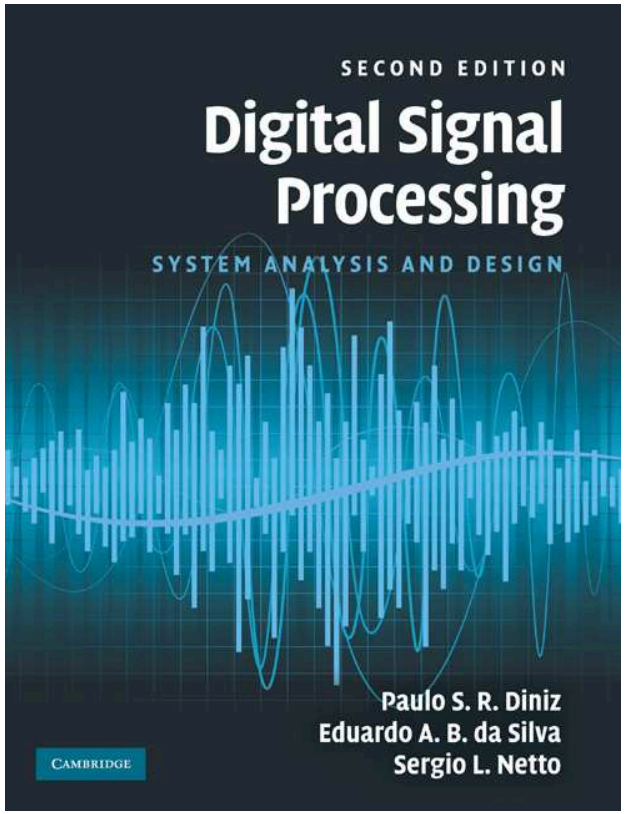
Spectral Estimation



Paulo S. R. Diniz

Eduardo A. B. da Silva

Sergio L. Netto

diniz,eduardo,sergioln@lps.ufrj.br

September 2010

Contents

- Estimation Theory
- Nonparametric spectral estimation
  - Periodogram
  - Minimum variance spectral estimator
- Modelling theory
  - Rational transfer-function models
  - Yule-Walker equations

Contents

- Parametric spectral estimation
  - Linear prediction
  - Covariance method
  - Autocorrelation method
  - Levinson-Durbin algorithm
  - Burg's method
  - Wiener filter

Spectral Estimation

- This chapter considers the very practical problem of estimating the power spectral density (PSD) of a given discrete-time signal $y(n)$.
- Typical applications include:
  - Radar/sonar systems.
  - Music transcription.
  - Speech modeling.
- The classical solution is first estimating the autocorrelation function, followed by a Fourier transform to obtain the desired spectral description.

Spectral Estimation

- Spectral estimation algorithms are classified as non-parametric or parametric methods.
- Non-parametric methods do not assume any particular structure behind the available data.
- Parametric schemes consider that the process follows some pattern characterized by a specific set of parameters pertaining to a given model.
- Parametric approaches tend to be simpler and more accurate, but they depend on some *a priori* information.

Spectral Estimation

Topics covered in this chapter:

- Basic concepts of estimation theory that are used to characterize the non-parametric methods.
- Minimum variance method.
- General theory of system modeling, characterizing the autocorrelation function for distinct system classes by the so-called Yule-Walker equations.
- PSD estimation for autoregressive systems, including the so-called autocorrelation, covariance, and Burg's methods.
- Parametric method based on linear prediction.
- Wiener solution.
- Do-It-Yourself section describing the PSD estimation of a synthetic signal.

Estimation theory

- Problem of estimation is classified in two groups.
  - The classic estimation problem: determine the value of a fixed (deterministic) unknown value.
  - The problem of estimating the value of a random parameter is referred to as Bayesian estimation.
- Consider the deterministic estimation problem. Let Θ be a real-valued parameter to be estimated from the available set of data $\{y(0), y(1), \ldots, y(L-1)\}$ associated to a random variable **y**.

Estimation theory

- The bias $\mathrm{B}(\hat{\Theta})$ of an estimate $\hat{\Theta}$ of a deterministic parameter Θ is defined as

$$\mathrm{B}(\hat{\Theta}) = \mathrm{E}\{\hat{\Theta}\} - \Theta \tag{1}$$

- If $\mathrm{B}(\hat{\Theta}) = 0$, then the estimate $\hat{\Theta}$ is called unbiased, otherwise it is referred to as biased.

- Other important characteristics of an estimator are its variance, standard deviation, and mean squared error (MSE), respectively defined as

$$\mathrm{var}\{\hat{\Theta}\} = \mathrm{E}\left\{\left(\hat{\Theta} - \mathrm{E}\{\hat{\Theta}\}\right)^2\right\} \tag{2}$$

$$\sigma_{\hat{\Theta}} = \sqrt{\mathrm{var}\{\hat{\Theta}\}} \tag{3}$$

$$\mathrm{MSE}\{\hat{\Theta}\} = \mathrm{E}\left\{\left(\hat{\Theta} - \Theta\right)^2\right\} \tag{4}$$

Estimation theory

- It is straightforward to show that

$$\text{MSE}\{\hat{\Theta}\} = \text{var}\{\hat{\Theta}\} + \text{B}^2(\hat{\Theta}) \tag{5}$$

in such a way that for an unbiased estimate we have that

$$\text{MSE}\{\hat{\Theta}\} = \text{var}\{\hat{\Theta}\} \tag{6}$$

Estimation theory

- An estimate $\hat{\Theta}$ is consistent if it converges (in probability) to the true parameter value

$$\lim_{L \to \infty} \text{Prob}\{|\hat{\Theta} - \Theta| > \epsilon\} = 0 \tag{7}$$

  where ϵ is a small positive number, with the associated variance also converging to zero.

- Consistency evaluates the estimator performance in the limiting case of sufficiently large L and is a desirable feature for an estimator.

Estimation theory

- We aim at an unbiased estimator with low variance or, equivalently, with low MSE.
- We can reduce bias at the cost of increasing the variance, and vice-versa. The natural way of reducing both simultaneously is increasing the amount of data L.
- The variance of any unbiased estimate is bounded from below by the so-called Cramer-Rao limit given by

$$\text{var}\{\hat{\Theta}\} \geq \frac{1}{E\left\{\left(\frac{\partial \ln p(\mathbf{y},\theta)}{\partial \theta}\right)^2\right\}} \tag{8}$$

  where $p(\mathbf{y}, \theta)$ is the probability density of the observations $\mathbf{y}$ including its dependence on θ, also known as likelihood function of θ.
- An unbiased estimate is referred to as efficient if its variance achieves the Cramer-Rao limit. In this case, we say that the estimate uses all available data in an efficient manner.

Nonparametric spectral estimation

- The nonparametric spectral estimation relies on the Wiener-Khinchin theorem that defines the relationship between the autocorrelation and its Fourier transform, namely the PSD.
- It explains how to estimate the PSD and/or the autocorrelation sequence of a WSS process, which is also ergodic, from a limited number of measured data. This type of estimator is called periodogram.

Periodogram

- The periodogram PSD estimator is defined by

$$\hat{\Gamma}_{Y,P}(e^{j\omega}) = \frac{1}{L}\left|\sum_{n=0}^{L-1} y(n)e^{-j\omega n}\right|^2 \tag{9}$$

- This is equivalent to using a rectangular window on the signal $y(n)$ for the time interval $0 \leq n \leq (L-1)$, squaring the absolute value of the Fourier transform of the truncated sequence, and normalizing the result by a factor L to obtain a density measure of spectral power.

Periodogram

- A simple algebraic development of equation (9) yields

$$
\begin{aligned}
\hat{\Gamma}_{Y,P}(e^{j\omega}) &= \frac{1}{L}\left(\sum_{m=0}^{L-1} y(m)e^{-j\omega m}\right)\left(\sum_{n=0}^{L-1} y(n)e^{j\omega n}\right) \\
&= \frac{1}{L}\sum_{m=0}^{L-1}\sum_{n=0}^{L-1} y(m)y(n)e^{-j\omega(m-n)} \\
&= \frac{1}{L}\Big[y(0)y(L-1)e^{-j\omega(-L+1)} \\
&\quad + (y(0)y(L-2)+y(1)y(L-1))\,e^{-j\omega(-L+2)} \\
&\quad + \ldots + \left(y^2(0)+y^2(1)+\ldots+y^2(L-1)\right)e^{-j\omega(0)} + \ldots \\
&\quad + (y(L-1)y(1)+y(L-2)y(0))\,e^{-j\omega(L-2)} \\
&\quad + y(L-1)y(0)e^{-j\omega(L-1)}\Big] \\
&= \sum_{\nu=-L+1}^{L-1} \hat{R}_{Y,b}(\nu)e^{-j\omega\nu} \qquad (10)
\end{aligned}
$$

Periodogram

- Where

$$\hat{R}_{Y,b}(\nu) = \frac{1}{L} \sum_{n=0}^{L-1-|\nu|} y(n)y(n+|\nu|) \tag{11}$$

for $\nu = -(L-1), -(L-2), \ldots, 0, \ldots, (L-2), (L-1)$, and the subscript b stands for biased.

- Therefore, the periodogram PSD applies the Wiener-Khinchin theorem in an estimated version of the autocorrelation function.

Periodogram

- By taking the expected value of equation (11), we get

$$\mathrm{E}\left\{\hat{R}_{Y,b}(\nu)\right\} = \frac{1}{L}\sum_{n=0}^{L-1-|\nu|}\mathrm{E}\{y(n)y(n+|\nu|)\} = \frac{L-|\nu|}{L}R_Y(\nu) \tag{12}$$

- Hence, the autocorrelation estimation associated to the periodogram method is, in average, the result of passing the Bartlett window

$$w_B(\nu) = \begin{cases} \frac{L-|\nu|}{L}, & \text{if } |\nu| \leq (L-1) \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

  on the true autocorrelation function $R_Y(\nu)$.

Periodogram

- In the frequency domain, the averaged periodogram PSD becomes the convolution of the true PSD function with the Fourier transform $W_B(e^{j\omega})$ of the Bartlett window, that is

$$E\{\hat{\Gamma}_{Y,P}(e^{j\omega})\} = \frac{1}{2\pi}\int_{-\pi}^{\pi} W_B(e^{j(\omega-\psi)})\Gamma_Y(e^{j\psi})\,d\psi \tag{14}$$

where

$$W_B(e^{j\omega}) = \frac{1}{L}\left(\frac{\sin\frac{\omega L}{2}}{\sin\frac{\omega}{2}}\right)^2 \tag{15}$$

Periodogram

- An important characteristic of the periodogram, as given in equation (9), is that it is readily implemented with the FFT algorithm.
- For finite L, the autocorrelation estimation $\hat{R}_{Y,b}(\nu)$ defined in equation (11) is clearly biased for $\nu \neq 0$, with the bias increasing for larger values of $|\nu|$.
- In addition, $\hat{R}_{Y,b}(\nu)$ becomes zero for all lags such as $|\nu| > (L-1)$.
- The variance of $\hat{R}_{Y,b}(\nu)$ also tends to increase with $|\nu|$, since the averaging sum is performed for less values of n.

Periodogram

- If the $\{Y\}$ process is a white noise, such that $\Gamma_Y(e^{j\omega}) = \sigma_Y^2$, the periodogram estimate is unbiased even for finite L, since

$$E\{\hat{\Gamma}_{Y,P}(e^{j\omega})\} = \frac{\sigma_Y^2}{2\pi}\int_{-\pi}^{\pi} W_B(e^{j(\omega-\psi)})\,d\psi = \sigma_Y^2 w_B(0) = \sigma_Y^2 = \Gamma_Y(e^{j\omega}) \quad (16)$$

- It can be verified that the periodogram PSD is biased for finite L, unbiased in the limiting case $L \to \infty$, and presents a constant variance, regardless the value of L, thus constituting a non-consistent estimator.

Periodogram variations

- A large set of data can be partitioned into $\frac{L}{K}$ blocks of length K each, yielding several PSD estimates that can be averaged to decrease the variance associated to the periodogram.
- This approach decreases the amount of data used in each estimate, increasing the associated bias. The averaged periodogram method exchanges bias for variance in the resulting PSD estimate.
- Another variation employs a non-rectangular window function on the whole data set $\{y(0), y(1), \ldots, y(L-1)\}$ being processed. This emphasizes the amplitude of PSD peaks, better detecting sinusoidal components in $\{Y\}$, but also widens these same peaks, what can make neighboring peaks to be seen as a single one.

Periodogram variations

- One can avoid the bias in the periodogram autocorrelation by using $(L - |\nu|)$ instead of L in the denominator of (12), yielding

$$\hat{R}_{Y,u}(\nu) = \frac{1}{L - |\nu|} \sum_{n=0}^{L-1-|\nu|} y(n)y(n + |\nu|) \tag{17}$$

such that

$$E\left\{\hat{R}_{Y,u}(\nu)\right\} = \frac{1}{L - |\nu|} \sum_{n=0}^{L-1-|\nu|} E\{y(n)y(n + |\nu|)\} = R_Y(\nu) \tag{18}$$

where u stands for unbiased.

Periodogram variations

- Replacing $\hat{R}_{Y,b}(\nu)$ with $\hat{R}_{Y,u}(\nu)$ in the last line of (10) may lead to negative values in the resulting PSD function.
- This can be overcome by introducing a window function $w(n)$, of length $(2K+1)$ with $K < L$, in the computation of the new PSD estimate

$$\hat{\Gamma}_{Y,BT}(e^{j\omega}) = \sum_{\nu=-K}^{K} w(\nu)\hat{R}_{Y,u}(\nu)e^{-j\omega\nu} \tag{19}$$

which is referred to as the Blackman-Tukey (BT) spectral estimator.

Periodogram variations

- Letting $K \ll L$ removes the noisier samples of the autocorrelation function in the computation of $\hat{\Gamma}_{Y,BT}(e^{j\omega})$, decreasing the variance in the resulting estimator, at the expense of a slight bias increase.
- To avoiding negative PSD values there are several window functions $w(n)$ for the BT spectral estimator, generating several different trade-offs between bias and variance.

Example 7.1

- Consider a signal $y(n)$ formed by three sinusoidal components as given by

$$y(n) = \sin\left(2\pi\frac{f_1}{f_s}n\right) + \sin\left(2\pi\frac{f_2}{f_s}n\right) + 5\sin\left(2\pi\frac{f_3}{f_s}n\right) \tag{20}$$

  with $f_1 = 45$ Hz, $f_2 = 55$ Hz, and $f_3 = 75$ Hz, sampled at $f_s = 400$ samples/s during a time interval of 200 ms.

- Use the periodogram, periodogram with windowed data, averaged periodogram, and BT methods to estimate the PSD function for this signal.

Example 7.1 - Solution

- From the example data, the total number of samples is $\mathrm{L} = 80$ and, where required, $\mathrm{K} = 20$.
- In Figure 1a, it is verified that the standard method yields a slightly biased estimation, since the 45- and 55-Hz peaks are slightly off mark, with the neighboring sidelobes of the 75-Hz peak almost masking the 55-Hz peak.
- The PSD estimate that applies a Hamming window to the data before calculating the autocorrelation function is able to avoid the masking effect, as seen in Figure 1b, but also widens the mainlobe of each peak, in this case merging the 45- and 55-Hz peaks into a single one.

Example 7.1 - Solution

- The averaged periodogram estimate is shown in Figure 1c where we observe the resulting variance reduction and a loss in resolution. The averaged periodogram can be shown to be asymptotically unbiased, however since the number of data samples utilized in each periodogram is smaller than L there is a loss of resolution in the order of $\frac{\mathrm{L}}{\mathrm{K}}$.
- A periodogram using the entire data record should have a resolution smaller than $\frac{2\pi}{\mathrm{L}}$ which corresponds to 5-Hz. Since $\mathrm{K} = 20$ the resolution is reduced by four explaining why the peak at 55-Hz could not be observed in Figure 1c.
- Figure 1d depicts the PSD estimate from the BT method, using the Hamming window on the estimated autocorrelation function, illustrating the excellent bias-variance compromise achieved by this technique.

