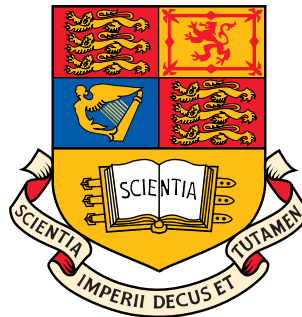

Advanced Signal Processing

Linear Stochastic Models

Danilo Mandic
room 813, ext: 46271



Department of Electrical and Electronic Engineering
Imperial College London, UK

d.mandic@imperial.ac.uk, URL: www.commsp.ee.ic.ac.uk/~mandic

Aims of this lecture

- To introduce stochastic models for real world data
- Learn how stochastic models shape the spectrum of white noise
- Understand how an ensemble of realisations of a stochastic process is created
- Learn how to derive parameters of linear stochastic models
- Introduce some special cases (autoregressive, moving average)
- Learn how to determine the model order
- Address optimal model order selection criteria
- Apply stochastic modelling to real world data (speech, environmental, finance)

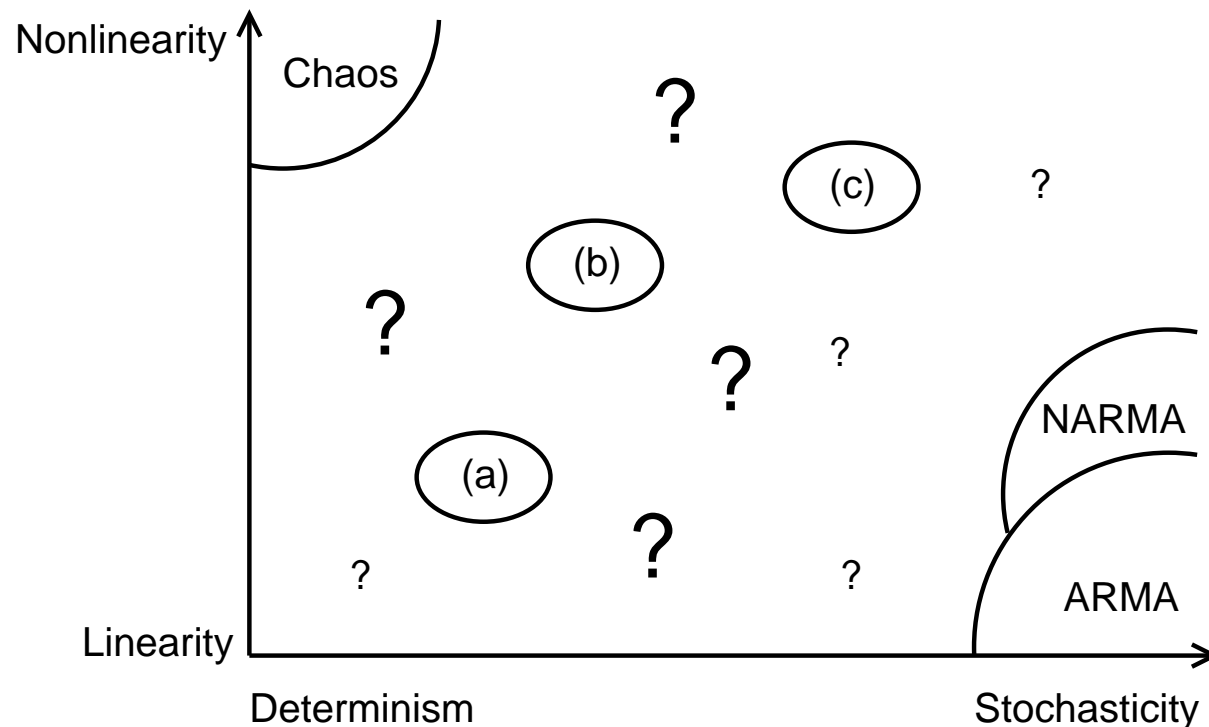
This material is a first fundamental step for the modelling of real world data

How can we categorise real-world measurements?

where would you place a DC level in WGN, $x[n] = A + w[n]$, $w \sim \mathcal{N}(0, \sigma_w^2)$

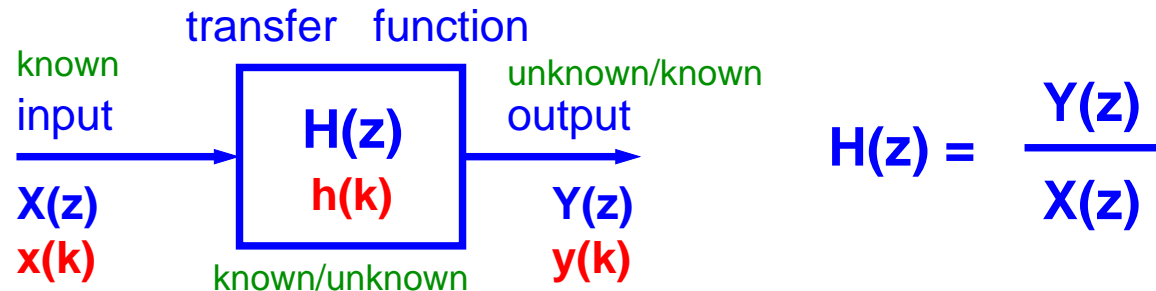
(a) Noisy oscillations, (b) Nonlinearity and noisy oscillations, (c) Random nonlinear process

(? left) Route to chaos, (? top) stochastic chaos, (? middle) mixture of sources



Our lecture is about ARMA models (linear stochastic)

Linear systems



Described by their impulse response $h(n)$ or the transfer function $H(z)$

In the frequency domain (remember that $z = e^{j\theta}$) the transfer function is

$$H(\theta) = \sum_{n=-\infty}^{\infty} h(n)e^{-jn\theta} \quad \{x[n]\} \rightarrow \left| \begin{array}{c} \{h(n)\} \\ H(\theta) \end{array} \right| \rightarrow \{y[n]\}$$

that is
$$y[n] = \sum_{r=-\infty}^{\infty} h(r)x[n-r] = h * x$$

The next two slides show how to calculate the power of the output, $y(k)$.

Linear systems – statistical properties \leadsto mean and variance

i) Mean

$$E\{y[n]\} = E\left\{\sum_{r=-\infty}^{\infty} h(r)x[n-r]\right\} = \sum_{r=-\infty}^{\infty} h(r)E\{x[n-r]\}$$

$$\Rightarrow \mu_y = \mu_x \sum_{r=-\infty}^{\infty} h(r) = \mu_x H(0)$$

[NB: $H(\theta) = \sum_{r=-\infty}^{\infty} h(r)e^{-jr\theta}$. For $\theta = 0$, then $H(0) = \sum_{r=-\infty}^{\infty} h(r)$]

ii) Cross-correlation

$$r_{yx}(m) = E\{y[n]x[n+m]\} = \sum_{r=-\infty}^{\infty} h(r)E\{x[n-r]x[n+m]\}$$

$$= \sum_{r=-\infty}^{\infty} h(r)r_{xx}(m+r) \quad \text{convolution of input ACF and } \{h\}$$

$$\Rightarrow \text{Cross-power spectrum } S_{yx}(f) = \mathcal{F}(r_{yx}) = S_{xx}(f)H(f)$$

Linear systems – statistical properties \leadsto crosscorrelation

These are key properties \leadsto used in AR spectrum estimation

From $r_{xy}(m) = r_{yx}(-m)$ we have

$r_{xy}(m) = \sum_{r=-\infty}^{\infty} h(r)r_{xx}(m-r)$. Now we write

$$\begin{aligned} r_{yy}(m) &= E\{y[n]y[n+m]\} = \sum_{r=-\infty}^{\infty} h(r)E\{x[n-r]y[n+m]\} \\ &= \sum_{r=-\infty}^{\infty} h(r)r_{xy}(m+r) = \sum_{r=-\infty}^{\infty} h(-r)r_{xy}(m-r) \end{aligned}$$

by taking Fourier transforms we have

$$S_{xy}(f) = S_{xx}(f)H(f)$$

$$S_{yy}(f) = S_{xy}(f)H(-f) \rightsquigarrow \text{function of } r_{xx}$$

or

$$\mathbf{S}_{yy}(\mathbf{f}) = \mathbf{H}(\mathbf{f})\mathbf{H}(-\mathbf{f})\mathbf{S}_{xx}(\mathbf{f}) = |\mathbf{H}(\mathbf{f})|^2\mathbf{S}_{xx}(\mathbf{f})$$

Output power spectrum = input power spectrum \times squared transfer function

Motivation: Wold decomposition theorem (also mentioned in your coursework)

The most fundamental justification for time series analysis is due to Wold's decomposition theorem, where it is explicitly proved that any (stationary) time series can be decomposed into two different parts: **deterministic** and **stochastic**.

Therefore, a general random process can be written a sum of two processes

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

$\Rightarrow x_r[n]$ – regular random process

$\Rightarrow x_p[n]$ – predictable process, with $x_r[n] \perp x_p[n]$,

$$E\{x_r[m]x_p[n]\} = 0$$

\Rightarrow we can treat **separately** the **predictable** part (e.g. a deterministic signal) and the **random** signal.

Our focus will be on the modelling of the random component

Linear stochastic processes

It therefore follows that the general form for the power spectrum of a WSS process is

$$P_x(e^{j\omega}) = P_{x_r}(e^{j\omega}) + \sum_{k=1}^N \alpha_k \delta(\omega - \omega_k)$$

We are interested in processes generated by **filtering white noise with a linear shift-invariant filter** that has a rational system function. This class of digital filters includes the following system functions:

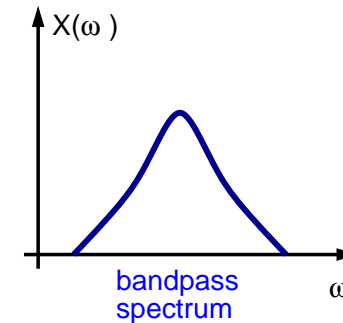
- Autoregressive (AR) \rightarrow all pole system $\rightarrow H(z) = 1/A(z)$
- Moving Average (MA) \rightarrow all zero system $\rightarrow H(z) = B(z)$
- Autoregressive Moving Average (ARMA) \rightarrow poles and zeros
 $\rightarrow H(z) = B(z)/A(z)$

Notice the difference between shift-invariance and time-invariance

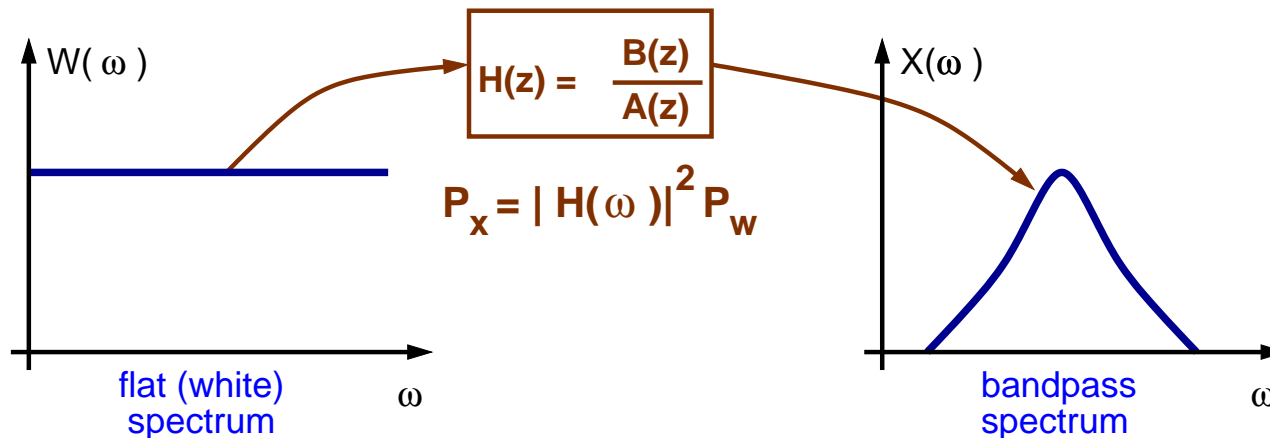
How do we model a real world signal?

Suppose the measured real world signal has e.g. a bandpass (any other) power spectrum

We desire to describe the whole long signal with very few parameters



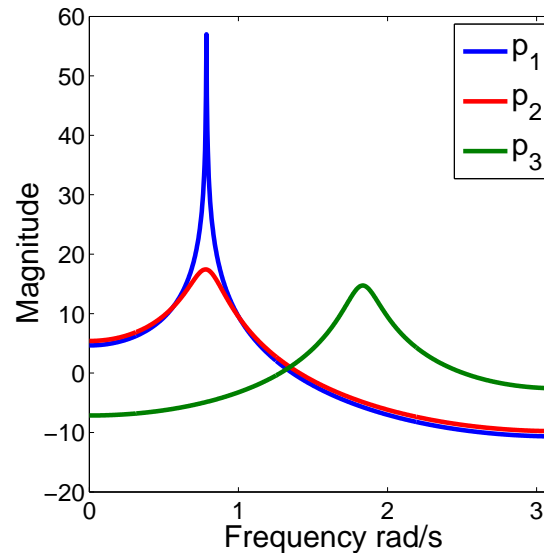
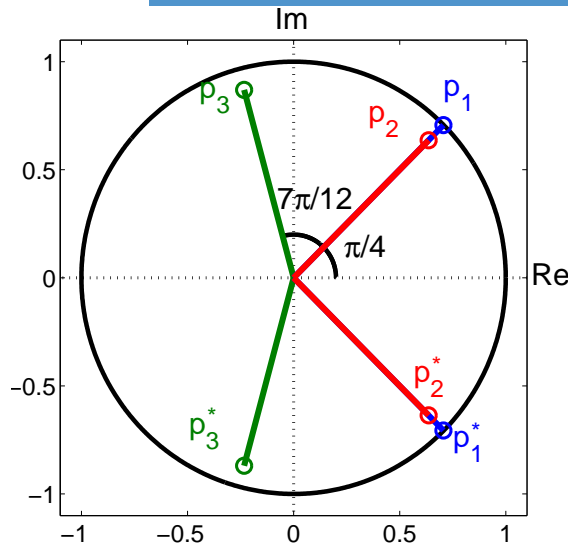
1. Can we model first and second statistics of real world signal by shaping the white noise spectrum using some transfer function?
2. Does this produce the same second order properties (mean, variance, ACF, spectrum) for any white noise input?



Can we use this linear stochastic model for prediction?

