BMEN E4420: Biomedical Signal Processing and Signal Modeling

Lecture 6

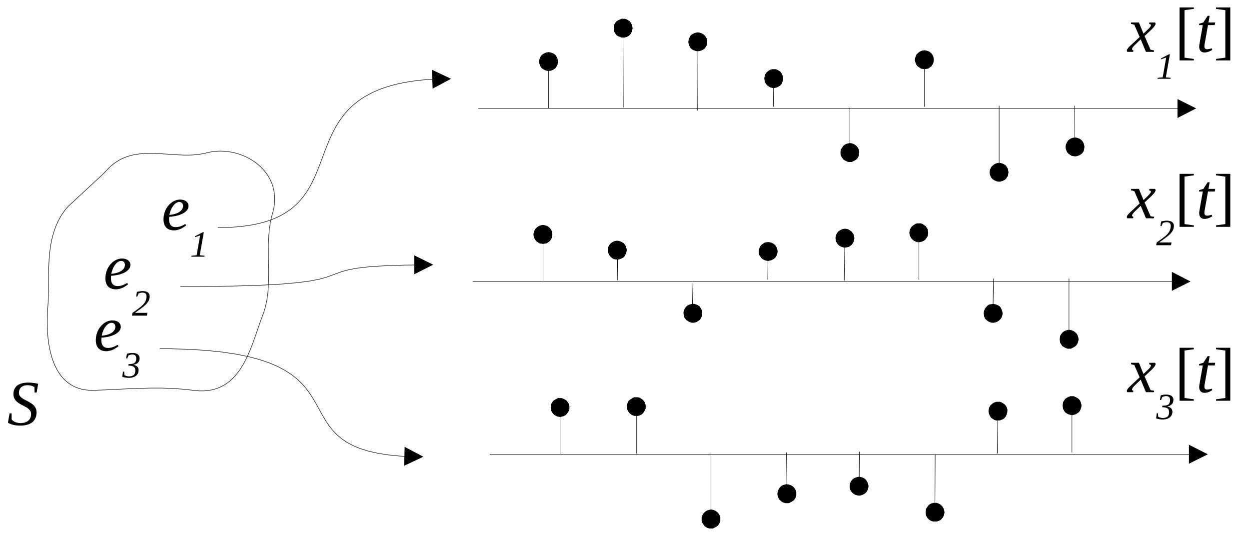
Stochastic Processes

Paul Sajda
psajda@columbia.edu

Date	Lecture	Readings	Assignments
1/22/2013	**Introduction and example physiological signal: EEG**	WvD Ch 1	HW 1
1/29/2013	**Linear Systems**	Bruce Ch: 3 Impulse Response Bruce Ch: 9 (sections 9.8-9.13) AR and ARMA WvD Ch: 5 Real & Complex Fourier Series	HW 2
2/5/2013	**Discrete Fourier Transform**	Bruce Ch: 4 Frequency Response Bruce Ch: 8.3-8.8 Z-transform Bruce Ch: 8.17 FIR filter, 8.19 Biomedical apps WvD Ch9: Laplace and Fourier Transforms	HW 3
2/12/2013	**More DFT**		HW4
2/19/2013	**Random Variables**	Bruce Ch 2: Memory and Correlation	HW 5
2/26/2013	**Stochastic Processes**	Bruce Ch. 9 Modeling Stochastic Signals as Filtered White Noise Especially 9.7-9.10	HW6
3/5/2013	**Review for Midterm**		
3/12/2013	**Midterm**		
3/19/2013	**Spring Break**		
3/26/2013	**Probabilistic Estimation & Harmonic Processes**		HW 7
4/2/2013	**Discrimination and Classification**	P. Sajda (2006) Machine Learning for Detection and Diagnosis of Disease, *Ann. Rev. Biomed. Eng.*	HW 8
4/9/2013	**Dynamic State Space Models: Hidden Markov Models and the Kalman Filter**		
4/16/2013	**Advanced Topics Introduced**		
4/23/2013	**Review of the Semester**		
4/30/2013	**Final Presentations**		

Stochastic Processes

A stochastic process is a random quantity that changes over time. More specifically it is a **set of random variables sorted in time** $X[t]$. Sampling a random process repeatedly gives an ensemble:



Examples: Spike train, position of a particle moving randomly, number of bacteria in a culture, thermal noise in an amplifier, etc.

Stochastic Process - Poisson Process

Poisson process arises when we count the number of occurrences $N(t)$ of **independent** random events where the likelihood of occurrence of an event is **proportional to the time** interval Δt.

$$P\left(\left(N\left(t+\Delta t\right)-N\left(t\right)\right)=1\right)=\nu\,\Delta t$$

One can prove by induction that the count follows a Poisson distribution with $\lambda=\nu t$

$$P\left(N\left(t\right)=k\right)=\frac{\left(\nu t\right)^k e^{-\nu t}}{k!}$$

Interesting is the first induction step. For k=0 using definition and limit $\Delta t \rightarrow 0$ it is easy to see that

$$P\left(N\left(t\right)=0\right)=e^{-\nu t}$$

Which means that the ISI is exponentially distributed.

Stochastic Process - Shot Noise

Shot noise: a Poisson process with events at times t_i convolved with an impulse response (IR).

$$X(t) = \sum_{i=1}^{N} h(t)\,\delta(t - t_i) = \sum_{i=1}^{N} h(t - t_i)$$

The expected value of $X(t)$ is

$$E[X(t)] = \nu \int_{-\infty}^{\infty} f(u)\,du$$

Event or spike rate ν can be estimated by convolving a spike train with a suff ciently long IR. This in effect is what neurons do, where the IR is given by the post-synaptic potential (PSP).

