Power Spectrum Estimation

- Introduction
- From finite duration observations of signals
- Nonparametric methods
- Parametric methods

Introduction

In many applications, the concerned signal $x(n)$ is of the form

$$x(n) = \sum_{k=1}^{N} a_k cos(\omega_k n + \phi_k) + w(n) \triangleq s(n) + w(n), \qquad (181)$$

where $w(n)$ is a noise, usually assumed statistically independent of $s(n)$.

Frequency estimation means to find out

- how many frequencies are underlying in $x(n)$,
- the values of $\{\omega_k\}$, and sometimes $\{a_k, \phi_k\}$.

Such kinds of problems are of importance in many areas such as

- multi-media signal processing systems, say speech synthesis, and
- telecommunications systems.

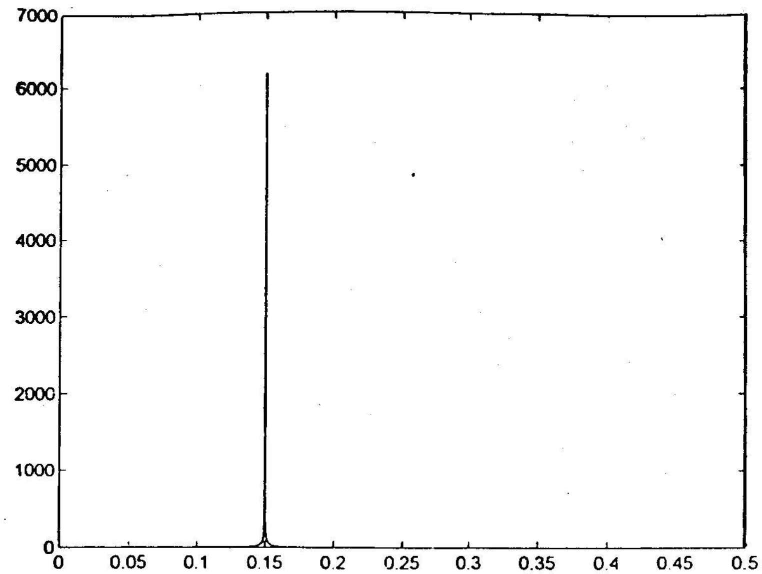
From finite duration observations of signals

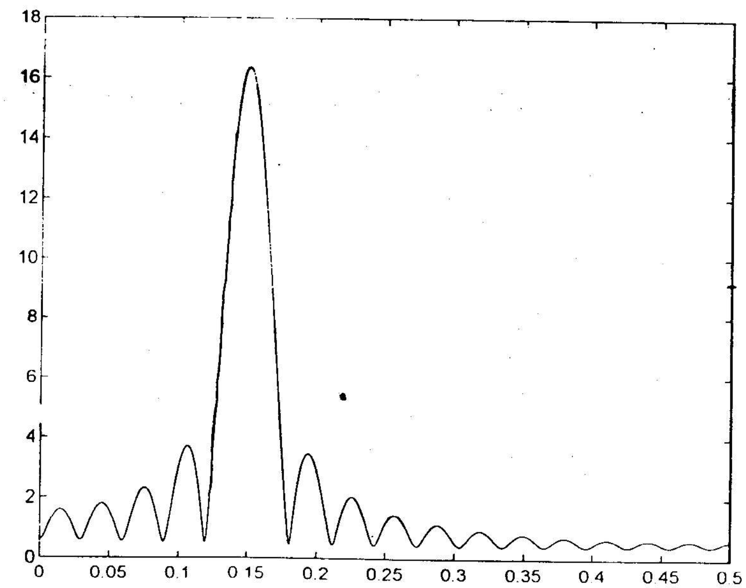
Traditionally, these problems can be attacked using DTFT (or DFT). As known, the DTFT of $x(n) = Acos(2\pi f_c n)$ is

$$X(e^{j2\pi f}) = \frac{A}{2}[\delta_c(f + f_c) + \delta_c(f - f_c)], \quad |f| < 1/2$$

suggesting to find the frequency from the peaks of $X(e^{j2\pi f})$.

Limitation 1: Effect of finite duration of signals
Denote $x_N(n) \triangleq cos(2\pi f_c n)[u(n+N) - u(n-N)]$, a sinusoidal but with a finite duration of $2N$ samples.
The following Fig. shows the magnitude responses for $N = 2^{13}, 2^4$ with $f_c = 0.15$.

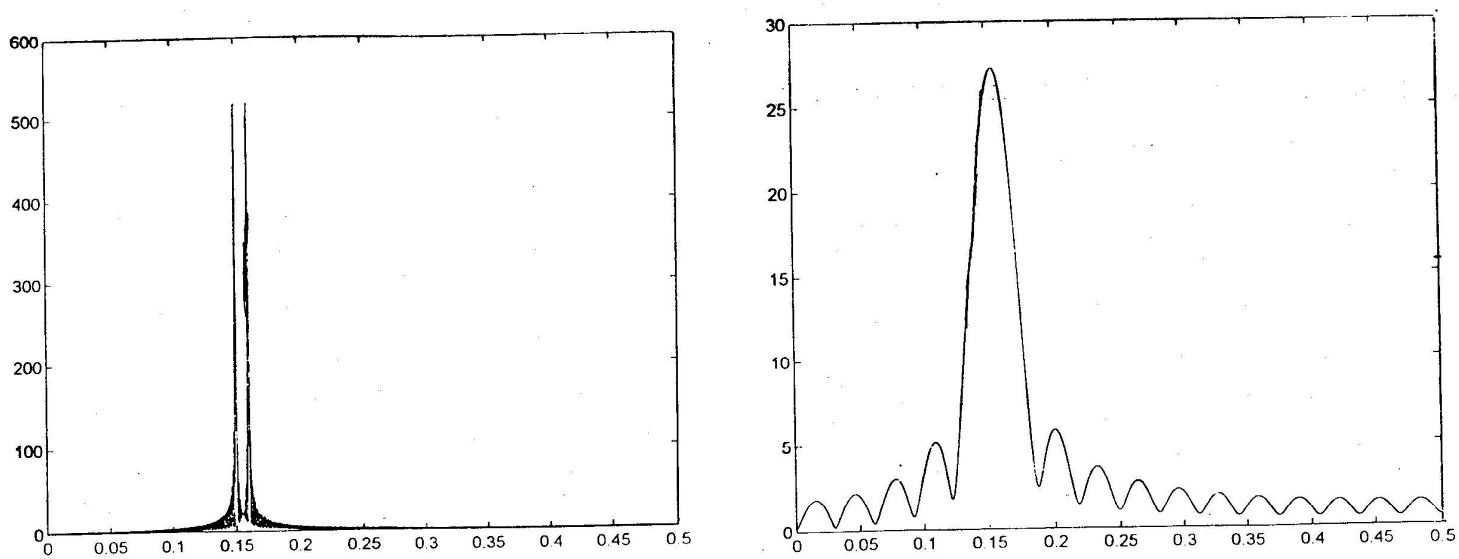




Magnitude responses: Left - $N = 2^{13}$; right - $N = 2^4$

When N decreases, some side-lobes appear. In that case, more than one frequency would be detected.

