Signal and Noise for Gravitational Antennas

ARMA models and frequency domain operations

by Sergio Frasca

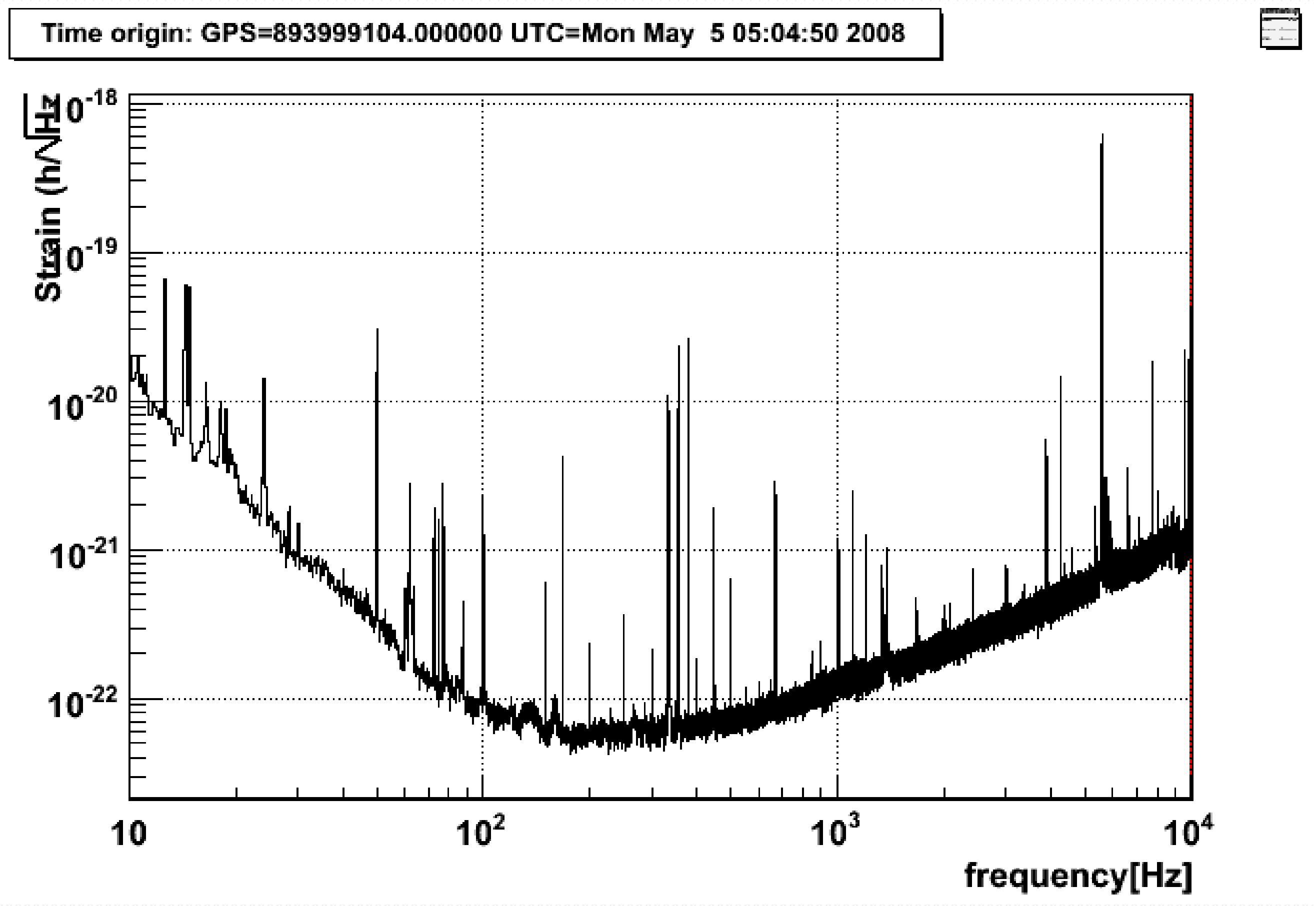
The Problem

- A gravitational antenna is a detector dominated by the noise
- The main part of this noise (when the detector is well-behaved) is gaussian and almost stationary, i.e. stationary on a certain time scale
- Summed to this gaussian noise there are also “disturbances” (mainly pulses or spectral lines). This part of noise is intrinsically non-stationary and difficult to model
- The basic problem is to describe, simulate and manage (filter) the gaussian noise.

Virgo h spectral density

- To show the gaussian noise characteristics of a gravitational antennas, we compute the **square root of the power spectrum** of the noise observed at the output of the detection system, as it was caused by gravitational waves.
- The power spectrum is the Fourier transform of the autocorrelation.
- Because of the Wiener-Kinchin teorem, it can be computed by a mean of many **periodograms**. (A periodogram is the absolute value of a segment of data $\{x_1, x_2, \ldots, x_N\}$)

Virgo h spectral density



Basic results for gaussian processes

- The model of the gaussian noise is a zero-mean gaussian stochastic process
- A gaussian stochastic process is completely described by the second-order distribution function

or, else, by the autocorrelation function

$$R_{xx}(t_1, t_2) = E\left[x(t_1) \cdot x(t_2)\right]$$

in case of stationarity, the autocorrelation depends only on the difference $\tau = t_2 - t_1$, so

$$R_{xx}(\tau) = E\left[x(t) \cdot x(t+\tau)\right]$$

Basic results for gaussian processes (cont)

- The power spectrum is the Fourier transform of the autocorrelation, and often it is more convenient (and more easy to understand) to use the power spectrum instead of the autocorrelation

- To summarize:

A gaussian stochastic process is completely defined if the power spectrum is given

Gaussian processes and linear systems

- Any linear operation on a gaussian process produces a gaussian process

- So, if a gaussian process x, defined by the power spectrum $S_x(\omega)$, passes through a linear system with transfer function $F(j\omega)$, it produces a gaussian process y with power spectrum

$$S_y(\omega) = S_x(\omega) \cdot |F(j\omega)|^2$$

If the input noise is white,

$$S_y(\omega) = k \cdot |F(j\omega)|^2$$

Gaussian processes and linear systems

- To summarize:

Any gaussian process can be modeled as the output of a proper linear system at whose input there is white gaussian noise.

And now let us see the systems. Because we use sampled data, the systems we are interested in are the discrete time systems, typically described by a difference equation.

Discrete-time linear system

A discrete-time system is a system that operates on discrete time sequences. If it is "linear", it is completely described by the **pulse response**, i.e. the response to a "discrete delta", a sequence $\{s_i\}$ composed by all zeros and a 1 at i = 0.

If $\{f_i\}$ is the pulse response, the output sequence to an input sequence $\{x_i\}$ is

$$y_i = \sum_{k=-\infty}^{\infty} x_{i-k} \cdot f_k$$

Note that the system can be a-causal.

We have neglected the initial "status" (we consider it null).

z-transform

The z-transform is a very useful instrument to study linear discrete-time system. The z-transform of a sequence $\{x_i\}$ is

$$X(z) = \sum_{i=-\infty}^{\infty} x_i \cdot z^{-i}$$

where z is a complex variable.

There are some similarities with the Laplace transform. As the variable **s** of the Laplace transform can be seen as the transform of the derivative operator, the variable **z** can be seen as the transform of the advance operator (and z^{-1} as the delay operator). So if the sequence is seen as the time function

$$x(t) = \sum_k x_k \cdot \delta(t - k \cdot \Delta t)$$

its Laplace transform is $X\left(e^{s \cdot \Delta t}\right)$ simply substituting $z = e^{s \cdot \Delta t}$.

Note that the frequency axis, that is the imaginary part of s, becomes the **unit circle** in the z plane.

General linear difference equation

A linear discrete-time system can be described also by the equation

$$\sum_{k=0}^{n} a_k \cdot y_{i-k} = \sum_{k=0}^{m} b_k \cdot x_{i-k}$$

Note: $a_0 = 1$

Where $\{x_i\}$ is the sequence of the input and $\{y_i\}$ is the sequence of the output.
Remembering the definition of z-transform, we put

$$X(z) = \sum_{i=-\infty}^{\infty} x_i \cdot z^{-i} \qquad Y(z) = \sum_{i=-\infty}^{\infty} y_i \cdot z^{-i}$$

$$A(z) = \sum_{i=0}^{n} a_i \cdot z^{-i} \qquad B(z) = \sum_{i=0}^{m} b_i \cdot z^{-i}$$

And we have

$$A(z)Y(z) = B(z)X(z)$$

Transfer function

We can reshape the equation as

$$Y(z) = \frac{B(z)}{A(z)} \cdot X(z)$$

and we have the (z domain) **transfer function**:

$$F(z) = \frac{B(z)}{A(z)} = \frac{\sum_{i=0}^{m} b_i \cdot z^{-i}}{1 + \sum_{i=1}^{n} a_i \cdot z^{-i}}$$

This is a division between two polynomials. If we perform this division, we obtain

$$F(z) = \sum_{i=0}^{\infty} f_i \cdot z^{-i}$$

that is the pulse response of the system and the ***inverse transform*** of the transfer function.

The frequency response

Because the relation between z and s, we easily solve the problem of the frequency response of the system: simply compute the transfer function on the unit circle, putting $z = e^{j\omega\Delta t} = e^{j\Omega}$.