Sometimes $(t_i - t_{i-1})^{-1}$ is considered "instantaneous" spike rate ν (t).

Stochastic Process - Gaussian Process

A stochastic process $x(t)$ is said to be a **Gaussian process** if $\boldsymbol{x}=[x(t_1),x(t_2),\ldots,x(t_n)]^T$ are jointly Gaussian distributed

$$p(\boldsymbol{x})=\frac{1}{\sqrt{(2\pi)^n|\boldsymbol{\Sigma}|}}\exp\left[-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\right]$$

In general the covariance matrix $\boldsymbol{\Sigma}$ and mean $\boldsymbol{\mu}$ depend on the time instances t_1, t_2, ... , t_n.

Classic example is Brownian motion

Stochastic Process - Wide Sense Stationary

If $p(x(t_1),x(t_2),...,x(t_n))$ does not depend on times t_1, t_2, ... , t_n the process is called **stationary.**

Any stochastic process for which the *mean and covariance* are *independent of time* is called **Wide Sense Stationary (WSS)**.

A WSS process is called **ergodic** if the sample average converges to the ensemble average. With samples $x[k] = x(t_k)$ this reads:

$$E[f(x(t))] = \lim_{N \to \infty} \frac{1}{N} \sum_{k=1}^{N} f(x[k])$$

Stochastic Process -Auto-correlation

If we have equally distant samples, $x[k] = x(kT)$, the second moments correspond to the **auto-correlation**

$$r_x[k,l] = E[x[k]x^*[l]]$$

For WSS stochastic processes:

The mean is independent of time $m_x = E[x[k]]$

The Auto-correlation $r_x[k] = E[x[n+k]x^*[n]]$

depends only on time differences $r_x[k,l] = r_x[k-l]$

it is symmetric, $r_x[k] = r_x^*[-k]$

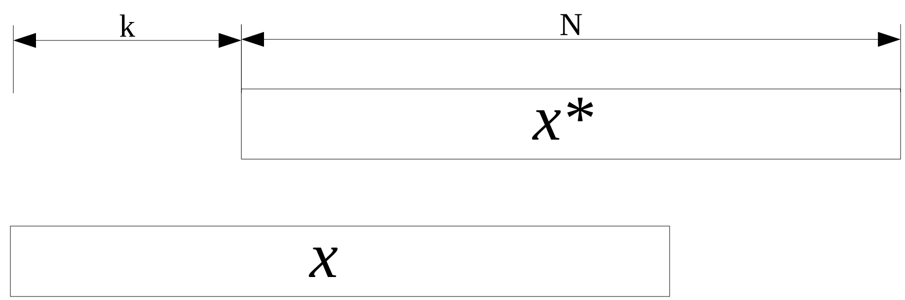
the zero lag is the power $r_x[0] = E[|x[k]|^2]$

upper bounded by the power $r_x[0] \geq |r_x[k]|$

Stochastic Process - Auto-correlation

Assuming ergodic WSS a direct estimate of the auto-correlation is the (unbiased) **sample auto-correlation**

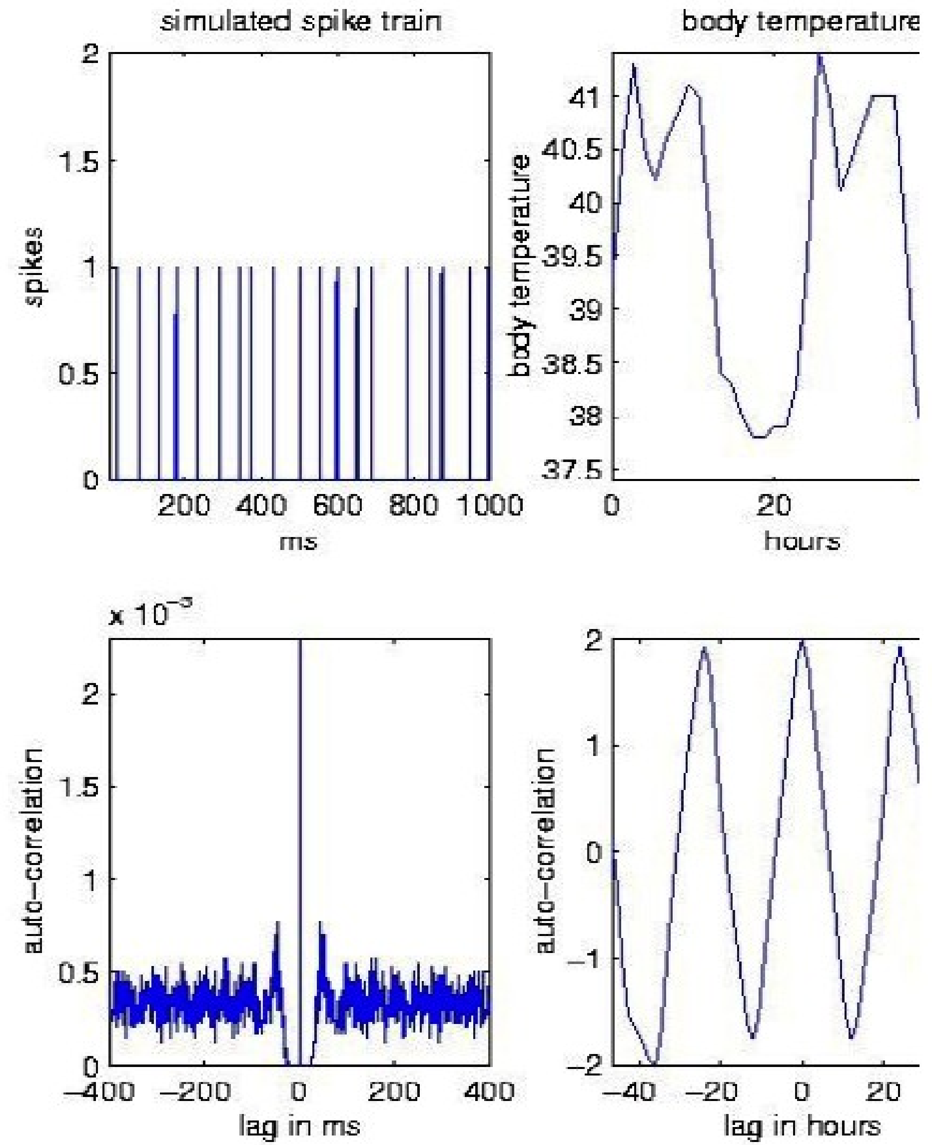
$$\hat{r}_x[k] = \frac{1}{N-k} \sum_{l=1}^{N-k} x[l+k]\, x^*[l]$$



