SPECTRAL ANALYSIS

This article is concerned with the spectral analysis problem: that of determining the distribution in frequency of the power of a time series from a finite set of measurements. Spectral analysis has found wide applications in diverse fields, such as radar, sonar, speech, biomedicine, economics, geophysics, and others in which the spectral contents of signals are of interest. For example, in radar and sonar systems, the locations of the sources or targets can be estimated by measuring the spectral contents of the received signals. In biomedicine, the spectral analysis of the signals from a patient provides doctors useful information for diagnosis.

In practice, the observed data are often of finite duration; hence, the quality of the spectral estimation is usually limited by the shortness of the data record available. As a general rule, for stationary random signals, the longer the data record, the better the spectral estimates that can be obtained. For deterministic signals, although the spectral characteristics are described by an arbitrary length of data, our goal is to select a data record as short as possible so that we can resolve different signal components.

There are two broad classes of spectral analysis approaches: nonparametric methods and parametric (model-based) methods. The nonparametric methods, such as periodogram, Blackman-Tukey, and minimum variance spectral estimators do not impose any model assumption on the data other than wide-sense stationarity. The parametric spectral estimation approaches, on the other hand, assume that the measurement data satisfy a generating model by which the spectral estimation problem is usually converted to that of determining the parameters of the assumed model. Two kinds of models are widely assumed and used within the parametric methods according to different spectral characteristics of the signals: the rational transfer function (RTF) model and the sinusoidal signal model. The RTF models, including autocorrelation (AR), moving average (MA), and autocorrelation moving average (ARMA) types are usually used to analyze the signals with continuous spectra, while the sinusoidal signal model is a good approximation to signals with discrete spectra.

Our discussion is divided into two parts: stationary spectral analysis and nonstationary spectral analysis. In the first part, we introduce the nonparametric spectral estimation methods and discuss the parametric methods for rational spectral analysis and sinusoidal spectral analysis. In the second part, we study two nonstationary spectral analysis examples: damped sinusoidal parameter estimation and instantaneous frequency measurement.

Energy Spectral Density of Deterministic Signals

Suppose that $x_C(t)$ represents a continuous-time signal of interest; $x(n)$ denotes the sequence obtained by sampling $x_C(t)$ at some uniform sampling rate F_S; that is, $x(n) = x_c(n/F_S)$.

If $x(n)$ has finite energy,

$$E = \sum_{n=-\infty}^{\infty} |x(n)|^2 < \infty \tag{1}$$

then its Fourier transform exists and is given by

$$X(\omega) = \sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n} \tag{2}$$

Using Parseval's theorem, we have

$$E = \sum_{n=-\infty}^{\infty} |x(n)|^2 = \frac{1}{2\pi}\int_{-\pi}^{\pi} |X(\omega)|^2\, d\omega \tag{3}$$

Let us define

$$S(\omega) = |X(\omega)|^2 \tag{4}$$

then the quantity $S(\omega)$ can be interpreted as the distribution of the signal energy as a function of frequency ω and, hence, it is called the energy spectral density of the signal. Here, the frequency ω is measured in radians per sampling interval, which corresponds to the physical frequency $\omega/2\pi F_s$ in hertz. Note that the total energy of the signal is the integral of $S(\omega)$ over the interval $(-\pi, \pi)$ (within a constant scale $1/2\pi$).

If we define the autocorrelation function of the deterministic signal $x(n)$ as

$$r(k) = \sum_{n=-\infty}^{\infty} x^*(n)x(n+k) \tag{5}$$

we have

$$\begin{aligned}\sum_{k=-\infty}^{\infty} r(k)e^{-j\omega k} &= \sum_{k=-\infty}^{\infty}\sum_{n=-\infty}^{\infty} x^*(n)x(n+k)e^{j\omega n}e^{-j\omega(n+k)} \\ &= \left[\sum_{n=-\infty}^{\infty} x(n)e^{-j\omega n}\right]^*\left[\sum_{m=-\infty}^{\infty} x(m)e^{-j\omega m}\right] \\ &= S(\omega)\end{aligned} \tag{6}$$

Eq. (6) means that the energy spectral density $S(\omega)$ may also be viewed as the Fourier transform of the autocorrelation function of the signal $x(n)$.

The above relations provide us two ways for computing the energy spectral density of a deterministic signal from its samples $x(n)$, $n = 0, 1, \ldots, N-1$.

1. *Direct Method [based on Eqs. (2) and (4)].* The direct method involves computing the Fourier transform of $x(n)$ then calculating the energy spectral density via Eq. (4).
2. *Indirect Method [based on Eqs. (5) and (6)].* The autocorrelation function $r(k)$ is estimated first; then the energy spectral density is computed by performing Fourier transform on $r(k)$.

Power Spectral Density of Random Signals

In practical applications, most of the signals encountered can be characterized as stationary random processes, which do not have finite energy, and, hence, do not possess a Fourier transform counterpart. However, such signals have finite average power and, hence, can be characterized by power spectral density functions.

Let $x(n)$ be a zero-mean stationary random process with autocorrelation function (ACF) given by

$$r(k) = E[x^*(n)x(n+k)] \tag{7}$$

Hereafter, $E[\cdot]$ denotes the expectation operator. From the Wiener-Khinchin theorem, the power spectral density $R(\omega)$ and the autocorrelation function $r(k)$ of the stationary random process $x(n)$ form a Fourier transform pair which is described by

$$R(\omega) = \sum_{k=-\infty}^{\infty} r(k)e^{-j\omega k} \tag{8}$$

and

$$r(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} R(\omega)e^{j\omega k}\,d\omega \tag{9}$$

Eq. (8), which is similar to Eq. (6) for the deterministic signals, is the definition of power spectral density (PSD) for random signals.

If the ACF $r(k)$ decays sufficiently rapidly, so that

$$\lim_{N\to\infty}\frac{1}{N}\sum_{k=-N}^{N}|k||r(k)| = 0 \tag{10}$$

then the PSD defined by Eq. (8) is equivalent to the following expression

$$R(\omega) = \lim_{N\to\infty} E\left[\frac{1}{N}\left|\sum_{n=0}^{N-1}x(n)e^{-j\omega n}\right|^2\right] \tag{11}$$

In fact,

$$\begin{aligned}\lim_{N\to\infty} E\left[\frac{1}{N}\left|\sum_{n=0}^{N-1}x(n)e^{-j\omega n}\right|^2\right] &= \lim_{N\to\infty}\frac{1}{N}\sum_{l=0}^{N-1}\sum_{m=0}^{N-1}E[x^*(l)x(m)]e^{-j\omega(m-l)}\\ &= \lim_{N\to\infty}\frac{1}{N}\sum_{k=-(N-1)}^{N-1}(N-|k|)r(k)e^{-j\omega k}\\ &= R(\omega)\end{aligned} \tag{12}$$

where we have used the definition $r(m-l) = E[x^*(l)x(m)]$ and the double summation formula, for any arbitrary function $f(\cdot)$,

$$\sum_{l=0}^{N-1}\sum_{m=0}^{N-1}f(l-m) = \sum_{k=-(N-1)}^{N-1}(N-|k|)f(k) \tag{13}$$

The objective of spectral analysis is to estimate the PSD of $x(n)$ from a finite-duration of observed samples $x(0)$, $x(1)$, . . ., $x(N-1)$. Before we turn our attention to the spectral estimation methods, let us present some useful properties of the ACF and PSD of a stationary random process.

1. $r(-k) = r^*(k)$, and $r(0) \geq |r(k)|$, for all k.
2. $R(\omega)$ is a real-valued and nonnegative function.
3. If $x(n)$ is a real-valued signal, $R(-\omega) = R(\omega)$, for $-\pi \leq \omega \leq \pi$; if $x(n)$ is a complex-valued signal, in general, $R(-\omega) \neq R(\omega)$, for $-\pi \leq \omega \leq \pi$.
4. Let $y(n)$ be generated by driving a stationary random process $x(n)$ through a linear time-invariant system with transfer function

$$H(\omega) = \sum_{k=-\infty}^{\infty} h(k)e^{-j\omega k} \tag{14}$$

where $h(k)$ is the unit impulse response of the system. Then, the relation between the input PSD $R_x(\omega)$ and the output PSD $R_y(\omega)$ is given by

$$R_y(\omega) = |H(\omega)|^2 R_x(\omega) \tag{15}$$

Eq. (15) will be used in developing the parametric methods for rational spectral estimation.

NONPARAMETRIC METHODS FOR SPECTRAL ESTIMATION

In this section, we shall discuss the nonparametric spectral estimation methods. We first introduce the periodogram estimator and analyze its statistical properties in terms of the bias and the variance of the PSD estimate. Since the periodogram estimator has high variance even for large sample length, several modified methods such as Bartlett (1), Welch (2), and Blackman–Tukey (3) methods are then discussed. Finally, the minimum variance spectral estimator is given.

Periodogram Method

Based upon Eq. (11), the periodogram spectral estimator is defined as

$$\hat{R}_P(\omega) = \frac{1}{N}\left|\sum_{n=0}^{N-1}x(n)e^{-j\omega n}\right|^2 = \frac{1}{N}|X(\omega)|^2 \tag{16}$$

where $X(\omega)$ is the Fourier transform of the sample sequence $x(n)$. Note that the implementation of the periodogram estimator involves performing discrete Fourier transform (DFT) on $x(n)$, followed by calculating the PSD directly. Specifically, given N data points $x(0)$, $x(1)$, . . ., $x(N-1)$, we compute the N-point DFT at frequency

$$\omega = \frac{2\pi}{N}k, \quad k = 0, 1, \ldots, N-1 \tag{17}$$

that yields the samples of the periodogram

$$\hat{R}_1\left(\frac{2\pi}{N}k\right) = \frac{1}{N}\left|\sum_{n=0}^{N-1}x(n)e^{-j2\pi n\frac{k}{N}}\right|^2, \quad k = 0, 1, \ldots, N-1 \tag{18}$$

In practice, however,when the data length N is small, the estimated PSD computed by Eq. (18) does not provide a good representation of the continuous spectrum estimate due to the small number of samples. In order to get a more complete description about the estimated PSD, it is necessary to evaluate $\hat{R}_P(\omega)$ at more dense frequencies. This can be achieved by increasing the sequence length via zero padding. Specifically,

if the data length is increased to L ($L > N$), evaluating L-point DFT yields

$$\hat{R}_2\left(\frac{2\pi}{L}k\right) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x(n)e^{-j2\pi n\frac{k}{L}}\right|^2, \quad k = 0, 1, \ldots, L-1 \quad (19)$$

We now turn our attention to the statistical properties of the periodogram estimator. It is easy to verify that the periodogram estimator defined in Eq. (16) is equivalent to

$$\hat{R}_{\mathrm{P}}(\omega) = \sum_{k=-N+1}^{N-1} \hat{r}(k)e^{-j\omega k} \quad (20)$$

where $\hat{r}(k)$ takes the standard biased ACF estimate, which is given by

$$\hat{r}(k) = \frac{1}{N}\sum_{n=0}^{N-|k|+1} x^*(n)x(n+k), \quad k \geq 0 \quad (21)$$

and $\hat{r}(-k) = \hat{r}^*(k)$. Equation (20) provides us a way by which the statistical performance expressions for the periodogram estimator can be obtained. In doing so, let us first consider the statistical properties of the biased ACF estimators.

The expected value of $\hat{r}(k)$ is given by

$$\begin{aligned} E[\hat{r}(k)] &= \frac{1}{N}\sum_{n=0}^{N-|k|+1} E[x^*(n)x(n+k)] \\ &= \left(1 - \frac{|k|}{N}\right) r(k) \end{aligned} \quad (22)$$

where $r(k) = E[x^*(n)x(n+k)]$ denotes the true ACF of $x(n)$. The variance of $\hat{r}(k)$ is shown to be

$$\mathrm{var}[\hat{r}(k)] \approx \frac{1}{N}\sum_{n=-\infty}^{\infty} [|r(n)|^2 + r^*(n-k)r(n+k)] \quad (23)$$

Since

$$\lim_{N\to\infty} E[\hat{r}(k)] = r(k) \quad \text{and} \quad \lim_{N\to\infty} \mathrm{var}[\hat{r}(k)] = 0 \quad (24)$$

the biased ACF estimate $\hat{r}(k)$ is a consistent estimate of $r(k)$.

