# Introduction to Spectral Analysis

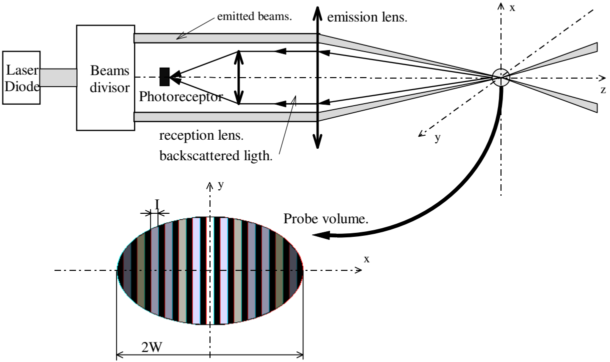
Olivier Besson

Institut Supérieur de l'Aéronautique et de l'Espace

## Some facts

- An ubiquitous problem, in many signal processing applications, is to recover some useful information from data in the time domain $\{x(n)\}_{n=0}^{N-1}$.
- Although time and frequency domains are dual (one goes between them using a Fourier transform), information is often more intuitively embedded in the spectral domain $\Rightarrow$ need for spectral analysis tools.
- In some cases (e.g., radar), the information itself consists of the frequencies of exponential signals.
- Spectral analysis can also serve as a pre-processing step to recognition and classification of signals, compression, filtering and detection.

# Laser anemometry




The received signal can be written as

$$x(t) = A \exp\left\{-2\alpha^2 f_d^2 t^2\right\} \cos(2\pi f_d t) + n(t)$$

with $f_d = v/I$ the information of most interest.

# Doppler effect

Assume a signal $s(t) = e^{i\omega_c t}$ is transmitted through an antenna and back-scattered by a moving target with radial velocity $v$. The received signal is given by

$$r(t) = As\left(t - 2\tau(t)\right) = As\left(t - 2\frac{d_0 - vt}{c}\right) = Ae^{i\omega_c t}e^{-i\omega_c \frac{2d_0}{c}}e^{i\frac{2\omega_c v}{c}t}.$$

After demodulation, one obtains

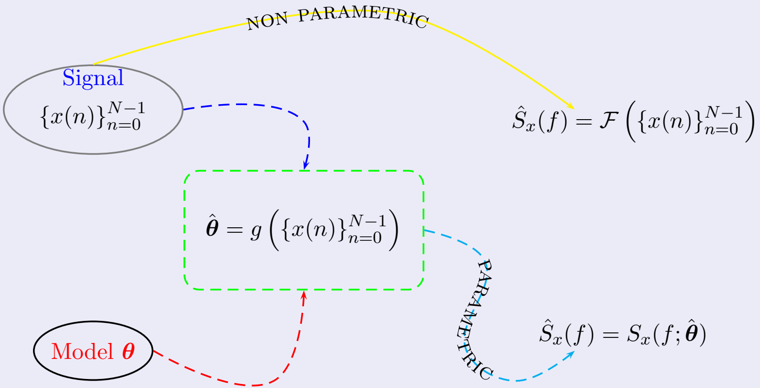
$$x(t) = Ae^{i\phi}e^{i2\pi\frac{2v}{\lambda}t} + n(t)$$

and hence the target velocity is directly related to the frequency of the useful signal.

## Problem statement

From the observation of $x(n)$, $n = 0, \cdots, N-1$, retrieve pertinent information about its spectral content.

## Parametric and non-parametric approaches



# Outline

## Power Spectral Density

Let $x(n)$ denote a 2nd-order ergodic and stationary process, with correlation function

$$r_{xx}(m) = \mathcal{E}\left\{x^*(n)x(n+m)\right\} = r^*_{xx}(-m).$$

The Power Spectral Density (PSD) can be defined in **2** different ways:

$$\begin{aligned} S_x(f) &= \sum_{m=-\infty}^{\infty} r_{xx}(m) e^{-i2\pi m f} \\ &= \lim_{N\to\infty} \mathcal{E}\left\{ \frac{1}{N} \left| \sum_{n=0}^{N-1} x(n) e^{-i2\pi n f} \right|^2 \right\}. \end{aligned}$$

# Principle

## From the theoretical PSD to its estimation

$$S_x(f) = \lim_{N\to\infty} \mathcal{E}\left\{ \frac{1}{N} \left| \sum_{n=0}^{N-1} x(n) e^{-i2\pi n f} \right|^2 \right\}$$

$$\downarrow$$

$$\hat{S}_p(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x(n) e^{-i2\pi n f} \right|^2 .$$

## Remark

The periodogram does not rely on any a priori information about the signal (hence it is robust) and can be computed efficiently using a fast Fourier transform (FFT).

# Performance

## Mean value

$$\mathcal{E}\left\{\hat{S}_p(f)\right\} = \int_{-1/2}^{1/2} W_B(f-u) S_x(u)\, du$$
$$\xrightarrow[N\to\infty]{} S_x(f)$$

with $W_B(f) = \frac{1}{N}\left[\frac{\sin(\pi N f)}{\sin(\pi f)}\right]^2$.

- Smearing of the main lobe $\propto \frac{0.9}{N}$
- Sidelobe levels $(-13\text{dB})$.

## Variance

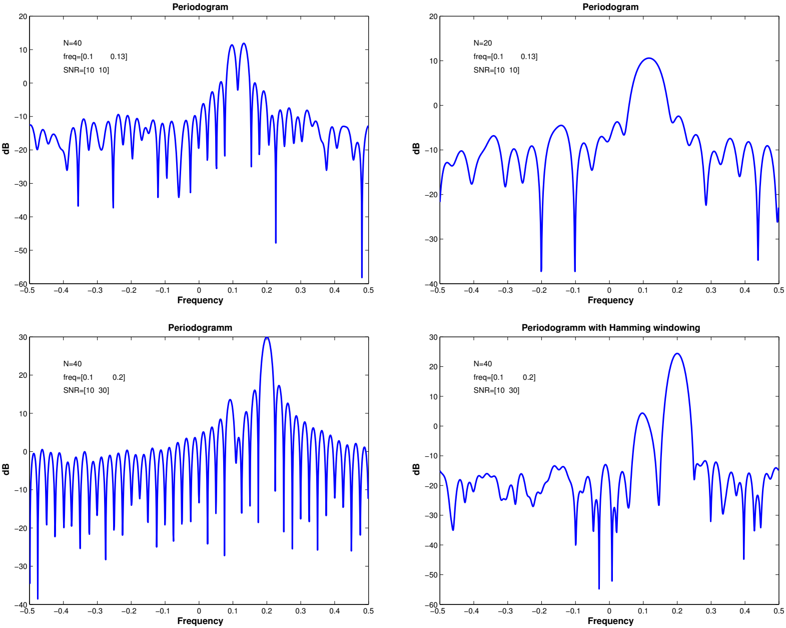
$$\text{var}\left\{\hat{S}_p(f)\right\} \simeq S_x(f)^2 \underset{N\to\infty}{\not\to} 0.$$

# Variations

- In order to decrease variance, one can compute several periodograms on shorter time intervals, and then average them: variance is decreased but resolution is poorer.
- Windows can be used, i.e.,

$$\hat{S}_{p-w}(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} w_n x(n) e^{-i2\pi n f} \right|^2$$

where $w_n$ is selected, e.g., to have lower sidelobe levels (at the price of a larger mainlobe).


Periodogram
N=40
freq=[0.1 0.13]
SNR=[10 10]
dB
Frequency
Periodogram
N=20
freq=[0.1 0.13]
SNR=[10 10]
dB
Frequency
Periodogramm
N=40
freq=[0.1 0.2]
SNR=[10 30]
dB
Frequency
Periodogramm with Hamming windowing
N=40
freq=[0.1 0.2]
SNR=[10 30]
dB
Frequency

## Periodogram-Correlogram

The periodogram can be rewritten as

$$\hat{S}_c(f) = \sum_{m=-(N-1)}^{N-1} \hat{r}_{xx}(m) e^{-i2\pi m f}$$

where $\hat{r}_{xx}(m) = N^{-1} \sum_{n=0}^{N-1-m} x^*(n)x(n+m)$ is a biased estimate of the correlation function. The variance of $\hat{S}_p(f)$ is due to a poor estimate $\hat{r}_{xx}(m)$ for large $m$.

## Remark

If the unbiased estimate $\hat{r}_{xx}(m) = (N-m)^{-1} \sum_{n=0}^{N-1-m} x^*(n)x(n+m)$ of $r_{xx}(m)$ is used in $\hat{S}_c(f)$, this may result in a non positive estimated PSD.

## Principle

$$S_x(f) = \sum_{m=-\infty}^{\infty} r_{xx}(m) e^{-i2\pi m f}$$

$$\downarrow$$

$$\hat{S}_{BT}(f) = \sum_{m=-M}^{M} w_m \hat{r}_{xx}(m) e^{-i2\pi m f}$$

where $\hat{r}_{xx}(m)$ is the biased estimate of the correlation function.

## Observations

One has

$$\hat{S}_{BT}(f) = \int_{-1/2}^{1/2} W(f-u)\hat{S}_p(u)\, du.$$

Use of a window $w_m$, $m = -M, \cdots, M$ enables one to achieve a good tradeoff between bias and variance: decreasing $M$ lowers variance (but increases bias and penalizes resolution).

# Usual windows and their characteristics

For each window $w(m)$ defined on $[-M, M]$, the table below gives the $-3$dB width of the mainlobe (in fraction of $N = 2M$) and the level of the first sidelobe compared to that of the main lobe.

| Window | Characteristics | $\frac{\text{amp. sidelobe}}{\text{amp. main lobe}}$ | $\Delta B_{3dB}$ |
|---|---|---|---|
| Rectangular | $w(m) = 1$ | -13dB | 0.89 |
| Bartlett | $w(m) = 1 - \frac{\lvert m \rvert}{M}$ | -26dB | 1.27 |
| Hanning | $w(m) = 0.5 + 0.5\cos(\pi\frac{m}{M})$ | -31.5dB | 1.41 |
| Hamming | $w(m) = 0.54 + 0.46\cos(\pi\frac{m}{M})$ | -42dB | 1.31 |
| Blackman | $w(m) = 0.42 + 0.5\cos(2\pi\frac{m}{M})$ $+0.08\cos(4\pi\frac{m}{M})$ | -58dB | 1.66 |

# Performances

## Mean value

$$\mathcal{E}\left\{\hat{S}_{BT}(f)\right\} = \int_{-1/2}^{1/2} W(f-u)\mathcal{E}\left\{\hat{S}_p(u)\right\} du$$
$$\simeq \int_{-1/2}^{1/2} W(f-u)S_x(u)\, du.$$

## Variance

The variance of the Blackman-Tuckey is given by

$$\text{var}\left\{\hat{S}_{BT}(f)\right\} \simeq \frac{S_x(f)^2}{N} \sum_{m=-M}^{M} w_m^2.$$

## Properties of Fourier-based methods

- Robust methods which require very few assumptions about the signal, hence applicable to a very large class of signals.
- Good performance, even at low signal to noise ratio.
- Simple and computationally effective algorithms (FFT).
- Estimated PSD proportional to actual signal power.
- Resolution is about $1/N \Longrightarrow$ problem to resolve two closely spaced spectral lines with short samples.
- Problem to recover weak signals in the presence of strong signals.

# Interpretation of the periodogram

- The periodogram can be interpreted as an estimate of **the power at the output of a filter tuned to $f$**.
- Assume that, for a given $f$, we wish to design a filter $\boldsymbol{w}(f) = \begin{bmatrix} w_0(f) & \cdots & w_{N-1}(f) \end{bmatrix}^T$ whose output

$$X(f) = \boldsymbol{w}^H(f)\boldsymbol{x} = \sum_{n=0}^{N-1} w_n^*(f)x(n)$$

  provides information about the signal power at frequency $f$.
- If the input signal is $x(n) = Ae^{i2\pi nf} + n(n)$, where $n(n)$ denotes white noise with power $\sigma^2$, the output is given by

$$X(f) = A\boldsymbol{w}^H(f)\boldsymbol{e}(f) + \boldsymbol{w}^H(f)\boldsymbol{n}$$

  with $\boldsymbol{e}(f) = \begin{bmatrix} 1 & e^{i2\pi f} & \cdots e^{i2\pi(N-1)f} \end{bmatrix}^T$.

- One looks for a filter that lets $\boldsymbol{e}(f)$ pass undistorted, i.e. $\boldsymbol{w}^H(f)\boldsymbol{e}(f) = 1$, while maximizing the output signal to noise ratio:

$$SNR = \frac{|A|^2 \left|\boldsymbol{w}^H(f)\boldsymbol{e}(f)\right|^2}{\mathcal{E}\left\{\left|\boldsymbol{w}^H(f)\boldsymbol{n}\right|^2\right\}} = \frac{|A|^2}{\sigma^2}\frac{\left|\boldsymbol{w}^H(f)\boldsymbol{e}(f)\right|^2}{\boldsymbol{w}^H(f)\boldsymbol{w}(f)}$$
$$\leq N\frac{|A|^2}{\sigma^2}$$

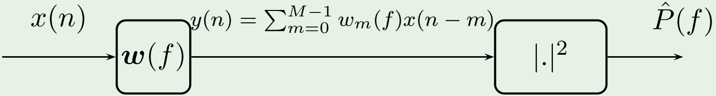
  with equality iif $\boldsymbol{w}(f) \propto \boldsymbol{e}(f)$. Since $\boldsymbol{w}^H(f)\boldsymbol{e}(f) = 1$ one finally gets $\boldsymbol{w}(f) = N^{-1}\boldsymbol{e}(f)$. The output power is thus

$$|X(f)|^2 = \frac{\left|\boldsymbol{e}^H(f)\boldsymbol{x}\right|^2}{N^2} = \frac{1}{N^2}\left|\sum_{n=0}^{N-1} x(n)e^{-i2\pi nf}\right|^2$$

  which coincides (up to a scaling factor) with the periodogram.
- The periodogram can be interpreted as **matched filter in white noise**.

## Principle (Capon)

For every frequency $f$, design a filter, tuned to $f$, which **eliminates all other spectral components contained in the signal**, and then compute output power:




## Problem formulation

$$\min_{\boldsymbol{w}(f)} \mathcal{E}\left\{ |y(n)|^2 \right\} \text{ subject to } \sum_{m=0}^{M-1} w_m(f) e^{-i2\pi m f} = 1$$

with $\boldsymbol{w}(f) = \begin{bmatrix} w_0(f) & \cdots & w_{M-1}(f) \end{bmatrix}^T$.

## Capon's minimization problem

Since $\mathcal{E}\left\{|y(n)|^2\right\} = \boldsymbol{w}^H(f)\boldsymbol{R}\boldsymbol{w}(f)$ with

$$\boldsymbol{R} = \begin{pmatrix} r_{xx}(0) & r_{xx}(-1) & \cdots & r_{xx}(-M+1) \\ r_{xx}(1) & r_{xx}(0) & \cdots & r_{xx}(-M+2) \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}(M-1) & r_{xx}(M-2) & \cdots & r_{xx}(0) \end{pmatrix}$$

one must solve

$$\min_{\boldsymbol{w}(f)} \boldsymbol{w}^H(f)\boldsymbol{R}\boldsymbol{w}(f) \text{ subject to } \boldsymbol{w}^H(f)\boldsymbol{e}(f) = 1$$

where $\boldsymbol{e}(f) = \begin{bmatrix} 1 & e^{i2\pi f} & \cdots e^{i2\pi(M-1)f} \end{bmatrix}^T$.