Note that correlation is convolution with opposite sign. It can be computed fast with the FFT. Use `xcorr()`

Stochastic Process - Auto-correlation

Examples



Stochastic Process -Auto-correlation

For WSS the **auto-correlation matrix** becomes Hermitian Toeplitz

$$\boldsymbol{R}_{\boldsymbol{xx}} = \begin{bmatrix} r_x(0) & r_x^*(1) & r_x^*(2) & \cdots & r_x^*(p) \\ r_x(1) & r_x(0) & r_x^*(1) & \cdots & r_x^*(p-1) \\ r_x(2) & r_x(1) & r_x(0) & \cdots & r_x^*(p-2) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_x(p) & r_x(p+1) & r_x(p+2) & \cdots & r_x(0) \end{bmatrix}$$

Important WSS process is **(Gaussian) uncorrelated noise**

$$\boldsymbol{R}_{\boldsymbol{xx}} = \sigma^2 \boldsymbol{I} \qquad \boldsymbol{m}_{\boldsymbol{x}} = \boldsymbol{0}$$

```
>> randn()
```

Stochastic Process - Power spectrum

Oscillations of the autocorrelation are best analyzed in the frequency domain, which leads to the **Power Spectrum**

$$P_x\left(e^{j\omega}\right)=\sum_{k=-\infty}^{\infty} r_x[k]e^{-jk\omega}$$

One can show that $P_x(e^{j\omega})$ is *real*, *even* and *positive*.

The auto-correlation can be recovered with the inverse Fourier transform

$$r_x[k]=\frac{1}{2\pi}\int_{-\pi}^{\pi} d\omega\, P_x\left(e^{j\omega}\right)e^{jk\omega}$$

Stochastic Process - Power spectrum

In particular, the total power is given by

$$r_x[0] = E[|x[k]|^2] = \frac{1}{2\pi}\int_{-\pi}^{\pi} d\omega\, P_x(e^{j\omega})$$

the power spectrum is sometimes called **spectral density** because it is *positive* and the signal power can always be *normalized* to $r(0) = (2\pi)^{-1}$.

Example: Uncorrelated noise has a constant power spectrum

$$r[k] = E[x[n+k]x^*[n]] = \sigma^2\delta(k)$$

$$P_x(e^{j\omega}) = \sum_{k=-\infty}^{\infty} \sigma^2\delta(k)e^{-jk\omega} = \sigma^2$$

Hence it is also called **white noise**.

Stochastic Process - Power spectrum

The power spectrum captures the spectral content of the random sequence. It can be estimated directly from the Fourier transform of the data:

$$\hat{P}_x\left(e^{j\omega}\right) = \frac{1}{N}\left|X\left(e^{j\omega}\right)\right|^2$$