We now evaluate the statistical properties of the periodogram estimator. The expected value of $\hat{R}_{\mathrm{P}}(\omega)$ is given by

$$\begin{aligned} E[\hat{R}_{\mathrm{P}}(\omega)] &= \sum_{k=-N+1}^{N-1} \left(1 - \frac{|k|}{N}\right) r(k)e^{-j\omega k} \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} R(\alpha)W_{\mathrm{B}}^{(N)}(\omega - \alpha)\, d\alpha \end{aligned} \quad (25)$$

where $R(\omega)$ is the true PSD of $x(n)$, and

$$W_{\mathrm{B}}^{(N)}(\omega) = \frac{1}{N}\left[\frac{\sin(\omega N/2)}{\sin(\omega/2)}\right]^2 \quad (26)$$

is the Fourier transform of the so-called Bartlett window with length N, which is described as

$$w_{\mathrm{B}}^{(N)}(k) = \begin{cases} 1 - \dfrac{|k|}{N}, & \text{if } |k| \leq N-1 \\ 0, & \text{otherwise} \end{cases} \quad (27)$$

When $N \to \infty$, $W_{\mathrm{B}}^{(N)}(\omega)$ tends to be an ideal Dirac function; thus,

$$\lim_{N\to\infty} E[\hat{R}_{\mathrm{P}}(\omega)] = R(\omega) \quad (28)$$

However, in general, the variance of $\hat{R}_{\mathrm{P}}(\omega)$ does not decay to zero as $N \to \infty$. Especially when the data sequence is Gaussian random process, the variance is given by

$$\mathrm{var}[\hat{R}_{\mathrm{P}}(\omega)] = R^2(\omega)\left[1 + \left(\frac{\sin\omega N}{N\sin\omega}\right)^2\right] \quad (29)$$

and when $N \to \infty$, it becomes

$$\lim_{N\to\infty} \mathrm{var}[\hat{R}_{\mathrm{P}}(\omega)] = R^2(\omega) \quad (30)$$

Hence, the periodogram is an asymptotically unbiased estimate of the PSD, but it is not a consistent estimate in the sense that the variance does not decrease to zero, no matter how large the data length is.

Although the periodogram is an asymptotically unbiased estimator of the PSD, as we can see from Eq. (25), this estimator is problematic when N is small. Specifically, in order to make $E[\hat{R}_{\mathrm{P}}(\omega)]$ as close as $R(\omega)$, $W_{\mathrm{B}}^{(N)}(\omega)$ should be a close approximation to a Dirac impulse. However, $W_{\mathrm{B}}^{(N)}(\omega)$ is different from an ideal Dirac impulse in two respects. First, $W_{\mathrm{B}}^{(N)}(\omega)$ has a main lobe with half-power (3 dB) width about $1/N$ in frequency. The effect of the main lobe is to smooth the estimated spectrum. In fact, when $R(\omega)$ has two peaks separated in frequency by less than $1/N$, these two peaks appear as a single broader peak in $E[\hat{R}_{\mathrm{P}}(\omega)]$ due to the filtering effect of $W_{\mathrm{B}}^{(N)}(\omega)$, as seen in Eq. (25). For this reason, $1/N$ is referred to as the spectral resolution limit of the periodogram method. Secondly, $W_{\mathrm{B}}^{(N)}(\omega)$ has a large number of sidelobes whose energies are leaked from the main lobe, which may obscure and distort other spectral responses that are present. In this case, weak signal spectral response can be masked by higher sidelobes from stronger spectral response.

We point out that while zero padding provides us a method for interpolating the values of the measured spectrum at more frequencies, it cannot improve the spectral resolution of the periodogram estimator since the continuous spectral estimate, $\hat{R}_{\mathrm{P}}(\omega)$, is the same for both the original data sequence and the sequence padded with zeros.

Due to the performance limitations of the periodogram, several modified methods, such as the Bartlett and Welch methods, are proposed to reduce either the bias or the variance of the spectral estimates.

Modified Periodogram Method

The Bartlett method (1) and the Welch method (2) are two modified periodogram methods. These methods aim at reducing the variance of the periodogram estimate by splitting up the N available observations into K segments, and then averaging the periodograms computed from each segments for each value of ω.

Let

$$x_{\mathrm{i}}(n) = x(n + iD), \quad i = 0, 1, \ldots, K-1; \quad n = 0, 1, \ldots, M-1 \quad (31)$$

denote the observations of the ith segment, where iD is the starting point of the ith segment. The Bartlett method takes $D = M$, and $N = LM$; thus, data samples in successive segments are not overlapped. In the Welch method, one chooses $D < M$ and obtains overlapped data samples in successive segments. For example, if $D = M/2$, there is 50% overlapping between successive data segments, and $K = 2L$ segments are obtained.

Let

$$\hat{R}^{(i)}(\omega) = \frac{1}{M}\left|\sum_{n=0}^{M-1} x_{\mathrm{i}}(n)e^{-j\omega n}\right|^2 \tag{32}$$

represent the periodogram of the ith segment. The Bartlett spectral estimator is defined as

$$\hat{R}_{\mathrm{B}}(\omega) = \frac{1}{L}\sum_{i=0}^{L-1}\hat{R}^{(i)}(\omega) \tag{33}$$

The Welch spectral estimator is defined as

$$\hat{R}_{\mathrm{W}}(\omega) = \frac{1}{K}\sum_{i=0}^{K-1}\hat{R}_{\mathrm{M}}^{(i)}(\omega) \tag{34}$$

where $\hat{R}_{\mathrm{M}}^{(i)}(\omega)$ is the windowed periodogram, given by

$$\hat{R}_{\mathrm{M}}^{(i)}(\omega) = \frac{1}{MP}\left|\sum_{n=0}^{M-1} x_{\mathrm{i}}(n)w(n)e^{-j\omega n}\right|^2 \tag{35}$$

with P the "power" of the time window $w(n)$,

$$P = \frac{1}{M}\sum_{n=0}^{M-1} w^2(n) \tag{36}$$

It is noted that in the Welch method, the data samples in each segment are windowed before they are performed via Fourier transform.

The statistical properties of the Bartlett estimator are easily obtained. First, the expected value of $\hat{R}_{\mathrm{B}}(\omega)$ is given by

$$\begin{aligned} E[\hat{R}_{\mathrm{B}}(\omega)] &= \frac{1}{L}\sum_{i=0}^{L-1} E[\hat{R}^{(i)}(\omega)] \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} R(\alpha)W_{\mathrm{B}}^{(M)}(\omega-\alpha)\,d\alpha \end{aligned} \tag{37}$$

where $W_{\mathrm{B}}^{(M)}(\omega)$ is the Fourier transform of the Bartlett window with length M. Compared with Eq. (25), the 3-dB spectral width of $W_{\mathrm{B}}^{(M)}(\omega)$ is now increased by a factor L since the window length is decreased by a factor L, which results in the reduction of frequency resolution by a factor L.

Secondly, the variance of the Bartlett spectral estimator is given by, when the data sequence is a Gaussian random process,

$$\begin{aligned} \mathrm{var}[\hat{R}_{\mathrm{B}}(\omega)] &= \frac{1}{L^2}\sum_{i=0}^{L-1}\mathrm{var}[\hat{R}^{(i)}(\omega)] \\ &= \frac{1}{L}R^2(\omega)\left[1+\left(\frac{\sin\omega M}{M\sin\omega}\right)^2\right] \end{aligned} \tag{38}$$

Eq. (38) shows that the variance of the Bartlett estimator has been reduced approximately by a factor L, as compared to that of the original periodogram method.

To evaluate the statistical properties of the Welch estimator, we first derive the expected value of the windowed periodogram. Taking expectation of $\hat{R}_{\mathrm{M}}^{(i)}(\omega)$ in Eq. (35) yields

$$\begin{aligned} E[\hat{R}_{\mathrm{M}}^{(i)}(\omega)] &= \frac{1}{MP}\sum_{n=0}^{M-1}\sum_{m=0}^{M-1} w(n)w(m)E[x_{\mathrm{i}}(n)x_{\mathrm{i}}^*(m)]e^{-j\omega(n-m)} \\ &= \frac{1}{MP}\sum_{n=0}^{M-1}\sum_{m=0}^{M-1} w(n)w(m)r(n-m)e^{-j\omega(n-m)} \\ &= \frac{1}{MP}\sum_{\tau=-M+1}^{M-1} w_{\mathrm{l}}(\tau)r(\tau)e^{-j\omega\tau} \end{aligned} \tag{39}$$

where

$$w_{\mathrm{l}}(\tau) = \frac{1}{MP}\sum_{n=\max\{0,\tau\}}^{\min\{M-1,M-1+\tau\}} w(n)w(n+\tau) \tag{40}$$

is called the lag window. Let $W_{\mathrm{l}}(\omega)$ be the Fourier transform of $w_{\mathrm{l}}(n)$; we can rewrite Eq. (39) as

$$E[\hat{R}_{\mathrm{M}}^{(i)}(\omega)] = \frac{1}{2\pi}\int_{-\pi}^{\pi} R(\alpha)W_{\mathrm{l}}(\omega-\alpha)\,d\alpha \tag{41}$$

Comparing Eq. (41) with Eq. (25), we observe that the windowed periodogram is simplified to the original periodogram when $w(n)$ is chosen to be 1 for $0 \le n \le M-1$, and the same data record is used. However, since different windows yield different lag windows which may have different main lobes and sidelobes, we may control the resolution and leakage properties of the estimated PSD by choosing different windows. We shall, in the next subsection, further discuss the lag windows in terms of their time and frequency domain characteristics.

The expected value of the Welch spectral estimator is given by

$$\begin{aligned} E[\hat{R}_{\mathrm{W}}(\omega)] &= \frac{1}{K}\sum_{i=0}^{K-1} E[\hat{R}_{\mathrm{M}}^{(i)}(\omega)] \\ &= E[\hat{R}_{\mathrm{M}}^{(i)}(\omega)] \\ &= \frac{1}{2\pi}\int_{-\pi}^{\pi} R(\alpha)W_{\mathrm{l}}(\omega-\alpha)\,d\alpha \end{aligned} \tag{42}$$

The variance of the Welch estimator is

$$\mathrm{var}[\hat{R}_{\mathrm{W}}(\omega)] = \frac{1}{K^2}\sum_{i=0}^{K-1}\sum_{j=0}^{K-1} E[\hat{R}_{\mathrm{M}}^{(i)}(\omega)\hat{R}_{\mathrm{M}}^{(j)}(\omega)] - \{E[\hat{R}_{\mathrm{W}}(\omega)]\}^2 \tag{43}$$

In the case of 50% overlapping between the successive data segments ($K = 2L$), the variance of the Welch estimator with the Bartlett time window is given by (2), for the Gaussian process,

$$\mathrm{var}[\hat{R}_{\mathrm{W}}(\omega)] \approx \frac{9}{8K}R^2(\omega) \tag{44}$$

From Eqs. (42) and (44), we have the following observations. First, data overlapping between the successive data segments yields more periodograms which can be used for averaging; hence, the variance of the Welch spectral estimator is further decreased as compared with that of the Bartlett estimator. Secondly, one may control the resolution and leakage properties of the estimated PSD by choosing different time windows.

Blackman–Tukey Method

As we have seen, the main problem with the periodogram method is the high variance (statistical variability) of the spectral estimator. This performance limitation may be attributed to the poor performance of the sample ACF estimates, from the equivalent definition of the periodogram given by Eqs. (20) and (21). In fact, the estimates of ACFs with larger lags will have poorer performances since they involve fewer data samples. In order to reduce the statistical variance of the periodogram estimator, Blackman and Tukey (3) suggested to window the sample ACF estimates and give less weight to the ACFs with larger lags before the Fourier transform is performed; that is,

$$\hat{R}_{\mathrm{T}}(\omega) = \sum_{k=-M+1}^{M-1} w_{\mathrm{l}}(k)\hat{r}(k)e^{-j\omega k} \tag{45}$$

where the lag window $w(k)$ has the following properties: (1) $0 \le w_{\mathrm{l}}(k) \le w(0) = 1$, (2) $w_{\mathrm{l}}(-k) = w_{\mathrm{l}}(k)$, and (3) $w_{\mathrm{l}}(k) = 0$ for $|k| > M$, and $M \le N - 1$. Some of the popular lag windows are listed in Table 1.

Table 1. Lag Windows

Name	Definition $w_{\mathrm{l}}(k)$, $W_{\mathrm{l}}(\omega)$
Rectangular	$w_{\mathrm{l}}(k) = 1, \quad \lvert k\rvert \le M$
	$W_{\mathrm{l}}(\omega) = W_{\mathrm{R}}(\omega) = \dfrac{\sin[\omega(2M+1)/2]}{\sin(\omega/2)}$
Bartlett	$w_{\mathrm{l}}(k) = 1 - \dfrac{\lvert k\rvert}{M}, \quad \lvert k\rvert \le M$
	$W_{\mathrm{l}}(\omega) = W_{\mathrm{B}}(\omega) = \dfrac{1}{M}\left(\dfrac{\sin M\omega/2}{\sin(\omega/2)}\right)^2$
Hanning	$w_{\mathrm{l}}(k) = 0.5 + 0.5\cos\dfrac{\pi k}{M}, \quad \lvert k\rvert \le M$
	$W_{\mathrm{l}}(\omega) = 0.25W_{\mathrm{B}}\left(\omega - \dfrac{\pi}{M}\right) + 0.5W_{\mathrm{B}}(\omega) + 0.25W_{\mathrm{B}}\left(\omega + \dfrac{\pi}{M}\right)$
Hamming	$w_{\mathrm{l}}(k) = 0.54 + 0.46\cos\dfrac{\pi k}{M}, \quad \lvert k\rvert \le M$
	$W_{\mathrm{l}}(\omega) = 0.23W_{\mathrm{B}}\left(\omega - \dfrac{\pi}{M}\right) + 0.54W_{\mathrm{B}}(\omega) + 0.23W_{\mathrm{B}}\left(\omega + \dfrac{\pi}{M}\right)$
Parzen	$w_{\mathrm{l}}(k) = \begin{cases} 2\left(1 - \dfrac{\lvert k\rvert}{M}\right)^3 - \left(1 - 2\dfrac{\lvert k\rvert}{M}\right)^3, & \lvert k\rvert \le \dfrac{M}{2} \\ 2\left(1 - \dfrac{\lvert k\rvert}{M}\right)^3, & \dfrac{M}{2} < \lvert k\rvert \le M \end{cases}$
	$W_{\mathrm{l}}(\omega) = \dfrac{8}{M^3}\left(\dfrac{3\sin^4 M\omega/4}{2\sin^4\omega/2} - \dfrac{\sin^4 M\omega/4}{\sin^2\omega/2}\right)$

Let $W_{\mathrm{l}}(\omega)$ be the Fourier transform of $w_{\mathrm{l}}(n)$,

$$W(\omega) = \sum_{k=-\infty}^{\infty} w(k)e^{-j\omega k} = \sum_{k=-M}^{M} w(k)e^{-j\omega k} \tag{46}$$

we have, from Eq. (45),

$$\hat{R}_{\mathrm{T}}(\omega) = \frac{1}{2\pi}\int_{-\pi}^{\pi} \hat{R}_{\mathrm{P}}(\alpha)W_{\mathrm{l}}(\omega - \alpha)\,d\alpha \tag{47}$$

where $\hat{R}_{\mathrm{P}}(\omega)$ is the periodogram spectral estimate defined by Eq. (16).