## Capons's solution (theoretical)

For any vector $\boldsymbol{w}(f)$ such that $\boldsymbol{w}^H(f)\boldsymbol{e}(f) = 1$, one has

$$\begin{aligned} 1 = \left|\boldsymbol{w}^H(f)\boldsymbol{e}(f)\right|^2 &= \left|\boldsymbol{w}^H(f)\boldsymbol{R}^{1/2}\boldsymbol{R}^{-1/2}\boldsymbol{e}(f)\right|^2 \\ &\leq \left[\boldsymbol{w}^H(f)\boldsymbol{R}\boldsymbol{w}(f)\right]\left[\boldsymbol{e}^H(f)\boldsymbol{R}^{-1}\boldsymbol{e}(f)\right] \end{aligned}$$

with equality if and only if $\boldsymbol{R}^{1/2}\boldsymbol{w}(f)$ and $\boldsymbol{R}^{-1/2}\boldsymbol{e}(f)$ are co-linear. The (minimal) output power becomes

$$\boxed{P_{\text{Capon}}(f) = \frac{1}{\boldsymbol{e}^H(f)\boldsymbol{R}^{-1}\boldsymbol{e}(f)}.}$$

# Implementation Capon

- In practice, implementation is based on an array processing model. More precisely, for every $f$, we let

$$\boldsymbol{x}(n) = \begin{bmatrix} x(n) \\ x(n+1) \\ \vdots \\ x(n+M-1) \end{bmatrix} = A(f)e^{i2\pi nf}\boldsymbol{e}(f) + \boldsymbol{n}(n).$$

  The objective is to estimate $A(f)$, which corresponds to the amplitude of the signal component at frequency $f$.

- One minimizes $\boldsymbol{w}^H(f)\hat{\boldsymbol{R}}\boldsymbol{w}(f)$ under the constraint that $\boldsymbol{w}^H(f)\boldsymbol{e}(f) = 1$ with

$$\hat{\boldsymbol{R}} = \frac{1}{N-M+1}\sum_{n=0}^{N-M} \boldsymbol{x}(n)\boldsymbol{x}^H(n).$$

# Implementation Capon

- $\boldsymbol{w}(f)$ is given by
$$\boldsymbol{w}(f) = \frac{\hat{\boldsymbol{R}}^{-1}\boldsymbol{e}(f)}{\boldsymbol{e}^H(f)\hat{\boldsymbol{R}}^{-1}\boldsymbol{e}(f)}.$$
- For each snapshot, we have $\boldsymbol{w}^H(f)\boldsymbol{x}(n) \simeq A(f)e^{i2\pi nf}$ and $A(f)$ is estimated by a coherent summation of the outputs $\boldsymbol{w}^H(f)\boldsymbol{x}(n)$, i.e.,
$$\hat{A}(f) = \frac{1}{N-M+1}\sum_{n=0}^{N-M} \boldsymbol{w}^H(f)\boldsymbol{x}(n)e^{-i2\pi nf} = \boldsymbol{w}^H(f)\boldsymbol{r}(f)$$
with $\boldsymbol{r}(f) = \frac{1}{N-M+1}\sum_{n=0}^{N-M}\boldsymbol{x}(n)e^{-i2\pi nf}$.

# Implementation Capon

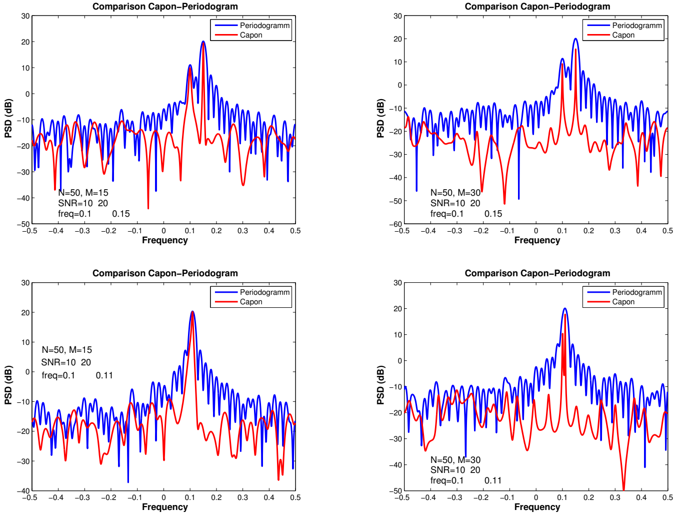
- In order to improve estimation (in particular that of $\boldsymbol{R}$), one might consider the snapshot

$$\boldsymbol{x}_b(n) = \begin{bmatrix} x^*(n+M-1) & x^*(n+M-2) & \cdots & x^*(n) \end{bmatrix}^T$$

whose correlation matrix is $\boldsymbol{R}$. The latter can therefore be estimated as

$$\hat{\boldsymbol{R}} = \frac{1}{2(N-M+1)} \sum_{n=0}^{N-M} \left[ \boldsymbol{x}(n)\boldsymbol{x}^H(n) + \boldsymbol{x}_b(n)\boldsymbol{x}_b^H(n) \right].$$

- Capon's method offers an improved resolution compared to the periodogram, at least for sufficiently large $M$.


Comparison Capon−Periodogram
Periodogramm
Capon
PSD (dB)
Frequency
N=50, M=15
SNR=10 20
freq=0.1 0.15
N=50, M=30
SNR=10 20
freq=0.1 0.15
N=50, M=15
SNR=10 20
freq=0.1 0.11
N=50, M=30
SNR=10 20
freq=0.1 0.11

# Amplitude and phase estimation (APES)

## Principle

Same approach as Capon: for every $f$, one looks for a filter $\boldsymbol{w}(f)$ which lets $\boldsymbol{e}(f)$ pass and such that the output is as close as possible to $\beta e^{i2\pi nf}$. The value of $\beta$ provides the signal amplitude at frequency $f$.

## Problem formulation

Let $\boldsymbol{x}(n) = \begin{bmatrix} x(n) & x(n+1) & \cdots & x(n+M-1) \end{bmatrix}^T$. One needs to solve

$$\min_{\boldsymbol{w}(f),\beta} \frac{1}{N-M+1} \sum_{n=0}^{N-M} \left| \boldsymbol{w}^H(f)\boldsymbol{x}(n) - \beta e^{i2\pi nf} \right|^2 \ / \ \boldsymbol{w}^H(f)\boldsymbol{e}(f) = 1.$$

## Minimization with respect to $\beta$

- Observe that

$$
\begin{aligned}
J &= \frac{1}{N-M+1} \sum_{n=0}^{N-M} \left| \boldsymbol{w}^H(f)\boldsymbol{x}(n) - \beta e^{i2\pi n f} \right|^2 \\
&= \boldsymbol{w}^H(f)\hat{\boldsymbol{R}}\boldsymbol{w}(f) - \beta \boldsymbol{r}^H(f)\boldsymbol{w}(f) - \beta^* \boldsymbol{w}^H(f)\boldsymbol{r}(f) + |\beta|^2 \\
&= \left|\beta - \boldsymbol{w}^H(f)\boldsymbol{r}(f)\right|^2 + \boldsymbol{w}^H(f)\left(\hat{\boldsymbol{R}} - \boldsymbol{r}(f)\boldsymbol{r}^H(f)\right)\boldsymbol{w}(f).
\end{aligned}
$$

- The solution for $\beta$ is $\beta = \boldsymbol{w}^H(f)\boldsymbol{r}(f)$ and it remains to solve

$$
\min_{\boldsymbol{w}(f)} \boldsymbol{w}^H(f)\left(\hat{\boldsymbol{R}} - \boldsymbol{r}(f)\boldsymbol{r}^H(f)\right)\boldsymbol{w}(f) \text{ subject to } \boldsymbol{w}^H(f)\boldsymbol{e}(f) = 1.
$$

## APES filter

The weight vector $\boldsymbol{w}(f)$ is hence given by

$$\boldsymbol{w}(f) = \frac{\left(\hat{\boldsymbol{R}} - \boldsymbol{r}(f)\boldsymbol{r}^H(f)\right)^{-1}\boldsymbol{e}(f)}{\boldsymbol{e}^H(f)\left(\hat{\boldsymbol{R}} - \boldsymbol{r}(f)\boldsymbol{r}^H(f)\right)^{-1}\boldsymbol{e}(f)}.$$

## APES amplitude

After some straightforward calculations, one finally gets

$$\beta(f) = \frac{\boldsymbol{e}^H(f)\hat{\boldsymbol{R}}^{-1}\boldsymbol{r}(f)}{\left(1 - \boldsymbol{r}^H(f)\hat{\boldsymbol{R}}^{-1}\boldsymbol{r}(f)\right)\boldsymbol{e}^H(f)\hat{\boldsymbol{R}}^{-1}\boldsymbol{e}(f) + \left|\boldsymbol{e}^H(f)\hat{\boldsymbol{R}}^{-1}\boldsymbol{r}(f)\right|^2}.$$

## Observation

APES has a lower resolution than Capon but provides more accurate estimates of the amplitude of complex exponentials.

## Modeling

The signal is modeled as **the output of a linear filter with rational transfer function**, whose input is a white noise:

$$u(n) \longrightarrow \boxed{H(z) = \frac{B(z)}{A(z)} = \frac{\sum_{k=0}^{q} b_k z^{-k}}{\sum_{k=0}^{p} a_k z^{-k}}} \longrightarrow x(n)$$

In order to guarantee a stable filter, all zeroes of $A(z)$ are assumed to lie *strictly* inside the unit circle.

## Temporal properties

The signal obeys the filtering equation

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + \sum_{k=0}^{q} b_k u(n-k).$$

## Spectral properties

The PSD is given by

$$S_x(z) = H(z)H^*(1/z^*)S_u(z) = \frac{B(z)B^*(1/z^*)}{A(z)A^*(1/z^*)} S_u(z)$$

$$S_x(f) = \sigma^2 \left|H(f)\right|^2 = \sigma^2 \frac{\left|\sum_{k=0}^{q} b_k e^{-i2\pi kf}\right|^2}{\left|\sum_{k=0}^{p} a_k e^{-i2\pi kf}\right|^2}.$$

## Influence of $A(z)$ and $B(z)$ on the PSD

The PSD depends entirely on $A(z)$ and $B(z)$. If we denote

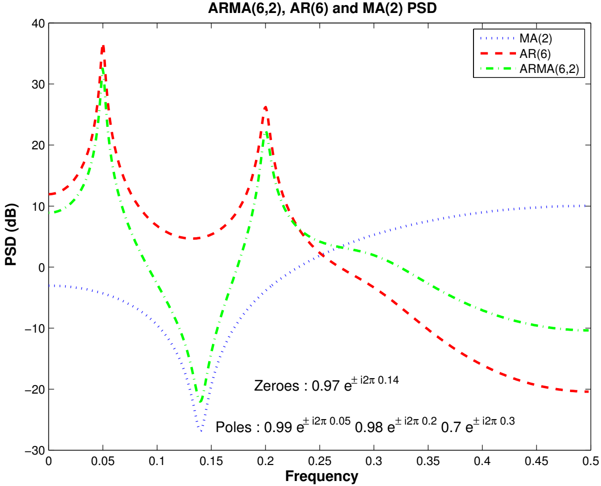
$$A(z) = \prod_{k=1}^{p} \left(1 - z_k z^{-1}\right) = \prod_{k=1}^{p} \left(1 - \rho_k e^{i\omega_k} z^{-1}\right)$$
$$B(z) = \prod_{k=1}^{q} \left(1 - \zeta_k z^{-1}\right) = \prod_{k=1}^{q} \left(1 - r_k e^{i\psi_k} z^{-1}\right)$$

then

- the poles $z_k$ correspond to "peaks" in the PSD, located at $(2\pi)^{-1}\,\omega_k$ and all the more sharp that $\rho_k$ is close to $1$, i.e. the pole is close to the unit circle.
- the zeroes $\zeta_k$ correspond to "nulls" in the PSD, located at $(2\pi)^{-1}\,\psi_k$ and all the more sharp that $r_k$ is close to $1$.

$\Longrightarrow$ an ARMA$(p, q)$ model enables one to approximate very accurately (depending on $p$ and $q$) any PSD.

# ARMA($p, q$) PSD example



# Relation between models

Every ARMA($p, q$) model can be approximated by an AR($\infty$) or MA($\infty$) model. For example,

$$\frac{B(z)}{A(z)} = \frac{1}{C(z)} \Leftrightarrow A(z) = B(z)C(z)$$

which implies that the $c_n$ are given by

$$c_n = \begin{cases} 1 & n = 0 \\ -\sum_{k=1}^{q} b_k c_{n-k} + a_n & 1 \leq n \leq p \\ -\sum_{k=1}^{q} b_k c_{n-k} & n > p \end{cases}$$

## Remark

The PSD depends only on $\sigma^2$, $\{a_k\}_{k=1}^p$ and $\{b_k\}_{k=1}^q$. Therefore, the correlation function $r_{xx}(m) = \mathcal{F}^{-1}(S_x(f))$ also depends on these parameters $\Longrightarrow$ **Yule-Walker equations**.

The filtering equation is the following

$$x(n) = -\sum_{k=1}^{p} a_k x(n-k) + \sum_{k=0}^{q} b_k u(n-k).$$

Pre-multiplying by $x^*(n-m)$ $(m \geq 0)$ and taking expectation, one obtains

$$r_{xx}(m) = -\sum_{k=1}^{p} a_k r_{xx}(m-k) + \sum_{k=0}^{q} b_k \mathcal{E}\left\{x^*(n-m)u(n-k)\right\}.$$

However,

$$\begin{aligned}
\mathcal{E}\left\{x^*(n-m)u(n-k)\right\} &= \sum_{\ell=0}^{\infty} h_\ell^* \mathcal{E}\left\{u^*(n-m-\ell)u(n-k)\right\} \\
&= \sigma^2 \sum_{\ell=0}^{\infty} h_\ell^* \delta(m+\ell-k) \\
&= \begin{cases} \sigma^2 h_{k-m}^* & k \geq m \\ 0 & \text{otherwise} \end{cases}
\end{aligned}$$

which implies that

$$\boxed{r_{xx}(m) = \begin{cases} r_{xx}^*(-m) & m < 0 \\ -\sum_{k=1}^{p} a_k r_{xx}(m-k) + \sigma^2 \sum_{k=m}^{q} b_k h_{k-m}^* & m \in [0,q] \\ -\sum_{k=1}^{p} a_k r_{xx}(m-k) & m > q \end{cases}}$$

*Yule-Walker Equations*

# Alternative proof

Taking the inverse $z$ transform of $A(z)S_x(z) = \sigma^2 B(z) H^*(1/z^*)$, and observing that $H^*(1/z^*) = \sum_{k=0}^{\infty} h_k^* z^k = \sum_{k=-\infty}^{0} h_{-k}^* z^{-k}$, it ensues

$$
\begin{aligned}
[a_n * r_{xx}(n)]_m &= \sum_{k=0}^{p} a_k r_{xx}(m-k) \\
&= \sigma^2 \left[ b_n * h_{-n}^* \right]_m \\
&= \sigma^2 \sum_{k=0}^{q} b_k h_{k-m}^* \\
&= \sigma^2 \sum_{k=m}^{q} b_k h_{k-m}^* .
\end{aligned}
$$

## Yule-Walker equations for an ARMA($p, q$) model

The coefficients $a_k$ can be obtained as the solution to the following **linear system of equations**:

$$\begin{pmatrix} r_{xx}(q) & r_{xx}(q-1) & \cdots & r_{xx}(q-p+1) \\ r_{xx}(q+1) & r_{xx}(q) & \cdots & r_{xx}(q-p+2) \\ \vdots & \vdots & \ddots & \vdots \\ r_{xx}(q+p-1) & r_{xx}(q+p-2) & \cdots & r_{xx}(q) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{pmatrix} = - \begin{pmatrix} r_{xx}(q+1) \\ r_{xx}(q+2) \\ \vdots \\ r_{xx}(q+p) \end{pmatrix}$$

$$\boldsymbol{R}\boldsymbol{a} = -\boldsymbol{r}$$

The relation between $b_k$ and $r_{xx}(m)$ is more complicated (non linear).

