Djuric, P.M. & Kay S.M. “Spectrum Estimation and Modeling”
Digital Signal Processing Handbook
Ed. Vijay K. Madisetti and Douglas B. Williams
Boca Raton: CRC Press LLC, 1999

14

Spectrum Estimation and Modeling

14.1 Introduction

14.2 Important Notions and Definitions
Random Processes • Spectra of Deterministic Signals • Spectra of Random Processes

14.3 The Problem of Power Spectrum Estimation

14.4 Nonparametric Spectrum Estimation
Periodogram • The Bartlett Method • The Welch Method • Blackman-Tukey Method • Minimum Variance Spectrum Estimator • Multiwindow Spectrum Estimator

14.5 Parametric Spectrum Estimation
Spectrum Estimation Based on Autoregressive Models • Spectrum Estimation Based on Moving Average Models • Spectrum Estimation Based on Autoregressive Moving Average Models • Pisarenko Harmonic Decomposition Method • Multiple Signal Classification (MUSIC)

14.6 Recent Developments

References

Petar M. Djurić
State University of New York at Stony Brook

Steven M. Kay
University of Rhode Island

14.1 Introduction

The main objective of spectrum estimation is the determination of the power spectrum density (PSD) of a random process. The PSD is a function that plays a fundamental role in the analysis of stationary random processes in that it quantifies the distribution of total power as a function of frequency. The estimation of the PSD is based on a set of observed data samples from the process. A necessary assumption is that the random process is at least wide sense stationary, that is, its first and second order statistics do not change with time. The estimated PSD provides information about the structure of the random process which can then be used for refined modeling, prediction, or filtering of the observed process.

Spectrum estimation has a long history with beginnings in ancient times [17]. The first significant discoveries that laid the grounds for later developments, however, were made in the early years of the eighteenth century. They include one of the most important advances in the history of mathematics, Fourier's theory. According to this theory, an arbitrary function can be represented by an infinite summation of sine and cosine functions. Later came the Sturm-Liouville spectral theory of differential equations, which was followed by the spectral representations in quantum and classical physics developed by John von Neuman and Norbert Wiener, respectively. The statistical theory of spectrum estimation started practically in 1949 when Tukey introduced a numerical method for computation of spectra from empirical data. A very important milestone for further development of the field was the reinvention of the fast Fourier transform (FFT) in 1965, which is an efficient algorithm for computation of the discrete Fourier transform. Shortly thereafter came the work of John Burg, who

proposed a fundamentally new approach to spectrum estimation based on the principle of maximum entropy. In the past three decades his work was followed up by many researchers who have developed numerous new spectrum estimation procedures and applied them to various physical processes from diverse scientific fields. Today, spectrum estimation is a vital scientific discipline which plays a major role in many applied sciences such as radar, speech processing, underwater acoustics, biomedical signal processing, sonar, seismology, vibration analysis, control theory, and econometrics.

14.2 Important Notions and Definitions

14.2.1 Random Processes

The objects of interest of spectrum estimation are random processes. They represent time fluctuations of a certain quantity which cannot be fully described by deterministic functions. The voltage waveform of a speech signal, the bit stream of zeros and ones of a communication message, or the daily variations of the stock market index are examples of random processes. Formally, a random process is defined as a collection of random variables indexed by time. (The family of random variables may also be indexed by a different variable, for example space, but here we will consider only random *time* processes.) The index set is infinite and may be continuous or discrete. If the index set is continuous, the random process is known as a continuous-time random process, and if the set is discrete, it is known as a discrete-time random process. The speech waveform is an example of a continuous random process and the sequence of zeros and ones of a communication message, a discrete one. We shall focus only on discrete-time processes where the index set is the set of integers.

A random process can be viewed as a collection of a possibly infinite number of functions, also called realizations. We shall denote the collection of realizations by $\{\tilde{x}[n]\}$ and an observed realization of it by $\{x[n]\}$. For fixed n, $\{\tilde{x}[n]\}$ represents a random variable, also denoted as $\tilde{x}[n]$, and $x[n]$ is the n-th sample of the realization $\{x[n]\}$. If the samples $x[n]$ are real, the random process is real, and if they are complex, the random process is complex. In the discussion to follow, we assume that $\{\tilde{x}[n]\}$ is a *complex* random process.