$$X\left(e^{j\omega}\right) = \sum_{k=0}^{N-1} x[k]e^{-j\omega k}$$

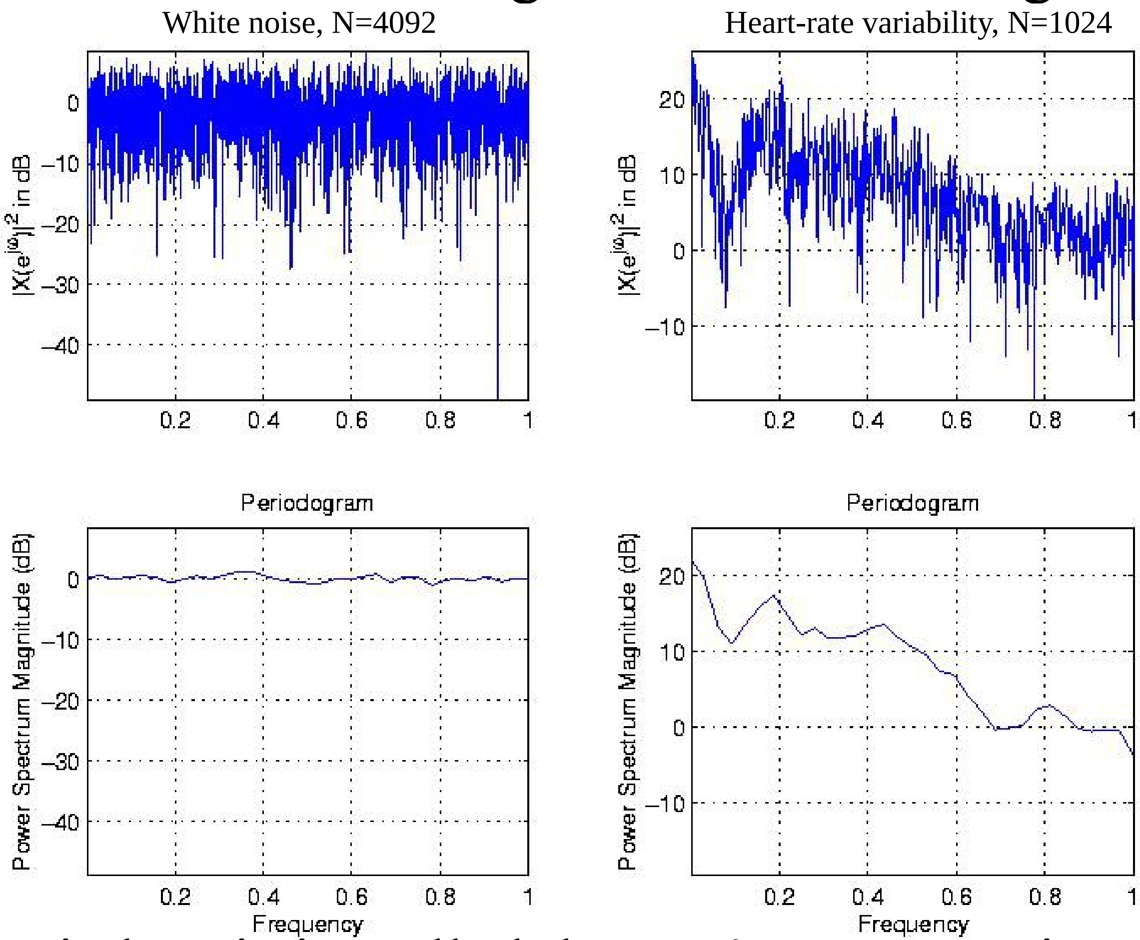
This estimate improves in the mean with increasing *N*.

$$P_x\left(e^{j\omega}\right) = \lim_{N\to\infty} E\left[\hat{P}_x\left(e^{j\omega}\right)\right]$$

Remember that when estimating auto-correlation and power spectrum we implicitly assume WSS processes!

Stochastic Process - Power spectrum

Unfortunately, the direct estimate is inconsistent, i.e. its variance does not converge to 0 for increasing N.



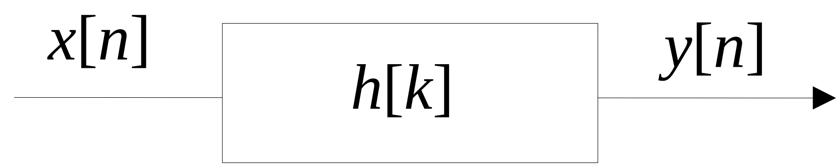

```
>> psd(x);
>> pwelch(x);
```

A classic heuristic, called the **periodogram**, is to smooth neighboring frequencies: Compute Fourier transform for window of size N/K and average over K windows.

Stochastic Process - Filtering

The effect of filtering on the statistics of a stochastic process:

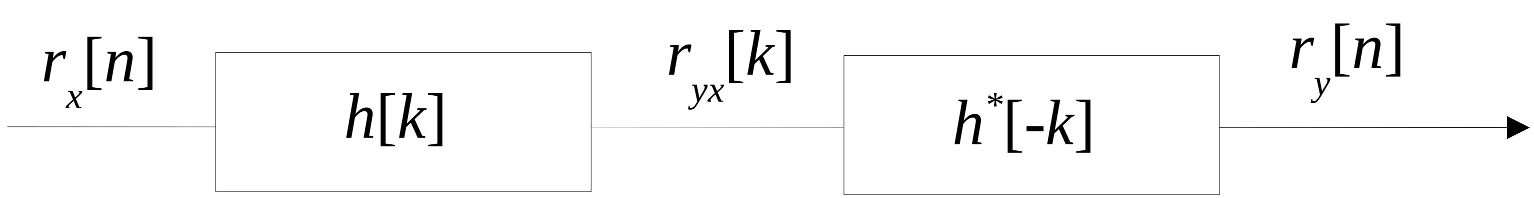
$$y[n] = \sum_{k=-\infty}^{\infty} h[k]x[n-k]$$



Filtering corresponds to convolution of the autocorrelation

$$r_y[k] = \sum_{l=-\infty}^{\infty} \sum_{m=-\infty}^{\infty} h[l] r_x[m-l+k] h^*(m)$$