So

$$\mathscr{F}(\Omega) = F\left(e^{j\cdot\Omega}\right) = \frac{\sum_{k=0}^{m} b_k \cdot e^{-j\cdot\Omega\cdot k}}{1 + \sum_{k=1}^{n} a_k \cdot e^{-j\cdot\Omega\cdot k}}$$

The MA (moving-average) model

If the dependence on the past outputs is absent, the system equation is

$$y_i = \sum_{k=0}^{m} b_k \cdot x_{i-k}$$

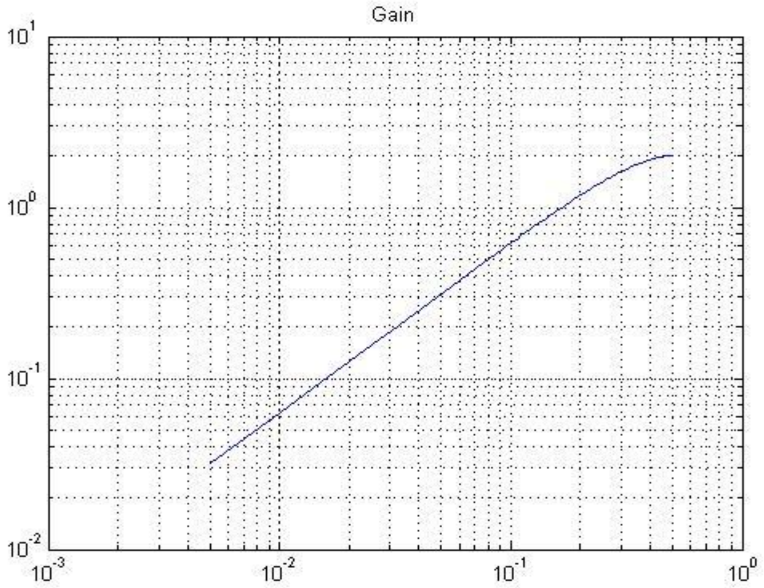
and the transfer function have only the numerator, i.e.

$$F(z) = \sum_{i=0}^{m} b_i \cdot z^{-i}$$

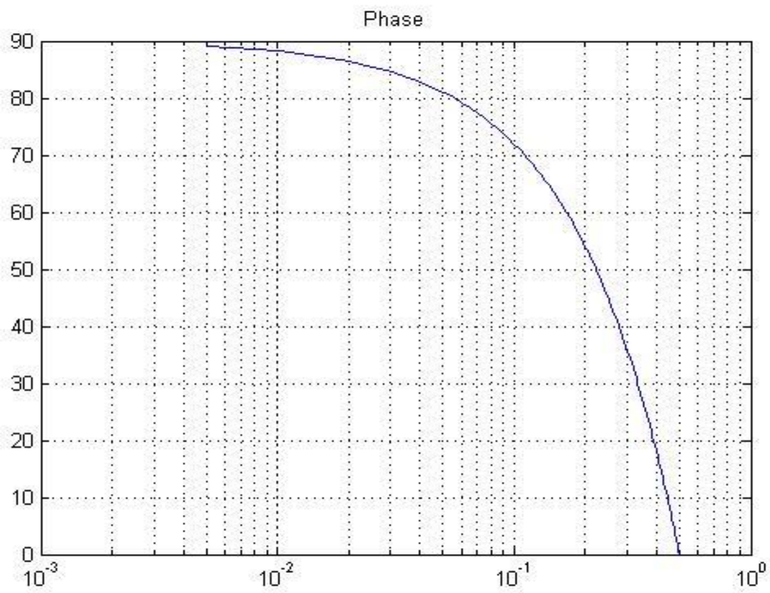
These systems are called FIR (finite impulse response) or MA systems. The impulse response is given by the m+1 coefficients.

Particular case of MA system: first order, b_0=1, b_1=-1

$$y_i = x_i - x_{i-1}$$



normalized frequency



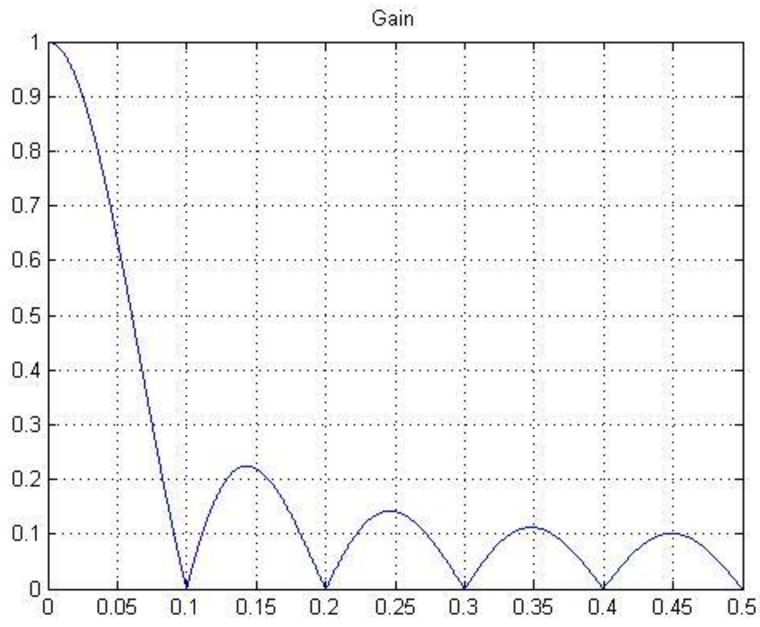
normalized frequency

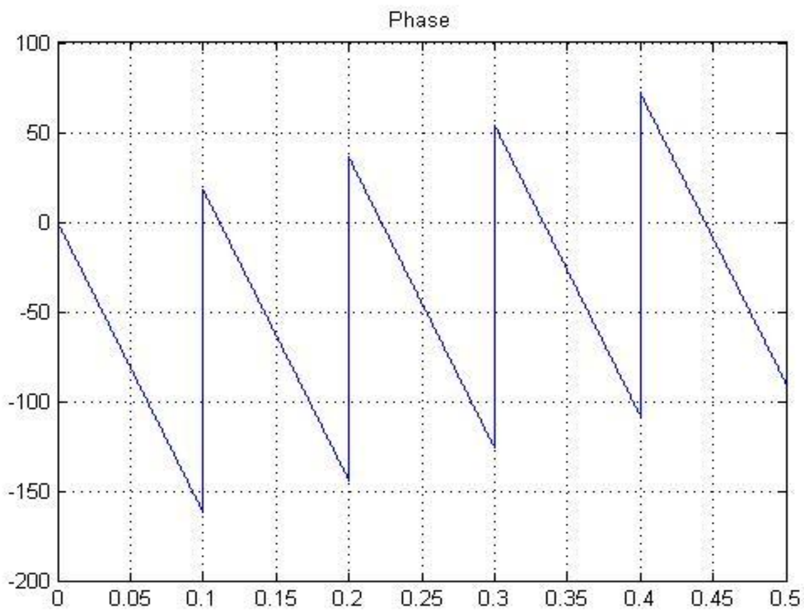
Here and in some subsequent slide the absolute value and the phase of the transfer function is shown

Particular case of MA system:
N-1 order, simple mean

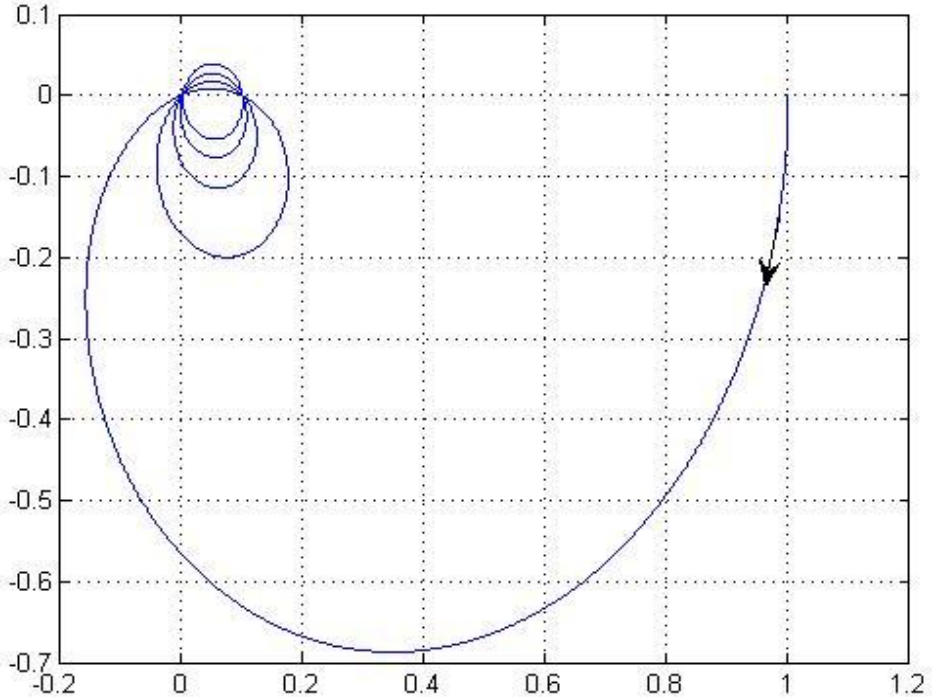
$$y_i = \frac{1}{N} \cdot \sum_{k=0}^{N-1} x_{i-k}$$

N=10





Nyquist diagram
(polar coordinates)



On the axis there is the real and imaginary part of the transfer function

The AR (auto-regressive) model

If the output depends only on the present input and some of the past outputs, i.e. we have

$$y_i = b_0 \cdot x_i - \sum_{k=1}^{n} a_k \cdot y_{i-k}$$

and the transfer function is

$$F(z) = \frac{b_0}{1 + \sum_{i=1}^{n} a_i \cdot z^{-i}}$$

the system is called IIR (infinite impulse response) or AR.