Now, let us consider
$x_N(n) = [cos(2\pi \times 0.15n) + cos(2\pi \times 0.16n)][u(n+N) - u(N-N)]$ with $N = 2^9$ and $N = 2^4$, respectively.



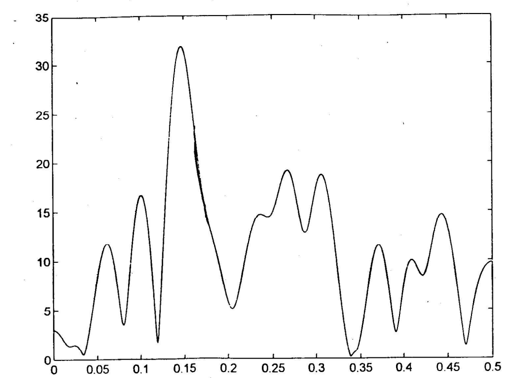
Magnitude responses: Left - $N = 2^9$; right - $N = 2^4$

One can see that when N decreases it becomes difficult to distinguish the two sinusoids, and they are indistinguishable when $N = 2^4$.

Limitation 2: Disturbance by the attached noise

Now, let us consider the situation when a sinusoid is corrupted with an additive white noise and $N = 2^4$:

$$x_N(n) = [cos(2\pi \times 0.15n) + w(n)][u(n+N) - u(n-N)]$$



Magnitude response with $N = 2^4$ when noise is attached

It is clear that more than one frequencies will be claimed from the spectrum, while there is only one sinusoid.

Conclusion: New approaches are needed!

As seen before, $x(n) = Acos(2\pi f_0 n + \phi) + w(n)$ has the auto-correlation $r_x(l)$ (or $\gamma_{xx}(m)$ previously)

$$r_x(l) = \frac{A^2}{2} cos(2\pi f_0 \ l) + \sigma_w^2 \delta(n)$$

and hence the *power spectral density* (PSD) $R_x(e^{j2\pi f})$ (or power density spectrum $\Gamma_{xx}(f)$ previously)

$$R_x(e^{j2\pi f}) = \frac{A^2}{4}[\delta_c(f + f_0) + \delta_c(f - f_0)] + \sigma_w^2, \quad |f| < 1/2$$

Obviously, it is very easy to detect the frequency f_0 from the peaks of the PSD, i.e, the DTFT of the auto-correlation $r_x(l)$, rather than $x(n)$ itself!

This is one of the motivations for studying *power spectrum estimation* (PSE) - an estimate of the PSD $R_x(e^{j2\pi f})$!

The problem of PSE can be stated as: with N samples of $x(n)$ given, to estimate $R_x(e^{j2\pi f})$.

In general, there are two different approaches to PSE:

- Non-parametric: Estimate $r_x(m)$ with $x(n)$, then compute its DFT.
- Parametric: Based on signal modelling, estimate the model parameters using $x(n)$, then compute its DFT.

Nonparametric methods

Intuitively, one may estimate $r_x(m)$ with $x_N(n)$ first, then take the DTFT of the estimate of $r_x(m)$ as the estimate of the true PSD.

A. Jenkins and Watts estimator
Naturally, use the time-average autocorrelation sequence

$$\check{r}_x(m) \triangleq \begin{cases} \frac{1}{N-m}\sum_{n=0}^{N-m-1} x(n+m)x^*(n), & 0 \le m < N \\ \check{r}^*{}_x(-m), & -N < m < 0 \\ 0, & elsewhere \end{cases} \qquad (182)$$

Performance analysis:

$$E[\check{r}_x(m)] = \begin{cases} r_x(m), & |m| < N \\ 0, & elsewhere \end{cases}$$

which means that $\breve{r}_x(m)$ is an *unbiased estimate* of $r_x(m)$ for the lags $|m| < N$.
It was shown by Jenkins and Watts in 1968

$$\begin{aligned} var[\breve{r}_x(m)] &\triangleq E\{|\breve{r}_x(m) - E[\breve{r}_x(m)]|^2\} \\ &\approx \frac{N}{(N-|m|)^2} \sum_{n=-\infty}^{+\infty} [|r_x(n)|^2 + r_x(n-m)r_x(n+m)]. \end{aligned}$$

Clearly, $\lim_{N\to+\infty} var[\breve{r}_x(m)] = 0$ provided that $r_x(m)$ is an energy sequence, that is $\sum_{m=-\infty}^{+\infty} |r_x(m)|^2 < +\infty$. Therefore, the estimator by (182) is *consistent*.

It is easy to see that for large values of the lag parameter m, the estimate $\check{r}_x(m)$ has a large variance, especially as m approaches N.
This is due to the fact that fewer data points enter into the estimate for large lags.

B. Schuster estimator

Alternatively, we have

$$\hat{r}_x(m) \triangleq \begin{cases} \frac{1}{N}\sum_{n=0}^{N-m-1} x(n+m)x^*(n), & 0 \leq m < N \\ \hat{r}^*(-m), & -N < m < 0 \\ 0, & \textit{elsewhere} \end{cases} \tag{183}$$

It is interesting to note that with

$$\mathbf{x}(n) \triangleq [\cdots\ 0\ x(0)\ x(1)\ \cdots\ x(N-1)\ 0\ \cdots]^{\mathcal{T}}$$

$\hat{r}_x(m)$ can be rewritten into

$$\hat{r}_x(m) = \frac{1}{N}\,\mathbf{x}^{\mathcal{T}}(n+m)\mathbf{x}^*(n)$$

It is easy to see that for $|m| < N$,

$$E[\hat{r}_x(m)] = \frac{N-|m|}{N} r_x(m) = (1-\frac{|m|}{N})r_x(m),$$

which implies that this estimator has a bias of $\frac{|m|}{N}\ r_x(m)$.
This estimator, however, has a smaller variance

$$var[\hat{r}_x(m)] \approx \frac{1}{N}\sum_{n=-\infty}^{+\infty}[|r_x(n)|^2 + r_x^*(n-m)r_x(n+m)].$$

If $r_x(m)$ is an energy sequence, the Schuster yields a consistent estimate of $r_x(m)$ as

$$\lim_{N\to+\infty} E[\hat{r}_x(m)] = r_x(m), \quad \lim_{N\to+\infty} var[\hat{r}_x(m)] = 0$$

Notes:

- Both $\check{r}_x(m)$ and $\hat{r}_x(m)$ yield a consistent estimate of $r_x(m)$,
- based on which the estimate of $R_x(e^{j2\pi f})$ is given respectively

