

# **SPECTRUM ESTIMATION (2)**

## **PARAMETRIC METHODS FOR POWER SPECTRUM ESTIMATION**

General consideration of parametric model spectrum estimation:

Autoregressive spectrum estimation:

- A. The autocorrelation method*
- B. The covariance method*
- C. Modified covariance method*
- D. Burg algorithm:*
- E. Selection of the model order:*

Moving average spectrum estimation:

Autoregressive moving average spectrum estimation:

**Minimum Variance Spectrum Estimation (a nonparametric method):**

- A. The relation of the variance with the power spectrum*
- B. FIR bandpass filter bank and the variances of the filters' outputs:*
- C. The FIR bandpass filters with minimum variance*
- D. The minimum variance spectral estimate*

**Maximum entropy method (an AR parametric method):**

- A. The concept of entropy*
- B. Extrapolation of the autocorrelation*
- C. The maximum entropy spectral estimate*

**Summary of spectral estimation methods:**

**Some useful MATLAB functions**

## PARAMETRIC METHODS FOR POWER SPECTRUM ESTIMATION

Parametric methods for power spectrum estimation are based on parametric models, and they include those of the autoregressive (AR) spectral estimation, the moving average (MA) spectral estimation, and the autoregressive moving average (ARMA) spectral estimation, which are, respectively, based on the AR, MA, and ARMA models. The maximum entropy method is of the same form of the AR spectral estimation.

In parametric methods, a parametric model for a random process is first selected and then the model parameters are determined.

The parametric spectral estimators are less biased and have a lower variance than the nonparametric spectral estimators. With parametric methods it is possible to significantly improve the resolution of the spectrum estimation unless the model used is consistent with the random process being analyzed. Otherwise inaccurate or misleading spectrum estimates may result.

### General consideration of parametric model spectrum estimation:

As we have learnt in the previous lectures concerning Signal Modeling, a random process can be modeled with an ARMA model, or an AR model, or an MA model. The AR and MA models are the special cases of the ARMA model. Supposing that a random process  $x(n)$  is modeled as an ARMA( $p, q$ ) process with an ARMA( $p, q$ ) model, then the system function of the model is

$$H(e^{j\omega}) = \frac{\sum_{k=0}^q b_q(k)e^{-jk\omega}}{1 + \sum_{k=1}^p a_p(k)e^{-jk\omega}} \quad (43)$$

In this case, the power spectrum of the process  $x(n)$  can be computed in the following manner,

$$P_x(e^{j\omega}) = \frac{\left| \sum_{k=0}^q b_q(k)e^{-jk\omega} \right|^2}{\left| 1 + \sum_{k=1}^p a_p(k)e^{-jk\omega} \right|^2} \quad (44)$$

Alternatively, if the autocorrelation  $r_x(k)$  is given, the power spectrum can be obtained from the Fourier transform of  $r_x(k)$ ,

$$P_x(e^{j\omega}) = \sum_{k=-\infty}^{\infty} r_x(k)e^{-jk\omega} \quad (45)$$

Eqs. (44) and (45) demonstrate two approaches to computing the power spectrum of an ARMA process, and also reveal two equivalent representations of an ARMA random process, that is, the process can be represented equivalently either by a finite sequence of model parameters  $a_p(k)$  and  $b_q(k)$ , or by an autocorrelation sequence  $r_x(k)$ . The equivalence of the two representations is because the autocorrelation and the model parameters are related with the Yule-Walker equations,

$$r_x(k) + \sum_{l=1}^p a_p(l)r_x(k-l) = \sigma_v^2 \sum_{l=k}^q b_q(l)h^*(l-k) \quad (46)$$

In practice, a random process  $x(n)$  is often given only over a finite interval,  $0 \leq n \leq N-1$ , and in this case the autocorrelation of  $x(n)$  must be estimated in a finite sum as follows,

$$\hat{r}_x(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x(n+k)x^*(n), \quad k=0, 1, \dots, N-1. \quad (47)$$

When the ARMA model in Eq. (43) is selected for modeling process  $x(n)$ , the model parameters in this case are determined from this estimated autocorrelation sequence  $\hat{r}_x(k)$ , and they are different from  $a_p(k)$  and  $b_q(k)$  determined from  $r_x(k)$  since  $\hat{r}_x(k)$  is, in general, not equal to  $r_x(k)$ . Such model parameters that are determined from  $\hat{r}_x(k)$  are denoted by  $\hat{a}_p(k)$  and  $\hat{b}_q(k)$ , which give an estimate of the power spectrum,

$$\hat{P}_x(e^{j\omega}) = \frac{\left| \sum_{k=0}^q \hat{b}_q(k) e^{-jk\omega} \right|^2}{\left| 1 + \sum_{k=1}^p \hat{a}_p(k) e^{-jk\omega} \right|^2} \quad (48)$$

Eq. (48) is a general case of the parametric spectral estimation methods. In this case, all we need to do for estimating the power spectrum is to find  $\hat{a}_p(k)$  and  $\hat{b}_q(k)$ . When  $\hat{a}_p(k)$  and  $\hat{b}_q(k)$  are determined,  $\hat{P}_x(e^{j\omega})$  is found.

Among these parametric spectral estimations, the AR estimation is the most popular. This is because the AR parameters can be found by solving a set of linear equations. Whereas, for the ARMA and MA parameters, a set of nonlinear equations need to be solved, which will be much more difficult.

