonormal basis for a space of functions is an orthonormal family of functions $\{\varphi_n\}_n$ such that any function f in the space can be written as sum

$$f = \sum_n c_n \varphi_n.$$

- The constants c_n is obtained by the inner product between the the functions f and φ_n

$$c_n = \int f(t) \overline{\varphi_n(t)} dt.$$

Bi-orthogonal basis

- A family o functions $\{\varphi_n\}_n$ in a space V and a family of functions $\{\tilde{\varphi}_n\}_n$ in the dual space $\tilde{V}$ are *bi-orthogonal bases* if if they are bases for V resp. $\tilde{V}$ and

$$(\tilde{\varphi}_n, \varphi_m) = \int \tilde{\varphi}_n(t)\overline{\varphi_m(t)}dt \begin{cases} \neq 0 \text{ when } n = m, \\ = 0 \text{ when } n \neq m. \end{cases}$$

  Then any function f in the space V can be written as sum

$$f = \sum_n c_n \varphi_n.$$

- The constants c_n is obtained by the inner product between the the functions f and φ_n

$$c_n = (f, \tilde{\varphi}_n)/(\tilde{\varphi}_n, \varphi_n).$$

- the dual space $\tilde{V}$ may, or may not be the same as V

Continuous versus discreet wavelet transform

- Continuous parameter family of wavelets

$$\psi_{a,b}(t) = \frac{1}{\sqrt{b}}\psi(\frac{t-a}{b}).$$

where a and b are real parameters, $b \neq 0$

- Orthonormal wavelet basis $\{\psi_{kj}\}_{k,j\in\mathbf{Z}}$ where

$$\psi_{kj}(t) = 2^{\frac{j}{2}}\psi(2^j - k).$$

Continuous wavelet transform

Let ψ be a function on the real line and

$$\psi_{a,b}(t) = \frac{1}{\sqrt{|b|}}\psi(\frac{t-a}{b}).$$

The wavelet let transform is defined by

$$f \quad \to \quad W_f(a,b) = \int f(t)\overline{\psi_{a,b}(t)}dt.$$

Inversion of the continuous wavelet transform

$$f(t)] = \frac{1}{C_\psi} \int\int_{\mathbf{R}x\mathbf{R}} W_f(a,b)\psi_{a,b}(t)\frac{da\,db}{a^2}.$$

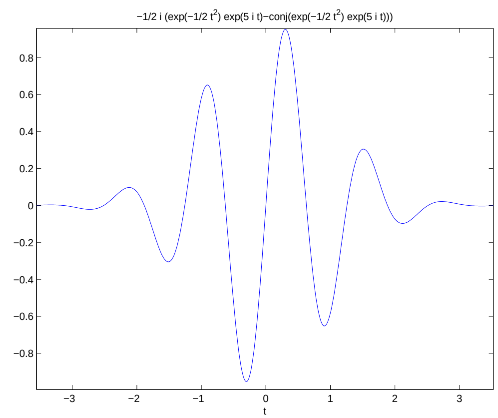
In contrast to the discrete wavelet transform we don't need very special function ψ. In general we need to have $\hat{\psi}(0) = 0$ and that

$$C_\psi = \int \hat{|}\psi(\omega)|^2 \frac{d\omega}{\omega} < \infty.$$

Example: Morlet wavelets

- Real Morlet-5 wavelet:

$$\psi(t) = \sin(5t)\mathrm{e}^{-\frac{t^2}{2}}.$$



Fourier transform

$$\hat{\psi}(\omega) = \frac{1}{2i\sqrt{2\pi}} \left(\mathrm{e}^{-\frac{(\omega-5)^2}{2}} - \mathrm{e}^{-\frac{(\omega+5)^2}{2}} \right).$$

- Complex Morlet-5

$$\psi(t) = \frac{\partial}{\partial t} \mathrm{e}^{i5t} \mathrm{e}^{-\frac{t^2}{2}}.$$

