Multiresolution schemes

Fondamenti di elaborazione del segnale multi-dimensionale

Stefano Ferrari

Università degli Studi di Milano
stefano.ferrari@unimi.it

Elaborazione dei Segnali Multi-dimensionali e Applicazioni
2011–2012

Motivations

- Transforms give an alternative view of a signal.
- The Fourier Transform of a signal $f(x)$ gives a representation of the same signal in the frequency domain.
- The processing should reveal information not directly accessible from the signal.
- The Fourier Transform gives information on the presence of frequency components.
  - It is impossible to localize them.
  - It can be critical for non-stationary signals.
- Wavelets perform what is called *space-frequency* representation:
  - information on the frequency components,
  - localized in space.

Motivations (2)

- Besides wavelet transform allows to describe the image at different resolutions:
  - features detection at different scales;
  - selective denoising;
  - compression and transmission.
- Multiresolution representation allows to perform linear filtering operations using the same (small) filter for each resolution, instead of using larger filter for coping with large features in the full resolution image.

Multiscale representation

- The concepts from which the wavelet transform can be derived root in several disciplines.
- In particular for images, multiscale representations have been proposed before the wavelets.
- Among them, at least the following have to be mentioned:
  - Gaussian pyramid
  - Laplacian pyramid
  - Scale spaces
  - Subband coding

Scale

Operating at different scales is a concept exploited in several approaches.

- Features are not independent of image scale:
  - their actual size depends on the resolution of the image and the distance from the camera.
- For example, Marr-Hildreth edge detector:
  1. filter with a Gaussian of a suitable scale;
  2. compute the Laplacian;
  3. find the zero-crossing.
- Also the Canny edge detector makes use of Gaussian smoothing.
- The size of the smoothing filter has to increase with the scale parameter.

Multiscale representation

Multiscale (or multigrid) representation are based on a simple observation:

- fine scales need high resolution,
- for coarse scales, low resolution copies.

Hence a M-levels multigrid representation of an image, f, can be obtained as:

- $f_0 = f$
- $f_{m+1} = S_{\downarrow 2}[f_m], \qquad n = 0, \ldots, M-1$

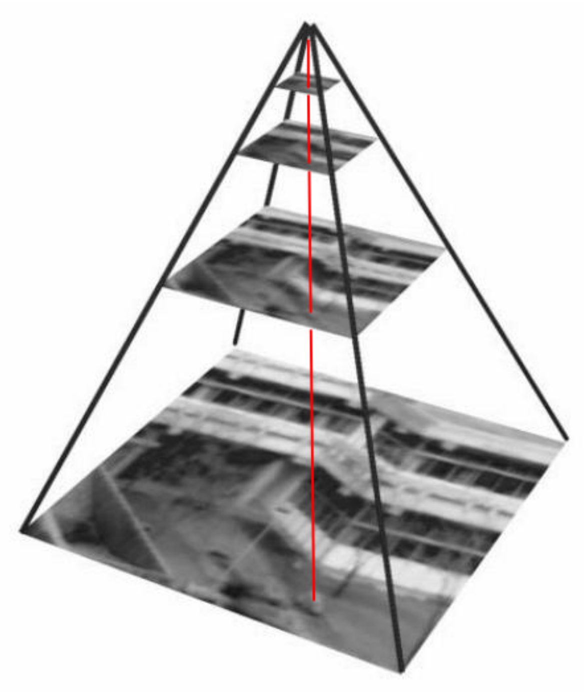
where $S_{\downarrow 2}[\cdot]$ is a suitable downsampling operator.

- Note: $\downarrow 2$ means that one sample out of two in each direction is discarded;
  - other downsampling rates are theoretically possible, but not useful in practice.

Multiscale representation (2)

- Multigrid representation does not increase too much the storage requirements:
  - given N^2 the number of pixel of f_0,
  - $N^2/4$ are required for f_1,
  - $N^2/16$ for f_2,
  - $N^2/2^m$ for f_m.
  - Less than $\frac{4}{3}N^2$ pixels are required.
- Since $S_{\downarrow 2}[\cdot]$ operates on the previous level, the total number of operations for obtaining the whole multigrid representation is proportional to $\frac{4}{3}N^2$.
- Multigrid are also known as *pyramidal* representations (Burt and Adelson, 1983).

Multiscale representation (3)



Downsampling operator

Naive choices for $S_{\downarrow 2}[\cdot]$ may produce undesirable effects.

- $S_{\downarrow 2}[\cdot] = \downarrow 2$ may produce aliasing.
- Smoothing is required.
  - Smoothing cancels high frequency variations that may cause aliasing when downsampled.
- Gaussian is generally chosen as smoothing operator.
  - Multigrid representation are called *Gaussian pyramid*.
- Smoothing and downsampling allow small filters or small spatial operators to operate on large scale features.

Upsampling operator

Some operations require the recovering of the original size of the lower levels of the pyramid.

- It happens when different levels content has to be compared.
- This operation is carried out by the upsampling operator, $R_{\uparrow 2}[\cdot]$
- A suitable rule for estimating the missing pixels is required.
  - Generally, every even pixel of each row is estimated from the odds;
  - then, the same procedure is applied considering the column direction.

Laplacian pyramid

A complementary representation can be derived from the Gaussian pyramid:

- $l_m = f_m - R_{\uparrow 2}[f_{m+1}]$
- $l_{M-1} = f_{M-1}$

where M is the number of pyramid levels.

This representation is called *Laplacian pyramid*.

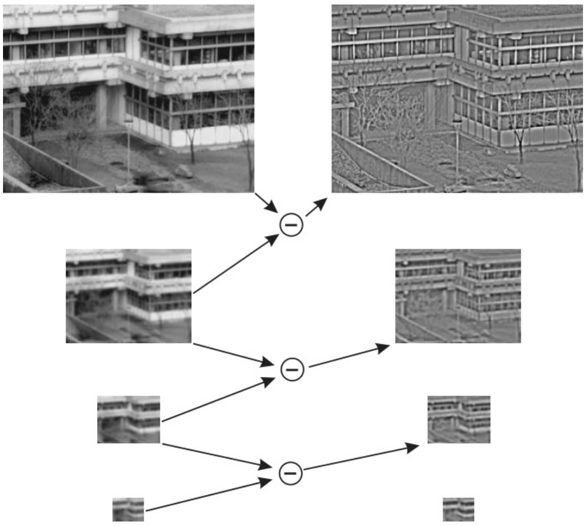
- It provides a bandpass decomposition of the image.
  - l_m contains those components belonging to f_m, but not in f_{m-1}.
  - l_{M-1} contains the low frequencies components (coarsest scale structures).

Laplacian pyramid (2)

From a Laplacian pyramid, the original image can be recursively reconstructed.