#### Autoregressive spectrum estimation:

The autoregressive spectrum estimation is based on the AR model. In this case, a random process  $x(n)$  is modeled as an AR( $p$ ) process. If the autocorrelation  $r_x(k)$  of a random process  $x(n)$  is given, the AR parameters,  $a_p(k)$  and  $b(0)$ , can be determined from  $r_x(k)$  using the AR model. Then the power spectrum of the AR process is

$$P_x(e^{j\omega}) = \frac{|b(0)|^2}{\left| 1 + \sum_{k=1}^p a_p(k) e^{-jk\omega} \right|^2} \quad (49)$$

If a random process  $x(n)$  is given over a finite interval  $0 \leq n \leq N-1$ , the autocorrelation of  $x(n)$  must be estimated, and it is denoted  $\hat{r}_x(k)$ . The AR parameters that are determined from the estimated autocorrelation  $\hat{r}_x(k)$  are defined as  $\hat{a}_p(k)$  and  $\hat{b}(0)$ . The power spectrum that is estimated based on  $x(n)$  for  $0 \leq n \leq N-1$  is

$$\hat{P}_x(e^{j\omega}) = \frac{|\hat{b}(0)|^2}{\left| 1 + \sum_{k=1}^p \hat{a}_p(k) e^{-jk\omega} \right|^2} \quad (50)$$

Several approaches are available for finding  $\hat{r}_x(k)$  from the finite data record of a process  $x(n)$  for  $0 \leq n \leq N-1$ , such as autocorrelation method, covariance method, and modified covariance method, which are presented below. When  $\hat{r}_x(k)$  are found,  $\hat{a}_p(k)$  and  $\hat{b}(0)$  can be found using the MYWE method or the EYWE method, which are studied extensively in the lectures on Signal Modeling.

#### A. The autocorrelation method

The AR parameters  $\hat{a}_p(k)$  are found by solving the autocorrelation normal equation

$$\begin{bmatrix} \hat{r}_x(0) & \hat{r}_x^*(1) & \dots & \hat{r}_x^*(p-1) \\ \hat{r}_x(1) & \hat{r}_x(0) & \dots & \hat{r}_x^*(p-2) \\ \dots & \dots & \dots & \dots \\ \hat{r}_x(p-1) & \hat{r}_x(p-2) & \dots & \hat{r}_x(0) \end{bmatrix} \begin{bmatrix} \hat{a}_p(1) \\ \hat{a}_p(2) \\ \dots \\ \hat{a}_p(p) \end{bmatrix} = - \begin{bmatrix} \hat{r}_x(1) \\ \hat{r}_x(2) \\ \dots \\ \hat{r}_x(p) \end{bmatrix} \quad (51)$$

where the autocorrelation estimate is given by

$$\hat{r}_x(k) = \frac{1}{N} \sum_{n=0}^{N-1-k} x(n+k)x^*(n) ; k=0, 1, \dots, p. \quad (52)$$

$$\left| \hat{b}(0) \right|^2 = \hat{r}_x(0) + \sum_{k=1}^p \hat{a}_p(k) \hat{r}_x^*(k). \quad (53)$$

Substituting  $\hat{a}_p(k)$  and  $\hat{b}(0)$  into Eq. (50) gives the estimate of the power spectrum of the process  $x(n)$ .

Note that Eq. (51) is the same in form as the modified Yule-Walker equations, but the autocorrelation values in Eq. (51) are the estimated ones,  $\hat{r}_x(k)$ , from a finite data record, i.e.,  $x(n)$  for  $0 \leq n \leq N-1$ . The autocorrelation estimate  $\hat{r}_x(k)$  is biased. This method that estimates the power spectrum using the autocorrelation method is also referred to as the *Yule-Walker method* (YWM).

Since Eq. (52) gives a biased estimate, a variation of Eq. (52) is of the following form

$$\hat{r}_x(k) = \frac{1}{N-k} \sum_{n=0}^{N-1-k} x(n+k)x^*(n) ; k=0, 1, \dots, p. \quad (52')$$

which may give an unbiased estimate. However, this can not guarantee the autocorrelation matrix to be positive definite and consequently, the variance of  $\hat{P}_x(e^{j\omega})$  tends to be large when the matrix is ill-conditioned or singular. Thus, the biased estimate in Eq. (52) is more preferable to the unbiased estimate in Eq. (52').

Like in a periodogram, the autocorrelation method uses the windowed data to estimate the autocorrelation and thus has window effect on the spectral estimate. Since the window effect will become severe for short data records, the autocorrelation method is not often used in the case of short data records. An artifact, called spectral line splitting that means the splitting of a single spectral peak into two or more separate and distinct peaks, may appear in the autocorrelation method when  $x(n)$  is overmodeled, i.e., when the model order  $p$  is too large, because a pole in the model, in general, may create a spectral peak. An example of such an artifact is shown in Fig. 8.24.