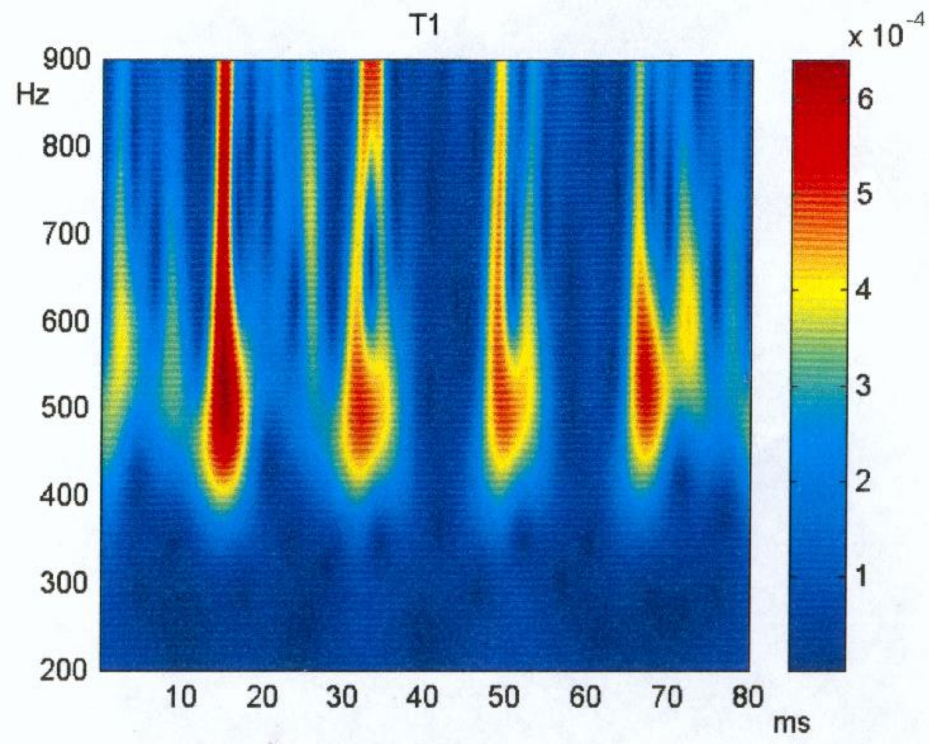
Fourier transform

$$\hat{\psi}(\omega) = i\omega \mathrm{e}^{-\frac{(\omega-5)^2}{2}}.$$

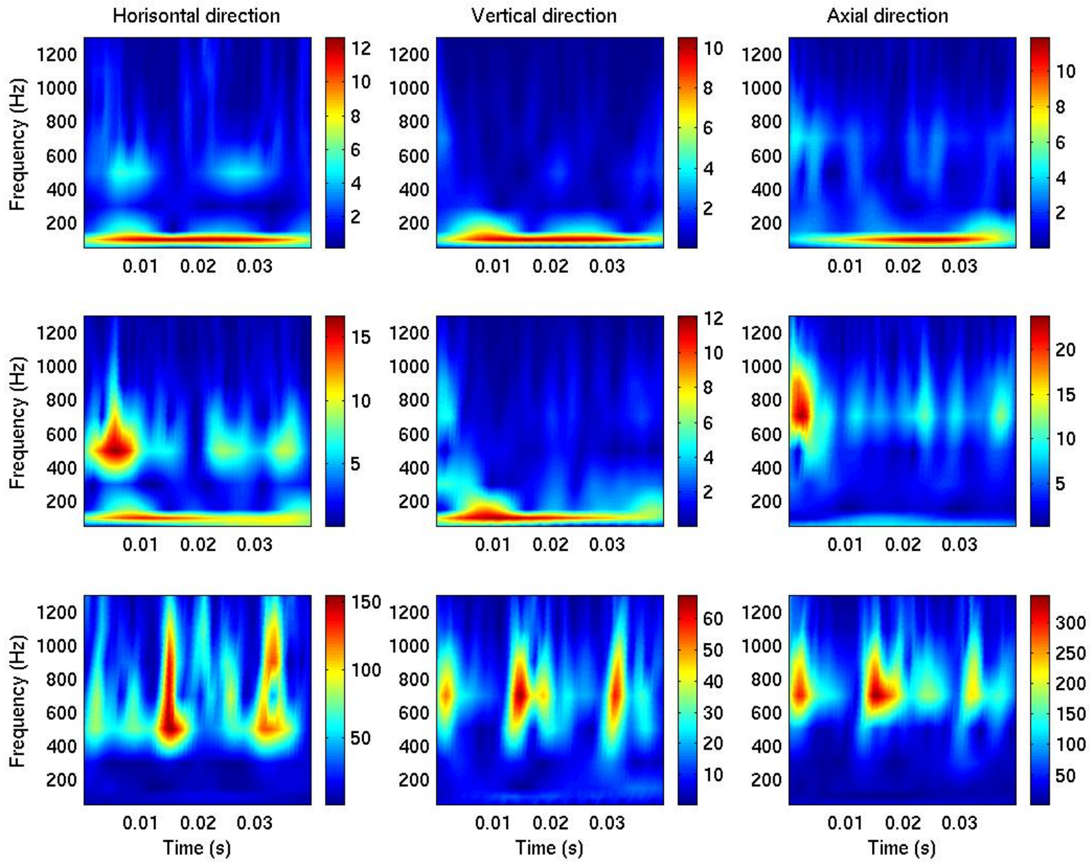
Application of Continuous Wavelet Transform

- In general not so useful when it involves reconstruction of functions, since it is too complex.
- Good for frequence analyze of functions since it gives information booth of time and frequency of events.
- As we saw with the scaling of the Gaussian one may easily gradually change the focusing in the analysis between frequency and time.