Particular case of AR system:

first order, $-1 < a_1 < 0 \quad a_1=-w$

equation :

$$y_i = b_0 \cdot x_i + w \cdot y_{i-1}$$

impulse response

$$x_0 = b_0$$

$$x_1 = b_0 \cdot w$$

$$x_2 = b_0 \cdot w^2$$

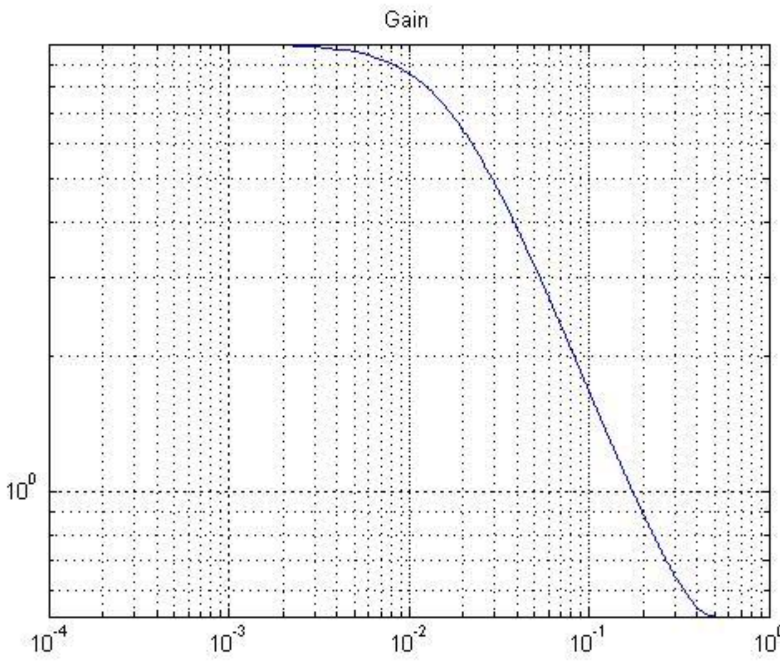
$$x_3 = b_0 \cdot w^3$$

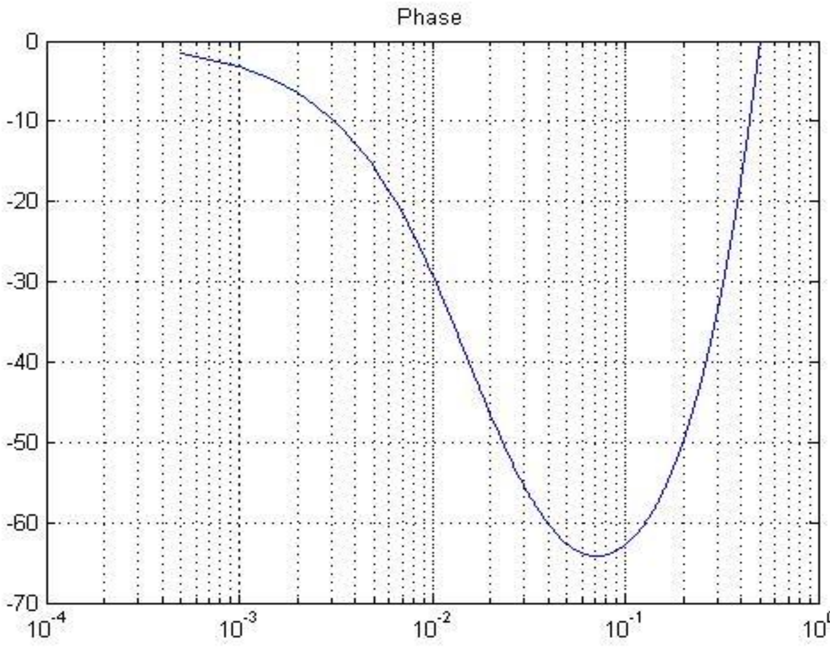
...

$$x_i = b_0 \cdot w^i$$

...

Particular case of AR system:
first order, $-1 < a_1 < 0$ $a_1=-w=0.9$





Note that the normalized frequency arrives at 0.5

Particular case of AR system: first order, a_1=-w complex

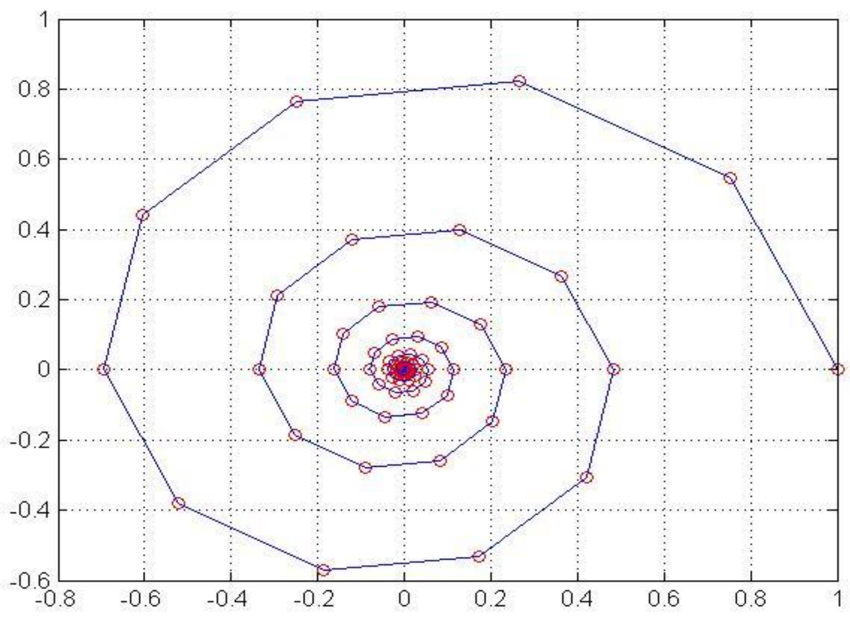
$$w = -a_1 = r \cdot e^{j \cdot \varphi}$$

$$x_k = b_0 \cdot w^k = b_0 \cdot \left(r^k \cdot e^{j \cdot k \cdot \varphi} \right)$$

$$b_0 = 1,\ r = 0.95,\ \varphi = 36°$$

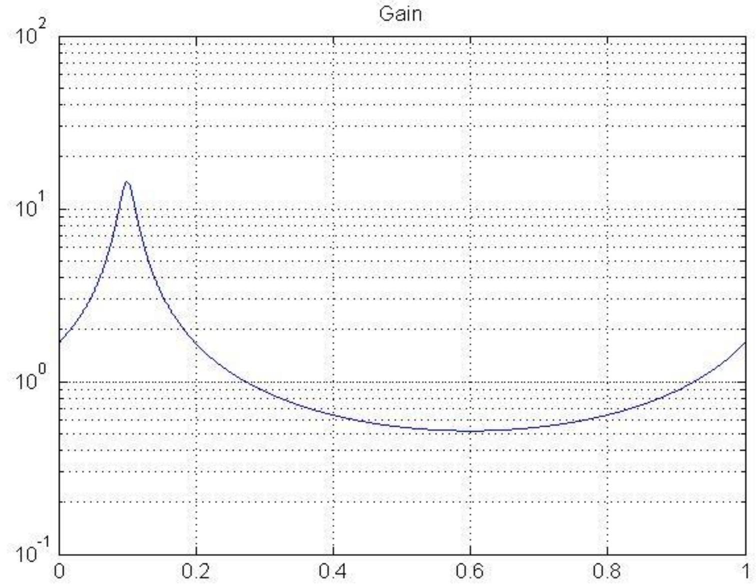
pulse response

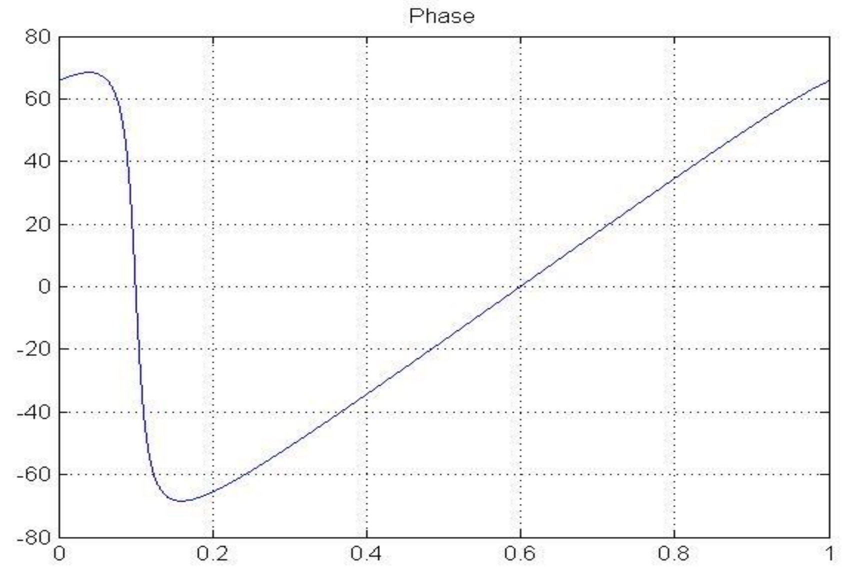
$\mathbf{w^k}$



Particular case of AR system:
first order, a_1=-w complex (cont.)

$$b_0 = 1,\ r = 0.95,\ \varphi = 36°$$





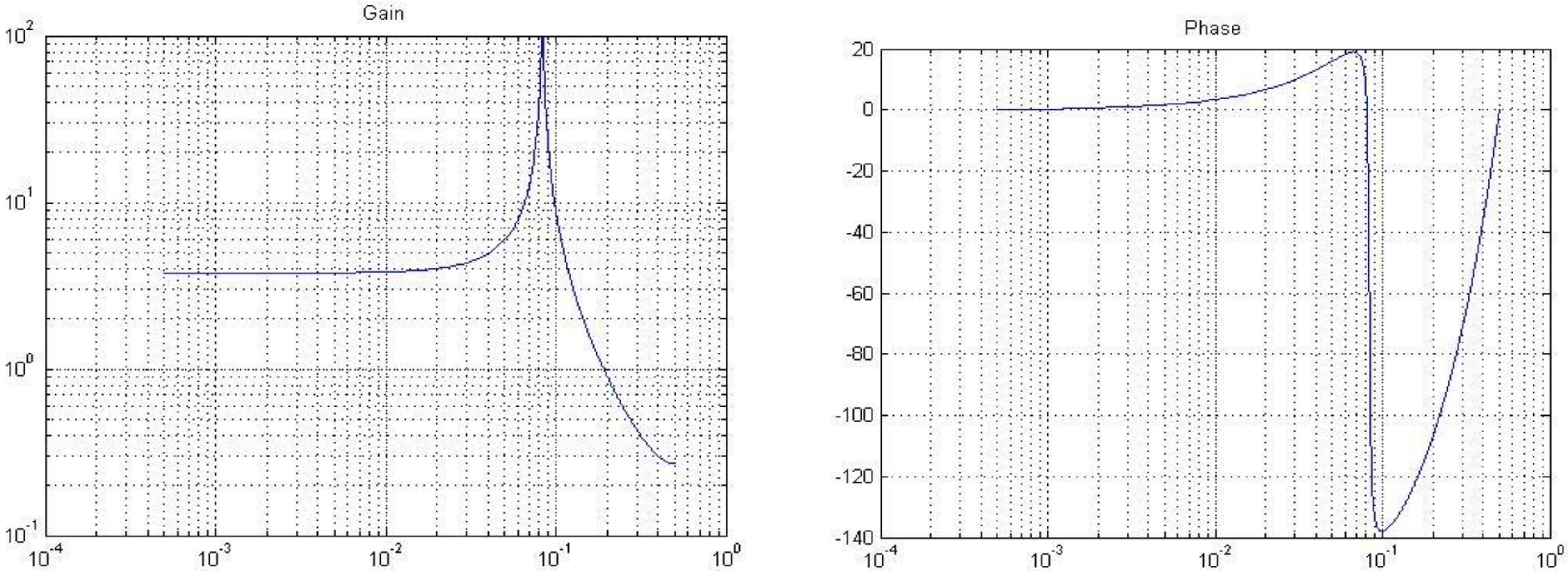
Note that the normalized frequency arrives at 1