Example 7.1 - Solution

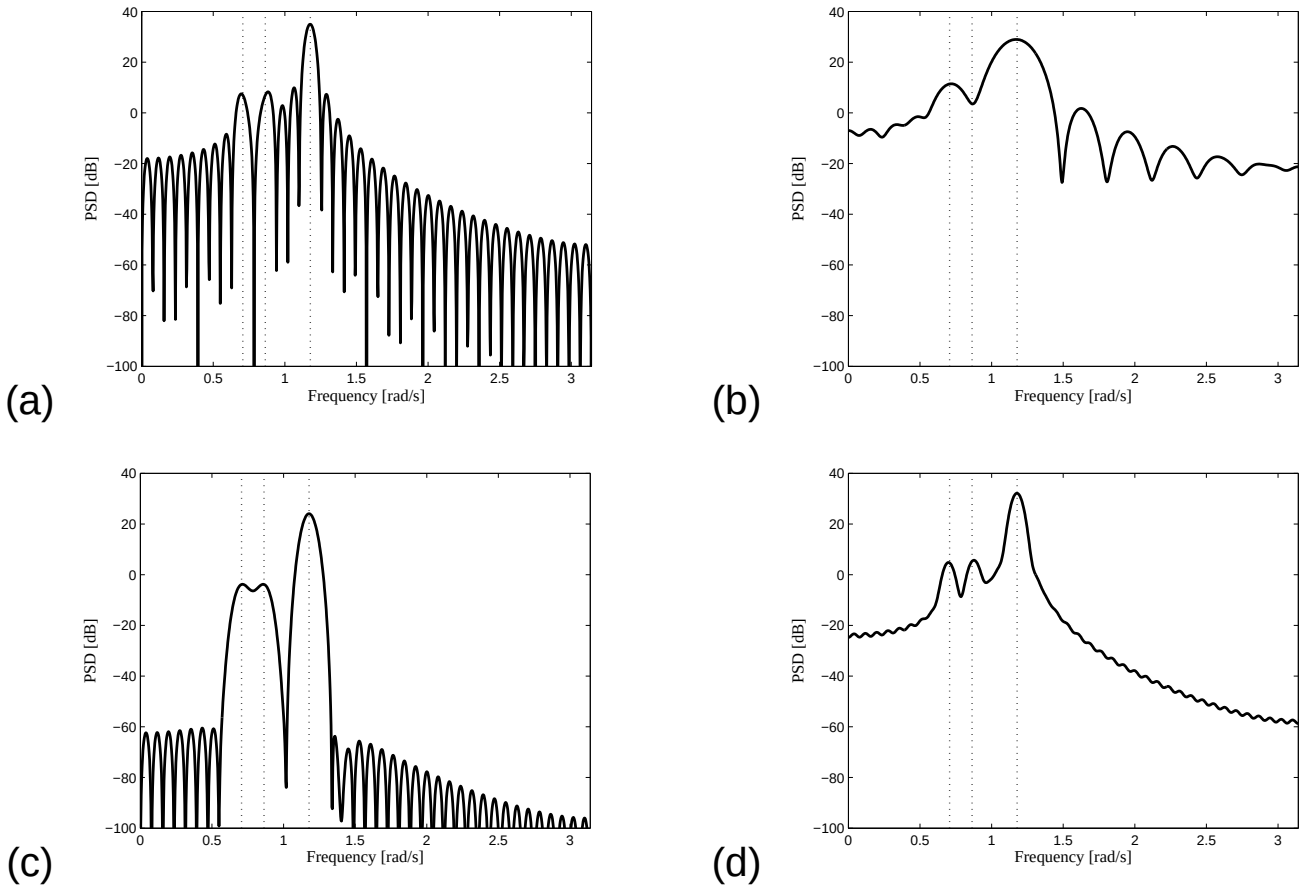


Figure 1: PSD: (a) standard periodogram; (b) periodogram with windowed data; (c) averaged periodogram; (d) Blackman-Tukey method.

Minimum variance spectral estimator

- In this section we derive yet another method to estimate the spectrum of a given signal based on an estimate of the signal power at arbitrary frequencies.
- The nonparametric spectral estimator methods presented so far are based on the periodogram as described in equation (9). This equation indicates that a rectangular window is applied to the signal $y(n)$ placed at the time interval $0 \leq n \leq (L-1)$, followed by the computation of its Fourier Transform.

Minimum variance spectral estimator

- The minimum-variance approach estimates the spectrum at frequency ω_c by filtering the signal with a narrow bandpass filter centered at ω_c and estimating the power at the output of the filter.
- The filter should be optimized for having minimum energy outside the passband. Consider the filter with center frequency ω_c is $w^*_{\omega_c}(n)$ and has length L, and its output is

$$y_{MV,\omega_c}(n) = \sum_{l=0}^{L-1} w^*_{\omega_c}(l) y(n-l) \tag{21}$$

where the subscript MV stands for minimum variance. We allow $w_{\omega_c}(n)$ to be complex.

Minimum variance spectral estimator

- The power at the filter output is

$$E\{|y_{MV,\omega_c}(n)|^2\} = R_{MV,\omega_c}(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} |W^*_{\omega_c}(e^{-j\omega})|^2 \Gamma_Y(e^{j\omega})\, d\omega \qquad (22)$$

where $W^*_{\omega_c}(e^{-j\omega})$ is the Fourier transform of the impulse response $w^*_{\omega_c}(n)$ of the passband filter centered at frequency ω_c.

Minimum variance spectral estimator

- If the passband $\Delta\omega$ is narrow enough and has unit gain, equation (22) can be written as

$$E\{|y_{MV,\omega_c}(n)|^2\} \approx \frac{\Delta\omega}{2\pi}\Gamma_Y(e^{j\omega_c}) \tag{23}$$

  which implies that the power of $y_{MV,\omega_c}(n)$ is proportional to the PSD of $y(n)$ around the frequency ω_c.

- Our aim is to obtain an accurate estimate of $\Gamma_Y(e^{j\omega_c})$ for any value of ω_c, by cleverly designing the appropriate filter.

- The key idea is to minimize the power of $y_{MV,\omega_c}(n)$ over the entire frequency range $-\pi \leq \omega \leq \pi$, while keeping the filter gain at the arbitrary central frequency ω_c equal to one.

- In this way we are able to reduce as much as possible the power contributions from frequencies away from the central frequency.

Minimum variance spectral estimator

- The objective of the MV spectral estimator is to minimize

$$\xi_{\omega_c} = E\{|y_{MV,\omega_c}(n)|^2\} \tag{24}$$

subject to

$$\sum_{l=0}^{L-1} w_l^* e^{-j\omega_c l} = 1 \tag{25}$$

for a given $-\pi \leq \omega_c \leq \pi$, where we have made $w_{\omega_c}(l) = w_l$ in order to simplify the notation.

Minimum variance spectral estimator

- By defining the auxiliary vectors

$$\mathbf{y}(n) = [y(n)\, y(n-1)\, \ldots\, y(n-L+1)]^T \quad (26)$$

$$\mathbf{w} = [w_0\, w_1\, \ldots\, w_{L-1}]^T \quad (27)$$

$$\mathbf{e}(e^{j\omega_c}) = \left[1\, e^{-j\omega_c}\, e^{-j2\omega_c}\, \ldots\, e^{-j(L-1)\omega_c}\right]^T \quad (28)$$

one may incorporate the constraint into the objective function using a Lagrange multiplier λ, generating an unconstrained minimization

$$\overline{\xi}_{\omega_c} = E[\mathbf{w}^{*^T}\mathbf{y}(n)\mathbf{y}^{*^T}(n)\mathbf{w}] + \lambda(\mathbf{w}^{*^T}\mathbf{e}(e^{j\omega_c}) - 1) \quad (29)$$

where $[\cdot]^{*^T}$ stands for complex conjugated and transposed.

Minimum variance spectral estimator

- The gradient of $\overline{\xi}_{\omega_c}$ with respect to $\mathbf{w}^*$ should be equal to

$$\nabla_{\mathbf{w}^*} \overline{\xi}_{\omega_c} = \mathbf{R}_Y \mathbf{w} + \lambda \mathbf{e}(e^{j\omega_c}) \tag{30}$$

  where $\mathbf{R}_Y = \mathrm{E}[\mathbf{y}(n)\mathbf{y}^{*^T}(n)]$.

- For a positive definite matrix $\mathbf{R}_Y$, the value of $\mathbf{w}$ that satisfies $\nabla_{\mathbf{w}^*} \overline{\xi}_{\omega_c} = \mathbf{0}$ is unique and is a minimum of $\overline{\xi}_{\omega_c}$.

Minimum variance spectral estimator

- If we denote this optimal solution by $\widetilde{\mathbf{w}}$, we have

$$\mathbf{R}_Y\widetilde{\mathbf{w}} + \lambda\mathbf{e}(e^{j\omega_c}) = \mathbf{0} \tag{31}$$

- Premultiplying the above equation by $\mathbf{e}^{*^T}(e^{j\omega_c})\mathbf{R}_Y^{-1}$, it follows that

$$\mathbf{e}^{*^T}(e^{j\omega_c})\widetilde{\mathbf{w}} + \lambda\mathbf{e}^{*^T}(e^{j\omega_c})\mathbf{R}_Y^{-1}\mathbf{e}(e^{j\omega_c}) = \mathbf{0} \tag{32}$$

and then

$$\lambda = -\frac{1}{\mathbf{e}^{*^T}(e^{j\omega_c})\mathbf{R}_Y^{-1}\mathbf{e}(e^{j\omega_c})} \tag{33}$$

considering that the constraint of equation (25) holds for $\widetilde{\mathbf{w}}$.

Minimum variance spectral estimator

- According to equation (31) the MV solution is

$$\widetilde{\mathbf{w}} = \frac{1}{\mathbf{e}^{*^{\mathrm{T}}}(e^{j\omega_c})\mathbf{R}_Y^{-1}\mathbf{e}(e^{j\omega_c})}\mathbf{R}_Y^{-1}\mathbf{e}(e^{j\omega_c}) \tag{34}$$

- For this solution, the minimum value of the objective function becomes

$$\begin{aligned} \xi_{\omega_c\,\min} &= \min\left\{E\{|y_{MV,\omega_c}(n)|^2\}\right\} \\ &= E[\widetilde{\mathbf{w}}^{*^{\mathrm{T}}}\mathbf{y}(n)\mathbf{y}^{*^{\mathrm{T}}}(n)\widetilde{\mathbf{w}}] \\ &= \frac{1}{\mathbf{e}^{*^{\mathrm{T}}}(e^{j\omega_c})\mathbf{R}_Y^{-1}\mathbf{e}(e^{j\omega_c})} \end{aligned} \tag{35}$$

Minimum variance spectral estimator

- This solution is valid for any $\omega = \omega_c$ so that according to (23), one gets

$$\hat{\Gamma}_{Y,MV}(e^{j\omega}) = \frac{2\pi}{\Delta\omega}\,\xi_{\omega\,min} \approx \frac{2\pi}{\Delta\omega\left(\mathbf{e}^{*^T}(e^{j\omega})\mathbf{R}_Y^{-1}\mathbf{e}(e^{j\omega})\right)} \tag{36}$$

Minimum variance spectral estimator

- Given that L is the window length, one may approximate the window bandwidth by

$$\Delta\omega \approx \frac{2\pi}{\mathrm{L}} \tag{37}$$

and the minimum-variance spectrum estimator is

$$\hat{\Gamma}_{Y,MV}(e^{j\omega}) \approx \frac{\mathrm{L}}{\mathbf{e}^{*^T}(e^{j\omega})\mathbf{R}_Y^{-1}\mathbf{e}(e^{j\omega})} \tag{38}$$

Example 7.2

- Repeat Example 7.1 using the minimum variance method and comment on the results.

Example 7.2 - Solution

- From Figure 2 we can observe that the minimum variance method is able to discriminate the presence of three sinusoids, however, the sinusoids at 45- and 55-Hz have rather biased estimations.
- It can also be noticed that the estimated spectrum presents a rather smooth behavior, as characteristic associated to the minimum variance method.
- Unfortunately, its resolution to discriminate the peaks is not as accurate as the parametric methods for the type of application where we wish to estimate the sinusoids frequencies.

Example 7.2 - Solution

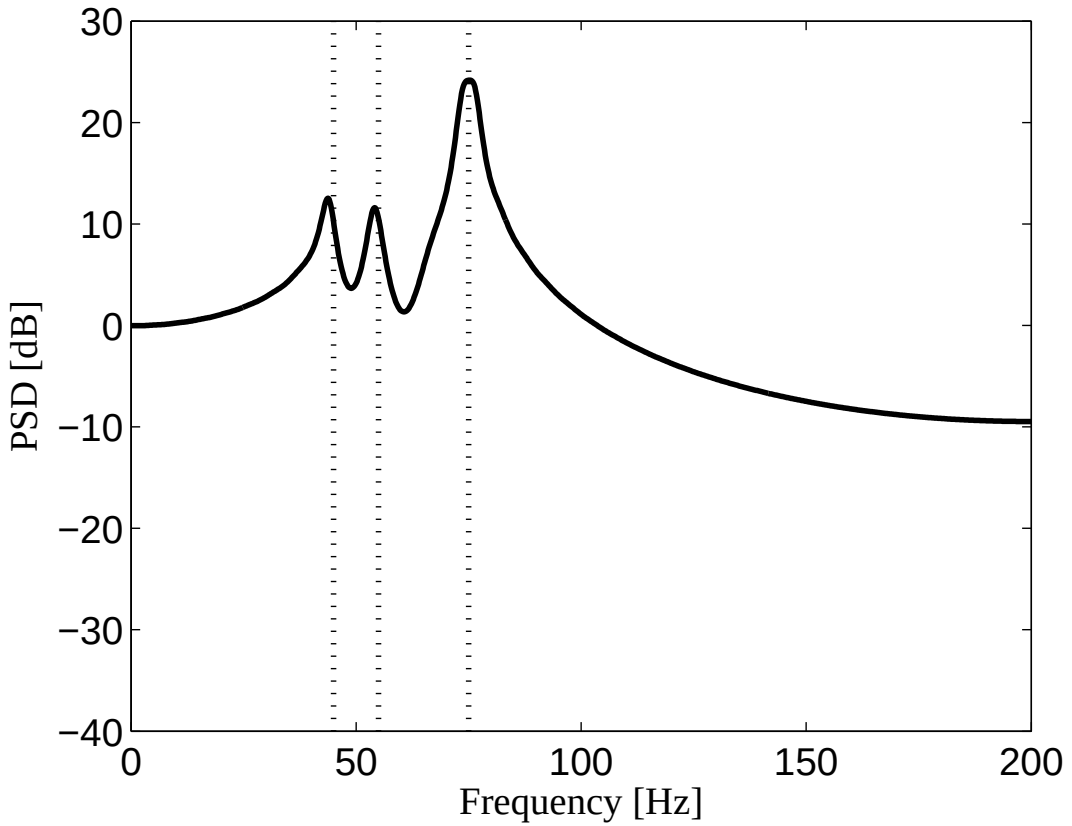


Figure 2: PSD estimate for MV method with true normalized frequencies indicated by vertical dotted lines.

Modeling theory

- In where applications there are models closely related to the process that we are willing to estimate the behavior the periodogram is not the method of choice.
- We now describe some classical modeling tools to be employed in applications allowing parametric estimation.
- The Wold decomposition theorem states that any WSS process can be expressed as the sum of a random and a deterministic processes.
- The deterministic component can be perfectly determined for all $0 \leq n$ based on the knowledge of its infinite past for $n < 0$.

Rational transfer-function models

- The classic concept of system modeling consists of writing a WSS process as the output signal of a linear, time-invariant, and causal system to a white-noise input.
- The modeling problem is commonly solved in a two-step procedure: The first step consists of choosing a particular input-output model that fits the available data.
- Then, we determine the parameter values of the particular model previously chosen.