Example 1: Second-order all-pole transfer functions

$$p_1 = 0.999\exp(j\pi/4), p_2 = 0.9\exp(j\pi/4), p_3 = 0.9\exp(j7\pi/12)$$



We have two conjugate complex poles, e.g. p_1 and p_1^* , therefore

$$H(z) = \frac{1}{(z - p_1)(z - p_1^*)} = \frac{z^{-2}}{(1 - p_1 z^{-1})(1 - p_1^* z^{-1})}$$

Transfer function for $p = \rho e^{j\theta}$ (ignoring z^{-2} in the numerator on the RHS):

$$H(z) = \frac{1}{(1 - \rho e^{j\theta} z^{-1})(1 - \rho e^{-j\theta} z^{-1})} = \frac{1}{1 - 2\rho \cos(\theta) z^{-1} + \rho^2 z^{-2}}$$

for the sinewave $\rho = 1 \Rightarrow H(z) = \frac{1}{1 - 2\cos(\theta)z^{-1} + z^{-2}} = \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2}}$

\Rightarrow **The sinewave can be modelled as an autoregressive process (later)**

Spectrum of ARMA models

recall that two conjugate complex poles of $A(z)$ give one peak in the spectrum

$ACF \equiv PSD$ in terms of the information available

In ARMA modelling we filter white noise $w[n]$ (so called driving input) with a causal linear shift-invariant filter with the transfer function $H(z)$, a rational system function with p poles and q zeros given by

$$X(z) = H(z)W(z) \quad \leadsto \quad H(z) = \frac{B_q(z)}{A_p(z)} = \frac{\sum_{k=0}^q b_k z^{-k}}{1 + \sum_{k=1}^p a_k z^{-k}}$$

For a stable $H(z)$, the **ARMA(p,q) stochastic process** $x[n]$ will be wide-sense stationary. For the **driving noise** power $P_w = \sigma_w^2$, the power of the stochastic process $x[n]$ is (recall $P_y = |H(z)|^2 P_x = H(z)H^*(z)P_x$)

$$P_x(z) = \sigma_w^2 \frac{B_q(z)B_q(z^{-1})}{A_p(z)A_p(z^{-1})} \quad \Rightarrow \quad P_z(e^{j\theta}) = \sigma_w^2 \frac{|B_q(e^{j\theta})|^2}{|A_p(e^{j\theta})|^2} = \sigma_w^2 \frac{|B_q(\omega)|^2}{|A_p(\omega)|^2}$$

Notice that “ $(\cdot)^*$ ” in analogue frequency corresponds to “ z^{-1} ” in “digital freq.”

Difference equation representation \leadsto the ACF follows the data model!

Random processes $x[n]$ and $w[n]$ are related by a linear difference equation with constant coefficients, given by

$$H(z) = \frac{X(z)}{W(z)} = \frac{B(z)}{A(z)} \quad \Leftrightarrow \text{ARMA}(p,q) \Leftrightarrow x[n] = \underbrace{\sum_{l=1}^p a_l x[n-l]}_{\text{autoregressive}} + \underbrace{\sum_{l=0}^q b_l w[n-l]}_{\text{moving average}}$$

Notice that the autocorrelation function of $x[n]$ and crosscorrelation between the **stochastic process** $x[n]$ and **the driving input** $w[n]$ follow the same difference equation, i.e. if we multiply both sides of the above equation by $x[n-k]$ and take the statistical expectation, we have

$$r_{xx}(k) = \underbrace{\sum_{l=1}^p a_l r_{xx}(k-l)}_{\text{easy to calculate}} + \underbrace{\sum_{l=0}^q b_l r_{xw}(k-l)}_{\text{can be complicated}}$$

Since x is WSS, it follows that $x[n]$ and $w[n]$ are jointly WSS

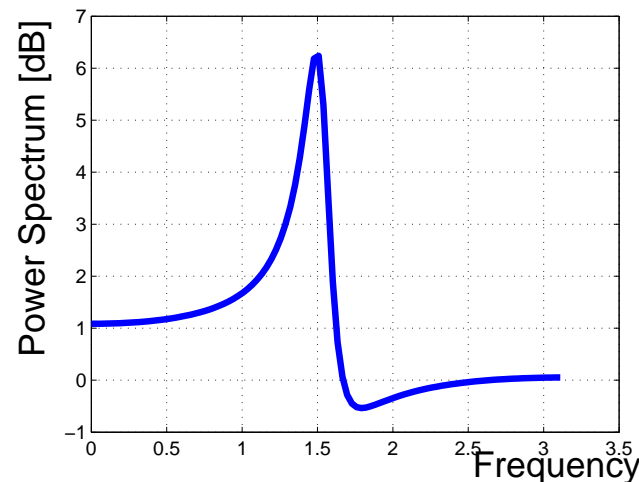
Example 2: Can the shape of power spectrum tell us about the order of the polynomials $B(z)$ and $A(z)$?

Plot the power spectrum of an ARMA(2,2) process for which

- the zeros of $H(z)$ are $z = 0.95e^{\pm j\pi/2}$
- poles are at $z = 0.9e^{\pm j2\pi/5}$

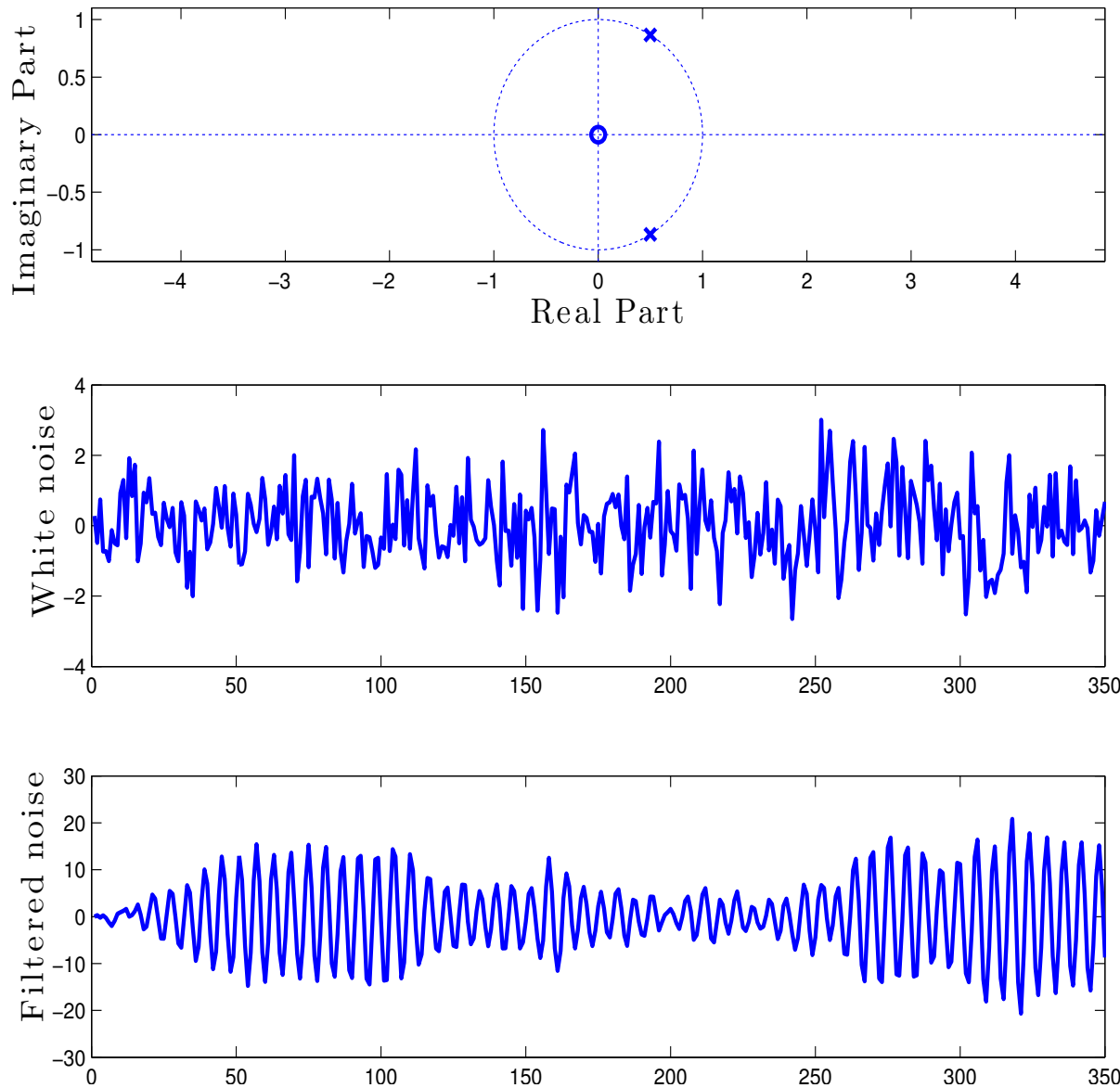
Solution: The system function is (poles and zeros – resonance & sink)

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



Example 3: How about a sinewave?

Is it nonlinear deterministic or linear stochastic?



Matlab code:

```
z1=0;
p1=[0.5+0.866i,0.5-0.866i];
[num1,den1]=zp2tf(z1,p1,1);
zplane(num1,den1);
s=randn(1,1000);
s1=filter(num1,den1,s);
figure;
subplot(311),plot(s),
subplot(313),plot(s1),
subplot(312),;
zplane(num1,den1)
```

The AR model of a sinewave

$$x(k) = a_1 x(k-1) + a_2 x(k-2) + w(k)$$
$$a_1 = -1, \quad a_2 = 0.98, \quad w \sim N(0,1)$$

General linear processes: Stationarity and invertibility

Consider a linear stochastic process \rightsquigarrow output from a linear filter, driven by WGN $w[n]$

$$x[n] = w[n] + b_1 w[n-1] + b_2 w[n-2] + \cdots = w[n] + \sum_{j=1}^{\infty} b_j w[n-j]$$

that is, **a weighted sum** of past samples of **driving white noise** $w[n]$.

For this process to be a valid stationary process, the coefficients must be **absolutely summable**, that is $\sum_{j=0}^{\infty} |b_j| < \infty$.

This also implies that under stationarity conditions, $x[n]$ is also a weighted sum of past values of x , plus an added shock $w[n]$, that is

$$x[n] = a_1 x[n-1] + a_2 x[n-2] + \cdots + w[n]$$

◦ Linear Process is *stationary* if $\sum_{j=0}^{\infty} |b_j| < \infty$

◦ Linear Process is *invertible* if $\sum_{j=0}^{\infty} |a_j| < \infty$

Recall that $H(\omega) = \sum_{n=0}^{\infty} h(n)e^{-j\omega n} \rightarrow \text{for } \omega = 0 \Rightarrow H(0) = \sum_{n=0}^{\infty} h(n)$

Autoregressive processes (pole-only)

A general $AR(p)$ process (autoregressive of order p) is given by

$$x[n] = a_1x[n-1] + \cdots + a_px[n-p] + w[n] = \sum_{i=1}^p a_i x[n-i] + w[n] = \mathbf{a}^T \mathbf{x}[n] + w[n]$$

Observe the auto-regression in $\{x[n]\}$ \nleftrightarrow the past of x is used to generate the future

Duality between AR and MA processes:

For example, the first order autoregressive process, $AR(1)$,

$$x[n] = a_1x[n-1] + w[n] \quad \Leftrightarrow \quad \sum_{j=0}^{\infty} b_j w[n-j]$$

has an MA representation, too (right hand side above).

This follows from the duality between IIR and FIR filters.

ACF and normalised ACF of AR processes

To obtain the autocorrelation function of an AR process, multiply the above equation by $x[n - k]$ to obtain

$$\begin{aligned} x[n - k]x[n] &= a_1x[n - k]x[n - 1] + a_2x[n - k]x[n - 2] + \cdots \\ &\quad + a_px[n - k]x[n - p] + x[n - k]w[n] \end{aligned}$$

Notice that $E\{x[n - k]w[n]\}$ vanishes when $k > 0$. Therefore we have

$$\begin{aligned} r_{xx}(0) &= a_1r_{xx}(1) + a_2r_{xx}(2) + \cdots + a_pr_{xx}(p) + \sigma_w^2, & k = 0 \\ r_{xx}(k) &= a_1r_{xx}(k - 1) + a_2r_{xx}(k - 2) + \cdots + a_pr_{xx}(k - p), & k > 0 \end{aligned}$$

On dividing throughout by $r_{xx}(0)$ we obtain

$$\rho(k) = a_1\rho(k - 1) + a_2\rho(k - 2) + \cdots + a_p\rho(k - p) \quad k > 0$$

Quantities $\rho(k)$ are called **normalised correlation coefficients**

Variance and spectrum of AR processes

Variance:

When $k = 0$ the contribution from the term $E\{x[n - k]w[n]\}$ is σ_w^2 , and

$$r_{xx}(0) = a_1 r_{xx}(-1) + a_2 r_{xx}(-2) + \cdots + a_p r_{xx}(-p) + \sigma_w^2$$

Divide by $r_{xx}(0) = \sigma_x^2$ to obtain

$$\sigma_x^2 = \frac{\sigma_w^2}{1 - \rho_1 a_1 - \rho_2 a_2 - \cdots - \rho_p a_p}$$

Power spectrum: (recall the expression for the output power of a linear system \rightarrow see Slides 4 – 6)

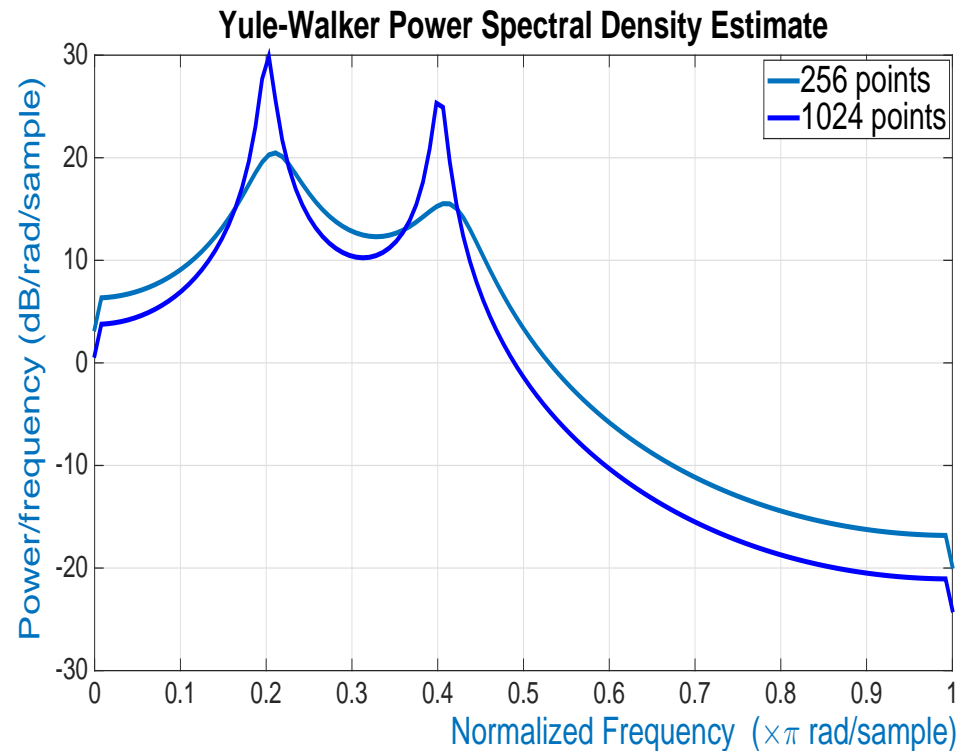
$$P_{xx}(f) = \frac{2\sigma_w^2}{|1 - a_1 e^{-j2\pi f} - \cdots - a_p e^{-j2\pi p f}|^2} \quad 0 \leq f \leq 1/2$$

Look at “*Spectrum of Linear Systems*” from Lecture 1: Course Introduction

Example 4: $AR(p)$ signal generation

$$\mathbf{a} = [2.2137, -2.9403, 2.1697, -0.9606]$$

- Generate the input signal \mathbf{x} by filtering white noise through the AR filter
- Estimate the PSD of \mathbf{x} based on a fourth-order AR model
- **Careful!** The Matlab routines require the AR coeff. \mathbf{a} in the format $\mathbf{a} = [1, -a_1, \dots, -a_p]$