## Yule-Walker equations for an AR($p$) model

$$r_{xx}(m) = -\sum_{k=1}^{p} a_k r_{xx}(m-k) + \sigma^2 \delta(m).$$

The coefficients $a_k$ obey a linear system of equations:

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

$$\boldsymbol{R}\boldsymbol{a} = -\boldsymbol{r}$$

The white noise power is simply

$$\sigma^2 = \sum_{k=0}^{p} a_k r_{xx}(-k).$$

## Remark

- The recurrence equation $r_{xx}(m) = -\sum_{k=1}^{p} a_k r_{xx}(m-k)$ admits as a solution

$$r_{xx}(m) = \sum_{k=1}^{p} A_k e^{i\phi_k} z_k^m = \sum_{k=1}^{p} A_k e^{i\phi_k} \rho_k^m e^{im\omega_k}$$

which is a sum of damped complex exponentials, with frequencies $\omega_k/(2\pi)$ and damping factors $\rho_k$. The closer $\rho_k$ to $1$, the longer the temporal support of $r_{xx}(m)$ and hence the spectral power is concentrated on a smaller frequency band. This is why AR modeling allows for high spectral resolution.

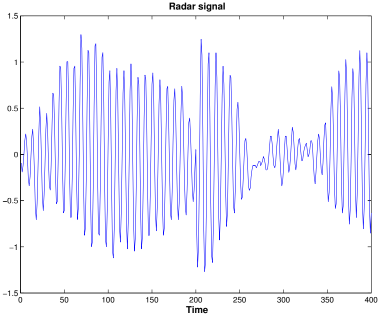
### Yule-Walker equations for a MA($q$) model

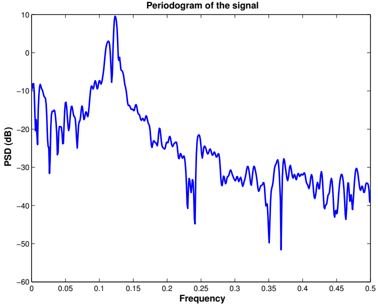
The coefficients $b_k$ now obey **non linear equations**

$$r_{xx}(m) = \begin{cases} \sigma^2 \sum_{k=m}^{q} b_k b_{k-m}^* & m \in [0, q] \\ 0 & m > q \end{cases}$$

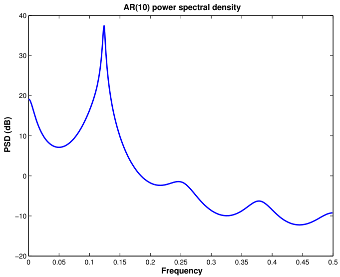
Since the correlation function is of finite duration, no way to perform high resolution spectral analysis with a MA($q$) model.

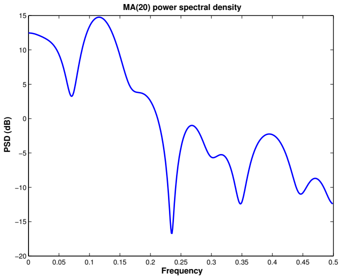
# Radar signal

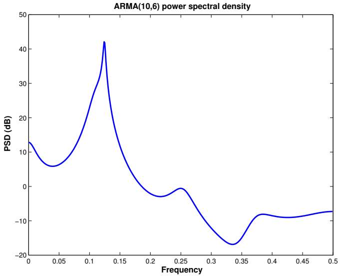


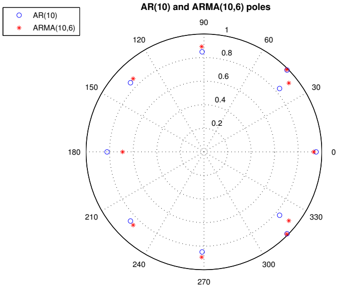



# The pitfalls of modeling



## Question

Let $x(n)$ be an AR($p$) process, with parameters $\sigma^2$, $a_1, \cdots, a_p$. Which is the **best linear predictor of order $p$** of $x(n)$:

$$\hat{x}(n) = -\sum_{k=1}^{p} \alpha_k x(n-k).$$

## Linear prediction error (LPE)

One looks for the coefficients $\alpha_k$ that minimize

$$\begin{aligned}
P_{\text{lpe}} &= \mathcal{E}\left\{|e(n)|^2\right\} = \mathcal{E}\left\{|\hat{x}(n) - x(n)|^2\right\} \\
&= \mathcal{E}\left\{\left[x(n) + \sum_{k=1}^{p} \alpha_k x(n-k)\right]\left[x^*(n) + \sum_{k=1}^{p} \alpha_k^* x^*(n-k)\right]\right\} \\
&= r_{xx}(0) + \sum_{k=1}^{p} \alpha_k r_{xx}(-k) + \sum_{k=1}^{p} \alpha_k^* r_{xx}(k) + \sum_{k=1}^{p}\sum_{m=1}^{p} \alpha_k \alpha_m^* r_{xx}(m-k).
\end{aligned}$$

## Linear prediction error

With $\boldsymbol{r} = \begin{bmatrix} r_{xx}(1) & r_{xx}(2) & \cdots & r_{xx}(p) \end{bmatrix}^T$ and $\boldsymbol{R}(k,\ell) = r_{xx}(k-\ell)$, one has

$$\begin{aligned} P_{\mathsf{lpe}} &= r_{xx}(0) + \boldsymbol{\alpha}^H \boldsymbol{r} + \boldsymbol{r}^H \boldsymbol{\alpha} + \boldsymbol{\alpha}^H \boldsymbol{R} \boldsymbol{\alpha} \\ &= \left(\boldsymbol{\alpha} + \boldsymbol{R}^{-1}\boldsymbol{r}\right)^H \boldsymbol{R} \left(\boldsymbol{\alpha} + \boldsymbol{R}^{-1}\boldsymbol{r}\right) + r_{xx}(0) - \boldsymbol{r}^H \boldsymbol{R}^{-1} \boldsymbol{r} \\ &\geq r_{xx}(0) - \boldsymbol{r}^H \boldsymbol{R}^{-1} \boldsymbol{r} \end{aligned}$$

with equality iif $\boldsymbol{\alpha} = -\boldsymbol{R}^{-1}\boldsymbol{r} = \boldsymbol{a}$: the best linear predictor is the AR parameter vector! Additionally,

$$P_{\mathsf{lpe\text{-}min}} = r_{xx}(0) - \boldsymbol{r}^H \boldsymbol{R}^{-1} \boldsymbol{r} = r_{xx}(0) + \boldsymbol{r}^H \boldsymbol{a} = \sigma^2.$$

$\Longrightarrow$ Solving the Yule-Walker equations is **equivalent** to minimizing the linear prediction error.

## Remark

The best predictor is the one for which the **prediction error** $e(n)$ **is orthogonal to the data** $\{x(n-k)\}_{k=1}^{p}$. Indeed,

$$\mathcal{E}\left\{e(n)x^*(n-k)\right\} = \mathcal{E}\left\{\sum_{\ell=0}^{p} \alpha_\ell x(n-\ell)x^*(n-k)\right\}$$
$$= \sum_{\ell=0}^{p} \alpha_\ell r_{xx}(k-\ell) = 0.$$

The optimal coefficients $\alpha_k$ make the prediction error $e(n)$ orthogonal (i.e. *uncorrelated*) to $\{x(n-1), \cdots, x(n-p)\}$. The innovation $e(n)$ can be viewed as the part of information in $x(n)$ which is not already contained in $\{x(n-1), \cdots, x(n-p)\}$.

## Theory

The parameters $a_1, \cdots, a_p$ are *theoretically* obtained in an equivalent way

1. by solving Yule-Walker equations $\boldsymbol{R}\boldsymbol{a} = -\boldsymbol{r}$
2. or by minimizing the linear prediction error $\mathcal{E}\left\{\left|x(n) + \sum_{k=1}^{p} a_k x(n-k)\right|^2\right\}$

## In practice

In practice the parameters $a_1, \cdots, a_p$ are *estimated* (in an *almost* equivalent way)

1. either by solving Yule-Walker equations $\hat{\boldsymbol{R}}\boldsymbol{a} = -\hat{\boldsymbol{r}}$
2. or by minimizing the linear prediction error $\sum_n \left|x(n) + \sum_{k=1}^{p} a_k x(n-k)\right|^2$

# Yule-Walker method

- The correlation function is first estimated

$$\hat{r}_{xx}(m) = \frac{1}{N-m} \sum_{n=0}^{N-m-1} x^*(n)x(n+m) \qquad m = 0, \cdots, p$$

- Then, one solves a linear system of $p$ equations in $p$ unknowns

$$\begin{pmatrix} \hat{r}_{xx}(0) & \hat{r}_{xx}(-1) & \cdots & \hat{r}_{xx}(-p+1) \\ \hat{r}_{xx}(1) & \hat{r}_{xx}(0) & \cdots & \hat{r}_{xx}(-p+2) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{r}_{xx}(p-1) & \hat{r}_{xx}(p-2) & \cdots & \hat{r}_{xx}(0) \end{pmatrix} \begin{pmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_p \end{pmatrix} = - \begin{pmatrix} \hat{r}_{xx}(1) \\ \hat{r}_{xx}(2) \\ \vdots \\ \hat{r}_{xx}(p) \end{pmatrix}$$

whose solution is

$$\boxed{\hat{\boldsymbol{a}} = -\hat{\boldsymbol{R}}^{-1}\hat{\boldsymbol{r}}}$$

# Minimization of the linear prediction error

- One seeks to minimize $\|\boldsymbol{X}\boldsymbol{a}+\boldsymbol{h}\|^2$ with

$$\boldsymbol{X}=\begin{pmatrix} x(p-1) & x(p-2) & \cdots & x(0) \\ x(p) & x(p-1) & \cdots & x(1) \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ x(N-2) & x(N-3) & \cdots & x(N-p-1) \end{pmatrix}, \boldsymbol{h}=\begin{pmatrix} x(p) \\ x(p+1) \\ \vdots \\ \vdots \\ x(N-1) \end{pmatrix}$$

- Since

$$\begin{aligned}\|\boldsymbol{X}\boldsymbol{a}+\boldsymbol{h}\|^2 &= (\boldsymbol{X}\boldsymbol{a}+\boldsymbol{h})^H(\boldsymbol{X}\boldsymbol{a}+\boldsymbol{h}) \\ &= \left[\boldsymbol{a}+\left(\boldsymbol{X}^H\boldsymbol{X}\right)^{-1}\boldsymbol{X}^H\boldsymbol{h}\right]^H\left(\boldsymbol{X}^H\boldsymbol{X}\right)\left[\boldsymbol{a}+\left(\boldsymbol{X}^H\boldsymbol{X}\right)^{-1}\boldsymbol{X}^H\boldsymbol{h}\right] \\ &+ \boldsymbol{h}^H\boldsymbol{h}-\boldsymbol{h}^H\boldsymbol{X}\left(\boldsymbol{X}^H\boldsymbol{X}\right)^{-1}\boldsymbol{X}^H\boldsymbol{h}\end{aligned}$$

the solution is given by $\hat{\boldsymbol{a}}=-\left(\boldsymbol{X}^H\boldsymbol{X}\right)^{-1}\boldsymbol{X}^H\boldsymbol{h}$.

## Remarks

- $\boldsymbol{X}^H\boldsymbol{X} \simeq \hat{\boldsymbol{R}}$ and $\boldsymbol{X}^H\boldsymbol{h} \simeq \hat{\boldsymbol{r}}$.
- In general, one avoids computing $\left(\boldsymbol{X}^H\boldsymbol{X}\right)^{-1}\boldsymbol{X}^H\boldsymbol{h}$: rather a decomposition (typically QR) of $\boldsymbol{X}$ is used to solve efficiently the linear least-squares problem $\min_{\boldsymbol{a}} \|\boldsymbol{X}\boldsymbol{a} + \boldsymbol{h}\|^2$.
- The previous algorithm uses only the available data making no assumption about the signal outside the observation interval. One could add rows to $\boldsymbol{X}$ assuming that $x(n) = 0$ for $n \notin [0, N-1]$.
- Fast, order recursive (which compute all predictors of order $k$, $k = 1, \cdots, p$) algorithms are available. They give access to the power of the linear prediction error for all predictors of order $k \leq p$ and can be useful in selecting the best model order.

## Levinson algorithm

**Inputs:** $r_{xx}(m)$, $m = 0, \cdots, p$

$a_1[1] = -\frac{r_x x(1)}{r_{xx}(0)}$, $P_{\mathsf{epl}}[1] = \left(1 - |a_1[1]|^2\right) r_{xx}(0)$

**for** $k = 1, \cdots, p$ **do**

$$a_k[k] = -\frac{r_{xx}(k) + \sum_{\ell=1}^{k-1} a_{k-1}[\ell] r_{xx}(k-\ell)}{P_{\mathsf{epl}}[k-1]}$$

$$a_k[\ell] = a_{k-1}[\ell] + a_k[k] a^*_{k-1}[k-\ell] \quad \ell = 1, \cdots, k-1$$

$$P_{\mathsf{epl}}[k] = \left(1 - |a_k[k]|^2\right) P_{\mathsf{epl}}[k-1]$$

**end for**

**Outputs:** $\boldsymbol{a}_k = -\boldsymbol{R}_k^{-1} \boldsymbol{r}_k$ et $P_{\mathsf{epl}}[k]$ pour $k = 1, \cdots, p$ où

$\boldsymbol{R}_k(\ell, n) = r_{xx}(\ell - n)$, $\boldsymbol{r}_k(\ell) = r_{xx}(\ell)$, $\ell, n = 1, \cdots, k$.

## Question

Why using only $p$ Yule-Walker equations while $r_{xx}(m) = -\sum_{k=1}^{p} a_k r_{xx}(m-k)$, for $m = 1, \cdots, \infty$?