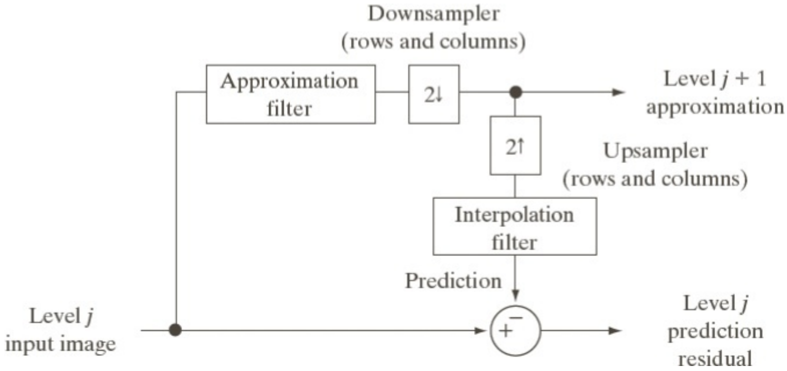
- The scheme for computing the Laplacian pyramid can be inverted:
  - $l_{M-1} = f_{M-1}$
  - $f_{m-1} = l_{m-1} + R_{\uparrow 2}[f_m]$
- Errors in the computation of l_{m-1} due to $R_{\uparrow 2}[\cdot]$ are absorbed in the reconstruction.

Laplacian pyramid (3)

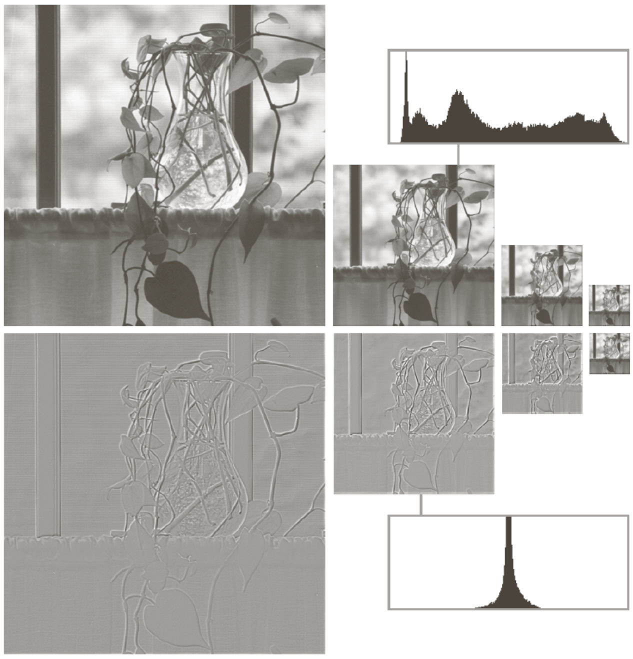


Since Laplacian images histograms are usually more dense in a small neighborhood of zero, compression algorithms can perform better on Laplacian pyramids than on Gaussian ones.

Pyramid scheme



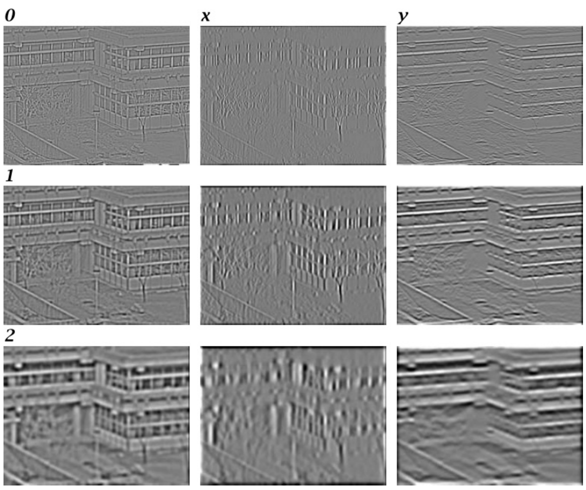
Pyramid example



Directional pyramids

A variant of the pyramidal schemes are the *directiopiramidal decomposition*:

- Instead of isotropic filters, a pool of directional filters are used for subsampling.

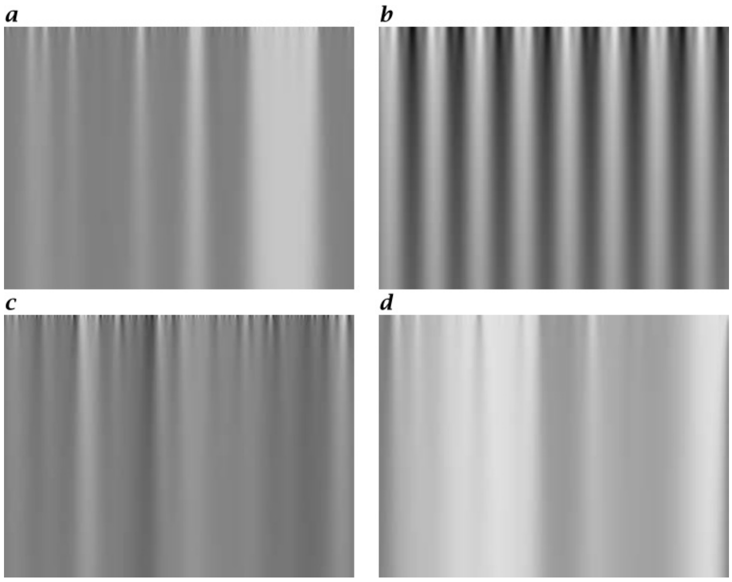


- For example, separable filters for horizontal and vertical directions can be used.
- More complex filters pool design is a difficult problem.

Scale spaces

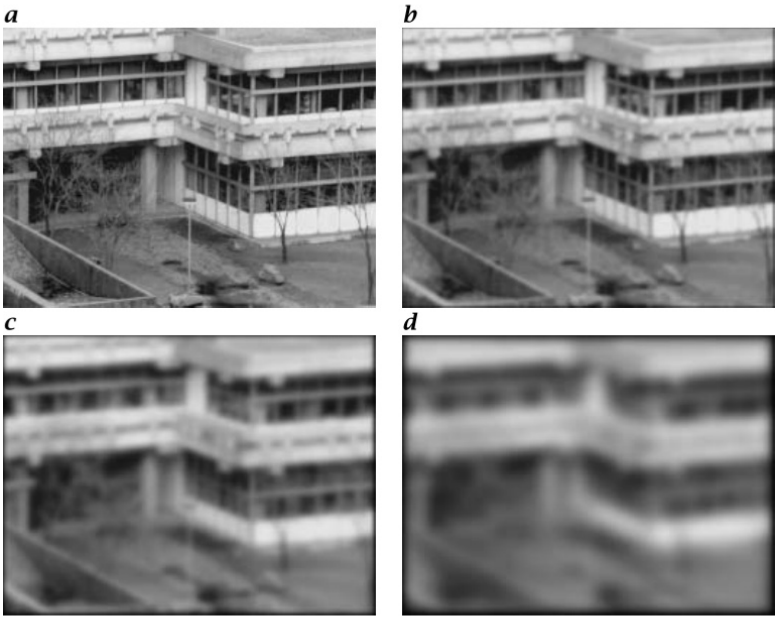
- Pyramidal representation allows for a very rigid multiscale processing.
  - The scale parameter can vary only of a factor of two between each lelel.
- Scale space scheme allows a continuous changing of the scale parameter.
- The scale space is generated by blurring the image at a given degree.
- This can be modelled as a diffusion process (such as heat).
  - Spatial concentration differences in the gray level are equalized.
  - The scale is modelled as the time and the diffusion process produces the scale space.
- It can be shown that for an homogeneous diffusion process, the version of the image f at the scale t, f_t, can be computed as the convolution of f with a Gaussian of variance t.