Rational transfer-function models

- FIR filters characterized by the input-output relationship

$$y(n) = b_0 x(n) + b_1 x(n-1) + \cdots + b_M x(n-M) \tag{39}$$

  are commonly referred to as moving average (MA) systems.

- The output can be seen as a weighted average of the input samples for $n, (n-1), \ldots, (n-M)$ that shifts with n.

- Autoregressive (AR) models are characterized by a purely recursive input-output relationship described by

$$y(n) = x(n) - a_1 y(n-1) - a_2 y(n-2) - \cdots - a_N y(n-N) \tag{40}$$

Rational transfer-function models

- Combining the two models above, we get the so-called ARMA model described by

$$\begin{aligned} y(n) &= b_0 x(n) + b_1 x(n-1) + \cdots + b_M x(n-M) \\ &\quad -a_1 y(n-1) - a_2 y(n-2) - \cdots - a_N y(n-N) \end{aligned} \tag{41}$$

  which we associate to the general IIR filter.

- In the z-transform domain, we notice that the MA, AR, and ARMA are associated to

$$\begin{aligned} H_{MA}(z) &= b_0 + b_1 z^{-1} + \cdots + b_M z^{-M} \\ &= \frac{b_0 z^M + b_1 z^{M-1} + \cdots + b_M}{z^M} \end{aligned} \tag{42}$$

Rational transfer-function models

- $$\begin{aligned} H_{AR}(z) &= \frac{1}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_N z^{-N}} \\ &= \frac{z^N}{z^N + a_1 z^{N-1} + a_2 z^{N-2} + \cdots + a_N} \end{aligned} \quad (43)$$

$$\begin{aligned} H_{ARMA}(z) &= \frac{b_0 + b_1 z^{-1} + \cdots + b_M z^{-M}}{1 + a_1 z^{-1} + a_2 z^{-2} + \cdots + a_N z^{-N}} \\ &= z^{N-M} \frac{b_0 z^M + b_1 z^{M-1} + \cdots + b_M}{z^N + a_1 z^{N-1} + a_2 z^{N-2} + \cdots + a_N} \end{aligned} \quad (44)$$

respectively.

Rational transfer-function models

- The MA, AR, and ARMA nomenclature originated in the field of system modeling and control, where the input signal is often assumed to be a white noise.
- **Theorem (System decomposition theorem):** Any stable ARMA system can be decomposed as a minimum-phase stable ARMA system cascaded by an all-pass system with constant-gain.

  In addition, any minimum-phase and stable ARMA model can be approximated by an infinite-order AR system.

Rational transfer-function models

- **Proof:** Consider the zero-pole constellation for the stable transfer function $H(z)$.
- All poles are inside the unit circle and the zeros can be anywhere in the complex plane.
- For each zero z_i outside the unit circle, multiply $H(z)$ by a factor $\frac{1-\frac{1}{z_i}z^{-1}}{1-\frac{1}{z_i}z^{-1}}$, which corresponds to a zero-pole pair inside the unit circle.
- Combining the z_i zero with its inverse $\frac{1}{z_i}$ forms a constant-gain all-pass system.

Rational transfer-function models

- The $\left(1 - \frac{1}{z_i} z^{-1}\right)$ term corresponds to a zero within the unit circle.
- The transfer function can be written as the cascade of an all-pass system (combining the constant-gain contributions of all original zeros outside the unit circle and the corresponding reciprocate poles) and a minimum-phase stable ARMA system (which includes all original poles and minimum-phase zeros and the newly introduced minimum-phase zeros).

Rational transfer-function models

- Let us now consider only the minimum-phase stable ARMA transfer function as

$$H(z) = b_0 \frac{\prod_{i=1}^{M}(1 - z_i z^{-1})}{\prod_{j=1}^{N}(1 - p_j z^{-1})} \tag{45}$$

with all $|z_i| < 1$ and $|p_j| < 1$.

Rational transfer-function models

- Representing the ARMA poles with an AR system is trivial. For each zero such that $|z_i| < 1$, using the summing formula for an infinite-length geometric series

$$1 - z_i z^{-1} = \frac{1}{1 + \dfrac{z_i z^{-1}}{1 - z_i z^{-1}}} = \frac{1}{1 + \displaystyle\sum_{k=1}^{\infty} z_i^k z^{-k}} \tag{46}$$

- Hence, each minimum-phase zero also can be modeled as an AR system and so can the complete ARMA model.

Rational transfer-function models

- This result indicates that the above minimum-phase and stable ARMA-AR equivalence, except for an all-pass factor, applies to a large set of linear systems.
- A similar result follows for minimum-phase and stable ARMA systems and MA models, the demonstration of which is left as an end-of-chapter exercise for the interested reader.
- The ARMA-AR and ARMA-MA equivalences indicate that we may not need to be sure about which model to employ in a particular application. As indicated above, the price to be payed for employing an MA or an AR model instead of an ARMA model is that we may be forced to work with a high-order model.

Example 7.3

- Approximate the ARMA system

$$H(z) = \frac{1 + 0.3z^{-1}}{1 - 0.9z^{-1}} \tag{47}$$

by an Nth-order AR model.

Example 7.3 - Solution

- Since the ARMA filter is minimum-phase, using equation (46) we can write that

$$
\begin{aligned}
H(z) &= \frac{1}{(1-0.9z^{-1})\left(1+\sum_{k=1}^{\infty}(-0.3)^k z^{-k}\right)} \\
&= \frac{1}{1+\left(\sum_{k=1}^{\infty}(-0.3)^k z^{-k}\right)-0.9z^{-1}-0.9\left(\sum_{k=1}^{\infty}(-0.3)^k z^{-(k+1)}\right)} \\
&= \frac{1}{1+\left(\sum_{k=1}^{\infty}(-0.3)^k z^{-k}\right)-0.9\left(\sum_{k=0}^{\infty}(-0.3)^k z^{-(k+1)}\right)}
\end{aligned}
$$

Example 7.3 - Solution

- Hence,

$$
\begin{aligned}
H(z) &= \frac{1}{1+\left(\sum_{k=1}^{\infty}(-0.3)^k z^{-k}\right)-0.9\left(\sum_{k'=1}^{\infty}\frac{(-0.3)^{k'}}{(-0.3)}z^{-k'}\right)} \\
&= \frac{1}{1+\left(\sum_{k=1}^{\infty}(-0.3)^k z^{-k}\right)+3\left(\sum_{k'=1}^{\infty}(-0.3)^{k'} z^{-k'}\right)} \\
&= \frac{1}{1+4\sum_{k=1}^{\infty}(-0.3)^k z^{-k}} \qquad (48)
\end{aligned}
$$

Example 7.3 - Solution

- Which corresponds to an infinite-order AR system with denominator coefficient

$$a_i = 4(-0.3)^i \Rightarrow \begin{cases} a_1 = -1.2 \\ a_2 = 0.36 \\ a_3 = -0.108 \\ \vdots \end{cases} \tag{49}$$

Example 7.3 - Solution

- Truncating the last summation in equation (48), we get a finite-order AR approximation

$$H(z) \approx \frac{1}{1 + 4\sum_{k=1}^{N} (-0.3)^k z^{-k}} \tag{50}$$

Example 7.3 - Solution

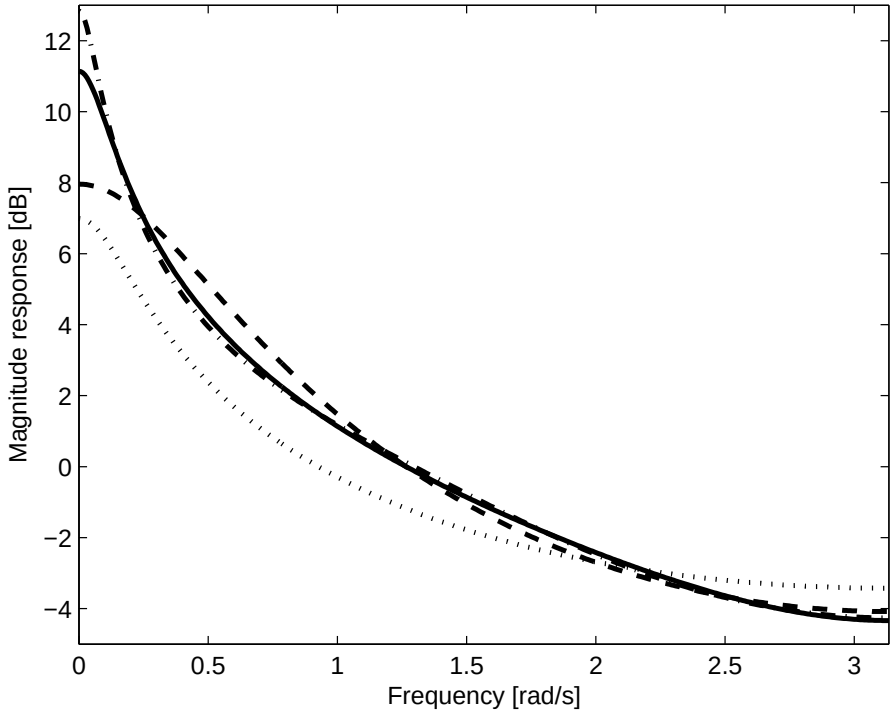


Figure 3: Magnitude responses of original ARMA system (solid line) and AR approximations of orders: $N = 1$ (dotted line), $N = 2$ (dashed line), and $N = 3$ (dash-dotted line).

Yule-Walker equations

- For the MA model, multiplying equation (39) by $y(n-v)$ and taking the expected value, we get

$$\begin{aligned} E\{y(n)y(n-v)\} &= E\left\{\sum_{j=0}^{M} b_j x(n-j)y(n-v)\right\} \\ &= \sum_{j=0}^{M} b_j E\{x(n-j)y(n-v)\} \end{aligned} \tag{51}$$

Yule-Walker equations

- The general term $E\{x(n-j)y(n-\nu)\}$ becomes null for $\nu > j$, since the output at time $(n-\nu)$ is independent of the future input $x(n-j)$, and then

$$E\{x(n-j)y(n-\nu)\} = E\{x(n-j)\}E\{y(n-\nu)\} = 0 \tag{52}$$

  assuming a zero-mean white noise $x(n)$.

Yule-Walker equations

- For $\nu \leq j$, however, we obtain

$$\begin{aligned} E\{x(n-j)y(n-\nu)\} &= E\left\{x(n-j)\sum_{l=0}^{M} b_l x(n-l-\nu)\right\} \\ &= \sum_{l=0}^{M} b_l E\{x(n-j)x(n-l-\nu)\} \\ &= b_{j-\nu}\sigma_X^2 \end{aligned} \tag{53}$$

for a white noise $x(n)$ with variance σ_X^2.

Yule-Walker equations

- Substituting equation (53) into (51), for the MA model, it is possible to show that

$$R_Y(\nu) = \begin{cases} \left(\sum_{j=\nu}^{M} b_j b_{j-\nu} \right) \sigma_X^2, & \text{for } \nu = 0, 1, \ldots, M \\ 0, & \text{for } \nu > M \end{cases} \tag{54}$$

Yule-Walker equations

- For the AR model, multiplying equation (40) by $y(n-\nu)$ and taking the expectation value, we get

$$\begin{aligned} E\{y(n)y(n-\nu)\} &= E\{x(n)y(n-\nu)\} - E\left\{\sum_{i=1}^{N} a_i y(n-i)y(n-\nu)\right\} \\ &= E\{x(n)y(n-\nu)\} - \sum_{i=1}^{N} a_i E\{y(n-i)y(n-\nu)\} \end{aligned} \quad (55)$$

Yule-Walker equations

- And then

$$R_Y(\nu) = \begin{cases} \sigma_X^2 - \sum_{i=1}^{N} a_i R_Y(\nu - i), & \text{for } \nu = 0 \\ -\sum_{i=1}^{N} a_i R_Y(\nu - i), & \text{for } \nu > 0 \end{cases} \tag{56}$$

since, for the AR model, we have that

$$E\{x(n)y(n-\nu)\} = \begin{cases} \sigma_X^2, & \text{for } \nu = 0 \\ 0, & \text{for } \nu > 0 \end{cases} \tag{57}$$

Yule-Walker equations

- For the general ARMA model, a similar relationship can be determined:

$$R_Y(\nu) = \begin{cases} \left(\sum_{j=\nu}^{M} b_j h(j-\nu) \right) \sigma_X^2 - \sum_{i=1}^{N} a_i R_Y(\nu - i), & \text{for } \nu = 0, 1, \ldots, M \\ -\sum_{i=1}^{N} a_i R_Y(\nu - i), & \text{for } \nu > M \end{cases} \tag{58}$$

  where $h(n)$ is the corresponding ARMA model impulse response.

- Equations (54), (56), and (58), relating the output autocorrelation function to the model coefficients, are the Yule-Walker equations for the MA, AR, and ARMA models.

Example 7.4

- Determine $R_Y(\nu)$ for the AR system

$$H(z) = \frac{1}{1 - 0.9z^{-1}} \tag{59}$$

assuming a unit-variance white-noise input.

Example 7.4 - Solution

- For the given AR system, we use $N = 1$ in equation (56), resulting in the following relationships for the correlation function

$$R_Y(\nu) = \begin{cases} \sigma_X^2 - a_1 R_Y(\nu - 1), & \text{for } \nu = 0 \\ -a_1 R_Y(\nu - 1), & \text{for } \nu > 0 \end{cases} \tag{60}$$

such that, for $\nu = 0, 1$,

$$\begin{cases} R_Y(0) = \sigma_X^2 - a_1 R_Y(-1) \\ R_Y(1) = -a_1 R_Y(0) \end{cases} \tag{61}$$

Example 7.4 - Solution

- Since $R_Y(-1) = R_Y(1)$, we have that

$$R_Y(0) = \frac{\sigma_X^2}{1-(-a_1)^2} \tag{62}$$

and then

$$R_Y(\nu) = \frac{\sigma_X^2(-a_1)^{|\nu|}}{1-(-a_1)^2} \tag{63}$$

which, for $\sigma_X^2 = 1$ and $a_1 = -0.9$, is depicted in Figure 4.

Example 7.4 - Solution

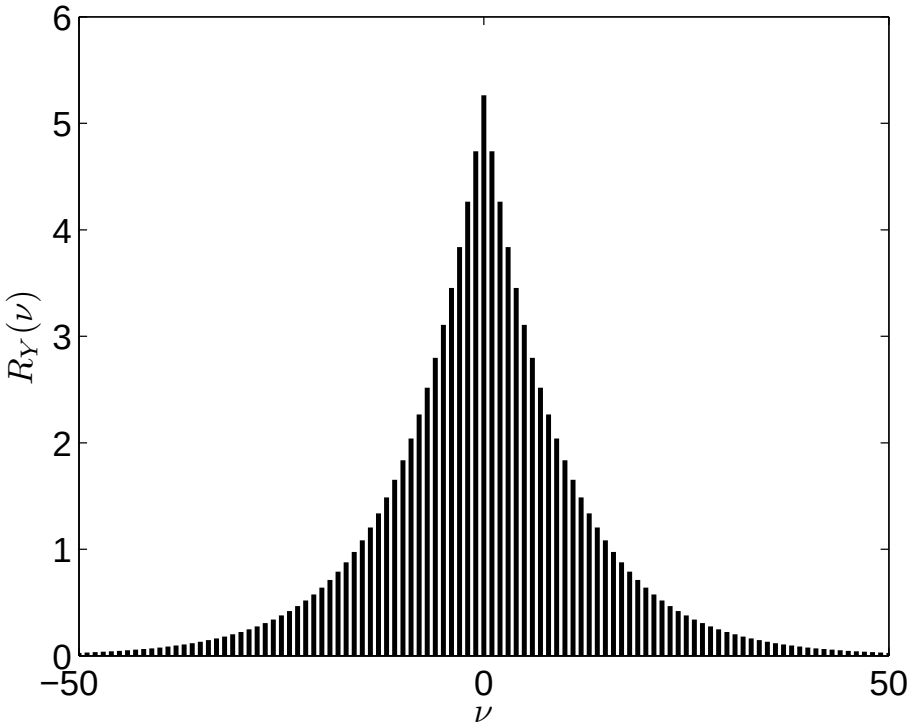


Figure 4: Autocorrelation function of AR system $H(z) = \frac{1}{1-0.9z^{-1}}$.

