Basic Definitions

and

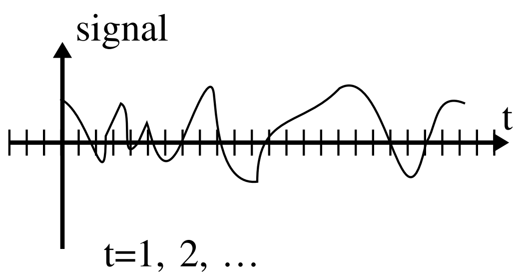
The Spectral Estimation Problem

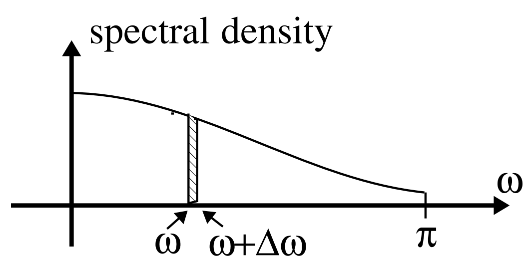
Lecture 1

Informal Definition of Spectral Estimation

Given: A finite record of a signal.

Determine: The distribution of signal power over frequency.





$$\begin{aligned} \omega &= \text{(angular) frequency in radians/(sampling interval)} \\ f &= \omega/2\pi = \text{frequency in cycles/(sampling interval)} \end{aligned}$$

Applications

Temporal Spectral Analysis

- Vibration monitoring and fault detection
- Hidden periodicity finding
- Speech processing and audio devices
- Medical diagnosis
- Seismology and ground movement study
- Control systems design
- Radar, Sonar

Spatial Spectral Analysis

- Source location using sensor arrays

Deterministic Signals

$\{y(t)\}_{t=-\infty}^{\infty} =$ discrete-time deterministic data sequence

If: $$\sum_{t=-\infty}^{\infty} |y(t)|^2 < \infty$$

Then: $$\boxed{Y(\omega) = \sum_{t=-\infty}^{\infty} y(t)e^{-i\omega t}}$$

exists and is called the **Discrete-Time Fourier Transform (DTFT)**

Energy Spectral Density

Parseval's Equality:

$$\sum_{t=-\infty}^{\infty} |y(t)|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} S(\omega)d\omega$$

where

$$\boxed{\begin{aligned} S(\omega) &\triangleq |Y(\omega)|^2 \\ &= \text{Energy Spectral Density} \end{aligned}}$$

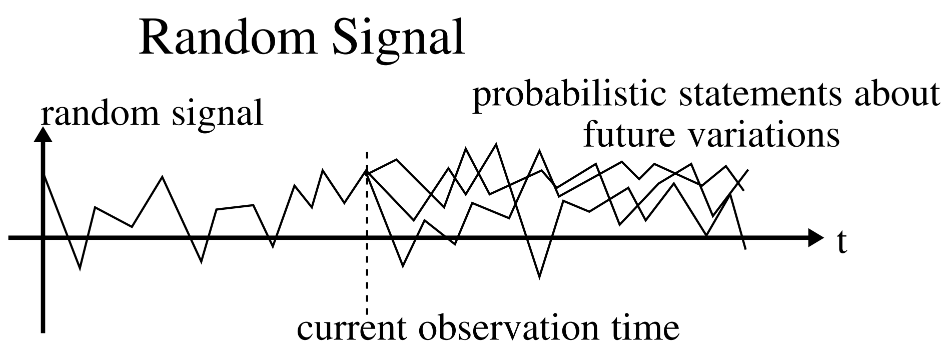
We can write

$$\boxed{S(\omega) = \sum_{k=-\infty}^{\infty} \rho(k)e^{-i\omega k}}$$

where

$$\rho(k) = \sum_{t=-\infty}^{\infty} y(t)y^*(t-k)$$

Random Signals



Here: $$\sum_{t=-\infty}^{\infty} |y(t)|^2 = \infty$$

But: $$\boxed{E\left\{|y(t)|^2\right\} < \infty}$$

$E\left\{\cdot\right\}$ = Expectation over the ensemble of realizations

$E\left\{|y(t)|^2\right\}$ = Average power in $y(t)$

PSD = (Average) power spectral density

First Definition of PSD

$$\boxed{\phi(\omega) = \sum_{k=-\infty}^{\infty} r(k)e^{-i\omega k}}$$

where $r(k)$ is the **autocovariance sequence (ACS)**

$$r(k) = E\left\{y(t)y^*(t-k)\right\}$$

$$r(k) = r^*(-k), \qquad r(0) \geq |r(k)|$$

Note that

$$r(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi(\omega)e^{i\omega k}d\omega \quad \text{(Inverse DTFT)}$$

Interpretation:

$$r(0) = E\left\{|y(t)|^2\right\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi(\omega)d\omega$$

so

$\phi(\omega)d\omega =$ infinitesimal signal power in the band $\omega \pm \frac{d\omega}{2}$

Second Definition of PSD

$$\boxed{\phi(\omega) = \lim_{N\to\infty} E\left\{\frac{1}{N}\left|\sum_{t=1}^{N} y(t)e^{-i\omega t}\right|^2\right\}}$$

Note that

$$\phi(\omega) = \lim_{N\to\infty} E\left\{\frac{1}{N}|Y_N(\omega)|^2\right\}$$

where

$$Y_N(\omega) = \sum_{t=1}^{N} y(t)e^{-i\omega t}$$

is the finite DTFT of $\{y(t)\}$.

Properties of the PSD

P1: $\phi(\omega) = \phi(\omega + 2\pi)$ for all ω.

Thus, we can restrict attention to

$$\omega \in [-\pi, \pi] \iff f \in [-1/2, 1/2]$$

P2: $\phi(\omega) \geq 0$

P3: If $y(t)$ is real,

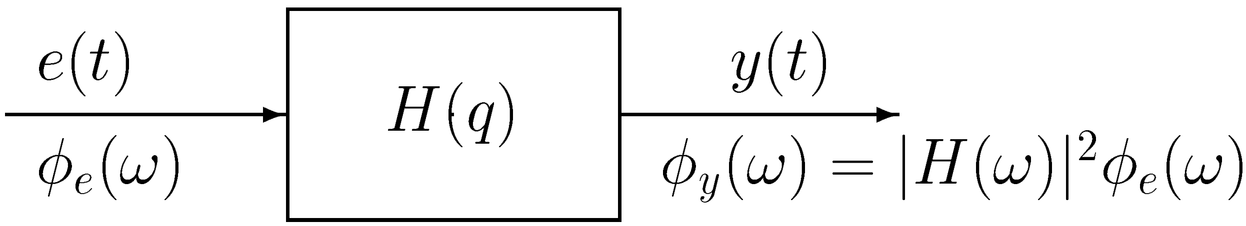
Then: $\phi(\omega) = \phi(-\omega)$

Otherwise: $\phi(\omega) \neq \phi(-\omega)$

Transfer of PSD Through Linear Systems

System Function: $H(q) = \sum_{k=0}^{\infty} h_k q^{-k}$

where q^{-1} = unit delay operator: $q^{-1}y(t) = y(t-1)$



Then

$$y(t) = \sum_{k=0}^{\infty} h_k \, e(t-k)$$

$$H(\omega) = \sum_{k=0}^{\infty} h_k \, e^{-i\omega k}$$

$$\boxed{\phi_y(\omega) = |H(\omega)|^2 \phi_e(\omega)}$$

The Spectral Estimation Problem

The Problem:

From a sample $\{y(1), \ldots, y(N)\}$

Find an estimate of $\phi(\omega)$: $\{\hat{\phi}(\omega),\ \omega \in [-\pi, \pi]\}$

Two Main Approaches :

- **Nonparametric:**
  - Derived from the PSD definitions.
- **Parametric:**
  - Assumes a parameterized functional form of the PSD

Periodogram and Correlogram Methods

Lecture 2

Periodogram

Recall 2nd definition of $\phi(\omega)$:

$$\phi(\omega) = \lim_{N\to\infty} E\left\{ \frac{1}{N} \left| \sum_{t=1}^{N} y(t) e^{-i\omega t} \right|^2 \right\}$$

Given : $\{y(t)\}_{t=1}^{N}$

Drop "$\lim_{N\to\infty}$" and "$E\{\cdot\}$" to get

$$\boxed{\hat{\phi}_p(\omega) = \frac{1}{N} \left| \sum_{t=1}^{N} y(t) e^{-i\omega t} \right|^2}$$

- Natural estimator
- Used by Schuster ($\sim$1900) to determine "hidden periodicities" (hence the name).

Correlogram

Recall 1st definition of $\phi(\omega)$:

$$\phi(\omega) = \sum_{k=-\infty}^{\infty} r(k) e^{-i\omega k}$$

Truncate the "$\sum$" and replace "$r(k)$" by "$\hat{r}(k)$":

$$\boxed{\hat{\phi}_c(\omega) = \sum_{k=-(N-1)}^{N-1} \hat{r}(k) e^{-i\omega k}}$$

Covariance Estimators
(or Sample Covariances)

Standard unbiased estimate:

$$\hat{r}(k) = \frac{1}{N-k} \sum_{t=k+1}^{N} y(t) y^*(t-k), \quad k \geq 0$$

Standard biased estimate:

$$\hat{r}(k) = \frac{1}{N} \sum_{t=k+1}^{N} y(t) y^*(t-k), \quad k \geq 0$$

For both estimators:

$$\hat{r}(k) = \hat{r}^*(-k), \quad k < 0$$

Relationship Between $\hat{\phi}_p(\omega)$ and $\hat{\phi}_c(\omega)$

If: the biased ACS estimator $\hat{r}(k)$ is used in $\hat{\phi}_c(\omega)$,

Then:

$$
\begin{aligned}
\hat{\phi}_p(\omega) &= \frac{1}{N}\left|\sum_{t=1}^{N} y(t)e^{-i\omega t}\right|^2 \\
&= \sum_{k=-(N-1)}^{N-1} \hat{r}(k)e^{-i\omega k} \\
&= \hat{\phi}_c(\omega)
\end{aligned}
$$

$$\boxed{\hat{\phi}_p(\omega) = \hat{\phi}_c(\omega)}$$

Consequence:

Both $\hat{\phi}_p(\omega)$ and $\hat{\phi}_c(\omega)$ can be analyzed simultaneously.

Statistical Performance of $\hat{\phi}_p(\omega)$ and $\hat{\phi}_c(\omega)$

Summary:

- Both are asymptotically (for large N) unbiased:

$$E\left\{\widehat{\phi}_p(\omega)\right\} \to \phi(\omega) \text{ as } N \to \infty$$

- Both have "large" variance, even for large N.

Thus, $\widehat{\phi}_p(\omega)$ and $\widehat{\phi}_c(\omega)$ have **poor performance**.

Intuitive explanation:

- $\widehat{r}(k) - r(k)$ may be large for large $|k|$

- Even if the errors $\{\widehat{r}(k) - r(k)\}_{|k|=0}^{N-1}$ are small, there are "so many" that when summed in $[\widehat{\phi}_p(\omega) - \phi(\omega)]$, the PSD error is large.

Bias Analysis of the Periodogram

$$
\begin{aligned}
E\left\{\hat{\phi}_p(\omega)\right\} &= E\left\{\hat{\phi}_c(\omega)\right\} = \sum_{k=-(N-1)}^{N-1} E\left\{\hat{r}(k)\right\} e^{-i\omega k} \\
&= \sum_{k=-(N-1)}^{N-1} \left(1 - \frac{|k|}{N}\right) r(k) e^{-i\omega k} \\
&= \sum_{k=-\infty}^{\infty} w_B(k) r(k) e^{-i\omega k}
\end{aligned}
$$

$$
\begin{aligned}
w_B(k) &= \begin{cases} \left(1 - \frac{|k|}{N}\right), & |k| \le N-1 \\ 0, & |k| \ge N \end{cases} \\
&= \text{Bartlett, or triangular, window}
\end{aligned}
$$

Thus,

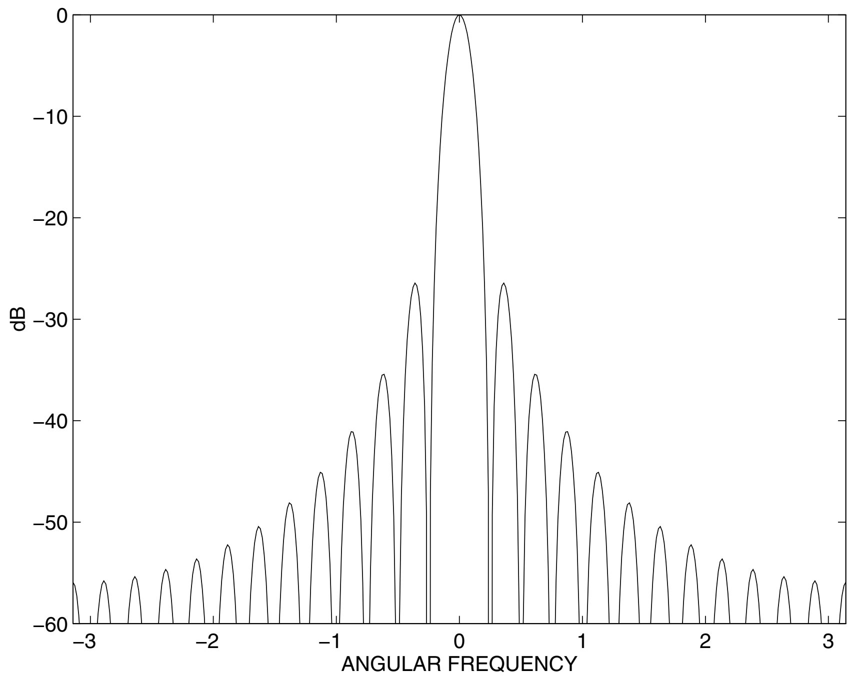
$$
\boxed{E\left\{\hat{\phi}_p(\omega)\right\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} \phi(\zeta) W_B(\omega - \zeta)\, d\zeta}
$$

Ideally: $W_B(\omega) =$ Dirac impulse $\delta(\omega)$.

Bartlett Window $W_B(\omega)$

$$W_B(\omega) = \frac{1}{N}\left[\frac{\sin(\omega N/2)}{\sin(\omega/2)}\right]^2$$

$W_B(\omega)/W_B(0)$, for $N = 25$



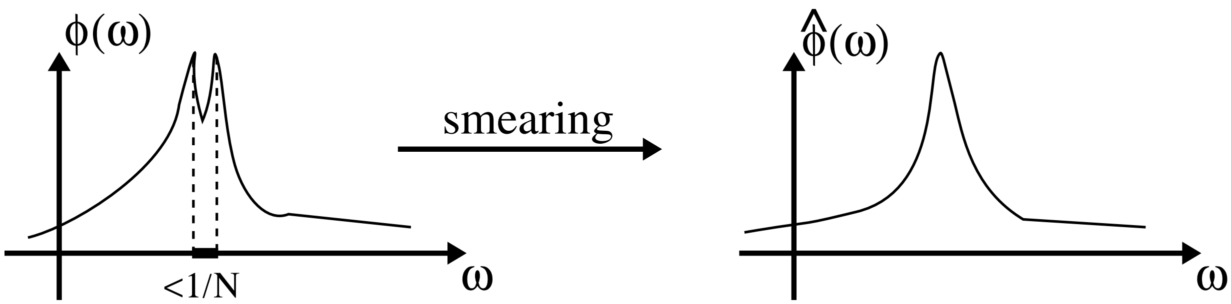
$\boxed{\text{Main lobe 3dB width} \sim 1/N.}$

For "small" N, $W_B(\omega)$ may differ quite a bit from $\delta(\omega)$.

