Noise analysis

Brief introduction to Signal Processing

Elena Cuoco1

1European Gravitational Observatory

Winter VESF School, 2010

Outline

1. Why signal processing
2. Basic concepts
   - Time series, Autocorrelation function, Spectral Density
3. Frequency domain
   - Non Parametric and Parametric spectral estimator
4. Non stationary Noise
   - Slow and Fast not stationarity noise
5. Time-Frequency Domain
   - The TF analysis
   - The wavelet based detection method
   - Example of Event Trigger Generator
6. Exercises
   - Tools

The goal of signal processing

The setting up of physical esperiment aims to prove some hyptohesis, theory, idea. To this goal it is important to analyze the data we acquired in our experiment.
What we want to obtain is the extraction of as much as possible information from our data.
A signal could be represented in different orthonormal basis. We can choice in which domain analyze the signal depending on the kind of information we would extract.
The signal we have to analyze have different features which could come up looking at them in different domain.
In this lesson I would like to show how we can use data in

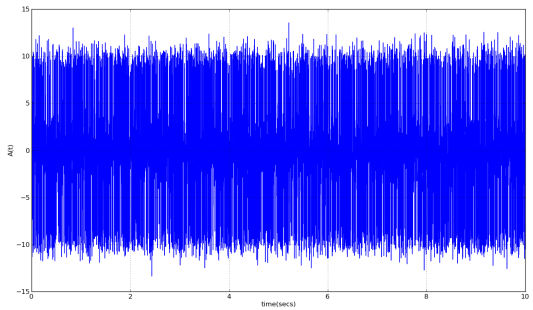
- time-domain
- frequency-domain
- time-frequency-domain

Outline

1. Why signal processing
2. Basic concepts
   - Time series, Autocorrelation function, Spectral Density
3. Frequency domain
   - Non Parametric and Parametric spectral estimator
4. Non stationary Noise
   - Slow and Fast not stationarity noise
5. Time-Frequency Domain
   - The TF analysis
   - The wavelet based detection method
   - Example of Event Trigger Generator
6. Exercises
   - Tools

Stochastic process and Time Series.

A stochastic discrete process $x[n]$ is a sequence of random variable for each value of n. If n represents the time, we call it time series. A time series is a sequence of data points measuring a physical quantity at successive times spaced at uniform time intervals.



We say that $x[n]$ is a stationary process, if its statistical description does not depend on n.

That is the moments $\mathscr{E}(x^{k_0}[n_0], x^{k_1}[n_1], \ldots x^{k_M}[n_M])$ do not depend on the value of x_M but only on the distance L between the value of $x[n]$ and $x[n+L]$.

The sampling or Shannon theorem

Theorem

A bandlimited analog signal can be perfectly reconstructed from an infinite sequence of samples if the sampling rate is greater than 2 times the highest frequency of the original signal

Let $x(t)$ be a continuous-time signal and $X(f)$its Fourier transform

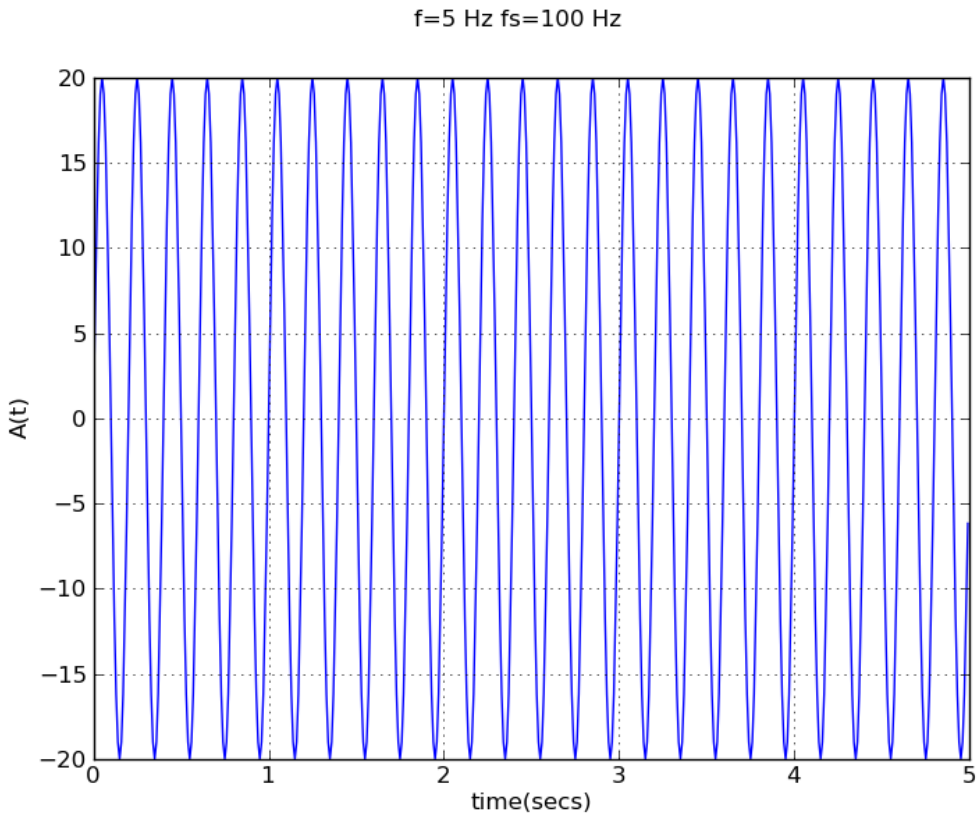
$$X(f) = \int_{-\infty}^{\infty} x(t)\exp(-2i\Pi ft)dt$$

The signal $x(t)$is band-limited to a one-sided band B if

$$X(f) = 0$$

for all $f > |B|$.

Then the sufficient condition for exact reconstructability from samples at a uniform sampling rate f_s (in samples per unit time) is that $f_s > 2B$ $f_s/2$ is called the Nyquist frequency. The time interval between successive samples is referred to as the sampling interval: $T = \frac{1}{f_s}$



See the animation

Aliasing

Suppose the true input signal has frequency f and that the sampling frequency is f_s.
The following rules allow the alias frequency f_a to be calculated.

- $f < 1/2f_s$ As the signal frequency lies below the limit set by the sampling theorem, no aliasing occurs; the sampled signal has the correct frequency.
- $1/2f_s < f < f_s$ The signal undergoes aliasing, with an alias frequency $f_a = f - f_s$
- $f > f_s$ Aliasing again occurs. To determine the alias frequency, first calculate the signal frequency f modulo the sampling frequency f_s (i.e. keep subtracting f_s from f until the result is below f_s). Then apply either of the previous two rules, depending on whether the resulting frequency value is below or above $1/2f_s$.

Autocorrelation function.

Definition

Given a discrete random process $x[n]$ we define the *mean* as

$$\mathscr{E}\{x[n]\} = \mu_x$$

Definition

The autocorrelation function (ACF)

$$r_{xx}[k] = \mathscr{E}\{x^*[n]x[n+k]\}$$

ACF

Definition

The *autocovariance* function is defined as

$$c_{xx}[k] = \mathscr{E}\{(x^*[n] - \mu_x)(x[n+k] - \mu_x)\} = r_{xx}[k] - |\mu_x|^2$$

Similar definition for cross-correlation bewteen $x[n]$ and $y[n]$.
Some properties of ACF:

$$r_{xx}[0] \geq |r_{xx}[k]| \qquad r_{xx}[-k] = r^*_{xx}[k] \qquad r_{xy}[-k] = r^*_{yx}[k]$$

Power Spectral Density.