**Fig. 8.24.** Spectral line splitting of an AR(2) process,  $x(n) = -0.9x(n-2) + w(n)$ . Two all-pole spectrum estimates were computed using the autocorrelation method with order  $p = 4$  (solid line) and  $p = 12$  (dashed-dot line)

### B. The covariance method

In the covariance method the AR parameters  $\hat{a}_p(k)$  are determined by solving the normal equations,

$$\begin{bmatrix} \hat{r}_x(1,1) & \hat{r}_x(2,1) & \dots & \hat{r}_x(p,1) \\ \hat{r}_x(1,2) & \hat{r}_x(2,2) & \dots & \hat{r}_x(p,2) \\ \dots & \dots & \dots & \dots \\ \hat{r}_x(1,p) & \hat{r}_x(2,p) & \dots & \hat{r}_x(p,p) \end{bmatrix} \begin{bmatrix} \hat{a}_p(1) \\ \hat{a}_p(2) \\ \dots \\ \hat{a}_p(p) \end{bmatrix} = - \begin{bmatrix} \hat{r}_x(0,1) \\ \hat{r}_x(0,2) \\ \dots \\ \hat{r}_x(0,p) \end{bmatrix} \quad (54)$$

where the autocorrelation estimate is given by

$$\hat{r}_x(k,l) = \sum_{n=p}^{N-1} x(n-l)x^*(n-k), \quad (55)$$

which is different from the autocorrelation method in that no windowing of the data is required since the values of  $x(n)$  used for finding  $\hat{r}_x(k,l)$  in Eq. (55) are all in the interval  $0 \leq n \leq N-1$  and thus no zero-padding is needed. This means that there is no windowing effect in the variance method. Therefore, for short data records the variance method generally gives higher resolution spectrum estimates than the autocorrelation method.

### C. Modified covariance method

In the modified covariance method the AR parameters  $\hat{a}_p(k)$  are also determined by solving the normal equations in Eq. (54). But the autocorrelation estimate is found in a different way as follows

$$\hat{r}_x(k,l) = \sum_{n=p}^{N-1} [x(n-l)x^*(n-k) + x(n-p+l)x^*(n-p+k)] \quad (56)$$

which is derived by minimizing the sum of the squares of the forward and backward prediction errors, that is,

$$\mathcal{E}_p(n) = \mathcal{E}_p^+(n) + \mathcal{E}_p^-(n) = \sum_{n=p}^{N-1} \left[ |e_p^+(n)|^2 + |e_p^-(n)|^2 \right] \quad (57)$$

where

$$e_p^+(n) = x(n) + \sum_{k=1}^p a_p(k)x(n-k) \quad (58)$$

$$e_p^-(n) = x(n-p) + \sum_{k=1}^p a_p^*(k)x(n-p+k) \quad (59)$$

are the forward and backward prediction errors, respectively (see pp. 322-324 in the Hayes' book). In contrast to the autocorrelation and covariance methods, the modified covariance method is observed to give statistically stable spectrum estimates with high resolution, and not to be subject to spectral line splitting.

### Example 5. Estimation of the power spectrum of an AR(4) process.

Consider the AR(4) process generated by the difference equation

$$x(n) = 2.7377x(n-1) - 3.7476x(n-2) + 2.6293x(n-3) - 0.9224x(n-4) + w(n)$$

where  $w(n)$  is unit variance white Gaussian noise. The filter generating  $x(n)$  has a pair of poles at  $z = 0.98e^{\pm j0.2\pi}$  and a pair of poles at  $z = 0.98e^{\pm j0.3\pi}$ . Using the data records of length  $N = 128$ , an ensemble of 50 spectrum estimates were calculated using the Yule-Walker method, the covariance method, and the modified covariance method, and the Burg's method. The overlay plots of the 50 estimates from the four methods are shown in part (a) in 8.25 to 8.28, and the ensemble average of the 50 estimates and the true power spectrum are shown in part (b) in 8.25 to 8.28.

**Figs. 8. 25 to 8.28.**

*E. Selection of the model order:*

The selection of the model order in the AR spectral estimation is critical in the parametric methods. Fig. 8.24 shows that spectral line splitting artifact appears in the autocorrelation method when the model order  $p$  selected is too large, because each of the poles in the model, in general, may create one spectral peak. This brings up the question of how to select an appropriate model order  $p$  for an AR spectrum estimation. If the model order  $p$  used is too small, then the resulting spectrum will be smoothed because the spectral peaks in a true power spectrum can not be represented with an enough number of poles. If, on the other hand, the model order  $p$  used is too large, then the power spectrum may contain more spectral peaks than those in a true power spectrum, and in this case the so-called spectral line splitting artifact is present, as the example in Fig. 8.24. This reveals a necessity to have some approaches to selecting an appropriate model order that gives the parametric model a best fit of a given data record. A relevant idea would be to adjust the model order until a certain modeling error becomes minimum. There are several approaches to selecting model order that were established based on such an idea.

One of them is the Akaike Information Criterion expressed as

$$\text{AIC}(p) = N \log \varepsilon_p + 2p \quad (60)$$

and the other is the minimum description length

$$\text{MDL}(p) = N \log \varepsilon_p + (\log N) p \quad (61)$$

Two other often used criteria are Akaike's Final Prediction Error

$$\text{FPE}(p) = \varepsilon_p \frac{N + p + 1}{N - p - 1} \quad (62)$$

and Parzen's Criterion Autoregressive Transfer function

$$\text{CAT}(p) = \left[ \frac{1}{N} \sum_{k=1}^p \frac{N - k}{N \varepsilon_k} \right] + \frac{N - p}{N \varepsilon_p} \quad (63)$$

In Eqs. (60) to (63),  $p$  is the model order to be selected,  $N$  is the length of the data record, and  $\varepsilon_p$  is the prediction error. For short data records, none of the criteria tend to work particularly well, and thus, these criteria should only be used as indicators of the model order. Eqs. (60) to (63) show that all these criteria depend on  $\varepsilon_p$ . Since different modeling techniques, e.g., the autocorrelation and covariance methods, may have different prediction errors, then the model order may be different even for the same data.