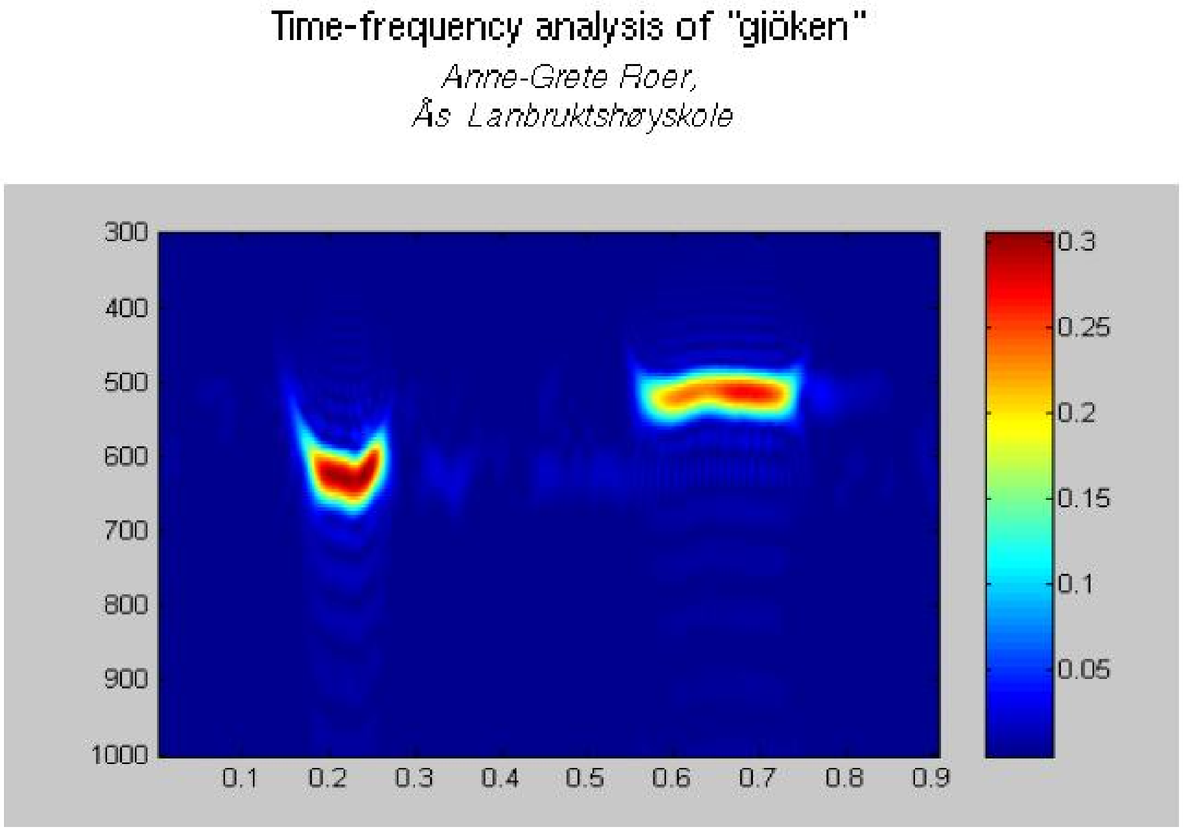
Vibration analysis of defect bear-rings



Classification of bear-ring signals



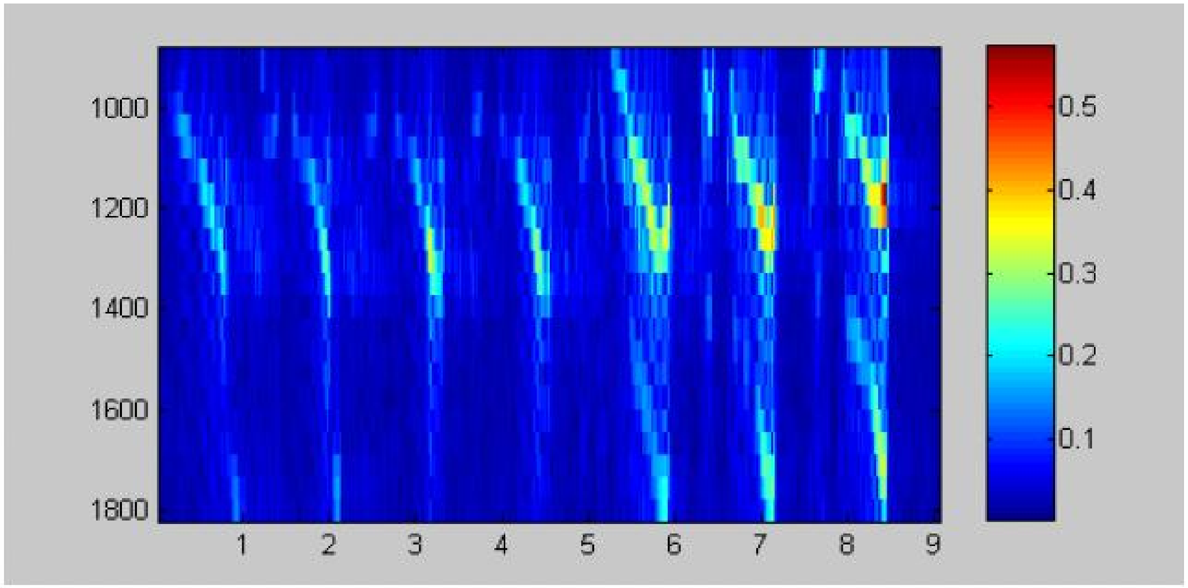
TF analysis of singing birds



TF analysis of singing birds (continued)

Time-frequency analysis of "storlom"

Anne-Grete Roer,
Ås Lanbruktshøyskole



Discrete wavelet transform

We have two parallel descriptions of the discrete wavelet transform

- By an orthonormal basis of wavelets and

$$f(t) = \sum_{kj} c_{kj} \psi_{kj}(t),$$

and the wavelet transform of f is

$$f \quad \rightarrow \quad c_{kj} = \int f(t) \overline{\psi_{kj}(t)} dt.$$

- By a low-pass filter $\mathbf{h}$ and a high-pass filter $\mathbf{g}$ which are arranged in an algorithmic tree. The filter h g are such that they can make a **Quadratic Mirror Filter**

What properties do we want ψ to have

- good localization in time.
- good localization in frequency.
- vanishing moments, the more the better.
- smoothness properties
- easy to compute with – as filter finite.

Three main approaches for construction

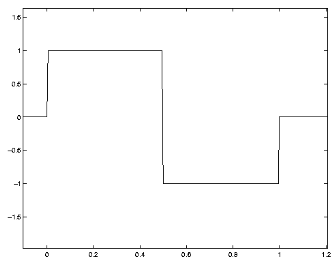
- Construction on the function side.
- Construction on the Fourier transform side.
- Construction based on construction of Quadratic Mirror filter

Haar basis (1910)

Construction on the function side

Haar function

$$H(t) = \begin{cases} 1 \text{ for } 0 < t < \frac{1}{2} \\ -1 \text{ for } \frac{1}{2} < t < 1 \\ 0 \text{ otherwise.} \end{cases}$$



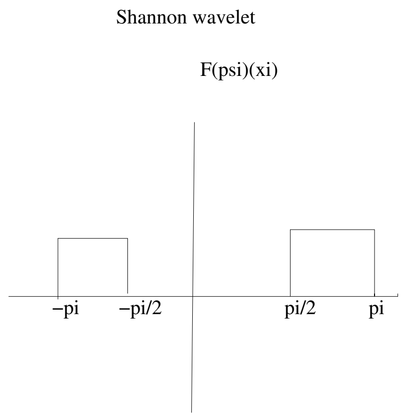
Haar functions have bad localization in frequency