The random process $\{\tilde{x}[n]\}$ is fully described if for any set of time indices n_1, n_2, ..., n_m, the joint probability density function of $\tilde{x}[n_1]$, $\tilde{x}[n_2]$, ..., and $\tilde{x}[n_m]$ is given. If the statistical properties of the process do not change with time, the random process is called stationary. This is always the case if for any choice of random variables $\tilde{x}[n_1]$, $\tilde{x}[n_2]$, ..., and $\tilde{x}[n_m]$, their joint probability density function is identical to the joint probability density function of the random variables $\tilde{x}[n_1 + k]$, $\tilde{x}[n_2 + k]$, ..., and $\tilde{x}[n_m + k]$ for any k. Then we call the random process strictly stationary. For example, if the samples of the random process are independent and identically distributed random variables, it is straightforward to show that the process is strictly stationary. Strict stationarity, however, is a very severe requirement and is relaxed by introducing the concept of wide-sense stationarity. A random process is wide-sense stationary if the following two conditions are met:

$$E\left(\tilde{x}[n]\right) = \mu \tag{14.1}$$

and

$$\begin{aligned} r[n, n+k] &= E\left(\tilde{x}^*[n]\tilde{x}[n+k]\right) \\ &= r[k] \end{aligned} \tag{14.2}$$

where $E(\cdot)$ is the expectation operator, $\tilde{x}^*[n]$ is the complex conjugate of $\tilde{x}[n]$, and $\{r[k]\}$ is the autocorrelation function of the process. Thus, if the process is wide-sense stationary, its mean value μ is constant over time, and the autocorrelation function depends only on the lag k between the random variables. For example, if we consider the random process

$$\tilde{x}[n] = a\cos(2\pi f_0 n + \tilde{\theta}) \tag{14.3}$$

where the amplitude a and the frequency f_0 are constants, and the phase $\tilde{\theta}$ is a random variable that is uniformly distributed over the interval $(-\pi, \pi)$, one can show that

$$E(\tilde{x}[n]) = 0 \tag{14.4}$$

and

$$\begin{aligned} r[n, n+k] &= E\left(\tilde{x}^*[n]\tilde{x}[n+k]\right) \\ &= \frac{a^2}{2}\cos(2\pi f_0 k) \,. \end{aligned} \tag{14.5}$$

Thus, Eq. (14.3) represents a wide-sense stationary random process.

14.2.2 Spectra of Deterministic Signals

Before we define the concept of spectrum of a random process, it will be useful to review the analogous concept for deterministic signals, which are signals whose future values can be exactly determined without any uncertainty. Besides their description in the time domain, the deterministic signals have a very useful representation in terms of superposition of sinusoids with various frequencies, which is given by the discrete-time Fourier transform (DTFT). If the observed signal is $\{g[n]\}$ and it is not periodic, its DTFT is the complex valued function $G(f)$ defined by

$$G(f) = \sum_{n=-\infty}^{\infty} g[n]e^{-j2\pi fn} \tag{14.6}$$

where $j = \sqrt{-1}$, f is the normalized frequency, $0 \leq f < 1$, and $e^{j2\pi fn}$ is the complex exponential given by

$$e^{j2\pi fn} = \cos(2\pi fn) + j\sin(2\pi fn) \,. \tag{14.7}$$

The sum in Eq. (14.6) converges uniformly to a continuous function of the frequency f if

$$\sum_{n=-\infty}^{\infty} |g[n]| < \infty \,. \tag{14.8}$$

The signal $\{g[n]\}$ can be determined from $G(f)$ by the inverse DTFT defined by

$$g[n] = \int_0^1 G(f)e^{j2\pi fn}df \tag{14.9}$$

which means that the signal $\{g[n]\}$ can be represented in terms of complex exponentials whose frequencies span the continuous interval [0,1).