## Modifed Yule-Walker

One solves in the least-squares sense $\hat{\boldsymbol{R}}\boldsymbol{a} \simeq -\hat{\boldsymbol{r}}$ with

$$\hat{\boldsymbol{R}} = \begin{pmatrix} \hat{r}_{xx}(0) & \hat{r}_{xx}(-1) & \cdots & \hat{r}_{xx}(-p+1) \\ \vdots & \vdots & \vdots & \vdots \\ \hat{r}_{xx}(p-1) & \hat{r}_{xx}(p-2) & \cdots & \hat{r}_{xx}(0) \\ \vdots & \vdots & \vdots & \vdots \\ \hat{r}_{xx}(M-1) & \hat{r}_{xx}(M-2) & \cdots & \hat{r}_{xx}(M-p) \end{pmatrix}, \hat{\boldsymbol{r}} = \begin{pmatrix} \hat{r}_{xx}(1) \\ \vdots \\ \hat{r}_{xx}(p) \\ \vdots \\ \hat{r}_{xx}(M) \end{pmatrix}$$

The solution is obtained as

$$\hat{\boldsymbol{a}} = \arg\min_{\boldsymbol{a}} \left\| \hat{\boldsymbol{R}}\boldsymbol{a} + \hat{\boldsymbol{r}} \right\|^2 = -\left( \hat{\boldsymbol{R}}^H \hat{\boldsymbol{R}} \right)^{-1} \hat{\boldsymbol{R}}^H \hat{\boldsymbol{r}}$$

# Spectral analysis based on $\hat{a}_k$

Once $\hat{a}_k$, $k = 1, \cdots, p$ and $\hat{\sigma}^2$ are obtained, spectral information is made available:

1. either by estimating the **power spectral density**

$$\hat{S}_x(f) = \frac{\hat{\sigma}^2}{\left|1 + \sum_{k=1}^{p} \hat{a}_k e^{-i2\pi k f}\right|^2}$$

   and observing the peaks of the PSD.

2. or by estimating the **poles of the model**

$$\hat{A}(z) = 1 + \sum_{k=1}^{p} \hat{a}_k z^{-k} = \prod_{k=1}^{p} \left(1 - \hat{\rho}_k e^{i\hat{\omega}_k} z^{-1}\right)$$

   and retaining those which are closest to the unit circle.

## Question

Let $x(n) = Ae^{i(2\pi n f_0 + \varphi)} + w(n)$ where $\varphi$ is uniformly distributed on $[0, 2\pi[$ and $w(n)$ is a white noise with variance $\sigma_w^2$. What happens if an AR($p$) model is fitted to such a signal?

## Answer

In the case where $r_{xx}(m)$ is **known**, the PSD associated with an AR($p$) model of $x(n)$ achieves its maximum at $f = f_0$.

## Proof

One has $r_{xx}(m) = Pe^{i2\pi m f_0} + \sigma_w^2 \delta(m)$ which implies that $\boldsymbol{R} = P\boldsymbol{s}\boldsymbol{s}^H + \sigma_w^2 \boldsymbol{I}$, $\boldsymbol{r} = P\boldsymbol{s}$ with $\boldsymbol{s} = \begin{bmatrix} e^{i2\pi f_0} & \cdots & e^{i2\pi p f_0} \end{bmatrix}^T$.

## Proof (cont'd)

It can be deduced that

$$\boldsymbol{a} = -\frac{P}{\sigma_w^2 + pP}\boldsymbol{s}, \quad \sigma^2 = \sigma_w^2 \left[1 + \frac{P}{\sigma_w^2 + pP}\right]$$

Therefore, the PSD can be written as

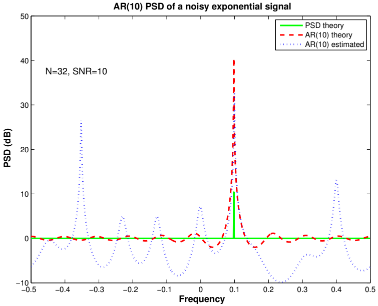
$$S_x(f) = \frac{\sigma^2}{\left|1 - \frac{P}{\sigma_w^2 + pP}\boldsymbol{e}^H(f)\boldsymbol{s}\right|^2}$$

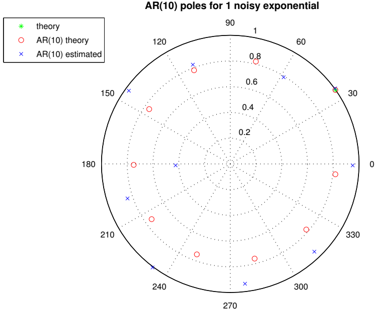
where $\boldsymbol{e}(f) = \begin{bmatrix} e^{i2\pi f} & \cdots & e^{i2\pi p f}\end{bmatrix}^T$ and its maximum is located at $f = f_0$. However

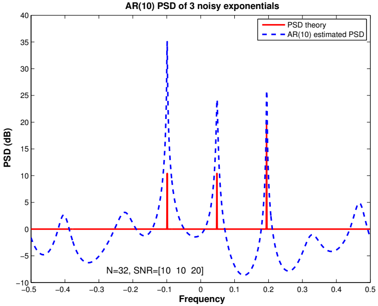
$$\begin{aligned} S_x(f_0) &= \sigma_w^2 \left[1 + (p+1)\frac{P}{\sigma_w^2}\right]\left[1 + p\frac{P}{\sigma_w^2}\right] \\ &\simeq p(p+1)\frac{P^2}{\sigma_w^2} \text{ for } \frac{P}{\sigma_w^2} \gg 1 \end{aligned}$$

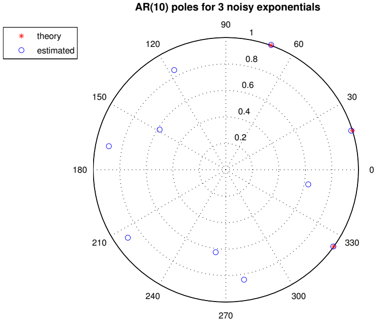
# Comments

- Even if a complex exponential is **not** an AR($p$) signal, an AR($p$) model enables one to recover the frequency of the exponential $\Longrightarrow$ one can use an AR($p$) model to estimate the frequency of a complex exponential signal (and, by extension, the frequencies of a sum of complex exponentials).
- The amplitude of the AR($p$) peak is not commensurate with the actual power of the exponential signal (contrary to Fourier analysis).

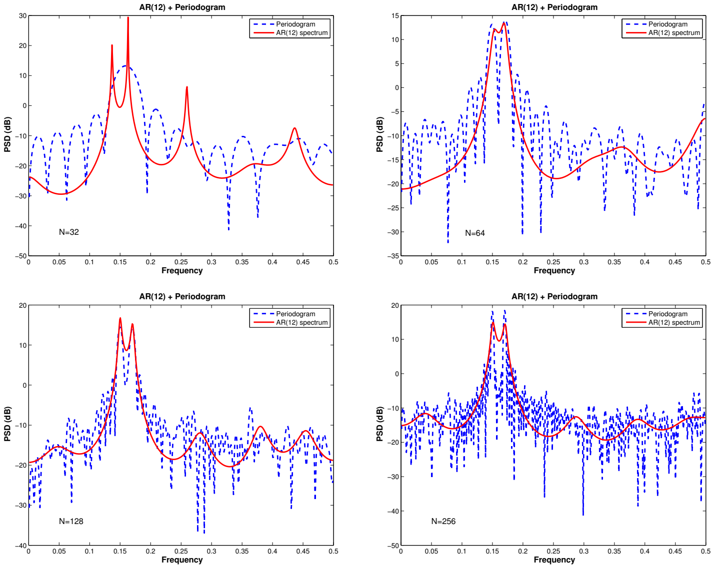

AR(10) PSD of a noisy exponential signal
PSD theory
AR(10) theory
AR(10) estimated
N=32, SNR=10
PSD (dB)
Frequency


AR(10) poles for 1 noisy exponential
theory
AR(10) theory
AR(10) estimated

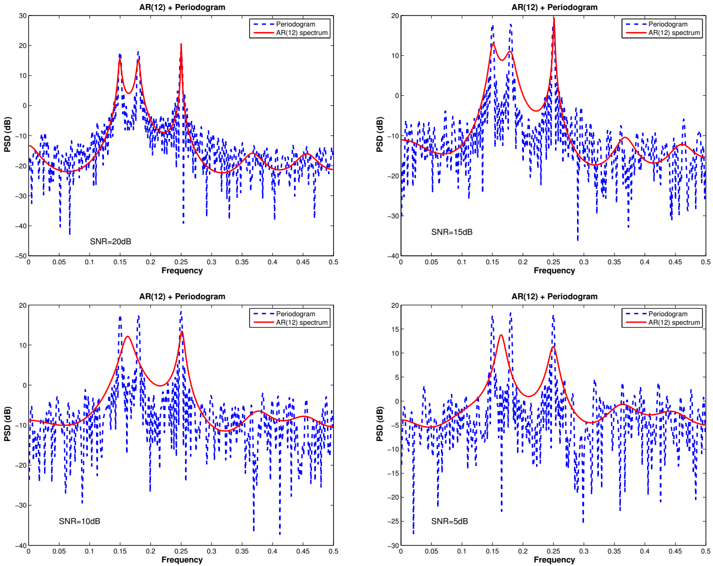

AR(10) PSD of 3 noisy exponentials
PSD theory
AR(10) estimated PSD
N=32, SNR=[10 10 20]
PSD (dB)
Frequency


AR(10) poles for 3 noisy exponentials
theory
estimated

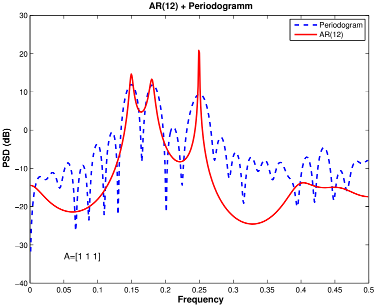
# Influence of $N$ on AR($p$) modeling

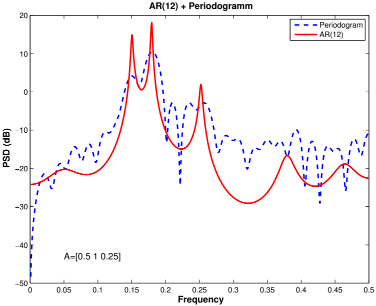


# Influence of SNR on AR($p$) modeling



# Problem of differences between components amplitudes



# Properties of AR($p$) modeling

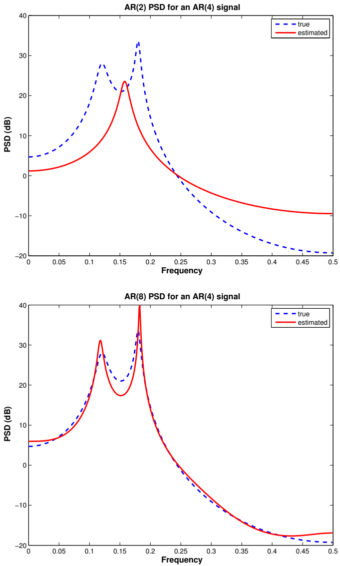
- Better resolution than periodogram, at least for small $N$ and high SNR

$$\delta f_{AR} \simeq \frac{1.03}{p\left[(p+1)\,SNR\right]^{0.31}}$$
$$\delta f_{PER} \simeq \frac{0.86}{N}$$

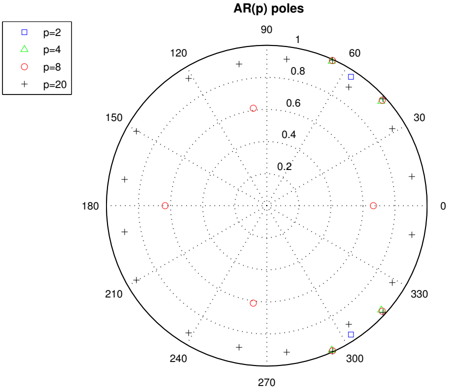
$\Longrightarrow$ interest only for short samples and large signal to noise ratio.

- Contrary to the periodogram, for complex sine waves, the amplitude of the AR peaks is not proportional to the power of the exponentials.
- Contrary to the periodogram, no problem with strong signals masking weak signals.

# Model order selection



# Model order selection




- a too small order results in smoothing the spectrum.
- a too large order gives rise to spurious peaks.

Remark: in case of an AR$(4)$ model, $\boldsymbol{R}$ of size $20 \times 20$ is not inversible and $\hat{\boldsymbol{R}}$ is badly conditioned.

# Criteria for model order selection

Based on the power of the linear prediction error at order $k$:

Akaike Information Criterion

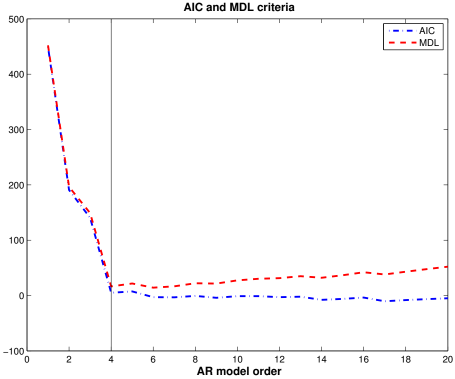
$$AIC(k) = N \ln(P_{\mathsf{epl}}[k]) + 2k$$

Final Prediction Error

$$FPE(k) = \frac{N+k+1}{N-k-1} P_{\mathsf{epl}}[k]$$

Minimum Description Length

$$MDL(k) = N \ln(P_{\mathsf{epl}}[k]) + p \ln(N)$$


AIC and MDL criteria
500
400
300
200
100
0
−100
0
2
4
6
8
10
12
14
16
18
20
AR model order
AIC
MDL

## Principle

One usually proceeds in 2 steps:

1. Estimation of parameters $a_1, \cdots, a_p$ using Yule-Walker equations

$$r_{xx}(m) = -\sum_{k=1}^{p} a_k r_{xx}(m-k) \quad m > q.$$

2. Estimation of parameters $b_1, \cdots, b_q$:
   - the signal $x(n)$ is filtered by $\hat{A}(z)$ to yield $y(n) = \sum_{k=0}^{p} \hat{a}_k x(n-k)$ which is theoretically MA($q$).
   - an AR($L$) (with $L$ "large") is fitted to $y(n)$, with coefficients $c_1, \cdots, c_L$, and one uses the equivalence between MA($q$) and AR($\infty$) models:

$$\left(\sum_{k=0}^{q} b_k z^{-k}\right)\left(\sum_{m=0}^{\infty} c_m z^{m}\right) = 1 \Leftrightarrow c_m = -\sum_{k=1}^{q} b_k c_{m-k} + \delta(m)$$

## Modified Yule-Walker

The linear system of $p$ equations in $p$ unknowns $\hat{\boldsymbol{R}}\hat{\boldsymbol{a}} = -\hat{\boldsymbol{r}}$ is solved, where

$$\hat{\boldsymbol{R}} = \begin{pmatrix} \hat{r}_{xx}(q) & \hat{r}_{xx}(q-1) & \cdots & \cdots & \hat{r}_{xx}(q-p+1) \\ \hat{r}_{xx}(q+1) & \hat{r}_{xx}(q) & \cdots & \cdots & \hat{r}_{xx}(q-p+2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{r}_{xx}(q+p-1) & \hat{r}_{xx}(q+p-2) & \cdots & \cdots & \hat{r}_{xx}(q) \end{pmatrix}$$

$$\hat{\boldsymbol{r}} = \begin{pmatrix} \hat{r}_{xx}(q+1) \\ \hat{r}_{xx}(q+2) \\ \vdots \\ \hat{r}_{xx}(q+p) \end{pmatrix}$$