Particular case of AR system:
second order, two poles complex conjugate

$$y_i = b_0 \cdot x_i - a_1 \cdot y_{i-1} - a_2 \cdot y_{i-2}$$

$$p_{1,2} = r \cdot e^{\pm j\theta}$$

$$b_0 = 1,\ r = 0.99,\ \theta = \frac{\pi}{6}$$

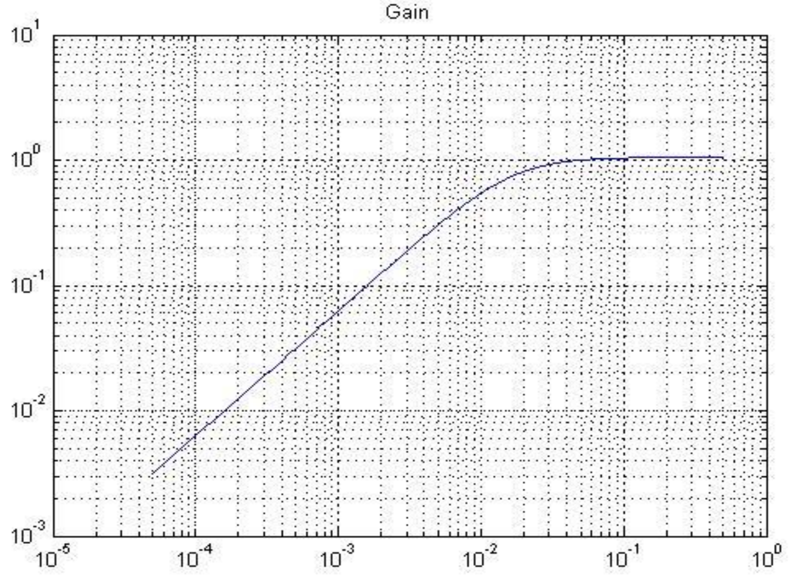


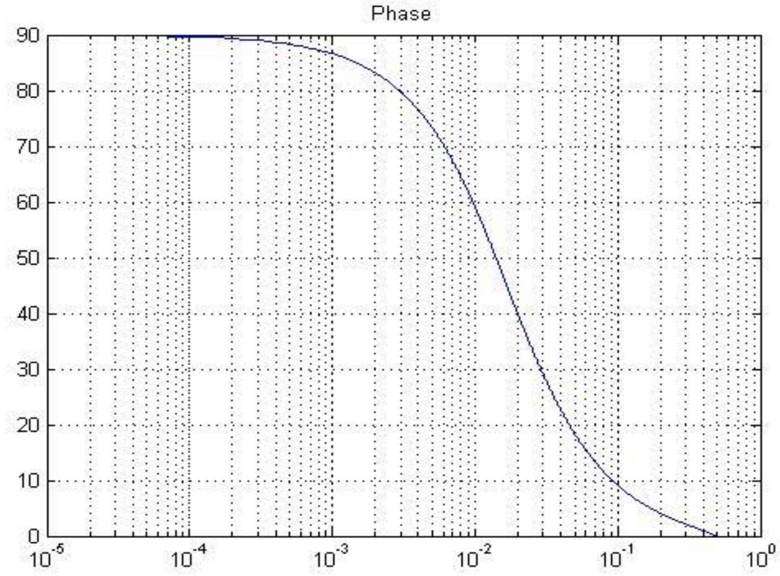
Note that the normalized frequency arrives at 0.5

Simple ARMA filter: the high pass

one pole, one zero

$$F(z) = \frac{1 - z^{-1}}{1 - w \cdot z^{-1}}$$





Operations on systems

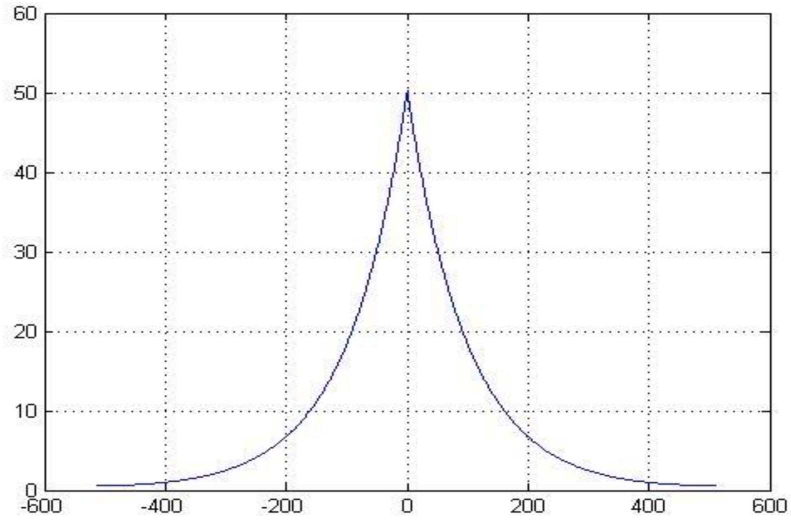
- Two or more systems can be put “in series”, obtaining a new system whose transfer function is the product of the transfer functions
- Two or more systems can be put “in parallel”, obtaining a new system whose transfer function is the sum of the transfer functions
- One system can be “inverted”, taking as the inverse the system that has the inverse of the transfer function of the given system: the series of a system and its inverse is the identity system

Linear systems with white noise at the input

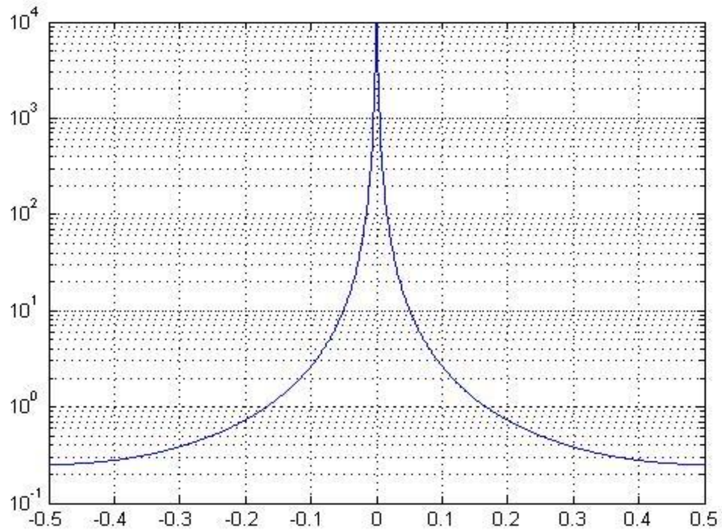
first order AR system, w>0

- In this case we have a so-called “first-order” process, with exponential autocorrelation and lorentzian zero centered spectrum