Var(function frequency) $= \infty$

Shannon wavelet

The other extreme, constructed on Fourier side.

Shannon basis around 1940.



Shannon wavelets have bad localization in time Var(function time) $= \infty$

Franklin system - asymptotically a wavelet

- Philip Franklin (1926):Construction of orthonormal spline system of piecewise polynomial of order m on a bounded interval. Away from the endpoints the function is approximatively a wavelet to any precision.
- Strömberg 1981. Transferring Franklin's construction to spline systems on the whole real line getting Franklin's asymptotic limit function as a wavelet generating an orthonormal basis. This wavelet function is exponentially decreasing.

Wavelet theory appear

- Yves Meyer (1985): Construction a wavelet on the Fourier side

$$\hat{\psi}(\xi) = b(\xi)\mathrm{e}^{i\frac{\xi}{2}},$$

where $b(\xi)$ an even function (the function $\chi_{[-\pi,\pi]}$ smoothed out in a special way) This wavelet is C^∞ smooth, of course compactly supported Fourier transform and it decreasing polynomially of any order.

- Stephan Mallat (1986) Multi-scale analysis and construction of wavelet from Quadratic Mirror-filter.

- Ingrid Daubechies (1987): Construction of wavelets with compact support by construction wavelet filter with finite length.

I want to go through the construction of Daubechies finite wavelet filter. I will relate the construction more to the spaces of spline function whereas Daubechies makes the construction of the filter on their Fourier transform side.

We need some more notation.