## Least-squares Yule-Walker

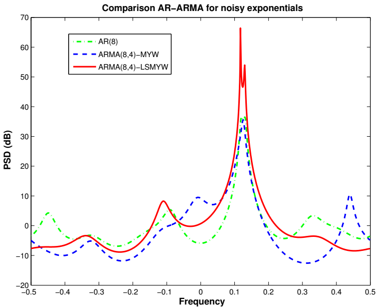
One solves, in the least-squares sense, a linear system of $M > p$ Yule-Walker equations with $p$ unknowns

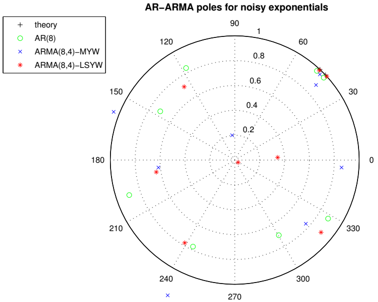
$$\hat{\boldsymbol{a}} = \arg\min_{\boldsymbol{a}} \left\| \hat{\boldsymbol{R}}\boldsymbol{a} + \hat{\boldsymbol{r}} \right\|^2 = -\left( \hat{\boldsymbol{R}}^H \hat{\boldsymbol{R}} \right)^{-1} \hat{\boldsymbol{R}}^H \hat{\boldsymbol{r}}$$

where

$$\hat{\boldsymbol{R}} = \begin{pmatrix} \hat{r}_{xx}(q) & \hat{r}_{xx}(q-1) & \cdots & \cdots & \hat{r}_{xx}(q-p+1) \\ \hat{r}_{xx}(q+1) & \hat{r}_{xx}(q) & \cdots & \cdots & \hat{r}_{xx}(q-p+2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{r}_{xx}(M+q-1) & \hat{r}_{xx}(M+q-2) & \cdots & \cdots & \hat{r}_{xx}(M+q-p+1) \end{pmatrix}$$

$$\hat{\boldsymbol{r}} = \begin{pmatrix} \hat{r}_{xx}(q+1) \\ \hat{r}_{xx}(q+2) \\ \vdots \\ \hat{r}_{xx}(M+q) \end{pmatrix}$$


Comparison AR−ARMA for noisy exponentials
AR(8)
ARMA(8,4)−MYW
ARMA(8,4)−LSMYW
PSD (dB)
Frequency


AR−ARMA poles for noisy exponentials
theory
AR(8)
ARMA(8,4)−MYW
ARMA(8,4)−LSYW

## Summary

- An ARMA($p, q$) enables one to approximate very accurately the PSD of a large class of signals. The AR part deals with peaks in the spectrum while the MA part models the valleys.
- The model parameters are usually estimated solving the Yule-Walker equations (which involve the correlation function). These equations are linear with respect to the AR parameters, non linear with respect to the MA parameters.
- Information about the spectral content can be retrieved from the (rational) ARMA PSD or from examining the poles and zeroes of the model.
- For an AR($p$) model, solving Yule-Walker equations is equivalent to minimizing the linear prediction error.
- AR and ARMA models are suitable for frequency estimation of complex exponential signals, with ARMA offering an enhanced resolution.

## Damped exponential signals

We are now interested in (possibly damped) exponential signals embedded in noise:

$$x(n) = s(n) + w(n) = \sum_{k=1}^{p} A_k e^{i\phi_k} e^{(-\alpha_k + i2\pi f_k)n} + w(n)$$

## Relation to AR($p$) models

Although $s(n)$ is not an AR($p$) process, it obeys linear prediction equations, similar to those of an AR($p$) signal.

## Methods

The main approach consists in solving the linear prediction equations

- either in a least-squares sense (Prony).
- or using the fact that $s(n)$, a linear combination of $p$ modes, lies within a **subspace** of size $p$ (Tufts-Kumaresan).

## Original problem

Assume we observe $2p$ samples $\{x(n)\}_{n=0}^{2p-1}$ of the following signal

$$x(n) = \sum_{k=1}^{p} A_k e^{i\phi_k} e^{(-\alpha_k + i2\pi f_k)n} = \sum_{k=1}^{p} h_k z_k^n.$$

From these $2p$ samples can we recover the $4p$ unknown parameters $A_k$, $\phi_k$, $\alpha_k$ and $f_k$, $k = 1, \cdots, p$?

## Answer

Let $A(z) = \prod_{k=1}^{p}(1 - z_k z^{-1}) = 1 + \sum_{k=1}^{p} a_k z^{-k}$. One has

$$\begin{aligned}\sum_{k=0}^{p} a_k x(n-k) &= \sum_{k=0}^{p} a_k \left(\sum_{\ell=1}^{p} h_\ell z_\ell^{n-k}\right) \\ &= \sum_{\ell=1}^{p} h_\ell z_\ell^{n} \left(\sum_{k=0}^{p} a_k z_\ell^{-k}\right) = 0.\end{aligned}$$

## Obtaining $z_k$

$a_k$ is obtained by solving

$$\begin{pmatrix} x(p-1) & x(p-2) & \cdots & \cdots & x(0) \\ x(p) & x(p-1) & \cdots & \cdots & x(1) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ x(2p-2) & x(2p-3) & \cdots & \cdots & x(p-1) \end{pmatrix} \begin{pmatrix} a_1 \\ a_2 \\ \vdots \\ \vdots \\ a_p \end{pmatrix} = - \begin{pmatrix} x(p) \\ x(p+1) \\ \vdots \\ \vdots \\ x(2p-1) \end{pmatrix}$$

which yields $z_k$ as the roots of

$$A(z) = 1 + \sum_{k=1}^{p} a_k z^{-k} = \prod_{k=1}^{p} (1 - z_k z^{-1}).$$

## Obtaining $h_k$

Once the $z_k$'s are available, the following Vandermonde system is solved

$$\begin{pmatrix} 1 & 1 & \cdots & \cdots & 1 \\ z_1 & z_2 & \cdots & \cdots & z_p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ z_1^{p-1} & z_2^{p-1} & \cdots & \cdots & z_p^{p-1} \end{pmatrix} \begin{pmatrix} h_1 \\ h_2 \\ \vdots \\ \vdots \\ h_p \end{pmatrix} = \begin{pmatrix} x(0) \\ x(1) \\ \vdots \\ \vdots \\ x(p-1) \end{pmatrix}$$

$\Longrightarrow$ unique solution to this problem with $4p$ equations and $4p$ unknowns.

## Problem

In general $N > p$ noisy samples are available:

$$x(n) = \sum_{k=1}^{p} A_k e^{i\phi_k} e^{(-\alpha_k + i2\pi f_k)n} + w(n); \qquad n = 0, \cdots, N-1$$

from which one tries to estimate $h_k = A_k e^{i\phi_k}$ and $z_k = e^{-\alpha_k + i2\pi f_k}$.

## Maximum likelihood

Under the assumption of white Gaussian noise $w(n)$, the maximum likelihood estimator amounts to minimizing the approximation error:

$$\hat{\boldsymbol{h}}, \hat{\boldsymbol{z}} = \arg\min_{\boldsymbol{h},\boldsymbol{z}} \sum_{n=0}^{N-1} \left| x(n) - \sum_{k=1}^{p} h_k z_k^n \right|^2$$

$\Longrightarrow$ non linear least-squares problem with $p$ complex-valued unknowns $z_k$.

## Least-squares Prony

Instead of minimizing the approximation error, one minimizes the power of the linear prediction error $e(n) = x(n) + \sum_{k=1}^{p} a_k x(n-k)$, which is equivalent to solving, in a least-squares sense, the linear system of equations

$$\boldsymbol{X}\boldsymbol{a} \simeq -\boldsymbol{h}$$

$$\boldsymbol{X} = \begin{pmatrix} x(p-1) & x(p-2) & \cdots & x(0) \\ x(p) & x(p-1) & \cdots & x(1) \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ x(N-2) & x(N-3) & \cdots & x(N-p-1) \end{pmatrix}, \boldsymbol{h} = \begin{pmatrix} x(p) \\ x(p+1) \\ \vdots \\ \vdots \\ x(N-1) \end{pmatrix}$$

whose solution is given by

$$\hat{\boldsymbol{a}} = \arg\min_{\boldsymbol{a}} \left\| \boldsymbol{X}\boldsymbol{a} + \boldsymbol{h} \right\|^2.$$

This is equivalent to using an AR($p$) model for $x(n)$.

## Estimation of $z_k$

$$\hat{A}(z) = 1 + \sum_{k=1}^{p} \hat{a}_k z^{-k} = \prod_{k=1}^{p} \left(1 - \hat{z}_k z^{-1}\right)$$

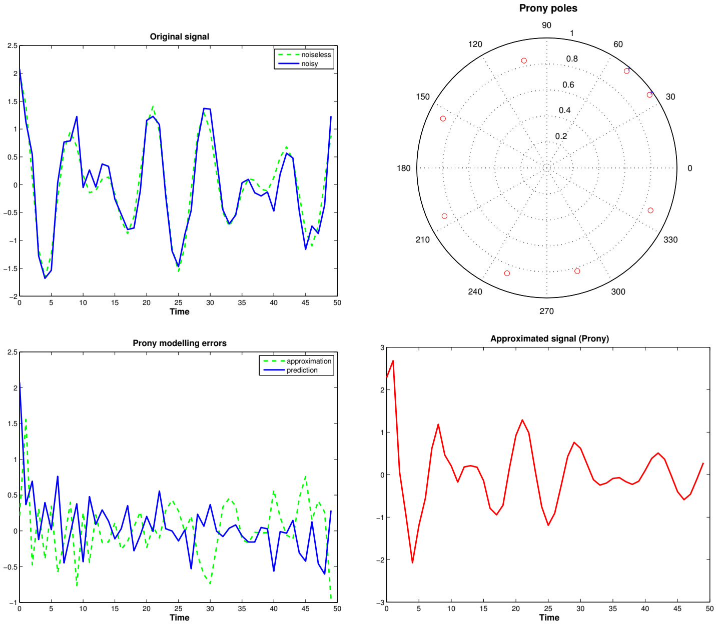
## Estimation of $h_k$

The Vandermonde system is solved in a least-squares sense

$$\begin{pmatrix} 1 & 1 & \cdots & \cdots & 1 \\ \hat{z}_1 & \hat{z}_2 & \cdots & \cdots & \hat{z}_p \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{z}_1^{N-1} & \hat{z}_2^{N-1} & \cdots & \cdots & \hat{z}_p^{N-1} \end{pmatrix} \begin{pmatrix} \hat{h}_1 \\ \hat{h}_2 \\ \vdots \\ \vdots \\ \hat{h}_p \end{pmatrix} = \begin{pmatrix} x(0) \\ x(1) \\ \vdots \\ \vdots \\ x(N-1) \end{pmatrix}$$

The solution can be written as

$$\hat{\boldsymbol{h}} = \left(\hat{\boldsymbol{Z}}^H \hat{\boldsymbol{Z}}\right)^{-1} \hat{\boldsymbol{Z}}^H \boldsymbol{x}.$$


Original signal
noiseless
noisy
Time
Prony poles
Prony modelling errors
approximation
prediction
Time
Approximated signal (Prony)
Time

# Prony's spectrum

Prony's spectrum is defined from the noiseless signal, in $2$ different ways:

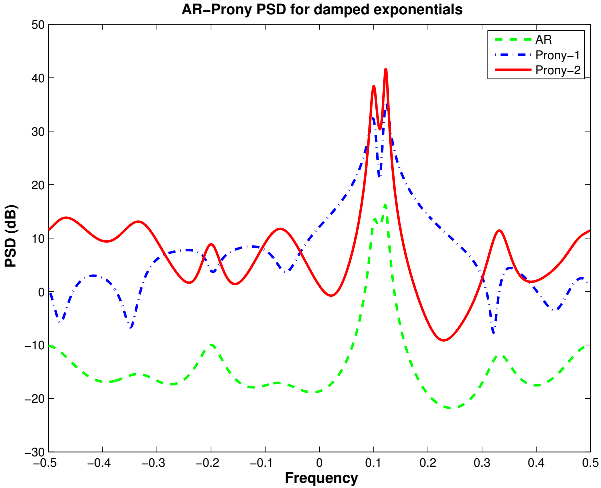
1. One assumes that

$$\hat{x}(n) = \begin{cases} \sum_{k=1}^{p} \hat{h}_k \hat{z}_k^n & n \geq 0 \\ 0 & n < 0 \end{cases} \xrightarrow{\mathcal{Z}} \hat{X}(z) = \sum_{k=1}^{p} \frac{\hat{h}_k}{1 - \hat{z}_k z^{-1}}$$

2. One assumes that

$$\hat{x}(n) = \begin{cases} \sum_{k=1}^{p} \hat{h}_k \hat{z}_k^n & n \geq 0 \\ \sum_{k=1}^{p} \hat{h}_k (\hat{z}_k^*)^{-n} & n < 0 \end{cases} \xrightarrow{\mathcal{Z}} \hat{X}(z) = \sum_{k=1}^{p} \frac{\hat{h}_k \left(1 - |\hat{z}_k|^2\right)}{\left(1 - \hat{z}_k z^{-1}\right)\left(1 - \hat{z}_k^* z\right)}$$

The "PSD" is then obtained as $\hat{S}(f) = \left| \hat{X}(e^{i2\pi f}) \right|^2$.


AR–Prony PSD for damped exponentials
AR
Prony–1
Prony–2
PSD (dB)
Frequency
50
40
30
20
10
0
–10
–20
–30
–0.5
–0.4
–0.3
–0.2
–0.1
0
0.1
0.2
0.3
0.4
0.5

# Prony correlation

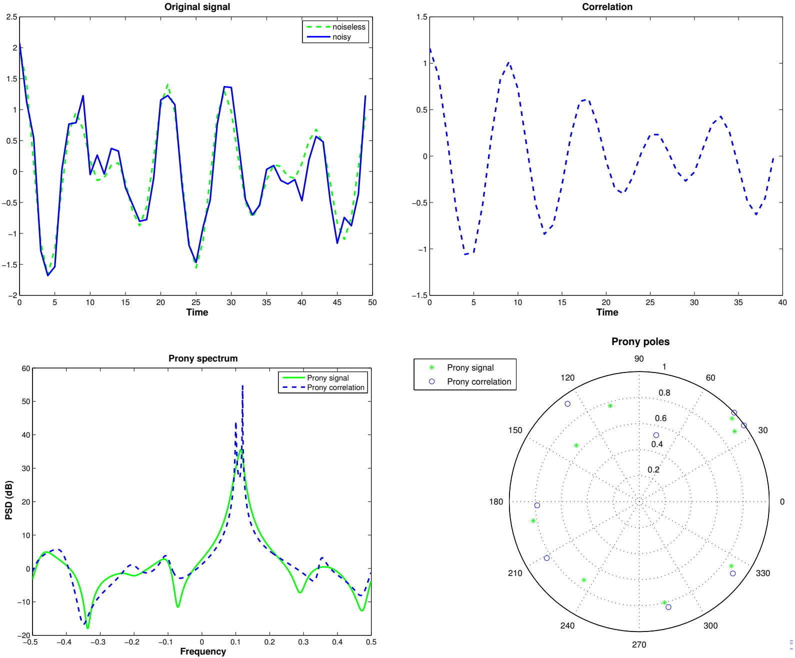
- We assume that the correlation function can be written as a sum of $p$ complex exponentials plus the correlation due to white noise:

$$r_{xx}(m) = \mathcal{E}\left\{x^*(n)x(n+m)\right\} = \sum_{k=1}^{p} P_k z_k^m + \sigma^2\delta(m).$$

- The correlation function hence verifies the following linear prediction equations

$$r_{xx}(m) = -\sum_{k=1}^{p} a_k r_{xx}(m-k) + \sigma^2 \sum_{k=1}^{p} a_k \delta(m-k)$$

which suggests estimating coefficients $a_k$ by minimization of the linear prediction error based on $r_{xx}(m)$.