Scale spaces (2)



- a edges and points
- b periodic pattern
- c random
- d small and large features

Scale spaces (3)



- a $f_0 = f$
- b f_1
- c f_4
- d f_{16}

Scale spaces (4)

A scale space filter must ensure two important properties to the generated scale space.

1. No new details have to be added as the scale parameter increases.
   - The image information content must decrease with the scale parameter.
   - It can be formalized with the maximum-minimum principle:
     - local extrema cannot be enhanced.
2. The scale space does not depend on the scale parameter from which the diffusion starts.
   - Scale invariance principle.
   - Starting with an image at the scale t_1 and applying the smoothing operator at scale t_2, the image at scale $t_1 + t_2$ is obtained.

Gaussian kernel is the only convolution kernel isotropic and homogeneous that meet these two properties.

Scale spaces (5)

Several variants stem from these schema.

- Quadratic and exponential increasing scale parameter;
  - accelerated diffusion process.
- Differential (Laplacian) scale spaces;
  - the change of the image with the scale is explicited.
- Discrete scale spaces;
  - the diffusion process is discretized.

Subband coding

- Subband coding operates on the frequency domain.
- The image is decomposed in bandlimited components, called *subbands*.
- The decomposition is invertible:
  - from the subbands, the original image can be recovered.
- The decomposition is realized by means of FIR digital filters.
  - FIR stands for Finite Impulse Response.

Digital filtering

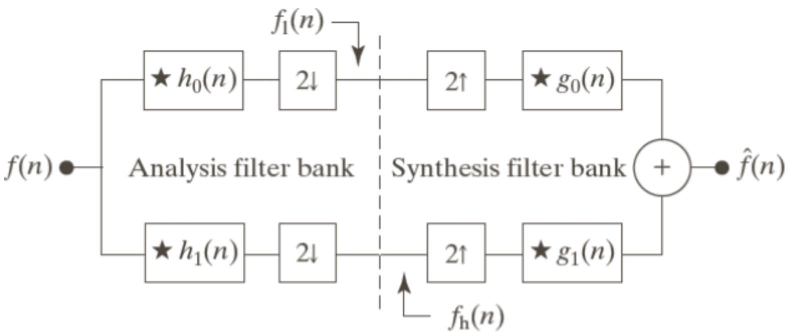
- Digital filtering is formalized by the convolution of the input signal, $f(\cdot)$, composed of discrete samples, with the filter, $h(\cdot)$ composed of a finite number, K, of samples:

$$\hat{f}(n) = \sum_{k=-\infty}^{\infty} h(k) * f(n-k)$$

where filter values out of $[0,\ K-1]$ are zero.
- When the impulse is input, the filter coefficients are output:

$$h(n) = \sum_{k=-\infty}^{\infty} h(k) * \delta(n-k)$$

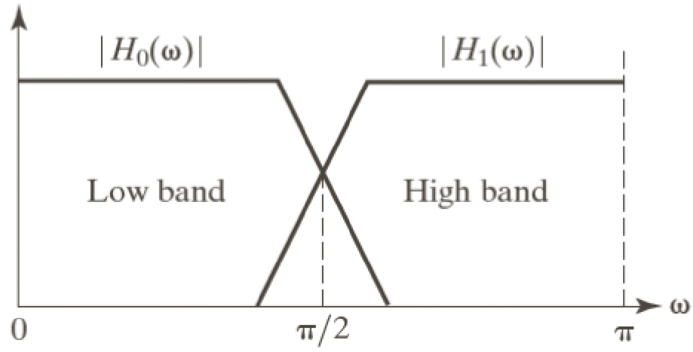
Subband coding and decoding



- Each module (coding/decoding) is composed of a filter banks, each containing two filters.
  - In a filter banks, the signal is passed through all the filters.
  - $\downarrow 2$ is the downsampling operator, which discard the odd index samples;
  - $\uparrow 2$ is the upsampling operator, which insert a 0 valued sample after each sample.

Subband coding and decoding (2)

- The *analysis* filter bank decompose the input sequence $f(n)$ in two (half length) subsequences, f_l and f_h.
- The filter h_0 is a lowpass filter, while h_1 is a highpass filter: f_l and f_h have a content in different frequency interval: the subbands.
  - f_l is called an approximation of f.
  - f_h is the detail of f.



Subband coding and decoding (3)

- The *synthesis* filter bank recombines the subband subsequences, f_l and f_h for producing the sequence $\hat{f}$.
- The two sequences, f_l and f_h, are upsampled and filtered through the filters g_0 and g_1 respectively, and then summed.
- If h_0, h_1, g_0, and g_1 are such that $\hat{f} = f$, they are called *perfect reconstruction filters*.

Filter design

- In order to achieve the *perfect reconstruction filters*, the filters must be related in two ways:

$$\begin{aligned} g_0(n) &= (-1)^n h_1(n) \\ g_1(n) &= (-1)^{n+1} h_0(n) \end{aligned} \qquad \text{or} \qquad \begin{aligned} g_0(n) &= (-1)^{n+1} h_1(n) \\ g_1(n) &= (-1)^n h_0(n) \end{aligned}$$

- The filters are *cross modulated*
  - $g_0 \leftrightarrow h_1$ and $g_1 \leftrightarrow h_0$
- and are *biorthogonal*:

$$\langle h_i(2n-k),\, g_j(k)\rangle = \delta(i-j)\delta(n), \qquad i, j = \{0, 1\}$$

Filter design (2)

- If they enjoy the following property:

$$\langle g_i(n),\, g_j(n+2m)\rangle = \delta(i-j)\delta(n), \qquad i, j = \{0, 1\}$$