Translation operators

For integer k an $f \in L^2$:

$$T_k f(x) = f(x-k).$$

For integer k and $f = \{f[j]\} \in l^2$:

$$(T_k f)[k] = f[j-k].$$

Adjoint notation

$$a^*[k] = \overline{a[-k]}.$$

Inner product

$$< a, T^k b >= \sum_j a[j]\overline{b[j-k]} = a * b^*[k].$$

Orthogonal complement

Definition: For h in l^2 define Orthcomp(h) by

$$\text{Orthcomp}(h)[j] = (-1)^j \overline{h}[1-j].$$

Lemma: If h in l^2 and $g = \text{Orthcom}(h)$. Then $\{T_{2k}h\}_k$ is orthogonal to $\{T_{2k}g\}_k$.

Multi-scale analysis

I will first do the MS-analysis i a bi-orthogonal setting

- $$\{0\} \leftarrow \ldots V_{j-1} \subset V_j \subset V_{j+1} \ldots \rightarrow \mathbf{L}^2(\mathbf{R}).$$

- $$f(x) \in V_j \text{ iff } f(2x) \in V_{j+1}.$$

- V_0 har an bi-orthogonal-basis $\{\varphi(x-k)\}_{k\in\mathbf{Z}}$ with dual bassis $\{\tilde{\varphi}(x-k)\}_k$ in the dual space $\tilde{V}_0$

Box spline function

Let

$$B^{(0)} = \text{Dirac delta function},$$

$$B^{(m)}(x) = B^{(m-1)} * \chi_{[0,1]}(x),$$

For fixed $m > 0$ let $\varphi_k(x) = B^{(m)}(x - k)$.

We define $V^{(m)_0}$ to be the closure of span(φ_k).
$V^{(m)_0}$ is the space of functions in C^{m-2} which are piecewise polynomial of degree less or equal to $m - 1$ on intervals $(n - 1, n)$

The scaling equations

$$\varphi(x) = c \sum_k h[k]\varphi(2x - k),$$

$$\psi(x) = c \sum_k g[k]\varphi(2x - k).$$

The cascade algorithm

Take the limit φ (if it exist) of the sequence $\varphi^{(m)}$ given by iteration formula

$$\varphi^{(m)}(x) = c \sum_k h[k] \varphi^{(m-1)}(2x-k),$$

where c is a suitable normalization constant, the starting function $\varphi^{(0)}$ could be almost any function with $\int \varphi dt \neq 0$

Observe: It is commutes with with convolutions:
if sequence h generates φ and g generates ψ with the sequence $h * g$ starting with $\varphi^{(0)} * \psi^{(0)}$ the outcome of the cascade algorithm is $\varphi * \psi$

$h = [1,1] \Rightarrow \varphi = \chi_{[0,1](x)}$

$h = [1,1]^2 = [1,2,1] \Rightarrow$ linear box-spline $B^{(2)}(x)$

$h = [1,1]^4 = [1,4,6,4,1,] \Rightarrow$ cubic box-spline $B^4(x)$,

and so on

$h = [1,1]^m = \Rightarrow m$ order box-spline $B^{(m)}(x)$

Decomposition of V_0

Suppose that the space V_0 has a bi-orthogonal basis $\{\varphi^1(x-k)\}$ with dual functions $\{\tilde{\varphi}^1(x-k)\}$ in the dual space $\tilde{V}_0$

We will find a complement W_{-1} of V_{-1} in the space V_0

- with a bi-orthogonal basis $\{\varphi(x-2k)\}$ of V_{-1} with dual functions $\{\tilde{\varphi}(x-2k)\}$ in a dual space $\tilde{V}_{-1}$
- with a bi-orthogonal basis $\{\psi(x-2k)\}$ of W_{-1} with dual functions $\{\tilde{\psi}(x-2k)\}$ in a dual space $\tilde{W}_{-1}$
- and where the functions $\{\varphi(x-2k)\} \cup \{\psi(x-2k)\}$ is an

bi-orthogonal basis of V_0 with dual basis $\{\tilde{\varphi}(x-2k)\} \cup \{\tilde{\psi}(x-2k)\}$ in $\tilde{V}_0$

Transferring the problem to l^2.

Using the bases of V_1 and $\tilde{V}_1$ we can transfer to the similar problem where V_0 is a subset of $V_1 = l^2$ and $\tilde{V}_1 = l^2$.