Notice the dependence on data length

Solution:

```
randn('state',1);  
x = filter(1,a,randn(256,1));  
pyulear(x,4)
```

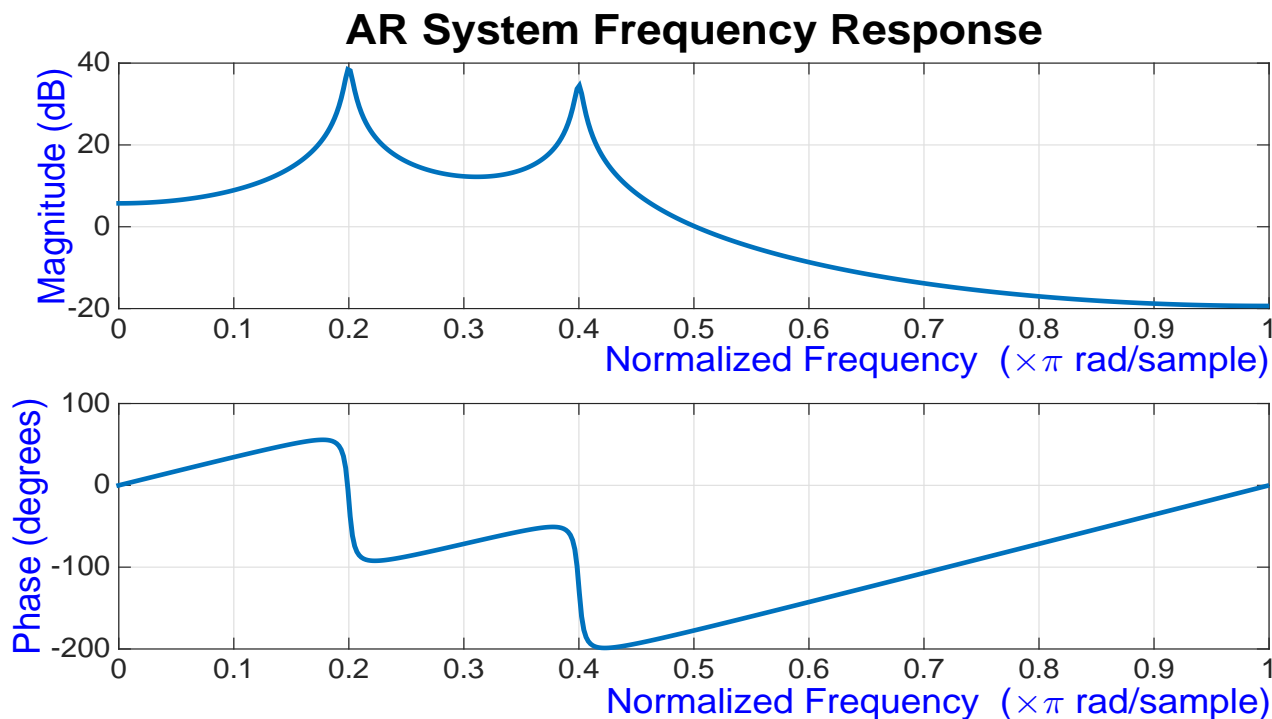
% AR system output
% Fourth-order estimate

Example 5: Alternative AR power spectrum calculation

From the AR(4) system given by

$$y[n] = 2.2137y[n-1] - 2.9403y[n-2] + 2.1697y[n-3] - 0.9606y[n-4] + w[n]$$

```
a = [1 -2.2137 2.9403 -2.1697 0.9606]; % AR filter coefficients  
freqz(1,a) % AR filter frequency response  
title('AR System Frequency Response')
```



Key: Finding AR coefficients \leadsto the Yule–Walker eqns

(there are several similar forms – we follow the most concise one)

For $k = 1, 2, \dots, p$ from the general autocorrelation function, we obtain a set of equations:

$$\begin{aligned} r_{xx}(1) &= a_1 r_{xx}(0) + a_2 r_{xx}(1) + \dots + a_p r_{xx}(p-1) \\ r_{xx}(2) &= a_1 r_{xx}(1) + a_2 r_{xx}(0) + \dots + a_p r_{xx}(p-2) \\ &\vdots \\ r_{xx}(p) &= a_1 r_{xx}(p-1) + a_2 r_{xx}(p-2) + \dots + a_p r_{xx}(0) \end{aligned}$$

These equations are called the **Yule–Walker or normal equations**.

Their solution gives us the set of **autoregressive parameters**

$$\mathbf{a} = [a_1, \dots, a_p]^T.$$

The above can be expressed in a vector–matrix form as

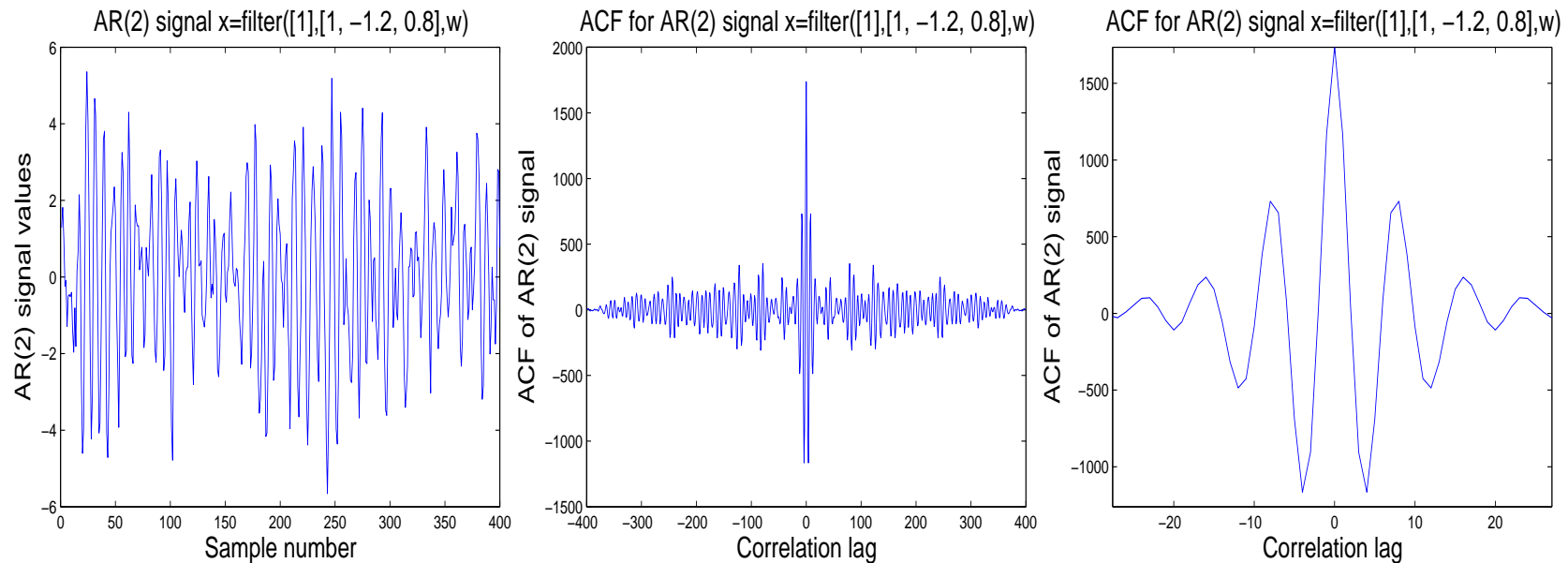
$$\mathbf{r}_{xx} = \mathbf{R}_{xx} \mathbf{a} \quad \Rightarrow \quad \mathbf{a} = \mathbf{R}_{xx}^{-1} \mathbf{r}_{xx}$$

The ACF matrix \mathbf{R}_{xx} is positive definite (Toeplitz) which guarantees matrix inversion

Example 6: Find the parameters of an AR(2) process

given by $x[n] = 1.2x[n-1] - 0.8x[n-2] + w[n]$

Coursework: comment on the shape of the ACF for large lags



Matlab: `for i=1:6; [a,e]=aryule(x,i); display(a);end`

$$\mathbf{a}^{(1)} = [0.6689] \quad \mathbf{a}^{(2)} = [1.2046, -0.8008]$$

$$\mathbf{a}^{(3)} = [1.1759, -0.7576, -0.0358]$$

$$\mathbf{a}^{(4)} = [1.1762, -0.7513, -0.0456, 0.0083]$$

$$\mathbf{a}^{(5)} = [1.1763, -0.7520, -0.0562, 0.0248, -0.0140]$$

$$\mathbf{a}^{(6)} = [1.1762, -0.7518, -0.0565, 0.0198, -0.0062, -0.0067]$$

Normal equations for the autocorrelation coefficients

For the autocorrelation coefficients

$$\rho_k = r_{xx}(k)/r_{xx}(0)$$

we have

$$\rho_1 = a_1 + a_2\rho_1 + \cdots + a_p\rho_{p-1}$$

$$\rho_2 = a_1\rho_1 + a_2 + \cdots + a_p\rho_{p-2}$$

$$\vdots = \vdots$$

$$\rho_p = a_1\rho_{p-1} + a_2\rho_{p-2} + \cdots + a_p$$

When does the sequence $\{\rho_0, \rho_1, \rho_2, \dots\}$ vanish?

Homework: Try command **xcorr** in Matlab

Example 7: Yule–Walker modelling in Matlab

In Matlab – Power spectral density using Y–W method *pyulear*

```
Pxx = pyulear(x,p)
[Pxx,w] = pyulear(x,p,nfft)
[Pxx,f] = pyulear(x,p,nfft,fs)
[Pxx,f] = pyulear(x,p,nfft,fs,'range')
[Pxx,w] = pyulear(x,p,nfft,'range')
```

Description:

```
Pxx = pyulear(x,p)
```

implements the Yule-Walker algorithm, and returns Pxx, an estimate of the power spectral density (PSD) of the vector x.

To remember for later → This estimate is also an estimate of the maximum entropy.

See also **aryule, lpc, pburg, pcov, peig, periodogram**

From data to an $AR(p)$ model

So far, we assumed the model (AR, MA, or ARMA) and analysed the ACF and PSD based on known model coefficients.

In practice: **DATA** \rightarrow **MODEL**

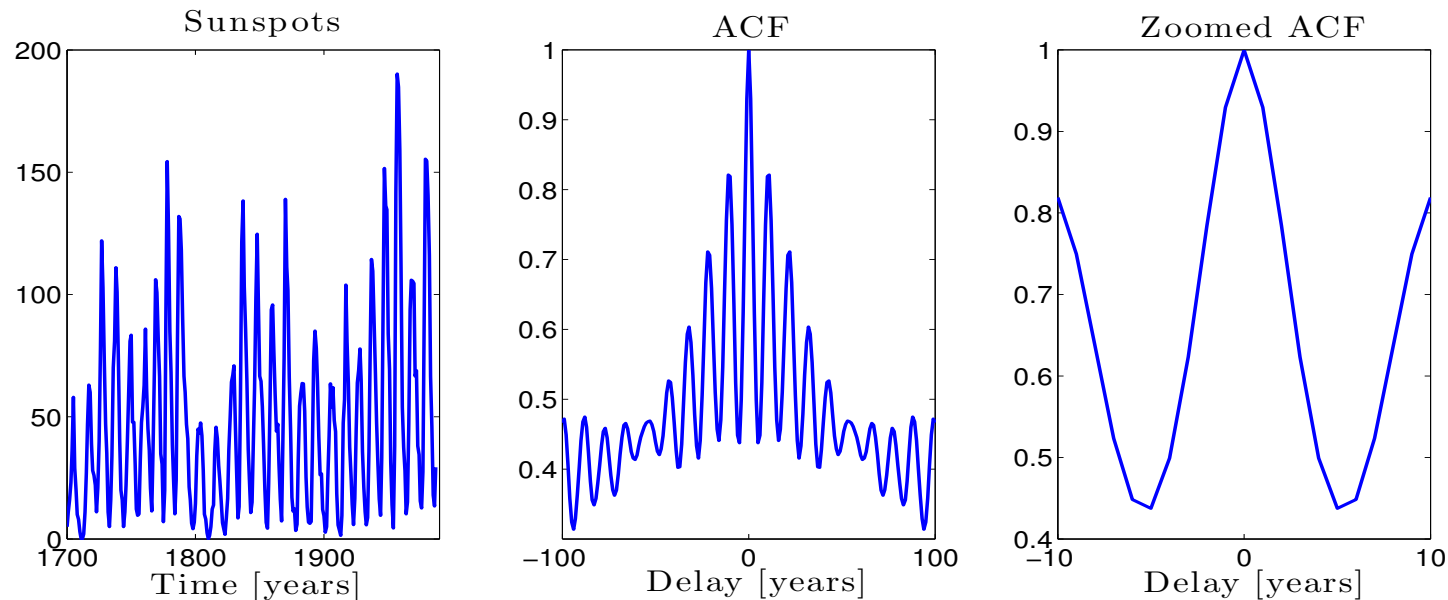
This procedure is as follows:

- * record data $x(k)$
- * find the autocorrelation of the data $ACF(x)$
- * divide by $r_{xx}(0)$ to obtain correlation coefficients $\rho(k)$
- * write down Yule-Walker equations
- * solve for the vector of AR parameters

The problem is that we do not know the model order p beforehand.
An illustration of the importance of the correct model order is given in the following example.