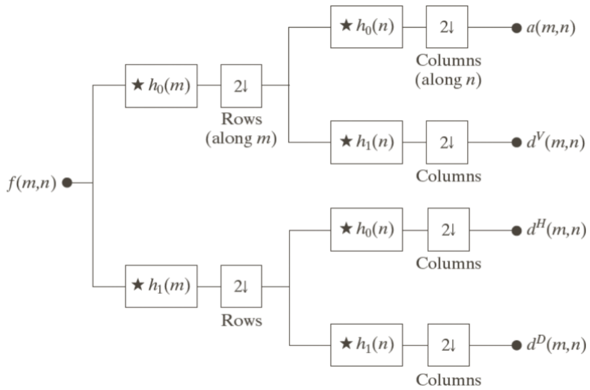
  the filter banks are *orthonormal*.
- Orthonormal filters, for even K, satisfy:

$$\begin{aligned} g_1(n) &= (-1)^n g_0(K-1-n) \\ h_i(n) &= g_i(K-1-n), \qquad i = \{0, 1\} \end{aligned}$$

- Orthonormal filter banks are defined from only one of its filters (*prototype*).

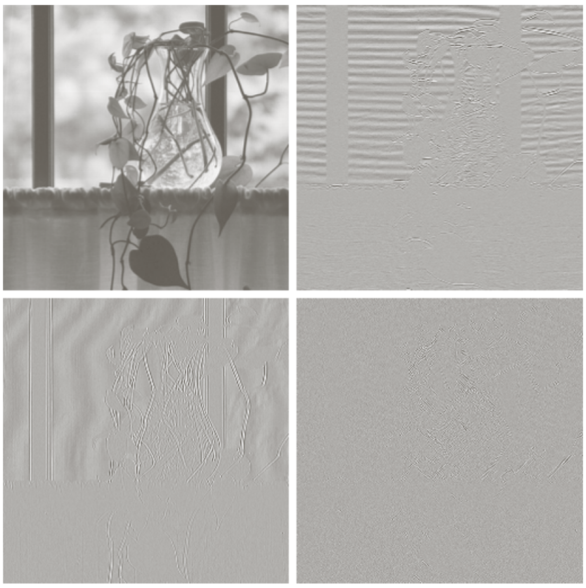
Image subband coding

- 1D orthogonal and biorthogonal filter banks can be used also for image (2D) processing.
- Considering 1D filtering as a separable transform, rows and columns can be processed in sequence.



- Four subimages are obtained:
  - $a(m,n)$, approximation
  - $d^V(m,n)$, vertical detail
  - $d^H(m,n)$, horizontal detail
  - $d^D(m,n)$, diagonal detail

Image subband coding (2)



- The schema can be iteratively used for obtaining a multiresolution representation.

Wavelets definition

- The wavelets, $\psi_{a,\,b}(\cdot)$, are scaled and translated copies of the same function, ψ:

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

- The function $\psi(\cdot)$ that generates the wavelet is called *mother wavelet*.
- The parameter a is the scale parameter.
  - It describes the length of space window embraced by $\psi_{a,\,b}$.
- The parameter b is the shift parameter.
  - It describes the position of the window along the space-line.

Continuous Wavelet Transform

- The Continuous Wavelet Transform (CWT) is defined as:

$$\begin{aligned} W(a,b) &= \langle f, \psi_{a,b}\rangle = \int_{-\infty}^{\infty} f(x)\psi^\star_{a,b}(x)\,\mathrm{d}x \\ &= \frac{1}{\sqrt{|a|}} \int_{-\infty}^{\infty} f(x)\psi^\star\left(\tfrac{x-b}{a}\right)\,\mathrm{d}x \end{aligned}$$

- The amplitude of $W(a,b)$ measures the similarity between f and $\psi_{a,b}$.
  - In this sense, the CWT *analyzes* $f(\cdot)$.

Space-frequency window

- The Fourier transform of $\psi_{a,b}(x)$ is:

$$\mathcal{F}(\psi) = \hat{\psi}_{a,b}(\nu) = \frac{a}{\sqrt{|a|}}\, e^{-\iota\nu b}\, \hat{\psi}(a\nu)$$

  - $\hat{\psi}_{a,b}(\cdot)$ embraces large intervals for small values of a, short intervals for large values of a
- $W(a, b)$ can be reframed as (Parseval):

$$2\pi\, W(a,\, b) = \left\langle \hat{f},\, \hat{\psi} \right\rangle$$

- It can be shown that CWT has:
  - high frequency resolution and low space resolution for high values of a
  - low frequency resolution and high space resolution for small values of a

Space-frequency window (2)

- Hence:
  - a is large $\Rightarrow$ CWT gives fine information on the FT of x, but poor localization in space;
  - a is small $\Rightarrow$ CWT gives very local information in space, but very general in frequency.
- Very useful property for real cases:
  - short events has only high frequency components;
  - long events are characterized by low frequencies.

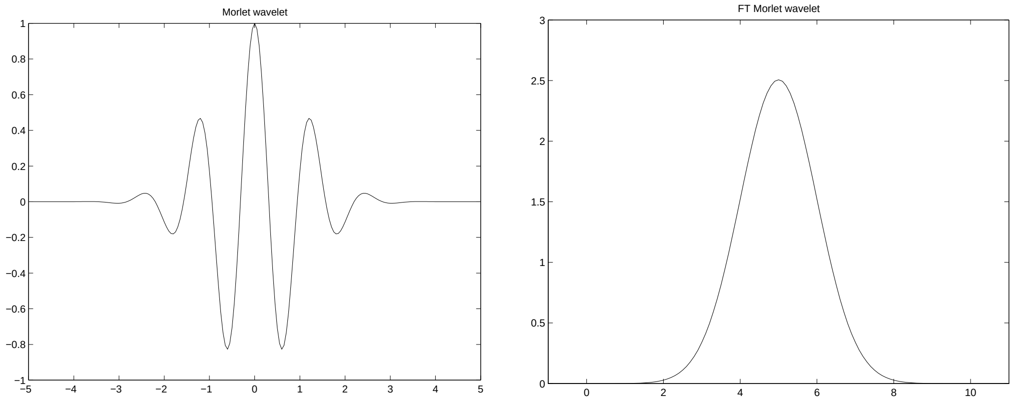
Invertibility

- Invertibility is a desirable property for a signal transform.
- It can be shown that if
$$C_\psi = \int_{-\infty}^{\infty} \frac{|\hat{\psi}(\nu)|^2}{\nu}\,\mathrm{d}\nu < \infty$$
the CWT $W(a,b)$ is invertible:
$$f(x) = \frac{1}{C_\psi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} W(a,b)\,\psi_{a,b}(x)\,\frac{\mathrm{d}a\,\mathrm{d}b}{2}$$
- Hence, it is possible to reconstruct $f(\cdot)$ from the coefficients of its CWT, $W(a,b)$.
  - This operation is often called *synthesis*.