Definition

We define the *Power Spectral Density* (PSD)

$$P_{xx}(f) = \sum_{k=-\infty}^{k=\infty} r_{xx}[k]\exp(-i2\pi fk) \quad P_{xy}(f) = \sum_{k=-\infty}^{k=\infty} r_{xy}[k]\exp(-i2\pi fk)$$

This relationship between PSD and ACF is often known as Wiener-Khinchin theorem.
The PSD describe the content in frequency in power of the signal $x[n]$.
In the following we will refer to $P_{xx}(f)$ as PSD
The PSD is periodic with period 1. The frequency interval $-1/2 \leq f \leq 1/2$ will be considered as the fundamental period.
The ACF is the inverse Fourier transform of the PSD and hence

$$r_{xx}[0] = \int_{-1/2}^{1/2} P_{xx}(f)df$$

White noise

One particular process is the discrete white noise. It si defined as a process having as ACF

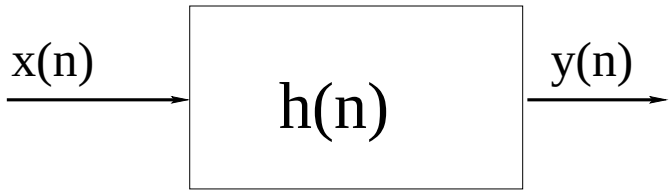
$$r_{xx}[k] = \sigma_x^2 \delta[k]$$

where $\delta[k]$ is the delta function.
The PSD of such a process is a flat function with the same value for all the frequency f

$$P_{xx}(f) = \sigma_x^2$$

Linear system

For a linear shift invariant system with impulse response $h(n)$



which has as input $x[n]$ a stationary process there are these relationship between the correlation of the input and the one of the output

$$x_{xy}[k] = h[k] * r_{xx}[k] = \sum_{l=-\infty}^{l=\infty} h[l] r_{xx}[k-l]$$

$$r_{yy}[k] = h[k] * r_{yx}[k] = \sum_{m=-\infty}^{m=\infty} h[k-m] \sum_{l=-\infty}^{l=\infty} h^*[-l] r_{xx}[m-l]$$

If we indicate with $\mathscr{H}(f)$ the Fourier Transform of $h(n)$, for the special case of a white noise input process for the linear system we have

$$P_{yy}(f) = |\mathscr{H}(f)|^2 \sigma_x^2.$$

Gaussian random process

A Gaussian sthocastic process is one for which each set $\{x[n_0], x[n_1] \ldots x[n_{N-1}]\}$ is distribuited as a multivariate Gaussian PDF. If we assume that the process is stationary with zero-mean, then the covariance matrix is the autocorrelation matrix $\mathbf{r}_{xx}$

$$\mathbf{r}_{xx} = \begin{bmatrix} r_{xx}[0] & r_{xx}[-1] & \ldots & r_{xx}[-(N-1)] \\ r_{xx}[1] & r_{xx}[0] & \ldots & r_{xx}[-(N-2)] \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}[N-1] & r_{xx}[N-2] & \ldots & r_{xx}[0] \end{bmatrix} \tag{1}$$

$$r_{xx}[k] = \mathscr{E}\{x^*[n]x[n+k]\} \, . \tag{2}$$

We can write the probability density function of a real random gaussian process a

$$P[\mathbf{x}] = \frac{1}{(2\pi)^{N/2}|\mathbf{r}_{xx}|^{1/2}} e^{\mathbf{x}T\mathbf{r}_{xx}^{-1}\mathbf{x}} \, . \tag{3}$$

Every linear operation which act on a stochastic gaussian process produces a gaussian distribuited process. For such a kind of process the second order statistics is enough to completely describe the process. The high order mode are zero.

Real white Gaussian random process

It is a process $x[n]$ with mean zero and variance σ_x^2 for which

$$x[n] \sim N(0, \sigma_x^2) \qquad -\infty < n < \infty$$

$$r_{xx}[m-n] = \mathscr{E}(x[n]x[m])) = 0 \qquad m \neq n$$

where $x \sim N(\mu_x, \sigma_x^2)$ means that $x[n]$ is Gaussian distributed with a probability density function

$$p(x) = \frac{1}{\sqrt{2\pi}\sigma_x} \exp[-\frac{1}{2}(\frac{x-\mu_x}{\sigma_x})^2] \qquad -\infty < x < \infty$$

Outline

Spectral Estimator

The way in which we can show the frequency spectral content of a signal is estimating the Power Spectral density.
We will see in the following in which way we can achivie this result.
Since we cannot deal with infinite data set, each estimation we can do will be the best compromise we can have using our data.
What we want is that our Estimator will converge to *True* value if $N \rightarrow \infty$
To estimate the PSD from our data we can use classical methods (which are called non parametric estimator) or modern spectral estimator (called parametric ones). We will go through the advantages and disavantages of both.

Non parametric Spectral estimator: Periodogram/1

The PSD is defined as

$$P_{xx}(f) : \lim_{M\to\infty} \mathscr{E}[\frac{1}{2M+1}|\sum_{n=-M}^{M} x[n]\exp(-2i\pi fn)|^2]$$

The Periodogram is defined as

$$P_{\mathrm{PER}} = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]\exp(-i2\pi fn)\right|^2 . \qquad (4)$$

It satisfies $P_{PER}(f) \to P(f)$ for $N \to \infty$, but its variance does not got to zero for $N \to \infty$; in particular it has a constant values for any N.

Periodogram/2

So the periodogram is an inconsistent estimate of the PSD of our process, being its variance as big as it mean. So, to better the statistics of periodogram we can introduce the avareged periodogram.

$$P_{average} = \frac{1}{K} \sum_{m=0}^{K-1} P_{\mathrm{PER}}^{(m)}(f), \tag{5}$$

where the periodogram is estimated on a set of data $L = N/K$.
In this way the variance is smaller of a factor $1/K$. We pay the cost for a higher *bias*, since we are using shorter data lenght set.

Periodogram/3

We can show [4] that the avareged value for the periodogram is

$$\mathscr{E}\{P_{\mathrm{PER}}\} = \int_{-1/2}^{1/2} W_B(f-\nu)P(\nu)d\nu\,, \tag{6}$$

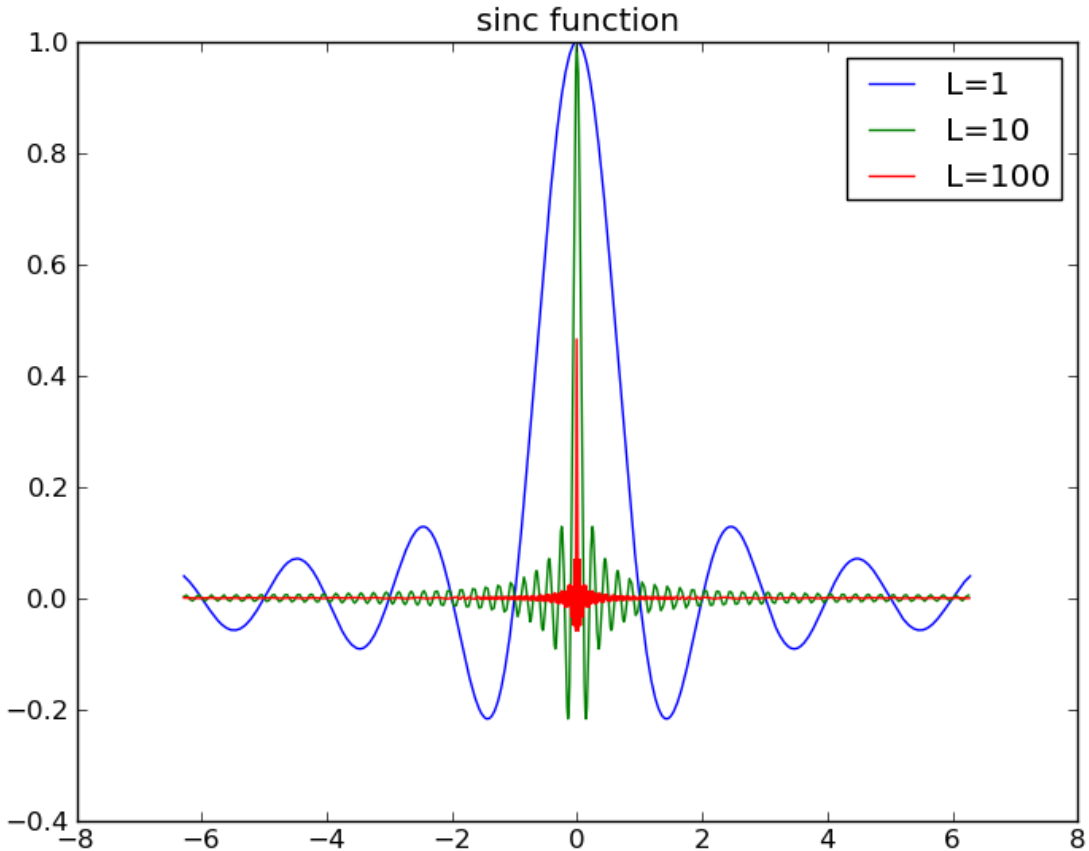
where P is PSD for the process and $W_B(f)$ is the Frourier Transform of Bartlett window