Parametric spectral estimation

- The periodogram-based spectral estimators perform a primary estimation of the autocorrelation function according to the Wiener-Khinchin theorem.
- In such case, the autocorrelation estimation is restricted to a lag interval delimited by the amount of available N or used M data. Outside that interval, the estimation becomes zero, leading to a biased PSD estimation.

Parametric spectral estimation

- In parametric spectral estimation, we model the random process at hand, and the resulting PSD estimate is obtained as the power spectrum of that model.
- The MA model also forces the estimated autocorrelation to become null for large lag values, see the Yule-Walker equations (54).
- Therefore, PSD estimation based on MA models presents similar properties to the periodogram-based methods.

Parametric spectral estimation

- By employing an AR or ARMA model, we do not impose any null constraint on the resulting autocorrelation function.
- In fact, for these types of models, the autocorrelation function is automatically adjusted to better fit the statistical behavior of the available data.
- The result is a PSD estimation with better statistical properties.
- The nonparametric PSD methods require large amount of data whereas the parametric methods are more suitable to applications where the data length is short.

Parametric spectral estimation

- The AR model presents interesting characteristics for practical system modeling:
  - It does not force the autocorrelation function to become zero for large lag values, as opposed to the MA model.
  - Apart from an all-pass component, it can be used to approximate any stable ARMA system.
  - Its Yule-Walker equations consist of a linear relationship between the autocorrelation values and the model coefficients, as opposed to the general ARMA model.
  - The resulting linear system presents a special structure yielding simplified numerical algorithm.

Linear prediction

- We now describe the classical approach, commonly referred to as linear prediction (LP), to estimate the parameters an AR model for a particular data set.
- The algebraic development is associated to the AR modeling problem with an interesting frequency-domain interpretation.
- Consider that we know a set of discrete-time data $\{y(0), y(1), \ldots, \}$ from a particular process. This data may have come from measurements of a stock price, some bacteria population, or a speech signal, for instance.

Linear prediction

- The idea behind the LP problem is to estimate the value of $y(n)$ as a linear combination of N past samples of the process, that is

$$\hat{y}(n) = \hat{a}_1 y(n-1) + \hat{a}_2 y(n-2) + \cdots + \hat{a}_N y(n-N) = \sum_{i=1}^{N} \hat{a}_i y(n-i) \quad (64)$$

where $\hat{a}_i$, for $i = 1, 2, \ldots, N$, are the LP coefficients and N is the LP model order.

Linear prediction

- By comparing the LP estimation to the true signal value, we can form the estimation error

$$e(n) = y(n) - \hat{y}(n) \tag{65}$$

and the associated mean-squared error (MSE) is given by

$$\xi = E\{e^2(n)\} \tag{66}$$

which, from equations (64) and (65), constitutes a quadratic function of the LP coefficients $\hat{a}_i$.

Linear prediction

- One may then determine that

$$
\begin{aligned}
\frac{\partial \xi}{\partial \hat{a}_i} &= \frac{\partial E\{e^2(n)\}}{\partial \hat{a}_i} \\
&= E\left\{\frac{\partial e^2(n)}{\partial \hat{a}_i}\right\} \\
&= E\left\{2e(n)\frac{\partial e(n)}{\partial \hat{a}_i}\right\} \\
&= E\left\{2e(n)\frac{\partial\left(y(n)-\sum_{j=1}^{N}\hat{a}_j y(n-j)\right)}{\partial \hat{a}_i}\right\} \\
&= -2E\{e(n)y(n-i)\} \qquad (67)
\end{aligned}
$$

Linear prediction

- By replacing $e(n)$ by its expression in equation (65), it follows that

$$\begin{aligned}\frac{\partial \xi}{\partial \hat{a}_i} &= -2E\left\{\left(y(n)-\sum_{j=1}^{N}\hat{a}_j y(n-j)\right)y(n-i)\right\} \\ &= -2\left\{E\{y(n)y(n-i)\}-\sum_{j=1}^{N}\hat{a}_j E\{y(n-j)y(n-i)\}\right\} \\ &= -2R_Y(i)+2\sum_{j=1}^{N}\hat{a}_j R_Y(i-j) \qquad (68)\end{aligned}$$

assuming that the process $\{Y\}$ is WSS.

Linear prediction

- By forming the gradient vector $\nabla_{\hat{\mathbf{a}}}\xi$, with respect to the LP coefficient vector $\hat{\mathbf{a}} = [\hat{a}_1 \;\; \hat{a}_2 \;\; \ldots \;\; \hat{a}_N]^T$ and equating it to zero, we determine the MSE stationary point $\hat{\mathbf{a}}^*$ as the solution of the following linear system:

$$\begin{bmatrix} R_Y(0) & R_Y(-1) & \cdots & R_Y(1-N) \\ R_Y(1) & R_Y(0) & \cdots & R_Y(2-N) \\ \vdots & \vdots & \ddots & \vdots \\ R_Y(N-1) & R_Y(N-2) & \cdots & R_Y(0) \end{bmatrix} \begin{bmatrix} \hat{a}_1^* \\ \hat{a}_2^* \\ \vdots \\ \hat{a}_N^* \end{bmatrix} = \begin{bmatrix} R_Y(1) \\ R_Y(2) \\ \vdots \\ R_Y(N) \end{bmatrix} \tag{69}$$

Linear prediction

- The previous relation is the so-called Wiener-Hopf equation, rewritten as

$$\mathbf{R}_Y \hat{\mathbf{a}}^* = \mathbf{p}_Y \tag{70}$$

where $\mathbf{R}_Y$ is the autocorrelation matrix for the process $\{Y\}$ and $\mathbf{p}_Y$ is the right-hand side vector in equation (69).

Linear prediction

- Using equation (68), we determine the MSE second-order differentiations as

$$\begin{aligned}
\frac{\partial^2 \xi}{\partial \hat{a}_i \partial \hat{a}_k} &= \frac{\partial \left(\frac{\partial \xi}{\partial \hat{a}_i} \right)}{\partial \hat{a}_k} \\
&= \frac{\partial \left(-2R_Y(i) + 2 \sum_{j=1}^{N} \hat{a}_j R_Y(i-j) \right)}{\partial \hat{a}_k} \\
&= 2R_Y(i-k)
\end{aligned} \tag{71}$$

Linear prediction

- The MSE Hessian matrix is given by

$$\mathbf{H} = \left[\frac{\partial^2 \xi}{\partial \hat{a}_i \partial \hat{a}_k}\right]_{i,k} = [2R_Y(i-k)]_{i,k} = 2\mathbf{R}_Y \tag{72}$$

  which is in general positive definite. The stationary point $\hat{\mathbf{a}}^\star$ in equation (70) is the MSE global minimum.

Linear prediction

- Analyzing equations (64) and (65) in the z-transform domain, for a given signal $y(n)$ we get

$$E(z) = \left(1 - \hat{a}_1 z^{-1} - \hat{a}_2 z^{-2} - \cdots - \hat{a}_N z^{-N}\right) Y(z) \tag{73}$$

or, equivalently,

$$H(z) = \frac{Y(z)}{E(z)} = \frac{1}{1 - \hat{a}_1 z^{-1} - \hat{a}_2 z^{-2} - \cdots - \hat{a}_N z^{-N}} \tag{74}$$

Linear prediction

- If the model order N is high enough, one may achieve the best possible prediction and the error process $\{E\}$ becomes a white noise.
- The LP problem corresponds to an AR modeling of $y(n)$ with the signal $e(n)$ as input.
- The Wiener-Hopf equation (69) can be obtained directly from the AR Yule-Walker equations by considering $e(n) \equiv x(n)$ and letting $-a_i = \hat{a}_i$ in equation (56), for $\nu = 1, 2, \ldots, N$.

Linear prediction

- AR modeling and LP can be seen as a pair of inverse problems: Given an AR system, using the Yule-Walker equations one is able to infer the autocorrelation function of the output process.
- In the LP context, given the autocorrelation samples, we can estimate the AR system that better fits these samples in the minimum MSE sense.

Example 7.5

- Determine the second-order linear predictor for a WSS process characterized by

$$R_Y(\nu) = 4(0.5)^{|\nu|} \tag{75}$$

Example 7.5 - Solution

- From equation (75), we have that

$$\left.\begin{aligned} R_Y(0) &= 4 \\ R_Y(1) &= 2 \\ R_Y(2) &= 1 \end{aligned}\right\} \tag{76}$$

which results in the Wiener-Hopf equation

$$\begin{bmatrix} 4 & 2 \\ 2 & 4 \end{bmatrix} \begin{bmatrix} \hat{a}_1^* \\ \hat{a}_2^* \end{bmatrix} = \begin{bmatrix} 2 \\ 1 \end{bmatrix} \tag{77}$$

- The solution of the above equation is given by $\hat{a}_1^* = 0.5$ and $\hat{a}_2^* = 0$.

Example 7.5 - Solution

- The obtained solution corresponds in fact to the first-order AR system

$$H(z) = \frac{1}{1 - 0.5z^{-1}} \tag{78}$$

whose autocorrelation function is of the form

$$R_Y(\nu) = \frac{\sigma_X^2(\hat{a}_1^*)^{|\nu|}}{1 - (\hat{a}_1^*)^2} = \frac{\sigma_X^2(0.5)^{|\nu|}}{1 - 0.5^2} = \frac{4\sigma_X^2(0.5)^{|\nu|}}{3} \tag{79}$$

with, in this case, σ_X^2 determined by

$$\sigma_X^2 = R_Y(0)(1 - (\hat{a}_1^*)^2) = 4(1 - 0.5^2) = 3 \tag{80}$$

turning equation (79) into (75), as expected.

Linear prediction

- An interesting frequency-domain interpretation for the LP problem arises from

$$E\{e^2(n)\} = R_E(0) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \Gamma_E(e^{j\omega})\, d\omega = \frac{1}{2\pi}\int_{-\pi}^{\pi} \frac{\Gamma_Y(e^{j\omega})}{|H(e^{j\omega})|^2}\, d\omega \qquad (81)$$

- For a particular process, $\Gamma_Y(e^{j\omega})$ is a fixed nonnegative function and, for an AR system, $|H(e^{j\omega})|^2$ is a strictly positive function.

- Hence, the last integral in equation (81) is equivalent to the area below the curve defined by the ratio of these two functions.

Linear prediction

- The best we can do to minimize this area, and also the MSE, is to choose the LP coefficients $\hat{a}_i$ in such a way that the denominator $|H(e^{j\omega})|^2$ becomes large when the numerator is large, small when the numerator is small, and so on.
- The optimal MSE solution is such that $|H(e^{j\omega})|^2$ follows the shape of $\Gamma_Y(e^{j\omega})$, as close as possible, considering the limited number N of variables.

Linear prediction

- In practice, the LP solution is such that the AR power spectrum becomes a smoothed version of the PSD function of the process $\{Y\}$.
- The level of approximation required by the application at hand determines the optimal value of N to be used.
- The implementation of the Wiener-Hopf equations differ in the algorithm for estimating the autocorrelation function. Each variation of estimation method results in a different AR model, as described in the following subsections.

Covariance method

- In the covariance method, the data set is windowed and the estimation error is minimized only within the window interval.
- Only part of the available data is used in estimating the autocorrelation function. This yields a different autocorrelation estimation for each window location k, disregarding the common WSS assumption, such that

$$\hat{R}_{Y,m}(\mu,\nu) = \frac{1}{K}\sum_{n=k}^{K+k-1} y(n-\mu)y(n-\nu) \tag{82}$$

where $K < N$ is the window length.

Covariance method

- This estimation has the advantage of being unbiased, since

$$\begin{aligned} E\left\{\hat{R}_{Y,m}(\mu,\nu)\right\} &= E\left\{\frac{1}{K}\sum_{n=k}^{K+k-1} y(n-\mu)y(n-\nu)\right\} \\ &= \frac{1}{K}\sum_{n=k}^{K+k-1} E\{y(n-\mu)y(n-\nu)\} \\ &= R_Y(\mu,\nu) \end{aligned} \tag{83}$$

with low variance for large values of $(\mu-\nu)$, since all autocorrelation values are the average of the same number K of non-zero terms.

Covariance method

- This modified estimator forms a Wiener-Hopf equation

$$\begin{bmatrix} R_Y(1,1) & R_Y(1,2) & \cdots & R_Y(1,N) \\ R_Y(2,1) & R_Y(2,2) & \cdots & R_Y(2,N) \\ \vdots & \vdots & \ddots & \vdots \\ R_Y(N,1) & R_Y(N,2) & \cdots & R_Y(N,N) \end{bmatrix} \begin{bmatrix} \hat{a}_1^* \\ \hat{a}_2^* \\ \vdots \\ \hat{a}_N^* \end{bmatrix} = \begin{bmatrix} R_Y(1,0) \\ R_Y(2,0) \\ \vdots \\ R_Y(N,0) \end{bmatrix} \tag{84}$$

- The autocorrelation matrix is symmetric and non-Toeplitz, which can be inverted by using the Cholesky or LU decompositions.
- The AR model obtained from the covariance method is not guaranteed to be stable.
- Practical problems often lead to stable poles and unbiased PSD estimation.

Autocorrelation method

- In the autocorrelation method for AR modeling, the LP problem (69) is solved by using the biased autocorrelation estimation $\hat{R}_{Y,b}(\nu)$.
- The AR model can be shown to be guaranteed stable, but the associated PSD estimation is biased and presents low resolution.
- The unbiased estimation $\hat{R}_{Y,u}(\nu)$ using all available data leads to an ill-conditioned autocorrelation matrix, inducing large variance in the final PSD estimation.
- The even symmetry in the $\hat{R}_{Y,b}(\nu)$ estimation, leads to autocorrelation matrix: symmetric, Toeplitz, and positive definite, so that the Levinson-Durbin recursion can be used to invert it and solve the linear system of equations.

Example 7.6

- Consider the output signal $y(n)$ of the Hamming-window FIR filter, to a white-noise input with zero mean and unit variance. Find the Nth-order AR model for such signal using the autocorrelation method for different values of N.

Example 7.6 - Solution

- This problem constitutes an interesting challenge for the AR model, since it is expected to approximate a system with all zeros on the unit circle and all poles at the origin of the complex plane.
- Using the autocorrelation method with $\mathrm{N} = 10, 20, 50$, as described in this section, results in the AR power spectra shown in Figure 5 along with the original FIR response (solid line).
- In this figure, all AR spectra were normalized to 0 dB at $\omega = 0$. Results clearly indicate that larger values of N yield AR models with power response closer to the ideal one.

Example 7.6 - Solution

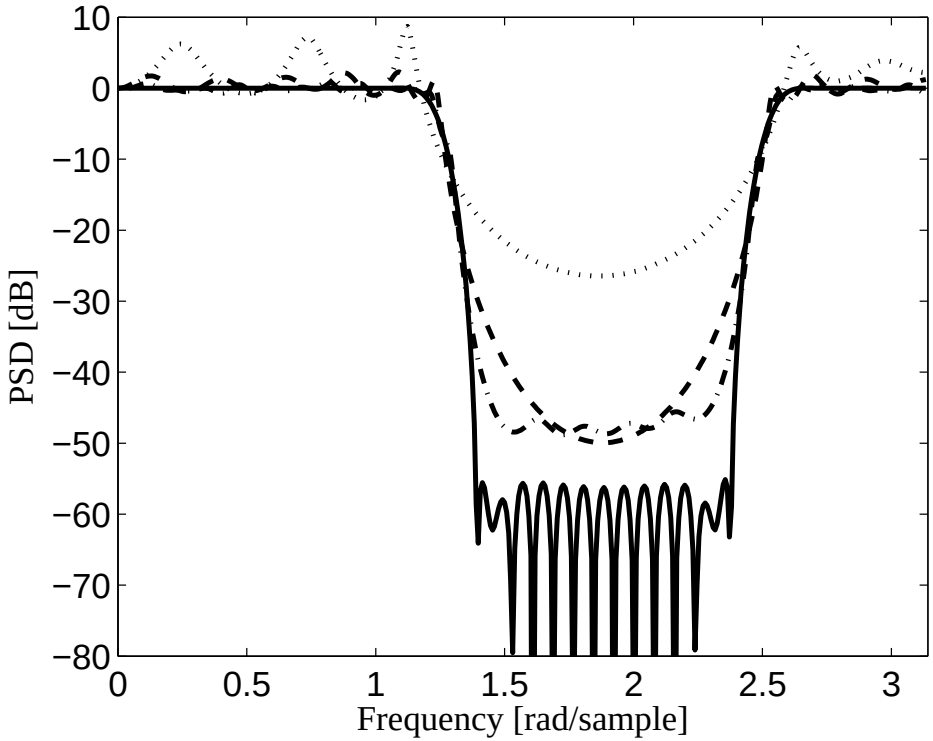


Figure 5: Power response for original MA system (solid line) and respective AR approximations using autocorrelation method: $\mathrm{N} = 10$ (dotted line), $\mathrm{N} = 20$ (dashed line), $\mathrm{N} = 50$ (dash-dotted line).