Smearing and Leakage

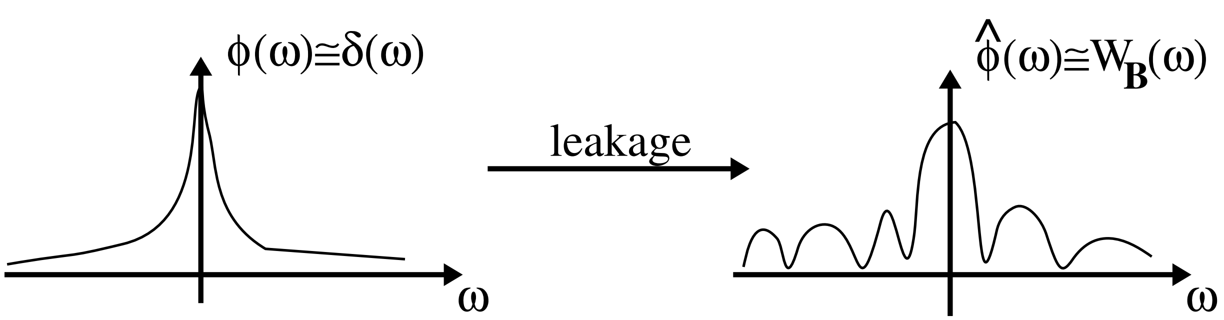
Main Lobe Width: *smearing* or *smoothing*

Details in $\phi(\omega)$ separated in f by less than $1/N$ are not resolvable.



Thus: $\boxed{\text{Periodogram resolution limit} = 1/N.}$

Sidelobe Level: *leakage*



Periodogram Bias Properties

Summary of Periodogram Bias Properties:

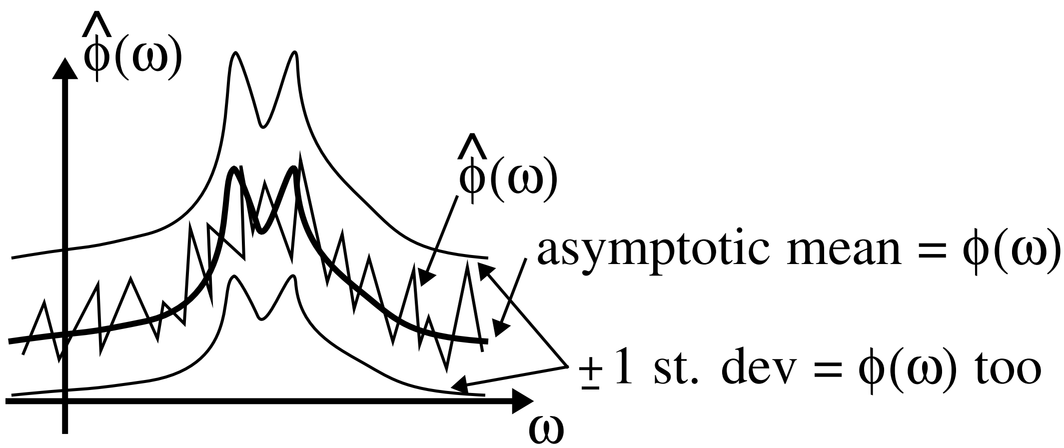
- For "small" N, severe bias
- As $N \to \infty$, $W_B(\omega) \to \delta(\omega)$,
  so $\hat{\phi}(\omega)$ is asymptotically unbiased.

Periodogram Variance

As $N \to \infty$

$$\boxed{\begin{gathered} E\left\{\left[\hat{\phi}_p(\omega_1) - \phi(\omega_1)\right]\left[\hat{\phi}_p(\omega_2) - \phi(\omega_2)\right]\right\} \\ = \begin{cases} \phi^2(\omega_1), & \omega_1 = \omega_2 \\ 0, & \omega_1 \neq \omega_2 \end{cases} \end{gathered}}$$

- Inconsistent estimate
- Erratic behavior



Resolvability properties depend on *both* bias and variance.

Discrete Fourier Transform (DFT)

Finite DTFT: $Y_N(\omega) = \sum_{t=1}^{N} y(t) e^{-i\omega t}$

Let $\omega = \frac{2\pi}{N}k$ and $W = e^{-i\frac{2\pi}{N}}$.

Then $Y_N(\frac{2\pi}{N}k)$ is the Discrete Fourier Transform (DFT):

$$Y(k) = \sum_{t=1}^{N} y(t) W^{tk}, \qquad k = 0, \ldots, N-1$$

Direct computation of $\{Y(k)\}_{k=0}^{N-1}$ from $\{y(t)\}_{t=1}^{N}$:
$O(N^2)$ flops

Radix–2 Fast Fourier Transform (FFT)

Assume: $N = 2^m$

$$
\begin{aligned}
Y(k) &= \sum_{t=1}^{N/2} y(t) W^{tk} + \sum_{t=N/2+1}^{N} y(t) W^{tk} \\
&= \sum_{t=1}^{N/2} [y(t) + y(t+N/2) W^{\frac{Nk}{2}}] W^{tk}
\end{aligned}
$$

with $W^{\frac{Nk}{2}} = \begin{cases} +1, & \text{for even } k \\ -1, & \text{for odd } k \end{cases}$

Let $\tilde{N} = N/2$ and $\tilde{W} = W^2 = e^{-i2\pi/\tilde{N}}$.

For $k = 0, 2, 4, \ldots, N-2 \stackrel{\triangle}{=} 2p$:

$$
Y(2p) = \sum_{t=1}^{\tilde{N}} [y(t) + y(t+\tilde{N})] \tilde{W}^{tp}
$$

For $k = 1, 3, 5, \ldots, N-1 = 2p+1$:

$$
Y(2p+1) = \sum_{t=1}^{\tilde{N}} \{[y(t) - y(t+\tilde{N})] W^t\} \tilde{W}^{tp}
$$

Each is a $\tilde{N} = N/2$-point DFT computation.

FFT Computation Count

Let $c_k =$ number of flops for $N = 2^k$ point FFT.

Then

$$\begin{aligned} c_k &= \frac{2^k}{2} + 2c_{k-1} \\ &\Rightarrow \quad c_k = \frac{k2^k}{2} \end{aligned}$$

Thus,

$$\boxed{c_k = \frac{1}{2} N \log_2 N}$$

Zero Padding

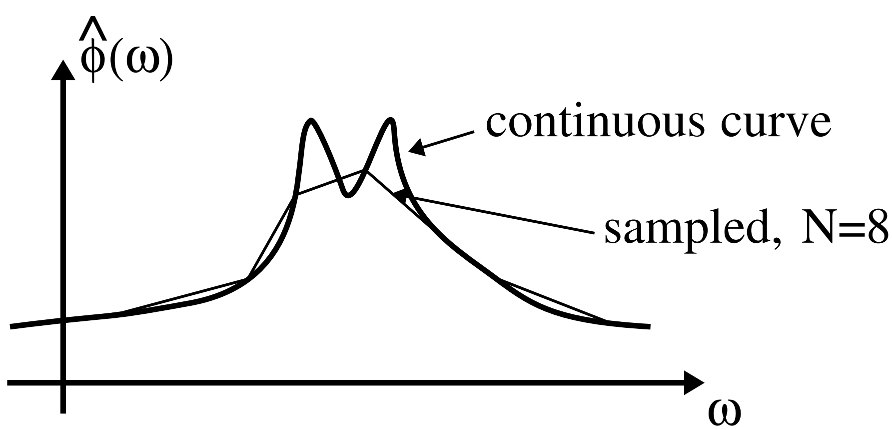
Append the given data by zeros prior to computing DFT (or FFT):

$$\{\underbrace{y(1), \ \ldots \ , y(N), 0, \ \ldots \ 0}_{\overline{N}}\}$$

Goals:

- Apply a radix-2 FFT (so $\overline{N}$ = power of 2)
- Finer sampling of $\hat{\phi}(\omega)$:

$$\left\{\frac{2\pi}{N}k\right\}_{k=0}^{N-1} \quad \rightarrow \quad \left\{\frac{2\pi}{\overline{N}}k\right\}_{k=0}^{\overline{N}-1}$$



Improved Periodogram-Based Methods

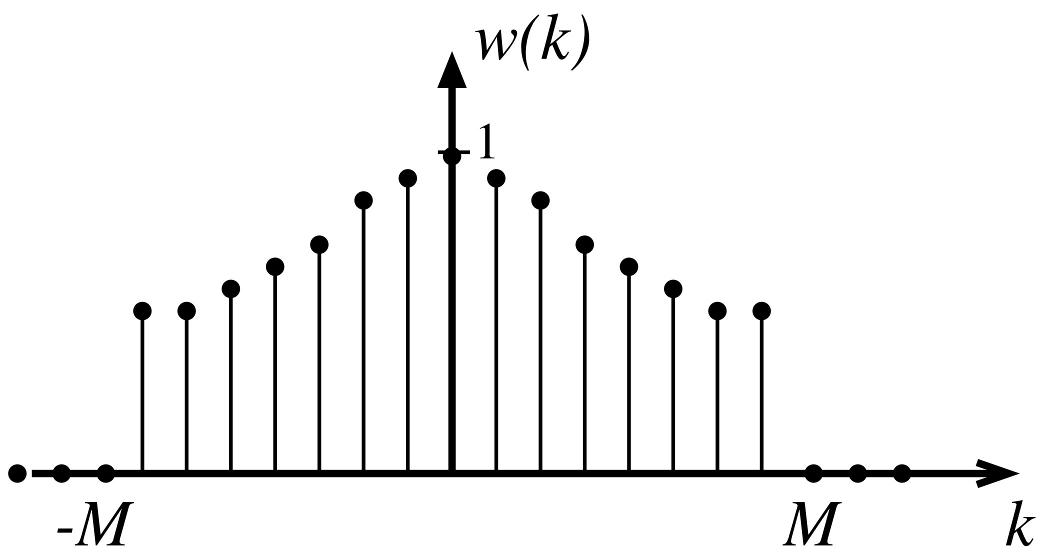
Lecture 3

Blackman-Tukey Method

Basic Idea: Weighted correlogram, with small weight applied to covariances $\hat{r}(k)$ with "large" $|k|$.

$$\boxed{\hat{\phi}_{BT}(\omega) = \sum_{k=-(M-1)}^{M-1} w(k)\hat{r}(k)e^{-i\omega k}}$$

$$\{w(k)\} = \text{ Lag Window}$$



Blackman-Tukey Method, con't

$$\boxed{\hat{\phi}_{BT}(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \hat{\phi}_p(\zeta) W(\omega - \zeta) d\zeta}$$

$$\begin{aligned} W(\omega) &= \text{DTFT}\{w(k)\} \\ &= \text{Spectral Window} \end{aligned}$$

Conclusion: $\hat{\phi}_{BT}(\omega) =$ "locally" smoothed periodogram

Effect:

- Variance decreases substantially
- Bias increases slightly

By proper choice of M:

$$\text{MSE} = \text{var} + \text{bias}^2 \rightarrow 0 \text{ as } N \rightarrow \infty$$

Window Design Considerations

Nonnegativeness:

$$\hat{\phi}_{BT}(\omega) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \underbrace{\hat{\phi}_p(\zeta)}_{\geq 0} W(\omega - \zeta) d\zeta$$

If $W(\omega) \geq 0$ ($\Leftrightarrow w(k)$ is a psd sequence)

Then: $\hat{\phi}_{BT}(\omega) \geq 0$ (which is desirable)

Time-Bandwidth Product

$$N_e = \frac{\sum_{k=-(M-1)}^{M-1} w(k)}{w(0)} = \text{equiv time width}$$

$$\beta_e = \frac{\frac{1}{2\pi}\int_{-\pi}^{\pi} W(\omega) d\omega}{W(0)} = \text{equiv bandwidth}$$

$$\boxed{N_e \, \beta_e = 1}$$

Window Design, con't

- $\beta_e = 1/N_e = \mathrm{O}(1/M)$
  is the BT resolution threshold.

- As M increases, bias decreases and variance increases.

  $\Rightarrow$ Choose M as a tradeoff between *variance* and *bias*.

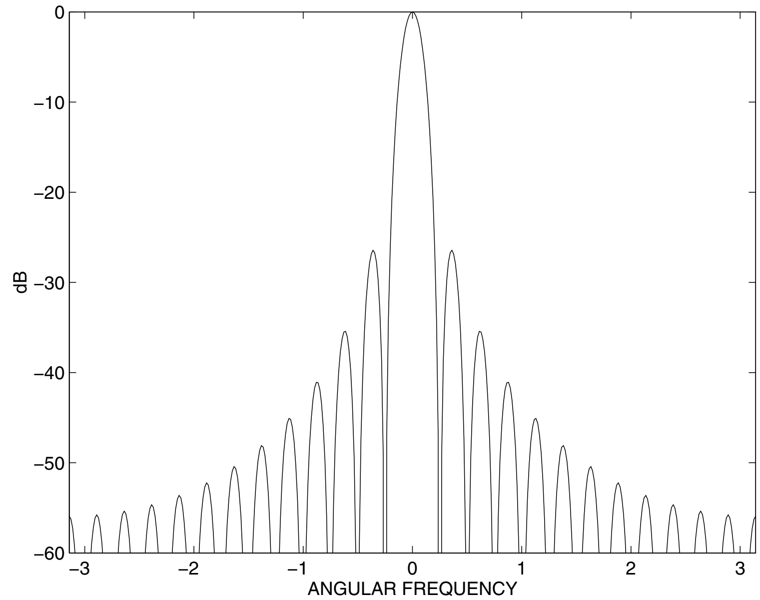
- Once M is given, N_e (and hence β_e) is essentially fixed.

  $\Rightarrow$ Choose window shape to compromise between *smearing* (main lobe width) and *leakage* (sidelobe level).

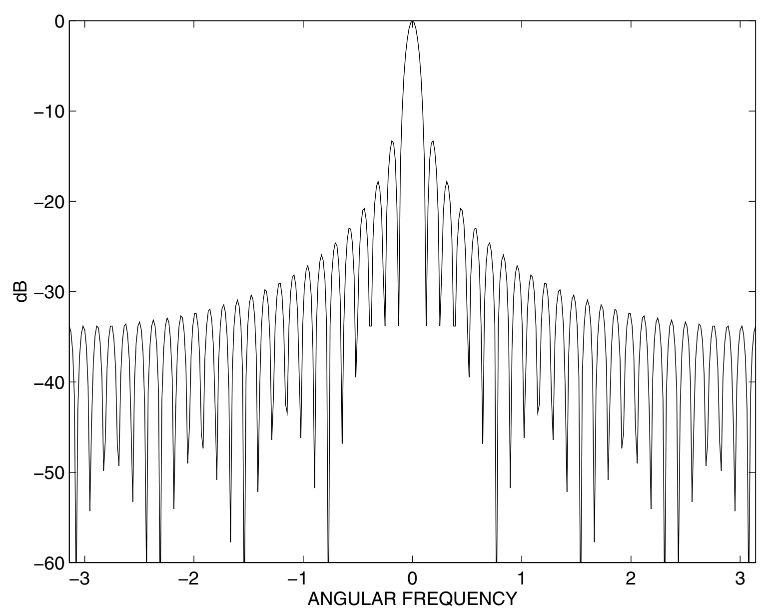
  The energy in the main lobe and in the sidelobes cannot be reduced *simultaneously*, once M is given.