The complex function $G(f)$ can be alternatively expressed as

$$G(f) = |G(f)|e^{j\phi(f)} \tag{14.10}$$

where $|G(f)|$ is called the amplitude spectrum of $\{g[n]\}$, and $\phi(f)$ the phase spectrum of $\{g[n]\}$. For example, if the signal $\{g[n]\}$ is given by

$$g[n] = \begin{cases} 1, & n = 1 \\ 0, & n \neq 1 \end{cases} \tag{14.11}$$

then

$$G(f) = e^{-j2\pi f} \tag{14.12}$$

and the amplitude and phase spectra are

$$\begin{aligned} |G(f)| &= 1, \qquad 0 \le f < 1 \\ \phi(f) &= -2\pi f, \quad 0 \le f < 1 \,. \end{aligned} \tag{14.13}$$

The total energy of the signal is given by

$$\mathcal{E} = \sum_{n=-\infty}^{\infty} |g[n]|^2 \tag{14.14}$$

and according to Parseval's theorem, it can also be obtained from the amplitude spectrum of the signal, i.e.,

$$\sum_{n=-\infty}^{\infty} |g[n]|^2 = \int_0^1 |G(f)|^2 df \,. \tag{14.15}$$

From Eq. (14.15), we deduce that $|G(f)|^2 df$ is the contribution to the total energy of the signal from the frequency band $(f, f + df)$. Therefore, we say that $|G(f)|^2$ represents the energy density spectrum of the signal $\{g[n]\}$.

When $\{g[n]\}$ is periodic with period N, that is

$$g(n) = g(n + N) \tag{14.16}$$

for all n, and where N is the period of $\{g[n]\}$, we use the discrete Fourier transform (DFT) to express $\{g[n]\}$ in the frequency domain, that is,

$$G(f_k) = \sum_{n=0}^{N-1} g[n] e^{-j2\pi f_k n}, \quad f_k = \frac{k}{N}, \quad k \in \{0, 1, \cdots, N-1\} \,. \tag{14.17}$$

Note that the frequency here takes values from a discrete set. The inverse DFT is defined by

$$g[n] = \frac{1}{N} \sum_{k=0}^{N-1} G(f_k) e^{j2\pi f_k n}, \quad f_k = \frac{k}{N} \,. \tag{14.18}$$

Now Parseval's relation becomes

$$\sum_{n=0}^{N-1} |g[n]|^2 = \frac{1}{N} \sum_{k=0}^{N-1} |G(f_k)|^2, \quad f_k = \frac{k}{N} \tag{14.19}$$

where the two sides are the total energy of the signal in one period. If we define the average power of the discrete-time signal by

$$\mathcal{P} = \frac{1}{N} \sum_{n=0}^{N-1} |g[n]|^2 \tag{14.20}$$

then from Eq. (14.19)

$$\mathcal{P} = \frac{1}{N^2} \sum_{k=0}^{N-1} |G(f_k)|^2, \quad f_k = \frac{k}{N} \,. \tag{14.21}$$

Thus, $|G(f_k)|^2/N^2$ is the contribution to the total power from the term with frequency f_k, and so it represents the power spectrum "density" of $\{g[n]\}$. For example, if the periodic signal in one period is defined by

$$g[n] = \begin{cases} 1, & n = 0 \\ 0, & n = 1, 2, \cdots, N-1 \end{cases} \tag{14.22}$$

its PSD $P(f_k)$ is

$$P(f_k) = \frac{1}{N^2}, \quad f_k = \frac{k}{N}, \quad k \in \{0, 1, \cdots, N-1\}\,. \tag{14.23}$$

Again, note that the PSD is defined for a discrete set of frequencies.

In summary, the spectra of deterministic aperiodic signals are energy densities defined on the continuous set of frequencies $\mathcal{C}_f = [0, 1)$. On the other hand, the spectra of periodic signals are power densities defined on the discrete set of frequencies $\mathcal{D}_f = \{0, 1/N, 2/N, \cdots, (N-1)/N\}$, where N is the period of the signal.