Moving average spectrum estimation:

The moving average spectrum estimation is carried out based on the MA model that models a random process  $x(n)$  as an MA process. The system function of the filter that generates the MA process by filtering unit variance white noise is of the form,

$$H(z) = \sum_{k=0}^q b_q(k) z^{-k} . \quad (64)$$

Thus, the power spectrum of the process  $x(n)$  is

$$P_x(e^{j\omega}) = \left| \sum_{k=0}^q b_q(k) e^{-jk\omega} \right|^2 . \quad (65)$$

In terms of the autocorrelation  $r_x(k)$  the power spectrum can be written as

$$P_x(e^{j\omega}) = \sum_{k=-q}^q r_x(k) e^{-jk\omega} . \quad (66)$$

where  $r_x(k)$  and  $b_q(k)$  are related with the Yule-Walker equations

$$r_x(k) = \sigma_v^2 b_q(k) * b_q^*(-k) = \sum_{l=0}^{q-|k|} b_q(l+|k|) b_q^*(l) \quad (67)$$

If a random process  $x(n)$  is given over a finite interval  $0 \leq n \leq N-1$ , we only have the estimate of the autocorrelation of  $x(n)$ ,  $\hat{r}_x(k)$ . The MA parameters that are determined from the estimated autocorrelation  $\hat{r}_x(k)$  are  $\hat{b}_q(k)$ . The estimate of the power spectrum of  $x(n)$  is

$$\hat{P}_{MA}(e^{j\omega}) = \left| \sum_{k=0}^q \hat{b}_q(k) e^{-jk\omega} \right|^2 . \quad (68)$$

Equivalently, directly using the autocorrelation estimate  $\hat{r}_x(k)$ , we may have the alternative form of the estimate of the power spectrum,

$$\hat{P}_{MA}(e^{j\omega}) = \sum_{k=-q}^q \hat{r}_x(k) e^{-jk\omega} . \quad (69)$$

Comparing the estimate in Eq. (69) with the Black-Tukey estimate in Eq. (41), we can see that the MA estimate is equivalent to the Black-Tukey estimate if the window  $w(n)$  used extends from  $-q$  to  $q$ . However, there is a subtle difference between the two estimates; for the MA spectral estimate in Eq. (69) the random process  $x(n)$  is modeled as an MA process of order  $q$ , and thus the autocorrelation sequence is zero for  $|k| > q$ . In this case, if the autocorrelation estimate  $\hat{r}_x(k)$  is unbiased for  $|k| \leq q$ , then  $E\{\hat{P}_{MA}(e^{j\omega})\} = P_x(e^{j\omega})$  so that  $\hat{P}_{MA}(e^{j\omega})$  is unbiased.

In the Blackman-Tukey method, no assumption is made about  $x(n)$ , and due to the windowing effect, thus, the Blackman-Tukey spectral estimate will be biased unless  $x(n)$  is an MA process.

Autoregressive moving average spectrum estimation:

The ARMA spectrum estimation is performed based on the ARMA model that models a random process  $x(n)$  as an ARMA process. This method has been dealt with in the earlier section, General consideration of parametric model spectrum estimation.

Minimum Variance Spectrum Estimation (a nonparametric method):

In the minimum variance (MV) method the power spectrum is estimated by filtering a random process with a bank of narrowband bandpass filters. The bandpass filters are designed to be optimum by minimizing the variance of the output of a narrowband filter that adapts to the spectral content of the input process at each frequency of interest.

*A. The relation of the variance with the power spectrum*

Consider a zero mean WSS process  $y(n)$ . The variance of  $y(n)$  is

$$\sigma_y^2(n) = E\{y(n)|^2\}, \quad (70)$$

which is the power of the process  $y(n)$ . For a given autocorrelation  $r_y(k)$  we have

$$r_y(0) = E\{y(n)|^2\}, \quad (71)$$

and then  $\sigma_y^2(n) = r_y(0)$  does not vary with  $n$ , and thus denote

$$\sigma_y^2 = E\{y(n)|^2\}. \quad (72)$$

Since  $r_y(0) = (1/2\pi) \int_{-\pi}^{\pi} P_y(e^{j\omega}) d\omega$ , the average power of such a WSS process  $y(n)$  is

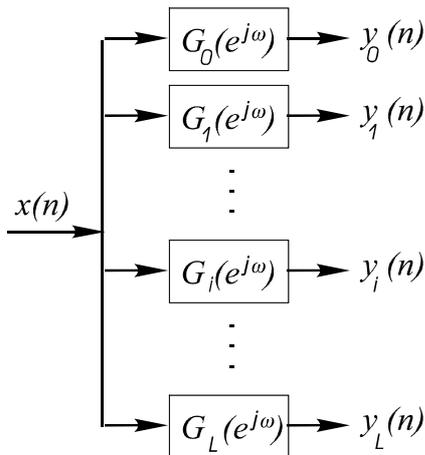
$$E\{y(n)|^2\} = \sigma_y^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_y(e^{j\omega}) d\omega, \quad (73)$$

which shows the relation of the variance with the power spectrum.