Levinson-Durbin algorithm

- We can generate the extended Wiener-Hopf equation of (56) by incorporating the AR Yule-Walker equation (69) for $\nu = 0$ into the standard form, obtaining

$$\begin{bmatrix} R_Y(0) & R_Y(-1) & R_Y(-2) & \dots & R_Y(-N) \\ R_Y(1) & R_Y(0) & R_Y(-1) & \dots & R_Y(1-N) \\ R_Y(2) & R_Y(1) & R_Y(0) & \dots & R_Y(2-N) \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ R_Y(N) & R_Y(N-1) & R_Y(N-2) & \dots & R_Y(0) \end{bmatrix} \begin{bmatrix} 1 \\ -\hat{a}^*_{1,[N]} \\ -\hat{a}^*_{2,[N]} \\ \vdots \\ -\hat{a}^*_{N,[N]} \end{bmatrix} = \begin{bmatrix} \sigma^2_{X,[N]} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \quad (85)$$

where the sub-index $[N]$ relates the associated variable to the Nth-order model.

Levinson-Durbin algorithm

- We must notice that the value of $R_Y(\nu)$ is independent of the model order, whereas the estimation error is highly dependent on N, and so is its variance σ_X^2.
- The extended Wiener-Hopf equation can also be written in a compact form as

$$\mathbf{R}_{Y,[N+1]} \begin{bmatrix} 1 \\ -\hat{\mathbf{a}}^*_{[N]} \end{bmatrix} = \begin{bmatrix} \sigma^2_{X,[N]} \\ \mathbf{0} \end{bmatrix} \tag{86}$$

Levinson-Durbin algorithm

- The Levinson-Durbin algorithm determines the LP coefficients of the Nth-order model from the $(N-1)$th-order model.
- This recursion is then initialized for the first-order model and then iterates up to desired order N.
- To obtain the order-recursive relationships consider the upgrade from the second- to the third-order AR model.

Levinson-Durbin algorithm

- We can express the third-order extended coefficient vector in equation (86) as

$$\begin{bmatrix} 1 \\ -\hat{\mathbf{a}}^*_{[3]} \end{bmatrix} = \begin{bmatrix} 1 \\ -\hat{a}^*_{1,[3]} \\ -\hat{a}^*_{2,[3]} \\ -\hat{a}^*_{3,[3]} \end{bmatrix} = \begin{bmatrix} 1 \\ -\hat{a}^*_{1,[2]} \\ -\hat{a}^*_{2,[2]} \\ 0 \end{bmatrix} - k_3 \begin{bmatrix} 0 \\ -\hat{a}^*_{2,[2]} \\ -\hat{a}^*_{1,[2]} \\ 1 \end{bmatrix} \tag{87}$$

where k_3 is an auxiliary parameter to be determined.

Levinson-Durbin algorithm

- Using this expression, the extended Wiener-Hopf equation for $N = 3$ becomes

$$\mathbf{R}_{Y,[4]} \begin{bmatrix} 1 \\ -\hat{a}^*_{1,[3]} \\ -\hat{a}^*_{2,[3]} \\ -\hat{a}^*_{3,[3]} \end{bmatrix} = \begin{bmatrix} & \mathbf{R}_{Y,[3]} & & R_Y(-3) \\ & & & R_Y(-2) \\ & & & R_Y(-1) \\ R_Y(3) & R_Y(2) & R_Y(1) & R_Y(0) \end{bmatrix} \begin{bmatrix} 1 \\ -\hat{a}^*_{1,[2]} \\ -\hat{a}^*_{2,[2]} \\ 0 \end{bmatrix}$$

$$-k_3 \begin{bmatrix} R_Y(0) & R_Y(-1) & R_Y(-2) & R_Y(-3) \\ R_Y(1) & & & \\ R_Y(2) & & \mathbf{R}_{Y,[3]} & \\ R_Y(3) & & & \end{bmatrix} \begin{bmatrix} 0 \\ -\hat{a}^*_{2,[2]} \\ -\hat{a}^*_{1,[2]} \\ 1 \end{bmatrix}$$

Levinson-Durbin algorithm

- Or equivalently

$$\mathbf{R}_{Y,[4]} \begin{bmatrix} 1 \\ -\hat{a}^*_{1,[3]} \\ -\hat{a}^*_{2,[3]} \\ -\hat{a}^*_{3,[3]} \end{bmatrix} = \begin{bmatrix} \mathbf{R}_{Y,[3]} \begin{bmatrix} 1 \\ -\hat{a}^*_{1,[2]} \\ -\hat{a}^*_{2,[2]} \end{bmatrix} \\ q_{[2]} \end{bmatrix} - k_{[3]} \begin{bmatrix} q_{[2]} \\ \mathbf{R}_{Y,[3]} \begin{bmatrix} -\hat{a}^*_{2,[2]} \\ -\hat{a}^*_{1,[2]} \\ 1 \end{bmatrix} \end{bmatrix}$$

$$= \begin{bmatrix} \sigma^2_{X,[3]} \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{88}$$

Levinson-Durbin algorithm

- With

$$\begin{aligned} q_{[2]} &= R_Y(3) - R_Y(2)\hat{a}^*_{1,[2]} - R_Y(1)\hat{a}^*_{2,[2]} \\ &= R_Y(-3) - R_Y(-2)\hat{a}^*_{1,[2]} - R_Y(-1)\hat{a}^*_{2,[2]} \end{aligned} \tag{89}$$

- Hence, we can write that

$$\begin{bmatrix} \sigma^2_{x,[2]} \\ 0 \\ 0 \\ q_{[2]} \end{bmatrix} - k_3 \begin{bmatrix} q_{[2]} \\ 0 \\ 0 \\ \sigma^2_{x,[2]} \end{bmatrix} = \begin{bmatrix} \sigma^2_{x,[3]} \\ 0 \\ 0 \\ 0 \end{bmatrix} \tag{90}$$

Levinson-Durbin algorithm

- Or, equivalently,

$$\begin{cases} \sigma^2_{X,[2]} - k_3 q_{[2]} = \sigma^2_{X,[3]} \\ q_{[2]} - k_3 \sigma^2_{X,[2]} = 0 \end{cases} \tag{91}$$

- Solving the system above, we get

$$k_3 = \frac{q_{[2]}}{\sigma^2_{X,[2]}} = \frac{R_Y(3) - R_Y(2)\hat{a}^*_{1,[2]} - R_Y(1)\hat{a}^*_{2,[2]}}{\sigma^2_{X,[2]}} \tag{92}$$

$$\sigma^2_{X,[3]} = (1 - k_3^2)\sigma^2_{X,[2]} \tag{93}$$

Levinson-Durbin algorithm

- Hence, from equation (87),

$$\begin{cases} \hat{a}^*_{1,[3]} = \hat{a}^*_{1,[2]} - k_3 \hat{a}^*_{2,[2]} \\ \hat{a}^*_{2,[3]} = \hat{a}^*_{2,[2]} - k_3 \hat{a}^*_{1,[2]} \\ \hat{a}^*_{3,[3]} = k_3 \end{cases} \tag{94}$$

Levinson-Durbin algorithm

- Equations (92), (93), and (94) illustrate that the third-order AR model can be determined from the second-order parameters $q_{[2]}$, $\sigma^2_{X,[2]}$, $\hat{a}^*_{1,[2]}$, and $\hat{a}^*_{2,[2]}$.
- Generalizing this recursion, we obtain an iterative algorithm with computational complexity proportional to N^2 and described by:

Levinson-Durbin algorithm

(i) For a given data set, determine $R_Y(0), R_Y(1), \ldots, R_Y(N)$.

(ii) Set $\sigma^2_{X,[0]} = R_Y(0)$.

(iii) For $i = 1, 2, \ldots, N$, compute:

$$k_i = \frac{R_Y(i) - \sum_{j=1}^{i-1} \hat{a}^*_{j,[i-1]} R_Y(i-j)}{\sigma^2_{X,[i-1]}} \tag{95}$$

$$\sigma^2_{X,[i]} = \left(1 - k_i^2\right) \sigma^2_{X,[i-1]} \tag{96}$$

$$\hat{a}^*_{j,[i]} = \begin{cases} \hat{a}^*_{j,[i-1]} - k_i \hat{a}^*_{i-j,[i-1]}, & \text{for } j = 1, 2, \ldots, (i-1) \\ k_i, & \text{for } j = i \end{cases} \tag{97}$$

Burg's method

- The auxiliary parameters k_i, for $i = 1, 2, \ldots, N$, are known as the reflection coefficients, and they constitute an alternative description of the Nth-order AR model, as opposed to the LP coefficients $\hat{a}_{i,[N]}$.
- The LP model given in equation (64) employs N past samples of the available signal to estimate the current sample $y(n)$.
- The estimation error between $y(n)$ and the optimal LP estimate is also known as the Nth-order forward estimation error and denoted by

$$x_{f,[N]}(n) = y(n) - \sum_{i=1}^{N} \hat{a}^*_{i,[N]} y(n-i) \tag{98}$$

Burg's method

- If we feed the LP model with a reversed-in-time version of the process $\{Y\}$, then future samples are combined to form an estimate of $y(n-N)$.
- Since $R_Y(\nu)$ is an even function, the corresponding Wiener-Hopf equations, and consequently the optimal LP model, remain the same.
- In such a case, the error between the past sample $y(n-N)$ and its estimate is referred to as the backward LP error and denoted by

$$x_{b,[N]}(n) = y(n-N) - \sum_{i=1}^{N} \hat{a}^*_{i,[N]} y(n-N+i) \tag{99}$$

Burg's method

- Applying the Levinson-Durbin order-recursion of equation (97) to the optimal LP coefficients in equation (98)

$$
\begin{aligned}
x_{f,[N]}(n) &= y(n) - \left(\sum_{i=1}^{N-1} \hat{a}^*_{i,[N]} y(n-i) \right) - \hat{a}^*_{N,[N]} y(n-N) \\
&= y(n) - \left(\sum_{i=1}^{N-1} \hat{a}^*_{i,[N-1]} y(n-i) \right) \\
&\quad + \left(\sum_{i=1}^{N-1} k_N \hat{a}^*_{N-i,[N-1]} y(n-i) \right) - k_N y(n-N) \\
&= x_{f,[N-1]}(n) - k_N \left[y(n-N) - \left(\sum_{j=1}^{N-1} \hat{a}^*_{j,[N-1]} y(n+j-N) \right) \right] \\
&= x_{f,[N-1]}(n) - k_N x_{b,[N-1]}(n-1) \qquad (100)
\end{aligned}
$$

Burg's method

- A similar development for the backward estimation error, as defined in equation (99), results in

$$
\begin{aligned}
x_{b,[N]}(n) &= y(n-N) - \left(\sum_{i=1}^{N-1} \hat{a}^*_{i,[N]} y(n-N+i) \right) - \hat{a}^*_{N,[N]} y(n) \\
&= y(n-N) - \left(\sum_{i=1}^{N-1} \hat{a}^*_{i,[N-1]} y(n-N+i) \right) \\
&\quad + \left(\sum_{i=1}^{N-1} k_N \hat{a}^*_{N-i,[N-1]} y(n-N+i) \right) - k_N y(n) \\
&= x_{b,[N-1]}(n-1) - k_N \left[y(n) - \left(\sum_{j=1}^{N-1} \hat{a}^*_{j,[N-1]} y(n-j) \right) \right] \\
&= x_{b,[N-1]}(n-1) - k_N x_{f,[N-1]}(n) \qquad (101)
\end{aligned}
$$

Burg's method

- We can then define the mean value of the forward and backward prediction-error powers for the ith-order LP model as

$$\xi_{B,[i]} = \frac{\xi_{f,[i]} + \xi_{b,[i]}}{2} \tag{102}$$

where

$$\xi_{f,[i]} = \frac{1}{L-i} \sum_{n=i}^{L-1} x_{f,[i]}^2(n) \tag{103}$$

$$\xi_{b,[i]} = \frac{1}{L-i} \sum_{n=i}^{L-1} x_{b,[i]}^2(n) \tag{104}$$

- Burg's method determines the reflection coefficients k_i that minimize $\xi_{B,[i]}$.

Burg's method

- Differentiating $\xi_{f,[i]}$ with respect to k_i, we get

$$\begin{aligned}
\frac{\partial \xi_{f,[i]}}{\partial k_i} &= \frac{2}{L-i}\sum_{n=i}^{L-1} x_{f,[i]}(n)\frac{\partial x_{f,[i]}(n)}{\partial k_i} \\
&= \frac{2}{L-i}\sum_{n=i}^{L-1} x_{f,[i]}(n)\frac{\partial\left(x_{f,[i-1]}(n)-k_i x_{b,[i-1]}(n-1)\right)}{\partial k_i} \\
&= -\frac{2}{L-i}\sum_{n=i}^{L-1} x_{f,[i]}(n)x_{b,[i-1]}(n-1) \qquad (105)
\end{aligned}$$

Burg's method

- Analogously, differentiating $\xi_{b,[i]}$ with respect to k_i, results in

$$\begin{aligned}
\frac{\partial \xi_{b,[i]}}{\partial k_i} &= \frac{2}{L-i}\sum_{n=i}^{L-1} x_{b,[i]}(n)\frac{\partial x_{b,[i]}(n)}{\partial k_i} \\
&= \frac{2}{L-i}\sum_{n=i}^{L-1} x_{b,[i]}(n)\frac{\partial\left(x_{b,[i-1]}(n-1)-k_i x_{f,[i-1]}(n)\right)}{\partial k_i} \\
&= -\frac{2}{L-i}\sum_{n=i}^{L-1} x_{b,[i]}(n)x_{f,[i-1]}(n) \qquad (106)
\end{aligned}$$

Burg's method

- Therefore, according to equations (100), (101), (105), and (106), we can write that

$$\begin{aligned}\frac{\partial \xi_{B,[i]}}{\partial k_i} &= \frac{\frac{\partial \xi_{f,[i]}}{\partial k_i} + \frac{\partial \xi_{b,[i]}}{\partial k_i}}{2} \\ &= -\frac{1}{L-i}\sum_{n=i}^{L-1}\left[\left(x_{f,[i-1]}(n) - k_i x_{b,[i-1]}(n-1)\right) x_{b,[i-1]}(n-1)\right] \\ &\quad -\frac{1}{L-i}\sum_{n=i}^{L-1}\left[\left(x_{b,[i-1]}(n-1) - k_i x_{f,[i-1]}(n)\right) x_{f,[i-1]}(n)\right] \quad (107)\end{aligned}$$

Burg's method

- Equating this result to zero, yields

$$k_i = \frac{2\sum_{n=i}^{L-1} x_{f,[i-1]}(n)x_{b,[i-1]}(n-1)}{\sum_{n=i}^{L-1}\left(x^2_{b,[i-1]}(n-1) + x^2_{f,[i-1]}(n)\right)} \tag{108}$$

  which is the reflection coefficient used in Burg's method.