Original signal
noiseless
noisy
Time
Correlation
Time
Prony spectrum
Prony signal
Prony correlation
PSD (dB)
Frequency
Prony poles
Prony signal
Prony correlation

## Reminder

For the signal $x(n) = \sum_{k=1}^{p} h_k z_k^n + w(n)$, Prony's method amounts to solving, in a least-squares sense, the linear system of $p$ linear prediction equations:

$$\underset{N-p|p}{\boldsymbol{X}} \; \underset{p|1}{\boldsymbol{a}} = - \underset{N-p|1}{\boldsymbol{h}}.$$

## Question

What happens, **in the noiseless case** where $x(n) = \sum_{k=1}^{p} h_k z_k^n$, **if one uses a linear prediction filter of order $L > p$**, that is if one tries to solve $\boldsymbol{X}\boldsymbol{a} = -\boldsymbol{h}$ with

$$\boldsymbol{X} = \begin{pmatrix} x(L-1) & x(L-2) & \cdots & x(0) \\ x(L) & x(L-1) & \cdots & x(1) \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ x(N-2) & x(N-3) & \cdots & x(N-L-1) \end{pmatrix}, \boldsymbol{h} = \begin{pmatrix} x(L) \\ x(L+1) \\ \vdots \\ \vdots \\ x(N-1) \end{pmatrix}$$

and $L > p$**?**

# Linear algebra reminders

Let $\boldsymbol{A} \in \mathbb{C}^{m\times n}$ be a complex matrix of size $m \times n$.

- The kernel (null space) and the range space of $\boldsymbol{A}$ are defined as

$$\mathcal{N}\left\{\boldsymbol{A}\right\} = \left\{\boldsymbol{x} \in \mathbb{C}^n / \boldsymbol{A}\boldsymbol{x} = \boldsymbol{0}\right\}$$
$$\mathcal{R}\left\{\boldsymbol{A}\right\} = \left\{\boldsymbol{b} \in \mathbb{C}^m / \boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}\right\}$$

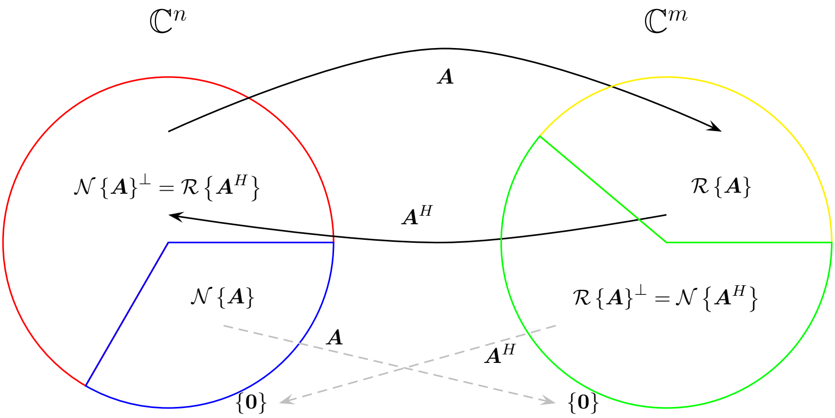
- The rank of $\boldsymbol{A}$ is defined as $\operatorname{rang}\left(\boldsymbol{A}\right) = \dim\left(\mathcal{R}\left\{\boldsymbol{A}\right\}\right) = \dim\left(\mathcal{R}\left\{\boldsymbol{A}^H\right\}\right)$.
- The **four subspaces** associated with $\boldsymbol{A}$ satisfy

$$\mathcal{N}\left\{\boldsymbol{A}\right\}^{\perp} = \mathcal{R}\left\{\boldsymbol{A}^H\right\}; \qquad \mathcal{R}\left\{\boldsymbol{A}\right\}^{\perp} = \mathcal{N}\left\{\boldsymbol{A}^H\right\}$$

  and, consequently,

$$\mathbb{C}^n = \mathcal{N}\left\{\boldsymbol{A}\right\} \oplus \mathcal{R}\left\{\boldsymbol{A}^H\right\}$$
$$\mathbb{C}^m = \mathcal{R}\left\{\boldsymbol{A}\right\} \oplus \mathcal{N}\left\{\boldsymbol{A}^H\right\}$$

# The subspaces associated with $\boldsymbol{A}$ et $\boldsymbol{A}^H$



- The *pseudo-inverse* of $\boldsymbol{A}$, $\boldsymbol{A}^{\#}$ (a matrix of size $n \times m$) is defined as:

$$\boldsymbol{x} \in \mathcal{R}\left\{\boldsymbol{A}^H\right\} \Rightarrow \boldsymbol{A}^{\#}\boldsymbol{A}\boldsymbol{x} = \boldsymbol{x}$$
$$\boldsymbol{x} \in \mathcal{N}\left\{\boldsymbol{A}^H\right\} \Rightarrow \boldsymbol{A}^{\#}\boldsymbol{x} = \boldsymbol{0}$$

  Therefore

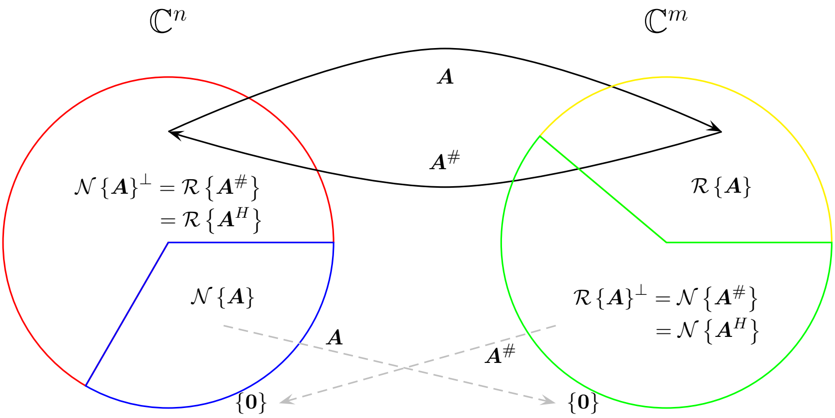
$$\mathcal{N}\left\{\boldsymbol{A}^{\#}\right\} = \mathcal{N}\left\{\boldsymbol{A}^H\right\} \qquad \mathcal{R}\left\{\boldsymbol{A}^{\#}\right\} = \mathcal{R}\left\{\boldsymbol{A}^H\right\}$$

- In the following cases, a direct expression can be obtained

$$\boldsymbol{A}^{\#} = \begin{cases} \left(\boldsymbol{A}^H\boldsymbol{A}\right)^{-1}\boldsymbol{A}^H & \text{if } \operatorname{rank}(\boldsymbol{A}) = n \\ \boldsymbol{A}^H\left(\boldsymbol{A}\boldsymbol{A}^H\right)^{-1} & \text{if } \operatorname{rank}(\boldsymbol{A}) = m \end{cases}$$

- $\boldsymbol{A}^{\#}$ appears naturally when it comes to solving $\boldsymbol{A}\boldsymbol{x} = \boldsymbol{b}$.

# Illustration of the pseudo-inverse $\boldsymbol{A}^{\#}$



# Singular Value Decomposition (SVD)

$\boldsymbol{A}$ can be decomposed as

$$\begin{aligned}
\boldsymbol{A} = \boldsymbol{U}\boldsymbol{\Sigma}\boldsymbol{V}^H &= \sum_{k=1}^{r} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^H \\
&= \begin{bmatrix} \underset{m|r}{\boldsymbol{U}_1} & \underset{m|m-r}{\boldsymbol{U}_2} \end{bmatrix} \begin{bmatrix} \underset{r|r}{\boldsymbol{\Sigma}_1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \begin{array}{c} \begin{bmatrix} \boldsymbol{V}_1^H \\ \boldsymbol{V}_2^H \end{bmatrix} \end{array} \begin{array}{l} r|n \\ n-r|n \end{array} \\
&= \boldsymbol{U}_1 \boldsymbol{\Sigma}_1 \boldsymbol{V}_1^H
\end{aligned}$$

where $\boldsymbol{U}(m \times m)$ and $\boldsymbol{V}(n \times n)$ are the **unitary** matrices of singular vectors, $\boldsymbol{\Sigma} = \mathrm{diag}\left\{\sigma_1, \sigma_2, \cdots, \sigma_r, 0, \cdots, 0\right\}$ is the quasi-diagonal matrix of singular values $(\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r)$ where $r$ stands for the rank of $\boldsymbol{A}$.

## SVD, subspaces and pseudo-inverse

- The SVD gives access to the $4$ **subspaces associated with $\boldsymbol{A}$** :

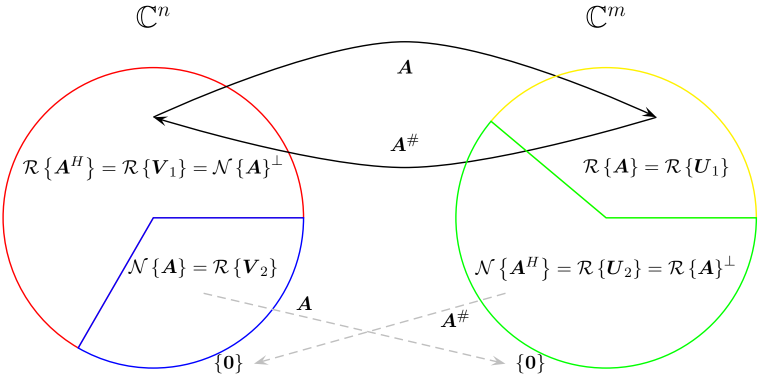
$$\begin{aligned}
\mathcal{N}\left\{\boldsymbol{A}\right\} &= \mathcal{R}\left\{\boldsymbol{V}_2\right\} \\
\mathcal{N}\left\{\boldsymbol{A}\right\}^{\perp} &= \mathcal{R}\left\{\boldsymbol{A}^H\right\} = \mathcal{R}\left\{\boldsymbol{V}_1\right\} \\
\mathcal{R}\left\{\boldsymbol{A}\right\} &= \mathcal{R}\left\{\boldsymbol{U}_1\right\} \\
\mathcal{R}\left\{\boldsymbol{A}\right\}^{\perp} &= \mathcal{N}\left\{\boldsymbol{A}^H\right\} = \mathcal{R}\left\{\boldsymbol{U}_2\right\}
\end{aligned}$$

- The pseudo-inverse can be written simply as

$$\boldsymbol{A}^{\#} = \boldsymbol{V}\boldsymbol{\Sigma}^{\#}\boldsymbol{U}^H = \sum_{k=1}^{r} \frac{1}{\sigma_k} \boldsymbol{v}_k \boldsymbol{u}_k^H = \boldsymbol{V}_1 \boldsymbol{\Sigma}_1^{-1} \boldsymbol{U}_1^H.$$

# The 4 subspaces associated with $\boldsymbol{A} \in \mathbb{C}^{m \times n}$

Let $\boldsymbol{A} = \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{0} \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1^H \\ \boldsymbol{V}_2^H \end{bmatrix} = \boldsymbol{U}_1 \boldsymbol{\Sigma}_1 \boldsymbol{V}_1^H$.



## Linear prediction equations (**noiseless case**)

Let $x(n) = \sum_{k=1}^{p} h_k z_k^n$ and assume that we wish to solve $\boldsymbol{X}\boldsymbol{a} = -\boldsymbol{h}$ with

$$\boldsymbol{X} = \begin{pmatrix} x(L-1) & x(L-2) & \cdots & x(0) \\ x(L) & x(L-1) & \cdots & x(1) \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ x(N-2) & x(N-3) & \cdots & x(N-L-1) \end{pmatrix}, \boldsymbol{h} = \begin{pmatrix} x(L) \\ x(L+1) \\ \vdots \\ \vdots \\ x(N-1) \end{pmatrix}$$

## Remarks

- the matrix $\boldsymbol{X}$ has rank $p$: every column after the $p$-th one is a linear combination of the first $p$ columns. $\mathcal{N}\left\{\boldsymbol{X}\right\}$ is of size $L-p$.
- $\boldsymbol{h} \in \mathcal{R}\left\{\boldsymbol{X}\right\} \Rightarrow \exists$ at least one solution.

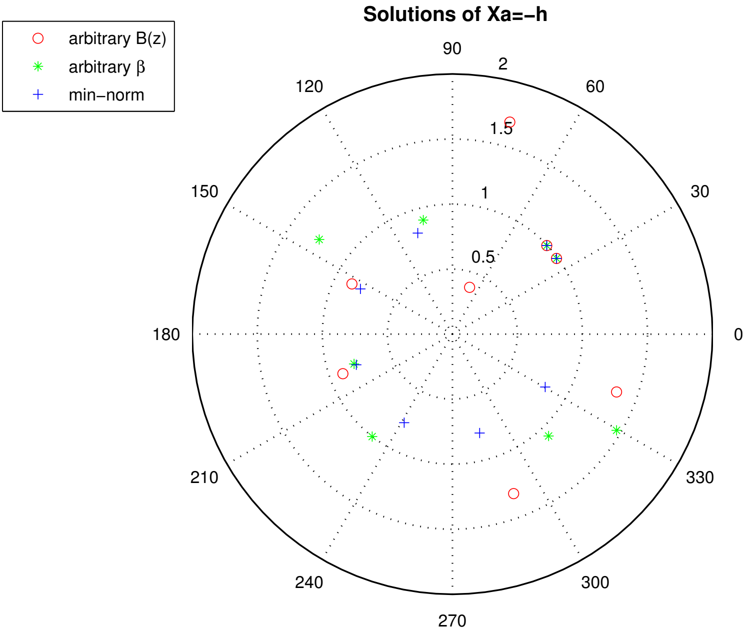
$\Rightarrow$ there exists **an infinite number of solutions** to the system.

# The solutions

The set of all possible solutions can be written in 2 ways :

1. If $A_p(z) = \sum_{k=0}^{p} a_k z^{-k} = \prod_{k=1}^{p}(1 - z_k z^{-1})$, then all solutions can be written as
$$\boxed{A(z) = A_p(z)B(z)}$$
where $B(z)$ is an **arbitrary** polynomial of degree $L - p$.