Window Examples

Triangular Window, $M = 25$

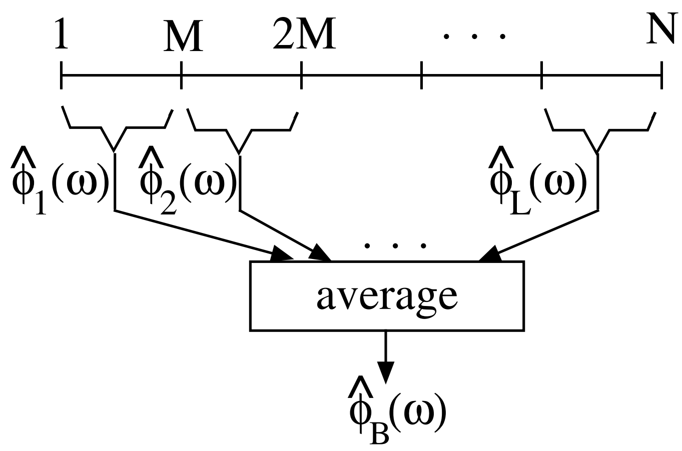


Rectangular Window, $M = 25$



Bartlett Method

Basic Idea:



Mathematically:

$$\begin{aligned} y_j(t) &= y((j-1)M+t) \quad t = 1,\ldots,M \\ &= \text{the } j\text{th subsequence} \\ &\quad (j = 1,\ldots,L \triangleq [N/M]) \end{aligned}$$

$$\hat{\phi}_j(\omega) = \frac{1}{M}\left|\sum_{t=1}^{M} y_j(t)e^{-i\omega t}\right|^2$$

$$\boxed{\hat{\phi}_B(\omega) = \frac{1}{L}\sum_{j=1}^{L}\hat{\phi}_j(\omega)}$$

Comparison of Bartlett and Blackman-Tukey Estimates

$$\begin{aligned}
\hat{\phi}_B(\omega) &= \frac{1}{L}\sum_{j=1}^{L}\left\{\sum_{k=-(M-1)}^{M-1}\hat{r}_j(k)e^{-i\omega k}\right\} \\
&= \sum_{k=-(M-1)}^{M-1}\left\{\frac{1}{L}\sum_{j=1}^{L}\hat{r}_j(k)\right\}e^{-i\omega k} \\
&\simeq \sum_{k=-(M-1)}^{M-1}\hat{r}(k)e^{-i\omega k}
\end{aligned}$$

Thus:

$$\boxed{\hat{\phi}_B(\omega) \simeq \hat{\phi}_{BT}(\omega) \text{ with a rectangular lag window } w_R(k)}$$

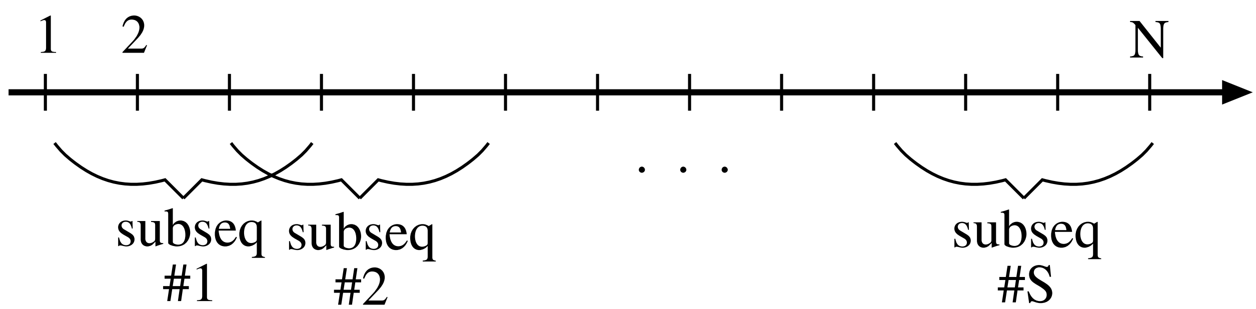
Since $\hat{\phi}_B(\omega)$ implicitly uses $\{w_R(k)\}$, the Bartlett method has

- High resolution (little smearing)
- Large leakage and relatively large variance

Welch Method

Similar to Bartlett method, but

- allow overlap of subsequences (gives more subsequences, and thus “better” averaging)
- use data window for each periodogram; gives mainlobe-sidelobe tradeoff capability



Let $S =$# of subsequences of length M.
(Overlapping means $S > [N/M] \Rightarrow$ “better averaging”.)

Additional flexibility:

The data in each subsequence are weighted by a *temporal* window

Welch is approximately equal to $\hat{\phi}_{BT}(\omega)$ with a non-rectangular lag window.

Daniell Method

By a previous result, for $N \gg 1$,

$\{\hat{\phi}_p(\omega_j)\}$ are (nearly) uncorrelated random variables for

$$\left\{\omega_j = \frac{2\pi}{N}\, j\right\}_{j=0}^{N-1}$$

Idea: "Local averaging" of $(2J+1)$ samples in the frequency domain should reduce the variance by about $(2J+1)$.

$$\boxed{\hat{\phi}_D(\omega_k) = \frac{1}{2J+1} \sum_{j=k-J}^{k+J} \hat{\phi}_p\left(\omega_j\right)}$$

Daniell Method, con't

As J increases:

- Bias increases (more smoothing)
- Variance decreases (more averaging)

Let $\beta = 2J/N$. Then, for $N \gg 1$,

$$\boxed{\hat{\phi}_D(\omega) \simeq \frac{1}{2\pi\beta} \int_{-\beta\pi}^{\beta\pi} \hat{\phi}_p(\overline{\omega}) d\overline{\omega}}$$

Hence: $\hat{\phi}_D(\omega) \simeq \hat{\phi}_{BT}(\omega)$ with a *rectangular spectral* window.

Summary of Periodogram Methods

- **Unwindowed periodogram**
  - reasonable bias
  - unacceptable variance

- **Modified periodograms**
  - Attempt to reduce the variance at the expense of (slightly) increasing the bias.

- **BT periodogram**
  - Local smoothing/averaging of $\hat{\phi}_p(\omega)$ by a suitably selected *spectral* window.
  - Implemented by truncating and weighting $\hat{r}(k)$ using a *lag* window in $\hat{\phi}_c(\omega)$

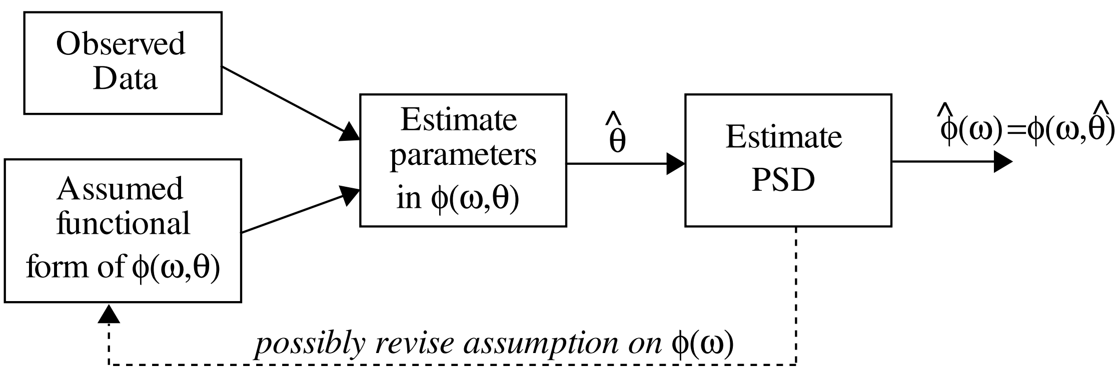
- **Bartlett, Welch periodograms**
  - Approximate interpretation: $\hat{\phi}_{BT}(\omega)$ with a suitable *lag* window (rectangular for Bartlett; more general for Welch).
  - Implemented by averaging subsample periodograms.

- **Daniell Periodogram**
  - Approximate interpretation: $\hat{\phi}_{BT}(\omega)$ with a rectangular *spectral* window.
  - Implemented by local averaging of periodogram values.

Parametric Methods for Rational Spectra

Lecture 4

Basic Idea of Parametric Spectral Estimation



Rational Spectra

$$\phi(\omega) = \frac{\sum_{|k|\leq m} \gamma_k e^{-i\omega k}}{\sum_{|k|\leq n} \rho_k e^{-i\omega k}}$$

$\phi(\omega)$ is a *rational function* in $e^{-i\omega}$.

By *Weierstrass theorem*, $\phi(\omega)$ can approximate arbitrarily well *any continuous PSD*, provided m and n are chosen sufficiently large.

Note, however:

- choice of m and n is not simple
- some PSDs are *not* continuous

By *Spectral Factorization* theorem, a rational $\phi(\omega)$ can be factored as

$$\phi(\omega) = \left| \frac{B(\omega)}{A(\omega)} \right|^2 \sigma^2$$

$$\begin{aligned} A(z) &= 1 + a_1 z^{-1} + \cdots + a_n z^{-n} \\ B(z) &= 1 + b_1 z^{-1} + \cdots + b_m z^{-m} \end{aligned}$$

and, e.g., $A(\omega) = A(z)|_{z=e^{i\omega}}$

Signal Modeling Interpretation:

$$\underset{\text{white noise}}{\xrightarrow[\phi_e(\omega) = \sigma^2]{e(t)}} \boxed{\frac{B(q)}{A(q)}} \underset{\text{filtered white noise}}{\xrightarrow[\phi_y(\omega) = \left|\frac{B(\omega)}{A(\omega)}\right|^2 \sigma^2]{y(t)}}$$

ARMA:	$A(q)y(t) = B(q)e(t)$
AR:	$A(q)y(t) = e(t)$
MA:	$y(t) = B(q)e(t)$

ARMA Covariance Structure

ARMA signal model:

$$y(t) + \sum_{i=1}^{n} a_i y(t-i) = \sum_{j=0}^{m} b_j e(t-j), \qquad (b_0 = 1)$$

Multiply by $y^*(t-k)$ and take $E\{\cdot\}$ to give:

$$\begin{aligned} r(k) + \sum_{i=1}^{n} a_i r(k-i) &= \sum_{j=0}^{m} b_j E\{e(t-j) y^*(t-k)\} \\ &= \sigma^2 \sum_{j=0}^{m} b_j h^*_{j-k} \\ &= 0 \text{ for } k > m \end{aligned}$$

where $H(q) = \frac{B(q)}{A(q)} = \sum_{k=0}^{\infty} h_k q^{-k}, \quad (h_0 = 1)$

AR Signals: Yule-Walker Equations

AR: $m = 0$.

Writing covariance equation in matrix form for $k = 1 \ldots n$:

$$\begin{bmatrix} r(0) & r(-1) & \ldots & r(-n) \\ r(1) & r(0) & & \vdots \\ \vdots & & \ddots & r(-1) \\ r(n) & \ldots & & r(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_1 \\ \vdots \\ a_n \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$$R \begin{bmatrix} 1 \\ \theta \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \end{bmatrix}$$

These are the **Yule–Walker (YW) Equations**.

Yule-Walker Method:

Replace $r(k)$ by $\hat{r}(k)$ and solve for $\{\hat{a}_i\}$ and $\hat{\sigma}^2$:

$$\begin{bmatrix} \hat{r}(0) & \hat{r}(-1) & \dots & \hat{r}(-n) \\ \hat{r}(1) & \hat{r}(0) & & \vdots \\ \vdots & & \ddots & \hat{r}(-1) \\ \hat{r}(n) & \dots & & \hat{r}(0) \end{bmatrix} \begin{bmatrix} 1 \\ \hat{a}_1 \\ \vdots \\ \hat{a}_n \end{bmatrix} = \begin{bmatrix} \hat{\sigma}^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

Then the PSD estimate is

$$\hat{\phi}(\omega) = \frac{\hat{\sigma}^2}{|\hat{A}(\omega)|^2}$$

AR Spectral Estimation: LS Method

Least Squares Method:

$$
\begin{aligned}
e(t) &= y(t) + \sum_{i=1}^{n} a_i y(t-i) = y(t) + \varphi^T(t)\theta \\
&\triangleq y(t) + \hat{y}(t)
\end{aligned}
$$

where $\varphi(t) = [y(t-1), \ldots, y(t-n)]^T$.

Find $\theta = [a_1 \; \ldots \; a_n]^T$ to minimize

$$
f(\theta) = \sum_{t=n+1}^{N} |e(t)|^2
$$

This gives $\hat{\theta} = -(Y^*Y)^{-1}(Y^*y)$ where

$$
y = \begin{bmatrix} y(n+1) \\ y(n+2) \\ \vdots \\ y(N) \end{bmatrix}, \; Y = \begin{bmatrix} y(n) & y(n-1) & \cdots & y(1) \\ y(n+1) & y(n) & \cdots & y(2) \\ \vdots & & & \vdots \\ y(N-1) & y(N-2) & \cdots & y(N-n) \end{bmatrix}
$$

Levinson–Durbin Algorithm

Fast, order-recursive solution to YW equations

$$\underbrace{\begin{bmatrix} \rho_0 & \rho_{-1} & \cdots & \rho_{-n} \\ \rho_1 & \rho_0 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \rho_{-1} \\ \rho_n & \cdots & \rho_1 & \rho_0 \end{bmatrix}}_{R_{n+1}} \begin{bmatrix} 1 \\ \\ \theta_n \\ \\ \end{bmatrix} = \begin{bmatrix} \sigma_n^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}$$

$\rho_k =$ either $r(k)$ or $\hat{r}(k)$.

Direct Solution:

- For one given value of n: $O(n^3)$ flops
- For $k = 1, \ldots, n$: $O(n^4)$ flops

Levinson–Durbin Algorithm:

Exploits the Toeplitz form of R_{n+1} to obtain the solutions for $k = 1, \ldots, n$ in $O(n^2)$ flops!