Burg's method

- This estimate is employed in the Levinson-Durbin recursions of equations (95), (96), and (97), to determine the corresponding LP model and the desired PSD estimation.
- The resulting estimate parameters $\hat{a}^*_{j,[i]}$ correspond to the solution of the so-called modified correlation method.
- The spectral estimation literature associates Burg's method to a maximum entropy PSD estimation. Such association, however, should be made to all AR modeling schemes.

Example 7.7

- Find an AR model for the same process using the covariance, autocorrelation, and Burg's method with $\mathrm{N} = 40$.

Example 7.7 - Solution

- The PSD estimates from the autocorrelation (dashed line), and Burg's (dash-dotted line) methods are shown in Figure 6 for $\mathrm{N} = 40$.
- The PSD estimate from the covariance method is indistinguishable to the one from Burg's method, which, due to the unbiased estimate for the autocorrelation function, is superior to the one yielded by the autocorrelation method.

Example 7.7 - Solution

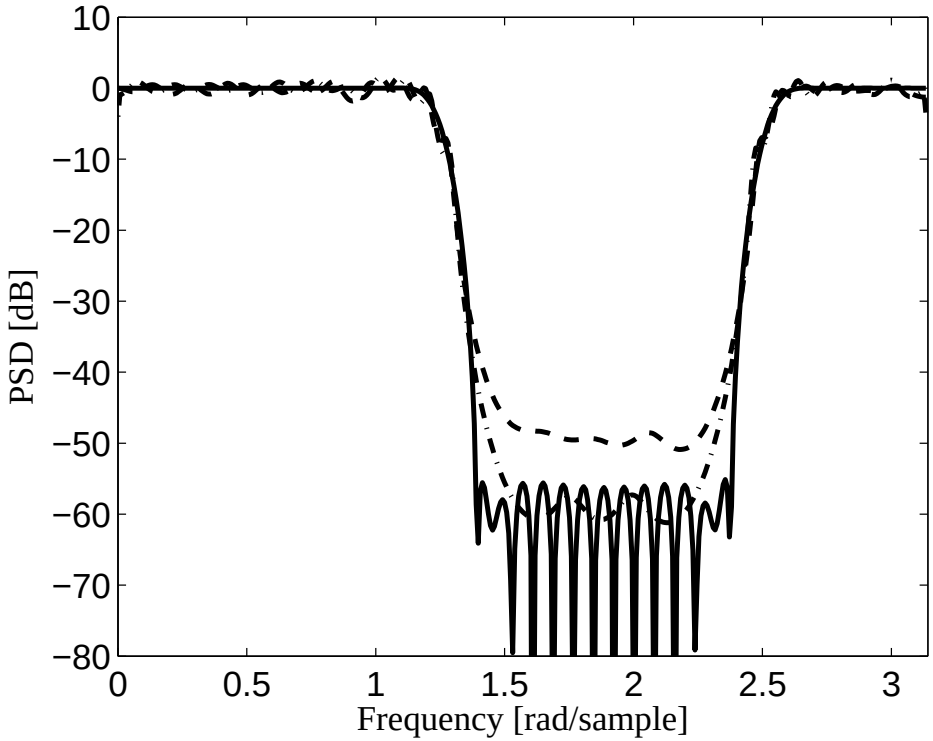


Figure 6: Power response for original MA system (solid line) and respective AR approximations of order $N = 40$: autocorrelation (dashed line) and Burg and covariance (dash-dotted line) methods.

Relationship of Levinson-Durbin and a lattice structure

- We describe how the parameters of the Levinson-Durbin recursion can be related to a digital filter structure known as lattice structure.
- Equations (100) and (101), repeated here for the reader's convenience,

$$\left.\begin{aligned} x_{f,[N]}(n) &= x_{f,[N-1]}(n) - k_N x_{b,[N-1]}(n-1) \\ x_{b,[N]}(n) &= x_{b,[N-1]}(n-1) - k_N x_{f,[N-1]}(n) \end{aligned}\right\} \tag{109}$$

relate the Nth-order estimation errors with their $(N-1)$th-order counterparts.

Relationship of Levinson-Durbin and a lattice structure

- Taking these relationships to the z domain, we obtain

$$\left.\begin{aligned} X_{f,[N]}(z) &= X_{f,[N-1]}(z) - k_N z^{-1} X_{b,[N-1]}(z) \\ X_{b,[N]}(z) &= z^{-1} X_{b,[N-1]}(z) - k_N X_{f,[N-1]}(z) \end{aligned}\right\} \tag{110}$$

  which can be readily associated to the basic digital-filter cell depicted in Figure 7.

- Concatenating several of these cells yields the two-multiplier lattice structure seen in Figure 8.

Relationship of Levinson-Durbin and a lattice structure

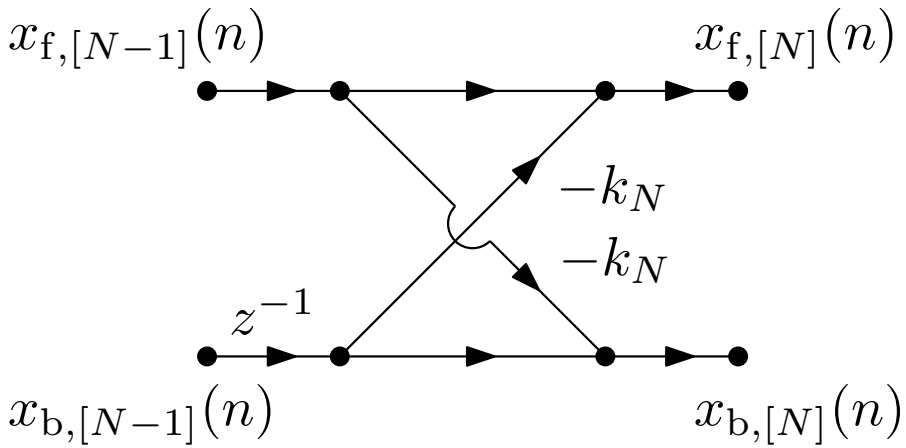


Figure 7: Basic cell equivalent to one-level Levinson-Durbin recursion.

Relationship of Levinson-Durbin and a lattice structure

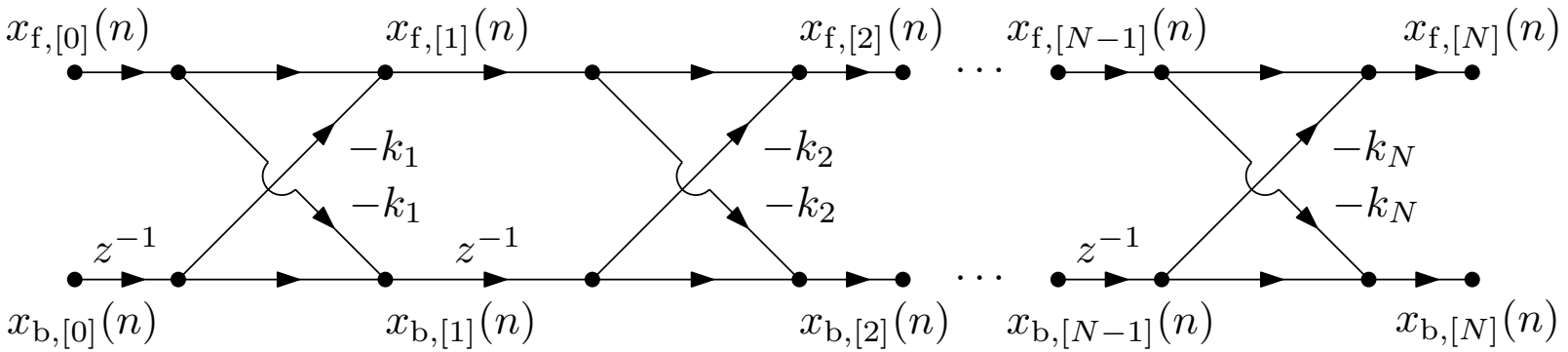


Figure 8: Two-multiplier lattice structure equivalent to N-level Levinson-Durbin recursion.

Wiener filter

- We generalize the linear prediction problem to the case where we wish to characterize the stochastic relationship between two processes represented by the realizations $y(n)$ and $x(n)$. The resulting linear model is referred as Wiener filter.
- Assume that an estimate $\hat{y}(n)$ of $y(n)$ can be determined as a linear combination of samples of $x(n)$, that is

$$\begin{aligned}\hat{y}(n) &= \hat{w}_0 x(n) + \hat{w}_1 x(n-1) + \cdots + \hat{w}_{N-1} x(n-M) \\ &= \sum_{i=0}^{M} \hat{w}_i x(n-i) \end{aligned} \tag{111}$$

where $\hat{w}_i$, for $i = 0, 1, \ldots, M$, are the so-called Wiener filter coefficients and M is the filter order.

Wiener filter

- The error signal in this case is given by

$$e(n) = y(n) - \hat{y}(n) \tag{112}$$

and the associated MSE is given by

$$\xi = E\{e^2(n)\} \tag{113}$$

Wiener filter

- Similarly to the linear prediction case, equation (113) represents a quadratic function of the Wiener filter coefficients $\hat{w}_i$.
- It is then possible to deduce that

$$\begin{aligned}
\frac{\partial \xi}{\partial \hat{w}_i} &= \frac{\partial E\{e^2(n)\}}{\partial \hat{w}_i} \\
&= E\left\{\frac{\partial e^2(n)}{\partial \hat{w}_i}\right\} \\
&= E\left\{2e(n)\frac{\partial e(n)}{\partial \hat{w}_i}\right\} \\
&= -2E\{e(n)x(n-i)\} \qquad (114)
\end{aligned}$$

Wiener filter

- Hence, by replacing $e(n)$ by its expression using equations (111) and (112), it follows that

$$\begin{aligned}
\frac{\partial \xi}{\partial \hat{w}_i} &= -2E\left\{\left(y(n) - \sum_{j=0}^{M} \hat{w}_j x(n-j)\right) x(n-i)\right\} \\
&= -2\left\{E\{y(n)x(n-i)\} - \sum_{j=0}^{M} \hat{w}_j E\{x(n-j)x(n-i)\}\right\} \\
&= -2p_{YX}(i) + 2\sum_{j=0}^{M} \hat{w}_j R_X(i-j) \qquad (115)
\end{aligned}$$

where $p_{YX}(i)$ is the cross-correlation between $y(n)$ and $x(n-i)$, assuming that the processes $\{Y\}$ and $\{X\}$ are jointly WSS.

Wiener filter

- By forming the gradient vector $\nabla_{\hat{\mathbf{w}}}\xi$ with respect to the Wiener filter coefficients $\hat{\mathbf{w}} = [\hat{w}_0 \;\; \hat{w}_1 \;\; \ldots \;\; \hat{w}_M]^T$ and equating it to zero, it is possible to compute the MSE stationary point $\hat{\mathbf{w}}^*$ by solving the system

$$\begin{bmatrix} R_X(0) & R_X(-1) & \ldots & R_X(-M) \\ R_X(1) & R_X(0) & \ldots & R_X(-M+1) \\ \vdots & \vdots & \ddots & \vdots \\ R_X(M) & R_X(M-1) & \ldots & R_X(0) \end{bmatrix} \begin{bmatrix} \hat{w}_0^* \\ \hat{w}_1^* \\ \vdots \\ \hat{w}_M^* \end{bmatrix} = \begin{bmatrix} p_{YX}(0) \\ p_{YX}(1) \\ \vdots \\ p_{YX}(M) \end{bmatrix} \quad (116)$$

Wiener filter

- Tthat is

$$\hat{\mathbf{w}}^* = \mathbf{R}_X^{-1}\mathbf{p}_{YX} \tag{117}$$

where $\mathbf{R}_X$ is the autocorrelation matrix for the process $\{X\}$ and $\mathbf{p}_{DY}$ is the cross-correlation vector in the right-hand side of equation (116).

- The minimum MSE yielded by the Wiener solution is given by

$$\xi_{min} = E\{y^2(n)\} - \mathbf{p}_{YX}^T\mathbf{R}_X^{-1}\mathbf{p}_{XY} \tag{118}$$

Example 7.8

- Let $y(n)$ and $v(n)$ be realizations of two first-order AR processes characterized by

$$y(n) = \alpha_1 w_1(n) + 0.8y(n-1) \quad (119)$$

$$v(n) = \alpha_2 w_2(n) - 0.8v(n-1) \quad (120)$$

  where $w_1(n)$ and $w_2(n)$ are uncorrelated white noise signals.

- Assume that $y(n)$, $v(n)$, $w_1(n)$, and $w_2(n)$ all have unit variance.

- Determine the second-order Wiener filter for the jointly WSS processes $\{Y\}$ and $\{X\}$, where

$$x(n) = \frac{1}{2}y(n) + \frac{\sqrt{3}}{2}v(n) \quad (121)$$

Example 7.8 - Solution

- From equation (63), as $\sigma_{W_1}^2 = \sigma_{W_2}^2 = 1$, the autocorrelations of $y(n)$ and $v(n)$ are such that

$$E[y(n)y(n-m)] = \alpha_1^2 \frac{(0.8)^{|m|}}{1-0.8^2} \quad (122)$$

$$E[v(n)v(n-m)] = \alpha_2^2 \frac{(-0.8)^{|m|}}{1-0.8^2} \quad (123)$$

and since $\sigma_Y^2 = \sigma_V^2 = 1$, one has $\alpha_1^2 = \alpha_2^2 = (1-0.8^2)$.

Example 7.8 - Solution

- In addition, since $w_1(n)$ and $w_2(n)$ are uncorrelated, so are $y(n)$ and $v(n)$, and then

$$R_X(0) = \left(\frac{1}{2}\right)^2 R_Y(0) + \left(\frac{\sqrt{3}}{2}\right)^2 R_V(0) = \frac{1}{4} + \frac{3}{4} = 1 \quad (124)$$

$$R_X(1) = \left(\frac{1}{2}\right)^2 R_Y(1) + \left(\frac{\sqrt{3}}{2}\right)^2 R_V(1) = \frac{1}{4}(0.8) + \frac{3}{4}(-0.8) = -0.4 \quad (125)$$

$$R_{YX}(0) = \frac{1}{2} R_Y(0) = 0.5 \quad (126)$$

$$R_{YX}(1) = \frac{1}{2} R_Y(1) = 0.4 \quad (127)$$

Example 7.8 - Solution

- Therefore, as given by equation (116), the Wiener filter coefficients are the solution of

$$\begin{bmatrix} 1 & -0.4 \\ -0.4 & 1 \end{bmatrix} \hat{\mathbf{w}}^* = \begin{bmatrix} 0.5 \\ 0.4 \end{bmatrix} \tag{128}$$

  which corresponds to the Wiener filter

$$W(z) = \frac{11}{14} + \frac{10}{14} z^{-1} \tag{129}$$

Wiener filter

- The Wiener filter determines the best closed-form relationship, in the MSE sense, between the realizations $y(n)$ and $x(n)$ of two joint WSS processes.
- In that context, the LP problem can be seen as a special case where $y(n)$ represents a future sample $x(n+1)$ of the other process.
- Due to its generality, the Wiener filter finds wide application in a variety of problems such as system identification, channel equalization, and noise cancellation, besides, of course, linear prediction.

Other methods for spectral estimation

- The list of PSD estimation methods is almost endless.
- For AR modeling, for instance, we may modify Burg's method by incorporating a window function onto the reflection coefficient estimate given in equation (108). The result is a PSD estimate less prone to bias due to phase dependence on sinusoidal components of the input signal.

Other methods for spectral estimation

- Another technique, also based on the Levinson-Durbin algorithm, is due to Itakura and Saito, and employs a reflection coefficient estimate given by

$$k_i = -\frac{2\sum_{n=i}^{L-1} x_{f,[i-1]}(n)x_{b,[i-1]}(n-1)}{\sqrt{\left(\sum_{n=i}^{L-1} x^2_{b,[i-1]}(n-1)\right)\left(\sum_{n=i}^{L-1} x^2_{f,[i-1]}(n)\right)}} \tag{130}$$

- Alternatively, the same idea of combining forward and backward prediction errors can be applied directly to the LP coefficients, originating a modified version of the covariance method.