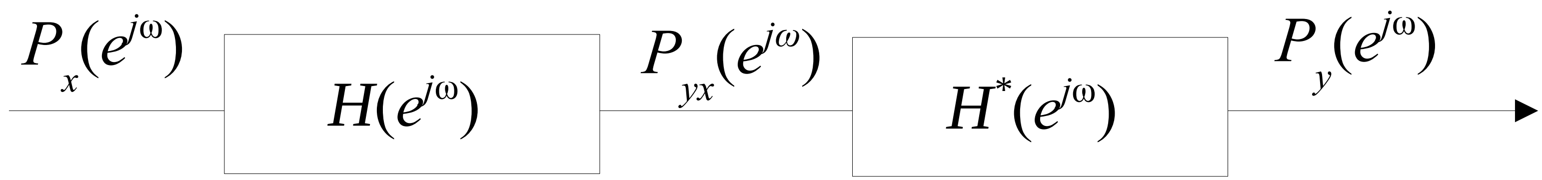
$$= r_x[k] * h[k] * h^*[-k]$$

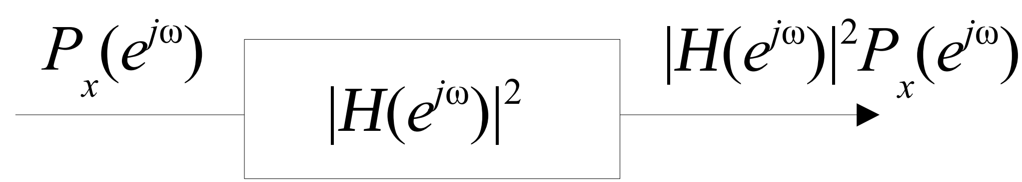


Stochastic Process - Filtering

With the convolution theorem of the Fourier transform we find for the power spectrum

$$P_y\left(e^{j\omega}\right) = P_x\left(e^{j\omega}\right) H\left(e^{j\omega}\right) H^*\left(e^{-j\omega}\right) = P_x\left(e^{j\omega}\right) \left|H\left(e^{j\omega}\right)\right|^2$$

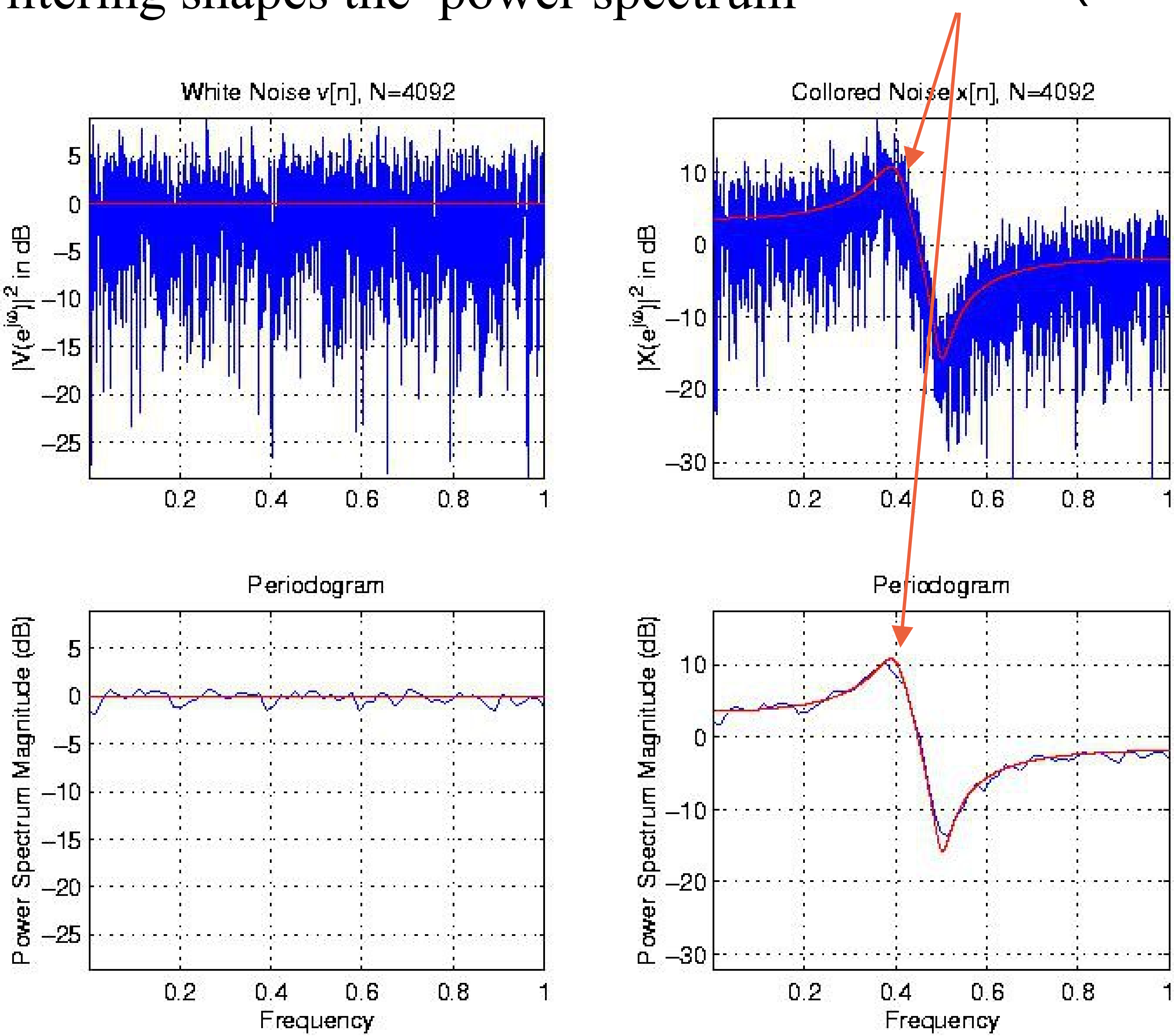




Stochastic Process - Filtering

Filtering shapes the power spectrum

`abs(fft(h,N)).^2`



Stochastic Process - Regular process

Colored noise process $x[n]$ is filtered white noise $v[n]$

$$v[n] \longrightarrow \boxed{h[n]} \longrightarrow x[n]$$

$$P_v(e^{j\omega}) = 1 \longrightarrow \boxed{|H(e^{j\omega})|^2} \longrightarrow P_x(e^{j\omega}) = \sigma_v^2 \, |H(e^{j\omega})|^2$$

If $h[n]$ is *causal* and *invertible* $x[n]$ is called a **regular process** and $v[n]$ the corresponding **innovation process** with power σ_v^2.