We now analyze the statistical properties of the Blackman-Tukey spectral estimator. From Eq. (47), we have

$$E[\hat{R}_{\mathrm{T}}(\omega)] = \frac{1}{2\pi}\int_{-\pi}^{\pi} E[\hat{R}_{\mathrm{P}}(\alpha)]W_{\mathrm{l}}(\omega - \alpha)\,d\alpha \tag{48}$$

Substituting $E[\hat{R}_{\mathrm{P}}(\alpha)]$ from Eq. (25) into Eq. (48), we obtain

$$E[\hat{R}_{\mathrm{T}}(\omega)] = \frac{1}{4\pi^2}\int_{-\pi}^{\pi}\int_{-\pi}^{\pi} R(\theta)W_{\mathrm{B}}^{(N)}(\alpha - \theta)W_{\mathrm{l}}(\omega - \alpha)\,d\alpha\,d\theta \tag{49}$$

If the window length of $w(n)$ is chosen such that $M \ll N$, Eq. (49) becomes

$$E[\hat{R}_{\mathrm{T}}(\omega)] \approx \frac{1}{2\pi}\int_{-\pi}^{\pi} R(\alpha)W_{\mathrm{l}}(\omega - \alpha)\,d\alpha \tag{50}$$

The variance of the Blackman-Tukey spectral estimator may be shown to be

$$\mathrm{var}[\hat{R}_{\mathrm{T}}(\omega)] \approx \frac{1}{2\pi N}R^2(\omega)\int_{-\pi}^{\pi} W_{\mathrm{l}}^2(\alpha)\,d\alpha \tag{51}$$

when $R(\omega)$ is smooth over the main lobe of the spectral window ($\approx 4\pi/M$, $M \ll N$), and the random process is Gaussian. By Parseval's theorem, we may rewrite Eq. (51) as

$$\mathrm{var}[\hat{R}_{\mathrm{T}}(\omega)] \approx \frac{1}{N}R^2(\omega)\sum_{k=-M}^{M} w_{\mathrm{l}}^2(k) \tag{52}$$

If $w_{\mathrm{l}}(n)$ is chosen to be Bartlett window with length M, we have

$$\mathrm{var}[\hat{R}_{\mathrm{T}}(\omega)] \approx \frac{2M}{3N}R^2(\omega) \tag{53}$$

Eq. (54) shows that the effect of windowing the autocorrelation is to smooth the periodogram estimate, thus to decrease the variance of the spectral estimates. On the other hand, from Eq. (50), the windowing also results in the reduction of spectral resolution since the lag window length is now decreased to M. Note that the smaller the M, the larger the reduction in variance and the lower the resolution.

Minimum-Variance Spectral Estimation

Capon (4) proposed a minimum variance spectral estimator (MVSE) for estimating the PSD of a random process by measuring the power of the output of a set of narrowband filters.

Let us consider an FIR filter with coefficients

$$\mathbf{a} = [a(0), a(1), \ldots, a(p)]^T \tag{54}$$

Suppose the observed data $x(n)$, $n = 0, 1, \ldots, N-1$ are passed through the filter; we obtain the response

$$y(n) = \sum_{k=0}^{p} a(k)x(n-k) = \mathbf{x}^T(n)\mathbf{a} \tag{55}$$

where

$$\mathbf{x}(n) = [x(n), x(n-1), \ldots, x(n-p)]^T \tag{56}$$

The coefficients are chosen so that the frequency response of the filter is unity at the frequency under consideration ω_0, and the variance of the output process is minimized. Thus, the filter should adjust itself to reject components of the spectrum not near ω_0 so that the output power is due mainly to the frequency components close to ω_0. If the process $x(n)$ is zero mean, the filter coefficients are estimated by minimizing the variance:

$$\sigma^2 = E[|y(n)|^2] = \mathbf{a}^H \mathbf{R} \mathbf{a} \tag{57}$$

subject to the unity frequency constraint:

$$\mathbf{e}^H(\omega_0)\mathbf{a} = 1 \tag{58}$$

where $\mathbf{R} = E[\mathbf{x}(n)\mathbf{x}^H(n)]$ is the autocorrelation matrix of the sequence $x(n)$, and $\mathbf{e}(\omega_0)$ is the vector

$$\mathbf{e}(\omega_0) = [1, e^{j\omega_0}, \ldots, e^{jp\omega_o}]^T \tag{59}$$

The solution for the filter coefficients can be shown to be (5)

$$\hat{\mathbf{a}} = \frac{\mathbf{R}^{-1}\mathbf{e}(\omega_0)}{\mathbf{e}^H(\omega_0)\mathbf{R}^{-1}\mathbf{e}(\omega_0)} \tag{60}$$

and the minimum output variance is

$$\sigma_{\mathrm{m}}^2 = \frac{1}{\mathbf{e}^H(\omega_0)\mathbf{R}^{-1}\mathbf{e}(\omega_0)} \tag{61}$$

Given finite duration of data $x(n)$, $n = 0, 1, \ldots, N-1$, we may first compute the autocorrelation matrix of $x(n)$, denoted as $\hat{\mathbf{R}}$. Then, the MVSE is obtained by

$$R(\omega) = \frac{1}{\mathbf{e}^H(\omega)\hat{\mathbf{R}}^{-1}\mathbf{e}(\omega)} \tag{62}$$

PARAMETRIC METHODS FOR RATIONAL SPECTRAL ESTIMATION

In the preceding section, we have studied the nonparametric spectral estimation methods which are usually implemented by the FFT technique. These methods are computationally efficient and yield reasonable spectral estimates when long data records are available. However, there are two main performance limitations involved with them. First, the frequency resolution in hertz is roughly the reciprocal of the time interval in seconds over which sampled data is available; thus, these methods are troublesome when analyzing short data records. Second, they suffer from spectral leakage effects due to windowing that is inherent in finite-length data records. In fact, weak signal spectral response can be masked by higher sidelobes from stronger spectral responses.

In order to alleviate the inherent limitations of the nonparametric methods, many spectral estimation procedures have been proposed. These methods assume that the signal of interest satisfies a generating model with known functional form and, hence, are referred to as model-based or parametric methods. Two broad classes of models are widely used and studied: the rational spectral model and the sinusoidal spectral model. The former is employed to analyze the signals with continuous spectra, while the latter is a candidate for describing signals with discrete spectra. In both models, the signal's spectra can be represented in terms of the model parameters; thus, the spectral estimation problem is usually converted to the model parameter estimation problem.

In this section, we focus on the parametric methods for rational spectral estimation. We assume that the signal is generated by passing a zero-mean white noise process $u(n)$ through a linear time invariant system; that is,

$$x(n) = -\sum_{k=1}^{p} a(k)x(n-k) + \sum_{k=0}^{q} b(k)u(n-k) \tag{63}$$

where $u(n)$ is called driving noise, and without loss of generality, $b(0) = 1$. The corresponding system transfer function is

$$H(z) = \frac{B(z)}{A(z)} \tag{64}$$

where

$$A(z) = 1 + \sum_{k=1}^{p} a(k)z^{-k} \tag{65}$$

and

$$B(z) = \sum_{k=0}^{q} b(k)z^{-k} \tag{66}$$

From Eq. (63), three types of rational models are readily derived:

1. *Autoregressive Moving–Average (ARMA) Model.* The pole-zero model in Eq. (63) is said to be an ARMA model of orders p and q and is denoted as ARMA(p, q). $a(k)$'s and $b(k)$'s (p and q) are referred to as AR and MA coefficients (orders, respectively.
2. *Autoregressive (AR) Model.* If $q = 0$, the model in Eq. (63) is simplified to an all-pole model with order p and is referred to as an AR(p) model.
3. *Moving-Average (MA) Model.* If $p = 0$, the model in Eq. (63) is reduced to an all-zero model with order q, and is called an MA(q) model.

In power spectral estimation, the input sequence $u(n)$ is not available. However, $u(n)$ is often assumed to be a zero-mean white noise process with variance σ^2. From Eq. (15), the PSD

of the observed data is related to the model parameters by

$$R(\omega) = \sigma^2 |H(e^{j\omega})|^2 = \sigma^2 \left| \frac{\sum_{k=0}^{q} b(k) e^{-j\omega k}}{1 + \sum_{k=1}^{p} a(k) e^{-j\omega k}} \right|^2 \tag{67}$$

To estimate the PSD, we need only estimate the parameters $\{a(1), \ldots, a(p), b(1), \ldots, b(q), \sigma^2\}$ and substitute the estimated values into Eq. (67).

Given a finite-duration of data samples, the first step toward spectral estimation is to select an appropriate model to fit the observed data. According to Wold decomposition and the Kolmogorov theorem, any ARMA or MA process may be represented uniquely by an AR model of possibly infinite order; likewise, any ARMA or AR process may be represented by an MA model of possible infinite order. However, using higher order appropriate models may not only result in spurious spectral peaks but also require more complex computations. Thus, our objective is to select the model that requires the smallest number of parameters which are also easily estimated. In view of the spectral characteristics of the signal, AR models are usually used to describe the signals with narrow spectral peaks, MA models are suitable for representing signals with broad spectral peaks and sharp nulls, while ARMA models are employed for representing signals with both sharp spectral peaks and deep nulls with relatively small orders.

Once a model is selected, the spectral estimation problem is converted to a model parameter estimation problem. In this section, we shall discuss this problem in terms of AR, MA, and ARMA models. Before describing the methods for estimating the parameters of AR(p), MA(q), and ARMA(p, q) models, let us first establish an important relation between the ACFs of the observed data and the model parameters.

Multiplying both sides of Eq. (63) by $x^*(n - m)$ and taking expectations yield

$$r(m) = -\sum_{k=1}^{p} a(k) r(m-k) + \sum_{k=0}^{q} b(k) E[u(n-k) x^*(n-m)] \tag{68}$$

where $r(i) = E[x^*(n)x(n + i)]$. Suppose that the filter $H(z)$ is asymptotically stable and causal; that is,

$$H(z) = \sum_{k=0}^{\infty} h(k) z^{-k} \tag{69}$$

or

$$x(n) = \sum_{i=0}^{\infty} h(k) u(n-i) \tag{70}$$

Then, the term $E[u(n - k)x^*(n - m)]$ becomes

$$E[u(n-k)x^*(n-m)] = E\left[u(n-k) \cdot \sum_{i=0}^{\infty} h^*(i) u^*(n-m-i) \right] = \sigma^2 h^*(k-m) \tag{71}$$

Since the filter is causal, that is, $h(i) \equiv 0$ for $i < 0$, Eq. (68) becomes

$$r(m) = \begin{cases} -\sum_{k=1}^{p} a(k) r(m-k) + \sigma^2 \sum_{k=0}^{q-m} h(k) b(k+m), & \\ \qquad\qquad\qquad\qquad\qquad\qquad m = 0, 1, \ldots, q & \\ -\sum_{k=1}^{p} a(k) r(m-k), \qquad\qquad m \geq q+1 & \end{cases} \tag{72}$$

The above relation is referred to as the Yule–Walker equations, which are the basis for determining the AR coefficients of AR or ARMA processes.

AR Spectral Estimation

For an AR model, the observed data satisfy the following difference equation:

$$x(n) = -\sum_{k=1}^{p} a(k) x(n-k) + u(n) \tag{73}$$

where $u(n)$ is a zero-mean white noise process with variance σ^2.