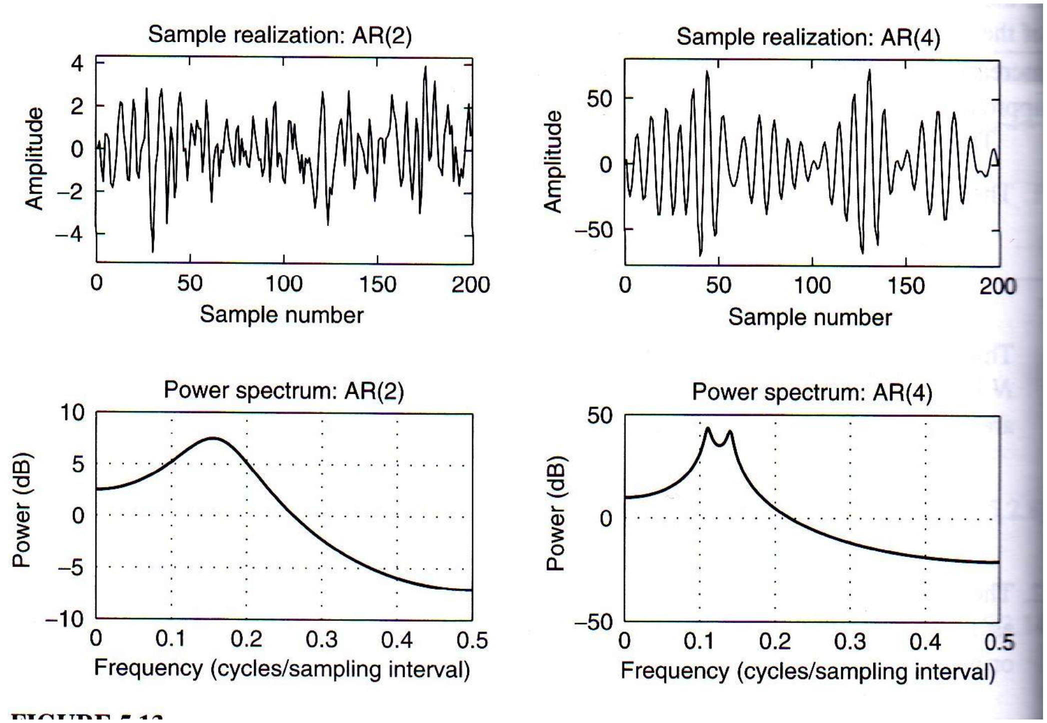
$$\begin{cases} \check{R}_x(e^{j2\pi f}) = \sum_{m=-(N-1)}^{N-1} \check{r}_x(m)e^{-j2\pi fm} \\ \hat{R}_x(e^{j2\pi f}) = \sum_{m=-(N-1)}^{N-1} \hat{r}_x(m)e^{-j2\pi fm} \end{cases} \quad (184)$$

However, the two estimators[4] by (184)

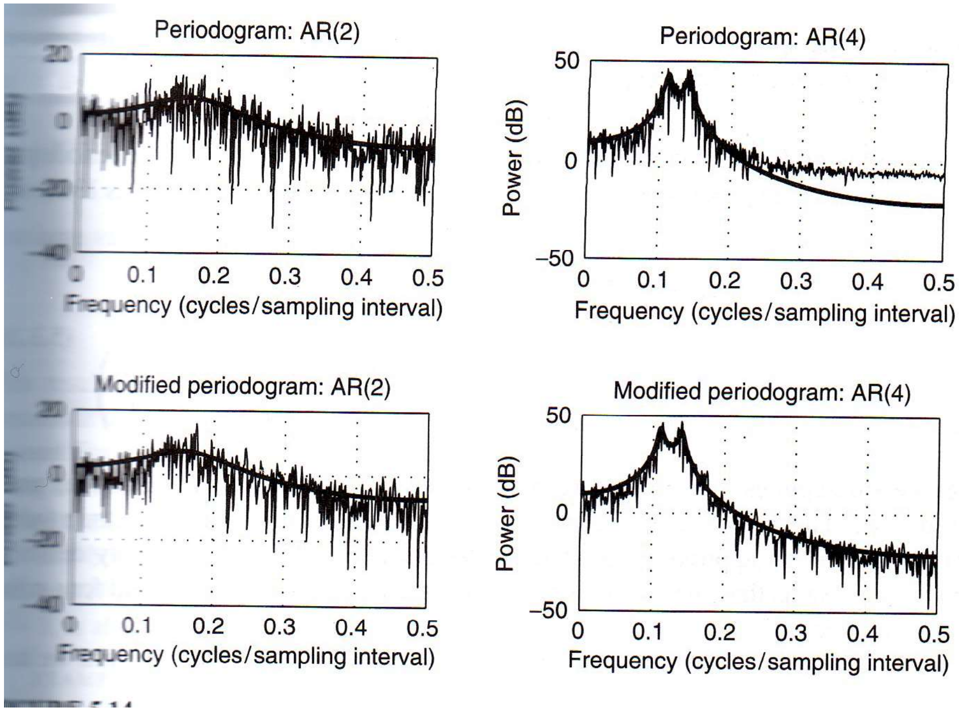
- are NOT a consistent estimate of the true $R_x(e^{j2\pi f})$ and
- suffer from the finite duration of data. This ultimately limits the ability to resolve closed spaced spectra.

[4]Note, it can be shown that $\hat{R}_x(e^{j2\pi f}) = \frac{1}{N}|\mathbf{X}(e^{j2\pi f})|^2$ and the latter is known as *periodogram* of the sequence $\mathbf{x}(n)$ defined above.

Example : Two signals AR(2) and AR(4) are generated with a white noise $w(n)$ of $N = 1024$ samples.



Sample realizations and the true power spectra of the AR(2) and AR(4) processes.



Estimates of power spectral density using two different methods.

There are some alternative non-parametric estimators (see the textbooks).

Parametric methods

Basic idea behind:

- Consider the PSD $R_x(e^{j2\pi f})$ as a function of the frequency variable f, characterized with a parameter vector θ, i.e., $P_x(\theta, f)$ and then
- determine the optimal θ, say θ_{opt} with the measurements $x_N(n)$ such that $P_x(\theta_{opt}, f)$ (or in short, $P_x(f)$) is as close to $R_x(e^{j2\pi f})$ as possible in a certain sense.

Is that possible?