Its power spectrum is continuous and given by

$$P_x\left(e^{j\omega}\right) = \sigma_v^2 \left|H\left(e^{j\omega}\right)\right|^2$$

The inverse filter $1/H(z)$ is called the **whitening filter** of $x[n]$.

$$x[n] \longrightarrow \boxed{1/H(z)} \longrightarrow v[n]$$

Stochastic Process - Predictive process

A process $x[n]$ is called **predictive** or **singular** if it can be predicted exactly from a linear combination of its past:

$$x[n] = \sum_{k=1}^{\infty} a[k]\, x[n-k]$$

Examples are sinusoids or sums thereof:

$$x[n] = \sum_{k=1}^{\infty} A_k e^{i n \omega_k + \phi_k}$$

The spectrum of a predictive process has only discrete lines:

$$P_x\left(e^{j\omega}\right) = \sum_{k=1}^{\infty} \alpha_k \delta\left(\omega - \omega_k\right)$$

Stochastic Process - Wold decomposition

The **Wold decomposition theorem** tells us that *any WSS process* $x[n]$ can be decomposed into a sum of a regular process $x_r[n]$ and predictive process $x_p[n]$:

$$x[n] = x_p[n] + x_r[n]$$

where the processes are **orthogonal**:

$$E[x_p[n] x_r^*[m]] = 0$$

Hence the spectrum of any WSS is composed of a continuous component with invertible $H(z)$ plus discrete lines:

$$P_x(e^{j\omega}) = |H(e^{j\omega})|^2 + \sum_{k=1}^{\infty} \alpha_k \delta(\omega - \omega_k)$$

Stochastic Process - Wold decomposition

Note that for orthogonal processes the power spectrum is additive.

If

$$z[n] = x[n] + y[n]$$

with

$$E[x[n]\, y^*[m]] = 0$$

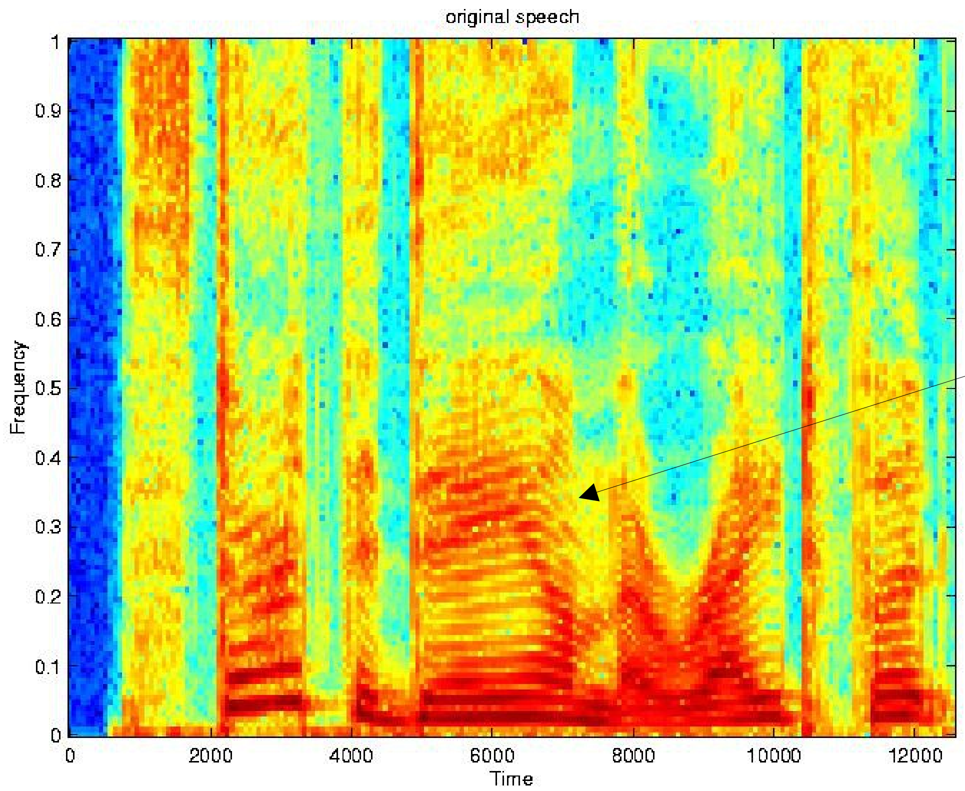
Then

$$P_z(e^{j\omega}) = P_x(e^{j\omega}) + P_y(e^{j\omega})$$

Stochastic Process -Spectrogram

Spectrogram: Power spectrum with strong spectral component can be estimate on short sequences, and hence, followed as it develops over time.