*B. FIR bandpass filter bank and the variances of the filters' outputs:*

Consider a bank of FIR bandpass filters (Fig. 3), all having order  $p$  and the frequency responses (or the system function) of the following form,

$$G_i(e^{j\omega}) = \sum_{n=0}^p g_i(n)e^{-jn\omega}, \quad i=0, 1, \dots, L. \quad (74)$$



**Fig. 3.** A bank of bandpass filters in the minimum variance spectrum estimation

The input to the filters is  $x(n)$ , and the outputs of the bandpass filters are  $y_i(n)$  for  $i=0, 1, \dots, L$ . To use such a filter bank to estimate the power spectrum of  $x(n)$  with a finite-length data record, we should constrain all bandpass filters that, at their center frequencies  $\omega_i$ , have a unit gain,

$$G_i(e^{j\omega_i}) = \sum_{n=0}^p g_i(n)e^{-jn\omega_i} = 1 \quad (75)$$

so that the power spectra of the filters' outputs  $y_i(n)$  are  $P_{y_i}(e^{j\omega_i}) = |G_i(e^{j\omega_i})|^2 P_x(e^{j\omega_i}) = P_x(e^{j\omega_i})$ .

Using the vector notations,

$$\mathbf{g}_i = [g_i(0), g_i(1), \dots, g_i(p)]^T \quad (76)$$

and

$$\mathbf{e}_i = [1, e^{j\omega_i}, \dots, e^{jp\omega_i}]^T \quad (77)$$

Eq. (75) can be written as

$$G_i(e^{j\omega_i}) = \mathbf{e}_i^H \mathbf{g}_i = 1 \quad (78a)$$

or equivalently as

$$[G_i(e^{j\omega_i})]^* = \mathbf{g}_i^H \mathbf{e}_i = 1 \quad (78b)$$

Since the autocorrelations of the output processes  $y_i(n)$  and the input process  $x(n)$  are related in the following manner (see Eq. (91) in *DISCRETE-TIME RANDOM PROCESS (3)*)

$$r_{y_i}(k) = r_x(k) * g_i(k) * g_i^*(-k) = \sum_{l=0}^p \sum_{m=0}^p g_i(l)r_x(m-l+k)g_i^*(m), \quad (79)$$

then the variance of the output process  $y_i(n)$ , which is equal to  $r_{y_i}(0)$ , is of the form

$$\sigma_{y_i}^2 = r_{y_i}(0) = \sum_{l=0}^p \sum_{m=0}^p g_i(l)r_x(m-l)g_i^*(m), \quad (80)$$

and its matrix form will be

$$\sigma_{y_i}^2 = \mathbf{g}_i^H \mathbf{R}_x \mathbf{g}_i, \quad (81)$$

which shows the relation of the variance of the output of the  $i$ th filter with the filter coefficients  $g_i(n)$ , for a given random process with the autocorrelation matrix  $\mathbf{R}_x$ .

### C. The FIR bandpass filters with minimum variance

Designing a filter is just determining the filter coefficients based on a certain criterion. The criterion that we use here is the minimum variance of  $y_i(n)$ , which is obtained by minimizing  $\sigma_{y_i}^2$  in Eq. (81) under the constraint given by Eq. (78). The approach to this constrained minimization problem is given in Section 2.3.10 in the Hayes' textbook. Using this approach, the coefficients of the optimum filter in terms of minimum variance are found as follows

$$\mathbf{g}_i = \frac{\mathbf{R}_x^{-1} \mathbf{e}_i}{\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i} \quad (82)$$

which obviously satisfies  $\mathbf{e}_i^H \mathbf{g}_i = 1$ , and inserting Eq. (82) into Eq. (81) the minimum variance is, thus, determined as follows

$$\min\{\sigma_{y_i}^2\} = \mathbf{g}_i^H \mathbf{R}_x \mathbf{g}_i = \frac{1}{\mathbf{e}_i^H \mathbf{R}_x^{-1} \mathbf{e}_i} \quad (83)$$

which is the best estimate of the variance of the process  $x(n)$  at frequency  $\omega_i$  in terms of minimum-variance, that is,  $\hat{\sigma}_x^2(\omega_i) = \min\{\sigma_{y_i}^2\}$ . Since Eqs. (82) and (83) are derived at an arbitrary frequency  $\omega_i$ , then these two equations hold for all  $\omega$ . Therefore, the optimum filter and the estimate of the variance of  $x(n)$  can be written, respectively, as

$$\mathbf{g} = \frac{\mathbf{R}_x^{-1} \mathbf{e}}{\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e}} \quad (84)$$

and

$$\hat{\sigma}_x^2(\omega) = \frac{1}{\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e}} \quad (85)$$

which is frequency dependent, and where

$$\mathbf{R}_x = E\{\mathbf{xx}^H\} = \begin{bmatrix} r_x(0) & r_x^*(1) & \dots & r_x^*(p) \\ r_x(1) & r_x(0) & \dots & r_x^*(p-1) \\ r_x(2) & r_x(1) & \dots & r_x(p-2) \\ \dots & \dots & \dots & \dots \\ r_x(p) & r_x(p-1) & \dots & r_x(0) \end{bmatrix} \quad (86)$$

$$\mathbf{g} = [g(0), g(1), \dots, g(p)]^T \quad (87)$$

and

$$\mathbf{e} = [1, e^{j\omega}, \dots, e^{jp\omega}]^T. \quad (88)$$

Till now we have found the variance estimate of the process  $x(n)$  but not the power spectrum estimate yet.