System-based signal models

This is supported by the *innovations representation* of WSS processes discussed in Chapter 1. In fact, the $ARMA(p,q)$ model yields a very large class of WSS processes and

$$R_x(e^{j2\pi f}) = \sigma_w^2 |H(e^{j2\pi f})|^2 \quad (185)$$

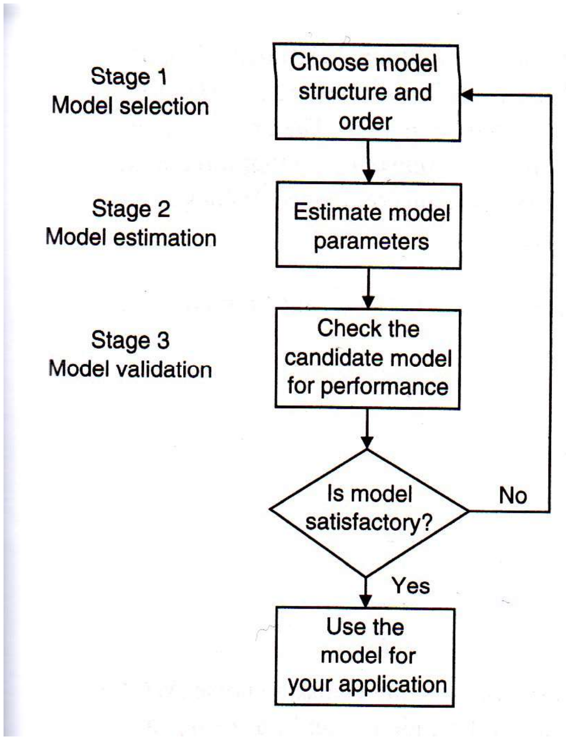
where

$$H(z) = \frac{1 + \sum_{k=1}^{q} b_k z^{-k}}{1 + \sum_{k=1}^{p} a_k z^{-k}} = \frac{B(z)}{A(z)}$$

If $H(z)$ is characterized with $\{a_k, b_k\}$, then

$$\theta = [a_1 \; \cdots \; a_p \; b_1 \; \cdots \; b_q \; \sigma_w^2]^T$$

Issues involved



Issues in system-based PSE.

Algorithms for AR parameter estimation

For $AR(p)$ models, we have $b_k = 0, \ \forall \ k \geq 1$. Therefore, the parameters to be estimated are $\{a_k\}$ and σ_w^2. Let $\{\hat{a}_k\}$ and $\hat{\sigma}_w^2$ be the corresponding estimates, the obtained PSD is

$$R_x^{AR}(\theta, f) = \frac{\hat{\sigma}_w^2}{|1 + \sum_{k=1}^{p} \hat{a}_k e^{-j2\pi fk}|^2}. \tag{186}$$

where the optimal parameters

$$\theta = [\hat{a}_1 \ \cdots \ \hat{a}_p \ \hat{\sigma}_w^2]^{\mathcal{T}}$$

Depending on how to compute θ, we have three different methods.

A1. The Yule-Walker method

Let $\hat{r}_x(m)$ be the estimate of $r_x(m)$, obtained using an estimator with $x_N(n)$ (say, the Schuster). With $q = 0$, $m = 1, \dots, p$ and $r_x(m)$ replaced by $\hat{r}_x(m)$, (73) becomes

$$\begin{bmatrix} \hat{r}_x(0) & \hat{r}_x(-1) & \dots & \hat{r}_x(-p+1) \\ \hat{r}_x(1) & \hat{r}_x(0) & \dots & \hat{r}_x(-p+2) \\ \vdots & \vdots & \vdots & \vdots \\ \hat{r}_x(p-1) & \hat{r}_x(p-2) & \dots & \hat{r}_x(0) \end{bmatrix} \begin{bmatrix} \hat{a}_1 \\ \hat{a}_2 \\ \vdots \\ \hat{a}_p \end{bmatrix} = - \begin{bmatrix} \hat{r}_x(1) \\ \hat{r}_x(2) \\ \vdots \\ \hat{r}_x(p) \end{bmatrix} \tag{187}$$

from which the *AR* parameters can be computed.

The estimate of σ_w^2 can be obtained from (73) with $h_0 = b_0 = 1, q = 0, m = 0$:

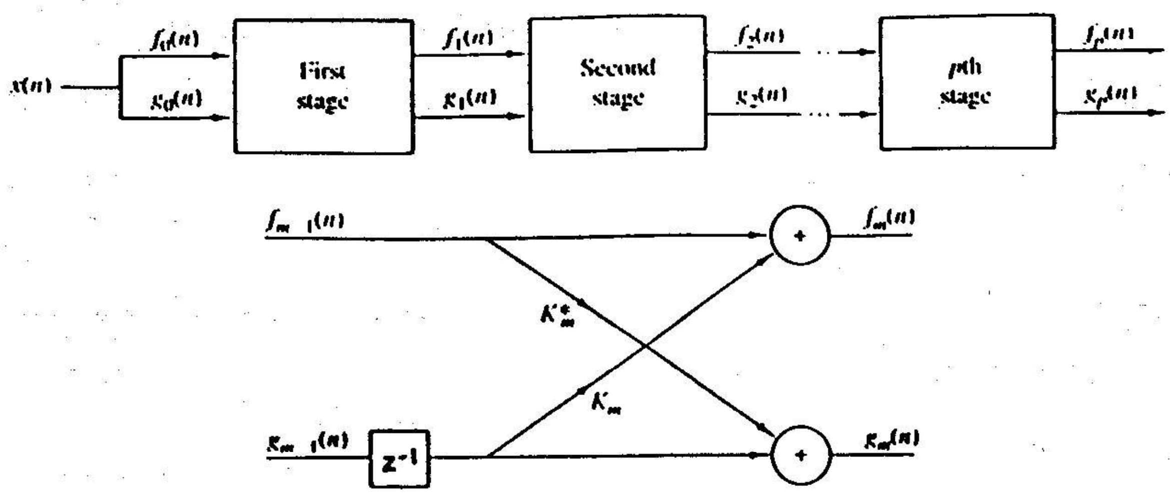
$$\hat{\sigma}_w^2 = \hat{r}_x(0) + \sum_{k=1}^{p} \hat{a}_k \hat{r}_x(-k). \tag{188}$$

In the Yule-Walker method, the Schuster estimator (183) is used to get the estimate $\hat{r}_x(m)$ of $r_x(m)$. By doing so,

- the estimate of the autocorrelation matrix is positive-definite, leading to a stable *AR* model, and
- this matrix is also of *Toeplitz* form, which allows us to use the Levinson-Durbin algorithm to solve (187) efficiently.

A2. Burg method (1968)

This method is based on the lattice structure depicted below:



The structure is specified by

$$\begin{cases} f_0(n) & = & g_0(n) = x(n) \\ f_m(n) & = & f_{m-1}(n) + K_m g_{m-1}(n-1) \\ g_m(n) & = & K_m^* f_{m-1}(n) + g_{m-1}(n-1) \end{cases}$$

for $m = 1, 2, \cdots, p$, where $\{K_m\}$ are called *reflection* or *partial* coefficients.

The set of lattice parameters k_m is one-to-one mapped with the the set of coefficients a_l of

$$A(z) = 1 + \sum_{l=1}^{P} a_l z^{-l}$$

$\{k_m\} \quad \Rightarrow \quad \{a_l\}$

With $A_0(z) = 1$, compute

$$A_m(z) = A_{m-1}(z) + k_m z^{-m} A^*_{m-1}(1/z^*), \;\; m = 1, 2, \cdots, P$$

The final stage yields

$$A(z) = A_P(z) = 1 + a_P(1)z^{-1} + \cdots + a_P(P)z^{-P}$$

Then, we obtain $a_P(1), \ldots, a_P(P)$.