Autocorrelation



spectrum



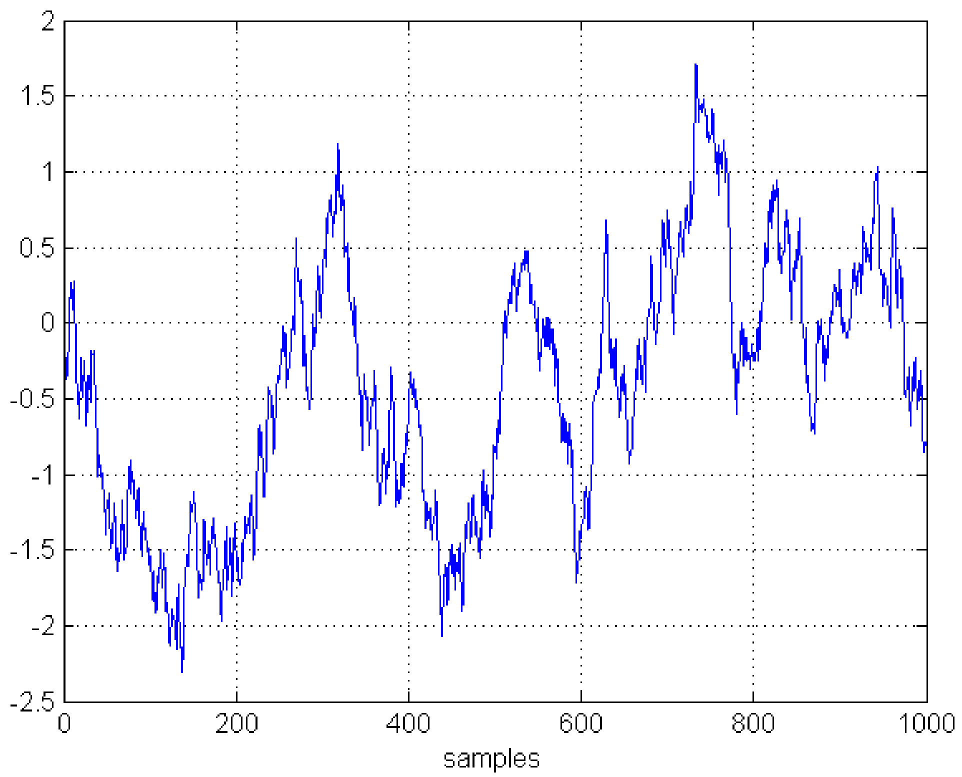
$b_0=1$, $a_0=1$, $a_1=-0.99$

Linear systems with white noise at the input

first order AR system, w>0

First order process

a piece of the created process

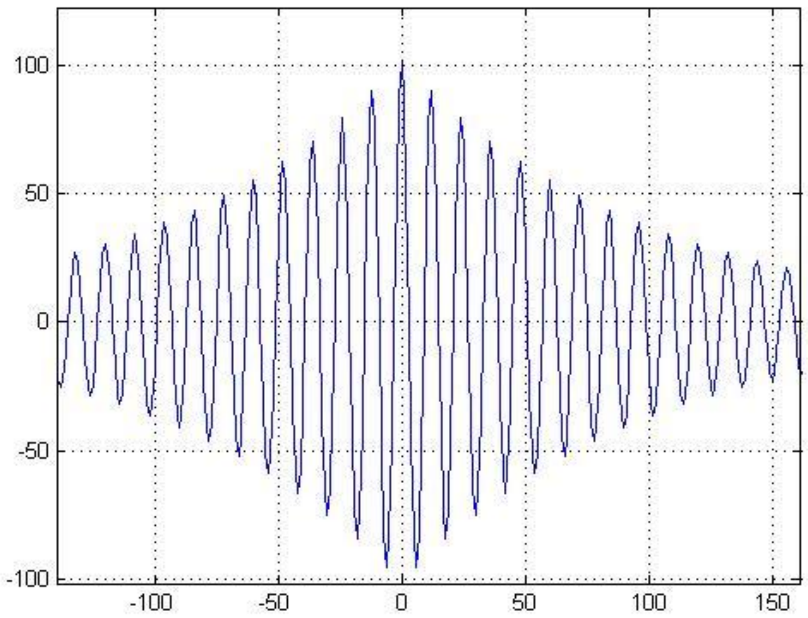


Linear systems with white noise at the input

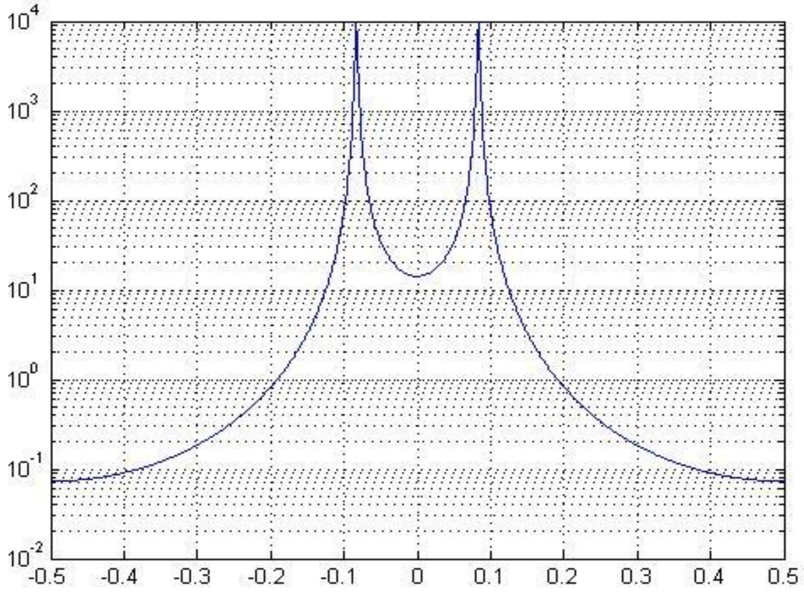
second order AR system, complex conjugate poles

Second order process

Autocorrelation



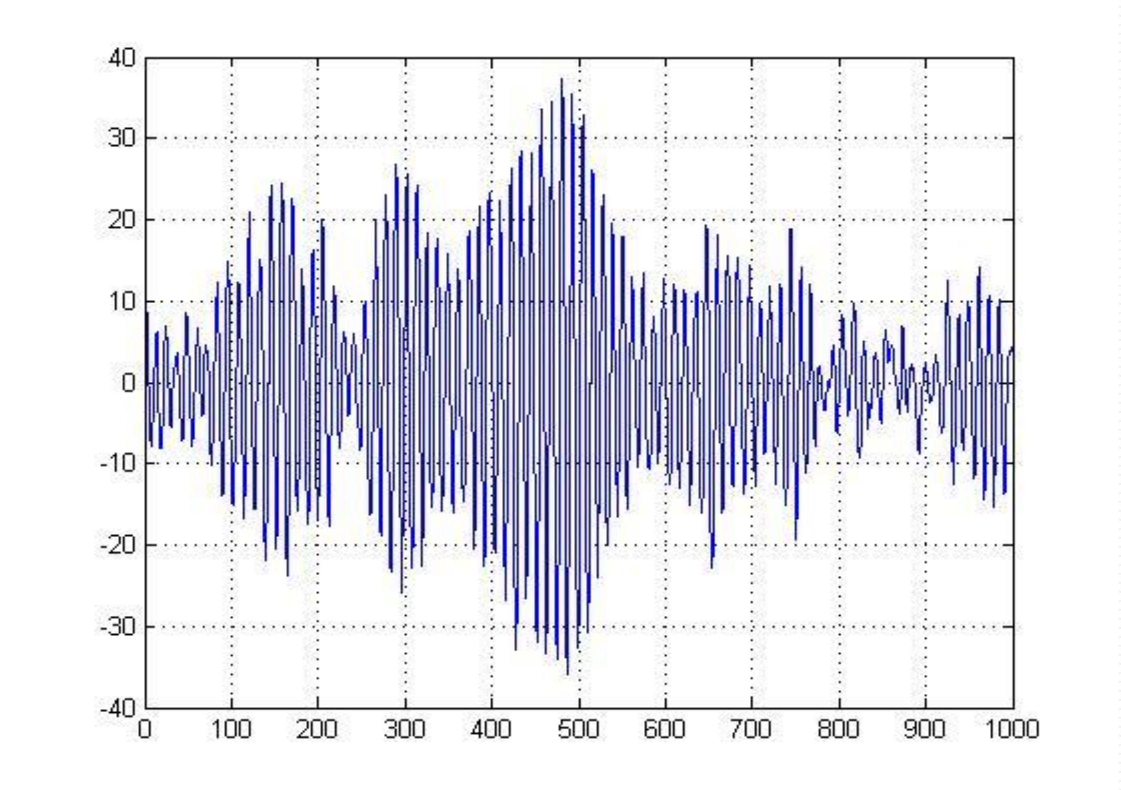
Spectrum



Linear systems with white noise at the input

second order AR system, complex conjugate poles

a piece of the produced data



time samples

The synthesis problem

How to construct a system such that if we put a white gaussian noise at its input, we have at the output a normal stochastic process with given spectrum ?

- AR systems (by Yule and Walker equation, starting from the autocorrelation)
- MA systems (operating in the frequency domain, starting from the power spectrum)

Toward the Yule and Walker equations

Consider the AR equation

$$y_i = b_0 \cdot x_i + w_1 \cdot y_{i-1} + w_2 \cdot y_{i-2} \ldots + w_n \cdot y_{i-n}$$

now multiply the left and right members for y_{i-k} and take the expected value.

We have, for the definition of autocorrelation and because $E[x_i y_{i-k}]=0$ for $k>0$,

$$R_{yy}(k) = w_1 \cdot R_{yy}(k-1) + w_2 \cdot R_{yy}(k-2) \ldots + w_n \cdot R_{yy}(k-n)$$

and putting $r_k = R_{xx}(k)$, we have, for $1 \leq k \leq n$,

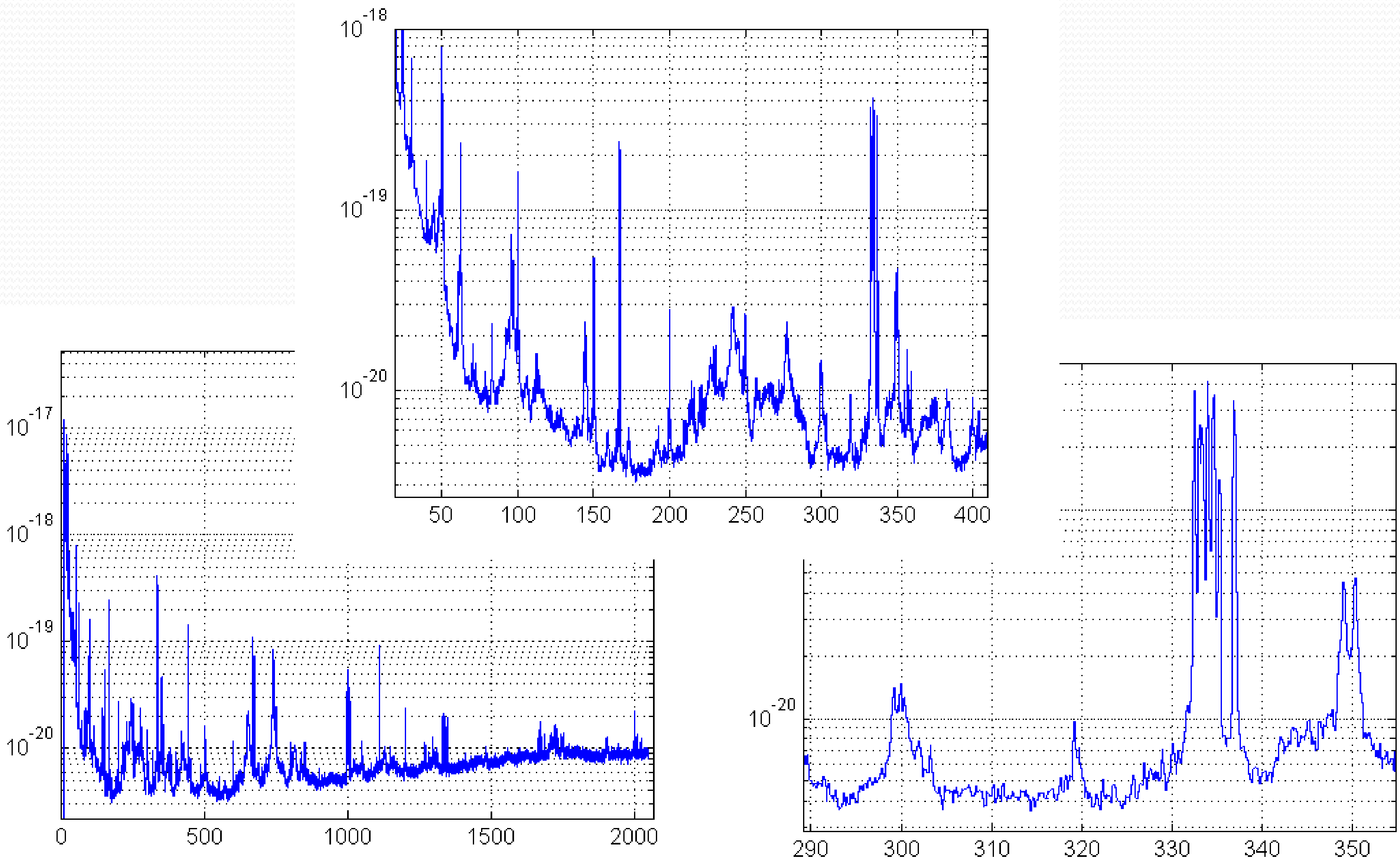
The Yule and Walker equations

$$\begin{pmatrix} r_0 & r_1 & \cdots & r_{n-1} \\ r_1 & r_0 & \cdots & r_{n-2} \\ \vdots & \vdots & & \vdots \\ r_{n-1} & r_{n-2} & \cdots & r_0 \end{pmatrix} \cdot \begin{pmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{pmatrix} = \begin{pmatrix} r_1 \\ r_2 \\ \vdots \\ r_n \end{pmatrix}$$