Let $q = 0$ in Eq. (72); we obtain

$$r(m) = -\sum_{k=1}^{p} a(k) r(m-k), \quad m = 1, 2, \ldots \tag{74}$$

Eq. (74) is referred to as the Yule–Walker equation for the AR process. Choosing $m = 1, 2, \ldots, p$, we have the following normal equation:

$$\begin{bmatrix} r(0) & r(-1) & \cdots & r(-p+1) \\ r(1) & r(0) & \cdots & r(-p+2) \\ \vdots & \vdots & \vdots & \vdots \\ r(p-1) & r(p-2) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ a(p) \end{bmatrix} = - \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(p) \end{bmatrix} \tag{75}$$

On the other hand, setting $q = 0$ and $m = 0$ in Eq. (75) yields

$$\sigma^2 = r(0) + \sum_{k=1}^{p} a(k) r(-k) \tag{76}$$

Eqs. (75) and (76) are the basis for estimating the AR coefficients and the noise variance. Given finite-duration of data $x(n)$, $n = 0, 1, \ldots, N - 1$, the sample ACF estimates $\{\hat{r}(k)\}_{k=0}^{p}$ are first computed by using the standard biased ACF estimator, which is defined in Eq. (21). Then, the AR coefficients $\hat{a}(k)$'s are estimated via Eq. (75), and the noise variance $\hat{\sigma}^2$ via Eq. (79). Finally, the AR spectra are computed by

$$\hat{R}(\omega) = \frac{\hat{\sigma}^2}{|1 + \sum_{k=1}^{p} \hat{a}(k) e^{-j\omega k}|^2} \tag{77}$$

Since the autocorrelation matrix in Eq. (75) is positive definite for any p, the solution for the AR parameters is unique by solving Eq. (75). When the standard biased ACF estimates are inserted into Eq. (75) in place of the true ACFs, the matrix is still positive definite; thus, the Yule–Walker method yields a unique solution to the AR parameter estimation.

Levinson-Durbin Algorithm. Although the above procedure yields the desired parameter estimates, the computation of matrix inversion requires the order of p^3 multiplication if the standard procedures are used. Computationally efficient algorithms can be derived by taking advantage of the structure properties of the autocorrelation matrix. In fact, if Eqs. (75) and (76) are combined, we may obtain a single matrix equation of the form

$$\begin{bmatrix} r(0) & r(-1) & \cdots & r(-p) \\ r(1) & r(0) & \cdots & r(-p+1) \\ \vdots & \vdots & \vdots & \vdots \\ r(p) & r(p-2) & \cdots & r(0) \end{bmatrix} \begin{bmatrix} 1 \\ a(1) \\ \vdots \\ a(p) \end{bmatrix} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{78}$$

Since the autocorrelation matrices in Eqs. (75) and (78) are both complex conjugate symmetric and Toepliz, Eq. (78) can be solved efficiently via order-recursive technique, which requires only the order of p^2 multiplication. This algorithm is referred to as the Levinson–Durbin algorithm (LDA) (6,7).

The basic idea of the LDA is to recursively compute the parameter sets $\{a(k, 1), a(k, 2), \ldots, a(k, k), \sigma^2(k), k = 1, 2, \ldots, p\}$. Here, we add an additional variable k to denote the order. The final set at order $k = p$ determines the desired solution. Especially, the recursive algorithm is initialized by

$$a(1,1) = -r(1)/r(0) \tag{79}$$

$$\sigma^2(1) = (1 - |a(1,1)|^2)r(0) \tag{80}$$

and updated by, for $k = 2, \ldots, p$,

$$a(k,k) = -\left[r(k) + \sum_{l=1}^{k-1} a(k-1,l)r(k-l)\right] \Big/ \sigma^2(k-1) \tag{81}$$

$$a(k,i) = a(k-1,i) + a(k,k)a^*(k-1,k-i) \tag{82}$$

$$\sigma^2(k) = (1 - |a(k,k)|^2)\sigma^2(k-1) \tag{83}$$

LS Algorithm. For AR models, the Yule–Walker equations in Eq. (74) are satisfied for any $m \geq 1$. However, the above approaches adopt only the first p linear equations (i.e., $m = 1, 2, \ldots, p$). Note when finite data is available, the standard biased ACF estimates are used in order to guarantee the positive definite property of the autocorrelation matrix. The errors in the ACF estimates result in the errors of the AR estimates. To obtain better AR parameter estimates, one may increase m in Eq. (74) to obtain an overdetermined system of linear equations. Specifically, letting $m = 1, 2, \ldots, t, t > p$ in Eq. (74) yields

$$\begin{bmatrix} r(0) & r(-1) & \cdots & r(-p+1) \\ r(1) & r(0) & \cdots & r(-p+2) \\ \vdots & \vdots & \vdots & \vdots \\ r(t-1) & r(t-2) & \cdots & r(t-p) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ a(p) \end{bmatrix} = - \begin{bmatrix} r(1) \\ r(2) \\ \vdots \\ r(t) \end{bmatrix} \tag{84}$$

or

$$\mathbf{Ra} = -\mathbf{b} \tag{85}$$

the LS solution to Eq. (85) is given by

$$\mathbf{a} = -(\mathbf{R}^T\mathbf{R})^{-1}\mathbf{R}^T\mathbf{b} \tag{86}$$

AR Order Determination. In practice, the AR order p is usually unknown *a priori;* it is necessary to determine this parameter when AR modeling is used. Following are four objective criteria for AR model order determination.

1. *Final prediction criterion (FPE).* The FPE is based on selecting the order that minimizes the performance index

$$\text{FPE}(p) = \hat{\sigma}_{\text{W}}^2(p)\left(\frac{N+p+1}{N-p-1}\right) \tag{87}$$

where $\hat{\sigma}_{\text{W}}^2(p)$ is the estimated variance of the linear prediction error, and N is the number of data samples.

2. *Akaike Information Criterion (AIC).* The AIC method determines the model order by minimizing an information theoretical function, which is defined as

$$\text{AIC}(p) = \ln \hat{\sigma}_{\text{W}}^2(p) + \frac{2p}{N} \tag{88}$$

3. *Minimum Description Length (MDL) Method.* The MDL method is another information criterion which selects the order by minimizing the description length (MDL):

$$\text{MDL}(p) = N \ln \hat{\sigma}_{\text{W}}^2(p) + p \ln N \tag{89}$$

4. *Criterion Autoregressive Transfer (CAT) Function Method.* The CAT chooses the order by minimizing the function

$$\text{CAT}(p) = \frac{1}{N}\sum_{j=1}^{p} \frac{1}{\overline{\sigma}_{\text{W}}^2(j)} - \frac{1}{\hat{\sigma}_{\text{W}}^2(p)} \tag{90}$$

where $\overline{\sigma}_{\text{W}}^2(j) = [N/(N-j)]\hat{\sigma}_{\text{W}}^2(p)$.

MA Spectral Estimation

An MA(q) signal is obtained by filtering a white noise process through an all-zero system; that is,

$$x(n) = \sum_{k=0}^{q} b(k)u(n-k) \tag{91}$$

where $u(n)$ is the aforementioned noise process with variance σ^2. MA models are usually used to characterize the processes with broad peaks and sharp nulls.

For MA(q) processes, we have $h(k) = b(k)$ for $0 \leq k \leq q$, and $h(k) = b(k) = 0$ for $k < 0$ and $k > q$; thus, the MA coefficients are related to the ACFs of the process by

$$r(m) = \begin{cases} \sigma^2 \sum_{k=0}^{q-|m|} b(k)b(k+m), & |m| \leq q \\ 0 & |m| > q \end{cases} \tag{92}$$

As contrasted to the AR signal case, there does not exist a system of linear equations which link the MA parameters and the ACFs of the observed data. If the ACFs $\{r(0), r(1), \ldots, r(q)\}$ are known or estimated, the parameters of the MA(q) process can be determined by solving the set of nonlinear equations from Eq. (92). This is implemented via iterative optimization techniques which are usually computationally expensive and are not guaranteed to converge to the optimal solution.

If only a spectral estimate is desired, however, there is no need to solve for the MA parameters, but only to determine the PSD by

$$\hat{R}(\omega) = \sum_{m=-q}^{q} \hat{r}(m) e^{-j\omega m} \tag{93}$$

where $\hat{r}(m)$ is the sample ACF estimate obtained from the finite-duration of observation data. Compared with Eq. (45), the MA spectral estimator is with the form of the Blackman–Tukey estimator. More precisely, Eq. (93) coincides with the Blackman–Tukey estimator using a rectangular window of length $2q + 1$. We point out that the Blackman–Tukey estimator is applicable to any random process, while the MA spectral estimator is not.

Finally, if we have to estimate the MA parameters, an alternative linear method (Durbin's method) can be used, which is based upon a higher order AR model approximation to the MA process. Let the MA(q) process be modeled by an AR(p) model with parameters $\{a(1), a(2), \ldots, a(p)\}$, where $p \gg q$. Then, we have

$$\sum_{k=0}^{q} b(k) z^{-k} = \frac{1}{1 + \sum_{k=1}^{p} a(k) z^{-k}} \tag{94}$$

or

$$a(n) + \sum_{k=1}^{q} b(k) a(n-k) = \begin{cases} 1, & n = 0 \\ 0, & n \neq 0 \end{cases} \tag{95}$$

The fitted AR(p) model parameters can be estimated via a linear method discussed earlier, while the MA parameters $b(k)$'s are linearly related to the estimated AR parameters.

Durbin's method is summarized as follows.

Step 1. Use a high-order AR(p) model ($p \gg q$) to fit the observed data, and obtain the AR parameters $\hat{a}(k)$, $k = 1, 2, \ldots, p$, and the noise variance $\hat{\sigma}^2$.

Step 2. Solve Eq. (95) using least-squares error criterion, and obtain the MA parameter estimates $\hat{b}(k)$, $k = 1, 2, \ldots, q$.

Since the MA order is not generally known a priori, it is usually necessary to determine the MA order when the above approaches are used. From Eq. (92), one may determine the MA order by testing (8)

$$r(q) \neq 0, \quad \text{and} \quad r(q+m) \equiv 0, \quad \text{for} \quad m = 1, 2, \ldots \tag{96}$$

Given a finite-duration of data, $r(k)$ is replaced by the standard unbiased ACF estimate $\hat{r}(k)$ in this test, which is defined as

$$\hat{r}(k) = \frac{1}{N-k} \sum_{n=0}^{N-k-1} x^*(n) x(n+k), \quad k = 0, 1, \ldots \tag{97}$$

ARMA Spectral Estimation

According to the definition in Eq. (63), an ARMA signal is obtained by filtering a white noise process through a pole-zero system. ARMA models are suitable for describing signals whose spectra have both sharp peaks and deep nulls by relatively lower orders.

As we have seen in Eq. (72), the ARMA parameters appear in a nonlinear fashion through the unknown impulse response $h(n)$. If the optimum modeling is required, it is necessary to solve the least mean square solution of the highly nonlinear Yule-Walker equations. To obtain such a solution, nonlinear iterative techniques are employed, which not only are computationally expensive but also suffer from the local convergence.

A considerable simplicity in computation may be achieved via the suboptimal techniques in which the AR and MA part coefficients are estimated separately. With that, it is possible to estimate the AR parameters via a linear procedure. After the AR parameters are obtained, we may use the AR polynomial to filter the observed data and obtain a pure MA(q) process, whose parameters can be estimated via the approaches developed in the preceding subsection.

AR Parameter Estimation. Choosing $m \geq q + 1$ in Eq. (72), we obtain

$$\sum_{k=0}^{p} a(k) r(m-k) = 0, \quad m = q+1, q+2, \ldots \tag{98}$$

Eq. (98) establishes a linear relation between the AR parameters and the ACFs of the observed signals.

To determine the AR parameters, one may adopt the first p linear equations (i.e., $q + 1 \leq m \leq q + p$) and then solve the resultant system of equations. When the ACFs are truly known, this set of equations is enough to yield a unique and accurate solution to the AR parameter estimates. In practice, since the sample ACF estimates are used, the AR parameter estimates obtained by this method may be poor due to the estimation errors of the sample ACF estimates. This deficiency may also be interpreted by the fact that only subset lags of ACFs are used. In fact, Eq. (98) is satisfied for any $m \geq q + 1$. To obtain better AR parameter estimates, one reasonable choice is to employ more than the minimal number (i.e., p) of the extended Yule–Walker equations. This results in an overdetermined set of linear equations which can be solved via least square (LS) or total least square (TLS) techniques.