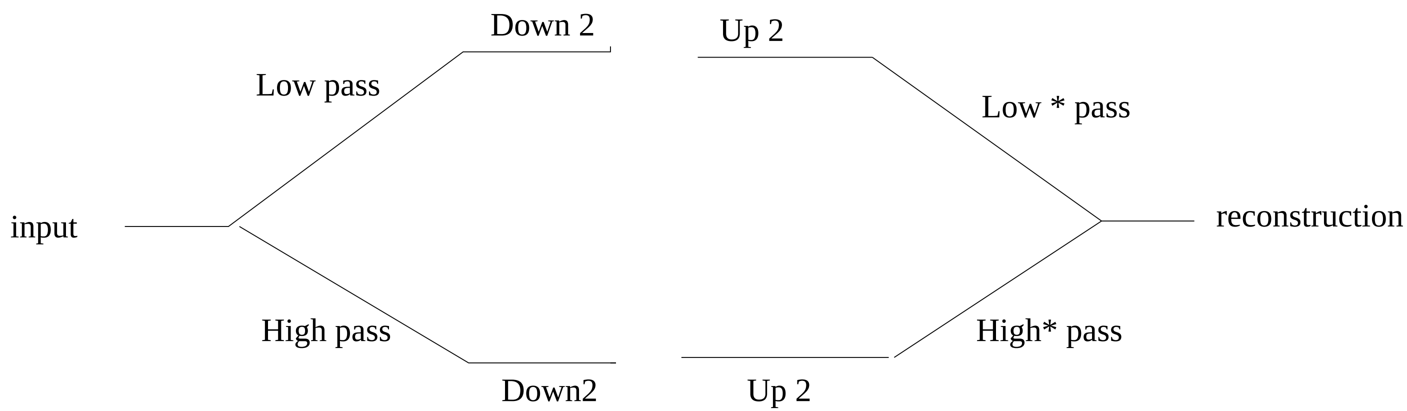
Observe: It might happen that the dual space V_0 and $\tilde{V}_0$ may be different spaces and also that W_0 and $\tilde{W}_0$ are different spaces

Basis representation and filters

(Assuming we have normalized so that $<\varphi,\tilde{\varphi}>=<\psi,\tilde{\psi}>=1$) we have the representation of function $f \in V_0$

$$f(x) = \sum_k < f, T^{2k}\tilde{\varphi} > T^{2k}\varphi(x) \quad + \quad \sum_k < f, T^{2k}\tilde{\psi} > T^{2k}\psi(x).$$

Quadratic Mirror Filter

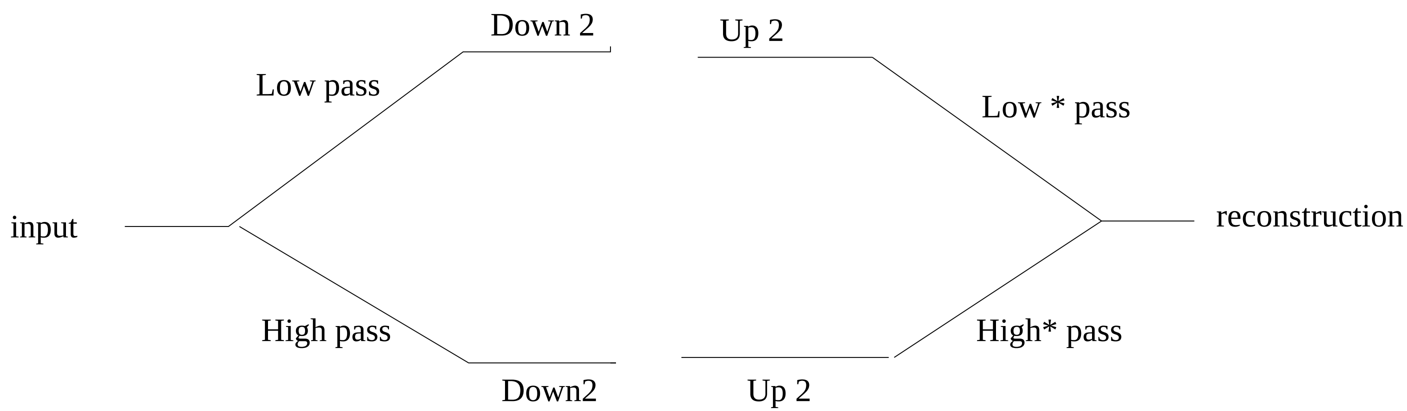


Remark: To be Mirror filters the filters should be orthonormal i.e $\tilde{h} = h$ and $\tilde{g} = g$.

- When m is even will find the bi-orthogonal filters: h with dual $\tilde{h}$ for V_{-1} bi-orthogonal. $\tilde{V}_0$
  g with dual $\tilde{g}$ for W_{-1} respectively. $\tilde{W}_0$
- (Assuming we have normalized the filter such that $< h, \tilde{h} >=< g, \tilde{g} >= 1$ we have for data set $f \in l^2$

$$f[j] = \sum_k < f, T^{2k}\tilde{h} > T^{2k}h[j] \quad + \quad \sum_k < f, T^{2k}g > T^{2k}g[j].$$

Quadratic Mirror Filter



Remark: To be called Mirror filters the filters should be orthonormal i.e $\tilde{h} = h$ and $\tilde{g} = g$.