Example 8: Sunspot number estimation

consistent with the properties of a second order AR process



$$\mathbf{a}_1 = [0.9295] \quad \mathbf{a}_2 = [1.4740, -0.5857]$$

$$\mathbf{a}_3 = [1.5492, -0.7750, 0.1284]$$

$$\mathbf{a}_4 = [1.5167, -0.5788, -0.2638, 0.2532]$$

$$\mathbf{a}_5 = [1.4773, -0.5377, -0.1739, 0.0174, 0.1555]$$

$$\mathbf{a}_6 = [1.4373, -0.5422, -0.1291, 0.1558, -0.2248, 0.2574]$$

↪ The sunspots model is $x[n] = 1.474 x[n-1] - 0.5857 x[n-2] + w[n]$

Special case #1: AR(1) process (Markov)

For Markov processes, instead of the iid condition, we have the **first order conditional expectation**, that is

$$p(x[n], x[n-1], x[n-2], \dots, x[0]) = p(x[n]|x[n-1])$$

Then $x[n] = a_1 x[n-1] + w[n] = w[n] + a_1 w[n-1] + a_1^2 w[n-2] + \dots$

Therefore: first order dependence, order-1 memory

- i) for the AR(1) process to be **stationary** $-1 < a_1 < 1$.
- ii) **Autocorrelation Function of AR(1)**: From the Yule-Walker equations

$$r_{xx}(k) = a_1 r_{xx}(k-1), \quad k > 0$$

In terms of the **correlation coefficients**, $\rho(k) = r(k)/r(0)$, with $\rho_0 = 1$

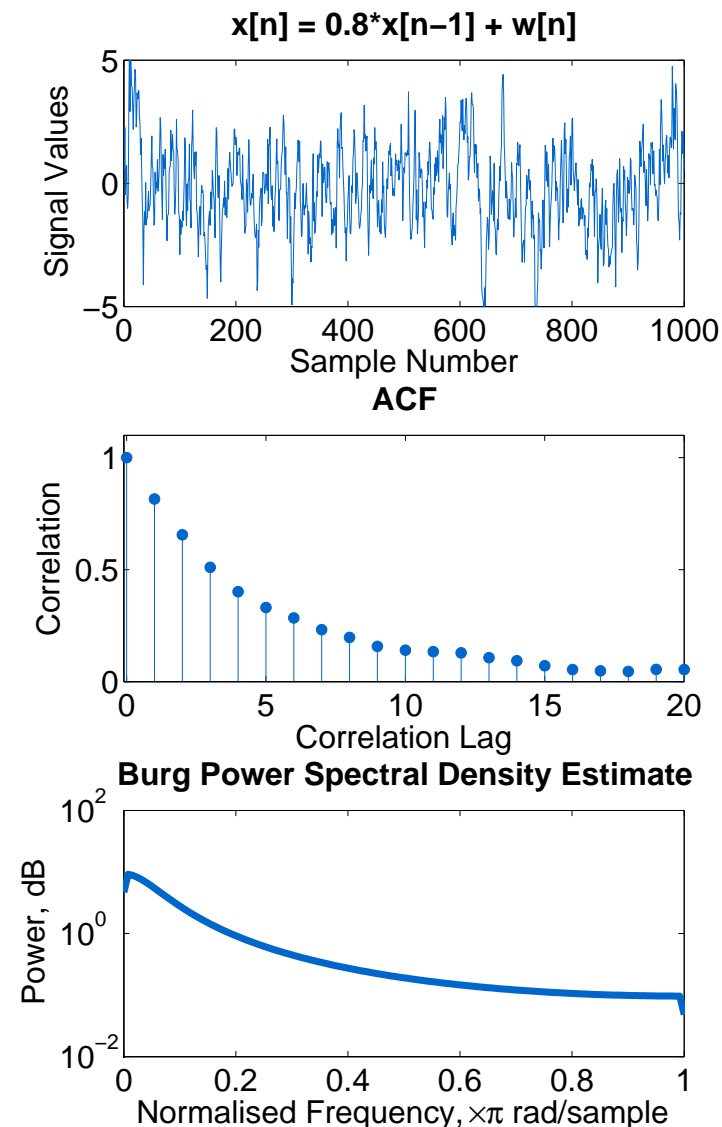
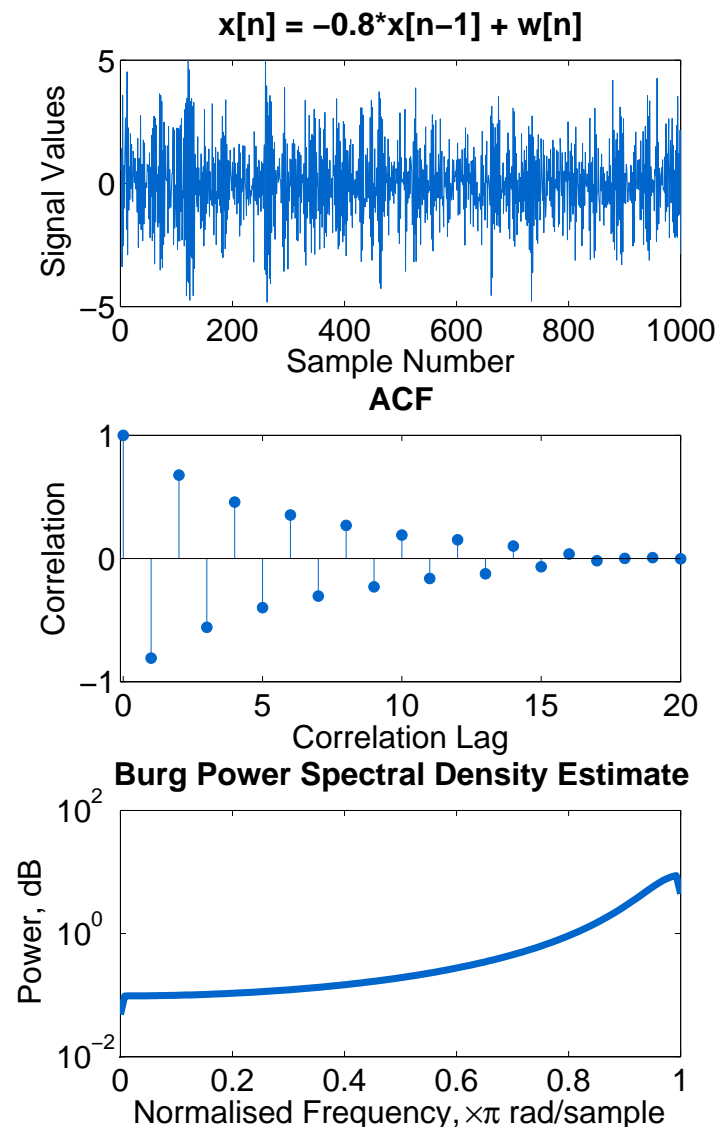
$$\rho_k = a_1^k, \quad k > 0$$

Notice the difference in the behaviour of the ACF for a_1 positive and negative

Example 9: ACF and spectrum of AR(1) for $a = \pm 0.8$

$a < 0 \rightarrow$ High Pass

$a > 0 \rightarrow$ Low Pass



Variance and power spectrum of AR(1) process

Both can be calculated directly from the general expression for the variance and spectrum of $AR(p)$ processes.

- **Variance:** Also from a general expression for the variance of linear processes from Lecture 1

$$\sigma_x^2 = \frac{\sigma_w^2}{1 - \rho_1 a_1} = \frac{\sigma_w^2}{1 - a_1^2}$$

- **Power spectrum:** Notice how the flat PSD of WGN is shaped according to the position of the pole of $AR(1)$ model (LP or HP)

$$P_{xx}(f) = \frac{2\sigma_w^2}{|1 - a_1 e^{-j2\pi f}|^2} = \frac{2\sigma_w^2}{1 + a_1^2 - 2a_1 \cos(2\pi f)}$$

Special case #2: Second order autoregressive processes, $p = 2, q = 0$, hence the notation **AR(2)**

The input-output functional relationship is given by ($w[n] \sim$ any white noise)

$$x[n] = a_1 x[n-1] + a_2 x[n-2] + w[n]$$

$$X(z) = (a_1 z^{-1} + a_2 z^{-2}) X(z) + W(z)$$

$$\Rightarrow H(z) = \frac{X(z)}{W(z)} = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2}}$$

$$H(\omega) = H(e^{j\omega}) = \frac{1}{1 - a_1 e^{-j\omega} - a_2 e^{-2j\omega}}$$

Y-W equations for $p=2$

$$\rho_1 = a_1 + a_2 \rho_1$$

$$\rho_2 = a_1 \rho_1 + a_2$$

Connecting a's and ρ 's

$$\rho_1 = \frac{a_1}{1 - a_2}$$

$$\rho_2 = a_2 + \frac{a_1^2}{1 - a_2}$$

when solved for a_1 and a_2 , we have

$$a_1 = \frac{\rho_1(1 - \rho_2)}{1 - \rho_1^2} \quad a_2 = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}$$

Since $\rho_1 < \rho(0) = 1 \Leftrightarrow$ a **stability condition** on a_1 and a_2

Variance and power spectrum

Both readily obtained from the general AR(2) process equation!

Variance

$$\sigma_x^2 = \frac{\sigma_w^2}{1 - \rho_1 a_1 - \rho_2 a_2} = \left(\frac{1 - a_2}{1 + a_2} \right) \frac{\sigma_w^2}{(1 - a_2)^2 - a_1^2}$$

Power spectrum

$$\begin{aligned} P_{xx}(f) &= \frac{2\sigma_w^2}{|1 - a_1 e^{-j2\pi f} - a_2 e^{-j4\pi f}|^2} \\ &= \frac{2\sigma_w^2}{1 + a_1^2 + a_2^2 - 2a_1(1 - a_2 \cos(2\pi f)) - 2a_2 \cos(4\pi f)}, \quad 0 \leq f \leq 1/2 \end{aligned}$$

Stability conditions \rightsquigarrow (Condition 1 can be obtained from the denominator of variance, Condition 2 from the expression for ρ_1 , etc.)

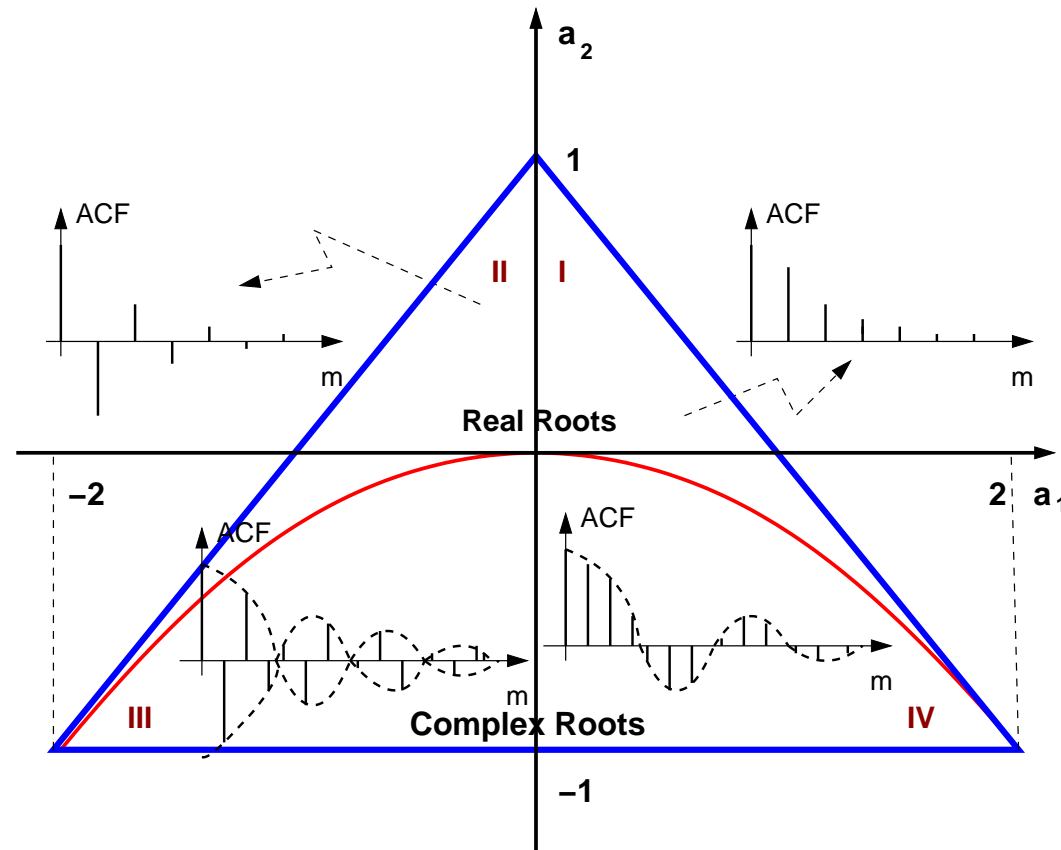
$$\text{Condition 1 : } a_1 + a_2 < 1$$

$$\text{Condition 2 : } a_2 - a_1 < 1$$

$$\text{Condition 3 : } -1 < a_2 < 1$$

This can be visualised within the so-called “stability triangle”

Stability triangle

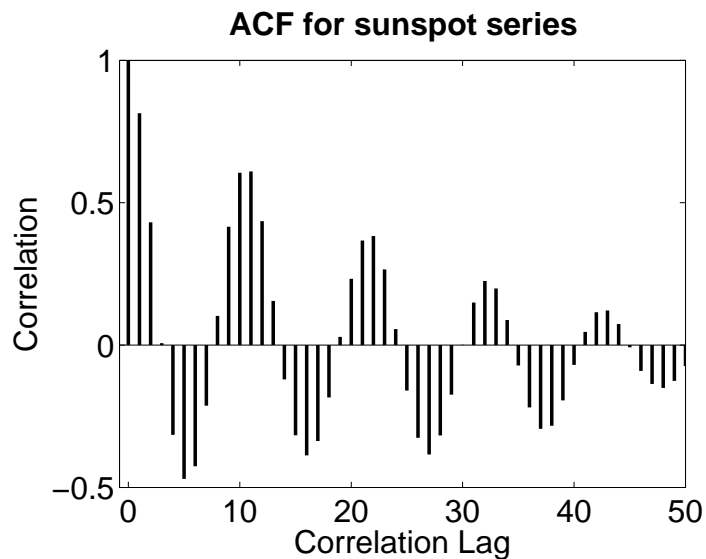
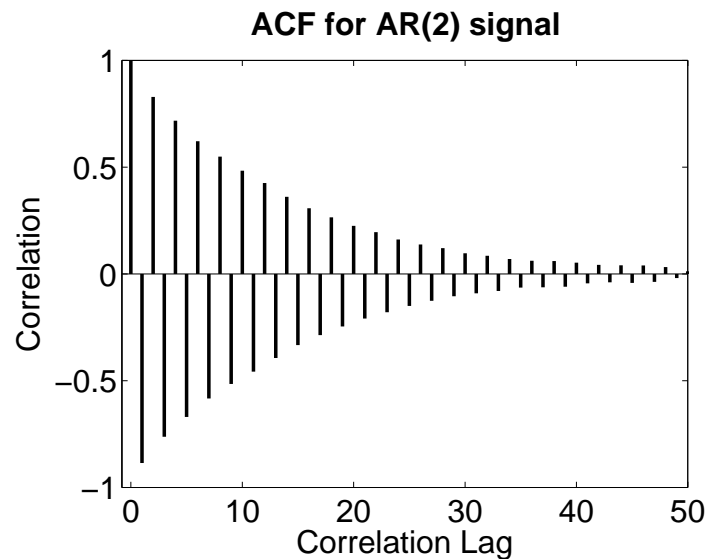
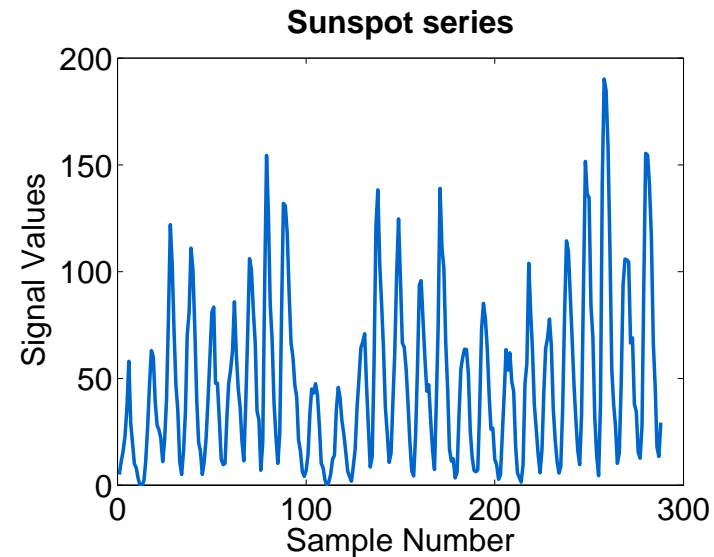
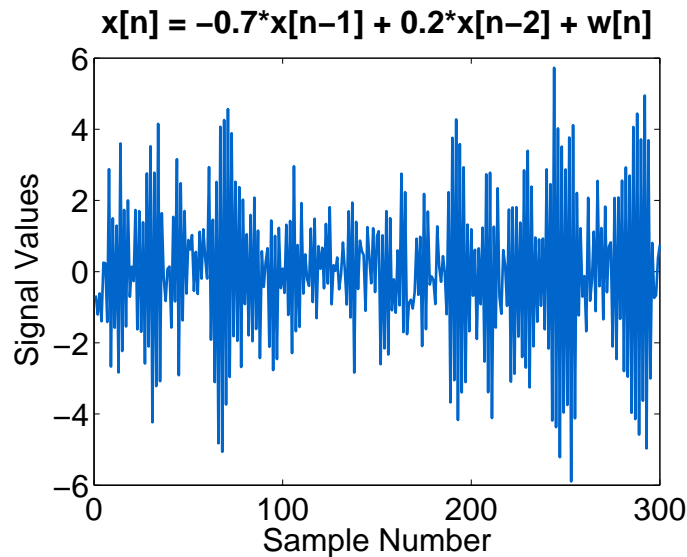


- i) **Real roots** Region 1: Monotonically decaying ACF
- ii) **Real roots** Region 2: Decaying oscillating ACF
- iii) **Complex roots** Region 3: Oscillating pseudo-periodic ACF
- iv) **Complex roots** Region 4: Pseudo-periodic ACF

Example 10: Stability triangle and ACFs of AR(2) signals

Left: $\mathbf{a} = [-0.7, 0.2]$ (region 3)

Right: $\mathbf{a} = [1.474, -0.586]$ (region 4)



Determining regions in the stability triangle

let us examine the autocorrelation function of AR(2) processes

The ACF

$$\rho_k = a_1\rho_{k-1} + a_2\rho_{k-2} \quad k > 0$$

- **Real roots:** $\Rightarrow (a_1^2 + 4a_2 > 0)$ ACF = mixture of damped exponentials
- **Complex roots:** $\Rightarrow (a_1^2 + 4a_2 < 0) \Rightarrow$ ACF exhibits a pseudo-periodic behaviour

$$\rho_k = \frac{D^k \sin(2\pi f_0 k + \Phi)}{\sin \Phi}$$

D - damping factor, of a sine wave with frequency f_0 and phase Φ .