Suppose that the ACFs can be estimated up to lag $q + t$, where $t > p$. Then, we may write the following matrix equation from Eq. (98)

$$\begin{bmatrix} r(q) & r(q-1) & \cdots & r(q-p+1) \\ r(q+1) & r(q) & \cdots & r(q-p+2) \\ \vdots & \vdots & \vdots & \vdots \\ r(q+t-1) & r(q+t-2) & \cdots & r(q+t-p) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ a(p) \end{bmatrix} = - \begin{bmatrix} r(q+1) \\ r(q+2) \\ \vdots \\ r(q+t) \end{bmatrix} \tag{99}$$

or equivalently,

$$\mathbf{R}\mathbf{a} = -\mathbf{b} \tag{100}$$

Since $\mathbf{R}$ is of dimension $t \times p$ where $t > p$, the LS solution for the AR parameter estimates is given by

$$\mathbf{a} = -(\mathbf{R}^T \mathbf{R})^{-1} \mathbf{R}^T \mathbf{b} \tag{101}$$

The LS technique yields the AR parameter estimates by minimizing the errors in the vector **b**. In fact, both the matrix **R** and the vector **b** in Eq. (100) are with errors when their elements are replaced by the sample ACF estimates. The TLS algorithm minimizes these errors simultaneously and is usually implemented by numerically robust techniques, such as singular value decomposition (SVD), and hence, has better estimation performance than the LS algorithm. Another advantage of the TLS algorithm is that it can estimate the AR order using effective rank determination when the model orders are not known *a priori.*

In order to develop the TLS method for AR parameter estimation, we first rewrite Eq. (99) as follows:

$$\begin{bmatrix} r(q+1) & r(q) & \cdots & r(q-p+1) \\ r(q+2) & r(q+1) & \cdots & r(q-p+2) \\ \vdots & \vdots & \vdots & \vdots \\ r(q+t) & r(q+t-1) & \cdots & r(q+t-p) \end{bmatrix} \begin{bmatrix} 1 \\ a(1) \\ \vdots \\ a(p) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix} \tag{102}$$

In practice, although the ARMA model orders are not known a priori, it will not be difficult to select the initial model orders such that $p_1 \gg p$ and $q_1 \gg q$. Let us consider the extended order ARMA(p_1, q_1) model. In accordance with Eq. (102), the $t \times (p_1 + 1)$ extended-order autocorrelation matrix associated with this ARMA(p_1, q_1) model may be expressed as

$$\mathbf{R}_1 = \begin{bmatrix} r(q_1+1) & r(q_1) & \cdots & r(q_1-p_1+1) \\ r(q_1+2) & r(q_1+1) & \cdots & r(q_1-p_1+2) \\ \vdots & \vdots & \vdots & \vdots \\ r(q_1+t) & r(q_1+t-1) & \cdots & r(q_1+t-p_1) \end{bmatrix} \tag{103}$$

By Cadzow (9), if the autocorrelation lag entries used in the $t \times (p_1 + 1)$ matrix correspond to an ARMA(p_1, q_1) process for which $q_1 - p_1 \geq q - p$, then the rank of $\mathbf{R}_1$ equals to p, and the effective rank of $\hat{\mathbf{R}}_1$ will be p, where $\hat{\mathbf{R}}_1$ is $\mathbf{R}_1$ with $r(l)$ replaced by its sample estimate $\hat{r}(l)$. Thus, when sample ACF estimates are used, the order estimation problem is equivalent to matrix effective rank determination problem, which can be implemented by SVD technique. Once the order is determined, a truncated matrix is constructed, by which the AR parameter estimates with significant improvements can be obtained.

The TLS algorithm for AR parameter estimation and order determination is now summarized as follows.

Step 1. Compute the biased ACF estimates from the given data samples $x(n)$, $n = 0, 1, \ldots, N-1$, and take $t > p, p_1 \gg p, q_1 \gg q, q_1 - p_1 \geq q - p$ to construct the sample autocorrelation matrix $\hat{\mathbf{R}}_1$ using Eq. (103).

Step 2. Compute the SVD of $\hat{\mathbf{R}}_1$:

$$\hat{\mathbf{R}}_1 = \sum_{k=1}^{p_1+1} \lambda(k)\mathbf{u}_k\mathbf{v}_k^H = \mathbf{U}\Lambda\mathbf{V}^H \tag{104}$$

where $\mathbf{u}_k$ and $\mathbf{v}_k$ are the kth column vectors of the unitary matrices **U** and **V**, respectively; and the singular values $\lambda(k)$ are ordered such that

$$\lambda(1) \geq \lambda(2) \geq \cdots \geq \lambda(p_1+1) \geq 0 \tag{105}$$

Calculate the normalized ratio $\beta(k)$, which is given by

$$\beta(k) = \left[\frac{\lambda^2(1) + \lambda^2(2) + \cdots + \lambda^2(k)}{\lambda^2(1) + \lambda^2(2) + \cdots + \lambda^2(p_1+1)}\right]^{1/2} \tag{106}$$

The effective rank of $\hat{\mathbf{R}}_1$, denoted as p, is set to be equal to the smallest value of k for which $\beta(k)$ is deemed "adequately" close to one.

Step 3. Compute the $(p + 1) \times (p + 1)$ truncated matrix:

$$\mathbf{S}^{(p)} = \sum_{n=1}^{p} \sum_{k=1}^{p_1-p+1} \lambda^2(n)\mathbf{v}_n^k(\mathbf{v}_n^k)^H \tag{107}$$

where $\mathbf{v}_n^k$ denotes the $(p + 1) \times 1$ vector as specified by

$$\begin{gathered} \mathbf{v}_n^k = [v_n(k), v_n(k+1), \ldots, v_n(k+p)]^T \\ 1 \leq k \leq p_1 - p + 1, \quad 1 \leq n \leq p \end{gathered} \tag{108}$$

The AR coefficients are obtained by computing

$$\hat{a}(i) = \mathbf{s}_1(i+1)/\mathbf{s}_1(1), \quad i = 1, 2, \ldots, p \tag{109}$$

where $\mathbf{s}_1(k)$ is the kth element of the first column vector $\mathbf{s}_1$ of the inverse of $\mathbf{S}^{(p)}$.

MA Parameter Estimation. In order to complete the ARMA modeling, it is necessary to determine the model's MA parameters. Using the AR polynomial

$$\hat{A}(z) = 1 + \sum_{k=1}^{p} \hat{a}(k)z^{-k} \tag{110}$$

to filter the observed data $x(n)$, we obtain

$$v(n) = x(n) + \sum_{k=1}^{p} \hat{a}(k)x(n-k), \quad n = 0, 1, \ldots, N-1 \tag{111}$$

If $\hat{a}(k) = a(k)$, the filtered output $w(n)$ is an MA(q) process with parameters $b(k)$'s. In fact, $\hat{a}(k)$ is an estimate of $a(k)$; thus, $w(n)$ is an approximate MA(q) process. With the methods developed for pure MA spectral estimation, the MA parameters of the ARMA model can be estimated by using a nonlinear programming technique or Durbin's method. If only the spectral estimate is desirable, we may calculate the MA spectral estimate from the filtered output process $v(n)$; that is,

$$\hat{R}_v(\omega) = \sum_{m=-q}^{q} \hat{r}_v(m)e^{-j\omega m} \tag{112}$$

where $\hat{r}_v(m)$ is the sample ACF estimate of $w(n)$. Then, the estimated ARMA power spectrum of $x(n)$ is given by

$$\hat{R}(\omega) = \frac{\hat{R}_v(\omega)}{|1 + \sum_{k=1}^{p} \hat{a}(k)e^{-j\omega k}|^2} \tag{113}$$

ARMA Order Selection. As we have seen, the AR order estimation can be implemented by determining the effective rank of an extended-order autocorrelation matrix. With that,

Zhang and Zhang (10) proposed an SVD-based algorithm for MA order determination.

Considering the matrix

$$\mathbf{R}_2 = \begin{bmatrix} r(q_1-p) & r(q_1-p+1) & \cdots & r(q_1) \\ r(q_1-p+1) & r(q_1-p+2) & \cdots & r(q_1+1) \\ \cdots & \cdots & \cdots & \cdots \\ r(q_1) & r(q_1+1) & \cdots & r(q_1+p) \end{bmatrix} \quad (114)$$

It can be shown that $\text{rank}(\mathbf{R}_2) = p$ when $q_1 > q$, and $\text{rank}(\mathbf{R}_2) = p + 1$ only when $q_1 = q$. Based on this fact, the MA order q may be determined as follows: starting with $Q = q_1 > q$ and successively reducing Q by one, the SVD is used in order to determine the rank of $\mathbf{R}_2$; the first transition from rank p to $p + 1$ occurs at $Q = q$ which is the lag of sample ACF appearing in the antidiagonal elements of the matrix $\mathbf{R}_2$.

Another method (11) for ARMA order determination is to choose the orders by minimizing the following index

$$\text{AIC}(p, q) = \ln \hat{\sigma}_{\text{W}}^2(p, q) + \frac{2(p+q)}{N} \quad (115)$$

where $\hat{\sigma}_{\text{W}}^2(p, q)$ is an estimate of the variance of the linear predict error, and N is the data length.

PARAMETRIC METHODS FOR SINUSOIDAL SPECTRAL ESTIMATION

The principal difference between spectral estimation methods of the preceding section and those in this section is that in the preceding section we assume that the signal of interest has rational (or continuous) spectra, while in this section, we focus our attention on the signal consisting of sinusoidal components whose spectrum is discrete in frequency.

Suppose that the signal consists of p sinusoids with the form

$$x(n) = \sum_{k=1}^{p} \alpha_{\text{k}} \cos(\omega_{\text{k}} n + \phi_{\text{k}}) \quad (116)$$

for the real-valued case, or

$$x(n) = \sum_{k=1}^{p} \alpha_{\text{k}} e^{j(\omega_{\text{k}} n + \phi_{\text{k}})} \quad (117)$$

for the complex-valued case, where α_{k}, ω_{k}, and ϕ_{k} are the amplitude, normalized frequency, and the initial phase of the kth sinusoidal component. We assume that ϕ_{k}s are statistically independent random variables uniformly distributed on $[-\pi, \pi)$, which implies that the signal is zero-mean wide-sense stationary with autocorrelation function

$$r_{\text{x}}(m) = \frac{1}{2} \sum_{k=1}^{p} \alpha_{\text{k}}^2 \cos(2\omega_{\text{k}} m) \quad (118)$$

for real-valued sinusoids, and

$$r_{\text{x}}(m) = \sum_{k=1}^{p} |\alpha_{\text{k}}|^2 e^{j\omega_{\text{k}} m} \quad (119)$$

for complex-valued sinusoids. The observed process $y(n)$ is given by

$$y(n) = x(n) + w(n) \quad (120)$$

where $w(n)$ is a zero-mean, additive white noise process, which is statistically independent of the signal $x(n)$.

We focus on the estimation of the frequency parameters from the observed data $y(n)$, $n = 0, 1, \ldots, N - 1$. Once the frequencies have been determined, the estimation of other parameters, the associated amplitudes, and noise variance becomes a simple linear regression problem.

ARMA Modeling Approach

For the clarity of statement, we shall only consider the real-valued signals in developing the ARMA modeling approach for sinusoidal frequency estimation. We first prove that sinusoids in additive white noise satisfy a special ARMA model by which an ARMA modeling approach is developed for estimating the sinusoidal parameters.

To motivate the selection of an ARMA process as the appropriate model for sinusoids in white noise, let us consider the following trigonometric identity:

$$\cos(\Omega n) = -2 \cos \Omega \cos[\Omega(n-1)] - \cos[\Omega(n-2)] \quad (121)$$

for $-\pi \le \Omega \le \pi$. Let $x(n) = \cos \Omega n$, $a(1) = 2 \cos(\Omega)$, and $a(2) = 1$; the single real sinusoidal component $x(n)$ can be generated via the second order difference equation

$$x(n) = -a(1)x(n-1) - a(2)x(n-2) \quad (122)$$

with the initial values to be $x(-1) = -1$, $x(-2) = 0$. This difference equation has the characteristics polynomial

$$1 + a(1)z^{-1} + a(2)z^{-2} \quad (123)$$

whose roots are $z_1 = e^{j\Omega}$ and $z_2 = z_1^* = e^{-j\Omega}$. The sinusoidal frequency is determined from the roots as follows:

$$\Omega = \tan^{-1}(\text{Im}\{z_1\}/\text{Re}\{z_1\}) \quad (124)$$

Eq. (122) is the limiting case of an AR(2) process in which the driving noise variance tends to be zero, and the poles tend to lie on the unit circle. Also, with only two coefficients and two successive samples, we may use Eq. (122) to perfectly predict the sinusoidal values at all time.

In general, a signal consisting of p real sinusoidal components satisfies the following $2p$th-order difference equation

$$\sum_{k=0}^{2p} a(k)x(n-k) = 0 \quad (125)$$

where $a(0) = 1$. The associated characteristics polynomial is given by

$$A(z) = \sum_{k=0}^{2p} a(k)z^{-k} \quad (126)$$

Note the roots of the above equation have unit modulus and occur in complex conjugate pairs whose phases are related to

the p sinusoidal frequencies. Specifically, let z_i, $i = 1, 2, \ldots, 2p$ be the roots of $A(z)$; the frequencies are determined by

$$\Omega_i = \tan^{-1}(\mathrm{Im}\{z_i\}/\mathrm{Re}\{z_i\}), \quad i = 1, 2, \ldots, p \tag{127}$$

For sinusoids in additive white noise, substituting $x(n) = y(n) - w(n)$ into Eq. (125) yields

$$\sum_{k=0}^{2p} a(k)y(n-k) = \sum_{k=0}^{2p} a(k)w(n-k) \tag{128}$$

Eq. (128) is a special ARMA($2p$, $2p$) model in which both the AR and MA parameters are identical, and all the poles and zeros are located exactly on the unit circle. Note that the frequency information of the signal is completely contained in the characteristics polynomial constructed via the AR coefficients of Eq. (128).