- Filter $h = [1,1]^m$ is given from the scaling equations.
- Filter $\tilde{g} = \text{Orthcomp}(h)$.
- Once filter $\tilde{h}$ is known we set $g = \text{Orthcomp}(\tilde{h})$.
- h has length $m+1$ we will find $\tilde{h}$ of length $m-1$ by straightforward orthogonalization in $\mathbf{R}^{m-1}$ against restriction of $T^{2k}h$ to $\mathbf{R}^{m-1}$.
- Example:

  When $m = 2$

$$h = [1,1]^2 = [1,2,1] \Rightarrow \tilde{h} = [1].$$

  When $m = 4$

$$h = [1,1]^2 = [1,4,6,4,1] \Rightarrow \tilde{h} = [-1,4,-1].$$

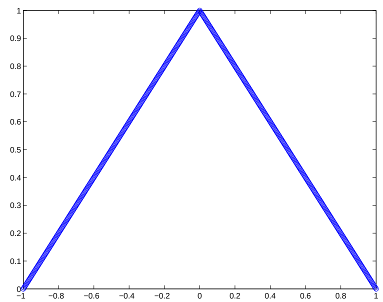
  When $m = 6$

$$h = [1,1]^6 = [1,6,15,20,15,6,1] \Rightarrow \tilde{h} = [3,-18,38,-18,3].$$

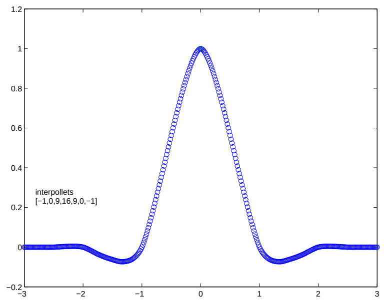
interpolets

The product $p = h * (\tilde{h})^*$ will be an interpolation filter which will generate what may be called interpolets.
Delaurier & Dubuc 1989.
$q = g * (\tilde{g})^*$ will be a "difference filter" and $q = \text{Orthcomp}(p)$.
The interpolating filter will be:

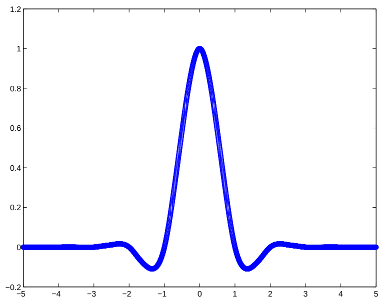
When $m = 2$ $p = [1, 2, 1]$, the generated intepolets i fig.:



When $m = 4$ $p = [-1, 0, 9, 16, 9, 0, 1]$:



When $m = 6$ $p = [3, 0, -25, 0, 150, 256, 150, 0, -25, 0, 3]$:



Factorization of the interpolation filter

- The filter p can be split up into factors that can be combined in many ways.
- For instance it may be written an autocorrelation filter

$$p = a * a^*,$$

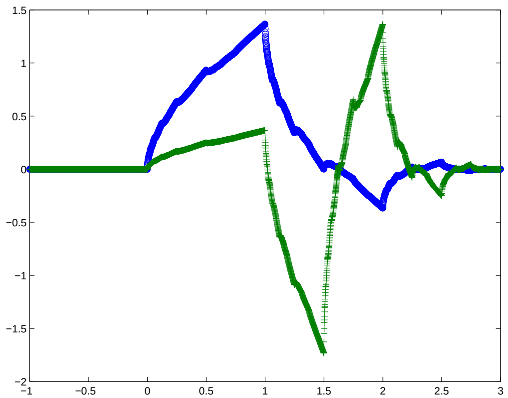
which means that $\tilde{a} = a$, i.e a is is an orthonormal filter.

- If we starting the cascade algorithm with something orthonormal as the Box-splines of order $m = 1$, then outcome will be a scaling function φ and a wavelet function ψ that generates an orthonormal wavelet.

When $m = 2$ we get the Haar filters $[1, 1]$ and $[1, -1]$.

When $m = 4$ we get Daubechies filters of length 4, the corresponding generated functions are in figure:

blue is wavelets and green is wavelets



When $m = 6$ we get Daubechies filters of length 6

short about the desired properties of constructed wavelets

- in general the factor $[1,1]$ implies in good properties of the wavelets it generates on the Fourier transform a sinc factor (decays like $|\xi|^{-1}$) and the other factor coming from the dual filter, kind of destroys some of these properties, but those factors are needed to obtain orthonormality.
  The $[1,1]$ factors win: the longer filter the smoother wavelets
- the factors $[1,-1]$ in the $\tilde{g}$ filter creates moments. The other factors (those from g) cannot destroy it, so the number of vanishing moments of the wavelet ψ is directly proportional to the length of the orthogonal wavelet filter.