Levinson-Durbin Alg, con't

Relevant Properties of R:

- $Rx = y \;\; \leftrightarrow \;\; R\tilde{x} = \tilde{y}$, where $\tilde{x} = [x_n^* \ldots x_1^*]^T$

- Nested structure

$$R_{n+2} = \left[\begin{array}{cc|c} \multicolumn{2}{c|}{R_{n+1}} & \begin{array}{c}\rho_{n+1}^* \\ \tilde{r}_n\end{array} \\ \hline \rho_{n+1} & \tilde{r}_n^* & \rho_0 \end{array}\right], \quad \tilde{r}_n = \begin{bmatrix} \rho_n^* \\ \vdots \\ \rho_1^* \end{bmatrix}$$

Thus,

$$R_{n+2}\left[\begin{array}{c} 1 \\ \theta_n \\ \hline 0 \end{array}\right] = \left[\begin{array}{cc|c} \multicolumn{2}{c|}{R_{n+1}} & \begin{array}{c}\rho_{n+1}^* \\ \tilde{r}_n\end{array} \\ \hline \rho_{n+1} & \tilde{r}_n^* & \rho_0 \end{array}\right] \left[\begin{array}{c} 1 \\ \theta_n \\ \hline 0 \end{array}\right] = \left[\begin{array}{c} \sigma_n^2 \\ 0 \\ \hline \alpha_n \end{array}\right]$$

where $\boxed{\alpha_n = \rho_{n+1} + \tilde{r}_n^* \theta_n}$

Levinson-Durbin Alg, con't

$$R_{n+2}\begin{bmatrix}1\\ \theta_n\\ 0\end{bmatrix}=\begin{bmatrix}\sigma_n^2\\ 0\\ \alpha_n\end{bmatrix}, \qquad R_{n+2}\begin{bmatrix}0\\ \tilde{\theta}_n\\ 1\end{bmatrix}=\begin{bmatrix}\alpha_n^*\\ 0\\ \sigma_n^2\end{bmatrix}$$

Combining these gives:

$$R_{n+2}\left\{\begin{bmatrix}1\\ \theta_n\\ 0\end{bmatrix}+k_n\begin{bmatrix}0\\ \tilde{\theta}_n\\ 1\end{bmatrix}\right\}=\begin{bmatrix}\sigma_n^2+k_n\alpha_n^*\\ 0\\ \alpha_n+k_n\sigma_n^2\end{bmatrix}=\begin{bmatrix}\sigma_{n+1}^2\\ 0\\ 0\end{bmatrix}$$

Thus, $\boxed{k_n=-\alpha_n/\sigma_n^2}$ $\Rightarrow$

$$\boxed{\begin{array}{l}\theta_{n+1}=\begin{bmatrix}\theta_n\\ 0\end{bmatrix}+k_n\begin{bmatrix}\tilde{\theta}_n\\ 1\end{bmatrix}\\ \sigma_{n+1}^2=\sigma_n^2+k_n\alpha_n^*=\sigma_n^2(1-|k_n|^2)\end{array}}$$

Computation count:

$\sim 2k$ flops for the step $k\to k+1$

$\Rightarrow$ $\boxed{\sim n^2 \text{ flops}}$ to determine $\{\sigma_k^2,\ \theta_k\}_{k=1}^n$

This is $O(n^2)$ times faster than the direct solution.

MA Signals

MA: $n = 0$

$$\begin{aligned} y(t) &= B(q)e(t) \\ &= e(t) + b_1 e(t-1) + \cdots + b_m e(t-m) \end{aligned}$$

Thus,

$$\boxed{r(k) = 0 \text{ for } |k| > m}$$

and

$$\phi(\omega) = |B(\omega)|^2 \sigma^2 = \sum_{k=-m}^{m} r(k) e^{-i\omega k}$$

MA Spectrum Estimation

Two main ways to Estimate $\phi(\omega)$:

1. Estimate $\{b_k\}$ and σ^2 and insert them in

$$\phi(\omega) = |B(\omega)|^2\sigma^2$$

- nonlinear estimation problem
- $\hat{\phi}(\omega)$ is guaranteed to be ≥ 0

2. Insert sample covariances $\{\hat{r}(k)\}$ in:

$$\phi(\omega) = \sum_{k=-m}^{m} r(k)e^{-i\omega k}$$

- This is $\hat{\phi}_{BT}(\omega)$ with a rectangular lag window of length $2m+1$.
- $\hat{\phi}(\omega)$ is *not* guaranteed to be ≥ 0

Both methods are special cases of ARMA methods described below, with AR model order $n = 0$.

ARMA Signals

ARMA models can represent spectra with both peaks (AR part) and valleys (MA part).

$$A(q)y(t) = B(q)e(t)$$

$$\phi(\omega) = \sigma^2 \left| \frac{B(\omega)}{A(\omega)} \right|^2 = \frac{\sum_{k=-m}^{m} \gamma_k e^{-i\omega k}}{|A(\omega)|^2}$$

where

$$\begin{aligned} \gamma_k &= E\left\{[B(q)e(t)][B(q)e(t-k)]^*\right\} \\ &= E\left\{[A(q)y(t)][A(q)y(t-k)]^*\right\} \\ &= \sum_{j=0}^{n} \sum_{p=0}^{n} a_j a_p^* \, r(k+p-j) \end{aligned}$$

ARMA Spectrum Estimation

Two Methods:

1. Estimate $\{a_i, b_j, \sigma^2\}$ in $\phi(\omega) = \sigma^2 \left|\frac{B(\omega)}{A(\omega)}\right|^2$

- nonlinear estimation problem; can use an approximate linear *two-stage least squares* method

- $\hat{\phi}(\omega)$ is guaranteed to be ≥ 0

2. Estimate $\{a_i,\ r(k)\}$ in $\phi(\omega) = \frac{\sum_{k=-m}^{m} \gamma_k e^{-i\omega k}}{|A(\omega)|^2}$

- linear estimation problem (the Modified Yule-Walker method).

- $\hat{\phi}(\omega)$ is *not* guaranteed to be ≥ 0

Two-Stage Least-Squares Method

Assumption: The ARMA model is invertible:

$$\begin{aligned} e(t) &= \frac{A(q)}{B(q)} y(t) \\ &= y(t) + \alpha_1 y(t-1) + \alpha_2 y(t-2) + \cdots \\ &= \text{AR}(\infty) \text{ with } |\alpha_k| \to 0 \text{ as } k \to \infty \end{aligned}$$

Step 1: Approximate, for some large K

$$e(t) \simeq y(t) + \alpha_1 y(t-1) + \cdots + \alpha_K y(t-K)$$

1a) Estimate the coefficients $\{\alpha_k\}_{k=1}^{K}$ by using AR modelling techniques.

1b) Estimate the noise sequence

$$\hat{e}(t) = y(t) + \hat{\alpha}_1 y(t-1) + \cdots + \hat{\alpha}_K y(t-K)$$

and its variance

$$\hat{\sigma}^2 = \frac{1}{N-K} \sum_{t=K+1}^{N} |\hat{e}(t)|^2$$

Two-Stage Least-Squares Method, con't

Step 2: Replace $\{e(t)\}$ by $\hat{e}(t)$ in the ARMA equation,

$$A(q)y(t) \simeq B(q)\hat{e}(t)$$

and obtain estimates of $\{a_i,\ b_j\}$ by applying least squares techniques.

Note that the a_i and b_j coefficients enter linearly in the above equation:

$$\begin{aligned} y(t) - \hat{e}(t) &\simeq [-y(t-1) \ldots - y(t-n), \\ &\qquad \hat{e}(t-1) \ldots \hat{e}(t-m)]\theta \\ \theta &= [a_1 \ldots a_n\, b_1 \ldots b_m]^T \end{aligned}$$

Modified Yule-Walker Method

ARMA Covariance Equation:

$$r(k) + \sum_{i=1}^{n} a_i r(k-i) = 0, \quad k > m$$

In matrix form for $k = m+1, \ldots, m+M$

$$\begin{bmatrix} r(m) & \ldots & r(m-n+1) \\ r(m+1) & & r(m-n+2) \\ \vdots & \ddots & \vdots \\ r(m+M-1) & \ldots & r(m-n+M) \end{bmatrix} \begin{bmatrix} a_1 \\ \vdots \\ a_n \end{bmatrix} = - \begin{bmatrix} r(m+1) \\ r(m+2) \\ \vdots \\ r(m+M) \end{bmatrix}$$

Replace $\{r(k)\}$ by $\{\hat{r}(k)\}$ and solve for $\{a_i\}$.

If $M = n$, fast Levinson-type algorithms exist for obtaining $\{\hat{a}_i\}$.

If $M > n$ *overdetermined* YW system of equations; least squares solution for $\{\hat{a}_i\}$.

Note: For narrowband ARMA signals, the accuracy of $\{\hat{a}_i\}$ is often better for $M > n$

Summary of Parametric Methods for Rational Spectra

Method	Computational Burden	Accuracy	Guarantee $\hat{\phi}(\omega) \geq 0$?	Use for
AR: YW or LS	low	medium	Yes	Spectra with (narrow) peaks but no valley
MA: BT	low	low-medium	No	Broadband spectra possibly with valleys but no peaks
ARMA: MYW	low-medium	medium	No	Spectra with both peaks and (not too deep) valleys
ARMA: 2-Stage LS	medium-high	medium-high	Yes	As above

Parametric Methods
for
Line Spectra — Part 1

Lecture 5

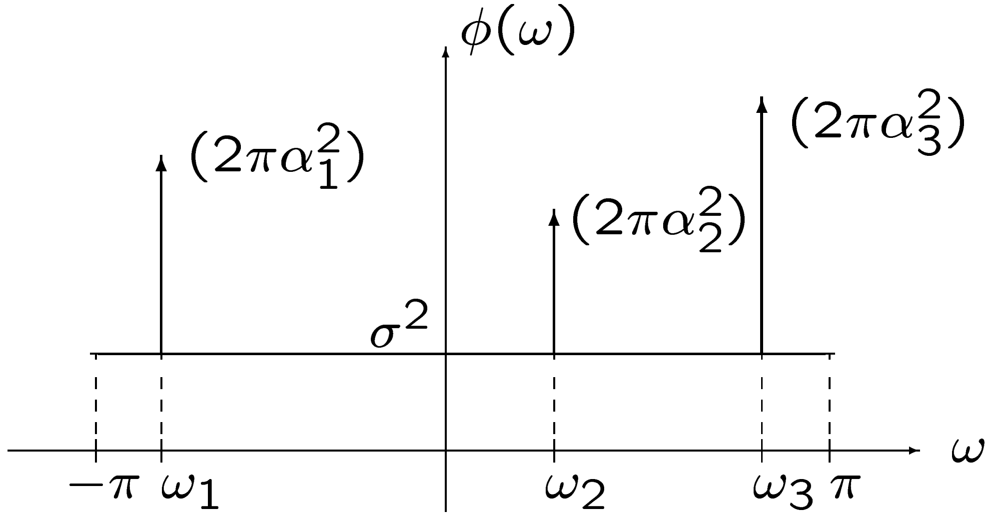
Line Spectra

Many applications have signals with (near) sinusoidal components. Examples:

- communications
- radar, sonar
- geophysical seismology

ARMA model is a *poor approximation*

Better approximation by *Discrete/Line Spectrum Models*



An "Ideal" line spectrum

Line Spectral Signal Model

Signal Model: Sinusoidal components of frequencies $\{\omega_k\}$ and powers $\{\alpha_k^2\}$, superimposed in white noise of power σ^2.

$$
\begin{aligned}
y(t) &= x(t) + e(t) \quad t = 1, 2, \ldots \\
x(t) &= \sum_{k=1}^{n} \underbrace{\alpha_k e^{i(\omega_k t + \phi_k)}}_{x_k(t)}
\end{aligned}
$$

Assumptions:

A1: $\alpha_k > 0 \quad \omega_k \in [-\pi, \pi]$
(prevents model ambiguities)

A2: $\{\varphi_k\} =$ independent rv's, uniformly distributed on $[-\pi, \pi]$
(realistic and mathematically convenient)

A3: $e(t) =$ circular white noise with variance σ^2

$$E\{e(t)e^*(s)\} = \sigma^2 \delta_{t,s} \quad E\{e(t)e(s)\} = 0$$

(can be achieved by "slow" sampling)

Covariance Function and PSD

Note that:

- $E\left\{e^{i\varphi_p}e^{-i\varphi_j}\right\} = 1$, for $p = j$

- $E\left\{e^{i\varphi_p}e^{-i\varphi_j}\right\} = E\left\{e^{i\varphi_p}\right\} E\left\{e^{-i\varphi_j}\right\}$
$$= \left|\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\varphi}\, d\varphi\right|^2 = 0, \text{ for } p \neq j$$

Hence,

$$E\left\{x_p(t)x_j^*(t-k)\right\} = \alpha_p^2\, e^{i\omega_p k}\, \delta_{p,j}$$

$$\boxed{\begin{array}{rcl} r(k) & = & E\left\{y(t)y^*(t-k)\right\} \\ & = & \sum_{p=1}^{n} \alpha_p^2 e^{i\omega_p k} + \sigma^2\delta_{k,0} \end{array}}$$

and

$$\boxed{\phi(\omega) = 2\pi \sum_{p=1}^{n} \alpha_p^2 \delta(\omega - \omega_p) + \sigma^2}$$

Parameter Estimation

Estimate either:

- $\{\omega_k, \alpha_k, \varphi_k\}_{k=1}^n,\ \sigma^2$ (Signal Model)
- $\{\omega_k, \alpha_k^2\}_{k=1}^n,\ \sigma^2$ (PSD Model)

Major Estimation Problem: $\{\hat{\omega}_k\}$

Once $\{\hat{\omega}_k\}$ are determined:

- $\{\hat{\alpha}_k^2\}$ can be obtained by a least squares method from

$$\hat{r}(k) = \sum_{p=1}^{n} \alpha_p^2 e^{i\hat{\omega}_p k} + \text{residuals}$$

OR:

- Both $\{\hat{\alpha}_k\}$ and $\{\hat{\varphi}_k\}$ can be derived by a least squares method from

$$y(t) = \sum_{k=1}^{n} \beta_k e^{i\hat{\omega}_k t} + \text{residuals}$$

with $\beta_k = \alpha_k e^{i\varphi_k}$.

Nonlinear Least Squares (NLS) Method

$$\min_{\{\omega_k,\alpha_k,\varphi_k\}} \underbrace{\sum_{t=1}^{N} \left| y(t) - \sum_{k=1}^{n} \alpha_k e^{i(\omega_k t+\varphi_k)} \right|^2}_{F(\omega,\alpha,\varphi)}$$

Let:

$$\begin{aligned}
\beta_k &= \alpha_k e^{i\varphi_k} \\
\beta &= [\beta_1 \ldots \beta_n]^T \\
Y &= [y(1) \ldots y(N)]^T \\
B &= \begin{bmatrix} e^{i\omega_1} & \ldots & e^{i\omega_n} \\ \vdots & & \vdots \\ e^{iN\omega_1} & \ldots & e^{iN\omega_n} \end{bmatrix}
\end{aligned}$$

Nonlinear Least Squares (NLS) Method, con't

Then:

$$
\begin{aligned}
F &= (Y - B\beta)^*(Y - B\beta) = \|Y - B\beta\|^2 \\
&= [\beta - (B^*B)^{-1}B^*Y]^*[B^*B] \\
&\qquad [\beta - (B^*B)^{-1}B^*Y] \\
&\qquad + Y^*Y - Y^*B(B^*B)^{-1}B^*Y
\end{aligned}
$$

This gives:

$$\boxed{\hat{\beta} = (B^*B)^{-1}B^*Y\Big|_{\omega=\hat{\omega}}}$$

and

$$\boxed{\hat{\omega} = \arg\max_{\omega} Y^*B(B^*B)^{-1}B^*Y}$$

NLS Properties

Excellent Accuracy:

$$\text{var}\ (\hat{\omega}_k) = \frac{6\sigma^2}{N^3\alpha_k^2} \quad (\text{for } N \gg 1)$$

Example: $N = 300$

$$\text{SNR}_k = \alpha_k^2/\sigma^2 = 30 \text{ dB}$$

Then $\sqrt{\text{var}(\hat{\omega}_k)} \sim 10^{-5}$.