On the other hand
$\{a_l\} \quad \Rightarrow \quad \{k_m\}$
With $A_P(z) = A(z)$ and $k_P = a_P$. Then

$$A_{m-1}(z) \quad = \quad \frac{A_m(z) - k_m z^{-m} A_m^*(1/z^*)}{1 - |k_m|^2}$$

Denote $A_{m-1}(z) = 1 + \sum_{k=1}^{m-1} a_{m-1}(k) z^{-k}$, then

$$k_{m-1} = a_{m-1}(m-1), \quad m = P, \cdots, 2$$

For this method, the optimal coefficients K_l are obtained one by one and

$$\hat{K}_l = \frac{-\sum_{n=l}^{N-1} f_{l-1}(n) g_{l-1}^*(n-1)}{\frac{1}{2}\sum_{n=l}^{N-1}[|f_{l-1}(n)|^2 + |g_{l-1}(n-1)|^2]}, \tag{189}$$

which minimizes

$$\epsilon_l(K_l) \triangleq \frac{1}{2} \sum_{n=l}^{N-1} [|f_l(n)|^2 + |g_l(n)|^2], \ \forall l \tag{190}$$

with $f_l(n)$ and $g_l(n)$ computed with $K_i = \hat{K}_i$ for $i = 1, \cdots, l-1$, while σ_w^2 is estimated using $\hat{\sigma}_w^2 = \epsilon_p(\hat{K}_p)$, with which and the converted $\{\hat{a}_l\}$ from $\{\hat{K}_l\}$, the power density spectrum can be computed with (186).

The major advantages of the Burg method for estimating the parameters of the AR model are

- it results in high frequency resolution due to the fact that the power density spectrum is less sensitive to the errors in $\{\hat{K}_l\}$ then those in $\{\hat{a}_l\}$.
- it yields a stable AR model due to $\hat{K}_l$ given by (189) satisfying $|\hat{K}_l| < 1, \forall l$, and
- it is computationally efficient.

A3. Unconstrained least-squares method

As seen in the Burg method, the optimal reflection coefficient $\hat{K}_l$ is obtained with the assumption that $K_i = \hat{K}_i, \forall i < l$ are known. Therefore, the minimization is just *suboptimal*.

In contrast to this approach, we may use an unconstrained least-squares algorithm to determine the AR parameters $\{a_k\}$.

For the last stage of the lattice,

$$\epsilon_p(\mathbf{a}) = \frac{1}{2} \sum_{n=p}^{N-1} [|f_p(n)|^2 + |g_p(n)|^2],$$

with $\mathbf{a} = [a_1 \ \cdots \ a_p]^{\mathcal{T}}$.

Minimizing $\epsilon_p(\mathbf{a})$ with respect to $\mathbf{a}$, the optimal AR parameters $\widehat{\mathbf{a}}$ and $\hat{\sigma}_w^2$ is taken as $\epsilon_p(\widehat{\mathbf{a}})$. The estimated power density spectrum can be computed with (186). Its performance has been found to be superior to the Burg method.

It should be pointed out that the frequencies underlying $x(n)$ can be detected from the significant peaks of $R_x^{AR}(\widehat{\theta}, f)$ by (185). These peaks are usually strongly related to the roots of $z^p \hat{A}(z) = z^p + \sum_{k=1}^{p} \hat{a}_k z^{p-k}$. If $z = re^{j\phi}$ is a root with $r > 0$ close one, then the power spectrum should have a peak at the frequency very close to $\hat{f} = \frac{\phi}{2\pi}$.

Selection of AR model order

One of the most important aspects on using the *AR* model is the section of the order p.

As a general rule, if a model with a too low order, a highly smoothed spectrum is obtained. On the other hand, if p is selected too high, we run the risk of introducing spurious low-level peaks in the estimated spectrum.

Two well known criteria for selection of p have been proposed by Akaike (1969, 1974):

- *Final prediction error (FPE) criterion*:
  The order is selected to minimize the following index

$$FPE(p) \triangleq \hat{\sigma}_w^2(p)\frac{N+p+1}{N-p-1}, \tag{191}$$

  where $\hat{\sigma}_w^2(p)$ is the estimate of σ_w^2, changing with the order p of AP model.

- *Akaike information criterion*:
  This criterion is based on selecting the order p that minimizes the following index:

$$AIC(p) \triangleq ln\hat{\sigma}_w^2(p) + \frac{2p}{N}. \tag{192}$$

Power spectrum estimation of *MA* processes

For an *MA* processes of order q, one would need to estimate $\{\hat{b}_k\}$ with $\hat{r}_x(m)$.
With $\hat{r}_x(m)$ replaced with $r_x(m)$, $a_k = 0$, for $k = 1, \ldots, p$, $h[k] = b_k$, (73) becomes

$$\hat{\sigma}_w^2 \sum_{k=0}^{q-m} \hat{b}_{k+m}\hat{b}_k = \hat{r}_x(m), \; m = 0, 1, \cdots, q \tag{193}$$

If $\{\hat{b}_k\}$ and $\hat{\sigma}_w^2$ are obtained from the above equation, the corresponding estimated *MA* power density spectrum is

$$R_x^{MA}(e^{j2\pi f}) = |\sum_{k=0}^{q} \hat{b}_k e^{-j2\pi fk}|^2 \hat{\sigma}_w^2$$

Simple calculation shows that

$$R_x^{MA}(e^{j2\pi f}) \quad = \quad \sum_{m=-q}^{q} (\hat{\sigma}_w^2 \sum_{k=0}^{q-m} \hat{b}_{k+m}\hat{b}_k)e^{-j2\pi fm}.$$

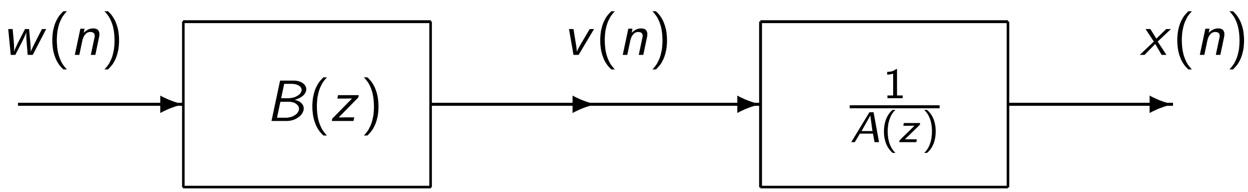
It then follows from (193) that

$$R_x^{MA}(e^{j2\pi f}) = \sum_{m=-q}^{q} \hat{r}_x(m)e^{-j2\pi fm}. \tag{194}$$

This means that to estimate the power density spectrum of an MA process with $\hat{r}_x(m)$ given, one can just use (194) rather than solving (193).