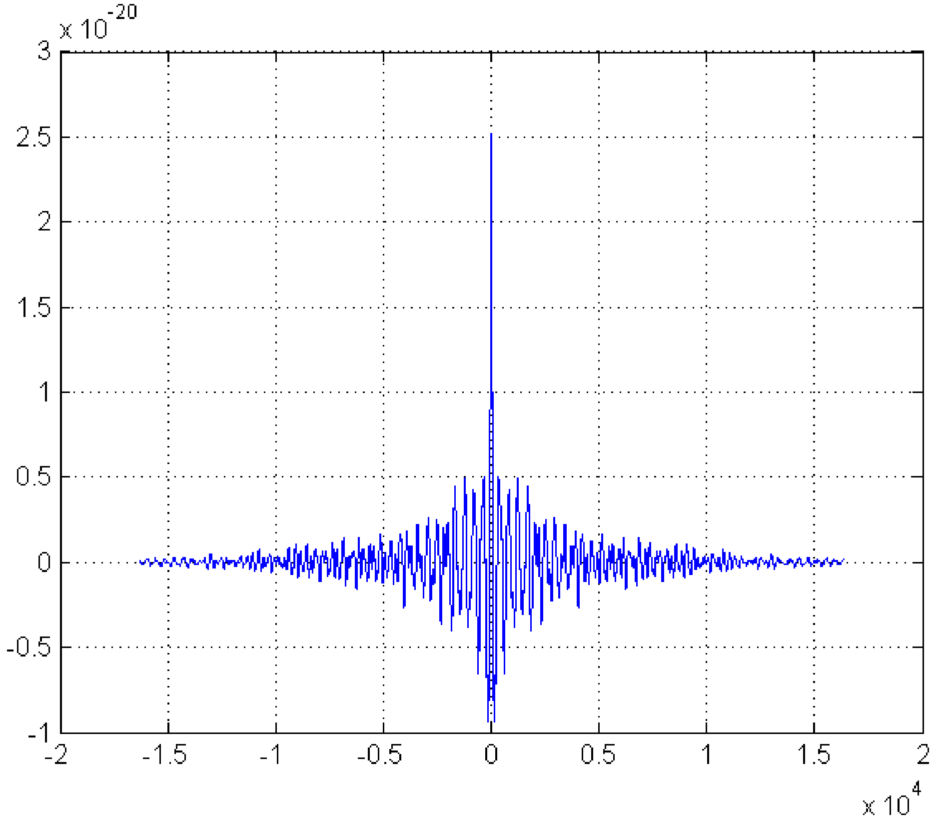
and we can obtain the AR coefficients.
The symmetric matrix is of the Toeplitz type and the system can be solved by the Levinson and Durbin algorithm.

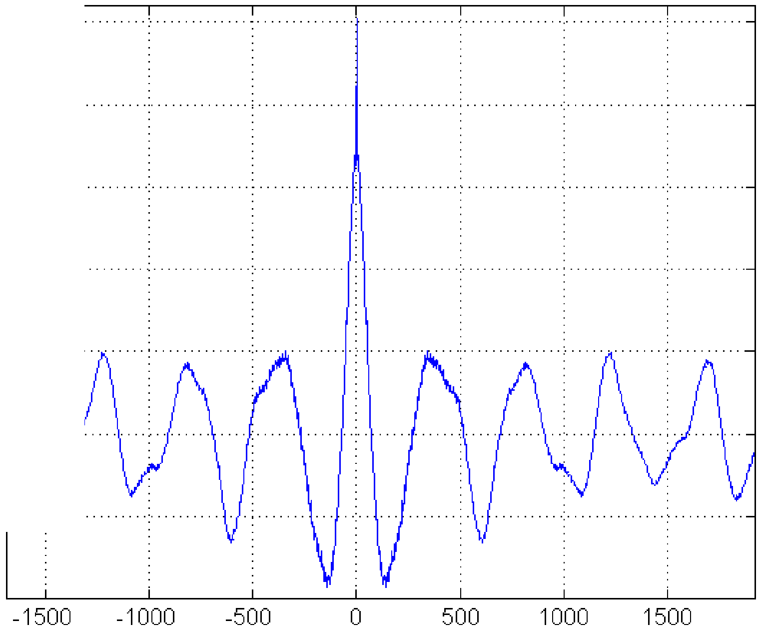
The problems arise if we don't have the autocorrelation, but an estimation of the autocorrelation, with estimation errors, and this is particularly delicate if the number of coefficients is high.

The real data (power spectrum)



The real data (autocorrelation)





The real data

- Note the complexity of the spectrum. It is not just a simple collection of first order processes and a certain number of lorentzian peaks.
- On the other hand the complexity of the spectrum, uniformly scattered on the frequency axis, affects the autocorrelation mostly on the low amplitude samples of the autocorrelation.
- If there is also a little estimation error (unavoidable anyway, also for the non-stationarities), the computation of the AR coefficients can have big errors and the computed system can be also unstable.

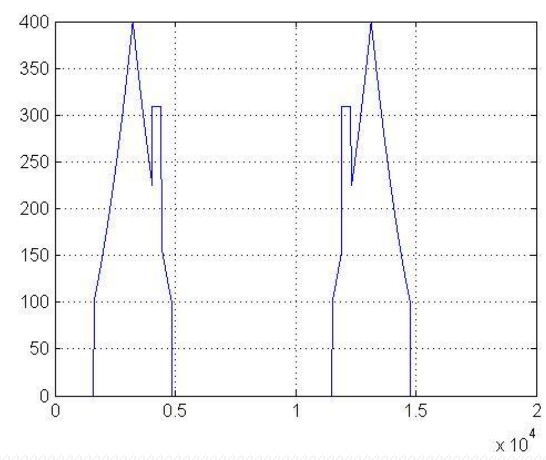
MA model and frequency domain operations

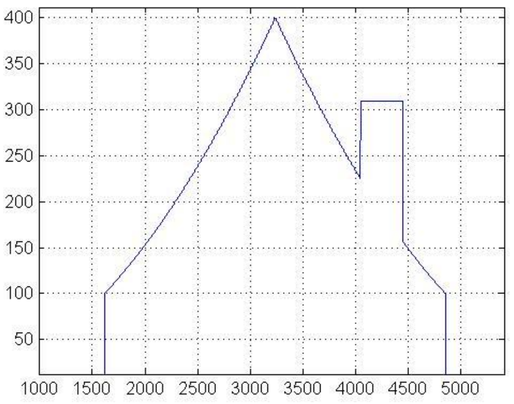
- An alternative way to operate is to consider a (much more) complex MA model, i.e. a model with a very big number N of coefficients (also more then 10^6)
- This would need high computer power , but the use of the frequency domain operation (with the use of the FFT) reduce drastically this needs (for each sample we have a computing time proportional to log(N) instead of N).
- Operating in the frequency domain means that a convolution is substituted by a scalar product.
- Operating in the frequency domain also the computation of the MA coefficients is easy, derived in a simple way from the noise power spectrum.

MA model and frequency domain operations

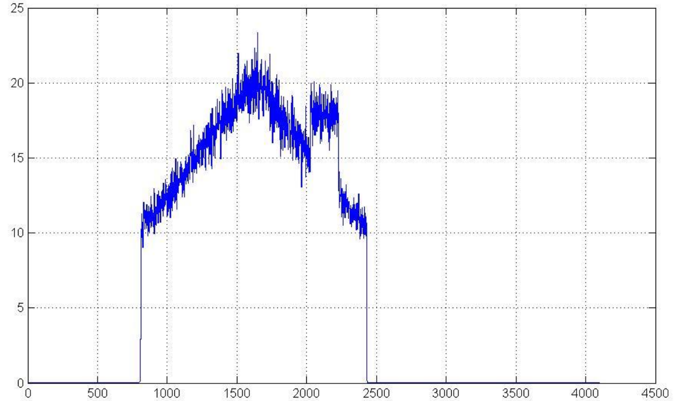
- With the frequency domain operation, it is easy to construct non-causal, non de-phasing systems, that are often useful. This is accomplished using systems that are symmetric in t in the pulse response.
- One of the problems of operating in the frequency domain is that we have to operate on pieces (segments) of data and not in a continuous way
- The segments of data are seen as “rings”, so the filter in the time domain should have (typically) at least half of the values null (typically the central part) and (in this case) only the central part of the output segment is correctly processed.
- For this reason we need to take segments interlaced by the half, and for each segment only the half central part