Example: Speech   `>> specgram(x);`



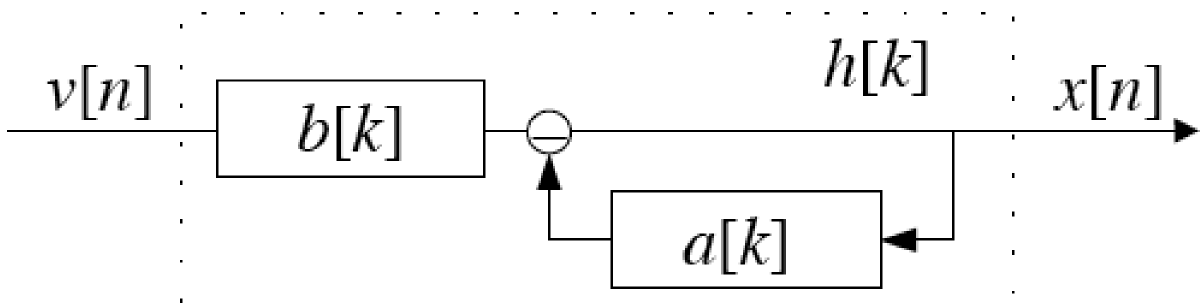
Note harmonic components

Assumption: WSS within each small window

Stochastic Process - ARMA process

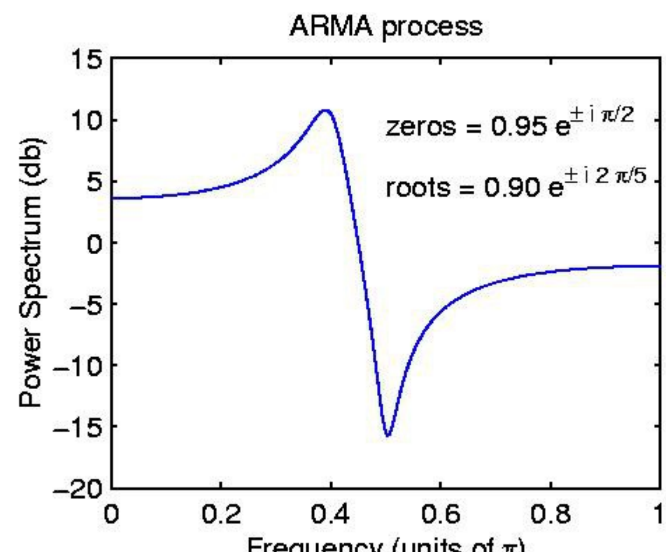
An important regular process is the **ARMA process:** white noise $v[n]$ filtered with a rational transfer function

$$H(z) = \frac{B(z)}{A(z)}$$



Example:

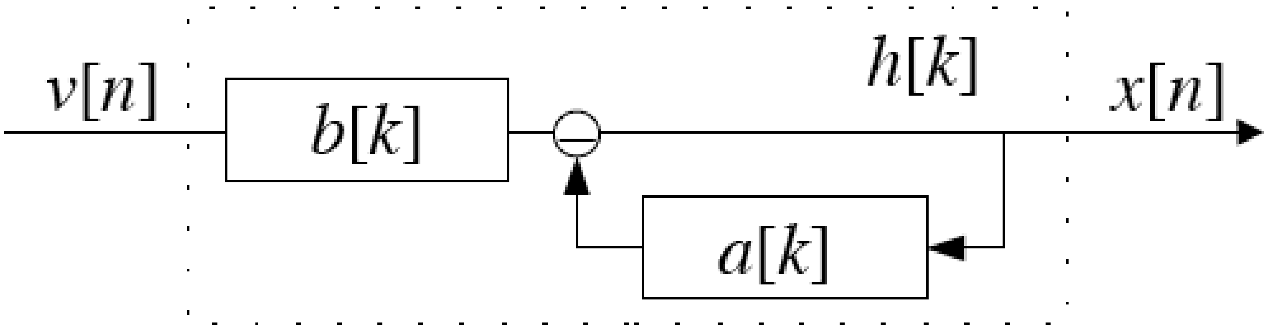
$$H(z) = \frac{1+0.9025\,z^{-2}}{1-0.5562\,z^{-1}+0.81\,z^{-2}}$$



Stochastic Process - ARMA modeling

ARMA modeling means to find $a[k]$, and $b[k]$ that approximate an observed $P_x(e^{j\omega})$ for a given $x[n]$

$$P_x\left(e^{j\omega}\right)=\frac{\left|B\left(e^{j\omega}\right)\right|^2}{\left|A\left(e^{j\omega}\right)\right|^2}$$



$$x[n]=-\sum_{l=1}^{P} a[l]x[n-l]+\sum_{l=0}^{Q} b[l]v[n-l]$$

Note that now the input is not given. Instead only statistical assumptions on the input are known (white WSS process).