Power spectrum estimation of *ARMA* processes

Note that an $ARMA(p,q)$ process $x(n)$ can be considered being generated.



Block diagram of generation for ARMA processes.

Basic idea:

- Consider $v(n)$ as a white process, equivalently, $x(n)$ as an $AR(p)$ process and then get the estimate $\hat{A}(z)$ with $r_x(m)$.
- The sequence $\hat{v}(n)$, generated as the output of the system $\hat{A}(z)$ excited by $x(n)$, is the estimate of $v(n)$.

- Compute $\hat{r}_{\hat{v}}(m)$ and hence true $|B(e^{j2\pi f})|^2$ can be approximated with $R_{\hat{v}}^{MA}(e^{j2\pi f})$, obtained using (194) with $\hat{r}_v(m)$ replaced by $\hat{r}_{\hat{v}}(m)$.

Therefore, the estimated $ARMA(p, q)$ power spectrum is

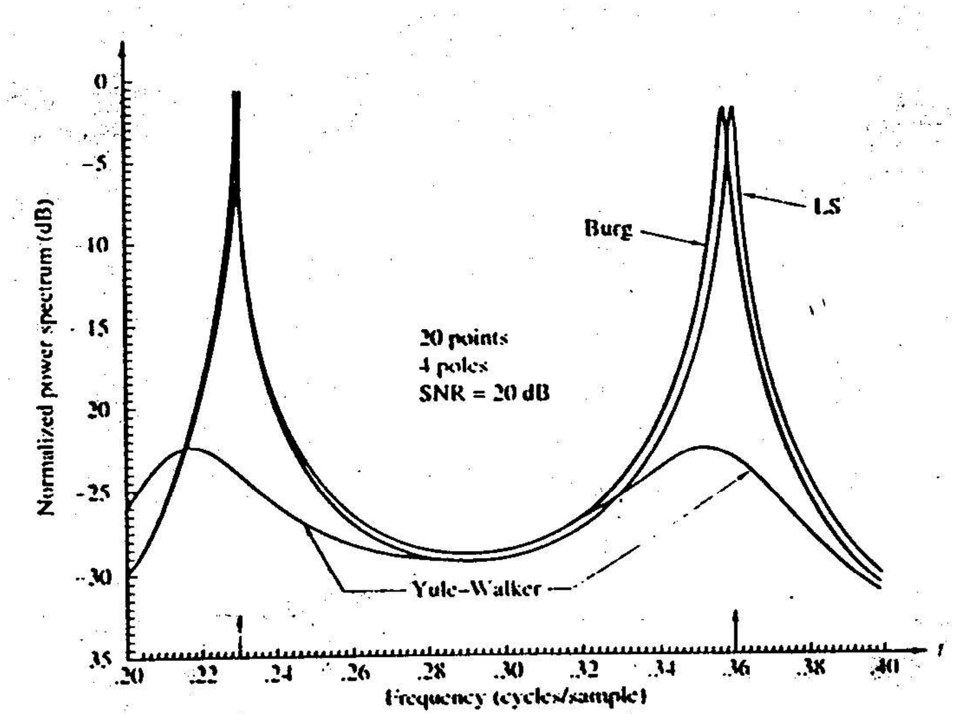
$$R_x^{ARMA}(e^{j2\pi f}) = \frac{R_{\hat{v}}^{MA}(e^{j2\pi f})}{|\hat{A}(e^{j2\pi f})|^2}. \tag{195}$$

Simulation results

We now present some experimental results to demonstrate their performance. Our objective is to compare the spectral estimation algorithms on the basis of their frequency resolution, bias, and their robustness in the presence of additive noise.

The data consists of one or two sinusoids and additive Gaussian noise. The two sinusoids are spaced Δf apart. The SNR is defined as $SNR = 10log_{10}\frac{M^2}{2\sigma_e^2}$, where M is the amplitude of the sinusoid. For the two-sinusoid situations, their amplitudes are set the same. σ_e^2 is the variance of the additive noise $e(n)$.

Case 1: We illustrate the results for $N = 20$ data points based on an $AR(4)$ model with an $SNR = 20dB$ and $\Delta f = f_2 - f_1 = 0.36 - 0.23 = 0.13$.



Comparison of AR power spectrum estimation methods

- *Yule-Walker method:* smooth (broad) spectral estimate, small peaks. Correct no. of frequencies but poor accuracy.
- *Burg and LS:* much better than Yule-Walker. For this example, *LS* slightly superior over Burg method in terms of frequency estimate accuracy.

When $\Delta f = 0.33 - 0.26 = 0.07$, the Yule-Walker method no longer resolves the peaks while the two other methods can still yield a satisfactory performance.