#### D. The minimum variance spectral estimate

To find the power spectrum estimate, let us look at the bandpass filter bank again. Since the bandpass filters are narrowband and the bandwidth of the  $i$ th filter  $G_i(e^{j\omega})$  is assumed to be  $\Delta$ , then in the bandwidth  $\Delta$ , that is,  $\omega_i - \Delta/2 \leq \omega \leq \omega_i + \Delta/2$ , we may assume  $|G_i(e^{j\omega})| \approx 1$  (due to the given constraint in Eq. (75)), and out of the bandwidth  $\Delta$ ,  $|G_i(e^{j\omega})| \approx 0$ . In this case, the relation of the variance of  $y_i(n)$  with the power spectrum of  $x(n)$  in Eq. (73) becomes

$$\sigma_{y_i}^2 = \frac{1}{2\pi} \int_{-\pi}^{\pi} P_{y_i}(e^{j\omega}) d\omega = \frac{1}{2\pi} \int_{-\pi}^{\pi} |G_i(e^{j\omega})|^2 P_x(e^{j\omega}) d\omega \approx \frac{1}{2\pi} \int_{\omega_i - \Delta/2}^{\omega_i + \Delta/2} P_x(e^{j\omega}) d\omega \approx \frac{\Delta}{2\pi} P_x(e^{j\omega_i}) \quad (89)$$

Since the estimate of the variance of  $x(n)$  is equal to the minimum variance of  $y_i(n)$ , that is,  $\hat{\sigma}_x^2 = \min\{\sigma_{y_i}^2\}$ ,

then the (best) estimate of the power spectrum of  $x(n)$  in terms of minimum variance can be expressed, from Eq. (89), as

$$\hat{P}_x(e^{j\omega}) = \frac{2\pi}{\Delta} \min\{\sigma_{y_i}^2\} = \frac{2\pi}{\Delta} \hat{\sigma}_x^2(\omega) = \frac{2\pi}{\Delta} \frac{1}{\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e}}. \quad (90)$$

In Eq. (90) the bandwidth  $\Delta$  is still unknown. To find  $\Delta$ , we consider estimating the power spectrum of a white noise process with a zero mean and a variance of  $\sigma_x^2$ . The autocorrelation matrix of the white noise is

$\mathbf{R}_x = \sigma_x^2 \mathbf{I}$ , and then the bandpass filters with minimum variances are

$$\mathbf{g} = \frac{\mathbf{R}_x^{-1} \mathbf{e}}{\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e}} = \frac{\sigma_x^{-2} \mathbf{I} \mathbf{e}}{\mathbf{e}^H \sigma_x^{-2} \mathbf{I} \mathbf{e}} = \frac{\sigma_x^{-2} \mathbf{e}}{\sigma_x^{-2} \mathbf{e}^H \mathbf{e}} = \frac{1}{p+1} \mathbf{e}, \quad (91)$$

which is frequency dependent, and the estimate of the variance of  $x(n)$  is

$$\hat{\sigma}_x^2(\omega) = \frac{1}{\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e}} = \frac{1}{\mathbf{e}^H \sigma_x^{-2} \mathbf{I} \mathbf{e}} = \frac{\sigma_x^2}{\mathbf{e}^H \mathbf{e}} = \frac{1}{p+1} \sigma_x^2 \quad (92)$$

which is independent of frequency. Substituting Eq. (92) into Eq. (90) yields the minimum variance estimate of the power spectrum,

$$\hat{P}_x(e^{j\omega}) = \frac{2\pi}{\Delta} \hat{\sigma}_x^2(\omega) = \frac{2\pi}{\Delta} \frac{1}{p+1} \sigma_x^2. \quad (93)$$

Since the power spectrum of a white noise process is equal to its variance  $\sigma_x^2$ , then we set the power spectrum of white noise to be equal to its estimate, i.e.,

$$\hat{P}_x(e^{j\omega}) = P_x(e^{j\omega}) = \sigma_x^2, \quad (94)$$

which is actually *an assumption we impose* here. The bandwidth  $\Delta$  can, thus, be determined,

$$\Delta = \frac{2\pi}{p+1}. \quad (95)$$

For a general WSS random process  $x(n)$ , we adopt this bandwidth  $\Delta$  for the white noise case, and then the estimate of the power spectrum of the general process  $x(n)$  in Eq. (90) becomes

$$\hat{P}_x(e^{j\omega}) = \hat{P}_{MV}(e^{j\omega}) = \frac{p+1}{\mathbf{e}^H \mathbf{R}_x^{-1} \mathbf{e}}. \quad (96)$$

Usually  $\mathbf{R}_x$  is unknown, and then  $\mathbf{R}_x$  may be replaced with an estimate,  $\hat{\mathbf{R}}_x$ .

Since the optimum bandpass filters are established based on the autocorrelation  $\hat{\mathbf{R}}_x$  whose values are determined from the data, then the minimum variance spectrum estimation may be thought of as a data-adaptive modification to the periodogram. Generally the minimum variance spectrum estimation offers higher resolution than the periodogram and Blackman-Tukey methods.

It should be noted that, although the MV method is established using a bank of filters, the MV spectral estimate does not need to use the filters in the end. One of the reasons is that the bandwidth found for white noise in Eq. (95) is adopted to a general WSS random process so that the MV spectral estimate is independent of the filters whose bandwidth should be determined from the general random process. Since in the MV spectral estimate no filter model and thus no model parameters need to be found and used, then the MV spectral estimation falls into the category of the nonparametric methods.

Note that the inverse transform of the MV estimate does not match the autocorrelation sequence that is used to create the MV estimate, unlike the autoregressive spectrum estimate that does match.