$$w_B[k] = \begin{cases} 1-\dfrac{|k|}{L} & \text{for} \quad |k| \leq L-1 \\ 0 & \text{for} \quad |k| > L \end{cases}\,, \tag{7}$$

that is

$$W_B(f) = \frac{1}{L}\left(\frac{\sin \pi f L}{\sin \pi f}\right)\,. \tag{8}$$

A shorter value for L corresponds to a narrower Bartlett (7) window and so to a broader peak in $W_B(f)$. So as L dicreases it decrease the spectral resolution too since we cannot distinguish details in the PSD smaller than $1/L$.



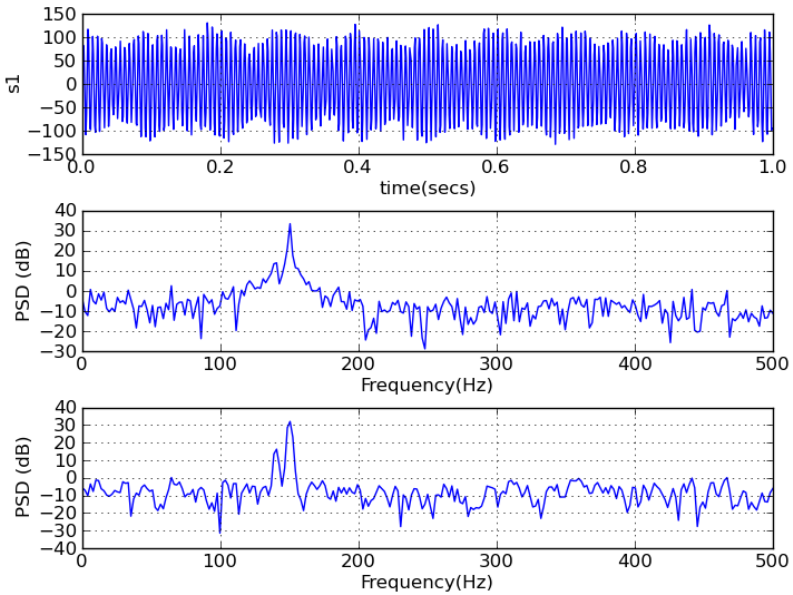
Periodogram/4: the leakage

In general, the signals we should analyze consist of sinusoidal plus colored noise. When we estimate the periodogram it could be useful to apply a data window to the process $x[n]$ before computing periodogram. This could help to do not loose narrow weak signal in the noise floor due to the presence of a higher level signal which produces sidelobes.

In order to resolve two sinusoids that are realatively close together in frequency, it is necessary for the difference between the two frequency to be greater than the width of the mainlobe of the leaked spectra for either one of these sinusoides

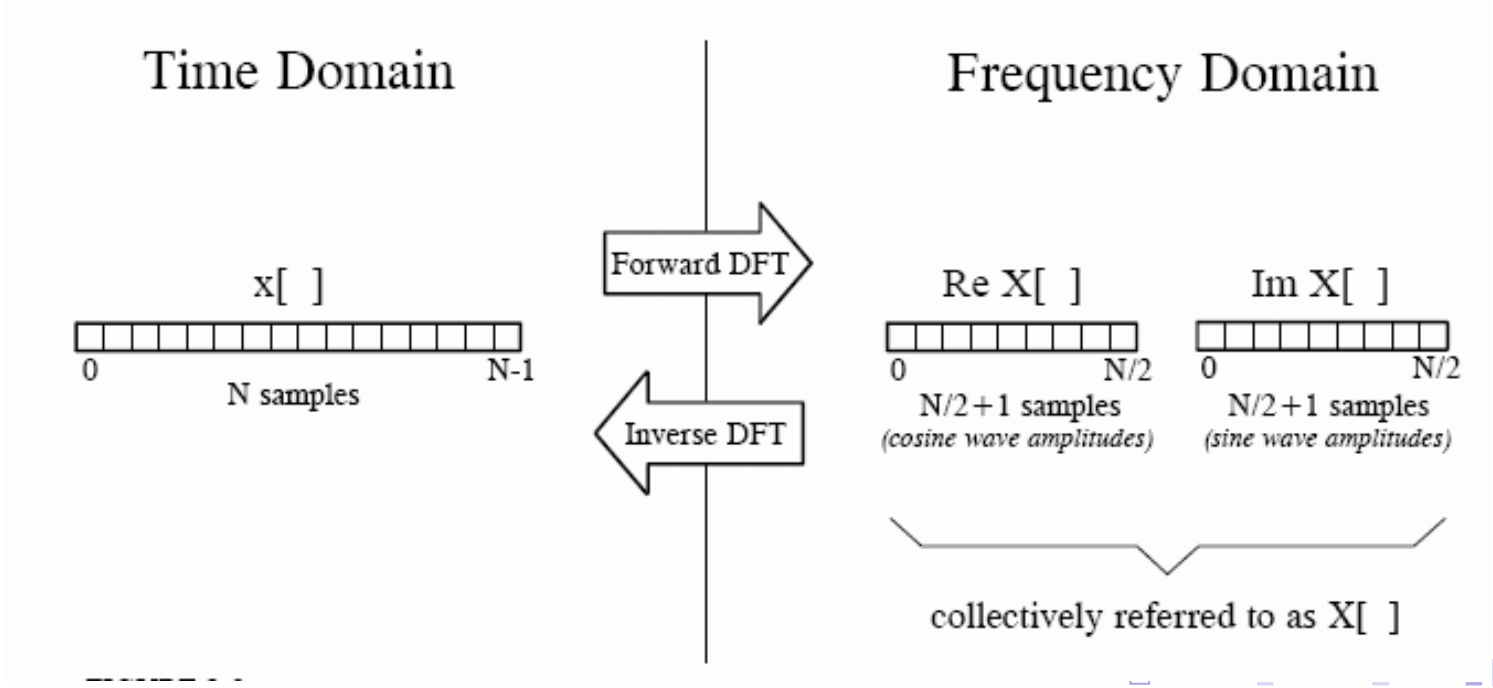
$$\delta f = (f_1 - f_2) > \frac{f_s}{L}$$

being $\frac{f_s}{L}$ the width of the mainlobe.



Discrete Fourier Transform (DFT)

If we want to look for the presence of periodical signal, we have to Fourier transform the signal.
Fourier analysis is a family of mathematical techniques, all based on decomposing signals into sinusoids.
The frequency domain contains exactly the same information as the time domain, just in a different form.
Since we are dealing with digitized signal, so we need digital algorithm to evaluate the Fourier Transform.



Welch Periodogram

- The signal is split up into overlapping segments: The original data segment is split up into L data segments of length M, overlapping by D points.
- The overlapping segments are then windowed: After the data is split up into overlapping segments, the individual L data segments have a window applied to them (in the time domain).
- The periodogram is calculated by computing the discrete Fourier transform, and then computing the squared magnitude of the result.
- The individual periodograms are then time-averaged, which reduces the variance of the individual power measurements.
- The end result is an array of power measurements vs. frequency "bin".