Difficult Implementation:

The NLS cost function F is multimodal; it is difficult to avoid convergence to local minima.

Unwindowed Periodogram as an Approximate NLS Method

For a single (complex) sinusoid, the maximum of the unwindowed periodogram is the NLS frequency estimate:

Assume: $\boxed{n = 1}$

Then: $B^*B = N$

$$B^*Y = \sum_{t=1}^{N} y(t)e^{-i\omega t} = Y(\omega) \quad \text{(finite DTFT)}$$

$$\begin{aligned} Y^*B(B^*B)^{-1}B^*Y &= \frac{1}{N}\,|Y(\omega)|^2 \\ &= \hat{\phi}_p(\omega) \\ &= \text{(Unwindowed Periodogram)} \end{aligned}$$

So, with *no approximation*,

$$\boxed{\hat{\omega} = \arg\max_{\omega}\ \hat{\phi}_p(\omega)}$$

Unwindowed Periodogram as an Approximate NLS Method, con't

Assume: $\boxed{n > 1}$

Then:

$$\boxed{\{\hat{\omega}_k\}_{k=1}^{n} \simeq \text{the locations of the } n \text{ largest peaks of } \hat{\phi}_p(\omega)}$$

provided that

$$\inf |\omega_k - \omega_p| > 2\pi/N$$

which is the periodogram resolution limit.

If better resolution desired then use a *High/Super Resolution* method.

High-Order Yule-Walker Method

Recall:

$$y(t) = x(t) + e(t) = \sum_{k=1}^{n} \underbrace{\alpha_k e^{i(\omega_k t + \varphi_k)}}_{x_k(t)} + e(t)$$

"Degenerate" ARMA equation for $y(t)$:

$$\begin{aligned}(1 \;-\; & e^{i\omega_k} q^{-1}) x_k(t) \\ & = \alpha_k \left\{ e^{i(\omega_k t + \varphi_k)} - e^{i\omega_k}\, e^{i[\omega_k(t-1)+\varphi_k]} \right\} = 0\end{aligned}$$

Let

$$\begin{aligned} B(q) &= 1 + \sum_{k=1}^{L} b_k q^{-k} \triangleq A(q)\bar{A}(q) \\ A(q) &= (1 - e^{i\omega_1} q^{-1}) \cdots (1 - e^{i\omega_n} q^{-1}) \\ \bar{A}(q) &= \text{arbitrary} \end{aligned}$$

Then $B(q)x(t) \equiv 0 \;\Rightarrow$

$$\boxed{B(q)y(t) = B(q)e(t)}$$

High-Order Yule-Walker Method, con't

Estimation Procedure:

- Estimate $\{\hat{b}_i\}_{i=1}^{L}$ using an ARMA MYW technique
- Roots of $\hat{B}(q)$ give $\{\hat{\omega}_k\}_{k=1}^{n}$, along with $L-n$ "spurious" roots.

High-Order and Overdetermined YW Equations

ARMA covariance:

$$r(k) + \sum_{i=1}^{L} b_i r(k-i) = 0, \quad k > L$$

In matrix form for $k = L+1, \ldots, L+M$

$$\underbrace{\begin{bmatrix} r(L) & \ldots & r(1) \\ r(L+1) & \ldots & r(2) \\ \vdots & & \vdots \\ r(L+M-1) & \ldots & r(M) \end{bmatrix}}_{\triangleq \Omega} b = - \underbrace{\begin{bmatrix} r(L+1) \\ r(L+2) \\ \vdots \\ r(L+M) \end{bmatrix}}_{\triangleq \rho}$$

This is a high-order (if $L > n$) and overdetermined (if $M > L$) system of YW equations.

High-Order and Overdetermined YW Equations, con't

Fact: $\boxed{\text{rank}(\Omega) = n}$

SVD of Ω: $\Omega = U\Sigma V^*$

- $U = (M \times n)$ with $U^*U = I_n$
- $V^* = (n \times L)$ with $V^*V = I_n$
- $\Sigma = (n \times n)$, diagonal and nonsingular

Thus,

$$(U\Sigma V^*)b = -\rho$$

The Minimum-Norm solution is

$$b = -\Omega^{\dagger}\rho = -V\Sigma^{-1}U^*\rho$$

Important property: The additional $(L - n)$ spurious zeros of $B(q)$ are located strictly *inside* the unit circle, if the Minimum-Norm solution b is used.

HOYW Equations, Practical Solution

Let $\hat{\Omega} = \Omega$ but made from $\{\hat{r}(k)\}$ instead of $\{r(k)\}$.

Let $\hat{U}$, $\hat{\Sigma}$, $\hat{V}$ be defined similarly to U, Σ, V from the SVD of $\hat{\Omega}$.

Compute $\boxed{\hat{b} = -\hat{V}\hat{\Sigma}^{-1}\hat{U}^*\hat{\rho}}$

Then $\{\hat{\omega}_k\}_{k=1}^n$ are found from the n zeroes of $\hat{B}(q)$ that are closest to the unit circle.

When the SNR is low, this approach may give spurious frequency estimates when $L > n$; this is the price paid for increased accuracy when $L > n$.

Parametric Methods for Line Spectra — Part 2

Lecture 6

The Covariance Matrix Equation

Let:

$$\begin{aligned} a(\omega) &= [1\ e^{-i\omega}\ \ldots\ e^{-i(m-1)\omega}]^T \\ A &= [a(\omega_1)\ \ldots\ a(\omega_n)] \quad (m \times n) \end{aligned}$$

Note: $\boxed{\text{rank}(A) = n}$ (for $m \geq n$)

Define

$$\tilde{y}(t) \stackrel{\triangle}{=} \begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(t-m+1) \end{bmatrix} = A\tilde{x}(t) + \tilde{e}(t)$$

where

$$\begin{aligned} \tilde{x}(t) &= [x_1(t) \ldots x_n(t)]^T \\ \tilde{e}(t) &= [e(t) \ldots e(t-m+1)]^T \end{aligned}$$

Then

$$\boxed{R \stackrel{\triangle}{=} E\left\{\tilde{y}(t)\tilde{y}^*(t)\right\} = APA^* + \sigma^2 I}$$

with

$$P = E\left\{\tilde{x}(t)\tilde{x}^*(t)\right\} = \begin{bmatrix} \alpha_1^2 & & 0 \\ & \ddots & \\ 0 & & \alpha_n^2 \end{bmatrix}$$

Eigendecomposition of R and Its Properties

$$R = APA^* + \sigma^2 I \quad (m > n)$$

Let:

$\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$: eigenvalues of R

$\{s_1, \ldots s_n\}$: orthonormal eigenvectors associated with $\{\lambda_1, \ldots, \lambda_n\}$

$\{g_1, \ldots, g_{m-n}\}$: orthonormal eigenvectors associated with $\{\lambda_{n+1}, \ldots, \lambda_m\}$

$$\begin{aligned} S &= [s_1 \ldots s_n] & & (m \times n) \\ G &= [g_1 \ldots g_{m-n}] & & (m \times (m-n)) \end{aligned}$$

Thus,

$$R = [S\ G] \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_m \end{bmatrix} \begin{bmatrix} S^* \\ G^* \end{bmatrix}$$

Eigendecomposition of R and Its Properties, con't

As $\text{rank}(APA^*) = n$:

$$\begin{array}{ll} \lambda_k > \sigma^2 & k = 1, \ldots, n \\ \lambda_k = \sigma^2 & k = n+1, \ldots, m \end{array}$$

$$\mathring{\Lambda} = \begin{bmatrix} \lambda_1 - \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \lambda_n - \sigma^2 \end{bmatrix} = \text{ nonsingular}$$

Note:

$$RS = APA^*S + \sigma^2 S = S \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix}$$

$$\boxed{S = A(PA^*S\mathring{\Lambda}^{-1}) \triangleq AC}$$

with $\boxed{|C| \neq 0}$ (since $\text{rank}(S) = \text{rank}(A) = n$).
Therefore, since $S^*G = 0$,

$$\boxed{A^*G = 0}$$

MUSIC Method

$$A^*G = \begin{bmatrix} a^*(\omega_1) \\ \vdots \\ a^*(\omega_n) \end{bmatrix} G = 0$$

$$\Rightarrow \ \{a(\omega_k)\}_{k=1}^{n} \perp \mathcal{R}(G)$$

Thus,

$$\boxed{\{\omega_k\}_{k=1}^{n} \text{ are the } \textit{unique solutions} \text{ of } \quad a^*(\omega)GG^*a(\omega) = 0.}$$

Let:

$$\hat{R} = \frac{1}{N} \sum_{t=m}^{N} \tilde{y}(t)\tilde{y}^*(t)$$

$$\hat{S}, \hat{G} = S, G \text{ made from the eigenvectors of } \hat{R}$$

Spectral and Root MUSIC Methods

Spectral MUSIC Method:

$\{\hat{\omega}_k\}_{k=1}^n =$ the locations of the n highest peaks of the "pseudo-spectrum" function:

$$\boxed{\frac{1}{a^*(\omega)\hat{G}\hat{G}^* a(\omega)}, \quad \omega \in [-\pi, \pi]}$$

Root MUSIC Method:

$\{\hat{\omega}_k\}_{k=1}^n =$ the angular positions of the n roots of:

$$\boxed{a^T(z^{-1})\hat{G}\hat{G}^* a(z) = 0}$$

that are closest to the unit circle. Here,

$$a(z) = [1, z^{-1}, \ldots, z^{-(m-1)}]^T$$

Note: Both variants of MUSIC may produce spurious frequency estimates.

Pisarenko Method

Pisarenko is a special case of MUSIC with $m = n + 1$ (the minimum possible value).

If: $m = n + 1$

Then: $\widehat{G} = \hat{g}_1$,
$\Rightarrow \; \{\hat{\omega}_k\}_{k=1}^{n}$ can be found from the roots of

$$\boxed{a^T(z^{-1})\hat{g}_1 = 0}$$

- no problem with spurious frequency estimates

- computationally simple

- (much) less accurate than MUSIC with $m \gg n + 1$

Min-Norm Method

Goals: Reduce computational burden, and reduce risk of false frequency estimates.

Uses $m \gg n$ (as in MUSIC), but only *one* vector in $\mathcal{R}(G)$ (as in Pisarenko).

Let

$$\begin{bmatrix} 1 \\ \hat{g} \end{bmatrix} = \text{the vector in } \mathcal{R}(\hat{G})\text{, with first element equal to one, that has minimum Euclidean norm.}$$

Min-Norm Method, con't

Spectral Min-Norm

$\{\hat{\omega}\}_{k=1}^{n} =$ the locations of the n highest peaks in the "pseudo-spectrum"

$$\boxed{1 \,/\, \left| a^*(\omega) \begin{bmatrix} 1 \\ \hat{g} \end{bmatrix} \right|^2}$$

Root Min-Norm

$\{\hat{\omega}\}_{k=1}^{n} =$ the angular positions of the n roots of the polynomial

$$\boxed{a^T(z^{-1}) \begin{bmatrix} 1 \\ \hat{g} \end{bmatrix}}$$

that are closest to the unit circle.

Min-Norm Method: Determining $\hat{g}$

Let $\widehat{S} = \begin{bmatrix} \alpha^* \\ \bar{S} \end{bmatrix} \begin{matrix} \} \; 1 \\ \} \; m-1 \end{matrix}$

Then:

$$\begin{bmatrix} 1 \\ \hat{g} \end{bmatrix} \in \mathcal{R}(\widehat{G}) \;\Rightarrow\; \widehat{S}^* \begin{bmatrix} 1 \\ \hat{g} \end{bmatrix} = 0$$

$$\Rightarrow\; \bar{S}^* \hat{g} = -\alpha$$

Min-Norm solution: $\hat{g} = -\bar{S}(\bar{S}^*\bar{S})^{-1}\alpha$

As: $I = \widehat{S}^*\widehat{S} = \alpha\alpha^* + \bar{S}^*\bar{S}$, $(\bar{S}^*\bar{S})^{-1}$ exists iff

$$\boxed{\alpha^*\alpha = \|\alpha\|^2 \neq 1}$$

(This holds, at least, for $N \gg 1$.)

Multiplying the above equation by α gives:

$$\begin{aligned} &\alpha(1 - \|\alpha\|^2) = (\bar{S}^*\bar{S})\alpha \\ &\Rightarrow\; (\bar{S}^*\bar{S})^{-1}\alpha = \alpha/(1 - \|\alpha\|^2) \\ &\Rightarrow\; \boxed{\hat{g} = -\bar{S}\alpha/(1 - \|\alpha\|^2)} \end{aligned}$$

ESPRIT Method

Let
$$\begin{aligned} A_1 &= [I_{m-1} \ \ 0]A \\ A_2 &= [0 \ \ I_{m-1}]A \end{aligned}$$

Then $A_2 = A_1 D$, where

$$D = \begin{bmatrix} e^{-i\omega_1} & & 0 \\ & \ddots & \\ 0 & & e^{-i\omega_n} \end{bmatrix}$$

Also, let
$$\begin{aligned} S_1 &= [I_{m-1} \ \ 0]S \\ S_2 &= [0 \ \ I_{m-1}]S \end{aligned}$$

Recall $S = AC$ with $|C| \neq 0$. Then

$$S_2 = A_2 C = A_1 D C = S_1 \underbrace{C^{-1} D C}_{\phi}$$

So ϕ has the same eigenvalues as D. ϕ is uniquely determined as

$$\phi = (S_1^* S_1)^{-1} S_1^* S_2$$

ESPRIT Implementation

From the eigendecomposition of $\hat{R}$, find $\hat{S}$, then $\hat{S}_1$ and $\hat{S}_2$.

The frequency estimates are found by:

$$\{\hat{\omega}_k\}_{k=1}^{n} = -\arg(\hat{\nu}_k)$$

where $\{\hat{\nu}_k\}_{k=1}^{n}$ are the eigenvalues of

$$\boxed{\hat{\phi} = (\hat{S}_1^* \hat{S}_1)^{-1} \hat{S}_1^* \hat{S}_2}$$

ESPRIT Advantages:

- computationally simple
- no extraneous frequency estimates (unlike in MUSIC or Min–Norm)
- accurate frequency estimates

Summary of Frequency Estimation Methods

Method	Computational Burden	Accuracy / Resolution	Risk for False Freq Estimates
Periodogram	small	medium-high	medium
Nonlinear LS	very high	very high	very high
Yule-Walker	medium	high	medium
Pisarenko	small	low	none
MUSIC	high	high	medium
Min-Norm	medium	high	small
ESPRIT	medium	very high	none

Recommendation:

- Use **Periodogram** for medium-resolution applications
- Use **ESPRIT** for high-resolution applications

Filter Bank Methods

Lecture 7

Basic Ideas

Two main PSD estimation approaches:

1. *Parametric Approach:* Parameterize $\phi(\omega)$ by a finite-dimensional model.