**Example 5. MV estimate of the power spectrum of an AR(1) process** (Example. 8.3.2)

### Maximum entropy method (an AR method):

The maximum entropy spectral estimation is established based on an explicit extrapolation of a finite length sequence of a known autocorrelation of a random process  $x(n)$ . The extrapolation should be chosen so that the random process characterized by the extrapolated autocorrelation sequence has maximum entropy. The random process treated here is assumed to be Gaussian so that the concerned problem becomes solvable.

#### A. The concept of entropy

Entropy is a measure of randomness or uncertainty. For a Gaussian random process  $x(n)$  with power spectrum  $P_x(e^{j\omega})$ , the entropy of the random variable  $x(n)$  is expressed by

$$H(x) = \frac{1}{2\pi} \int_{-\pi}^{\pi} \ln P_x(e^{j\omega}) d\omega \quad (97)$$

#### B. Extrapolation of the autocorrelation

Given the autocorrelation  $r_x(k)$  of a WSS process for  $|k| \leq p$ , we want to extrapolate  $r_x(k)$  for  $|k| > p$ . Supposing that the extrapolated autocorrelation is  $r_e(k)$ , the power spectrum of  $x(n)$  can be written as

$$P_x(e^{j\omega}) = \sum_{k=-p}^p r_x(k) e^{-jk\omega} + \sum_{|k|>p} r_e(k) e^{-jk\omega} \quad (98)$$

Now the question is how or what criterion should be used to determine the extrapolated autocorrelation. As the name of the method indicates, the maximum entropy is the criterion for performing the extrapolation. A maximum entropy extrapolation is equivalent to finding the sequence of the extrapolated autocorrelations that make  $x(n)$  as white (random) as possible. From the power spectrum point of view, this maximum entropy extrapolation makes the power spectrum as flat as possible.

#### C. The maximum entropy spectral estimate

If a random process  $x(n)$  is assumed to be a Gaussian process with a given segment of the autocorrelation  $r_x(k)$  for  $|k| \leq p$ , then the extrapolated autocorrelation  $r_e(k)$  that maximizes the entropy in Eq. (97) can be found by setting  $\partial H(x)/\partial r_e^*(k) = 0$ , specifically

$$\frac{\partial H(x)}{\partial r_e^*(k)} = \frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{P_x(e^{j\omega})} \frac{\partial P_x(e^{j\omega})}{\partial r_e^*(k)} d\omega = 0, \quad |k| > p \quad (99)$$

Using the conjugate symmetry  $r_x^*(k) = r_x(-k)$  in Eq. (99), the derivative in the integral becomes

$$\frac{\partial P_x(e^{j\omega})}{\partial r_e^*(k)} = \frac{\partial P_x(e^{j\omega})}{\partial r_e(-k)} = \frac{\partial}{\partial r_e(-k)} \sum_{|k|>p} r_e(k) e^{-jk\omega} = \frac{\partial}{\partial r_e(-k)} \sum_{|k|>p} r_e(-k) e^{jk\omega} = e^{jk\omega} \quad (100)$$

Inserting Eq. (100) into Eq. (99), we have

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} \frac{1}{P_x(e^{j\omega})} e^{jk\omega} d\omega = 0, \quad |k| > p \quad (101)$$

Defining  $Q_x(e^{j\omega}) = 1/P_x(e^{j\omega})$ , Eq. (101) becomes

$$\frac{1}{2\pi} \int_{-\pi}^{\pi} Q_x(e^{j\omega}) e^{jk\omega} d\omega = q_x(k) = 0, \quad |k| > p \quad (102)$$

which shows that the inverse Fourier transform of  $Q_x(e^{j\omega})$ , namely  $q_x(k)$ , is equal to zero for  $|k| > p$ . Therefore, the Fourier transform of  $q_x(k)$  is

$$Q_x(e^{j\omega}) = \frac{1}{P_x(e^{j\omega})} = \sum_{k=-\infty}^{\infty} q_x(k)e^{-jk\omega} = \sum_{k=-p}^p q_x(k)e^{-jk\omega}. \quad (103)$$

From Eq. (103), we may define the maximum entropy (MEM) estimate of the power spectrum  $P_x(e^{j\omega})$  for a Gaussian process, as follows

$$\hat{P}_{mem}(e^{j\omega}) = \frac{1}{\sum_{k=-p}^p q_x(k)e^{-jk\omega}} \quad (104)$$

Noting that  $q_x(-k) = q_x^*(k)$  and  $\sum_{k=-p}^p q_x(k)e^{-jk\omega} = \sum_{k=1}^p q_x^*(k)e^{jk\omega} + q_x(0) + \sum_{k=1}^p q_x(k)e^{-jk\omega}$ , we can perform spectral factorization on Eq. (104) and may have  $\hat{P}_{mem}(e^{j\omega})$  in the following form

$$\hat{P}_{mem}(e^{j\omega}) = \frac{b(0)b^*(0)}{\left[1 + \sum_{k=1}^p a_p(k)e^{-jk\omega}\right] \left[1 + \sum_{k=1}^p a_p^*(k)e^{jk\omega}\right]} = \frac{|b(0)|^2}{\left|1 + \sum_{k=1}^p a_p(k)e^{-jk\omega}\right|^2} = \frac{|b(0)|^2}{A_p(e^{j\omega})A_p^*(e^{-j\omega})}, \quad (105)$$

which is the same as the AR spectrum estimate. Since  $r_x(k)$  are given for  $|k| \leq p$ , the coefficients  $a_p(k)$  can be found by solving the following normal equations

$$\begin{bmatrix} r_x(0) & r_x^*(1) & \dots & r_x^*(p) \\ r_x(1) & r_x(0) & \dots & r_x^*(p-1) \\ \dots & \dots & \dots & \dots \\ r_x(p) & r_x(p-1) & \dots & r_x(0) \end{bmatrix} \begin{bmatrix} 1 \\ a_p(1) \\ \dots \\ a_p(p) \end{bmatrix} = \begin{bmatrix} 1 \\ 0 \\ \dots \\ 0 \end{bmatrix} \quad (106)$$

and  $b(0)$  can be determined from

$$|b(0)|^2 = r_x(0) + \sum_{k=1}^p a_p(k)r_x^*(k) = \varepsilon_p. \quad (107)$$

Defining  $\mathbf{a}_p = [1, a_p(1), \dots, a_p(p)]^T$  and  $\mathbf{e} = [1, e^{j\omega}, \dots, e^{jp\omega}]^T$ , the MEM may written as