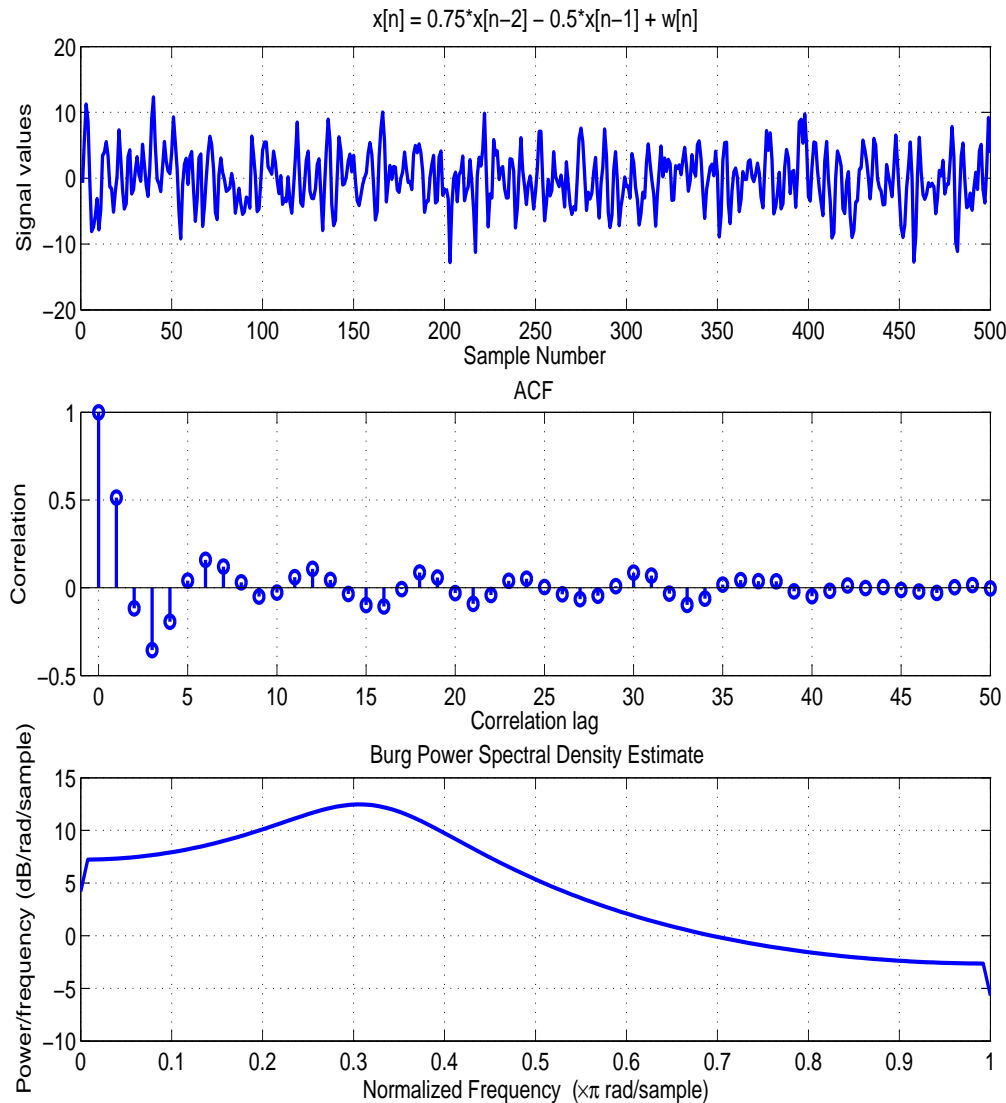
$$D = \sqrt{-a_2}$$

$$\cos(2\pi f_0) = \frac{a_1}{2\sqrt{-a_2}}$$

$$\tan(\Phi) = \frac{1 + D^2}{1 - D^2} \tan(2\pi f_0)$$

Example 11: AR(2) where $a_1 > 0$, $a_2 < 0 \quad \leadsto \quad$ Region 4

Consider: $x[n] = 0.75x[n-1] - 0.5x[n-2] + w[n]$



The damping factor

$$D = \sqrt{0.5} = 0.71,$$

Frequency

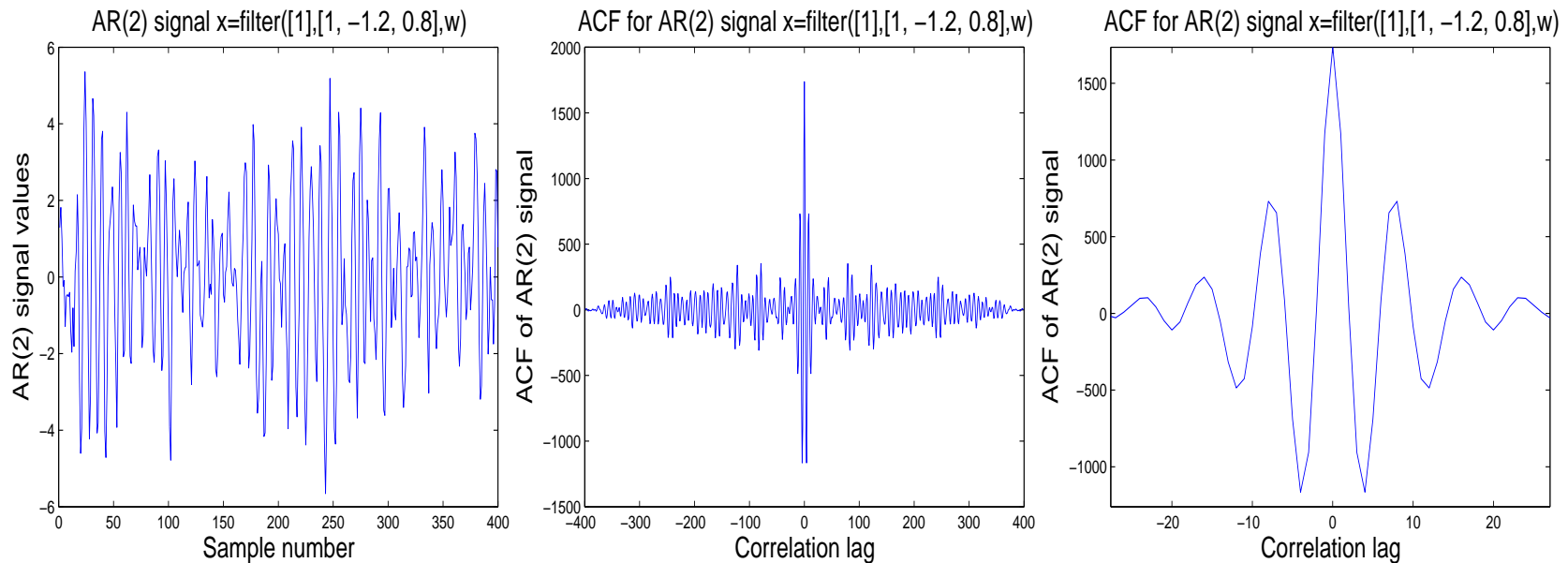
$$f_0 = \frac{\cos^{-1}(0.5303)}{2\pi} = \frac{1}{6.2}$$

The fundamental period of the autocorrelation function is therefore

$$T_0 = 6.2.$$

Model order selection: Partial autocorrelation function

Consider an earlier example using a slightly different notation for AR coefficients



To find p , first re-write AR coeffs. of order p as $[a_{p1}, \dots, a_{pp}]$

$$\mathbf{p} = \mathbf{1} \mapsto [0.6689] = a_{11} \quad \mathbf{p} = \mathbf{2} \mapsto [1.2046, -0.8008] = [a_{21}, a_{22}]$$

$$\mathbf{p} = \mathbf{3} \mapsto [1.1759, -0.7576, -0.0358] = [a_{31}, a_{32}, a_{33}]$$

$$\mathbf{p} = \mathbf{4} \mapsto [1.1762, -0.7513, -0.0456, 0.0083] = [a_{41}, a_{42}, a_{43}, a_{44}]$$

$$\mathbf{p} = \mathbf{5} \mapsto [1.1763, -0.7520, -0.0562, 0.0248, -0.0140] = [a_{51}, \dots, a_{55}]$$

$$\mathbf{p} = \mathbf{6} \mapsto [1.1762, -0.7518, -0.0565, 0.0198, -0.0062, -0.0067]$$

Partial autocorrelation function: Motivation

Notice: ACF of $AR(p)$ infinite in duration, **but** can be described in terms of p nonzero functions ACFs.

Denote by a_{kj} the j th coefficient in an autoregressive representation of order k , so that a_{kk} is the last coefficient. Then

$$\rho_j = a_{kj}\rho_{j-1} + \cdots + a_{k(k-1)}\rho_{j-k+1} + a_{kk}\rho_{j-k} \quad j = 1, 2, \dots, k$$

leading to the Yule–Walker equation, which can be written as

$$\begin{bmatrix} 1 & \rho_1 & \rho_2 & \cdots & \rho_{k-1} \\ \rho_1 & 1 & \rho_1 & \cdots & \rho_{k-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho_{k-1} & \rho_{k-2} & \rho_{k-3} & \cdots & 1 \end{bmatrix} \begin{bmatrix} a_{k1} \\ a_{k2} \\ \vdots \\ a_{kk} \end{bmatrix} = \begin{bmatrix} \rho_1 \\ \rho_2 \\ \vdots \\ \rho_k \end{bmatrix}$$

The only difference from the standard Y-W equations is the use of the symbols a_{ki} to denote the AR coefficient a_i \nrightarrow indicating the model order.

Partial ACF coefficients

Solving these equations for $k = 1, 2, \dots$ successively, we obtain

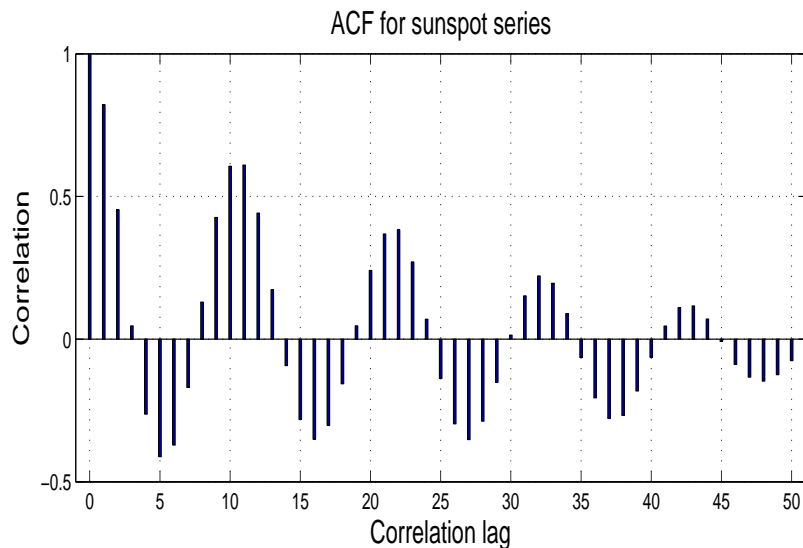
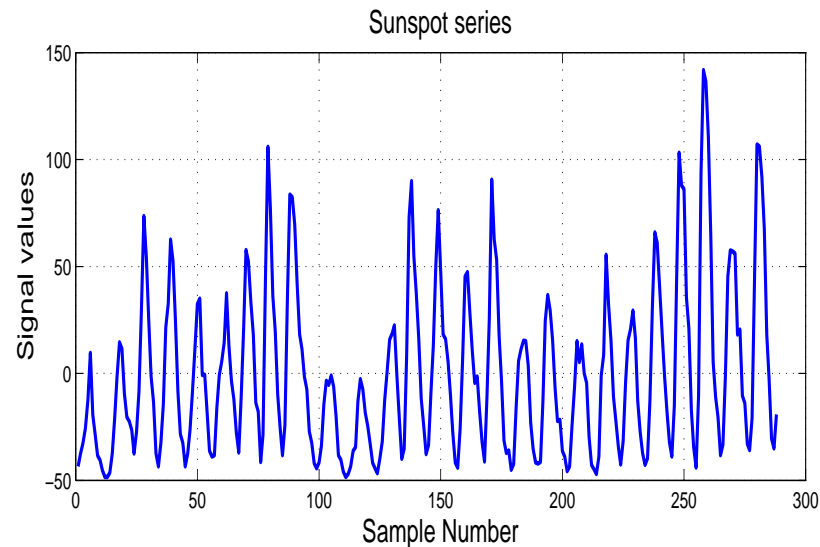
$$a_{11} = \rho_1, \quad a_{22} = \frac{\rho_2 - \rho_1^2}{1 - \rho_1^2}, \quad a_{33} = \frac{\begin{vmatrix} 1 & \rho_1 & \rho_1 \\ \rho_1 & 1 & \rho_2 \\ \rho_2 & \rho_1 & \rho_3 \end{vmatrix}}{\begin{vmatrix} 1 & \rho_1 & \rho_2 \\ \rho_1 & 1 & \rho_1 \\ \rho_2 & \rho_1 & 1 \end{vmatrix}}, \quad \text{etc}$$

- The quantity a_{kk} , regarded as a function of lag k , is called the **partial autocorrelation function**.
- For an AR(p) process, the PAC a_{kk} will be nonzero for $k \leq p$ and zero for $k > p \Rightarrow$ **indicates the order of an AR(p) process**.