To establish the relation between the AR coefficients and the ACFs of the observed process $y(n)$, we multiply Eq. (128) by $y(n - l)$ and take the expection; it follows that

$$\sum_{k=0}^{2p} a(k)r_y(l-k) = \sum_{k=0}^{2p} a(k)r_w(l-k) = \sigma^2 a(l) \tag{129}$$

where we have used the assumption that $x(n)$ and $w(n)$ are zero-mean and statistically independent, and that $E[w(n) \times w(n + k)] = \sigma^2\delta(k)$. Since $a(l) = 0$ for $l < 0$ and $l > 2p$, we have

$$\sum_{k=0}^{2p} a(k)r_y(l-k) = 0, \quad l > 2p \tag{130}$$

We refer to Eq. (130) as the Higher-Order Yule–Walker (HOYW) equation, which is the basis of the ARMA modeling approach for sinusoidal parameter estimation.

One algorithm for AR parameter estimation is to adopt the first $2p$ equations by setting $l = 2p + 1, 2p + 2, \ldots, 4p$ in Eq. (130). In practice, the sample ACF estimates are employed in place of the true ACFs, which result in some errors in estimating the AR parameters. Similar to the general ARMA modeling case, we may construct an overdetermined system of equations and then use the LS or TLS techniques to estimate the AR parameters.

Setting $l = 2p + 1, 2p + 2, \ldots, t$ ($t > 4p$) in Eq. (130), we have the following matrix equation:

$$\begin{bmatrix} r_y(2p) & r_y(2p-1) & \cdots & r_y(1) \\ r_y(2p+1) & r_y(2p) & \cdots & r_y(2) \\ \vdots & \vdots & \vdots & \vdots \\ r_y(t-1) & r_y(t-1) & \cdots & r_y(t-2p) \end{bmatrix} \begin{bmatrix} a(1) \\ a(2) \\ \vdots \\ a(p) \end{bmatrix} = -\begin{bmatrix} r_y(2p+1) \\ r_y(2p+2) \\ \vdots \\ r_y(t) \end{bmatrix} \tag{131}$$

or

$$\mathbf{Ra} = -\mathbf{b} \tag{132}$$

The LS solution to Eq. (132) is given by

$$\mathbf{a} = -(\mathbf{R}^T\mathbf{R})^{-1}\mathbf{R}^T\mathbf{b} \tag{133}$$

The TLS algorithm can also be used for estimating the AR coefficients of the special ARMA($2p$, $2p$) process. The TLS algorithm considers both errors in $\mathbf{R}$ and $\mathbf{b}$; in addition, it yields not only the AR parameter estimates but also the AR order determination. Thus, the TLS algorithm has better performance than the LS algorithm. The detailed steps of the TLS algorithm can be referred to as those for general ARMA modeling.

ARMA modeling approach is summarized as follows.

Step 1. Compute the sample ACF estimates $\{\hat{r}_y(m), m = 1, 2, \ldots, t\}$ where $t > 4p$, from the observations $y(n)$, $n = 0, 1, \ldots, N - 1$.

Step 2. Estimate the AR coefficients using LS or TLS algorithms.

Step 3. Compute the roots of Eq. (126), and obtain the frequency estimates as the angular positions of these roots.

Once the frequency parameters have been determined, the associated amplitude parameters and the noise variance can be estimated as follows. For p real-valued sinusoids in additive white noise, the ACFs of the observed process $y(n)$ are

$$r_y(0) = \sigma^2 + \sum_{k=1}^{p} P_k \tag{134}$$

$$r_y(m) = \sum_{k=1}^{p} P_k \cos(\omega_k m), \quad \text{for} \quad m \neq 0 \tag{135}$$

where $P_k = \alpha_k^2 / 2$ is the average power of the kth sinusoid. Evaluating the above equation at $m = 1, 2, \ldots, t$ ($t \geq p$), we may obtain the LS estimates of the powers of the sinusoids with ω_k replaced by $\hat{\omega}_k$ and $r_y(m)$ by $\hat{r}_y(m)$. Once the powers are known, the noise variance can be estimated by

$$\hat{\sigma}^2 = \hat{r}_y(0) - \sum_{k=1}^{p} \hat{P}_k \tag{136}$$

Pisarenko Method

In Eq. (129), if we choose $l = 0, 1, \ldots, 2p$, it follows that

$$\mathbf{Ra} = \sigma^2\mathbf{a} \tag{137}$$

where

$$\mathbf{R} = \begin{bmatrix} r_y(0) & r_y(1) & \ldots & r_y(-2p) \\ r_y(1) & r_y(0) & \ldots & r_y(-2p+1) \\ \vdots & \vdots & \vdots & \vdots \\ r_y(2p) & r_y(2p-1) & \ldots & r_y(0) \end{bmatrix} \tag{138}$$

and

$$\mathbf{a} = [a(0), a(1), \ldots, a(2p)]^T \tag{139}$$

Eq. (139) is an eigenequation in which the noise variance σ^2 is an eigenvvalue of the autocorrelation matrix $\mathbf{R}$; while AR

parameter vector $\mathbf{a}$ is the eigenvector associated with the eigenvalue σ^2, scaled so that the first element is unity. It may be shown (12) that for a process consisting of p real-value sinusoids in additive white noise, the variance σ^2 corresponds to the minimum eigenvalue of $\mathbf{R}$ when its dimension is $(2p + 1) \times (2p + 1)$.

Pisarenko harmonic decomposition method (12) is based upon the above eigenanalysis. It can be summarized as follows.

Step 1. Compute the sample ACF estimates $\{\hat{r}_y(m), m = 0, 1, \ldots, 2p\}$, and construct the autocorrelation matrix $\hat{\mathbf{R}}$ using Eq. (138) where $r_y(m)$ is replaced by $\hat{r}_y(m)$.

Step 2. Find the minimum eigenvalue and the corresponding eigenvector, thus the AR coefficients of Eq. (128).

Step 3. Compute the roots of the AR polynomial and obtain the frequence estimates.

MUSIC Method

The Pisarenko method is an eigendecomposition technique which employs the eigenstructure of the autocorrelation matrix $\mathbf{R}$ with dimension $(2p + 1) \times (2p + 1)$. When the signal consists of p complex-valued sinusoids, the matrix dimension is $(p + 1) \times (p + 1)$. In the sequel, we shall explore the eigenstructure properties of the general autocorrelation matrix with higher dimension by which the multiple signal classification (MUSIC) method (13) is developed for sinusoidal parameter estimation.

For mathematical convenience, we now consider the complex-valued sinusoids as assumed in Eq. (117). Take $m > p$, and let

$$\mathbf{A} = [\mathbf{a}(\omega_1), \mathbf{a}(\omega_2), \ldots, \mathbf{a}(\omega_p)] \tag{140}$$

with

$$\mathbf{a}(\omega_i) = [1, e^{j\omega_i}, \ldots, e^{j(m-1)\omega_i}]^T, \quad \text{for} \quad i = 1, 2, \ldots, p \tag{141}$$

and denote

$$\mathbf{y}(n) = [y(n), y(n+1), \ldots, y(n+m-1)]^T \tag{142}$$

$$\mathbf{x}(n) = [\alpha_1 e^{j(\omega_1 n + \phi_1)}, \alpha_2 e^{j(\omega_2 n + \phi_2)}, \ldots, \alpha_p e^{j(\omega_p n + \phi_p)}]^T \tag{143}$$

$$\mathbf{w}(n) = [w(n), w(n+1), \ldots, w(n+m-1)]^T \tag{144}$$

Then, along with Eqs. (117) and (120), $\mathbf{y}(n)$ can be expressed as

$$\mathbf{y}(n) = \mathbf{A}\mathbf{x}(n) + \mathbf{w}(n) \tag{145}$$

Note $\mathbf{A}$ is a Vandermonde matrix which has the property $\text{rank}(\mathbf{A}) = p$ if $m \geq p$, and $\omega_i \neq \omega_j$ for $i \neq j$.

The autocorrelation matrix of $\mathbf{y}(n)$ is

$$\mathbf{R} = E[\mathbf{y}(n)\mathbf{y}^H(n)] = \mathbf{A}\mathbf{P}\mathbf{A}^H + \sigma^2\mathbf{I} \tag{146}$$

where $\mathbf{P} = \text{diag}(|\alpha_1|^2, |\alpha_2|^2, \ldots, |\alpha_p|^2)$, and $\mathbf{I}$ is the $m \times m$ identity matrix. On the other hand, direct calculation of the autocorrelation matrix yields

$$\mathbf{R} = \begin{bmatrix} r_y(0) & r_y(1) & \cdots & r_y(m-1) \\ r_y^*(1) & r_y(0) & \cdots & r_y(m-2) \\ \vdots & \vdots & \vdots & \vdots \\ r_y^*(m-1) & r_y^*(m-2) & \cdots & r_y(0) \end{bmatrix} \tag{147}$$

Let $\lambda(1) \geq \lambda(2) \geq \ldots \geq \lambda(m)$ denote the eigenvalues of $\mathbf{R}$, and let the corresponding eigenvectors be denoted as $\{\mathbf{v}(i), i = 1, 2, \ldots, m\}$. Since

$$\text{rank}(\mathbf{A}\mathbf{P}\mathbf{A}^H) = p \tag{148}$$

it follows that $\mathbf{A}\mathbf{P}\mathbf{A}^H$ has p strictly positive eigenvalues $\tilde{\lambda}(1) \geq \tilde{\lambda}(2) \geq \ldots \geq \tilde{\lambda}(p) > 0$, and $(m - p)$ zero eigenvalues. Hence, performing eigendecomposition on $\mathbf{R}$ yields

$$\mathbf{R} = \sum_{i=1}^{p} [\tilde{\lambda}(i) + \sigma^2]\mathbf{v}(i)\mathbf{v}^H(i) + \sum_{j=p+1}^{m} \sigma^2 \mathbf{v}(j)\mathbf{v}^H(j) \tag{149}$$

We may split the eigenvalues of $\mathbf{R}$ into two subsets.

$$\lambda(i) = \tilde{\lambda}(i) + \sigma^2 > \sigma^2, \quad \text{for} \quad i = 1, 2, \ldots, p \tag{150}$$

and

$$\lambda(i) = \sigma^2, \quad \text{for} \quad i = p+1, p+2, \ldots, m \tag{151}$$

Similarly, the corresponding eigenvectors can be separated into two subsets:

$$\mathbf{S} = [\mathbf{v}(1), \mathbf{v}(2), \ldots, \mathbf{v}(p)] \tag{152}$$

and

$$\mathbf{G} = [\mathbf{v}(p+1), \mathbf{v}(p+2), \ldots, \mathbf{v}(m)] \tag{153}$$

From the definition of eigendecomposition, it follows that

$$\mathbf{S}^H\mathbf{G} = \mathbf{0} \tag{154}$$

and

$$\mathbf{R}\mathbf{G} = \sigma^2\mathbf{G} \tag{155}$$

On the other hand, from Eq. (146), we have

$$\mathbf{R}\mathbf{G} = \mathbf{A}\mathbf{P}\mathbf{A}^H\mathbf{G} + \sigma^2\mathbf{G} \tag{156}$$

Combination of Eq. (155) and Eq. (156) yields

$$\mathbf{A}^H\mathbf{G} = \mathbf{0} \tag{157}$$

where we have used the fact that the matrix $\mathbf{A}\mathbf{P}$ has full column rank. Eq. (157) shows that the columns $\{\mathbf{v}(i), i = p + 1, p + 2, \ldots, m\}$ of $\mathbf{G}$ belong to the null space of $\mathbf{A}^H$, or called noise subspace; while the combination of Eq. (154) and Eq. (157) implies that the columns $\{\mathbf{v}(i), i = 1, 2, \ldots, p\}$ of $\mathbf{S}$, which are the principle eigenvectors, span the signal subspace.

The multiple signal classification (MUSIC) method employs the noise subspace information to estimate the fre-

quency parameters. Specifically, along with Eqs. (140), (141), and (157), we obtain

$$\mathbf{a}^H(\omega_i)\mathbf{G}\mathbf{G}^H\mathbf{a}(\omega_i) = 0, \quad i = 1, 2, \ldots, p \tag{158}$$

which means that the true frequency values ω_k, $k = 1, 2, \ldots, p$ are the solutions of the equation

$$P(\omega) = \mathbf{a}^H(\omega)\mathbf{G}\mathbf{G}^H\mathbf{a}(\omega) = 0, \quad \text{for any} \quad m > p \tag{159}$$

On the other hand, it has been proved that ω_k, $k = 1, 2, \ldots, p$ are the only solutions to Eq. (158). Hence, the reciprocal of $P(\omega)$ has sharp peaks at the true frequencies ω_ks.

The MUSIC method is summarized as follows.

Step 1. From the observed data $y(n)$, $n = 0, 1, \ldots, N-1$, compute the sample ACF estimates $\hat{r}_y(m)$, $m = 0, 1, \ldots, m-1$, and form the autocorrelation matrix $\hat{\mathbf{R}}$.

Step 2. Perform eigendecomposition on matrix $\hat{\mathbf{R}}$, and obtain the estimates of $\mathbf{S}$ and $\mathbf{G}$, denoted as $\hat{\mathbf{S}}$ and $\hat{\mathbf{G}}$, respectively.