2. Let $\boldsymbol{X} = \boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1^H$ be the SVD of $\boldsymbol{X}$. Since $\boldsymbol{h} \in \mathcal{R}\{\boldsymbol{X}\} = \mathcal{R}\{\boldsymbol{U}_1\}$, one has $\boldsymbol{h} = \boldsymbol{U}_1\boldsymbol{U}_1^H\boldsymbol{h}$ and hence $\boldsymbol{a}_{\text{mn}} = -\boldsymbol{V}_1\boldsymbol{\Sigma}_1^{-1}\boldsymbol{U}_1^H\boldsymbol{h} = -\boldsymbol{X}^{\#}\boldsymbol{h}$ verifies
$$\boldsymbol{X}\boldsymbol{a}_{\text{mn}} = -\left[\boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{V}_1^H\right]\left[\boldsymbol{V}_1\boldsymbol{\Sigma}_1^{-1}\boldsymbol{U}_1^H\boldsymbol{h}\right] = -\boldsymbol{U}_1\boldsymbol{U}_1^H\boldsymbol{h} = -\boldsymbol{h}.$$
The set of solutions is given by
$$\boxed{-\boldsymbol{V}_1\boldsymbol{\Sigma}_1^{-1}\boldsymbol{U}_1^H\boldsymbol{h} + \boldsymbol{V}_2\boldsymbol{\beta}; \quad \boldsymbol{\beta} \in \mathbb{C}^{L-p}}$$
$\boldsymbol{a}_{\text{mn}}$ is **the minimum norm solution**. It ensures that all zeroes of $B(z)$ are strictly inside the unit circle.


Solutions of Xa=−h
arbitrary B(z)
arbitrary β
min−norm
90
2
60
120
1.5
1
30
150
0.5
180
0
210
330
240
300
270

## Linear prediction equations (**noisy case**)

If now $x(n) = \sum_{k=1}^{p} h_k z_k^n + w(n)$ then

- $\boldsymbol{X}$ is full-rank
- $\boldsymbol{h} \notin \mathcal{R}\left\{\boldsymbol{X}\right\}$

$\Rightarrow$ there is no solution to $\boldsymbol{X}\boldsymbol{a} = -\boldsymbol{h}$.

## Solution

One can

1. either solve in the least-squares sense, i.e., $\min_{\boldsymbol{a}} \left\|\boldsymbol{X}\boldsymbol{a} + \boldsymbol{h}\right\|^2$ (Prony).
2. or "recover" the noiseless case, viz that of a rank-deficient matrix $\boldsymbol{X}$ (Tufts-Kumaresan).

# Tufts-Kumaresan

## Principle

Let

$$\boldsymbol{X} = \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\Sigma}_1 & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{\Sigma}_2 \end{bmatrix} \begin{bmatrix} \boldsymbol{V}_1^H \\ \boldsymbol{V}_2^H \end{bmatrix} = \sum_{k=1}^{L} \sigma_k \boldsymbol{u}_k \boldsymbol{v}_k^H = \boldsymbol{U}_1 \boldsymbol{\Sigma}_1 \boldsymbol{V}_1^H + \boldsymbol{U}_2 \boldsymbol{\Sigma}_2 \boldsymbol{V}_2^H$$

where $\boldsymbol{U}_1 \in \mathbb{C}^{N-L \times p}$ and $\boldsymbol{V}_1 \in \mathbb{C}^{p \times L}$. Tufts and Kumaresan have proposed not to solve $\boldsymbol{X}\boldsymbol{a} = -\boldsymbol{h}$ but

$$\boldsymbol{X}_p \boldsymbol{a} = -\boldsymbol{h}$$

where $\boldsymbol{X}_p = \boldsymbol{U}_1 \boldsymbol{\Sigma}_1 \boldsymbol{V}_1^H$ is the best rank-$p$ approximant of $\boldsymbol{X}$.
Tufts-Kumaresan's method performs filtering of the least singular values and hence noise-cleaning of $x(n)$.

## Solution

Since $\boldsymbol{h} \notin \mathcal{R}\{\boldsymbol{X}_p\}$ there is no solution to $\boldsymbol{X}_p\boldsymbol{a} = -\boldsymbol{h}$. One can solve in a least-squares sense, i.e.,

$$\min_{\boldsymbol{a}} \|\boldsymbol{X}_p\boldsymbol{a} + \boldsymbol{h}\|^2$$

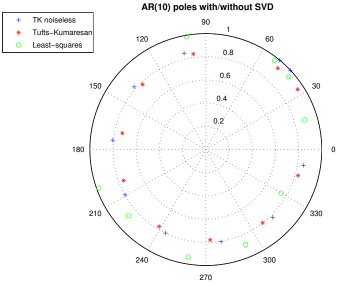
The solution is of the form $\boldsymbol{a} = \boldsymbol{V}_1\boldsymbol{\alpha}_1 + \boldsymbol{V}_2\boldsymbol{\alpha}_2$. However, $\mathcal{N}\{\boldsymbol{X}_p\} = \mathcal{R}\{\boldsymbol{V}_2\}$ and hence $\boldsymbol{X}_p\boldsymbol{a} = \boldsymbol{X}_p\boldsymbol{V}_1\boldsymbol{\alpha}_1 = \boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{\alpha}_1$. Consequently, $\boldsymbol{\alpha}_2$ has no influence on $\|\boldsymbol{X}_p\boldsymbol{a} + \boldsymbol{h}\|^2$. The minimum norm solution is thus obtained for $\boldsymbol{\alpha}_2 = \boldsymbol{0}$ and

$$\hat{\boldsymbol{\alpha}}_1 = \arg\min_{\boldsymbol{\alpha}_1} \|\boldsymbol{U}_1\boldsymbol{\Sigma}_1\boldsymbol{\alpha}_1 + \boldsymbol{h}\|^2 = -\boldsymbol{\Sigma}_1^{-1}\boldsymbol{U}_1^H\boldsymbol{h}.$$

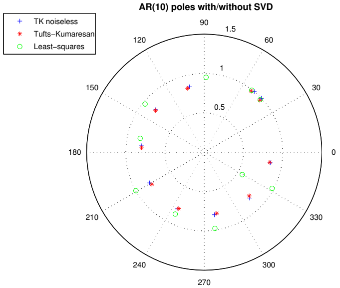
Finally

$$\boxed{\boldsymbol{a}_{\mathsf{TK}} = -\boldsymbol{V}_1\boldsymbol{\Sigma}_1^{-1}\boldsymbol{U}_1^H\boldsymbol{h} = -\boldsymbol{X}_p^{\#}\boldsymbol{h}.}$$

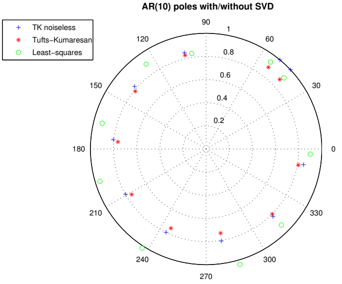
This is also the minimum norm solution to $\boldsymbol{X}_p\boldsymbol{a} = -\boldsymbol{U}_1\boldsymbol{U}_1^H\boldsymbol{h}$.



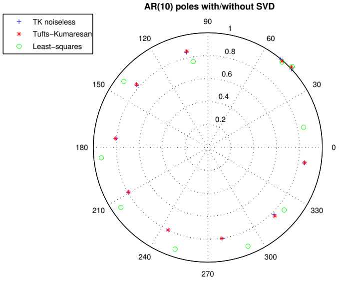

N=24, SNR=10dB




N=24, SNR=20dB




N=32, SNR=10dB




N=32, SNR=20dB

## A word on backward linear prediction

Let $x(n) = \sum_{k=1}^{p} h_k z_k^n$ and let

$$A^b(z) = \sum_{k=0}^{p} a_k^b z^{-k} = \prod_{k=1}^{p} \left(1 - e^{(\alpha_k + i2\pi f_k)} z^{-1}\right) = \prod_{k=1}^{p} \left(1 - \frac{1}{z_k^*} z^{-1}\right).$$

It can be shown that $x(n)$ verifies the backward linear prediction equations

$$\begin{aligned}
\sum_{k=0}^{p} a_k^b x^*(n+k) &= \sum_{k=0}^{p} a_k^b \left(\sum_{\ell=1}^{p} h_\ell (z_\ell^*)^{n+k}\right) \\
&= \sum_{\ell=1}^{p} h_\ell (z_\ell^*)^n \left(\sum_{k=0}^{p} a_k^b (z_\ell^*)^k\right) \\
&= \sum_{\ell=1}^{p} h_\ell (z_\ell^*)^n \left(\sum_{k=0}^{p} a_k^b \left(\frac{1}{z_\ell^*}\right)^{-k}\right) \\
&= 0.
\end{aligned}$$

## Backward linear prediction

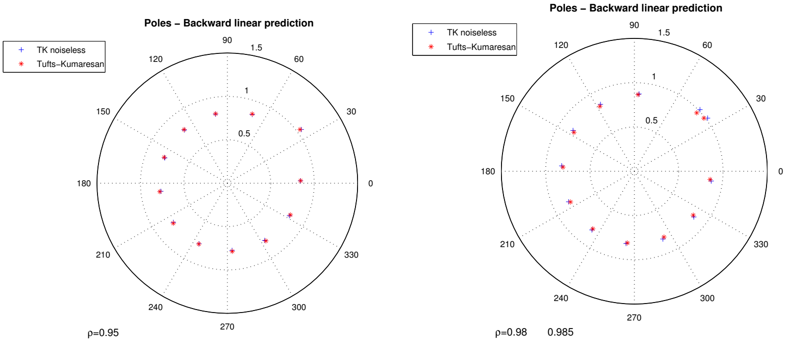
The minimum norm solution of $\boldsymbol{X}\boldsymbol{a} = -\boldsymbol{h}$ with

$$\boldsymbol{X} = \begin{pmatrix} x^*(1) & x^*(2) & \cdots & x^*(L) \\ x^*(2) & x^*(3) & \cdots & x^*(L+1) \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ x^*(N-L) & x^*(N-L+1) & \cdots & x^*(N-1) \end{pmatrix}, \boldsymbol{h} = \begin{pmatrix} x^*(0) \\ x^*(1) \\ \vdots \\ \vdots \\ x^*(N-L-1) \end{pmatrix}$$

results in a polynomial $A(z) = \sum_{k=0}^{p} a_k z^{-k}$ such that

- $p$ roots are located at $1/z_k^*$ (outside the unit circle)
- $L-p$ roots are strictly inside the unit circle.

$\Longrightarrow$ natural separation between poles due to signal and poles due to noise.


Poles – Backward linear prediction
TK noiseless
Tufts−Kumaresan
90
120
60
150
30
180
0
210
330
240
300
270
1.5
1
0.5
ρ=0.95
Poles – Backward linear prediction
TK noiseless
Tufts−Kumaresan
ρ=0.98
0.985

## Summary

- Estimation of damped complex exponentials is mainly based on minimizing the linear prediction error, a computationally more efficient solution than maximum likelihood.
- The linear prediction error minimization can be conducted in $2$ ways:
  1. conventional least-squares (Prony) which is equivalent to AR modeling.
  2. Tufts-Kumaresan's method which consists in filtering the least significant singular values so as to come close to the noiseless case.
- Tufts-Kumaresan's method is very performant but computationally intensive. Moreover, it needs a good signal to noise ratio and requires knowledge of the number of exponentials.

## Signal model

Let us consider a sum of complex exponential signals buried in noise:

$$x(n) = \sum_{k=1}^{p} A_k e^{i\phi_k} e^{i2\pi n f_k} + w(n) \qquad n = 0, \cdots, N-1$$

where $\phi_k$ is uniformly distributed on $[0, 2\pi[$ and independent of $\phi_\ell$, $w(n)$ is assumed to be a white noise with variance $\sigma^2 = \mathcal{E}\left\{w^*(n)w(n)\right\}$. One is interested in estimating $f_k$ (or equivalently $\omega_k = 2\pi f_k$).

## Correlation function

The correlation function is given by

$$
\begin{aligned}
r_{xx}(m) &= \mathcal{E}\left\{x^*(n)x(n+m)\right\} \\
&= \mathcal{E}\left\{\left[\sum_{k=1}^{p} A_k e^{-i\phi_k} e^{-in\omega_k} + w^*(n)\right]\left[\sum_{\ell=1}^{p} A_\ell e^{i\phi_\ell} e^{i(n+m)\omega_\ell} + w(n+m)\right]\right\} \\
&= \sum_{k=1}^{p} P_k e^{im\omega_k} + \sigma^2\delta(m)
\end{aligned}
$$

with $P_k = |A_k|^2$.

## Correlation matrix

Let us define the following matrix

$$\begin{aligned}\boldsymbol{R} &= \begin{pmatrix} r_{xx}(0) & r_{xx}(-1) & \cdots & \cdots & r_{xx}(-M+1) \\ r_{xx}(1) & r_{xx}(0) & \cdots & \cdots & r_{xx}(-M+2) \\ \vdots & \vdots & \ddots & & \vdots \\ \vdots & \vdots & & \ddots & \vdots \\ r_{xx}(M-1) & r_{xx}(M-2) & \cdots & \cdots & r_{xx}(0) \end{pmatrix} \\ &= \sum_{k=1}^{p} P_k \boldsymbol{a}_k \boldsymbol{a}_k^H + \sigma^2 \boldsymbol{I} = \boldsymbol{A}(\boldsymbol{\omega}) \boldsymbol{P} \boldsymbol{A}(\boldsymbol{\omega})^H + \sigma^2 \boldsymbol{I} \\ &= \boldsymbol{R}_s + \sigma^2 \boldsymbol{I}\end{aligned}$$

where $\boldsymbol{a}_k = \begin{bmatrix} 1 & e^{i\omega_k} & \cdots & e^{i(M-1)\omega_k} \end{bmatrix}^T$, $\boldsymbol{A}(\boldsymbol{\omega}) = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_p \end{bmatrix}$, $\boldsymbol{\omega} = \begin{bmatrix} \omega_1 & \omega_2 & \cdots & \omega_p \end{bmatrix}^T$ and $\boldsymbol{P} = \operatorname{diag}(P_1, P_2, \cdots, P_p)$.

# Properties of $\boldsymbol{R}_s$

- One have

$$\boldsymbol{R}_s \boldsymbol{\alpha} = \sum_{k=1}^{p} P_k \left( \boldsymbol{a}_k^H \boldsymbol{\alpha} \right) \boldsymbol{a}_k$$

and hence $\mathcal{R}\left\{\boldsymbol{R}_s\right\} = \mathcal{R}\left\{\boldsymbol{A}(\boldsymbol{\omega})\right\}$. Consequently, assuming vectors $\boldsymbol{a}_k$ are linearly independent, it follows that $\mathrm{rank}\left(\boldsymbol{R}_s\right) = p$.