In practice, we introduce a small threshold, as for real world data it is difficult to guarantee that $a_{kk} = 0$ for $k > p$. (see your coursework)

Example 12: Work by Yule \rightarrow model of sunspot numbers

Recorded for > 300 years. To study them in 1927 Yule invented the $AR(2)$ model



We first center the data, as we do not wish to model the DC offset (deterministic component), but the stochastic component (AR model driven by white noise)!

Using the Y-W equations we obtain:

$$\mathbf{a}_1 = [0.9295]$$

$$\mathbf{a}_2 = [1.4740, -0.5857]$$

$$\mathbf{a}_3 = [1.5492, -0.7750, 0.1284]$$

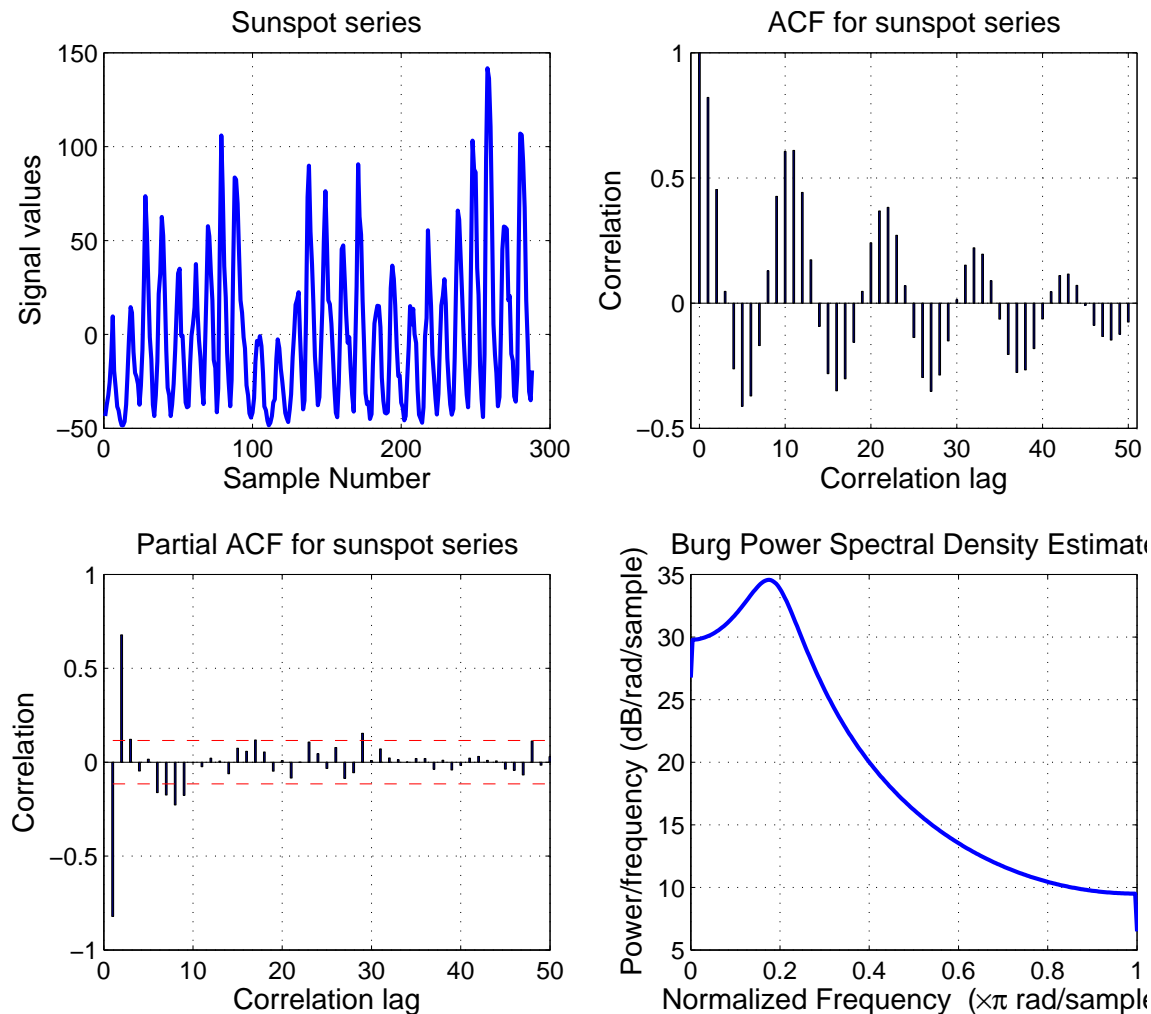
$$\mathbf{a}_4 = [1.5167, -0.5788, -0.2638, 0.2532]$$

$$\mathbf{a}_5 = [1.4773, -0.5377, -0.1739, 0.0174, 0.1555]$$

$$\mathbf{a}_6 = [1.4373, -0.5422, -0.1291, 0.1558, -0.2248, 0.2574]$$

Example 12 (contd.): Model order for sunspot numbers

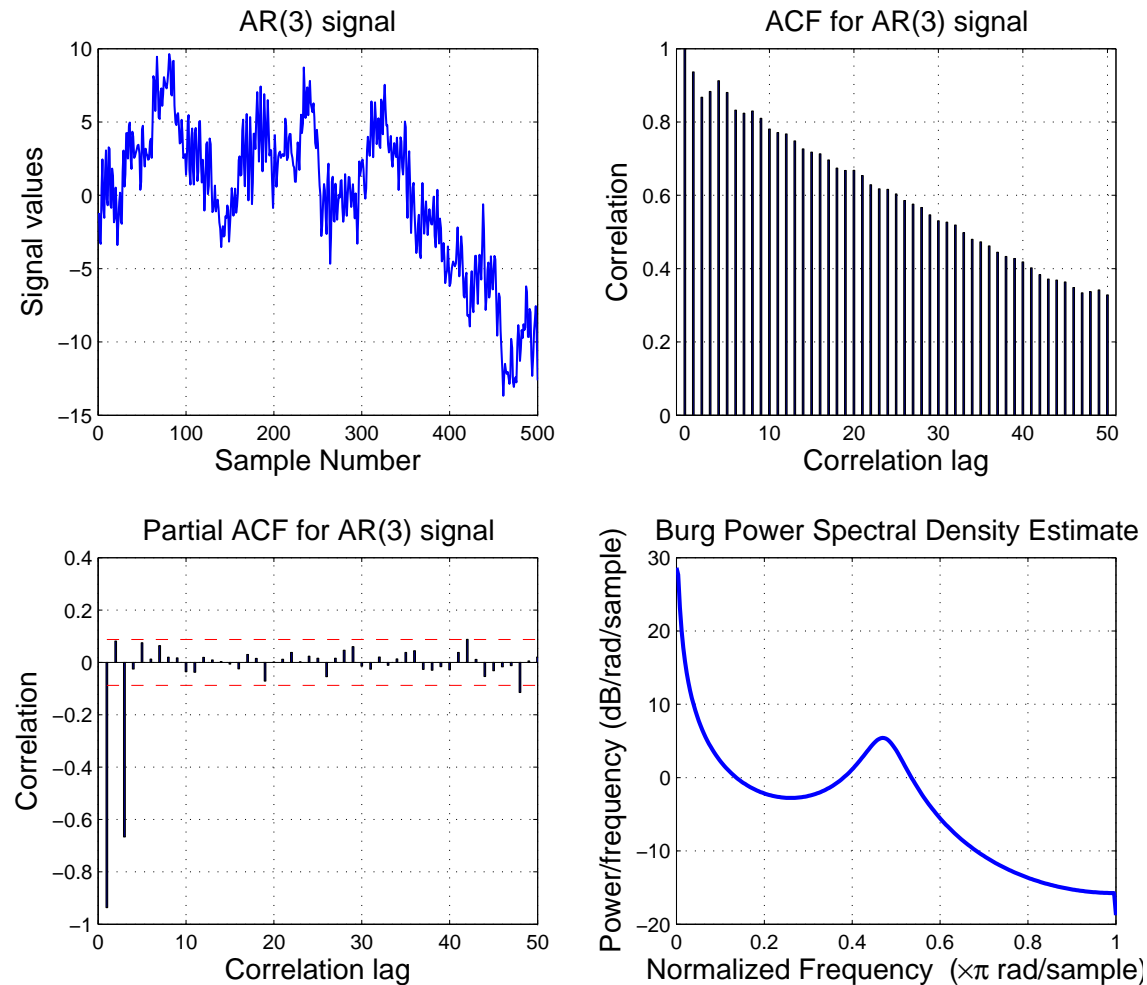
Sunspot numbers, their ACF and partial autocorrelation (PAC)



After lag $k = 2$, the PAC becomes very small

Example 13: Model order for an AR(3) process

AR(3) signal, its ACF, and partial autocorrelation (PAC)

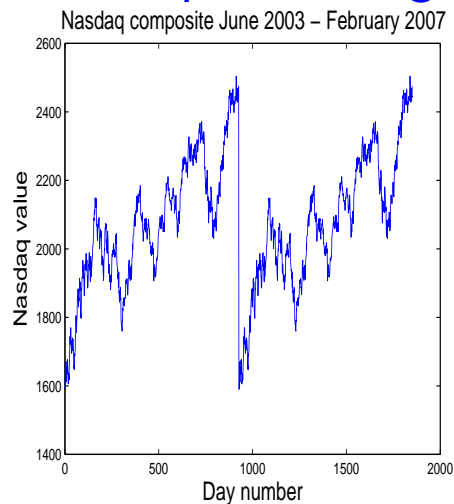


After lag $k = 3$, the PAC becomes very small

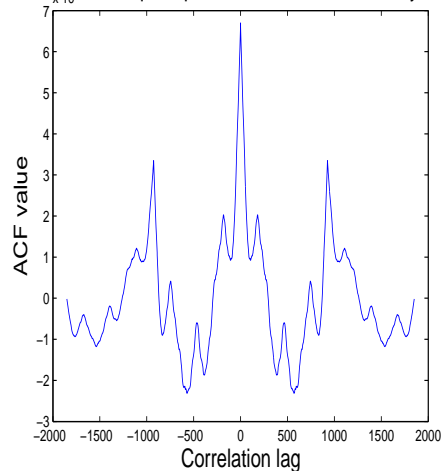
Example 14: Model order selection for a financial NASDAQ time series (the 'correct' and 'time-reversed' time series)

<http://finance.yahoo.com/q/ta?s=%5EIXIC&t=1d&l=on&z=m&q=b&p=v&a=&c=>

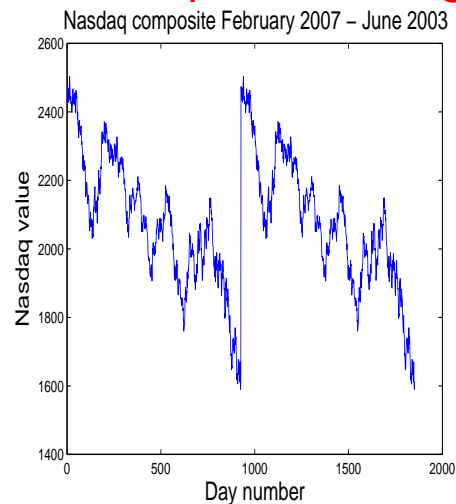
Nasdaq ascending



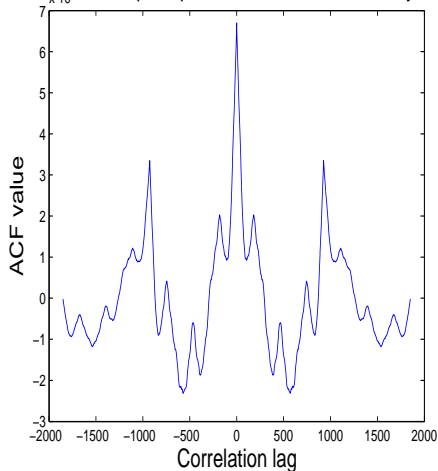
ACF of Nasdaq composite June 2003 – February 2007



Nasdaq descending



ACF of Nasdaq composite June 2003 – February 2007



Partial correlations:

$$\text{AR}(1): \mathbf{a} = [0.9994]$$

$$\text{AR}(2): \mathbf{a} = [.9982, .0011]$$

$$\text{AR}(3): \mathbf{a} = [.9982, -0.0086, .0097]$$

$$\text{AR}(4): \mathbf{a} = [.9983, -0.0086, .0128, -0.003]$$

$$\text{AR}(5): \mathbf{a} = [.9983, -0.0086, .0128, -0.0026, -0.0005]$$

$$\text{AR}(6): \mathbf{a} = [.9983, -0.0086, .0127, -0.0026, -0.0017, .0012]$$

AR model based prediction: Importance of model order

For a zero mean process $x[n]$, the best **linear predictor** in the **mean square error** sense of $x[n]$ based on $x[n-1], x[n-2], \dots$ is

$$\hat{x}[n] = a_{k-1,1}x[n-1] + a_{k-1,2}x[n-2] + \dots + a_{k-1,k-1}x[n-k+1]$$

(apply the $E\{\cdot\}$ operator to the general $AR(p)$ model expression, and recall that $E\{w[n]\} = 0$)

(Hint:

$$E\{x[n]\} = \hat{x}[n] = E\{a_{k-1,1}x[n-1] + \dots + a_{k-1,k-1}x[n-k+1] + w[n]\} = a_{k-1,1}x[n-1] + \dots + a_{k-1,k-1}x[n-k+1])$$

whether the process is an AR or not

In MATLAB, check the function:

ARYULE

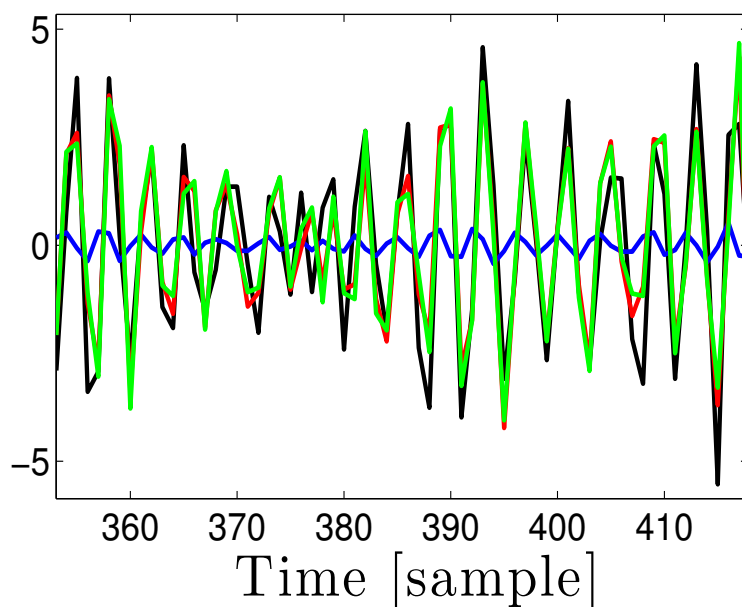
and functions

PYULEAR, ARMCOV, ARBURG, ARCOV, LPC, PRONY

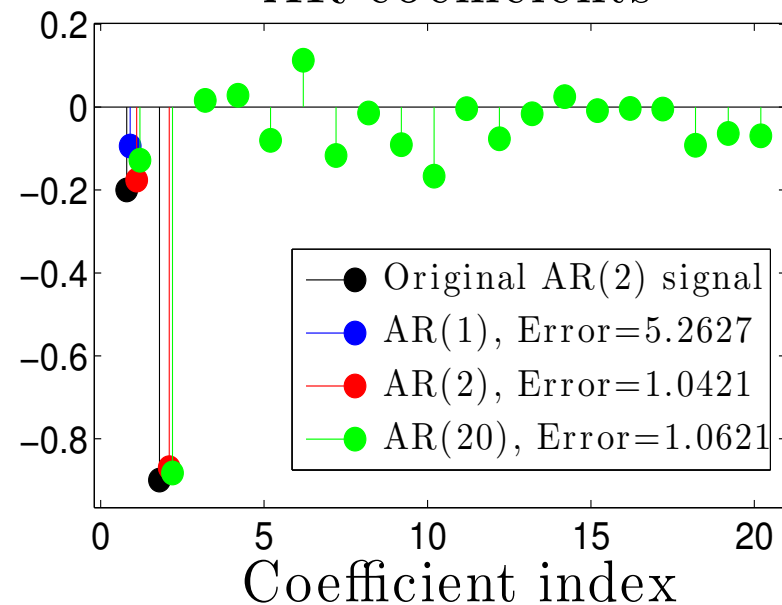
Estimation of the parameters of an AR(2) process using models of different orders (under- vs. over-fitting)