Other methods for spectral estimation

- All these AR-modeling modifications yield minor differences on the resulting PSD estimation, and may be seen as variations of the methods described here.
- We focused on AR models due to the simplicity of the associated algorithms, their capacity of modeling ARMA systems, and to the excellent quality of the obtained PSD estimates.

Other methods for spectral estimation

- All these AR-modeling modifications yield minor Algorithms for ARMA and MA models do exist and can be effectively employed when the spectrum at hand presents significant number of zeros.
- Working directly with ARMA or MA models, we avoid the high order required to approximate these functions by AR systems. The price to be paid is the computational complexity of the ARMA and MA modeling algorithms, which may converge to local-optimal solutions due to the highly nonlinear equations associated.

Other methods for spectral estimation

- The problem of PSD estimation may be particularized for several types of random processes.
- The practical case of sinusoids embedded in noise is of crucial interest since it is related to several communications or radar/sonar systems.
- Algorithms that are customized for this problem, including Prony's, MUSIC, or subspace methods, are classified as parametric, since they assume a particular structure for the data at hand.
- Due to their general complexity, these methods are not covered in this book, and the reader is once again directed to the vast literature on the subject to find out more about them.

Other methods for spectral estimation

- There is another very important group of methods that can be employed for PSD estimation.
- The so-called adaptive algorithms are often employed in real-time situations, where we do not have access to a significant amount of data in advance, or when the process presents nonstationary statistical properties.
- Adaptive filters repeatedly adjust their transfer-function coefficients, generating a time-varying model with an associated PSD function.

Do-It-Yourself: Spectral estimation

- **Experiment 7.1:** In this experiment, we consider an ARMA process z generated by applying a white Gaussian noise, with zero mean and unit variance, to a filter with transfer function

$$H(z) = \frac{z^2 - \sqrt{3}z + 1}{10(z^2 + 1.8z + 0.96)} \tag{131}$$

- This transfer function has zeros on the unit circle with angles ± 0.5236 radians and the poles are placed at angles ± 2.7352 radians with radius 0.9798.

- The corresponding power response is shown in Figure 9.

Do-It-Yourself: Spectral estimation

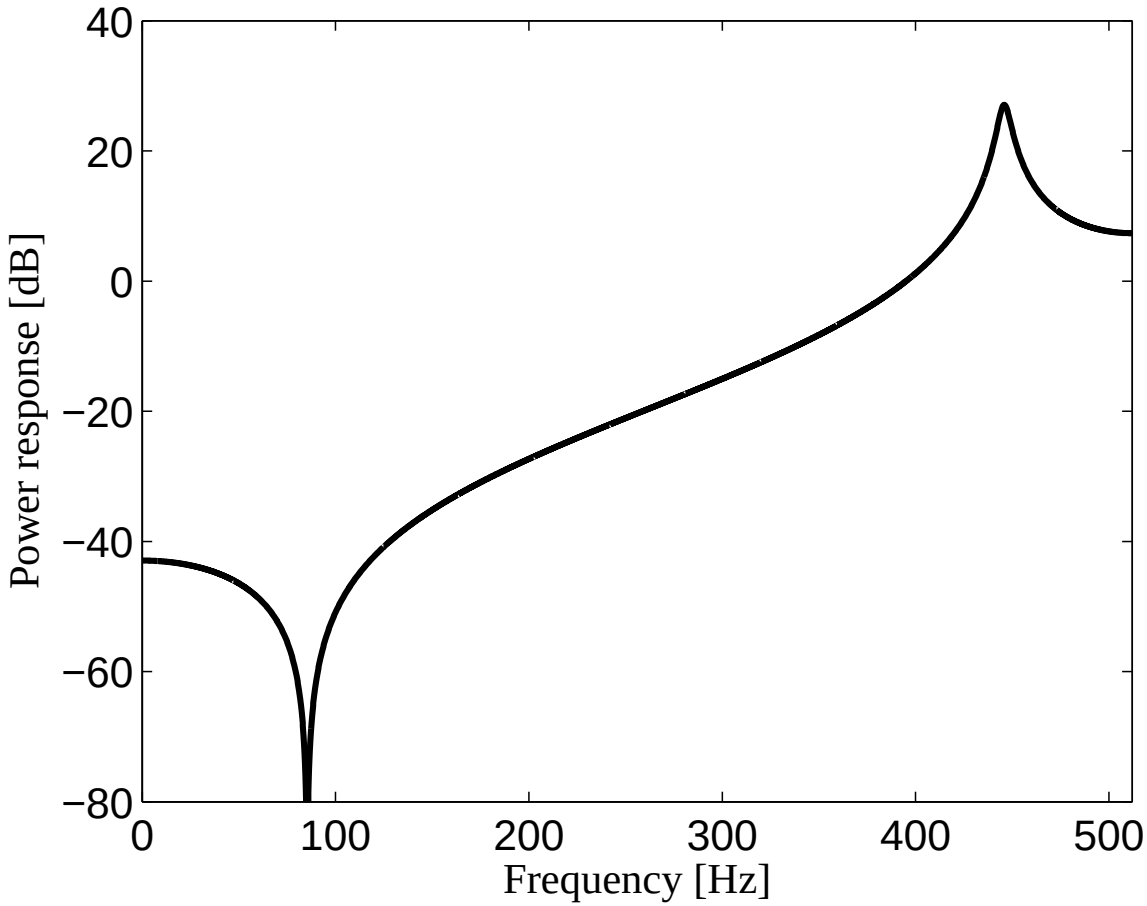


Figure 9: Power response associated to $\mathrm{H}(z)$ in Experiment 7.1.

Do-It-Yourself: Spectral estimation

- Assume that this ARMA process is corrupted by a sinusoidal component `s` of frequency `f1 = fs/8`, with `fs = 1024` Hz, such that 250 ms of the measured signal `y` can be determined in MATLAB as

```
fs = 1024; Ts = 1/fs; t = 0:Ts:(0.25-Ts); L =
length(t);
f1 = 128; s1 = sin(2*pi*f1.*t);
x = randn(1,L); x = x-mean(x); x = x./sqrt(var(x));
b = [1 -sqrt(3) 1.0]; a = 10*[1 1.8 0.96]; z =
filter(b,a,x);
y = z + s1;
```

- Therefore, the PSD associated to the process $\{Y\}$ incorporates a delta function with area 0.5 at $f = f1$ to the power spectrum of $\{Z\}$ seen in Figure 9.

Do-It-Yourself: Spectral estimation

- The PSD estimate `Gamma_Y_P` using the periodogram algorithm, based on an `Lf`-point FFT, with `Lf = 2048`, can be determined as

```
Ry = xcorr(y,'biased');
Gamma_Y_P = 10*log10(abs(fft(Ry,Lf)));
```

- The estimated PSD is shown in Figure 10a, where it can be observed that both the sinusoid frequency and the filter resonance frequency, indicated by the dotted vertical lines, are located at their proper positions.

Do-It-Yourself: Spectral estimation

- By dividing the data into `nb = 4` blocks of length `L/nb`, we may obtain distinct periodogram estimates, which can be averaged to generate the PSD estimate `Gamma_Y_AP` using the script

```
YBlock = reshape(y,L/nb,nb);
RYBlock = zeros(2*L/nb-1,nb);
for i = 1:nb,
   RYBlock(:,i) = xcorr(YBlock(:,i),'biased');
end;
Gamma_Y_AP = 10*log10(mean(abs(fft(RYBlock,LRy)')));
```

- The result, as seen in Figure 10b, is a smoothed estimate less able to discriminate the sinusoidal component.

Do-It-Yourself: Spectral estimation

- Another periodogram variation uses windowed data as given by

```
w = hamming(L)'; wy = w.*y;
Rwy = xcorr(wy,'biased');
Gamma_WY = 10*log10(abs(fft(Rwy,Lf)));
```

  the PSD estimate of which is shown in Figure 10c.

- From this plot, one observes that the main effect of the data-windowing operation is to discriminate the sinusoid with less resolution and turn the estimated curve a bit smoother, as compared to the original periodogram estimate.

Do-It-Yourself: Spectral estimation

- Using an unbiased estimate for the autocorrelation function of the windowed data, as given by

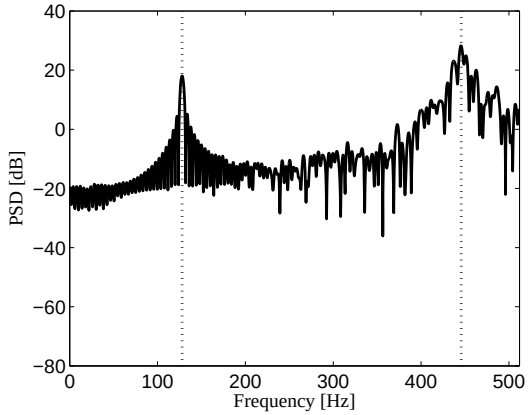
```
Rwyu = xcorr(wy,'unbiased');
Gamma_WYU = 10*log10(abs(fft(Rwyu,Lf)));
```

  results in the PSD estimate seen in Figure 10d.

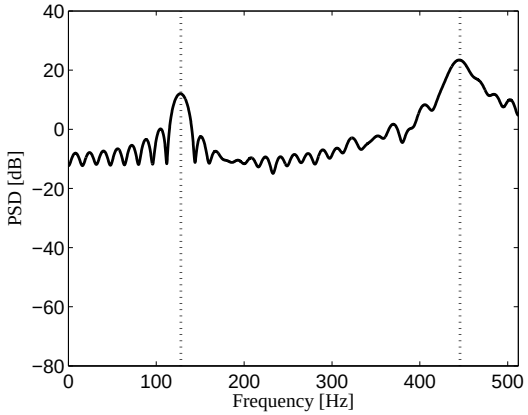
- As expected, this approach leads to a sharp peak in the vicinity of the sinusoid frequency with a very high variance around the entire frequency range.

Do-It-Yourself: Spectral estimation

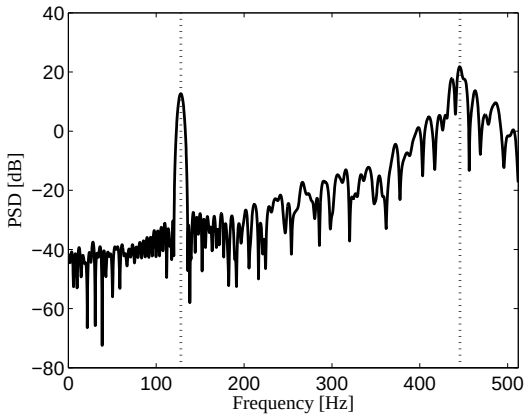
(a)



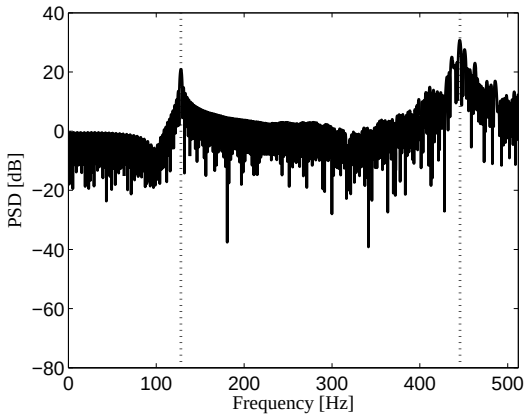
(b)



(c)

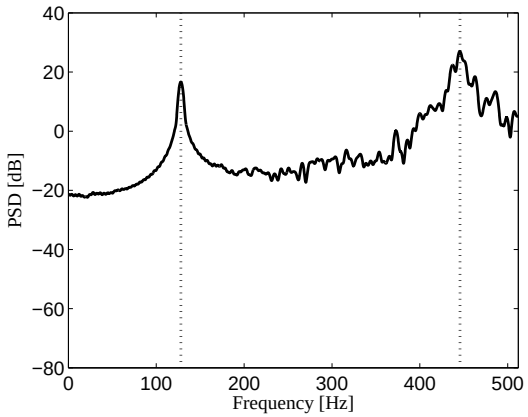


(d)



Do-It-Yourself: Spectral estimation

(e)



(f)

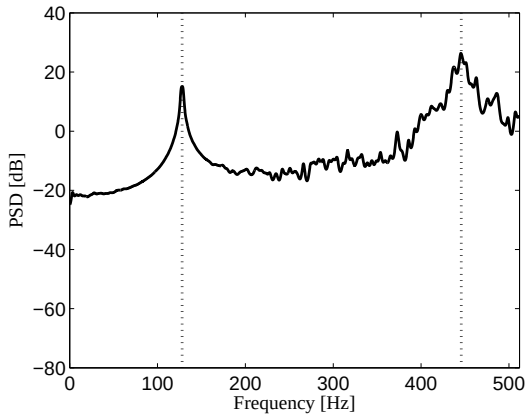


Figure 10: PSD estimates for periodogram-based methods: (a) standard periodogram; (b) averaged periodogram; (c) windowed-data periodogram; (d) windowed-data periodogram with unbiased autocorrelation; (e) Blackman-Tukey scheme; (f) minimum-variance method.

Do-It-Yourself: Spectral estimation

- Using the autocorrelation function `Ry` determined for the standard periodogram method, the Blackman-Tukey estimator can be implemented as `RyBT =`

```
hamming(2*L-1)'.*Ry;
Gamma_BT = 10*log10(abs(fft(RyBT,Lf)));
```

- As can be seen in Figure 10e, the PSD for the Blackman-Tukey estimator is quite smooth and presents a broad peak associated to the sinusoidal component.

Do-It-Yourself: Spectral estimation

- The MV method can be implemented in MATLAB according to the script:

```
wy = y' - mean(y);
wy_pad = [zeros(L-1,1);wy;zeros(L-1,1)]';
for i=1:L,
   XU(:,i) = wy_pad(L-i+1:3*L-1-i);
end;
Rwy = XU'*XU/(L-1);
[eigvec,eigval] = eig(Rwy);
U = abs(fft(eigvec,LRy)).^2; V = diag(inv(eigval +
eps));
Gamma_MV = 10*log10(L)-10*log10(U*V);
```

Do-It-Yourself: Spectral estimation

- This implementation follows decomposes $\mathbf{R}_Y^{-1}$ by its spectral decomposition

$$\mathbf{R}_Y^{-1} = \sum_{l=0}^{L-1} \frac{\mathbf{q}_l \mathbf{q}_l^{*^T}}{\lambda_l} \tag{132}$$

  where $\mathbf{q}_l$ and λ_l, for $l = 1, 2, \ldots, L$, are the eigenvectors and eigenvalues of $\mathbf{R}_Y$.

- As a result, one can rewrite equation (38) as

$$\Gamma_{Y,MV}(e^{j\omega}) = \frac{L}{\sum_{l=0}^{L-1} \frac{|\mathbf{e}^{*^T}(e^{j\omega})\mathbf{q}_l|^2}{\lambda_l}} \tag{133}$$

- The terms $\mathbf{e}^{*^T}(e^{j\omega})\mathbf{q}_l$ represent the DFT of the lth eigenvector.

Do-It-Yourself: Spectral estimation

- In the script above, matrix U stores the terms $|\text{FFT}[\mathbf{q}_l]|^2$ and the diagonal matrix V stores the inverse of the eigenvalues of $\mathbf{R}_Y^{-1}$, with a conditioning factor eps to avoid division by zero.
- Using the MV method, the PSD estimate for a single run of the recorded data is depicted in Figure 10f, where we can observe a strong similarity, in this case, to the Blackman-Tukey estimate.

Do-It-Yourself: Spectral estimation

- **Experiment 2:** In this experiment, we investigate the performances of some parametric methods.

- For the autocorrelation algorithm with order `N`, one may employ the script

```
Ry = xcorr(y2,N,'biased');
Ryy = toeplitz(Ry(N+1:2*N));
pyy = Ry(N+2:2*N+1);
a = inv(Ryy)*pyy';
H_AC_dB = 20*log10(abs(freqz(1,[1; -a],LRy/2+1));
```

  with `LRy` as defined in Experiment 1, making the resulting PSD estimate compatible to the frequency vector `freq` determined in the same experiment.