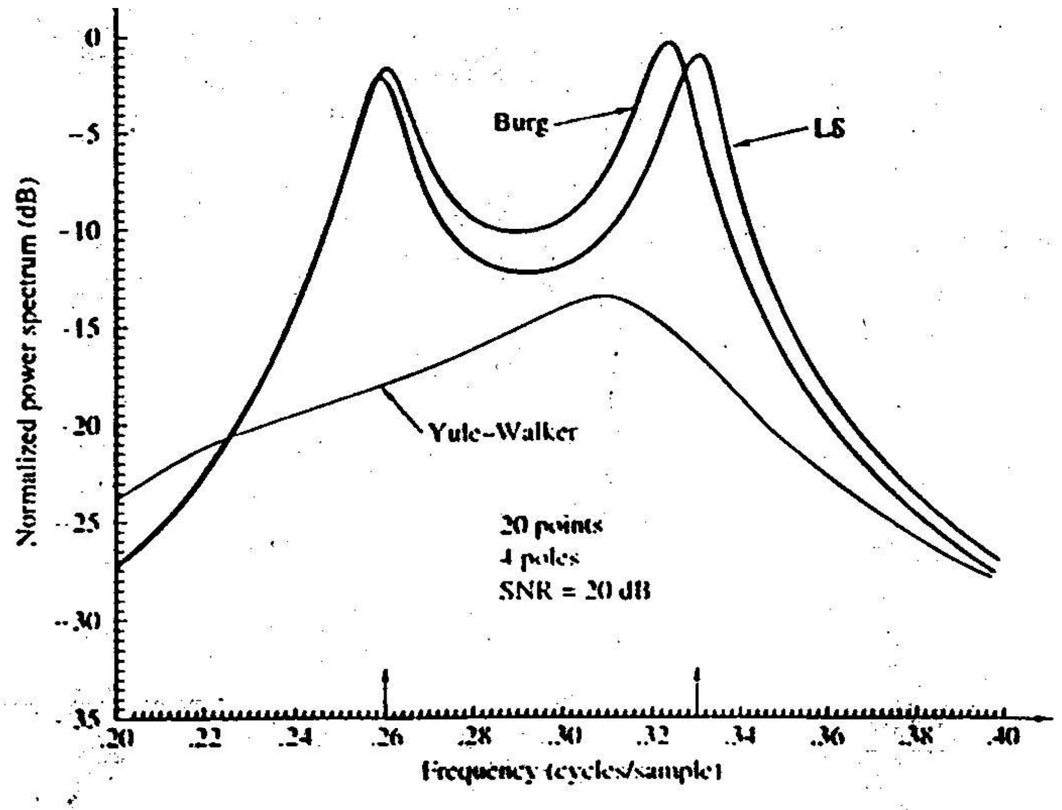
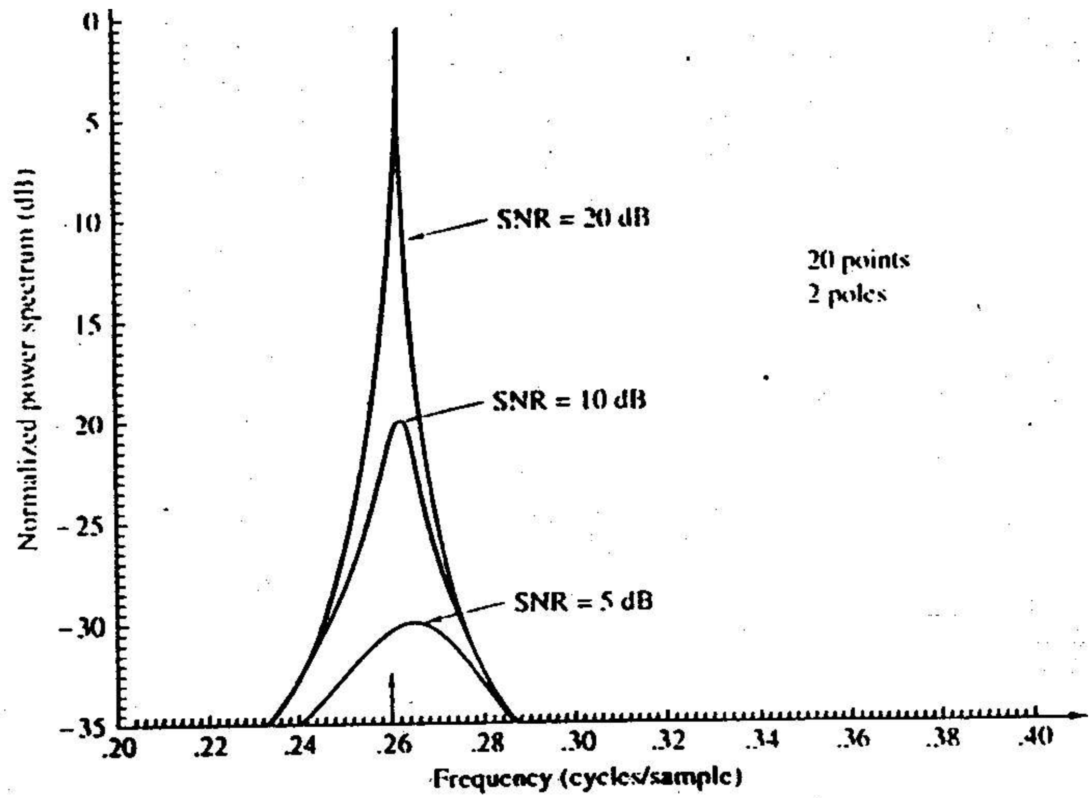


Fig. 5.10: Comparison of AR power spectrum estimation methods

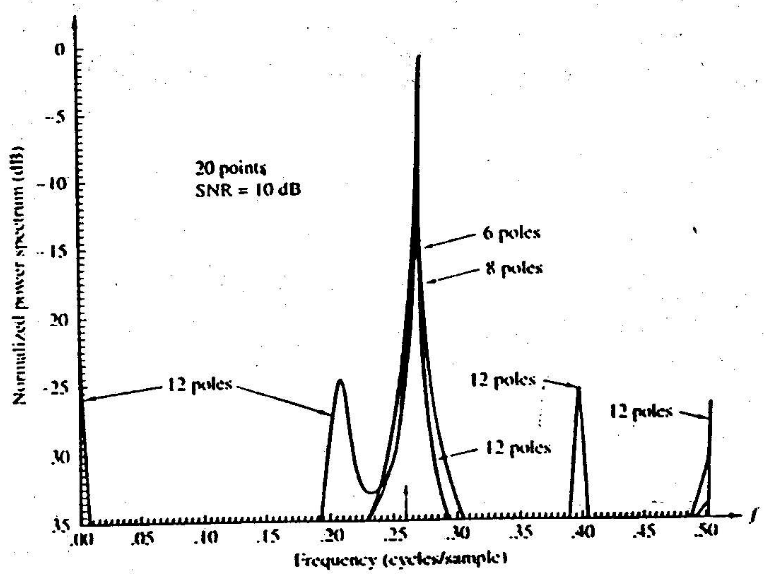
One can conclude the Burg and *LS* methods are clearly superior to the Yule-Walker method for short data records.

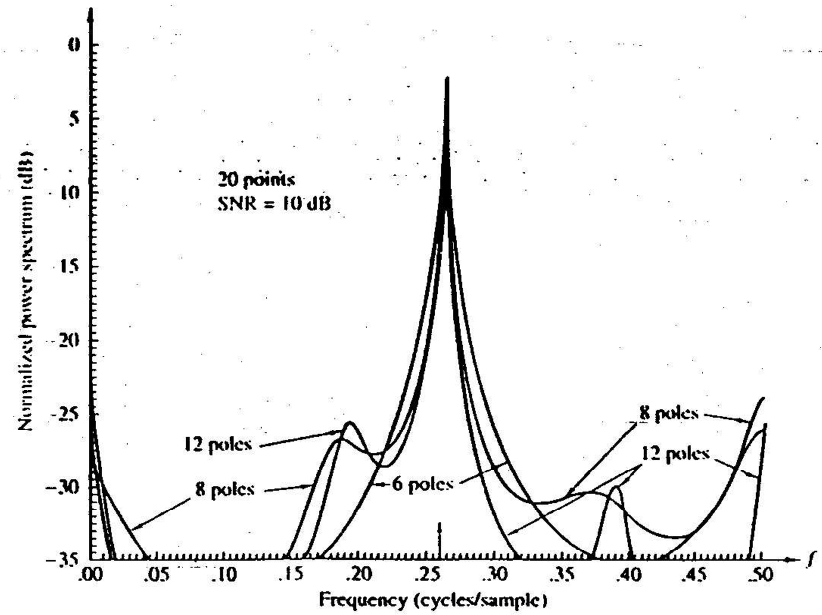
Case 2: In this example, we have one frequency component underlying the process. The frequency is 0.26. The effect of additive noise on the estimate is illustrated for least-squares method.