2. *Nonparametric Approach:* Implicitly smooth $\{\phi(\omega)\}_{\omega=-\pi}^{\pi}$ by assuming that $\phi(\omega)$ is nearly constant over the bands

$$[\omega - \beta\pi,\ \omega + \beta\pi],\ \beta \ll 1$$

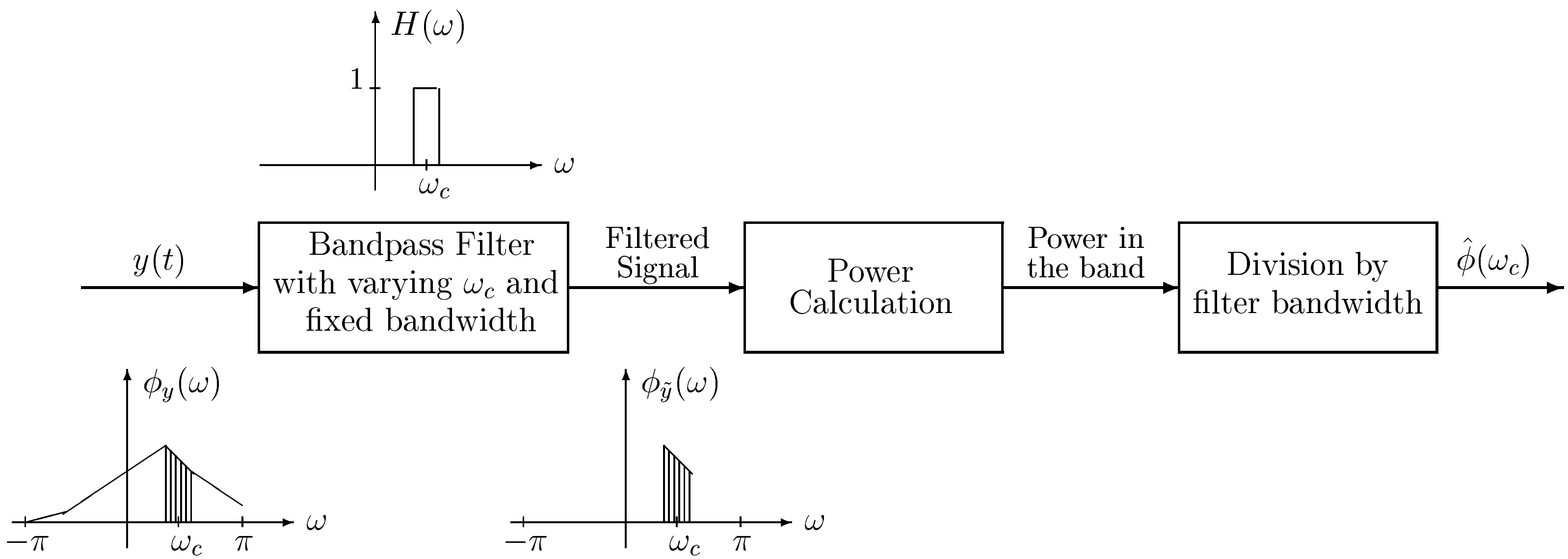
2 is more general than 1, but 2 requires

$$\boxed{N\beta > 1}$$

to ensure that the number of estimated values ($= 2\pi/2\pi\beta = 1/\beta$) is $< N$.

$N\beta > 1$ leads to the variability / resolution compromise associated with all nonparametric methods.

Filter Bank Approach to Spectral Estimation



$$\hat{\phi}_{FB}(\omega) \overset{(a)}{\simeq} \frac{1}{2\pi}\int_{-\pi}^{\pi} |H(\tau)|^2 \phi(\tau) d\tau/\beta \overset{(b)}{\simeq} \frac{1}{2\pi}\int_{\omega-\pi\beta}^{\omega+\pi\beta} \phi(\tau) d\tau/\beta \overset{(c)}{\simeq} \phi(\omega)$$

(a) consistent power calculation

(b) Ideal passband filter with bandwidth β

(c) $\phi(\tau)$ constant on $\tau \in [\omega - 2\pi\beta, \omega + 2\pi\beta]$

Note that assumptions (a) and (b), as well as (b) and (c), are conflicting.

Filter Bank Interpretation of the Periodogram

$$
\begin{aligned}
\hat{\phi}_p(\tilde{\omega}) &\triangleq \frac{1}{N}\left|\sum_{t=1}^{N} y(t)e^{-i\tilde{\omega}t}\right|^2 \\
&= \frac{1}{N}\left|\sum_{t=1}^{N} y(t)e^{i\tilde{\omega}(N-t)}\right|^2 \\
&= N\left|\sum_{k=0}^{\infty} h_k y(N-k)\right|^2
\end{aligned}
$$

where

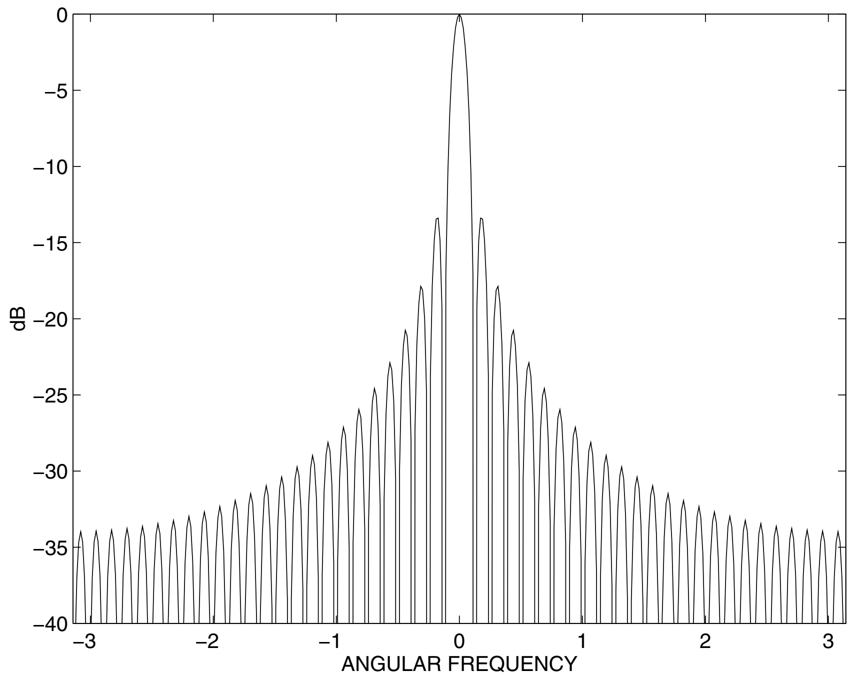
$$
h_k = \begin{cases} \frac{1}{N}e^{i\tilde{\omega}k}, & k = 0,\ldots,N-1 \\ 0, & \text{otherwise} \end{cases}
$$

$$
H(\omega) = \sum_{k=0}^{\infty} h_k e^{-i\omega k} = \frac{1}{N}\,\frac{e^{iN(\tilde{\omega}-\omega)}-1}{e^{i(\tilde{\omega}-\omega)}-1}
$$

- center frequency of $H(\omega) = \tilde{\omega}$
- 3dB bandwidth of $H(\omega) \simeq 1/N$

Filter Bank Interpretation of the Periodogram, con't

$|H(\omega)|$ as a function of $(\tilde{\omega} - \omega)$, for $N = 50$.



Conclusion: The periodogram $\hat{\phi}_p(\omega)$ is a filter bank PSD estimator with bandpass filter as given above, and:

- narrow filter passband,

- power calculation from only **1** sample of filter output.

Possible Improvements to the Filter Bank Approach

1. *Split the available sample*, and bandpass filter each subsample.

   - more data points for the power calculation stage.

   This approach leads to Bartlett and Welch methods.

2. Use *several bandpass filters on the whole sample.* Each filter covers a small band centered on $\tilde{\omega}$.

   - provides several samples for power calculation.

   This "multiwindow approach" is similar to the Daniell method.

Both approaches *compromise bias for variance*, and in fact are quite related to each other: splitting the data sample can be interpreted as a special form of windowing or filtering.

Capon Method

Idea: Data-dependent bandpass filter design.

$$\begin{aligned} y_F(t) &= \sum_{k=0}^{m} h_k y(t-k) \\ &= \underbrace{[h_0\ h_1\ \ldots\ h_m]}_{h^*} \underbrace{\begin{bmatrix} y(t) \\ \vdots \\ y(t-m) \end{bmatrix}}_{\tilde{y}(t)} \end{aligned}$$

$$E\left\{|y_F(t)|^2\right\} = h^* R h, \quad R = E\left\{\tilde{y}(t)\tilde{y}^*(t)\right\}$$

$$H(\omega) = \sum_{k=0}^{m} h_k e^{-i\omega k} = h^* a(\omega)$$

where $a(\omega) = [1, e^{-i\omega}\ \ldots\ e^{-im\omega}]^T$

Capon Method, con't

Capon Filter Design Problem:

$$\boxed{\min_h (h^* R h) \quad \text{subject to } h^* a(\omega) = 1}$$

Solution: $\boxed{h_0 = R^{-1}a / a^* R^{-1} a}$

The power at the filter output is:

$$E\left\{|y_F(t)|^2\right\} = h_0^* R h_0 = 1/a^*(\omega) R^{-1} a(\omega)$$

which should be the power of $y(t)$ in a passband centered on ω.

The Bandwidth $\simeq \frac{1}{m+1} = \frac{1}{(\text{filter length})}$

Conclusion Estimate PSD as:

$$\boxed{\hat{\phi}(\omega) = \frac{m+1}{a^*(\omega)\hat{R}^{-1}a(\omega)}}$$

with

$$\hat{R} = \frac{1}{N-m} \sum_{t=m+1}^{N} \tilde{y}(t)\tilde{y}^*(t)$$

Capon Properties

- m is the user parameter that controls the compromise between bias and variance:
  - as m increases, bias decreases and variance increases.

- Capon uses one bandpass filter only, but it splits the N-data point sample into $(N - m)$ subsequences of length m with maximum overlap.

Relation between Capon and Blackman-Tukey Methods

Consider $\hat{\phi}_{BT}(\omega)$ with *Bartlett window*:

$$
\begin{aligned}
\hat{\phi}_{BT}(\omega) &= \sum_{k=-m}^{m} \frac{m+1-|k|}{m+1} \hat{r}(k) e^{-i\omega k} \\
&= \frac{1}{m+1} \sum_{t=0}^{m} \sum_{s=0}^{m} \hat{r}(t-s) e^{-i\omega(t-s)} \\
&= \frac{a^*(\omega) \hat{R} a(\omega)}{m+1}; \quad \hat{R} = [\hat{r}(i-j)]
\end{aligned}
$$

Then we have

$$
\begin{aligned}
\hat{\phi}_{BT}(\omega) &= \frac{a^*(\omega) \hat{R} a(\omega)}{m+1} \\
\hat{\phi}_{C}(\omega) &= \frac{m+1}{a^*(\omega) \hat{R}^{-1} a(\omega)}
\end{aligned}
$$

Relation between Capon and AR Methods

Let

$$\hat{\phi}_k^{\mathrm{AR}}(\omega) = \frac{\hat{\sigma}_k^2}{|\hat{A}_k(\omega)|^2}$$

be the kth order AR PSD estimate of $y(t)$.

Then

$$\boxed{\hat{\phi}_C(\omega) = \frac{1}{\frac{1}{m+1}\sum_{k=0}^{m} 1/\hat{\phi}_k^{\mathrm{AR}}(\omega)}}$$

Consequences:

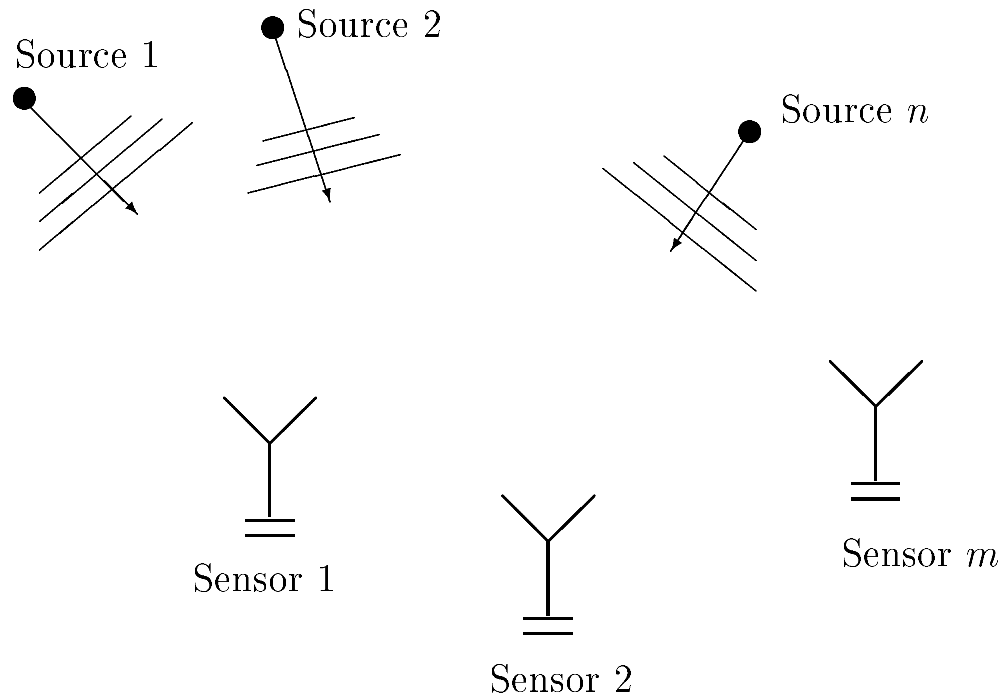
- Due to the average over k, $\hat{\phi}_C(\omega)$ generally has *less statistical variability* than the AR PSD estimator.

- Due to the low-order AR terms in the average, $\hat{\phi}_C(\omega)$ generally has *worse resolution and bias properties* than the AR method.

Spatial Methods — Part 1

Lecture 8

The Spatial Spectral Estimation Problem



Problem: *Detect* and *locate* n radiating sources by using an array of m passive sensors.

Emitted energy: Acoustic, electromagnetic, mechanical

Receiving sensors: Hydrophones, antennas, seismometers

Applications: Radar, sonar, communications, seismology, underwater surveillance

Basic Approach: Determine energy distribution over *space* (thus the name "spatial spectral analysis")

Simplifying Assumptions

- Far-field sources in the same plane as the array of sensors
- Non-dispersive wave propagation

Hence: The waves are planar and the only location parameter is **direction of arrival (DOA)** (or angle of arrival, AOA).

- The number of sources n is known. (We do not treat the detection problem)
- The sensors are linear dynamic elements with *known transfer characteristics* and *known locations*

  (That is, the array is *calibrated.*)

Array Model — Single Emitter Case

$x(t) =$ the signal waveform as measured at a reference point (*e.g.*, at the "first" sensor)

$\tau_k =$ the delay between the reference point and the kth sensor

$h_k(t) =$ the impulse response (weighting function) of sensor k

$\bar{e}_k(t) =$ "noise" at the kth sensor (*e.g.*, thermal noise in sensor electronics; background noise, etc.)

Note: $t \in \mathcal{R}$ (continuous-time signals).

Then the output of sensor k is

$$\bar{y}_k(t) = h_k(t) * x(t - \tau_k) + \bar{e}_k(t)$$

($* =$ convolution operator).

Basic Problem: Estimate the *time delays* $\{\tau_k\}$ with $h_k(t)$ known but $x(t)$ unknown.

This is a *time-delay estimation problem* in the unknown input case.

Narrowband Assumption

Assume: The emitted signals are narrowband with known carrier frequency ω_c.