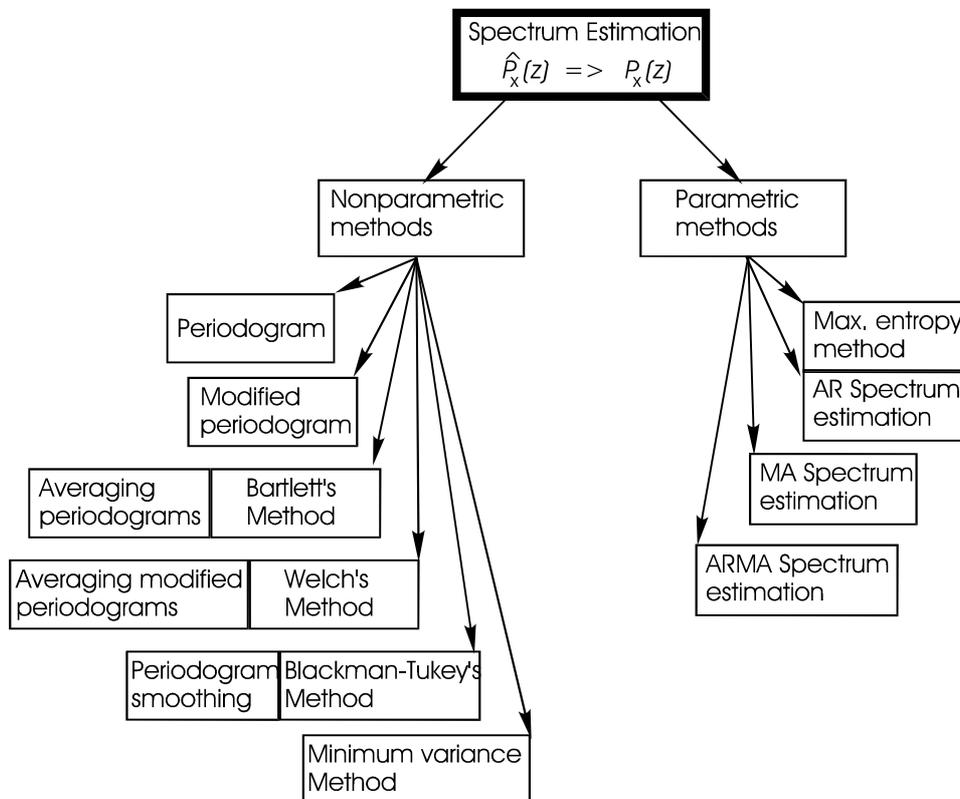
$$\hat{P}_{mem}(e^{j\omega}) = \frac{\varepsilon_p}{|\mathbf{e}^H \mathbf{a}_p|^2} \quad (108)$$

The maximum entropy method is equivalent to the Yule-Walker method (YWM), an AR spectrum estimation method. The difference between the two methods lies in that in the MEM the random process  $x(n)$  is assumed to be Gaussian, whereas in the YWM  $x(n)$  is assumed to be an AR process.

**Example 6. MEM estimation of the power spectrum of a complex exponential in noise** (Example 8.4.1)

Summary of spectral estimation methods:

- Power spectrum of a WSS process  $x(n)$ :  $P_x(e^{j\omega}) = \sum_{k=-\infty}^{\infty} r_x(k)e^{-jk\omega}$ , (note  $-\infty < k < \infty$ ) which is the Fourier transform of the autocorrelation sequence  $r_x(k)$ .
  - $P_x(e^{j\omega})$  can only be estimated when the data available for a random process  $x(n)$  are of finite-length or the data are contaminated with noise.
  - Estimating the power spectrum is equivalent to estimating the autocorrelation.
  - Two classes of methods for power spectrum estimation: nonparametric methods and parametric methods
- In each class there are a set of methods. Since  $P_x(e^{j\omega})$  is estimated with this or that method, the performance of the estimating method must be evaluated by looking into the expected value (bias) and the variance of the estimate  $\hat{P}_x(e^{j\omega})$ .



**Some useful MATLAB functions**

- >> Pxx = PYULEAR(X,ORDER) % returns the PSD estimate of a discrete-time signal vector X in % the vector Pxx, using Yule-Walker's method.
- >> Pxx = PCOV(X,ORDER) % returns the PSD estimate of a discrete-time signal vector X in the % vector Pxx, using the Covariance method.
- >> Pxx = PBURG(X,ORDER) % returns the PSD estimate of a discrete-time signal vector X in the % vector Pxx using Burg's method.
- >> Pxx = PMCOV(X,ORDER) % returns the PSD estimate of a discrete-time signal vector X in the % vector Pxx, using the Modified Covariance method.