Stochastic Process - ARMA modeling

We derive now statistical conditions for $a[k]$ and $b[k]$. Multiply the difference equation by $x^*[n-k]$ and take the expectation:

$$\sum_{l=0}^{P} a[l]\, r_x[k-l] = \sum_{l=0}^{Q} b[l]\, r_{vx}[k-l]$$

with cross-correlation: $r_{vx}[k-l] = E\left[v[n-l]\,x^*[n-k]\right]$

Using

$$x[n] = h[n] * v[n]$$

$$E\left[v[n]\,v^*[m]\right] = \delta(n-m)$$

we obtain (non-linear) conditions for $a[k]$ and $b[k]$

$$\sum_{l=0}^{P} a[l]\, r_x[k-l] = \sum_{l=0}^{Q} b[l]\, h^*[l-k]$$

Stochastic Process - ARMA modeling

These conditions are called the **Yule-Walker equations**

$$\sum_{l=0}^{P} a[l]\, r_x[k-l] = \sum_{l=0}^{Q-k} b[l+k]\, h^*[l] \equiv c[k]$$

In matrix form and using Hermitian symmetry of the auto-correlation we can write for lags 0...L:

$$\begin{bmatrix} r_x[0] & r_x^*[1] & r_x^*[2] & \cdots & r_x^*[P] \\ r_x[1] & r_x[0] & r_x^*[1] & \cdots & r_x^*[P-1] \\ \vdots & \vdots & \vdots & & \vdots \\ r_x[Q] & r_x[Q-1] & r_x[Q-2] & \cdots & r_x[Q-P] \\ r_x[Q+1] & r_x[Q] & r_x[Q-1] & \cdots & r_x[Q-P+1] \\ \vdots & \vdots & \vdots & & \vdots \\ r_x[L] & r_x[L-1] & r_x[L-2] & \cdots & r_x[L-P] \end{bmatrix} \begin{bmatrix} 1 \\ a[1] \\ a[2] \\ \vdots \\ a[P] \end{bmatrix} = \begin{bmatrix} c[0] \\ c[1] \\ \vdots \\ c[Q] \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Since $c[k]$ depend on $h[k]$ these equations are non-linear in $a[k]$ and $b[k]$. They simplify for Q=0, i.e. AR model only.

Stochastic Process - AR modeling

For Q=0 the Yule-Walker equations simplify to

$$\sum_{l=0}^{P} a[l]\, r_x[k-l] = |b[0]|^2 \delta(k)$$

In matrix form and using Hermitian symmetry of the auto-correlation we obtain the AR **normal equations**:

$$\begin{bmatrix} r_x[0] & r_x^*[1] & r_x^*[2] & \cdots & r_x^*[P] \\ r_x[1] & r_x[0] & r_x^*[1] & \cdots & r_x^*[P-1] \\ r_x[2] & r_x[1] & r_x[0] & \cdots & r_x^*[P-2] \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ r_x[P] & r_x[P-1] & r_x[P-2] & \cdots & r_x[0] \end{bmatrix} \begin{bmatrix} 1 \\ a[1] \\ a[2] \\ \vdots \\ a[P] \end{bmatrix} = \begin{bmatrix} |b[0]|^2 \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Which are a set of linear equations for the unknown $a[k]$.

Stochastic Process - AR modeling

The AR normal equations can be rewritten as:

$$\begin{bmatrix} r_x[0] & r_x^*[1] & \cdots & r_x^*[P-1] \\ r_x[1] & r_x[0] & \cdots & r_x^*[P-2] \\ \vdots & \vdots & \ddots & \vdots \\ r_x[P-1] & r_x[P-2] & \cdots & r_x[0] \end{bmatrix} \begin{bmatrix} a[1] \\ a[2] \\ \vdots \\ a[P] \end{bmatrix} = - \begin{bmatrix} r_x[1] \\ r_x[2] \\ \vdots \\ r_x[P] \end{bmatrix}$$

$$\boldsymbol{R}_{\boldsymbol{xx}} \boldsymbol{a} = -\boldsymbol{r}_{\boldsymbol{x}}$$

Linear equations with Toeplitz Hermitian form can be solved efficiently with the Levinson-Durbin recursion.

Computes the auto-correlations and solves this linear equation:

```
>> a = lpc(x,P);
```

Stochastic Process - Linear prediction

The goal of linear prediction is to predict sample $x[n]$ from its past P samples:

$$\hat{x}[n] = -\sum_{k=1}^{P} a[k]x[n-k]$$

And $a[k]$ are called the **linear prediction coefficients** (LPC). The prediction error is given by

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

The LPC that minimize the expected error

$$a = argmin\, E[\| e[n] \|^2]$$

satisfy the same linear equations as the AR model parameters:

$$\sum_{l=0}^{P} a[l] r_x[k-l] = \epsilon \delta(k)$$

which we obtain by taking derivatives with respect to $a[k]$.

Stochastic Process - LPC and Wold Decomp.

In matlab all this is done with:

```
>> a = lpc(x,P);
```

The error process $e[n]$ represents the "new" or "innovative" part over the linear prediction, hence the name innovation process.

One can show that the innovation in linear prediction is white and uncorrelated to the prediction,

$$E[e[n]e^*[m]] = \delta(n-m)$$

$$E[e[n]\hat{x}^*[m]] = 0$$

Stochastic Process - AR and LPC

Assignment 6:

- Pick either speech **or** heart rate

```
>> load hrs.mat; x=hrs;
>> [x,fs]=wavread('speech.wav');x=x(5501:6012);
```

- Generate an AR model by solving normal equations.
- Plot an estimated power spectrum together with the AR model spectrum.
- Show the innovation process and its power spectrum.

- Compute the LPC coefficient using `lpc(x,P)`
- Compare your AR results with LPC.
- Show the linear prediction estimation error.
- Compute the signal to noise ratio (SNR) of your LPC model:

$$SNR = \frac{\sum_n |\hat{x}[n]|^2}{\sum_n |e[n]|^2}$$

- Show SNR in dB as a function of model order *P*.