- This command sequence is equivalent to

```
[b,E] = lpc(y2,N);
```

  with `b = [1; -a]`.

Do-It-Yourself: Spectral estimation

- For other parametric approaches, we may resort to a powerful spectral-analysis tool in MATLAB whose usage is exemplified here for Burg's algorithm:

```
h = spectrum.burg(N,'UserDefined');
set(hopts,'NFFT',LRy);
HBurg = psd(h,y2,hopts);
HBurg_dB = 10*log10(abs(HBurg.Data));
```

- Figure 11 shows the resulting PSDs for the autocorrelation and Burg's methods for `N = 4` and `N = 8` in each case, indicating the good performances achieved by both methods particularly for larger values of `N`.

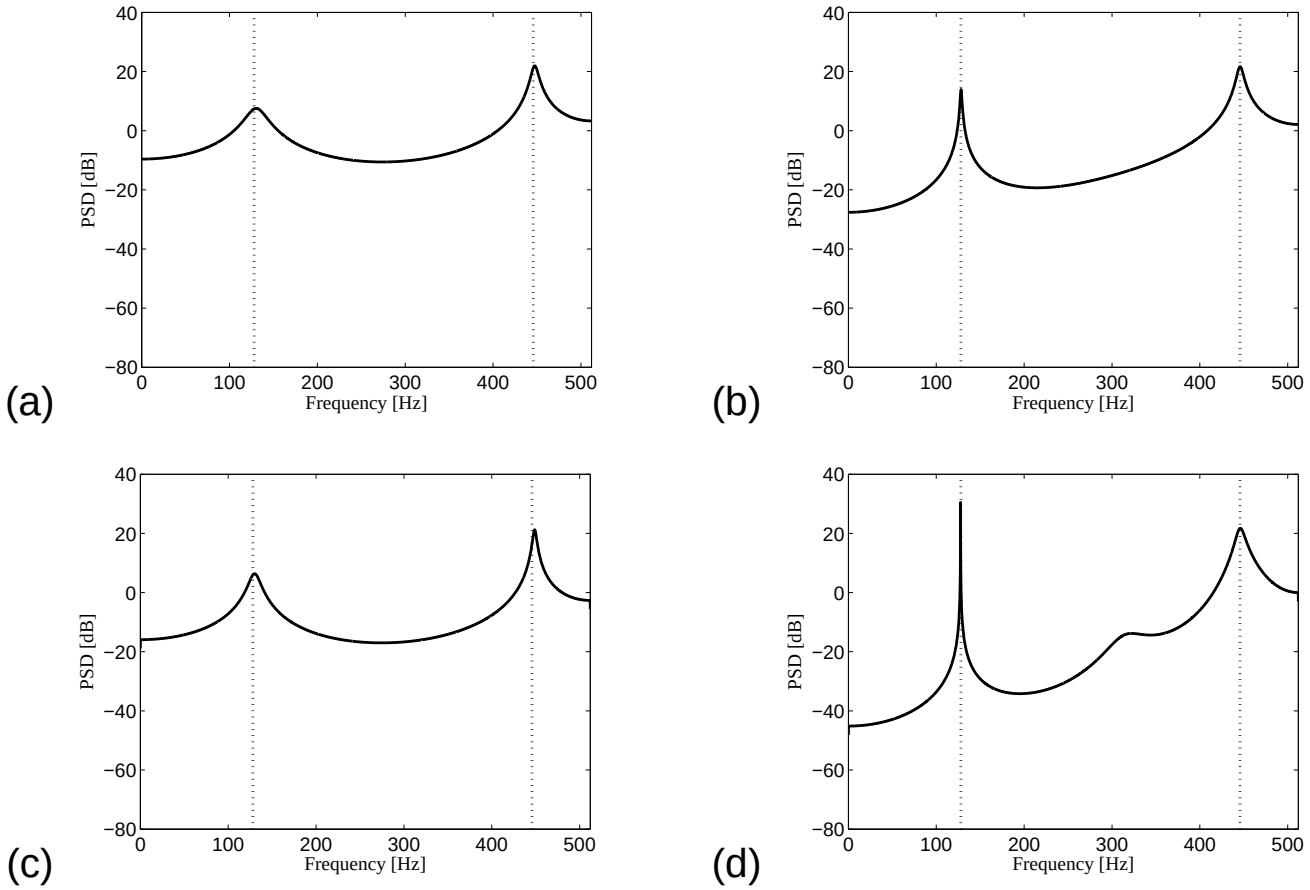


Figure 11: PSD estimates for AR methods when `f1 = fs/8`: (a) autocorrelation with `N = 4`; (b) autocorrelation with `N = 8`; (c) Burg's algorithm with `N = 4`; (d) Burg's algorithm with `N = 8`.

Do-It-Yourself: Spectral estimation

- If, however, we move the sinusoid frequency `f2` closer to the ARMA pole frequencies, by making, for instance,

```
f2 = 3.3*fs/8; s2 = sin(2*pi*f2.*t);
y2 = z + s2;
```

  the resulting PSD estimates for the two methods with `N = 8` and `N = 24` are depicted in Figure 161

- From these plots, one clearly notices how in this case both methods required a very large model order `N` to discriminate the two PSD peaks satisfactorily.

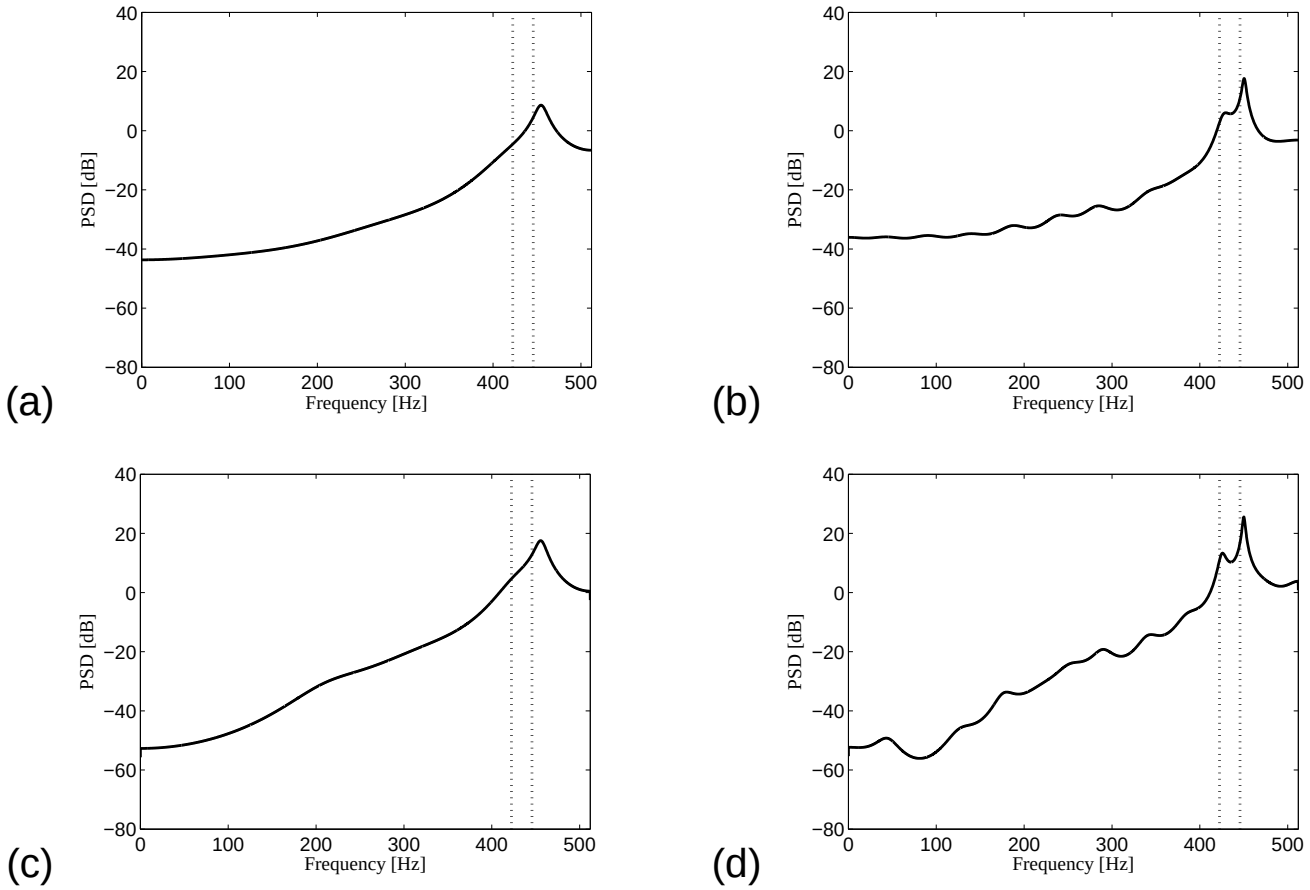


Figure 12: PSD estimates for AR methods when $f2 = 3.3fs/8$: (a) autocorrelation with $N = 8$; (b) autocorrelation with $N = 24$; (c) Burg's algorithm with $N = 8$; (d) Burg's algorithm with $N = 24$.

Do-It-Yourself: Spectral estimation

- Figure 13 shows the pole-zero constellations for the AR model generated by the autocorrelation method with $N = 8$ and $N = 24$.
- It can be observed that the AR estimate places some poles closer to the unit circle in the frequency range where the input signal has higher power.
- Although this phenomenon happened in both cases, when $N = 8$ the number of available poles was not enough to detect the presence of the sinusoid.

Do-It-Yourself: Spectral estimation

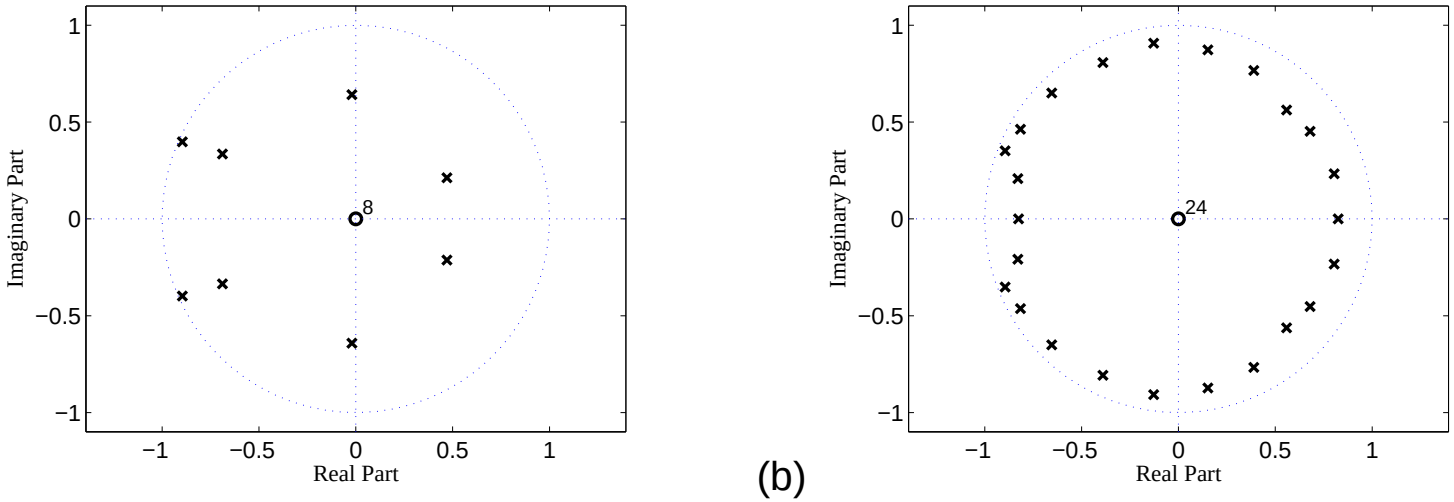


Figure 13: Pole-zero plot for the AR models generated by the autocorrelation algorithm for Experiment 2: (a) $N = 8$; (b) $N = 24$.

Do-It-Yourself: Spectral estimation

- Figure 14 shows 15 PSD estimates superimposed using the nonparametric Blackman-Tukey and the parametric Burg (with $N = 24$) methods.
- As can be observed from the overlaid plots, the result for the nonparametric method, despite being able to locate better the sinusoid, has much higher variance than the parametric method.

Do-It-Yourself: Spectral estimation

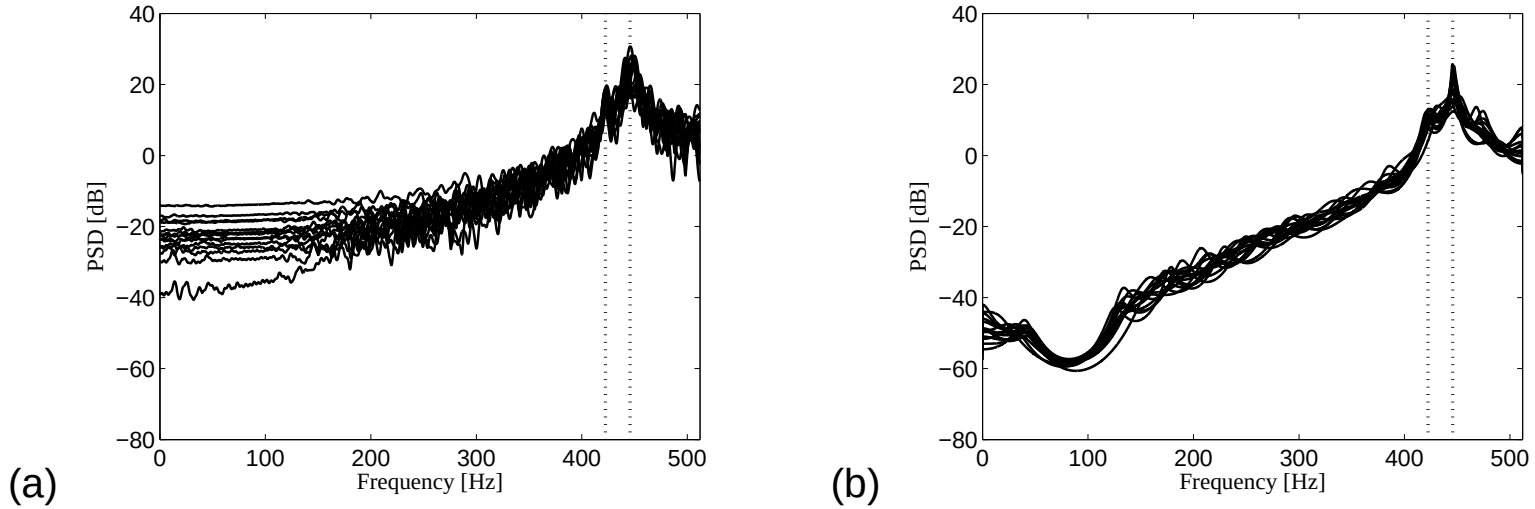


Figure 14: PSD estimates for 15 independent runs in Experiment 2: (a) Blackman-Tukey periodogram; (b) Burg's method with $N = 24$.

Summary

- The nonparametric periodogram method was introduced, along with several of its variations.
- The concept of window function was applied in the context of smoothing the data or the autocorrelation function to yield better (with less bias or with reduced variance) PSD estimates.
- We also addressed the minimum variance nonparametric method which, in general, results in smooth estimates of the PSD.
- The modeling problem, in which digital filters appear playing a major role, was then introduced as the basic tool for performing parametric spectral estimation.

Summary

- Our presentation focused on autoregressive (AR) modeling due to their general ability of mimicking moving-average (MA) and ARMA systems.
- Several methods for determining the AR model, including the so-called covariance, autocorrelation, and Burg's approaches, were presented with different converging characteristics.
- The resulting PSD estimate can then be determined as the power response for the obtained AR model.
- We addressed the Wiener filtering where the optimum filter output produces an estimate of a reference signal by filtering another signal related to the reference.