Why *wavelets*?

- From $C_\psi < \infty$, it can be derived that $\hat{\psi}(0) = 0$.
- Hence, $\psi(\cdot)$ must oscillate.
- It can be also shown that $\psi(\cdot) \in L^2(\mathbb{R})$;
  - $f \in L^2 \Leftrightarrow ||f|| = \sqrt{\int_{t\in\mathbb{R}} f^2(x)\,\mathrm{d}x} < \infty$
  - ψ have some limitations in space and frequency.
- The term *wavelet* (small wave) derives from these conditions;
  - *ondina* in Italian, *ondelette* in French.

Morlet wavelet

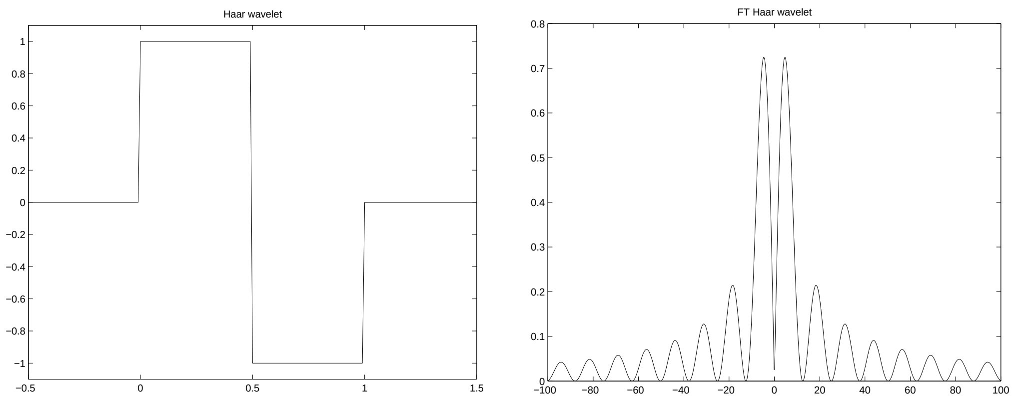


(a) (b)

The Morlet wavelet (a), and its Fourier transform (b).
It is a planar wave localized by a Gaussian.

$$\psi(x) = e^{\iota 5x} e^{-\frac{x^2}{2\sigma}}$$

Haar wavelet

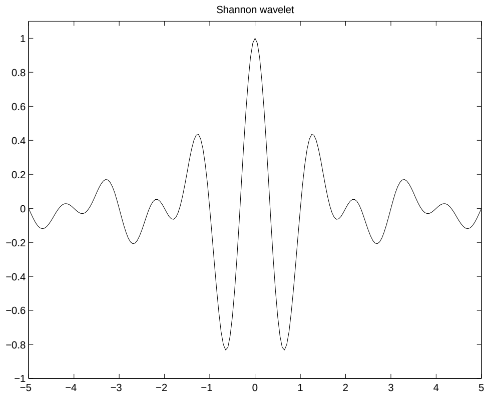


(a) (b)

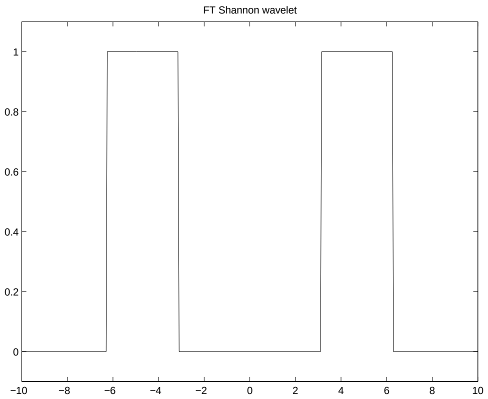
The Haar wavelet (a), and its Fourier transform (b).

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

Shannon wavelet



(a)

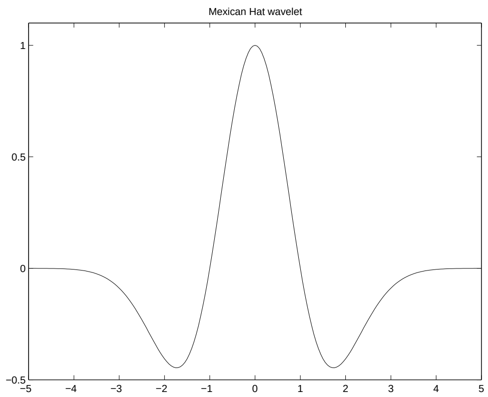


(b)

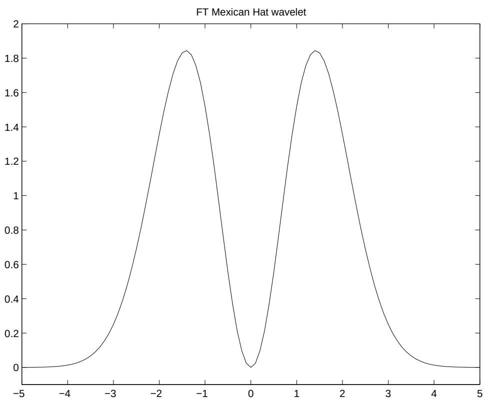
The Shannon (or sinc) wavelet (a), and its Fourier transform (b). It is a family parametrized by ν_b (bandwidth) and ν_c (center frequency).

$$\psi(x) = \sqrt{\nu_b}\,\mathrm{sinc}(\nu_b x)\, e^{2\iota\pi\nu_c x}$$

Mexican hat wavelet



(a)

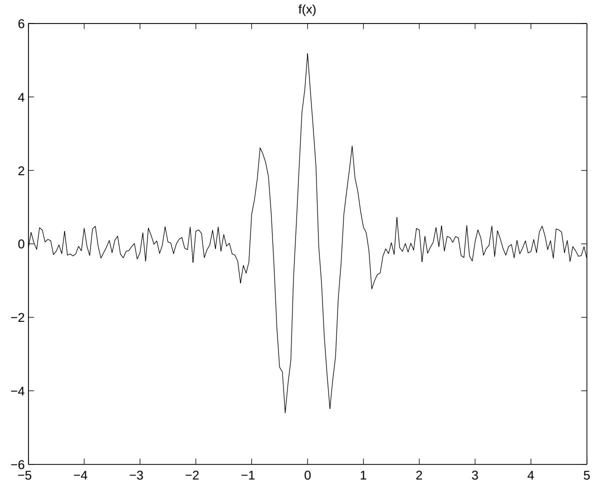


(b)

The Mexican hat wavelet (a), and its Fourier transform (b). It is the (negative normalized) second derivative of the Gaussian.