Parametric Spectral estimator:why?

Using a window to estimate the PSD means that we put to zero the correlation function beside a givend lag L. This is a strong approximation since we loose all the information contained in the autocorrelation for distance between data greater than L.
The procedure to estimate the PSD using parametric modeling is based on three steps:

- select the appropriate model for the process;
- estimate the model parameters from the given data;
- use these parameters in the theoretical power spectrum density for the model.

ARMA model

A general process described by a ARMA (Autoregressive Moving average) model satisfies the relation:

$$x[n] = -\sum_{k=1}^{p} a[k]x[n-k] + \sum_{k=0}^{q} b[k]w[n-k] \tag{9}$$

and its transfer function is given by $\mathscr{H}(z) = \frac{\mathscr{B}(z)}{\mathscr{A}(z)}$
where $\mathscr{A}(z) = \sum_{k=0}^{p} a[k]z^{-k}$ and $\mathscr{B}(z) = \sum_{k=0}^{q} b[k]z^{-k}$.

ARMA,AR,MA model

The PSD of the ARMA output process is

$$P_{ARMA}(f) = \sigma^2 |\frac{B(f)}{A(f)}|^2, \tag{10}$$

σ being the variance of driven white noise w, $A(f) = \mathscr{A}(2\pi i f)$ and $B(f) = \mathscr{B}(2\pi i f)$. An autoregressive (AR) process is governed by the relation

$$x[n] = -\sum_{k=1}^{p} a[k]x[n-k] + w[n], \tag{11}$$

and its PSD for a process of order P is given by

$$P_{AR}(f) = \frac{\sigma^2}{|1 + \sum_{k=1}^{P} a_k \exp(-i2\pi k f)|^2} \tag{12}$$

MA

A moving avarage (MA) process is described by

$$x[n] = \sum_{k=0}^{q} b[k]w[n-k]$$

and the PSD is

$$P_{MA}(f) = \sigma^2 |B(f)|^2$$

Estimation of the parameters

Once we selected the model for our process, we need to find the parameters for this model. The parameters of the ARMA model are linked to the autocorrelation function of the process by the Yule-Walker equations[4]. In the general case of an ARMA process we must solve a set of non linear equations while, if we specialize to an AR process (that is an all-poles model) the equations to be solved to find the AR parameters become linear.
The relationship between the parameters of the AR model and the autocorrelation function $r_{xx}(n)$ is given by the Yule–Walker equations

$$r_{xx}[k] = \begin{cases} -\sum_{l=1}^{p} a_l r_{xx}[k-l] & \text{for } k \geq 1 \\ -\sum_{l=1}^{p} a_l r_{xx}[-l] + \sigma^2 & \text{for } k = 0\,. \end{cases} \tag{13}$$

The stability of the filter

The problem of determining the AR parameters is the same of that of finding the optimal "weights vector" $\mathbf{w} = w_k$, for $k = 1, \ldots P$ for the problem of linear prediction [4]. In the linear prediction we would predict the sample $x[n]$ using the P previous observed data $\mathbf{x}[n] = \{x[n-1], x[n-2] \ldots x[n-P]\}$ building the estimate as a transversal filter:

$$\hat{x}[n] = \sum_{k=1}^{P} w_k x[n-k]\,. \tag{14}$$

We choose the coefficients of the linear predictor by minimizing a cost function that is the mean squares error $\varepsilon = \mathscr{E}[e[n]^2]$, being

$$e[n] = x[n] - \hat{x}[n] \tag{15}$$

the error we make in this prediction.

We obtain the so called Normal or Wiener-Hopf equations

$$\varepsilon_{min} = r_{xx}[0] - \sum_{k=1}^{P} w_k r_{xx}[-k], \tag{16}$$

which are identical to the Yule–Walker equations with

$$w_k = -a_k \tag{17}$$

$$\varepsilon_{min} = \sigma^2 \tag{18}$$

This relationship between AR model and linear prediction assures us to obtain a filter which is stable and causal [4].

The Burg algorithm

There are different way to find the AR parameter by a process which we assumed can be modeled as autoregressive one[4, 8]. It is possible to show that an equivalent representation for an AR process is based on the value of the autocorrelation function at lag 0 and a set of coefficient called reflection coefficient k_p, $p = 1, ..P$, being P the order of our model. The kth reflection coefficient is the partial correlation coefficient between $x[n]$ and $x[n-k]$, when the dependence of the samples in between has been removed.
The Burg algorithm, like the Levinson or Durbin ones estimates the reflection coefficients by autocorrelation of the data in a recursive way. It is important that the estimated autocorrelation function has the right form, that is a Toeplitz form (or diagonal-constant matrix). This assures the minimum-phase behavior of the filter.

The Burg Algorithm

The algorithm proceeds in the following way:

- Initialize the mean squares error as $\varepsilon_0 = r_{xx}[0]$.
- Introduce the reflection coefficients k_p, linked to the partial correlation between the $x[n]$ and $x[n-p]$ [8]:

$$k_p = \frac{1}{\varepsilon_{p-1}} \left[r_{xx}[p] - \sum_{j=1}^{p-1} a_j^{(p-1)} r_{xx}[p-j] \right] \tag{19}$$

- At the p stage the parameter of the model is equal at the pth reflection coefficient

$$a_p^{(p)} = k_p \tag{20}$$

Burg algorithm

- The other parameters are updated in the following way:

For $1 \leq j \leq p-1$

$$a_j^{(p)} = a_j^{(p-1)} - k_p a_{p-j}^{(p-1)} \tag{21}$$

$$\varepsilon_p = (1 - k_p^2)\varepsilon_{p-1} \tag{22}$$

- At the end of the p loop, when $p = P$, the final AR parameters are

$$a_j = a_j^{(P)}, \qquad \sigma^2 = \varepsilon_P \tag{23}$$

Burg algorithm:example

In figure 1 we report an example of Burg fit with an AR model of 512 parameters to Virgo-like simulated noise.

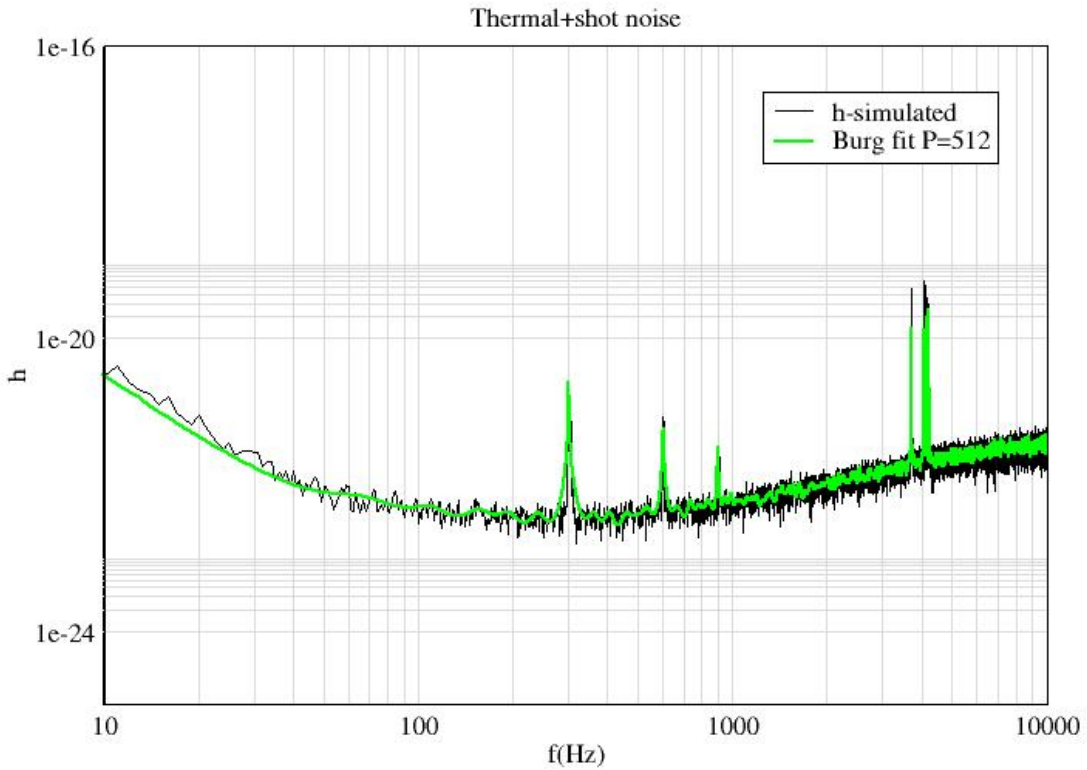


Figure: Example of Burg fit to Virgo-like simulated data

Whitening/1

We call whitening procedure the operation of making flat, that is delta correlated, a PSD which is colored, and thus characterized by different slopes with respect to the frequency.