Step 3. Determine frequency estimates by locating the p highest peaks of the function

$$\frac{1}{\mathbf{a}^H(\omega)\hat{\mathbf{G}}\hat{\mathbf{G}}^H\mathbf{a}(\omega)}, \quad -\pi \le \omega \le \pi \tag{160}$$

The MUSIC method is an extension of the Pisarenko eigendecomposition method. Specifically, when $m = p + 1$, the MUSIC method is reduced to the Pisarenko method (here, we consider the p complex-valued sinusoids). For the Pisarenko method, the involved matrix produces only one noise eigenvector which can be used for estimating the sinusoidal frequencies. However, the MUSIC method forms an $m \times m$ ($m > p + 1$) matrix which contains more information about the ACFs of the observed data and yields $(m - p)$ noise eigenvectors which are all useful for estimating the sinusoidal frequencies. Hence, the MUSIC method yields more accurate frequency estimates than Pisarenko does, especially when $m \gg p + 1$ is chosen.

ESPRIT Method

ESPRIT (Estimation of Signal Parameters via Rotational Invariance Techniques) (14,15) is another eigendecomposition method for estimating sinusoidal frequency parameters. It yields the sinusoidal frequency estimates by computing the generized eigenvalues of two well-constructed matrices.

We again consider the complex-valued case. Using the notations defined in the MUSIC method, and denoting

$$\mathbf{z}(n) = [y(n+1), y(n+2), \ldots, y(n+m)]^T \tag{161}$$

we have, along with Eqs. (117) and (120),

$$\mathbf{z}(n) = \mathbf{A}\mathbf{\Phi}\mathbf{x}(n) + \mathbf{w}(n+1) \tag{162}$$

where $\mathbf{\Phi}$ is a $p \times p$ diagonal matrix

$$\mathbf{\Phi} = \text{diag}[e^{j\omega_1}, e^{j\omega_2}, e^{j\omega_p}] \tag{163}$$

which relates the time-displaced vector $\mathbf{y}(n)$ and $\mathbf{z}(n)$, and hence, is called a rotation operator.

The cross-correlation matrix of the data vectors $\mathbf{y}(n)$ and $\mathbf{z}(n)$ is

$$\mathbf{R}_1 = E[\mathbf{y}(n)\mathbf{z}^H(n)] = \mathbf{A}\mathbf{P}\mathbf{\Phi}^H\mathbf{A}^H + \sigma^2\mathbf{Q} \tag{164}$$

where

$$\mathbf{Q} = \begin{bmatrix} 0 & 0 & 0 & \cdots & 0 & 0 \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \ddots & \ddots & \ddots & \cdots & \ddots & \ddots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{bmatrix} \tag{165}$$

On the other hand, direct calculation of $\mathbf{R}_1$ yields

$$\mathbf{R}_1 = \begin{bmatrix} r_y(1) & r_y(0) & \cdots & r_y(m) \\ r_y(0) & r_y(1) & \cdots & r_y(m-1) \\ \vdots & \vdots & \vdots & \vdots \\ r_y^*(m-2) & r_y^*(m-3) & \cdots & r_y(1) \end{bmatrix} \tag{166}$$

Let us construct the following two matrices:

$$\mathbf{C}_1 \triangleq \mathbf{R} - \sigma^2\mathbf{I} = \mathbf{A}\mathbf{P}\mathbf{A}^H \tag{167}$$

$$\mathbf{C}_2 \triangleq \mathbf{R}_1 - \sigma^2\mathbf{Q} = \mathbf{A}\mathbf{P}\mathbf{\Phi}^H\mathbf{A}^H \tag{168}$$

and consider the matrix $(\mathbf{C}_1 - \lambda\mathbf{C}_2)$; that is,

$$\mathbf{C}_2 - \lambda\mathbf{C}_1 = \mathbf{A}\mathbf{P}(\mathbf{I} - \lambda\mathbf{\Phi}^H)\mathbf{A}^H \tag{169}$$

Paularj, Roy, and Kailath (14) have shown that matrix pair $(\mathbf{C}_1, \mathbf{C}_2)$ has p generized eigenvalues at $\lambda(i) = e^{j\omega_i}$, $i = 1, 2, \ldots, p$, and $(m - p)$ generized eigenvalues being zero.

Using the above results, we may summarize the ESPRIT algorithm as follows.

Step 1. Calculate the sample ACFs $\hat{r}_y(m)$, $m = 0, 1, \ldots, m$ using a standard biased formula, and construct the matrices $\mathbf{R}$ and $\mathbf{R}_1$ using Eqs. (147) and (166).

Step 2. Compute the eigenvalues of $\mathbf{R}$, and obtain the estimate of noise variance $\hat{\sigma}^2$.

Step 3. Compute $\hat{\mathbf{C}}_1 = \mathbf{R} - \hat{\sigma}^2\mathbf{I}$ and $\hat{\mathbf{C}}_2 = \mathbf{R}_1 - \hat{\sigma}^2\mathbf{Q}$

Step 4. Compute the generized eigenvalues of the matrix pair $(\hat{\mathbf{C}}_1, \hat{\mathbf{C}}_2)$. The p generized eigenvalues which lie on (or near) the unit circle determine the digonal elements of $\mathbf{\Phi}$, and hence, the sinusoidal frequencies. The remaining $(m - p)$ eigenvalues will lie at (or near) the origin.

Sinusoidal Number Determination

From Eqs. (150) and (151), if there are p complex-valued sinusoidal components, performing eigendecomposition on the autocorrelation matrix $\mathbf{R}$ yields p eigenvalues which are larger than the noise variance and $(m - p)$ eigenvalues which are equal to σ^2. Based on this observation, the sinusoidal number can be determined by comparing the eigenvalues with a specific threshold and calculating the number of eigenvalues which are larger than the threshold. Obviously, the selected threshold is the key parameter within this method.

Another approach is based on the extension of the AIC information criterion, which was proposed by Wax and Kailath (16). One chooses the sinusoidal number p by minimizing the function:

$$\mathrm{MDL}(k) = -\log\left[\frac{G(k)}{Q(k)}\right]^N + E(k), \quad k = 0, 1, \ldots, m-1 \tag{170}$$

where

$$G(k) = \prod_{i=k+1}^{m} \lambda(i) \tag{171}$$

$$Q(k) = \left[\frac{1}{m-k}\sum_{i=k+1}^{m} \lambda(i)\right]^{m-k} \tag{172}$$

and

$$E(k) = \frac{1}{2}p(2m-k)\log N \tag{173}$$

with $\lambda(1) \geq \lambda(2) \geq \cdots \geq \lambda(m)$ being the eigenvalues of $\mathbf{R}$, and N being the number of data samples.

APPLICATIONS TO TIME-VARYING SIGNAL SPECTRAL ANALYSIS

In this section, we discuss two problems which are closely related to the spectral analysis problem: damped sinusoidal parameter estimation and instantaneous frequency measurement. The former problem is frequently encountered in magnetic resonance spectroscopy and radioastronomy; the latter one may be found in vibration measurements, doppler radar returns, geophysical processing, and surveillance observations of the electromagnetic spectrum. In both problems, the signals are nonstationary; hence, the approaches developed earlier are not applicable directly.

Damped Sinusoidal Parameter Estimation

A sequence $x(n)$ consisting of p damped sinusoidal signals can be expressed as

$$x(n) = \sum_{m=1}^{p} b_m z_m^n \tag{174}$$

where $b_m = A_m \exp(j\theta_m)$ is nonzero amplitude, $z_m = \exp(-\alpha_m + j\omega_m)$ and $\alpha_m > 0$, and $-\pi \leq \omega_m \leq \pi$ for $k = 1, 2, \ldots, p$. α_m is called the *damping factor* of the damped sinusoid with angular frequency ω_m. The larger the damping factor, the faster the amplitude of the sinusoid decays. The observed sequence $y(n)$ is given by

$$y(n) = x(n) + w(n) \tag{175}$$

where $w(n)$ is additive measurement noise, which is assumed to be a complex white Gaussian process. The problem of interest is to estimate the parameters $\{A_m, \theta_m, \alpha_m, \omega_m\}$ and p from finite data $y(n)$, $n = 0, 1, \ldots, N-1$.

The optimum solution to the unknown parameter estimation is based on minimizing the squared error

$$\epsilon = \sum_{n=0}^{N-1} |y(n) - x(n)|^2 \tag{176}$$

which is a difficult nonlinear least-squares problem. First, the computation will be very expensive since there are $4p$ unknown parameters. Second, the solution involves an iterative process in which a good initial guess of the unknown parameters is required; otherwise, the algorithm may not converge or may converge to the wrong solution. Based on this fact, some suboptimal methods, such as Prony's method, KT method, and MKT method, are developed which do not minimize Eq. (176) but still may provide satisfactory results.

Prony's Method. Define polynomial $A(z)$ as

$$A(z) = \prod_{k=1}^{p}(z - z_k) = \sum_{i=0}^{p} a(i) z^{p-i}, \quad a(0) = 1 \tag{177}$$

It has been shown that $x(n)$ satisfies the following difference equation

$$x(n) = -\sum_{m=1}^{p} a(m)x(n-m), \quad n = p, p+1, \ldots, N-1 \tag{178}$$

Note that the complex exponentials z_ks are the roots of the polynomial $A(z)$; hence, in order to estimate z_ks, we may first estimate the coefficients $a(m)$s. In doing so, substituting $x(n) = y(n) - w(n)$ into Eq. (178) yields

$$\begin{aligned} y(n) &= -\sum_{m=1}^{p} a(m)x(n-m) + w(n) \\ &= -\sum_{m=1}^{p} a(m)y(n-m) + \sum_{m=0}^{p} a(m)w(n-m) \end{aligned} \tag{179}$$

for $p \leq n \leq N-1$. Equation (179) represents the sum of damped sinusoids in additive noise in terms of the noise $w(n)$ and the observed sequence $y(n)$. Also, it is a special ARMA model with identical AR and MA parameters driven by noise process $w(n)$. Unlike Eq. (128), the coefficients $a(m)$s here are not constrained to produce polynomial roots of unit modulus (no damping).

The least-squares solution to the parameters $a(m)$s can be obtained by minimizing $\sum_{n=p}^{N-1} |w(n)|^2$. However, this procedure again leads to a set of nonlinear equations. An alternative method is to minimize the term

$$\begin{aligned} J &= \sum_{n=p}^{N-1} \left| \sum_{m=0}^{p} a(m)e(n-m) \right|^2 \\ &= \sum_{n=p}^{N-1} \left| y(n) + \sum_{m=1}^{p} a(m)y(n-m) \right|^2 \end{aligned} \tag{180}$$

which can be solved by a well-established least-squares algorithm.

Once $a(i)$s are determined, b_ms can be estimated by minimizing

$$Q = \sum_{n=0}^{N-1} \left| y(n) - \sum_{m=1}^{p} b_m \hat{z}_m^n \right|^2 \tag{181}$$

Finally, the ampitude A_i, phase θ_i, damping factor α_i, and the frequency ω_i can be computed as follows:

$$A_i = |b_i|, \quad \theta_i = \tan^{-1}[\mathrm{Im}(b_i)/\mathrm{Re}(b_i)] \tag{182}$$

$$\alpha_i = \ln |z_i|, \quad \omega_i = \tan^{-1}[\mathrm{Im}(z_i)/\mathrm{Re}(z_i)] \tag{183}$$

Kumaresan–Tufts (KT) Method. The Prony's method is conceptually simple and computationally efficient; however, it is highly sensitive to additive measurement noise. In fact, this method is statistically inefficient in the sense that the variances of the estimated parameters often exceed the Cramer-Rao (CR) lower bound.

Kumaresan and Tufts (17) proposed a method which has considerable performance improvements over Prony's method, at the expense of greater computational complexity. The KT method has the following features: (1) it uses an overdetermined set of linear equations and overestimates the order of the assumed linear model; (2) the solution of the linear set of equations involves singular value decomposition (SVD) of the data matrix, followed by a truncation of the set of the singular values; (3) the backward predictor polynomial is estimated, and its roots are used to determine the frequency parameters.