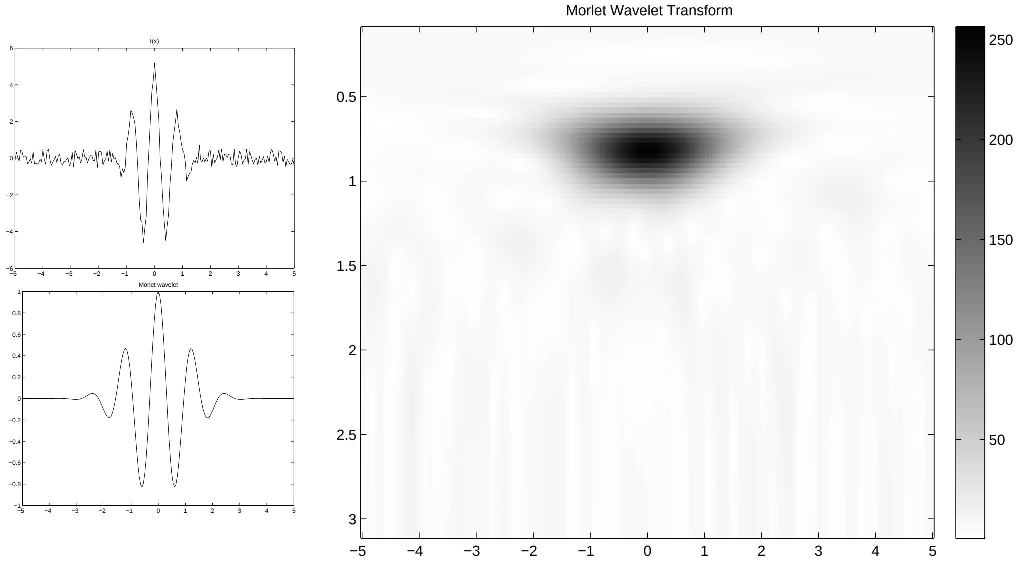
$$\psi(x) = \frac{2}{\sqrt{3\sigma\pi^{\frac{1}{4}}}}\left(1 - \frac{x^2}{\sigma^2}\right) e^{-\frac{x^2}{2\sigma^2}}$$

CWT examples

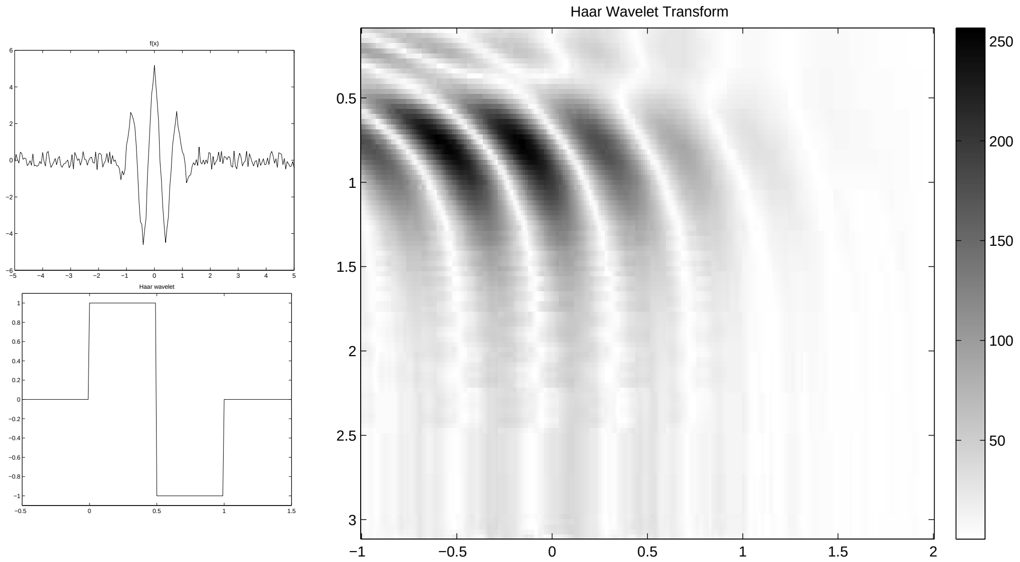


The signal $f(\cdot)$ is a Morlet wavelet shrinked by a factor of 0.667, multiplied by 5.