Moment conditions on wavelet function

Lemma Given the function φ satisfying:

$$\int \varphi(x)ds \neq 0,$$

and the filter h and define the function ϕ by

$$\phi = \sum h_k \varphi_k.$$

Let m_0, and m_1 be the order of moment condition of the filter h and of the function ϕ in other words:

$$\sum h_k k^l \begin{cases} = 0, & l < m_0, \\ \neq 0, & l = m_0 \end{cases}$$

$$\int \phi(x) x^l dx \begin{cases} = 0, & l < m_2, \\ \neq 0, & l = m_1 \end{cases}$$

Then

$$m_1 = m_0.$$

Estimate of wavelet coefficients

Let f be a m times continuous differentiable function on the line and assume that ϕ satisfies all moment condition up to order m then

$$|<f,\psi_{kj}>| \leq O(2^{-j(m+\frac{1}{2})}).$$

Wavelet basis in dimension 2

Corresponding to the wavelet functions $\{\psi_{k,j}\}$ are the three sets of tensor functions

$$\{\varphi_{kj}(x)\psi_{lj}(y)>\}_{k,l,j\in Z},$$

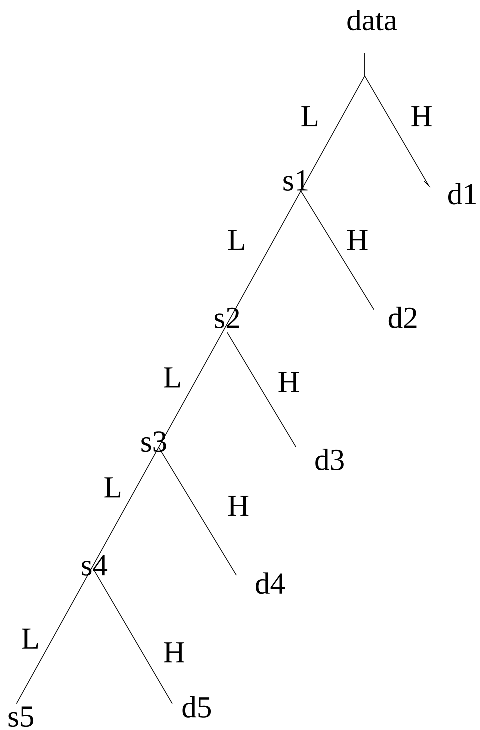
$$\{\psi_{kj}(x)\varphi_{lj}(y)>\}_{k,l,j\in Z},$$

$$\{\psi_{kj}(x)\psi_{lj}(y)>\}_{k,l,j\in Z},$$

and corresponding to the scaling functions $\{\varphi_{k,j}\}$ are the functions

$$\{\varphi_{kj}(x)\varphi_{lj}(y)>\}_{k,l,j\in Z}.$$

Wavelet-filter-tree



$L =$ low-pass filter, $\quad H =$ high-pass filter

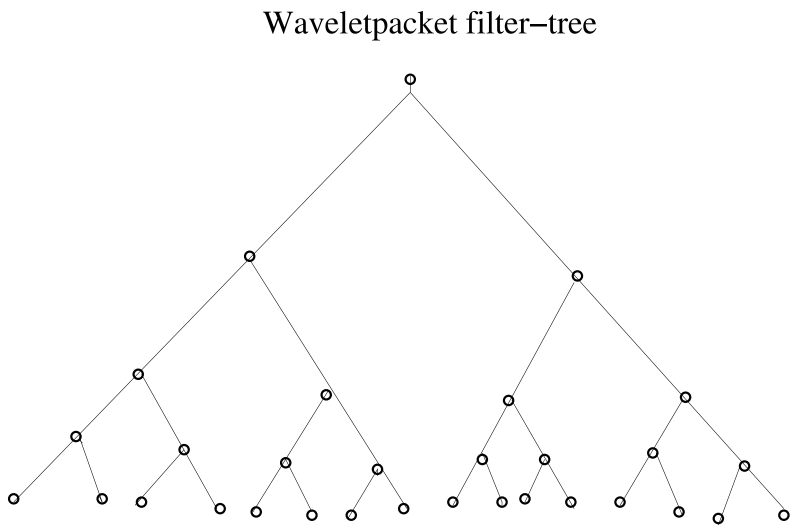
$d =$ "sum" $\quad d =$ "sum"

Applications of discrete wavelet transform

- Image processing, noise reduction of signals

Wavelet packets

We extend the wavelet-filter-tree to the full tree. With data of size $N = 2^M$ there will be $M + 1$ levels in the tree including the top.(Top node=input data).



wavelet packets (continued)

- There will be totally $(M+1)N$ different coefficients -including the N input data values.
- Each combination of stopping at nodes in the tree corresponds to an Orthonormal basis. There will be about 1.45^N different families of orthonormal bases.
- All coefficients may be computed with complexity MN

Best basis

- Assume we have a linear cost function about how good a basis is. Linear means for each basis B

$$Cost_B = \sum_{c_n \in B} Cost(c_n).$$

- There is an algorithm choosing the **best basis** B in the libraries of all those bases obtained from different combination of nodes.
- In the best basis we chose a few (< 10) most significant coefficients.
- The complexity the algorithm is of order MN