Let $\mathbf{A}$ and $\mathbf{h}$ denote the $(N - L) \times L$ $[\min(N - L, L) \geq p]$ conjugate backward prediction matrix and the $(N - L)$-component column vector, respectively

$$\mathbf{A} = \begin{bmatrix} y^*(1) & y^*(2) & \cdots & y^*(L) \\ y^*(2) & y^*(3) & \cdots & y^*(L+1) \\ \vdots & \vdots & \vdots & \cdots \\ y^*(N-L) & y^*(N-L+1) & \cdots & y^*(N-1) \end{bmatrix} \tag{184}$$

$$\mathbf{h} = [y^*(0), y^*(1), \ldots, y^*(N-L-1)]^T \tag{185}$$

The KT method is to find an L-component prediction vector $\mathbf{c} = [c(1), c(2), \ldots, c(L)]^T$ such that

$$\mathbf{Ac} \approx -\mathbf{h} \tag{186}$$

Then, z_ms can be estimated by calculating the roots of the prediction polynomial

$$C(z) = 1 + c(1)z^{-1} + \cdots + c(L)z^{-L} \tag{187}$$

To estimate $\mathbf{c}$, the optimum rank p approximation of $\mathbf{A}$ is first constructed by

$$\hat{\mathbf{A}} = \sum_{k=1}^{p} \lambda(k)\mathbf{u}(k)\mathbf{v}^H(k) \tag{188}$$

where $\{\lambda(1), \lambda(2), \ldots, \lambda(p)\}$ are the p largest singular values of $\mathbf{A}$, and $\mathbf{u}(k)$ and $\mathbf{v}(k)$ are the left right singular vector of $\mathbf{A}$ corresponding to the singular value $\lambda(k)$, respectively. To make the matrix equation

$$\hat{\mathbf{A}}\mathbf{c} = -\mathbf{h} \tag{189}$$

to have a solution, either $\mathbf{h}$ must be in the span $\{\mathbf{u}(1), \mathbf{u}(2), \ldots, \mathbf{u}(p)\}$ or $\hat{\mathbf{h}}$, which is the projection of $\mathbf{h}$ on span $\{\mathbf{u}(1), \mathbf{u}(2), \ldots, \mathbf{u}(p)\}$, must be used instead of $\mathbf{h}$ in Eq. (189). In either case, Eq. (189) can be written as

$$\hat{\mathbf{A}}\mathbf{c} = -\hat{\mathbf{h}} \tag{190}$$

where

$$\hat{\boldsymbol{h}} = \sum_{k=1}^{p} [\mathbf{u}^H(k)\mathbf{h}]\mathbf{u}(k) \tag{191}$$

Since $\mathrm{rank}(\hat{\mathbf{A}}) = p \leq L$, Eq. (190) is an underdetermined system of equation about $\mathbf{c}$, and there are multiple solutions. The solution minimizing $\|\mathbf{c}\|$ is given by

$$\mathbf{c} = -\sum_{k=1}^{p} \frac{1}{\lambda(k)} [\mathbf{u}^H(k)\mathbf{h}]\mathbf{v}(k) \tag{192}$$

It has been proven in (17) that if $\mathbf{c}$ is estimated using Eq. (192), then the prediction polynomial $C(z)$ has p zeros outside the unit circle, which are the reciprocals of z_i for $i = 1, 2, \ldots, p$, and $(L - p)$ zeros inside the unit circle. By means of this property, the zeros outside the unit circle uniquely determine the parameters α_ms and ω_ms.

MKT Method. The KT method uses the low-rank matrix approximation to reduce the noise effect and, thus, has much better performance than Prony's method. Indeed, when the SNR is high and enough data are available, the rank approximation in the KT method will reduce the measurement noise effect significantly; hence, the KT method in this case will almost attain the CR bound. However, if the SNR is reduced to a certain degree, the rank approximation in the KT method is unable to reduce the noise effect efficiently, and moreover, the noise threshold appears.

On the basis of the KT method, Li, Liu, and Razavilar (18) proposed a modified KT (MKT) algorithm which exploits not only the rank-deficient property but also the Hankel property of the prediction matrix.

In fact, if a data sequence $x(n)$ consists of p distinct sinusoids, as in Eq. (174), then for any L $(L \geq p)$, the $L \times L$ prediction matrix $\mathbf{P}_L = \{x(i + j)\}_{i,j=0}^{L-1}$ is a singular Hankel matrix with rank p and full rank $p \times p$ principle minor $\mathbf{P}_p = \{x(i + j)\}_{i,j=0}^{p-1}$. Conversely, for any $L \times L$ singular matrix $\mathbf{P}_L = \{x(i + j)\}_{i,j=0}^{L-1}$ with rank p, if its $p \times p$ principle minor $\mathbf{P}_p = \{x(i + j)\}_{i,j=0}^{p-1}$ is full rank, then $x(n)$ for $n = 0, 1, \ldots, (2L - 2)$ can be uniquely expressed as the summation of p distinct sinusoids as given by Eq. (174). These observations reveal a one-to-one correspondence between a data sequence consisting of damped sinusoids and a rank-deficient Hankel matrix. Therefore, parameter estimation of damped sinusoidal signals from noisy data is equivalent to performing the low-rank Hankel matrix approximation.

The MKT algorithm is summarized as follows:

Step 1. Choose L, and form the square prediction matrix $\mathbf{P}_L = \{y(i + j)\}_{i,j=0}^{L-1}$.

Step 2. Find rank-p approximation matrix $\overline{\mathbf{P}}_L = \{\overline{y}(i, j)\}_{i,j=0}^{L-1} = \sum_{k=1}^{p} \lambda(k)\mathbf{u}(k)\mathbf{v}^H(k)$, where $\lambda(k)$ for $k = 1, 2, \ldots, p$ are the p largest singular values of $\mathbf{P}_L$, and $\mathbf{u}(k)$ and $\mathbf{v}(k)$ are corresponding left and right singular vectors, respectively.

Step 3. Find a Hankel matrix $\hat{\mathbf{P}}_L = \{\hat{y}(i+j)\}_{i,j=0}^{L-1}$ to approximate the rank-p prediction matrix $\overline{\mathbf{P}}_L$.

Step 4. Repeat Steps 2 and 3 to get an estimation of $\hat{y}(n)$ from $\hat{\mathbf{P}}_L$, where in the repeation, $\mathbf{P}_L$ is replaced by $\hat{\mathbf{P}}_L$.

Step 5. Estimate the parameters using the KT algorithm to $\hat{y}(n)$.

As compared with the original KT algorithm which ignores the Hankel property of the prediction matrix, the MKT algorithm has a lower noise threshold and can estimate the parameters of signals with larger damping factors.

Instantaneous Frequency Measurement

In this subsection, we are concerned with the problem of instantaneous frequency measurement, that is, estimating the frequency content of a sinusoidal signal with time-varying frequencies or a narrow-band signal with time varying power spectrum.

As we know, AR models are usually used to represent signals with a narrow-band spectrum. Accordingly, a narrow-band signal with time-varying spectrum can be characterized by a time-varying AR model,

$$x(n) = -\sum_{k=1}^{2p} a(n;k)x(n-k) + w(n) \tag{193}$$

On the other hand, a signal consisting of p time-varying sinusoidal components satisfies the following prediction equation:

$$x(n) = -\sum_{k=1}^{2p} a(n;k)x(n-k) \tag{194}$$

where $a(n;\ k)$s are the time-varying model parameters, and $w(n)$ in Eq. (193) is a zero-mean white noise process with time-varying variance $\sigma^2(n)$. In both cases, the instantaneous frequencies are determined by locating the peak positions of the modified spectrum, which is defined as

$$P(n;\omega) = \frac{1}{|1 + \sum_{k=1}^{2p} a(n;k)e^{-j\omega k}|^2} \tag{195}$$

To measure the instantaneous frequencies, we need to design an adaptive algorithm, such as a least-mean-square (LMS) or recursive least square (RLS) algorithm, for tracking the model parameters at each time instant, say $\hat{a}(n;\ k)$. Then, the instantaneous frequencies at each time instant are computed by determining the peak positions of the following function

$$\hat{P}(n;\omega) = \frac{1}{|1 + \sum_{k=1}^{2p} \hat{a}(n;k)e^{-j\omega k}|^2} \tag{196}$$

LMS Algorithm. Define the prediction equation

$$\hat{x}(n) = -\sum_{k=1}^{2p} \hat{a}(n;k)x(n-k) \tag{197}$$

The LMS algorithm updates the coefficients $\hat{a}(n;\ k)$ by minimizing the mean square error:

$$J_1 = E[\epsilon_1^2(n)] \tag{198}$$

where $\epsilon_1(n)$ is prediction error at time n:

$$\epsilon_1(n) = x(n) - \hat{x}(n) \tag{199}$$

The LMS algorithm takes the update equation as given by

$$\hat{a}(n+1;k) = \hat{a}(n;k) + \mu x(n-k)\epsilon(n), \quad k = 1, 2, \ldots, 2p \tag{200}$$

where μ is the step size. This algorithm, which was proposed by Griffiths (19), can track the variation of the instantaneous frequency provided that the frequency is slowly time-varying, and that the step size μ is chosen optimally so that the prediction error $\epsilon_1(n)$ is small for all n.

RLS Algorithm. Let

$$\mathbf{y}(t) = [x(t-1), x(t-2), \ldots, x(t-2p)]^T \tag{201}$$

Suppose we have observed the vectors $\mathbf{y}(t)$, $t = 0, 1, \ldots, n$, and we wish to determine the coefficient vector $\mathbf{a}(n) = [a(n;\ 1), a(n;\ 2), \ldots, a(n;\ 2p)]^T$. The RLS solution is obtained by minimizing the time-average weighted square error

$$J_2 = \sum_{t=0}^{n} w^{n-t}|\epsilon_2(t;n)|^2 \tag{202}$$

where the error is defined as

$$\epsilon_2(t;n) = x(t) - \mathbf{a}^T(n)\mathbf{y}(t) \tag{203}$$

and w represents a weighting factor $0 < w < 1$. Minimizing J_2 with respect to the coefficient vector yields the RLS solution as follows

$$\mathbf{a}(n) = \mathbf{R}^{-1}(n)\mathbf{D}(n) \tag{204}$$

where

$$\mathbf{R}(n) = \sum_{t=0}^{n} w^{n-t}\mathbf{y}^*(t)\mathbf{y}^T(t) \tag{205}$$

and

$$\mathbf{D}(n) = \sum_{t=0}^{n} w^{n-t}x(t)\mathbf{y}^*(t) \tag{206}$$

The major advantage of the LMS algorithm lies in its computational simplicity. However, the price paid for this simplicity is slow convergence. The RLS algorithm, on the other hand, can achieve fast convergence; however, it involves more complex computations. Note the computation of matrix inversion in the RLS algorithm can be avoided by using the Kalman filtering approach.

BIBLIOGRAPHY

1. M. S. Bartlett, Smoothing periodograms for time series with continuous spectra, *Nature,* **161**: 686–687, 1948.
2. P. D. Welch, The use of the fast Fourier transform for the estimation of power spectra: A method based on time averaging over short modified periodograms, *IEEE Trans. Audio Electroacoust.,* **AU-15**: 70–76, 1967.

3. R. B. Blackman and J. W. Tukey, *The Measurement of Power Spectra from the Point of View of Communication Engineering,* New York: Dover, 1959.
4. J. Capon, Higher-resolution frequency-wavenumber spectrum analysis, *Proc. IEEE,* **57**: 1408–1418, 1969.
5. L. C. Pusey, High resolution spectral estimates, Lincoln Laboratory, M.I.T., Tech. Note, Jan. 21, 1975–7, 1975.
6. N. Levinson, The Wiener RMS (root mean square) criterion in filter design and prediction, *J. Math. Phys.,* **25**: 261–278, 1947.
7. J. Durbin, The fitting of time series models, *Rev. Inst. Int. Stat.,* **28**: 233–244, 1960.
8. J. C. Chow, On the estimation of the order of a moving-average process, *IEEE Trans. Autom. Control,* **AC-17**: 386–387, 1972.
9. J. A. Cadzow, Spectral estimation: An overdetermined rational model equation approach, *Proc. IEEE,* **70**: 907–939, 1982.
10. X. D. Zhang and Y. S. Zhang, Determination of the MA order of an ARMA process using sample correlations, *IEEE Trans. Signal Process.,* **SP-41**: 2277–2280, 1993.
11. J. C. Chow, On estimating the orders of an autoregressive moving-average process with uncertain observations, *IEEE Trans. Autom. Control,* **AC-17**: 707–709, 1972.
12. V. F. Pisarenko, The retrieval of harmonics from a covariance function, *Geophys. J. Roy. Astron. Soc.,* **33**: 347–366, 1973.
13. R. O. Schmidt, Multiple emitter location and signal parameter estimation, *Proc. RADC, Spectral Estimation Workshop,* Rome, NY, 243–258, 1979.
14. A. Paulraj, R. Roy, and T. Kailath, A subspace rotation approach to signal parameter estimation, *Proc. IEEE,* **74**: 1044–1054, 1986.
15. R. Roy and T. Kailath, ESPRIT—Estimation of signal parameters via rotational invariance techniques, *IEEE Trans. Acoust. Speech Signal Process.,* **ASSP-37**: 984–995, 1989.
16. M. Wax and T. Kailath, Detection of signals by information theoretic criteria, *IEEE Trans. Acoust. Speech Signal Process.,* **ASSP-33**: 387–392, 1985.
17. R. Kumaresan and R. W. Tufts, Estimating the parameters of exponentially damped sinusoids and pole-zero modeling in noise, *IEEE Trans. Acoust. Speech Signal Process.,* **ASSP-30**: 833–840, 1982.
18. Y. Li, K. J. R. Liu, and J. Razavilar, A parameter estimation scheme for damped sinusoidal signals based on low-rank Hankel approximation, *IEEE Trans. Signal Process.,* **SP-45**: 481–486, 1997.
19. L. J. Griffiths, Rapid measurement of digital instantaneous frequency, *IEEE Trans. Acoust. Speech Signal Process.,* **ASSP-23**: 207–222, 1975.

K. J. RAY LIU
YING-CHANG LIANG
University of Maryland

See also FOURIER ANALYSIS; SPEECH ANALYSIS.