- The eigenvalue decomposition of $\boldsymbol{R}_s$ can thus be written as

$$\boldsymbol{R}_s = \sum_{k=1}^{p} \lambda_k^s \boldsymbol{u}_k \boldsymbol{u}_k^H + 0 \sum_{k=p+1}^{M} \boldsymbol{u}_k \boldsymbol{u}_k^H = \boldsymbol{U}_s \boldsymbol{\Lambda}_s^s \boldsymbol{U}_s^H + \boldsymbol{U}_n \boldsymbol{0} \boldsymbol{U}_n^H$$

where $\begin{bmatrix} \boldsymbol{U}_s & \boldsymbol{U}_n \end{bmatrix}$ is the orthogonal basis of eigenvectors. Therefore,

$$\boxed{\mathcal{R}\left\{\boldsymbol{R}_s\right\} = \mathcal{R}\left\{\boldsymbol{U}_s\right\}; \quad \mathcal{N}\left\{\boldsymbol{R}_s\right\} = \mathcal{R}\left\{\boldsymbol{U}_n\right\}}$$

# Properties of $\boldsymbol{R}$

- The eigenvalue decomposition (EVD) of $\boldsymbol{R}$ follows from that of $\boldsymbol{R}_s$:

$$\begin{aligned}
\boldsymbol{R} &= \boldsymbol{R}_s + \sigma^2 \boldsymbol{I} \\
&= \boldsymbol{U}_s \boldsymbol{\Lambda}_s^s \boldsymbol{U}_s^H + \boldsymbol{U}_n \boldsymbol{0} \boldsymbol{U}_n^H + \sigma^2 \boldsymbol{I} \\
&= \boldsymbol{U}_s \boldsymbol{\Lambda}_s^s \boldsymbol{U}_s^H + \boldsymbol{U}_n \boldsymbol{0} \boldsymbol{U}_n^H + \sigma^2 \left( \boldsymbol{U}_s \boldsymbol{U}_s^H + \boldsymbol{U}_n \boldsymbol{U}_n^H \right) \\
&= \boldsymbol{U}_s \left( \boldsymbol{\Lambda}_s^s + \sigma^2 \boldsymbol{I}_p \right) \boldsymbol{U}_s^H + \sigma^2 \boldsymbol{U}_n \boldsymbol{U}_n^H \\
&= \boldsymbol{U}_s \boldsymbol{\Lambda}_s \boldsymbol{U}_s^H + \sigma^2 \boldsymbol{U}_n \boldsymbol{U}_n^H .
\end{aligned}$$

- The EVD gives access to $2$ subspaces:

$$\begin{aligned}
\mathcal{R}\left\{\boldsymbol{U}_s\right\} &= \mathcal{R}\left\{\boldsymbol{A}(\boldsymbol{\omega})\right\} \\
\mathcal{R}\left\{\boldsymbol{U}_n\right\} &= \mathcal{N}\left\{\boldsymbol{R}_s\right\} \perp \mathcal{R}\left\{\boldsymbol{A}(\boldsymbol{\omega})\right\}
\end{aligned}$$

## Subspace methods

Subspace-based methods exploit the fact that the correlation matrix can be decomposed into a **"signal" subspace** (corresponding to largest eigenvalues) which coincides with the subspace spanned by the exponential signals, and a **"noise" subspace** orthogonal to the signal subspace.

$\boldsymbol{\omega}$ can thus be estimated from

1. either $\boldsymbol{U}_s$ using the fact that $\boldsymbol{U}_s = \boldsymbol{A}(\boldsymbol{\omega})\boldsymbol{T} \Rightarrow$ ESPRIT.
2. or $\boldsymbol{U}_n$ using the fact that $\mathcal{R}\left\{\boldsymbol{U}_n\right\} \perp \mathcal{R}\left\{\boldsymbol{A}(\boldsymbol{\omega})\right\}$, or equivalently

$$\boldsymbol{a}^H(\omega_\ell)\left(\sum_{k=p+1}^{M} \alpha_k \boldsymbol{u}_k\right) = 0 \qquad \forall \ell \in [1, p]\,,\ \forall \alpha_k,\ k \in [p+1, M]$$

$\Rightarrow$ MUSIC.

# Relation with array processing

- The above result bears much resemblance with array processing since matrix $\boldsymbol{R}$ above shares the same algebraic properties as the spatial covariance matrix of $p$ signals impinging on a uniform linear array of $M$ antennas.
- This relation is better highlighted using the "pseudo-snapshot"

$$\begin{aligned}\boldsymbol{x}(n) &= \begin{bmatrix} x(n) & x(n+1) & \cdots & x(n+M) \end{bmatrix}^T \\ &= \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_p \end{bmatrix} \begin{bmatrix} A_1 e^{i\phi_1} e^{in\omega_1} \\ A_2 e^{i\phi_2} e^{in\omega_2} \\ \vdots \\ A_p e^{i\phi_p} e^{in\omega_p} \end{bmatrix} + \begin{bmatrix} b(n) \\ b(n+1) \\ \vdots \\ b(n+M) \end{bmatrix} \\ &= \boldsymbol{A}(\boldsymbol{\omega})\boldsymbol{s}(n) + \boldsymbol{b}(n)\end{aligned}$$

whose covariance matrix is $\mathcal{E}\left\{\boldsymbol{x}(n)\boldsymbol{x}^H(n)\right\} = \boldsymbol{R}$. Yet, the snapshots $\boldsymbol{x}(n)$ are not independent here.

# MUSIC

- MUSIC relies on the orthogonality between the noise (minor) eigenvectors and the exponential signals, i.e. $\boldsymbol{u}_k \perp \boldsymbol{a}_\ell$, for $k = p+1 \cdots, M$ and $\ell = 1, \cdots, p$. It is based on the following pseudo-spectrum

$$P_{\text{MUSIC}}(\omega) = \frac{1}{\boldsymbol{a}^H(\omega)\boldsymbol{U}_n\boldsymbol{U}_n^H\boldsymbol{a}(\omega)}$$

  by observing that $P_{\text{MUSIC}}(\omega_\ell) = \infty$ for $\ell = 1, \cdots p$.

- In practice $\boldsymbol{R}$ and hence $\boldsymbol{U}_n$ are estimated and one looks for the locations of the $p$ largest peaks in

$$P_{\text{MUSIC}}(\omega) = \frac{1}{\boldsymbol{a}^H(\omega)\hat{\boldsymbol{U}}_n\hat{\boldsymbol{U}}_n^H\boldsymbol{a}(\omega)}$$

## Remarks

- $\boldsymbol{U}_n \boldsymbol{U}_n^H$ is the projection matrix onto the noise subspace: hence, one looks for the exponentials whose projection onto the noise subspace has minimum norm.
- The pseudo-spectrum can be rewritten as

$$P_{\mathsf{MUSIC}}(\omega) = \frac{1}{\sum_{k=p+1}^{M} \left|\boldsymbol{a}^H(\omega)\hat{\boldsymbol{u}}_k\right|^2}$$

  and $\boldsymbol{a}^H(\omega)\hat{\boldsymbol{u}}_k$ corresponds to the Fourier transform of $\hat{\boldsymbol{u}}_k \Rightarrow$ possibly use FFT for computational gain.
- The pseudo-spectrum can alternatively be rewritten as

$$\begin{aligned} P_{\mathsf{MUSIC}}(\omega) &= \frac{1}{M - \boldsymbol{a}^H(\omega)\hat{\boldsymbol{U}}_s \hat{\boldsymbol{U}}_s^H \boldsymbol{a}(\omega)} \\ &= \frac{1}{M - \sum_{k=1}^{p} \left|\boldsymbol{a}^H(\omega)\hat{\boldsymbol{u}}_k\right|^2} \end{aligned}$$

# Root-MUSIC

- An alternative solution consists in finding the roots of the polynomial

$$P(z) = \boldsymbol{a}^T(z^{-1})\hat{\boldsymbol{U}}_n\hat{\boldsymbol{U}}_n^H\boldsymbol{a}(z)$$

with $\boldsymbol{a}(z) = \begin{bmatrix} 1 & z & \cdots & z^{M-1} \end{bmatrix}^T$.

- This polynomial of degree $2M-1$ verifies

$$P^*\left(1/z^*\right) = \left[\boldsymbol{a}^T(z^*)\hat{\boldsymbol{U}}_n\hat{\boldsymbol{U}}_n^H\boldsymbol{a}(1/z^*)\right]^* = P(z)$$

$\Rightarrow P(z)$ has $(M-1)$ roots $z_k$ inside the unit circle and $(M-1)$ roots $1/z_k^*$. Moreover, if $\hat{\boldsymbol{U}}_n$ is replaced by $\boldsymbol{U}_n$, then $P(e^{i\omega_k}) = 0$.

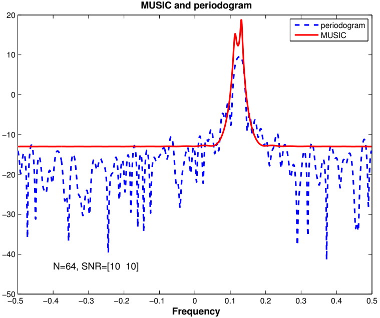
- In practice, $\boldsymbol{\omega}$ is estimated by picking the $p$ roots of $P(z)$ closest (and inside) the unit circle.

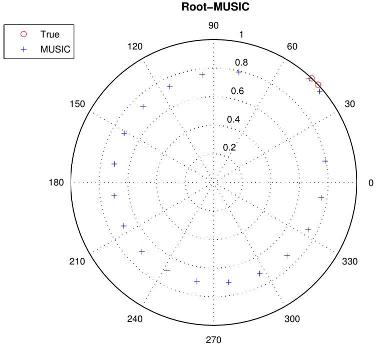
# Variations

- When $M = p + 1$ there is only one eigenvector in the noise subspace and one may look for the roots of $H(z) = \boldsymbol{a}^T(z^{-1})\boldsymbol{u}_{p+1}$ which are closest to the unit circle: this is referred to as Pisarenko's method.
- The pseudo-spectrum can be modified to

$$P(\omega) = \frac{1}{\sum_{k=p+1}^{M} w_k \left|\boldsymbol{a}^H(\omega)\hat{\boldsymbol{u}}_k\right|^2}$$

  where $w_{p+1} \leq w_{p+2} \leq \cdots \leq w_M$ in order to give more weight to the smallest eigenvectors (since we are pretty sure they belong to the noise subspace). For instance, one may select $w_k = \hat{\lambda}_k^{-1}$.
- Instead of using all $M - p$ noise eigenvectors, another method consists in finding the vector $\boldsymbol{d}$ with minimum norm (and such that $d_1 = 1$) which belongs to the noise subspace: this is referred to as **min-norm** method, which is closely related to Tufts-Kumaresan's method presented above.


MUSIC and periodogram
periodogram
MUSIC
N=64, SNR=[10 10]
Frequency


Root−MUSIC
True
MUSIC

# ESPRIT

- ESPRIT uses the fact that the subspaces spanned $\boldsymbol{U}_s$ and $\boldsymbol{A}(\boldsymbol{\omega})$ are identical, viz. $\boldsymbol{U}_s = \boldsymbol{A}(\boldsymbol{\omega})\boldsymbol{T}$.
- One can write

$$\boldsymbol{A}(\boldsymbol{\omega}) = \begin{pmatrix} 1 & 1 & \cdots & \cdots & 1 \\ e^{i\omega_1} & e^{i\omega_2} & \cdots & \cdots & e^{i\omega_p} \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ e^{i(M-1)\omega_1} & e^{i(M-1)\omega_2} & \cdots & \cdots & e^{i(M-1)\omega_p} \end{pmatrix}$$

$$\triangleq \begin{pmatrix} \boldsymbol{A}_1 \\ - \end{pmatrix} \triangleq \begin{pmatrix} - \\ \boldsymbol{A}_2 \end{pmatrix}$$

- Observe that

$$\boldsymbol{A}_2 = \boldsymbol{A}_1 \boldsymbol{\Phi} = \boldsymbol{A}_1 \begin{pmatrix} e^{i\omega_1} & & & & \\ & e^{i\omega_2} & & & \\ & & \ddots & & \\ & & & \ddots & \\ & & & & e^{i\omega_p} \end{pmatrix}$$

- Now, if we partition $\boldsymbol{U}_s$ as

$$\boldsymbol{U}_s = \begin{pmatrix} \boldsymbol{U}_{s1} \\ - \end{pmatrix} = \begin{pmatrix} - \\ \boldsymbol{U}_{s2} \end{pmatrix}$$

is there a similar relation between $\boldsymbol{U}_{s1}$ and $\boldsymbol{U}_{s2}$, knowing that $\boldsymbol{U}_s = \boldsymbol{A}(\boldsymbol{\omega})\boldsymbol{T}$?

- One has

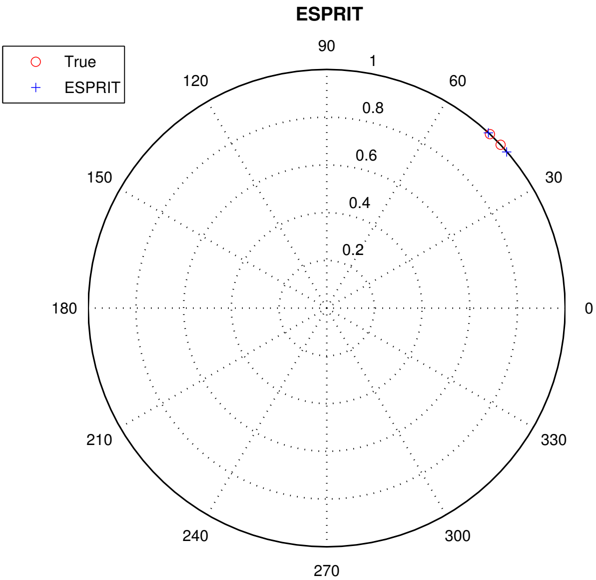
$$\boldsymbol{U}_{s2} = \boldsymbol{A}_2\boldsymbol{T} = \boldsymbol{A}_1\boldsymbol{\Phi}\boldsymbol{T} = \boldsymbol{U}_{s1}\boldsymbol{T}^{-1}\boldsymbol{\Phi}\boldsymbol{T} = \boldsymbol{U}_{s1}\boldsymbol{\Psi}.$$

  The matrices $\boldsymbol{\Phi}$ and $\boldsymbol{\Psi}$ share the same eigenvalues, namely $e^{i\omega_k}$!

- In practice, there is no matrix $\boldsymbol{\Psi}$ which satisfies $\hat{\boldsymbol{U}}_{s2} = \hat{\boldsymbol{U}}_{s1}\boldsymbol{\Psi}$. $\boldsymbol{\Psi}$ is then estimated using a least-squares approach as

$$\hat{\boldsymbol{\Psi}} = \arg\min_{\boldsymbol{\Psi}} \left\| \hat{\boldsymbol{U}}_{s2} - \hat{\boldsymbol{U}}_{s1}\boldsymbol{\Psi} \right\|^2 = \left( \hat{\boldsymbol{U}}_{s1}^H \hat{\boldsymbol{U}}_{s1} \right)^{-1} \hat{\boldsymbol{U}}_{s1}^H \hat{\boldsymbol{U}}_{s2}$$

  from which the eigenvalues $e^{i\hat{\omega}_k}$ of $\hat{\boldsymbol{\Psi}}$ are obtained.


ESPRIT
True
ESPRIT
90
120
60
150
30
180
0
210
330
240
300
270
1
0.8
0.6
0.4
0.2

N=64, SNR=[10 10]

## Summary

- Subspace-based methods enable one to estimate the frequencies of noisy exponential signals with *high resolution*.
- They rely on the partitioning between the subspace spanned by the exponentials and the orthogonal subspace, both of which being obtained from EVD of the correlation matrix.
- Drawbacks :
  1. high computational complexity (EVD).
  2. require knowledge of the number of exponential signals.
  3. require a high signal to noise ratio.

# References


- F. Castanié (Editeur), *Analyse Spectrale*, Hermès Lavoisier, 2003
- S. M. Kay, *Modern Spectral Estimation: Theory and Application*, Prentice Hall, Englewood Cliffs, NJ, 1988
- S. M. Kay, *Fundamentals of Statistical Signal Processing - Estimation Theory*, Prentice Hall, Upper Saddle River, NJ, 1993
- B. Porat, *Digital Processing of Random Signals*, Prentice Hall, Englewood Cliffs, NJ, 1994
- P. Stoica, R. L. Moses, *Spectral Analysis of Signals*, Pearson Prentice Hall, Upper Saddle River, NJ, 2005