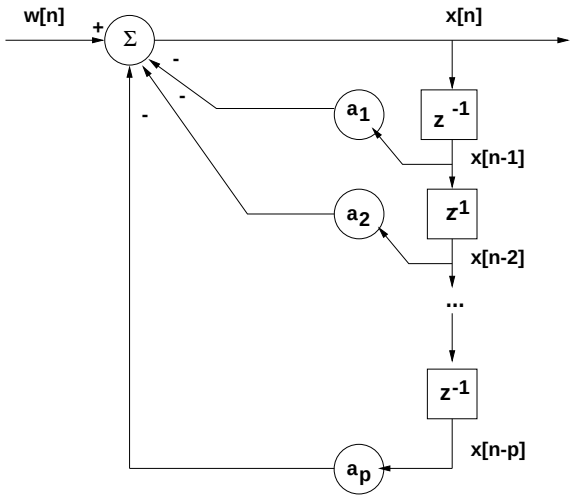
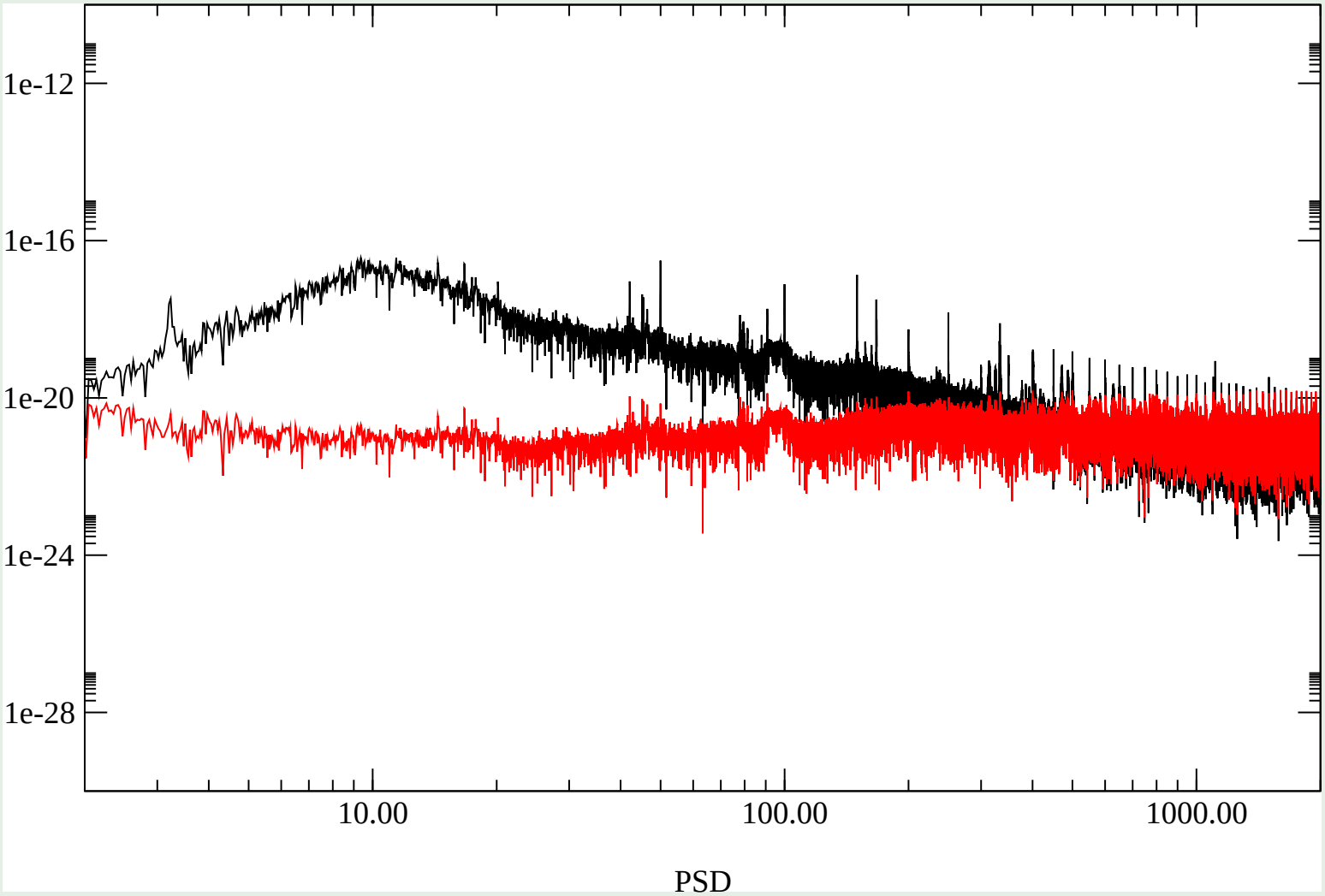


Figure: Link between AR filter and whitening filter.The tight relation between the AR filter and the whitening filter is clear in the figure. The figure describes how an AR process colors a white process at the input of the filter if you look at the picture from left to right. If you read the picture from right to left you see a colored process at the input that pass through the AR inverse filter coming out as a white process.

Whitening/2

Example



Outline

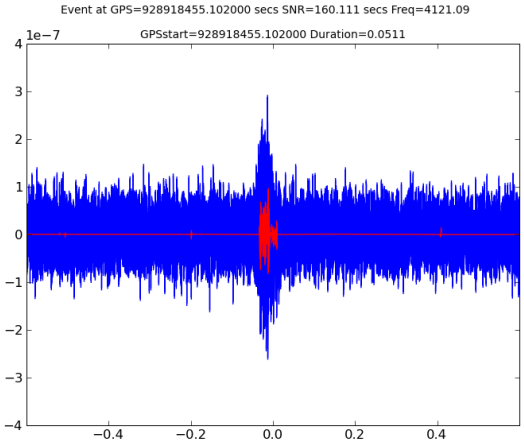
1. Why signal processing
2. Basic concepts
   - Time series, Autocorrelation function, Spectral Density
3. Frequency domain
   - Non Parametric and Parametric spectral estimator
4. Non stationary Noise
   - Slow and Fast not stationarity noise
5. Time-Frequency Domain
   - The TF analysis
   - The wavelet based detection method
   - Example of Event Trigger Generator
6. Exercises
   - Tools

Real data: non stationary noise

The data we have to analyze are far away from being white gaussian noise plus sinusoidal signals. The noise we have to face in the interferometer is full of features which make more difficult doing analysis.
Often the noise is not stationary, since some source of noise change in time.
We have to tune our analysis depending on the kind of non stationary noise, that is if we want to follow or understand slow not stationary or identify fast transient noise.
The latter are the most “dangeuros” for our detection analysis, since they can mimic real gravitational signal.