Original AR(2) process $x[n] = -0.2x[n-1] - 0.9x[n-2] + w[n]$, $w[n] \sim \mathcal{N}(0, 1)$, estimated using AR(1), AR(2) and AR(20) models:

Original and estimated signals



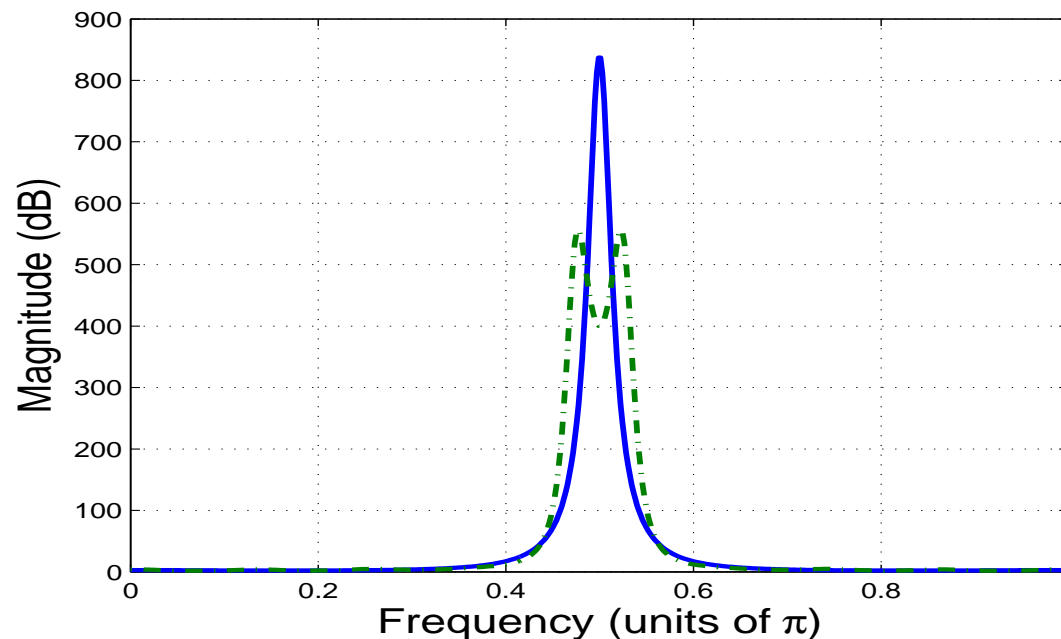
AR coefficients



The *higher order* coefficients of the AR(20) model are close to zero and therefore do not contribute significantly to the estimate, while the AR(1) does not have sufficient degrees of freedom.

Effects of over-modelling on autoregressive spectral estimation: Spectral line splitting

Consider an AR(2) signal $x(n) = -0.9x(n-2) + w(n)$ with $w \sim \mathcal{N}(0, 1)$. $N = 64$ data samples, model orders $p = 4$ (solid blue) and $p = 12$ (broken green).



Notice that this is an AR(2) model!

Although the true spectrum has a single spectral peak at $\omega = \pi/2$ (blue), when over-modelling using $p = 12$ this peak is split into two peaks (green).

Model order selection \leadsto practical issues

In practice: the greater the model order the greater accuracy & complexity

Q: When do we stop? What is the optimal model order?

Solution: To establish a trade-off between computational complexity and model accuracy, we introduce “penalty” for a high model order. Such criteria for model order selection are:

MDL: The minimum description length criterion (MDL) (by Rissanen),

AIC: The Akaike information criterion (AIC)

$$\text{MDL} \quad p_{opt} = \min_p \left[\log(E) + \frac{p * \log(N)}{N} \right]$$

$$\text{AIC} \quad p_{opt} = \min_p [\log(E) + 2p/N]$$

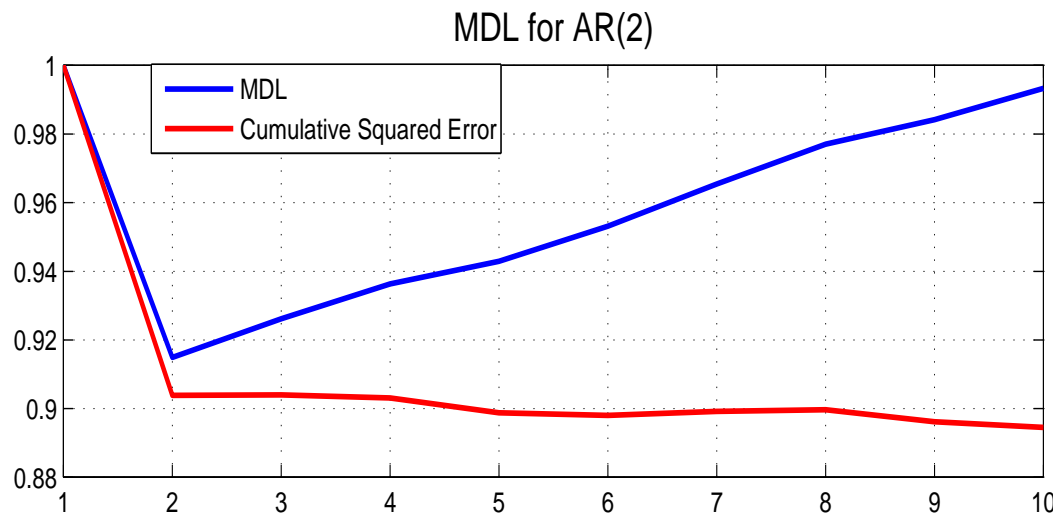
$E \rightsquigarrow$ the loss function (typically cumulative squared error,

$p \rightsquigarrow$ the number of estimated parameters (model order),

$N \rightsquigarrow$ the number of estimated data points.

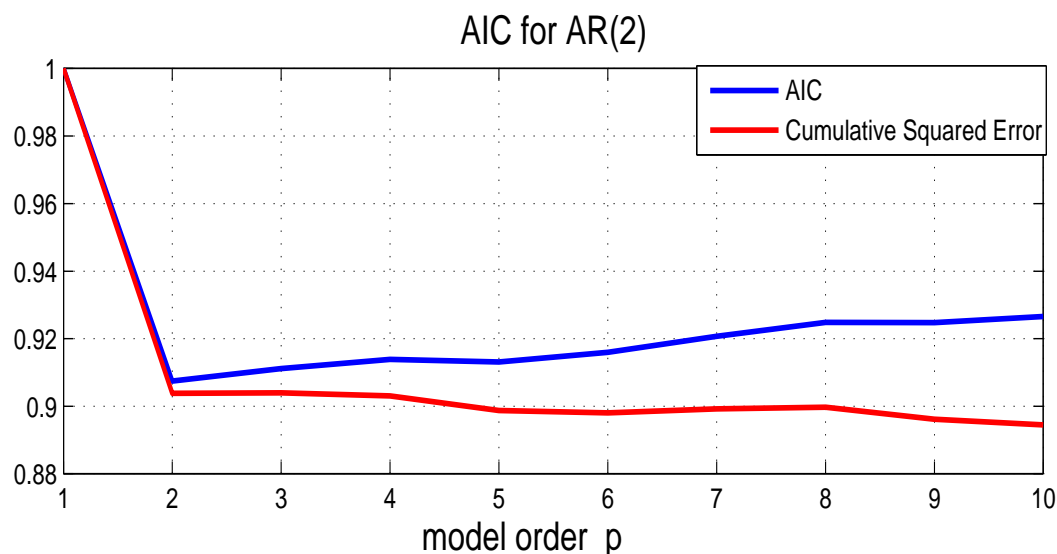
Example 15: Model order selection \leadsto MDL vs AIC

MDL and AIC criteria for an AR(2) model with $a_1 = 0.5$ $a_2 = -0.3$



The graphs on the left show the **(model error)²** (vertical axis) versus the **model order p** (horizontal axis). Notice that $p_{opt} = 2$.

The curves are **convex**, i.e. a monotonically decreasing **error²** with an increasing **penalty term** (MDL or AIC correction).



Hence, we have a **unique minimum at $p = 2$** , reflecting the correct model order (no over-modelling)

Moving average processes, MA(q)

A general MA(q) process is given by

$$x[n] = w[n] + b_1 w[n-1] + \cdots + b_q w[n-q]$$

Autocorrelation function: The autocovariance function of MA(q)

$$c_k = E[(w[n] + b_1 w[n-1] + \cdots + b_q w[n-q])(w[n-k] + b_1 w[n-k-1] + \cdots + b_q w[n-k-q])]$$

Hence the variance of the process

$$c_0 = (1 + b_1^2 + \cdots + b_q^2) \sigma_w^2$$

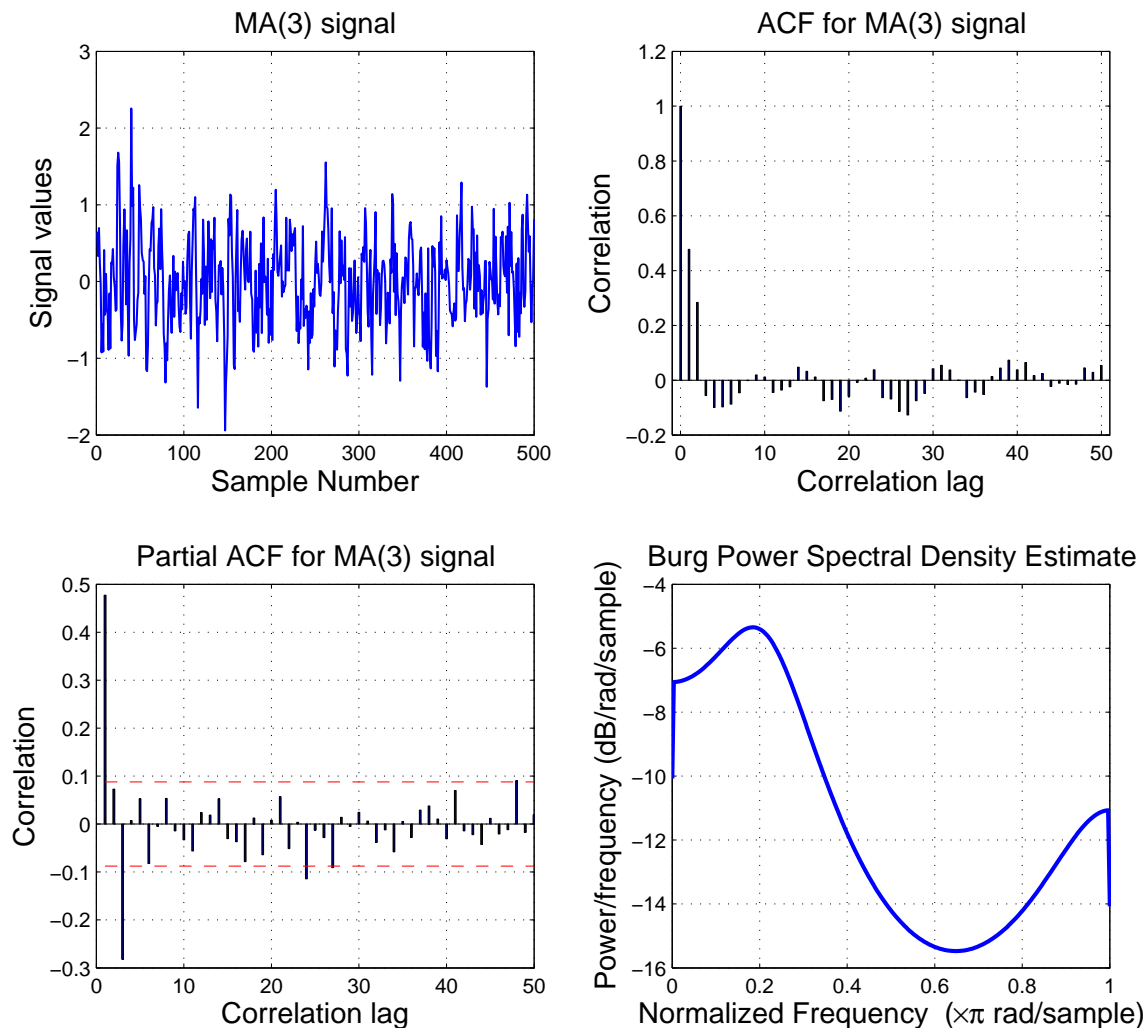
The ACF of an MA process has a cutoff after lag q .