Then: $x(t) = \alpha(t) \cos[\omega_c t + \varphi(t)]$

where $\alpha(t),\ \varphi(t)$ vary "slowly enough" so that

$$\alpha(t - \tau_k) \simeq \alpha(t), \qquad \varphi(t - \tau_k) \simeq \varphi(t)$$

Time delay is now $\simeq$ to a *phase shift* $\omega_c \tau_k$:

$$x(t - \tau_k) \simeq \alpha(t) \cos[\omega_c t + \varphi(t) - \omega_c \tau_k]$$

$$\begin{aligned} & h_k(t) * x(t - \tau_k) \\ & \quad \simeq |H_k(\omega_c)| \alpha(t) \cos[\omega_c t + \varphi(t) - \omega_c \tau_k + \arg\{H_k(\omega_c)\}] \end{aligned}$$

where $H_k(\omega) = \mathcal{F}\{h_k(t)\}$ is the kth sensor's transfer function

Hence, the kth sensor output is

$$\begin{aligned} \bar{y}_k(t) &= |H_k(\omega_c)| \alpha(t) \\ & \quad \cdot \cos[\omega_c t + \varphi(t) - \omega_c \tau_k + \arg H_k(\omega_c)] + \bar{e}_k(t) \end{aligned}$$

Complex Signal Representation

The noise-free output has the form:

$$\begin{aligned} z(t) &= \beta(t) \cos\left[\omega_c t + \psi(t)\right] = \\ &= \frac{\beta(t)}{2} \left\{ e^{i[\omega_c t + \psi(t)]} + e^{-i[\omega_c t + \psi(t)]} \right\} \end{aligned}$$

Demodulate $z(t)$ (translate to baseband):

$$2z(t)e^{-\omega_c t} = \beta(t)\{ \underbrace{e^{i\psi(t)}}_{\text{lowpass}} + \underbrace{e^{-i[2\omega_c t + \psi(t)]}}_{\text{highpass}} \}$$

Lowpass filter $2z(t)e^{-i\omega_c t}$ to obtain $\beta(t)e^{i\psi(t)}$

Hence, by low-pass filtering and sampling the signal

$$\begin{aligned} \tilde{y}_k(t)/2 &= \bar{y}_k(t)e^{-i\omega_c t} \\ &= \bar{y}_k(t)\cos(\omega_c t) - i\bar{y}_k(t)\sin(\omega_c t) \end{aligned}$$

we get the **complex representation**: (for $t \in \mathcal{Z}$)

$$y_k(t) = \underbrace{\alpha(t)\ e^{i\varphi(t)}}_{s(t)}\ \underbrace{|H_k(\omega_c)|\ e^{i\arg[H_k(\omega_c)]}}_{H_k(\omega_c)}\ e^{-i\omega_c \tau_k} + e_k(t)$$

or

$$\boxed{y_k(t) = s(t)H_k(\omega_c)\ e^{-i\omega_c \tau_k} + e_k(t)}$$

where $s(t)$ is the *complex envelope* of $x(t)$.

Vector Representation for a Narrowband Source

Let

$$\begin{aligned} \theta &= \text{the emitter DOA} \\ m &= \text{the number of sensors} \\ a(\theta) &= \begin{bmatrix} H_1(\omega_c)\, e^{-i\omega_c \tau_1} \\ \vdots \\ H_m(\omega_c)\, e^{-i\omega_c \tau_m} \end{bmatrix} \end{aligned}$$

$$y(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_m(t) \end{bmatrix} \qquad e(t) = \begin{bmatrix} e_1(t) \\ \vdots \\ e_m(t) \end{bmatrix}$$

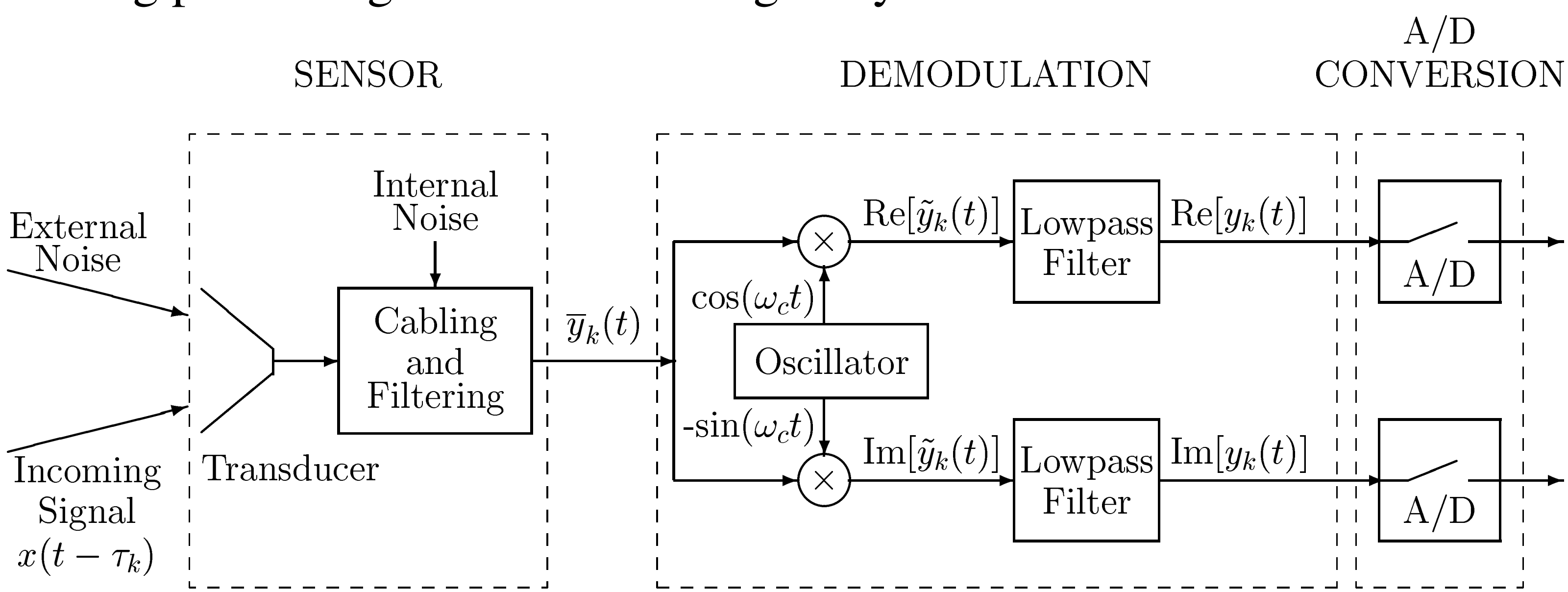
Then

$$\boxed{y(t) = a(\theta)s(t) + e(t)}$$

NOTE: θ enters $a(\theta)$ via both $\{\tau_k\}$ and $\{H_k(\omega_c)\}$. For *omnidirectional* sensors the $\{H_k(\omega_c)\}$ do not depend on θ.

Analog Processing Block Diagram

Analog processing for each receiving array element



Multiple Emitter Case

Given n emitters with

- received signals: $\{s_k(t)\}_{k=1}^{n}$
- DOAs: θ_k

Linear sensors $\Rightarrow$

$$y(t) = a(\theta_1)s_1(t) + \cdots + a(\theta_n)s_n(t) + e(t)$$

Let

$$A = [a(\theta_1) \ldots a(\theta_n)],\ (m \times n)$$

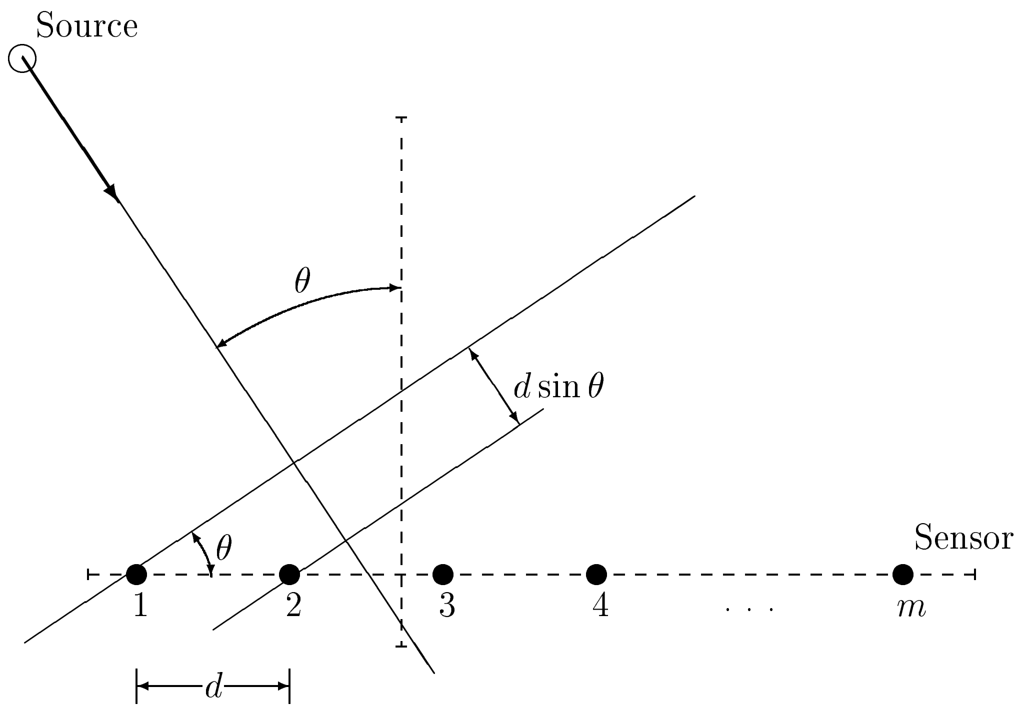
$$s(t) = [s_1(t) \ldots s_n(t)]^T,\ (n \times 1)$$

Then, the **array equation** is:

$$\boxed{y(t) = As(t) + e(t)}$$

Use the *planar wave* assumption to find the dependence of τ_k on θ.

Uniform Linear Arrays



ULA Geometry

Sensor #1 = time delay reference

Time Delay for sensor k:

$$\tau_k = (k-1)\,\frac{d\sin\theta}{c}$$

where c = wave propagation speed

Spatial Frequency

Let:

$$\begin{aligned}
\omega_s &\stackrel{\triangle}{=} \omega_c \frac{d \sin\theta}{c} = 2\pi \frac{d\sin\theta}{c/f_c} = 2\pi \frac{d\sin\theta}{\lambda} \\
\lambda &= c/f_c = \textit{signal wavelength} \\
a(\theta) &= [1, e^{-i\omega_s} \ \ldots \ e^{-i(m-1)\omega_s}]^T
\end{aligned}$$

By direct analogy with the vector $a(\omega)$ made from uniform samples of a *sinusoidal time series*,

$$\omega_s = \text{ spatial frequency}$$

The function $\omega_s \mapsto a(\theta)$ is *one-to-one* for

$$|\omega_s| \leq \pi \leftrightarrow \frac{d|\sin\theta|}{\lambda/2} \leq 1 \leftarrow \boxed{d \leq \lambda/2}$$

As

$$d = \text{spatial sampling period}$$

$d \leq \lambda/2$ is a **spatial** Shannon sampling theorem.

Spatial Methods — Part 2

Lecture 9

Spatial Filtering

Spatial filtering useful for

- DOA discrimination (similar to frequency discrimination of time-series filtering)
- Nonparametric DOA estimation

There is a strong analogy between temporal filtering and spatial filtering.

Analogy between Temporal and Spatial Filtering

Temporal FIR Filter:

$$
\begin{aligned}
y_F(t) &= \sum_{k=0}^{m-1} h_k u(t-k) = h^* y(t) \\
h &= [h_o \ldots h_{m-1}]^* \\
y(t) &= [u(t) \ldots u(t-m+1)]^T
\end{aligned}
$$

If $u(t) = e^{i\omega t}$ then

$$
y_F(t) = \underbrace{[h^* a(\omega)]}_{\text{filter transfer function}} u(t)
$$

$$
a(\omega) = [1, e^{-i\omega} \ \ldots \ e^{-i(m-1)\omega}]^T
$$

We can select h to enhance or attenuate signals with different frequencies ω.

Analogy between Temporal and Spatial Filtering

Spatial Filter:

$\{y_k(t)\}_{k=1}^{m} =$ the "spatial samples" obtained with a sensor array.

Spatial FIR Filter output:

$$y_F(t) = \sum_{k=1}^{m} h_k y_k(t) = h^* y(t)$$

Narrowband Wavefront: The array's (noise-free) response to a narrowband ($\sim$ sinusoidal) wavefront with complex envelope $s(t)$ is:

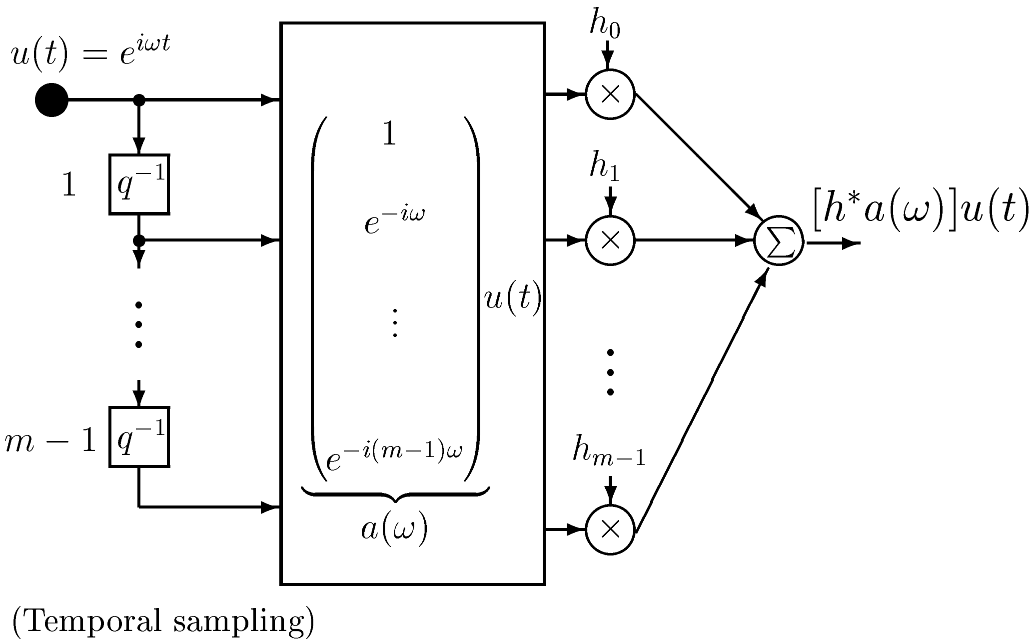
$$\begin{aligned} y(t) &= a(\theta)s(t) \\ a(\theta) &= [1, e^{-i\omega_c\tau_2} \; \ldots \; e^{-i\omega_c\tau_m}]^T \end{aligned}$$

The corresponding filter output is

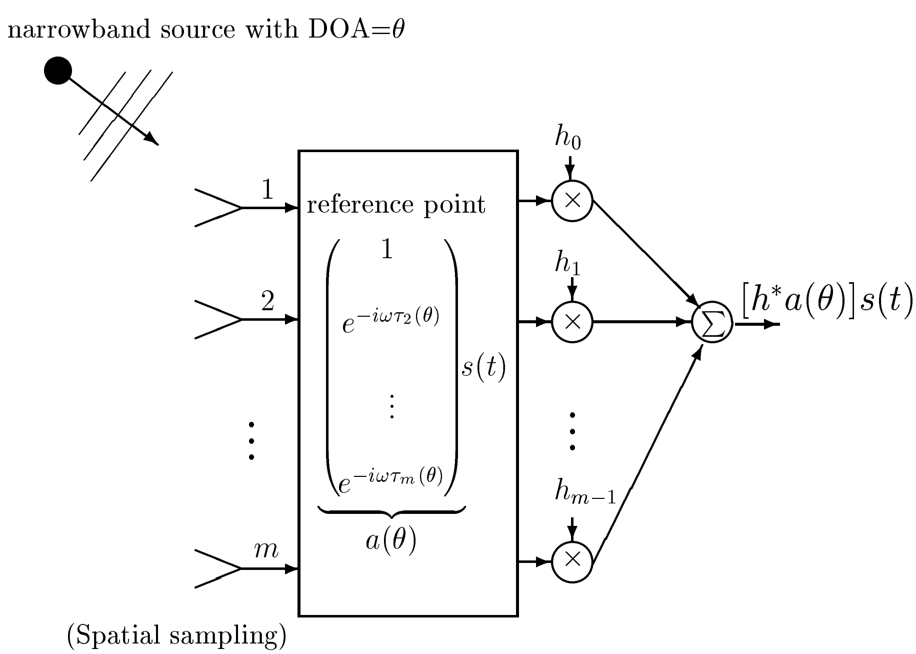
$$\boxed{y_F(t) = \underbrace{[h^* a(\theta)]}_{\text{filter transfer function}} s(t)}$$

We can select h to enhance or attenuate signals coming from different DOAs.