Real data: transient noise

In our data often there are transient signal (glitches) which are due only to noise.
For example a lightening, a thunder, an ariplane flying over us could produce in the data a signal like the one in the picture below
Real data



For this kind of signal we need the help of different analysis from time or frequency domain. We need to have information on how the energy content of the signal varies with time and in frequency.

Outline

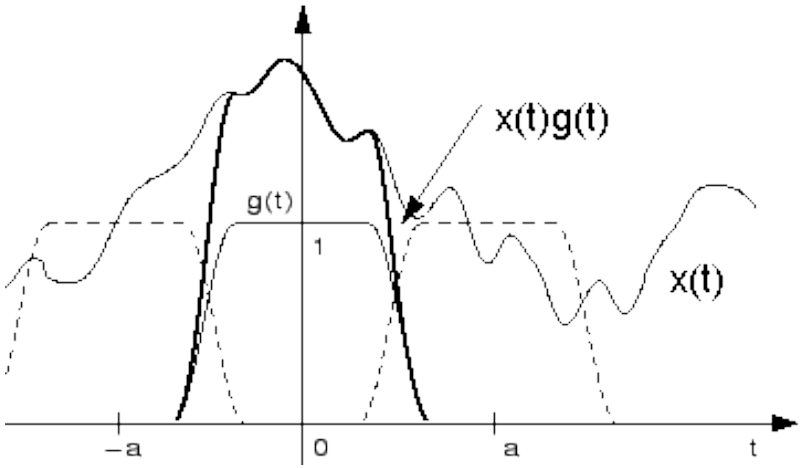
The time-frequency domain

We need for a combined time-frequency representation because neither time domain or frequency domain analysis can fully describe the nature of non-stationary signals. A time frequency distribution of a signal provides information about how the spectral content of the signal evolves with time This is performed by mapping a one dimensional signal in the time domain, into a two dimensional time-frequency representation of the signal. A variety of methods for obtaining the energy density of a function, simultaneously in the time and frequency have been devised. We will talk about the short time Fourier transform andthe wavelet transform.

The Short Fourier Transform

One approach which can give information on the time resolution of the spectrum is the short time fourier transform (STFT).
The short time Fourier transform (STFT) function is simply the Fourier transform operating on a small section of the data. Here a moving window is applied to the signal and the fourier transform is applied to the signal within the window as the window is moved.

$$STFT\{x(t)\} = X(\tau, f) = \int_{-\infty}^{\infty} x(t)g(t-\tau)\exp(-2i\pi ft)dt$$

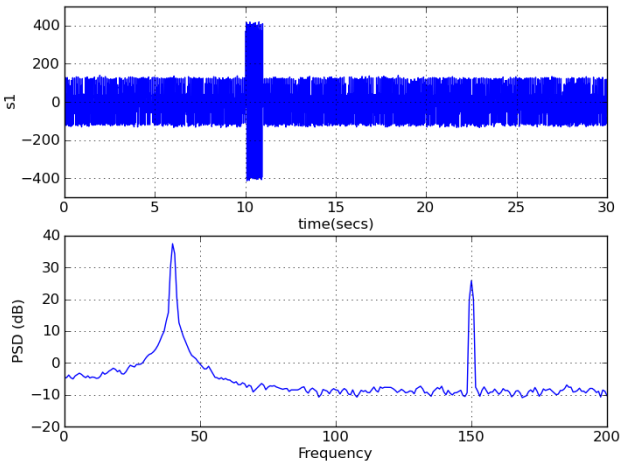


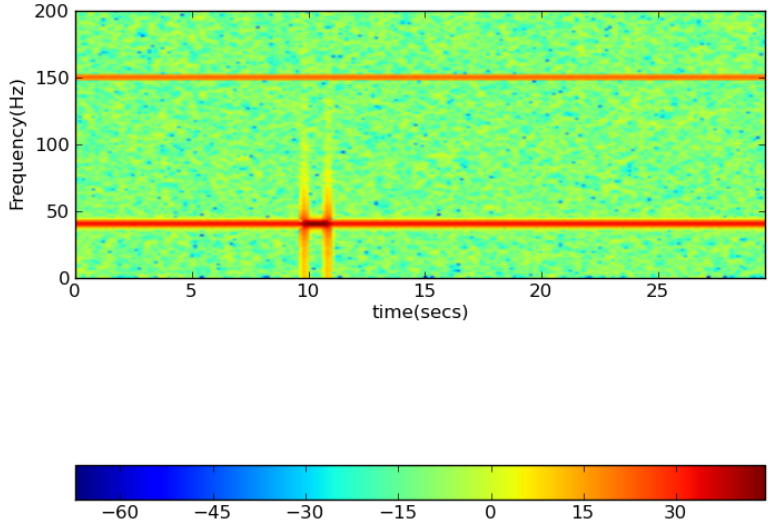
The Spectrogram

To have easy access to the information of the STFT we can plot the spectrogram.
It is defined as

$$Spectrogram(\tau, f) = |X(\tau, f)|^2$$

So we will have a bidimensional plot where on x-axis usually is plotted the time, on y-axis the frequency, while the color of the map is the the amplitude of a particular frequency at a particular time.

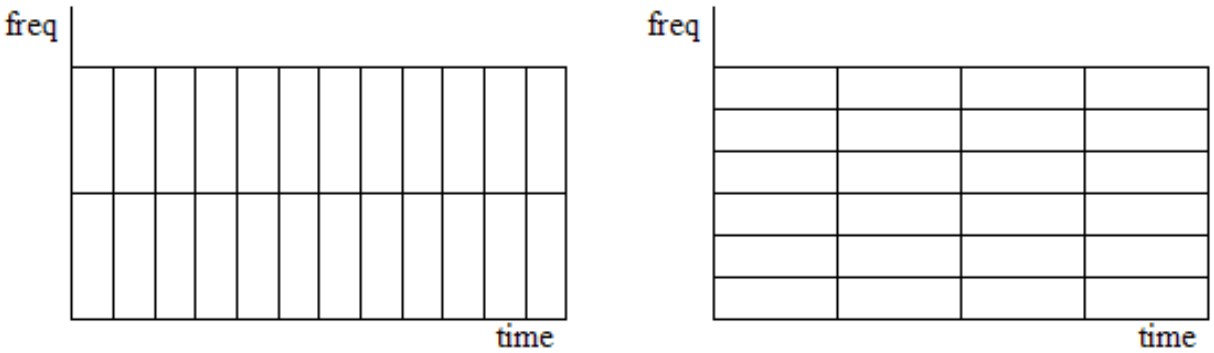




The limit of STFT

One of the problem of the STFT is that the resolution if fixed by the dimension of the window. If we choose a narrow window we will have better time resolution but worste frequency, viceversa to have a better frequency resolution we need a larger window.
This is a general limit given by time-frequency Heinsenberg uncertainty principle:

$$\Delta t \Delta \omega \geq \frac{1}{2}$$

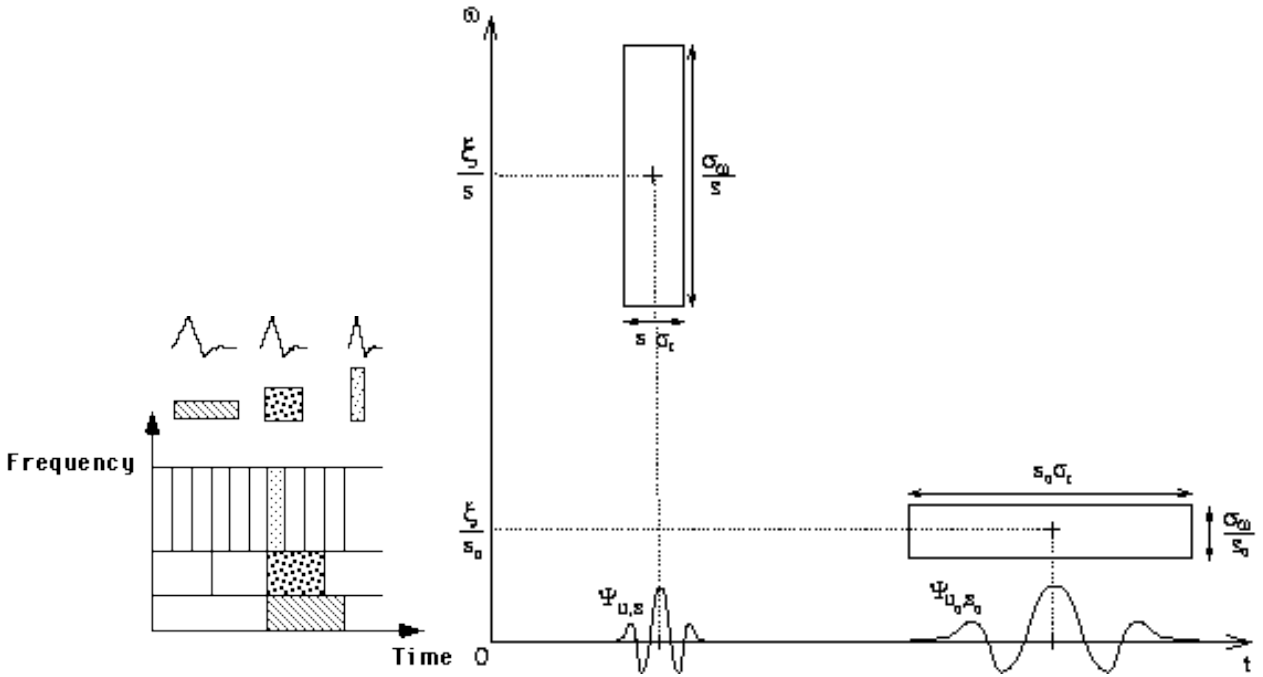


That's why in the last decades multiresolution analysis is used to analyze non stationary data.

Outline

Multiresolution analysis

Using multiresolution analysis we could choice to have different resolution in time and in frequency at different frequency.



The wavelets

The wavelet transform replaces the Fourier transform sinusoidal waves by a family generated by translations and dilations of a window called a wavelet.
The wavelet transform of a signal $f(t)$ is defined as

$$Wf(a,b) = < f, \psi_{a,b} > = \int_{-\infty}^{+\infty} f(t) \frac{1}{\sqrt{b}} \psi^*(\frac{t-a}{b}) \; dt \qquad (24)$$

where the base is a zero average function, centered around zero and with a finite energy. The entire base is obtained by translations and dilations of the base atom:

$$\psi_{ab}(t) \quad = \quad \frac{1}{\sqrt{b}} \psi(\frac{t-a}{b}) \qquad (25)$$

The wavelet transform has a time frequency resolution which depends on the scale b. Its time spread is proportional to b and its frequency spread is proportional to the inverse of b.

The discrete wavelet transform

The discrete wavelet transform is an implementation of the wavelet transform using a discrete set of the wavelet scales and translations. This transform decomposes the signal into mutually orthogonal set of wavelets.
The most popular known family of orthonormal wavelets are the Daubechies ones. The Daubechies wavelets are usually denominated by the number of non zero coefficients, for this reason we usually talk about Daubechies 4, Daubechies 6 etc. wavelets. Increasing the number of non zero coefficients the functions become smoother. In figure 3 are reported the wavelet functions for DaubC4 and DaubC20.

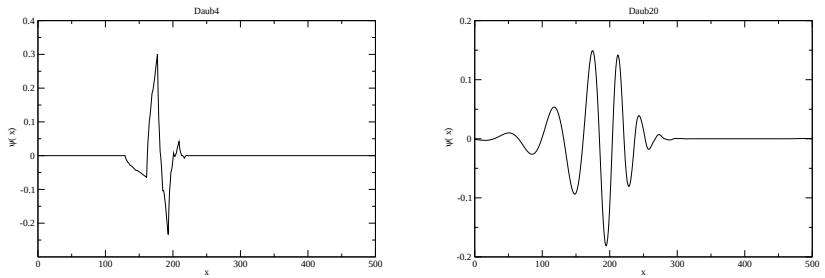
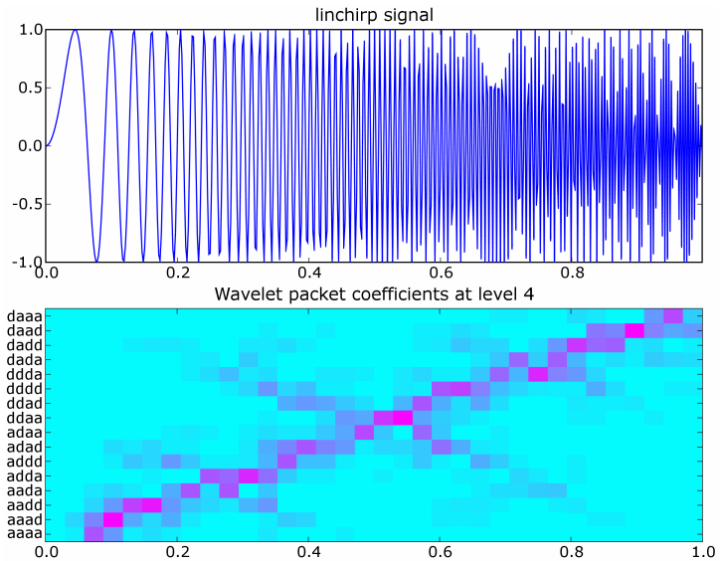


Figure: DaubC4 and DaubC20 wavelet function

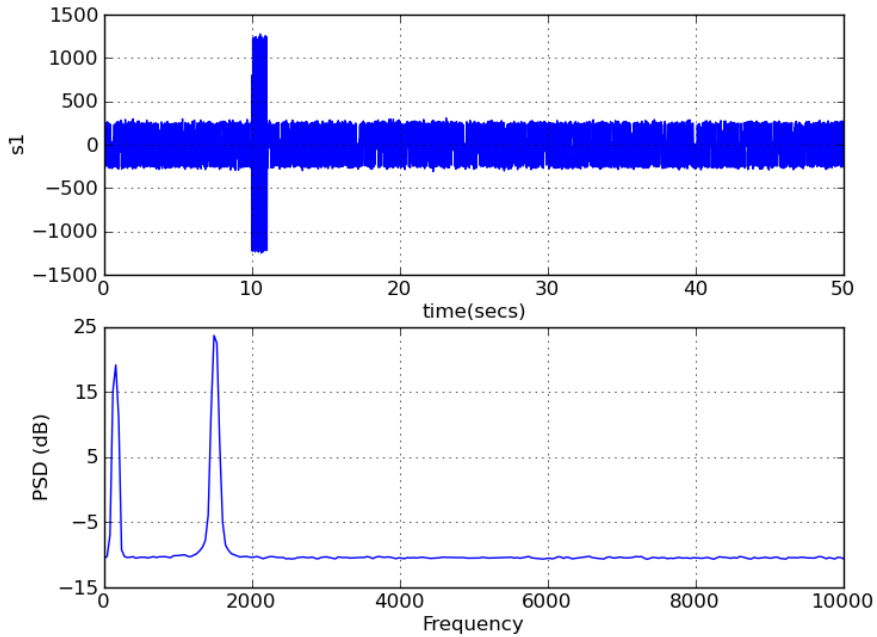
The scalogram

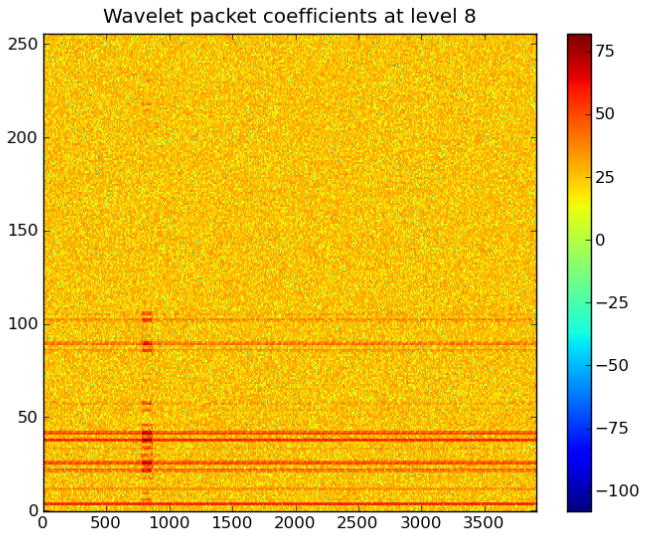
We can define the scalogram of function x .
If h denotes the frequency center of the base wavelet, then the frequency center of a dilated wavelet is f=h/s. The scalogram of a signal is defined by