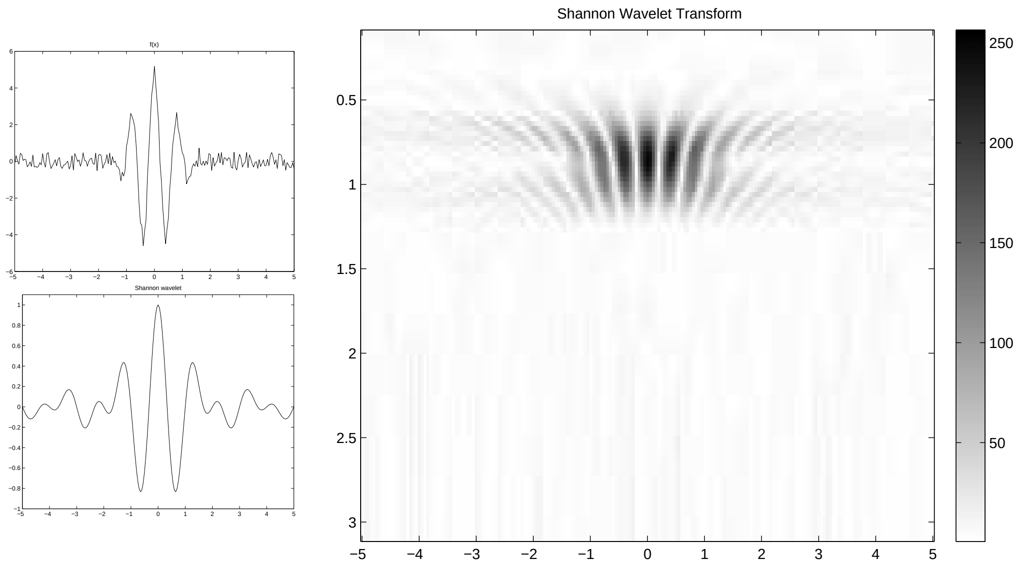
CWT examples: Morlet wavelet



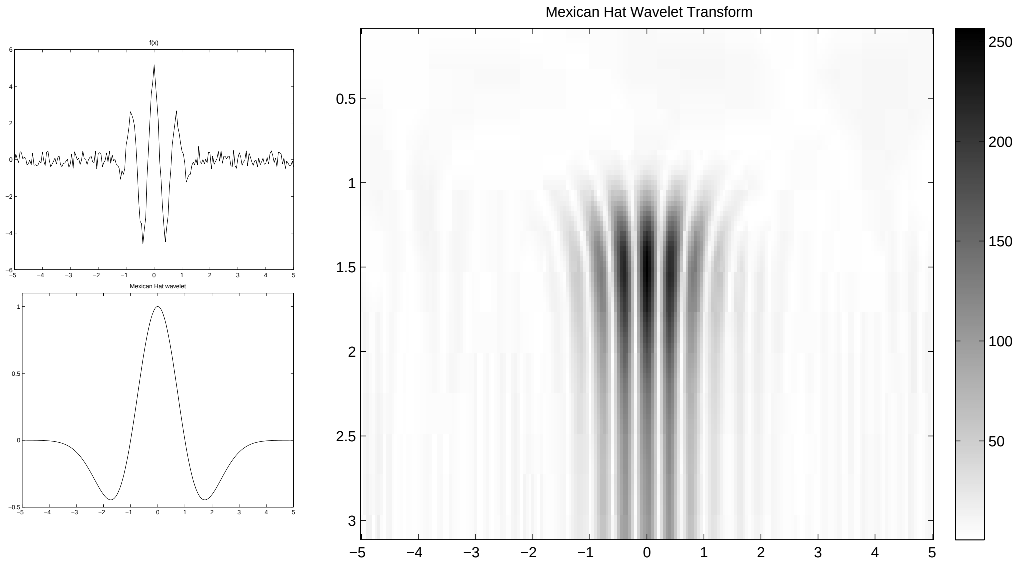
CWT examples: Haar wavelet



CWT examples: Shannon wavelet



CWT examples: Mexican hat wavelet



Space-frequency locality

The quantities $\bar{x}$, Δ_x, $\bar{\nu}$, Δ_ν

$$\bar{x} = \frac{1}{||\psi(x)||^2} \int_{-\infty}^{\infty} x\, |\psi(x)|^2 \, \mathrm{d}x$$

$$\Delta_x^2 = \frac{1}{||\psi(x)||^2} \int_{-\infty}^{\infty} (x - \bar{x})^2\, |\psi(x)|^2 \, \mathrm{d}x$$

$$\bar{\nu} = \frac{1}{||\hat{\psi}(\nu)||^2} \int_{-\infty}^{\infty} \nu\, |\hat{\psi}(\nu)|^2 \, \mathrm{d}\nu$$

$$\Delta_\nu^2 = \frac{1}{||\hat{\psi}(\nu)||^2} \int_{-\infty}^{\infty} (\nu - \bar{\nu})^2\, |\hat{\psi}(\nu)|^2 \, \mathrm{d}\nu$$

characterize the wavelet's distribution in the space and frequency domains.

Space-frequency locality (2)

- In fact, $\bar{\nu}$ is the center of mass of the wavelet in the space domain, and the energy of ψ is concentrated in a $2\Delta_x$ long neighborhood of $\bar{x}$.
- Same considerations hold for $\bar{\nu}$ and Δ_ν with respect to $\hat{\psi}$.
- Applying the above defined quantities to $\psi_{a,b}$, it can be shown that $\psi_{a,b}$ is concentrated around $b + a\bar{x}$ with radius $a\Delta_x$, while $\hat{\psi}$ is concentrated around $\frac{\bar{\nu}}{a}$ with radius $\frac{\Delta_\nu}{a}$.
- Hence, the region

$$[b + a\bar{x} - a\Delta_x,\ b + a\bar{x} + a\Delta_x] \times \left[\frac{\bar{\nu} - \Delta_\nu}{a}, \frac{\bar{\nu} + \Delta_\nu}{a}\right]$$

  is where the wavelet $\psi_{a,b}$ lives in the space-frequency domain.

Space-frequency locality (3)

- The region

$$[b + a\bar{x} - a\Delta_x,\ b + a\bar{x} + a\Delta_x] \times \left[\frac{\bar{\nu} - \Delta_\nu}{a}, \frac{\bar{\nu} + \Delta_\nu}{a}\right]$$

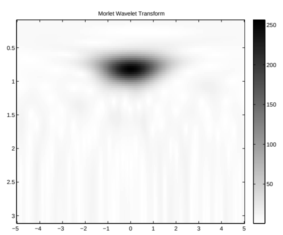
  is called the *space-frequency window* of the wavelet.
  - Actually, as usually $\bar{\nu} = 0$ and $\hat{\psi(0)} = 0$, the wavelet is localized in symmetric disjointed bands, such as $[-\frac{\nu_2}{a}, -\frac{\nu_1}{a}] \cup [\frac{\nu_1}{a}, \frac{\nu_2}{a}]$.
- The shape of the window, $2a\Delta_x \times 2\frac{\Delta_x}{a}$, depends on a.
- Its position in space depends also on b.
- The window area is constant: $4\Delta_x\Delta_\nu$.
- As it can be shown that $||\psi(x)||^2 \leq 2||x\,\psi(x)||\,||\nu\,\hat{\psi}(\nu)||$ (Heisenberg), the window size has a lower bound;
  - although it can depend upon the actual wavelet.

Space-frequency locality (4)

- Since the coefficients of the CWT $W(a,b)$ reflect the similarity of the signal f and $\psi_{a,b}$, they describe the behavior of the signal in the wavelet window.
- For $a > 1$, the wavelet is dilated, and its frequency content move toward lower frequencies.
- The opposite for $a < 1$ (the wavelet is shrinked).
- Hence, fast events can be captured by wavelets with a small a:
  - the window base is $2a\Delta_x$ wide, good localization in space
- and long events can be described by wavelet with large a:
  - the window height is $\frac{\Delta_\nu}{a}$, good frequency resolution.

Discrete representation of the CWT

- The CWT $W(a,b)$ can be represented as an image $I(i,j)$:



  - i represents a given scale, $a(i)$;
  - j represents a given position, $b(j)$;
  - the color in $I(i,j)$ is proportional to the value (modulus or phase) of the coefficient $W(a(i), b(j))$.
- This operation is a discretization of the CWT.
  - The parameters a and b are discretized.
  - It is not the Discrete Wavelet Transform.
  - It is just a sampling of $W(a,b)$.
- How choosing $a(i)$ and $b(j)$?
  - Do not loose critical information.
  - Can the signal $f(x)$ be obtained from the $\{W(a(i), b(j))\}$ sampling?

Dyadic sampling

- To be able to reconstruct a signal from its sampling, the sampling frequency have to be at least double of the maximum frequency component of the signal (Nyquist's theorem).
- The frequency content of $X(a, b)$ diminishes when a increases.
- Hence, the sampling frequency can be different for different scales.
- This allows to save computational resources.
- Usually, sampling on the *dyadic grid*:
  - logarithmic law for scales: $a = 2^{-j}$, $j \in \mathbb{Z}$;
  - translations proportional to scales: $b = ka$, $k \in \mathbb{Z}$;
  - the logarithmic law allows to cover a wide range of scales with a relatively small number of scales (and samples).
- In general, sparser the sampling, more restrictions on the wavelet.