Effect of additive noise on LS method

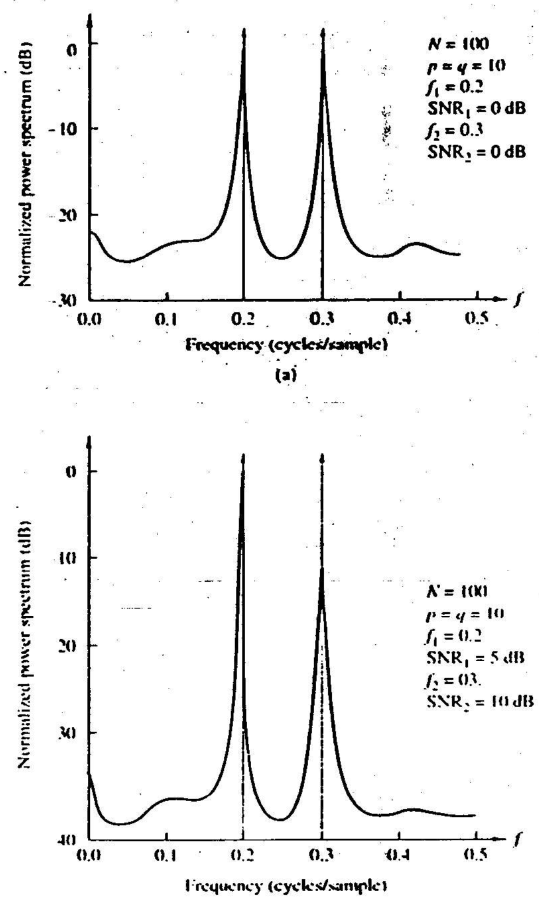
While the effect of filter order on the Burg and least-squares methods are demonstrated





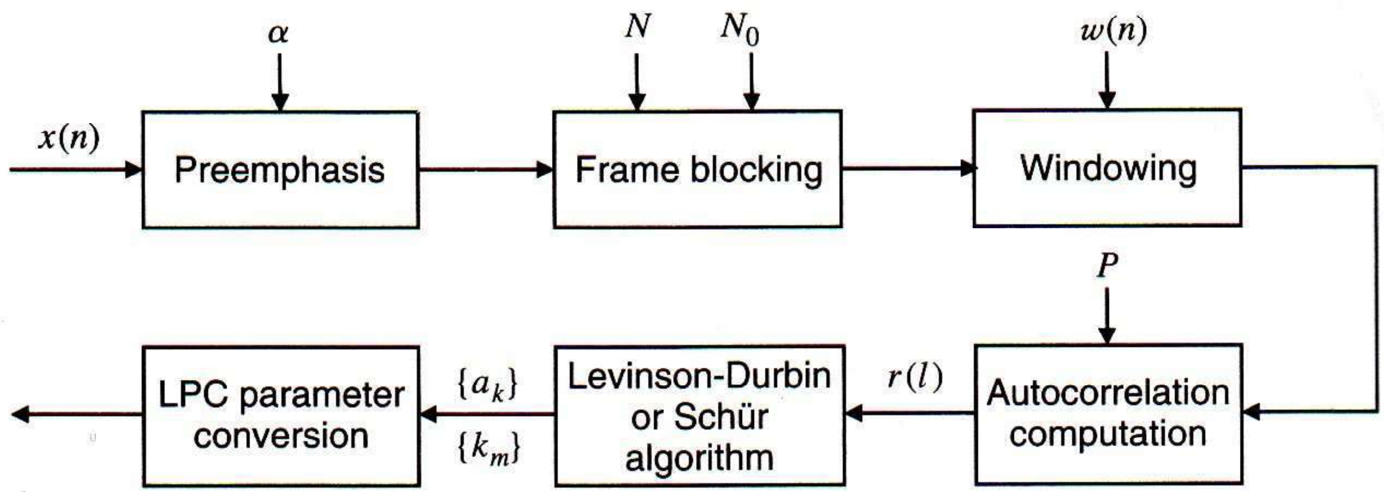
Effect of filter order on: Left - Burg method; right - LS method

Case 3: The following figure shows the spectral estimates for two sinusoids in noise using LS method with $ARMA(10, 10)$. An excellent quality in terms of frequency accuracy is demonstrated for low SNR values.



$ARMA(10, 10)$ power spectrum estimate for two sinusoids in additive noise

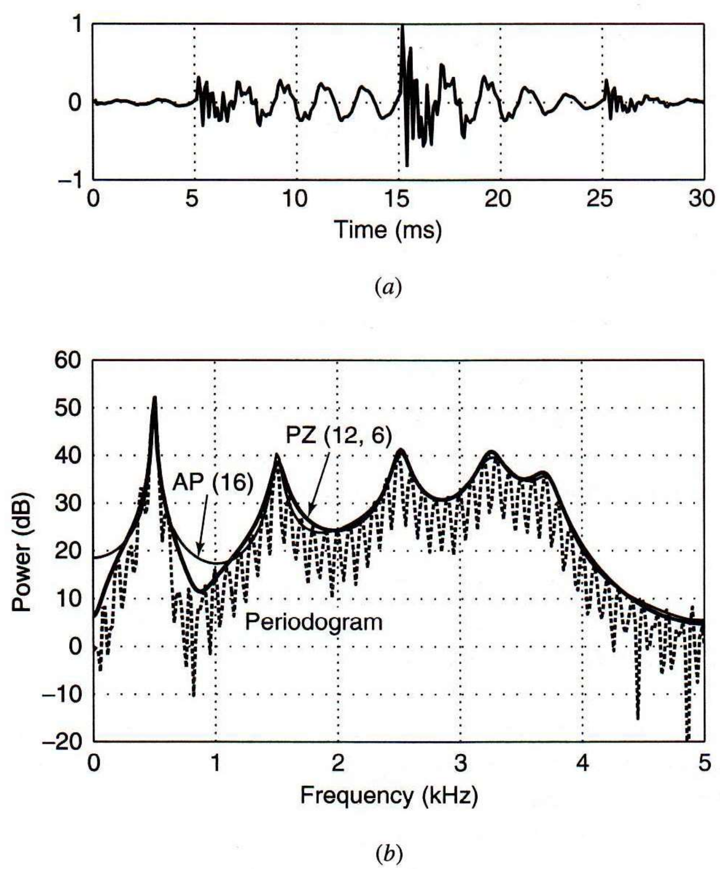
Application in speech signal processing
The following figure shows a *AR*-based speech processing scheme.



Block-diagram of an AP-based speech coding/recognition system.

1) Pre-emphasis: $H_p(z) = 1 - \alpha z^{-1}$ with $0.9 \leq \alpha \leq 1$ to flatten the spectrum; 2) Frame-blocking with an overlapping by $N_0 \approx N/3$.

A portion of speech signal and the PSDs obtained with different algorithms are shown in th following figure.



(a) A segment of speech signal; (b) PSDs estimated with different methods.