$$P_w(x(u,\xi)) = |Wx(u,s)|^2 = |Wx(u,\eta/\xi)|^2$$



Example of Wavelet application





Outline

Wavelet de-noising strategy

Let us consider a signal x_i which is corrupted by additive Gaussian random noise $n_i \sim N(0, \sigma^2)$ as follows

$$x_i = h_i + n_i \quad i = 0, 1, \ldots N-1$$

From this noise signal h_i, we want to find an approximation $\hat{h}_i$ to the original h_i, which minimizes the mean squared error

$$\|\mathbf{h} - \hat{\mathbf{h}}\|^2 = \frac{1}{N} \sum_{i=0}^{N-1} |h_i - \hat{h_i}|^2$$

Let W be an orthogonal wavelet transform. If we apply it to the sequence of data x_i we obtain

$$W(x) = W(h) + W(n)$$

Since W is an orthogonal transform, $W(n_i)$ are also Gaussian random variables with distribution $N(0, \sigma^2)$.
Now let T be a wavelet thresholding function. Then the wavelet thresholding based de-noising schemes can be written

$$\hat{h} = W^{-1}(T(Wx))$$

that is we first take the wavelet transform of our noisy signal and pass it through the thresholding function, then the output is inverse wavelet transformed. The effectiveness of this technique is influenced by the choice of wavelet used, the decomposition level, and the threshold (both amplitude and type).

Thresholding function

The idea of the thresholding procedure, is to retain in the wavelet transform only the highest coefficients, which are supposed to be due to the transient signals, putting to zero all the coefficients which are smaller than a given threshold, which gives a measure of the background noise.

- The Donoho-Johnston threshold Estimator

Dohone and Johnston proposed two different thresholding strategy: the soft thresholding and the hard thresholding. Given a threshold t and w the wavelet coefficient, the hard threshold for the signal is w if $|w| > t$, and is 0 if $|w| < t$. The soft threshold for the signal is $sign(w)(|w| - t)$ if $|w| > t$ and is 0 if $|w| < t$.
We used the universal Donoho and Johnstone [**?**] method

$$t = \sqrt{2 \log N} \hat{\sigma} \tag{26}$$

where N is the number of data points and $\hat{\sigma}$ is an estimate of the noise level σ.
They suggested as method for the estimation of σ, the value of the median of the empirical wavelet coefficients. In particular their threshold is

Transient detection method

The wavelet coefficients contain the energy of the signal at different scale. After the wavelet threshold, we have selected the highest coefficients of the wavelet transform which are supposed to contain only the signal and not the noise.
So we build the energy of the signal as the sum of the square of all the coefficients which are above the threshold of the noise as in the following equation

$$E_s = \sqrt{\sum_{k,j} w_{k,j}^2} \tag{28}$$

being $w_{k,j}$ the wavelet coefficients above the threshold.

In this way E_s represents the signal energy content, so we can build our signal to noise ratio, as

$$SNR = \frac{E_s}{\sigma} \tag{29}$$

being σ the variance of the noise, and use this function as our receiver. We select the events whose SNR , defined in eq. (29), is above a given value. In our case we decide to retain only events with SNR greater than a given value.

Denoising:High SNR example

In Figure 4 we display the results of applying the wavelet de-noising method to a high SNR transient event detected
We show in this Figure also the comparison for the reconstruction of different wavelet functions.

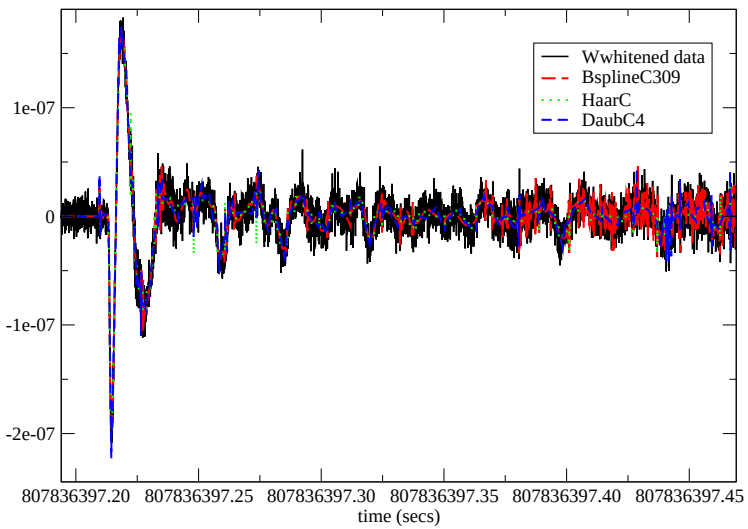


Figure: Reconstruction of high SNR transient waveform after wavelet de-noising.

The wavelet de-noising strategy is very effective for this high SNR event, and that the reconstructed waveform fits the transient signal well.

Outline

1. Why signal processing
2. Basic concepts
   - Time series, Autocorrelation function, Spectral Density
3. Frequency domain
   - Non Parametric and Parametric spectral estimator
4. Non stationary Noise
   - Slow and Fast not stationarity noise
5. Time-Frequency Domain
   - The TF analysis
   - The wavelet based detection method
   - Example of Event Trigger Generator
6. Exercises
   - Tools

Tools

1. python and matplotib tools
2. dataDisplay
3. NAP software
4. xmgrace to do plot from ascii files
5. A set of real data acquired during C7 is at disposal at /data/mdc/VESF_DA/ffl
6. A set of simulated gaussian noise is at disposal at /data/mdc/VESF_DA/C7gaussianNoise

Let's go on with exercises

For Further Reading I

- Beccaria M Cuoco E Curci G 1997 *Adaptive System Identification of VIRGO-like noise spectrum Proc. of 2nd Amaldi Conference* World Scientific
- E. Cuoco et al Class.Quant.Grav. 18 (2001) 1727-1752
- E. Cuoco et al Phys. Rev. D 64, 122002 (2001)
- Kay S 1988 *Modern spectral estimation:Theory and Application* Prentice Hall Englewood Cliffs
- Haykin S 1996*Adaptive Filter Theory* (Upper Saddle River: Prentice Hall)
- Alexander S T 1986 *Adaptive Signal Processing* (Berlin: Springer-Verlag)
- Hayes M H 1996*Statistical Digital Signal Processing and Modeling* Wiley

For Further Reading II

- Therrien C W 1992*Discrete Random Signals and Statistical Signal Processing* (Englewood Cliffs: Prentice Hall)
- Zubakov L A Wainstein V D 1962 *Extraction of signals from noise* (Englewood Cliffs: Prentice Hall)
- Parzen E 1977 *An approach to Time series modeling: determining the order of approximating autoregressive schemes in "Multivariate Analysis"* North Holland
- Widrow B and Stearns S D 1985 *Adaptive Signal Processing* (Englewood Cliffs: Prentice Hall)
- Orfanidis S J 1996 *Introduction to Signal Processing* (Englewood Cliffs: Prentice-Hall)
- http://wwwcascina.virgo.infn.it/DataAnalysis/Noise/nap_project.html
- http://www.python.org/

For Further Reading III


http://docs.scipy.org/doc/


http://matplotlib.sourceforge.net/