Example of simulation with frequency domain MA system

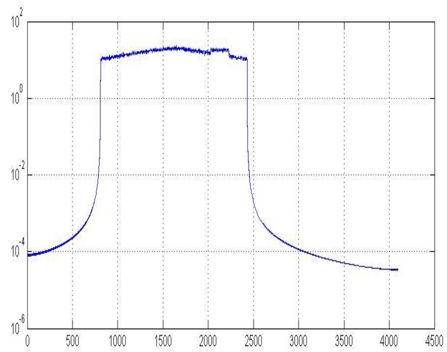




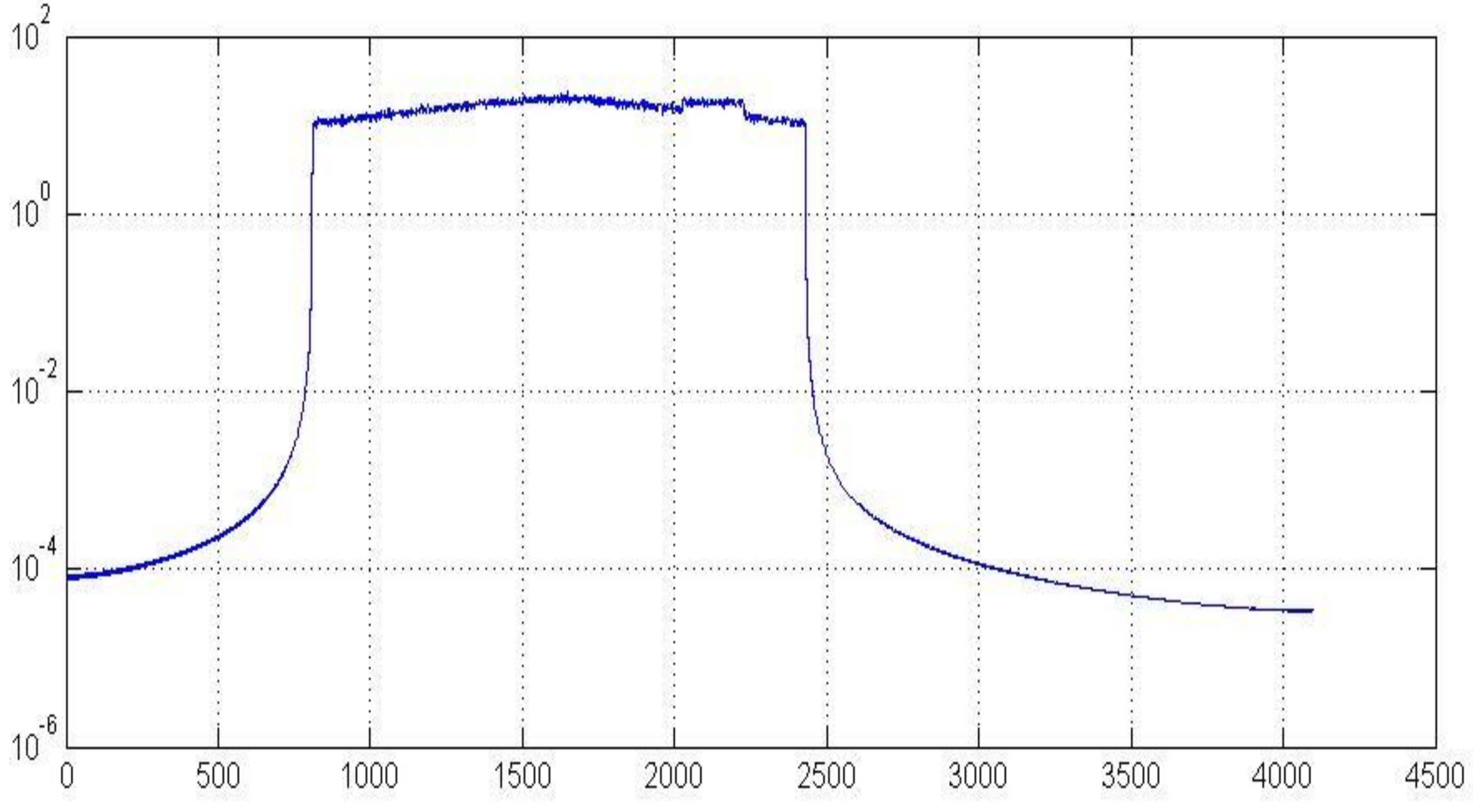
input spectrum



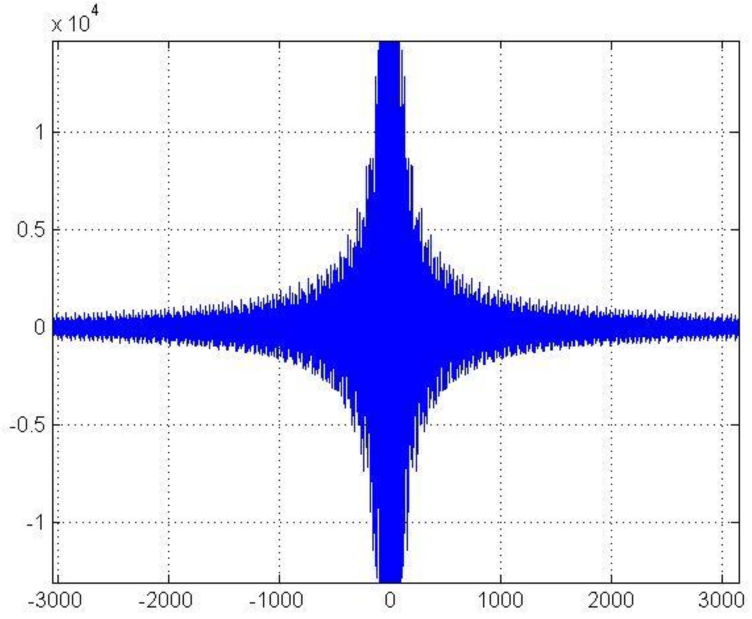
h-density of the produced data

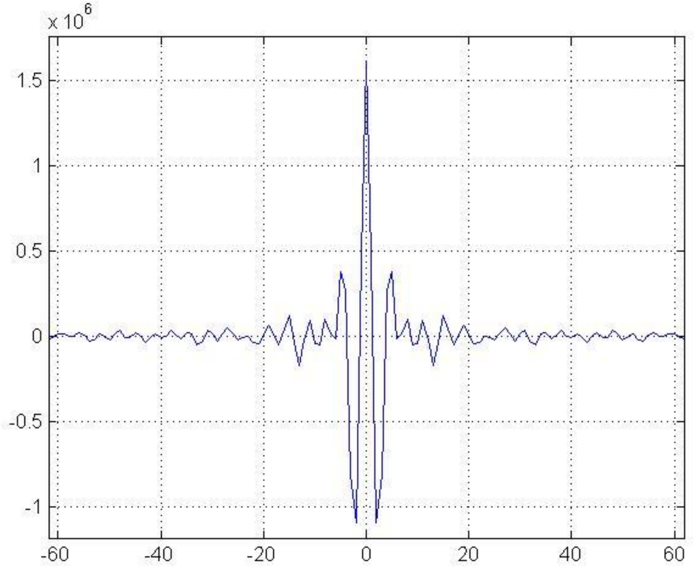


Example of simulation with frequency domain MA system
(log of h-density)



Example of simulation with frequency domain MA system (autocorrelation)





Discrete time system operations

- The discrete systems we described can be used not only to model (or simulate) a gaussian stochastic process, but also to do a variety of operations on them, often called “filtering” .
- This can be done indifferently with MA or AR (or FIR or IIR) systems. And, obviously, in the time or frequency domain.

Filtering: whitening filter

- A filter that is sometimes useful is the whitening filter, that it is used to “equalize” in the frequency domain the gaussian noise.
- If the power spectrum is S(ω), the “generator” system is

$$F(j\omega) = k \cdot \sqrt{S(\omega)}$$

- The whitening filter, in the frequency domain, is

$$B(j\omega) = \frac{1}{F(j\omega)}$$

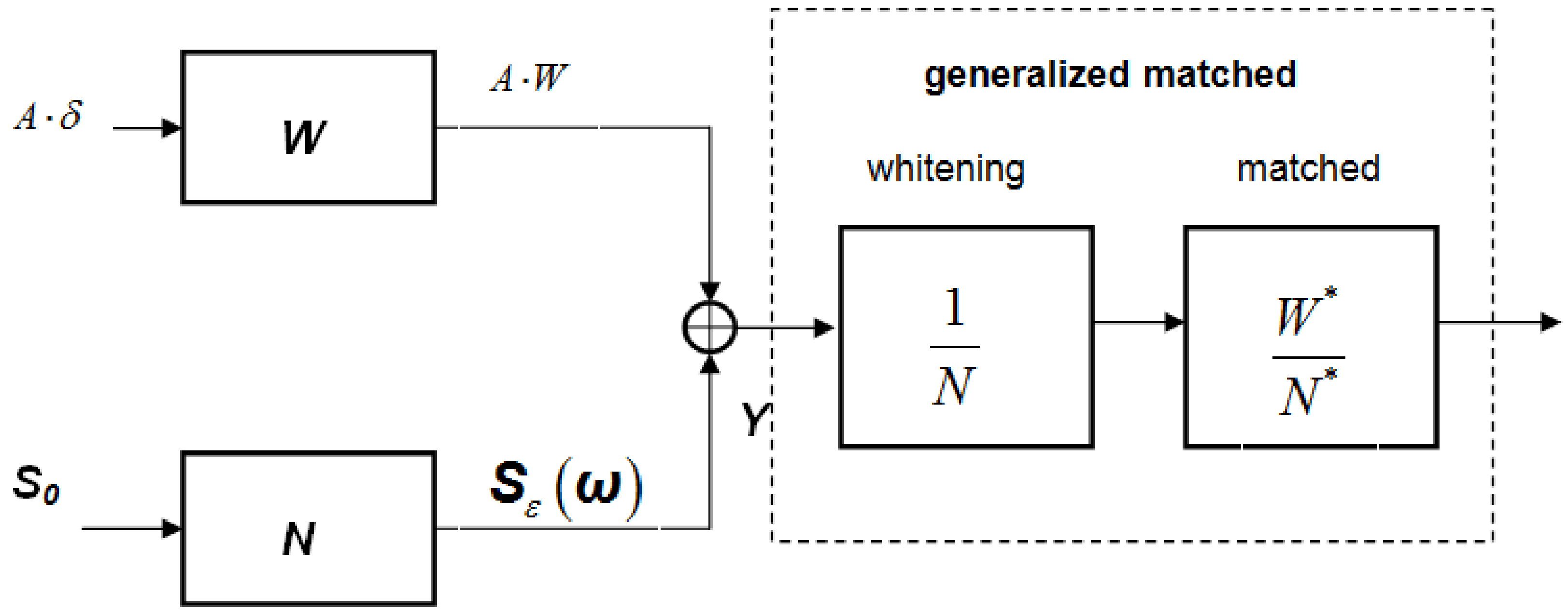
Filtering: Wiener filter

- In some cases we need the Wiener filter to estimate the innovation process. This filter is

$$W(j\omega) = \frac{a}{S(\omega)} = \frac{b}{\left|F(j\omega)\right|^2}$$

for this reason, in the Virgese dialect it is called "double whitening"