Short-time Fourier Transform (STFT)

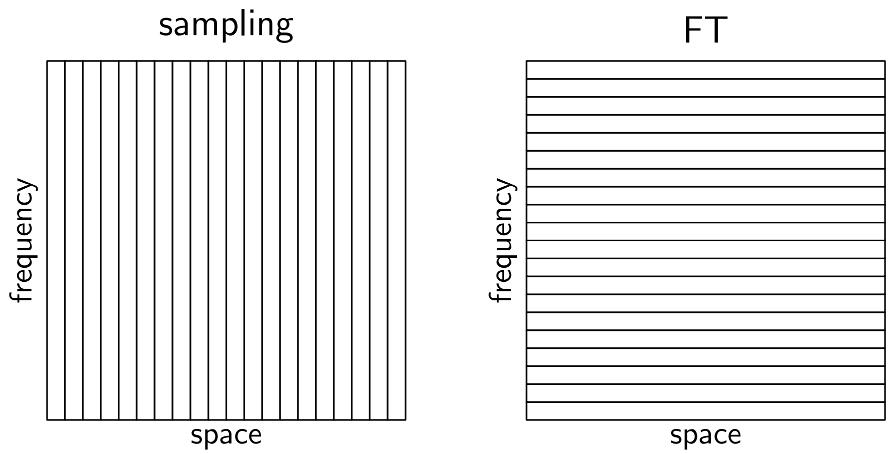
- The Short-time (or Short-term) Fourier Transform (STFT) is defined as:

$$STFT(f(x)) = W(\tau, \nu) = \int_{-\infty}^{\infty} f(x) w(x - \tau) e^{-\iota \nu x} \, \mathrm{d}x$$

- It is the FT of a signal windowed by the function $w(\cdot)$, while it slid along the space.
  - The function $w(\cdot)$ can be a Gaussian with a given width, σ.
  - This is the case of the *Gabor* transform.
- The STFT realizes a space-frequency transform.
  - How STFT compares with CWT?

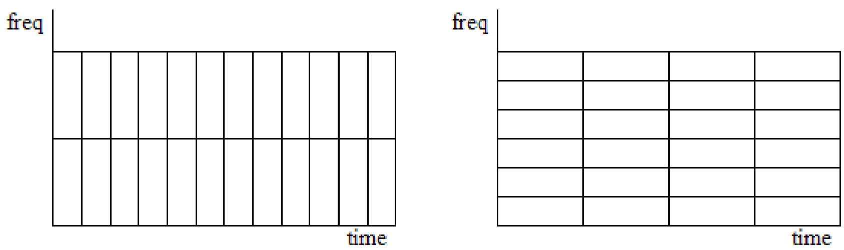
Space-frequency coverage

- Sampling vs. FT space-frequency coverage:

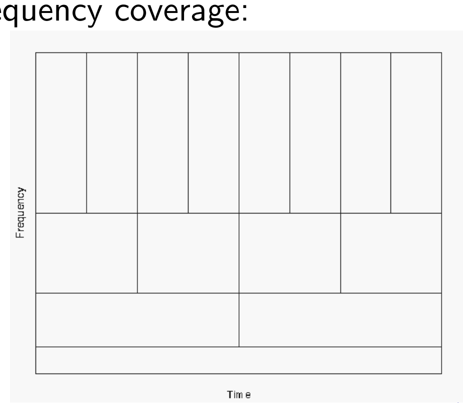


Space-frequency coverage (2)

- STFT space-frequency coverage:



- CWT space-frequency coverage:



Mathematical view of the signal transforms

- Signals can be seen as functions (real, complex).
- Usually a subset of the functions can be considered.
  - E.g., the continuous function (up to the n-th order), C^n, or the L^2 functions (finite energy).
- These sets are (infinite dimensional) vector spaces.
- The transform describes the signal as a linear combination of other functions (the basis vectors).
- Hence, the transformed signal is constituted of the coefficients of the linear combination.
  - I.e., the "importance" of each basis function for describing the signal.

Mathematical view of the signal transforms (2)

- The basis functions declared in the transforms limit the subset of the representable functions:
  - the vector space generated by the basis;
  - setting to zero some coefficients (e.g., for noise suppression) means considering only a subset of the vector space.
- Some other issues:
  - The decomposition always exists?
    - Which properties must have the signal to be transformed?
  - Is the decomposition unique?
    - Different coefficients can reconstruct the same signal?

Inner product

- The *inner product*, $\langle \cdot, \cdot \rangle$, on the vector space V is a function $V \times V \to \mathbb{R}$ such that, for each v_1, $v_2 \in V$ and $\alpha \in \mathbb{R}$:
  - $\langle v_1, v_1 \rangle \geq 0$, with $\langle v_1, v_1 \rangle = 0$ iif $v_1 = 0$;
  - $\langle v_1, v_2 \rangle = \langle v_2, v_1 \rangle$;
  - $\langle \alpha v_1, v_2 \rangle = \langle v_1, \alpha v_2 \rangle = \alpha \langle v_1, v_2 \rangle$
- An inner product induces the *norm*, $|| \cdot ||$: $||v_1|| = \sqrt{\langle v_1, v_1 \rangle}$
- Two vectors v_1 and v_2 are *orthogonal* if $\langle v_1, v_2 \rangle = 0$.
- A basis $\{v_k \,|\, v_k \in V\}$ is *orthogonal* if the vectors are orthogonal each others.
- A basis $\{v_k \,|\, v_k \in V\}$ is *orthonormal* if the vectors are orthogonal each others and have a unitary norm:

$$\langle v_k, v_j \rangle = \delta_{k-j} = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases}$$

Bases

- If $\{v_k \,|\, v_k \in V\}$ is an orthonormal basis, every vector $v \in V$ can be expressed as:

$$v = \sum_k \langle v, v_k \rangle v_k$$

- Two bases $\{v_k \,|\, v_k \in V\}$ and $\{w_k \,|\, w_k \in V\}$ are biorthogonal if:

$$\langle v_k, w_j \rangle = \delta_{k-j} = \begin{cases} 1, & k = j \\ 0, & k \neq j \end{cases}$$

  In this case, the following relations hold:

$$\forall v \in V \;\; v = \sum_k \langle v, w_k \rangle v_k \qquad \text{and} \qquad v = \sum_k \langle v, v_k \rangle w_k$$

Back to the functions

- The inner product of the functions f and g, $f, g \in L^2(\mathbb{R})$ is:

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

  - It represents the projection of a signal onto the other.
- Hence, the FT of a signal and its inverse can be seen as a decomposition in terms of basis composed of sinusoidal and cosinusoidal functions.
- Same apply for CWT and its inverse.