Analogy between Temporal and Spatial Filtering



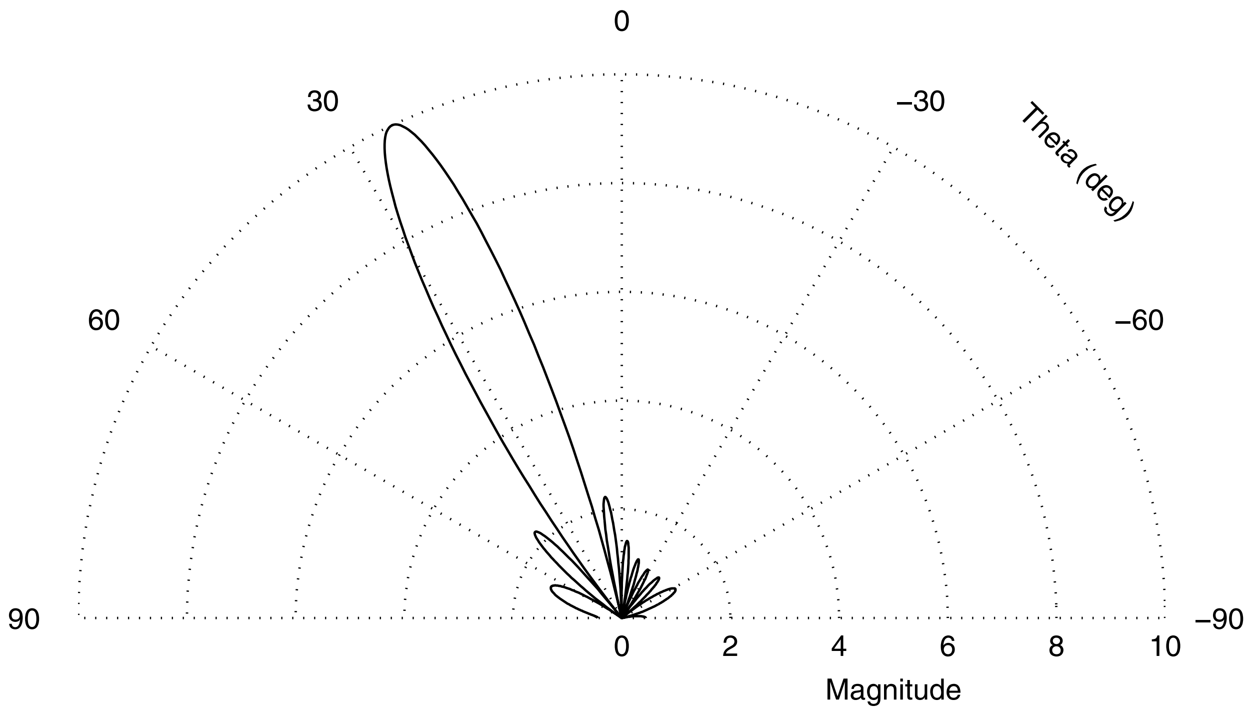
(a) Temporal filter



(b) Spatial filter

Spatial Filtering, con't

Example: The response magnitude $|h^* a(\theta)|$ of a spatial filter (or beamformer) for a 10-element ULA. Here, $h = a(\theta_0)$, where $\theta_0 = 25^\circ$



Spatial Filtering Uses

Spatial Filters can be used

- To pass the signal of interest only, hence filtering out interferences located outside the filter' s beam (but possibly having the same temporal characteristics as the signal).
- To locate an emitter in the field of view, by sweeping the filter through the DOA range of interest ("goniometer").

Nonparametric Spatial Methods

A *Filter Bank Approach* to DOA estimation.

Basic Ideas

- Design a filter $h(\theta)$ such that for each θ
  - It passes undistorted the signal with DOA = θ
  - It attenuates all DOAs $\neq \theta$
- Sweep the filter through the DOA range of interest, and evaluate the powers of the filtered signals:

$$\begin{aligned} E\left\{|y_F(t)|^2\right\} &= E\left\{|h^*(\theta)y(t)|^2\right\} \\ &= h^*(\theta)Rh(\theta) \end{aligned}$$

  with $R = E\{y(t)y^*(t)\}$.

- The (dominant) peaks of $h^*(\theta)Rh(\theta)$ give the DOAs of the sources.

Beamforming Method

Assume the array output is *spatially white*:

$$R = E\left\{y(t)y^*(t)\right\} = I$$

Then: $E\left\{|y_F(t)|^2\right\} = h^*h$

Hence: In direct analogy with the temporally white assumption for filter bank methods, $y(t)$ can be considered as impinging on the array from *all* DOAs.

Filter Design:

$$\boxed{\min_h \, (h^*h) \;\text{ subject to }\; h^*a(\theta) = 1}$$

Solution:

$$h = a(\theta)/a^*(\theta)a(\theta) = a(\theta)/m$$

$$E\left\{|y_F(t)|^2\right\} = a^*(\theta)Ra(\theta)/m^2$$

Implementation of Beamforming

$$\hat{R} = \frac{1}{N} \sum_{t=1}^{N} y(t)y^*(t)$$

The beamforming DOA estimates are:

$$\boxed{\begin{aligned}\{\hat{\theta}_k\} = \; & \text{the locations of the } n \text{ largest peaks of} \\ & a^*(\theta)\hat{R}a(\theta).\end{aligned}}$$

This is the direct spatial analog of the Blackman-Tukey periodogram.

Resolution Threshold:

$$\begin{aligned}\inf \; |\theta_k - \theta_p| \; & > \; \frac{\text{wavelength}}{\text{array length}} \\ & = \; \text{array beamwidth}\end{aligned}$$

Inconsistency problem:

Beamforming DOA estimates are consistent if $n = 1$, but inconsistent if $n > 1$.

Capon Method

Filter design:

$$\boxed{\min_h (h^* R h) \text{ subject to } h^* a(\theta) = 1}$$

Solution:

$$\begin{aligned} h &= R^{-1} a(\theta) / a^*(\theta) R^{-1} a(\theta) \\ E\left\{ |y_F(t)|^2 \right\} &= 1 / a^*(\theta) R^{-1} a(\theta) \end{aligned}$$

Implementation:

$$\boxed{\begin{aligned} \{\hat{\theta}_k\} = \; & \text{the locations of the } n \text{ largest peaks of} \\ & 1/a^*(\theta) \hat{R}^{-1} a(\theta). \end{aligned}}$$

Performance: Slightly superior to Beamforming.

Both Beamforming and Capon are *nonparametric* approaches. They do not make assumptions on the covariance properties of the data (and hence do not depend on them).

Parametric Methods

Assumptions:

- The array is described by the equation:

$$y(t) = As(t) + e(t)$$

- The noise is spatially white and has the same power in all sensors:

$$E\{e(t)e^*(t)\} = \sigma^2 I$$

- The signal covariance matrix

$$P = E\{s(t)s^*(t)\}$$

  is nonsingular.

Then:

$$R = E\{y(t)y^*(t)\} = APA^* + \sigma^2 I$$

Thus: The NLS, YW, MUSIC, MIN-NORM and ESPRIT methods of frequency estimation can be used, almost without modification, for DOA estimation.

Nonlinear Least Squares Method

$$\min_{\{\theta_k\},\,\{s(t)\}} \underbrace{\frac{1}{N}\sum_{t=1}^{N} \|y(t) - As(t)\|^2}_{f(\theta,s)}$$

Minimizing f over s gives

$$\hat{s}(t) = (A^*A)^{-1}A^*y(t), \quad t = 1, \ldots, N$$

Then

$$\begin{aligned} f(\theta, \hat{s}) &= \frac{1}{N}\sum_{t=1}^{N} \| \, [I - A(A^*A)^{-1}A^*]y(t)\|^2 \\ &= \frac{1}{N}\sum_{t=1}^{N} y^*(t)[I - A(A^*A)^{-1}A^*]y(t) \\ &= \operatorname{tr}\{[I - A(A^*A)^{-1}A^*]\hat{R}\} \end{aligned}$$

Thus, $\boxed{\{\hat{\theta}_k\} = \arg\max_{\{\theta_k\}} \operatorname{tr}\{[A(A^*A)^{-1}A^*]\hat{R}\}}$

For $N = 1$, this is precisely the form of the NLS method of frequency estimation.

Nonlinear Least Squares Method

Properties of NLS:

- Performance: high
- Computational complexity: high
- Main drawback: need for multidimensional search.

Yule-Walker Method

$$y(t) = \begin{bmatrix} \bar{y}(t) \\ \tilde{y}(t) \end{bmatrix} = \begin{bmatrix} \bar{A} \\ \tilde{A} \end{bmatrix} s(t) + \begin{bmatrix} \bar{e}(t) \\ \tilde{e}(t) \end{bmatrix}$$

Assume: $\boxed{E\left\{\bar{e}(t)\tilde{e}^*(t)\right\} = 0}$

Then:

$$\Gamma \stackrel{\triangle}{=} E\left\{\bar{y}(t)\tilde{y}^*(t)\right\} = \bar{A}P\tilde{A}^* \quad (M \times L)$$

Also assume:

- $M > n, \ L > n \quad (\Rightarrow \ m = M + L > 2n)$
- $\operatorname{rank}(\bar{A}) = \operatorname{rank}(\tilde{A}) = n$

Then: $\operatorname{rank}(\Gamma) = n$, and the SVD of Γ is

$$\Gamma = [\underbrace{U_1}_{n} \ \underbrace{U_2}_{M-n}] \begin{bmatrix} \Sigma_{n\times n} & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix} \begin{matrix} \} \ n \\ \} \ L-n \end{matrix}$$

Properties: $\boxed{\tilde{A}^* V_2 = 0}$ $\qquad$ $\boxed{V_1 \in \mathcal{R}(\tilde{A})}$

YW-MUSIC DOA Estimator

$$\boxed{\begin{aligned}\{\hat{\theta}_k\} = & \text{ the } n \text{ largest peaks of} \\ & 1/\tilde{a}^*(\theta)\hat{V}_2\hat{V}_2^*\tilde{a}(\theta)\end{aligned}}$$

where

- $\tilde{a}(\theta)$, $(L \times 1)$, is the "array transfer vector" for $\tilde{y}(t)$ at DOA θ
- $\hat{V}_2$ is defined similarly to V_2, using

$$\hat{\Gamma} = \frac{1}{N} \sum_{t=1}^{N} \bar{y}(t)\tilde{y}^*(t)$$

Properties:

- Computational complexity: medium
- Performance: satisfactory if $m \gg 2n$
- Main advantages:
  - weak assumption on $\{e(t)\}$
  - the subarray $\bar{A}$ need not be calibrated

MUSIC and Min-Norm Methods

Both MUSIC and Min-Norm methods for frequency estimation apply with only minor modifications to the DOA estimation problem.

- Spectral forms of MUSIC and Min-Norm can be used for arbitrary arrays
- Root forms can be used only with ULAs
- MUSIC and Min-Norm break down if the source signals are coherent; that is, if

$$\text{rank}(P) = \text{rank}(E\left\{s(t)s^*(t)\right\}) < n$$

  Modifications that apply in the coherent case exist.

ESPRIT Method

Assumption: The array is made from *two identical subarrays* separated by a *known displacement vector.*

Let

$$\begin{aligned} \bar{m} &= \text{\# sensors in each subarray} \\ A_1 &= [I_{\bar{m}}\ 0]A \quad \text{(transfer matrix of subarray 1)} \\ A_2 &= [0\ I_{\bar{m}}]A \quad \text{(transfer matrix of subarray 2)} \end{aligned}$$

Then $A_2 = A_1 D$, where

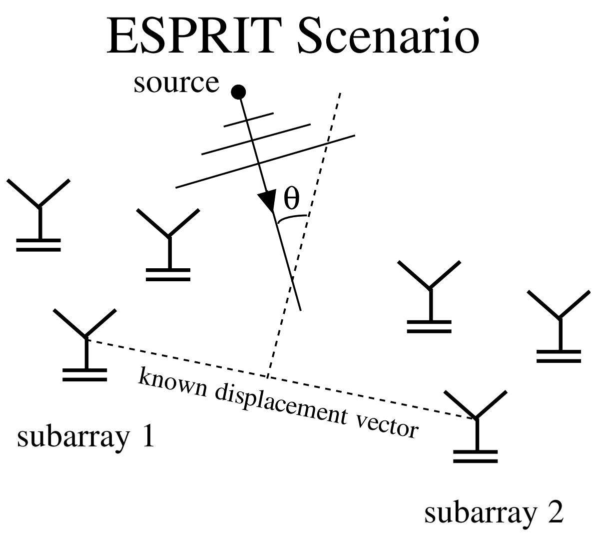
$$D = \begin{bmatrix} e^{-i\omega_c \tau(\theta_1)} & & 0 \\ & \ddots & \\ 0 & & e^{-i\omega_c \tau(\theta_n)} \end{bmatrix}$$

$\tau(\theta) =$ the time delay from subarray 1 to subarray 2 for a signal with DOA = θ:

$$\tau(\theta) = d \sin(\theta)/c$$

where d is the subarray separation and θ is measured from the perpendicular to the subarray displacement vector.

ESPRIT Method, con't



Properties:

- Requires special array geometry
- Computationally efficient
- *No risk* of spurious DOA estimates
- Does not require array calibration

Note: For a ULA, the two subarrays are often the first $m-1$ and last $m-1$ array elements, so $\bar{m} = m-1$ and

$$A_1 = [I_{m-1}\ 0]A, \qquad A_2 = [0\ I_{m-1}]A$$