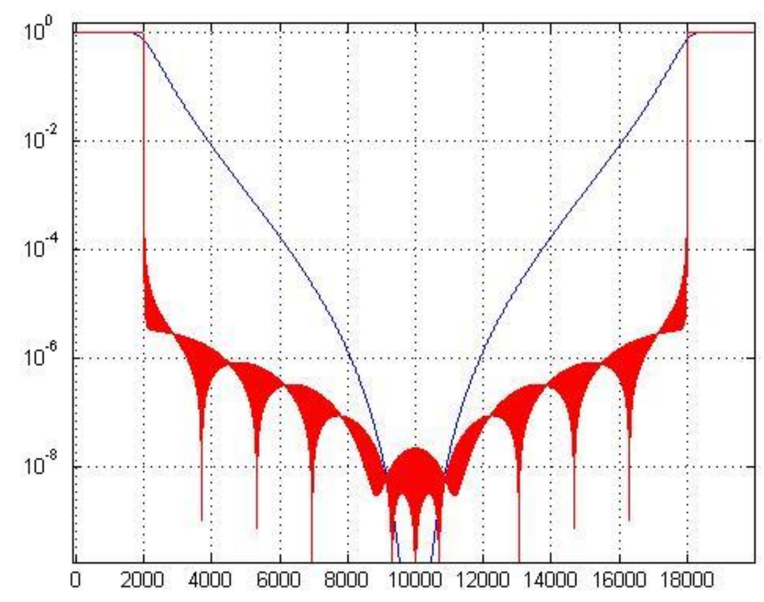
Filtering: matched filter



Filtering: anti-aliasing filter

- In order to reduce the sampling time, e.g. from 20 kHz to 4 kHz, we must use an anti-aliasing filter.
- This can be done by a Butterworth filter (an ARMA (6,6) low-pass filter, **red**) or by a frequency domain filter (16384 points, **blue**). Here are the results:

modulus of the transfer function



Note: the butterworth filter heavily dephases, the frequency domain filter is 0-phase

http://grwavsf.roma1.infn.it/PSS/reports/FrDom-AA.pdf

Non-linear operations: resampling

- In the frequency domain we can also perform non-linear operations.
- An example is the method used in Virgo in order to pass from 20 kHz of sampling time to 16384 and 4096.
- See the reports
  - http://grwavsf.roma1.infn.it/PSS/reports/rep_res_1.pdf
  - http://grwavsf.roma1.infn.it/PSS/reports/rep_res_1.pdf

Other uses of the AR systems for the gravitational wave detection

- An important feature of the gravitational wave antenna noise is that it is not stationary and the non-stationarities are not easily modeled.
- In order to construct better estimators of various characteristics of the noise, we can construct AR estimators, defined with a given "τ" (or "memory", or "length of stationarity").

Example: AR(1) estimation of the power spectrum

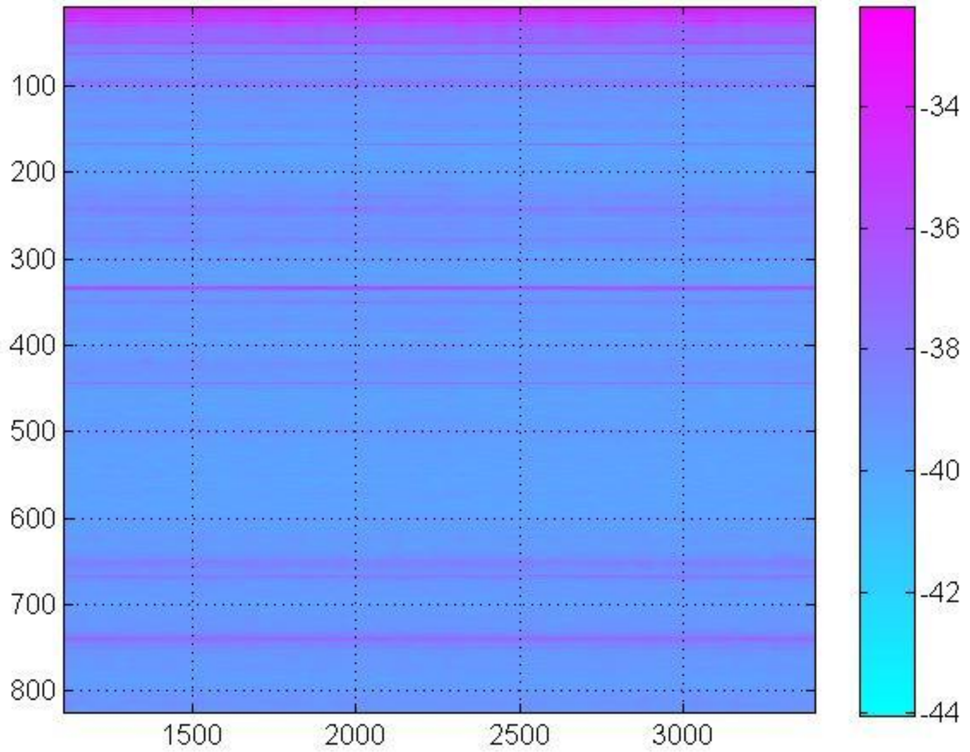
- If we compute a series of periodograms of subsequent segments of data $\mathbf{P}_i(\omega)$, if the process is stationary within a time scale of τ, the estimation of the (varying) power spectrum can be done in the following way:

$$\begin{cases} \hat{S}_i'(\omega) = P_i(\omega) + w \cdot \hat{S}_{i-1}'(\omega) \\ k_i = 1 + w \cdot k_{i-1} \end{cases} \qquad w = e^{-\frac{\Delta t}{\tau}}$$

$$\hat{S}_i(\omega) = \frac{\hat{S}_i'(\omega)}{k_i}$$

- A similar thing can be done with many type of estimation (mean, variance, frequency distribution,…).

Example: AR(1) estimation of the power spectrum



Example: power spectrum innovation

- Sometimes the estimation of the spectral innovation is very useful. It can be computed from the previous estimation of the power spectrum in the following way:

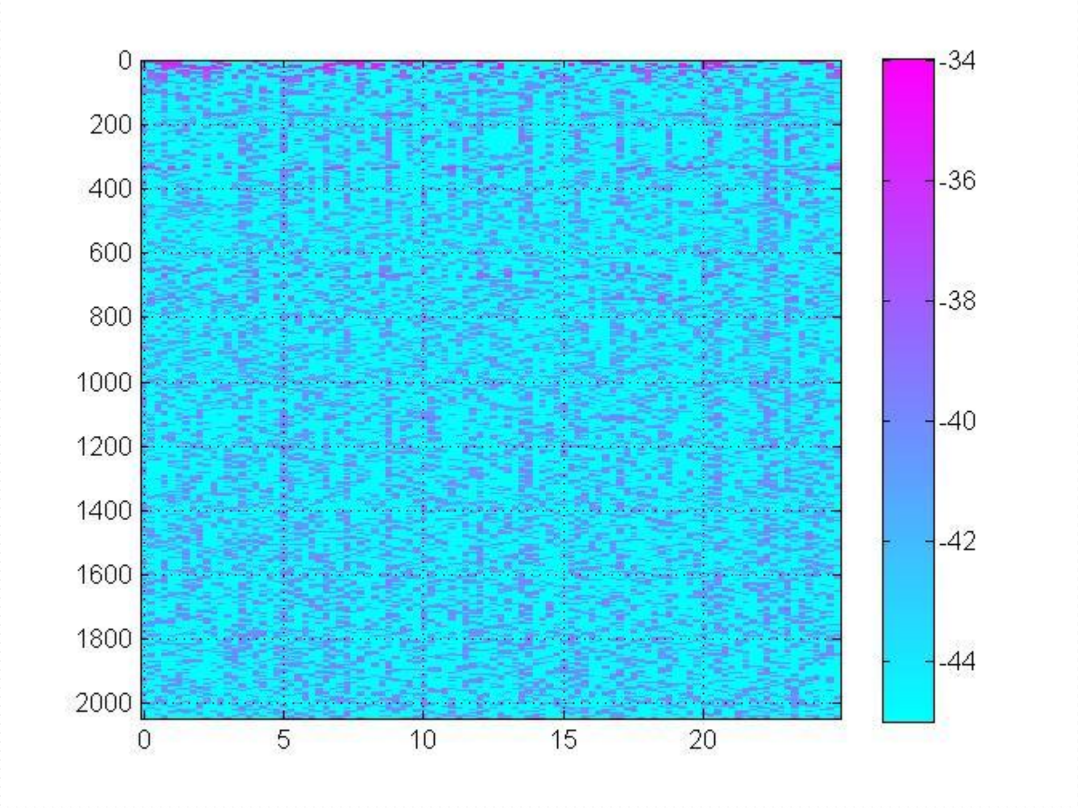
$$\hat{I}_i(\omega) = \hat{S}_i(\omega) - \hat{S}_{i-1}(\omega)$$

or

$$\hat{I}_i(\omega) = \frac{\hat{S}_i(\omega)}{\hat{S}_{i-1}(\omega)}$$

It shows what is the spectrum variation.

Power spectrum innovation



Example: adaptive threshold event finder

- If we do an AR estimation of the mean and of the standard deviation, we can compute an adaptive threshold in order to search for events with fixed CR (critical ratio).