Spectrum: All-zero transfer function \Rightarrow struggles to model 'peaky' PSDs

$$P(f) = 2\sigma_w^2 |1 + b_1 e^{-j2\pi f} + b_2 e^{-j4\pi f} + \cdots + b_q e^{-j2\pi qf}|^2$$

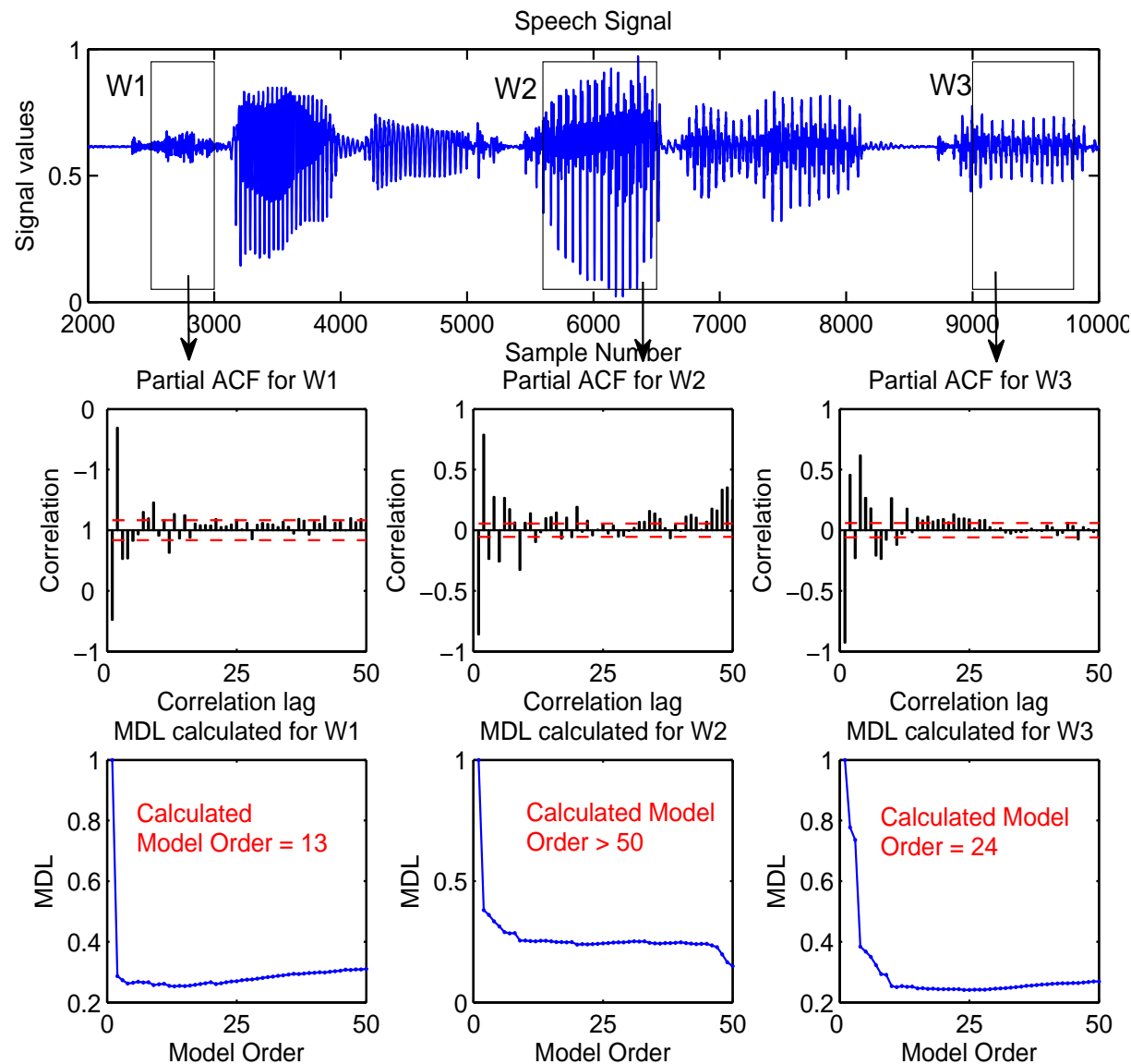
Example 16: Third order moving average MA(3) process

An MA(3) process and its autocorrel. (ACF) and partial autocorrel. (PAC) fns.



After lag $k = 3$, the ACF becomes very small

Analysis of nonstationary signals



○ Consider a real-world speech signal, and three different segments with different statistical properties

○ **Different AR model orders required for different segments of speech** → **opportunity for content analysis!**

○ To deal with nonstationarity we need short sliding data windows

Summary: AR and MA Processes

- i) A stationary finite AR(p) process can be represented as an infinite order MA process. A finite MA process can be represented as an infinite AR process.
- ii) The finite MA(q) process has an ACF that is zero beyond q . For an AR process, the ACF is infinite in length and consists of mixture of damped exponentials and/or damped sine waves.
- iii) Finite MA process are always stable, and there is no requirement on the coefficients of MA processes for stationarity. However, for invertibility, the roots of the characteristic equation must lie inside the unit circle.
- iv) AR processes produce spectra with sharp peaks (two poles of $A(z)$ per peak), whereas MA processes cannot produce peaky spectra.

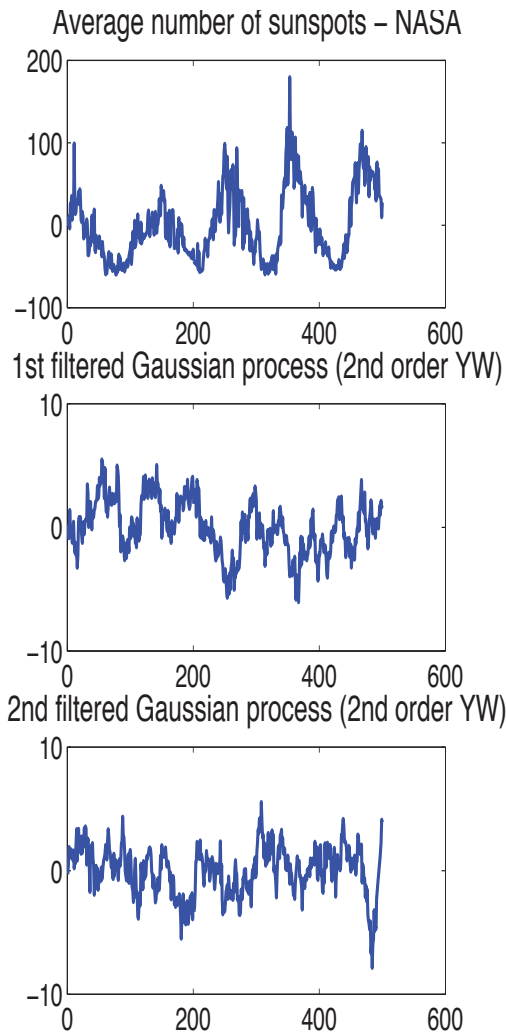
ARMA modelling is a classic technique which has found a tremendous number of applications

Appendix 1: More on numbers (recorded since 1874)

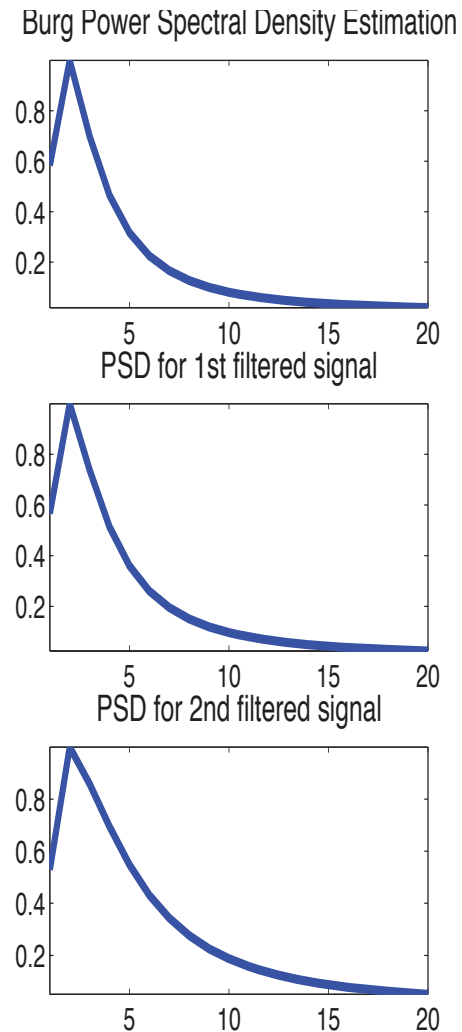
Top: original sunspots

Middle and Bottom: AR(2) representations

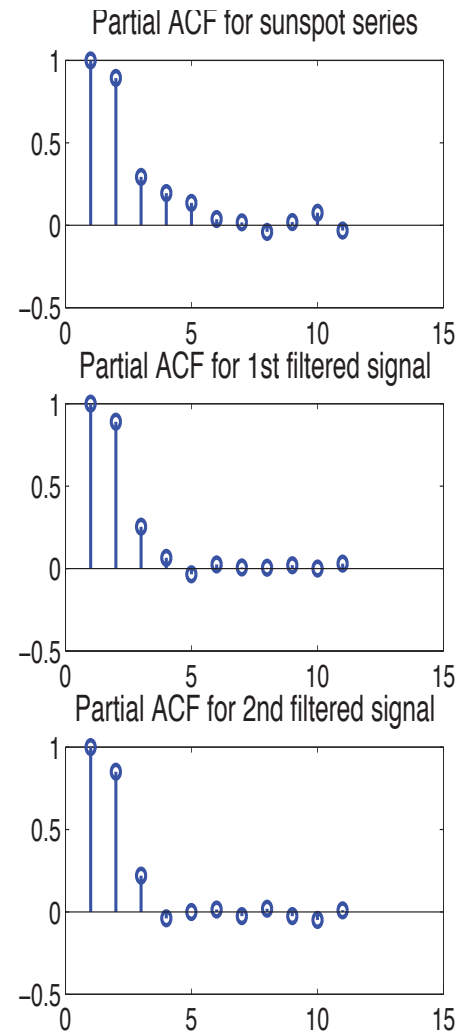
Left: time



Middle:spectrum



Right: autocorrelation



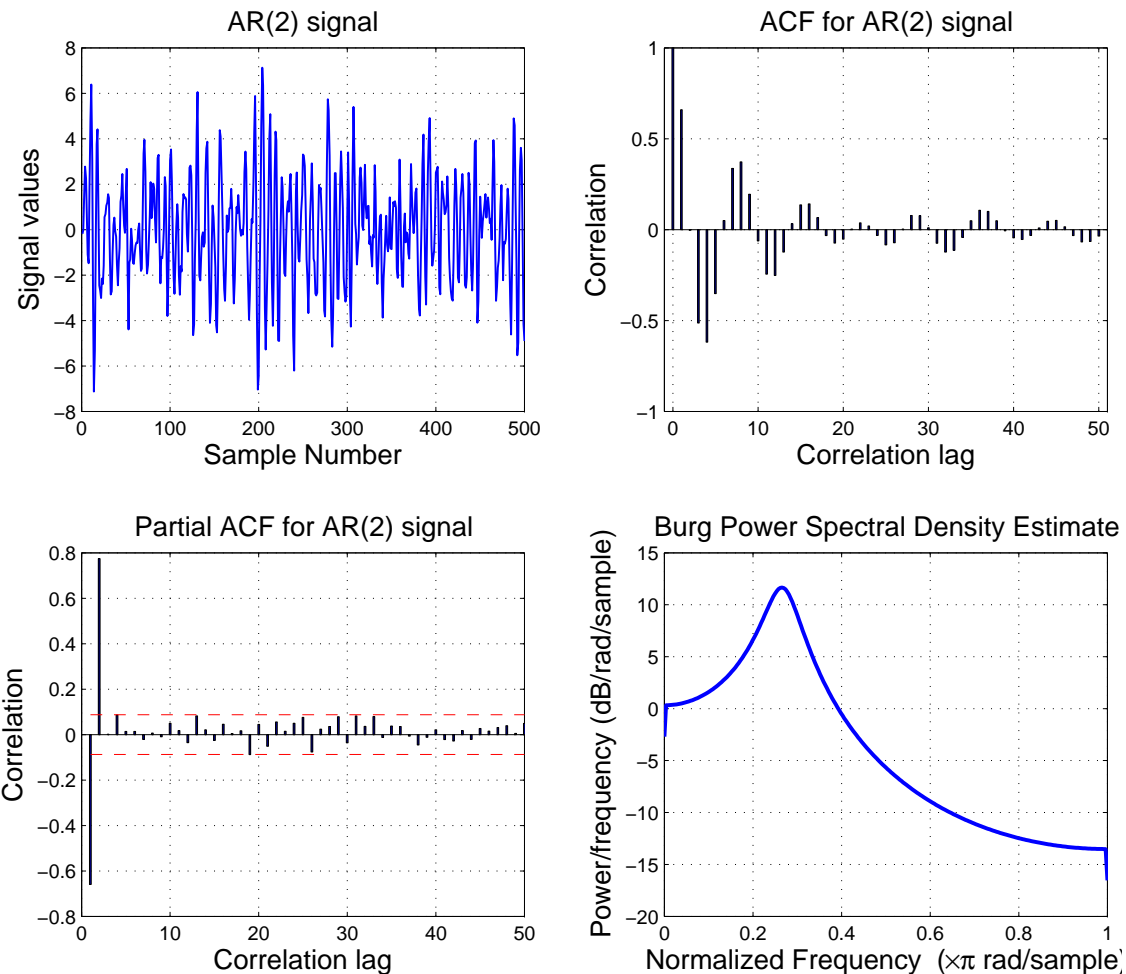
Top: original

Middle: first AR(2) model

Bottom: second AR(2) model

Appendix 2: Model order for an AR(2) process

An AR(2) signal, its ACF, and its partial autocorrelations (PAC)



After lag $k = 2$, the PAC becomes very small

Something to think about ...

- What would be the properties of a multivariate (MV) autoregressive, say $\text{MVAR}(1)$, process, where the quantities \mathbf{w} , \mathbf{x} , \mathbf{a} now become matrices, that is

$$\mathbf{X}(n) = \mathbf{A}\mathbf{X}(n-1) + \mathbf{W}(n)$$

- Would the inverse of the multichannel correlation matrix depend on 'how similar' the data channels are? Explain also in terms of eigenvalues and 'collinearity'.
- Threshold autoregressive (TAR) models allow for the mean of a time series to change along the blocks of data. What would be the advantages of such a model?
- Try to express an $\text{AR}(p)$ process as a state-space model. What kind of the transition matrix between the states do you observe?

Consider also: Fourier transform as a digital filter

1) Continuous FT. For a continuous FT $F(\omega) = \int_{-\infty}^{\infty} x(t)e^{-j\omega t} dt$

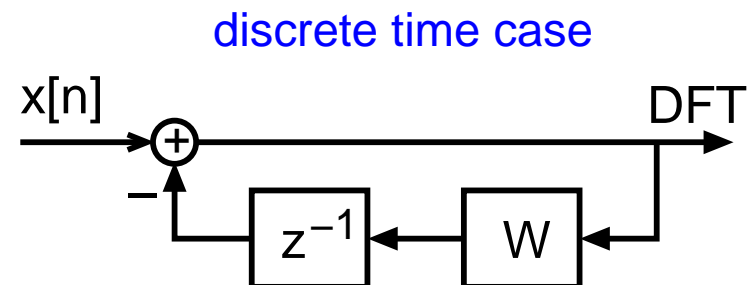
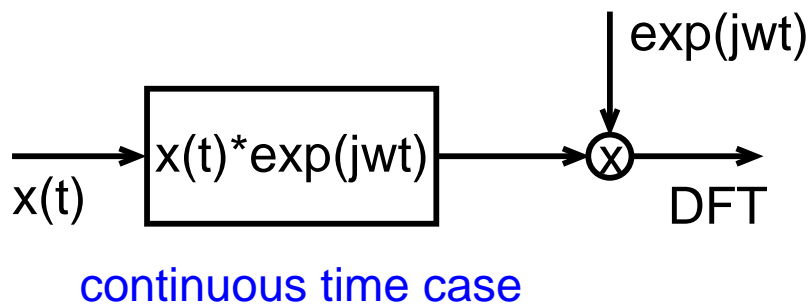
Let us now swap variables $t \rightarrow \tau$ and multiply by $e^{j\omega t}$, to give

$$e^{j\omega t} \int x(\tau)e^{-j\omega \tau} d\tau = \int x(\tau)e^{j\omega(t-\tau)} d\tau = x(t) * e^{j\omega t}$$

2) Discrete Fourier transform. For DFT, we have a filtering operation

$$X(k) = \sum_{n=0}^{N-1} x(n)e^{-j\frac{2\pi}{N}nk} = \underbrace{x(0) + W(x(1) + W(x(2) + \dots))}_{\text{cumulative add and multiply}} \quad W = e^{-j\frac{2\pi}{N}n}$$

with the transfer function $H(z) = \frac{1}{1 - z^{-1}W} = \frac{1 - z^{-1}W^*}{1 - 2\cos\theta_k z^{-1} + z^{-2}}$



Notes

Notes

Notes