14.2.3 Spectra of Random Processes

Suppose that we observe one realization of the random process $\{\tilde{x}[n]\}$, or $\{x[n]\}$. From the definition of the DTFT and the assumption of wide-sense stationarity of $\{\tilde{x}[n]\}$, it is obvious that we cannot use the DTFT to obtain $X(f)$ from $\{x[n]\}$ because Eq. (14.8) does not hold when we replace $g[n]$ by $x[n]$. And indeed, if $\{x[n]\}$ is a realization of a wide-sense stationary process, its energy is infinite. Its power, however, is finite as was the case with the periodic signals. So if we observe $\{x[n]\}$ from $-N$ to N, $\{x[n]\}_{-N}^{N}$, and assume that outside this interval the samples $x[n]$ are equal to zero, we can find its DTFT, $X_N(f)$ from

$$X_N(f) = \sum_{n=-N}^{N} x[n]e^{-j2\pi fn}. \tag{14.24}$$

Then according to Eq. (14.15), $|X_N(f)|^2 df$ represents the energy of the truncated realization that is contributed by the components whose frequencies are between f and $f + df$. The power due to these components is given by

$$\frac{|X_N(f)|^2 df}{2N+1} \tag{14.25}$$

and $|X_N(f)|^2/(2N+1)$ can be interpreted as power density. If we let $N \to \infty$, under suitable conditions [15],

$$\lim_{N\to\infty} \frac{|X_N(f)|^2}{2N+1} \tag{14.26}$$

is finite for all f, and this is then the PSD of $\{x[n]\}$. We would prefer to find, however, the PSD of $\{\tilde{x}[n]\}$, which we define as

$$P(f) = \lim_{N\to\infty} E\left\{\frac{|\tilde{X}_N(f)|^2}{2N+1}\right\} \tag{14.27}$$

where $\tilde{X}_N(f)$ is the DTFT of $\{\tilde{x}[n]\}_{-N}^{N}$. Clearly, $P(f)df$ is interpreted as the *average* contribution to the total power from the components of $\{\tilde{x}[n]\}$ whose frequencies are between f and $f + df$.

There is a very important relationship between the PSD of a wide-sense stationary random process and its autocorrelation function. By Wold's theorem, which is the analogue of Wiener-Khintchine theorem for continuous-time random processes, the PSD in Eq. (14.27) is the DTFT of the autocorrelation function of the process [15], that is,

$$P(f) = \sum_{k=-\infty}^{\infty} r[k]e^{-j2\pi fk} \tag{14.28}$$

where $r[k]$ is defined by Eq. (14.2).

For all practical purposes, there are three different types of $P(f)$ [15]. If $P(f)$ is an absolutely continuous function of f, the random process has a purely continuous spectrum. If $P(f)$ is identically equal to zero for all f except for frequencies $f = f_k$, $k = 1, 2, ...$, where it is infinite, the

random process has a line spectrum. In this case, a useful representation of the spectrum is given by the Dirac δ-functions,

$$P(f) = \sum_k P_k \delta(f - f_k) \tag{14.29}$$

where P_k is the power associated with the k line component. Finally, the spectrum of a random process may be mixed if it is a combination of a continuous and line spectra. Then $P(f)$ is a superposition of a continuous function of f and δ-functions.

14.3 The Problem of Power Spectrum Estimation

The problem of power spectrum estimation can be stated as follows: Given a set of N samples $\{x[0], x[1], ..., x[N-1]\}$ of a realization of the random process $\{\tilde{x}[n]\}$, denoted also by $\{x[n]\}_0^{N-1}$, estimate the PSD of the random process, $P(f)$. Obviously this task amounts to estimation of a function and is distinct from the typical problem in elementary statistics where the goal is to estimate a finite set of parameters.

Spectrum estimation methods can be classified into two categories: nonparametric and parametric. The nonparametric approaches do not assume any specific parametric model for the PSD. They are based solely on the estimate of the autocorrelation sequence of the random process from the observed data. For the parametric approaches on the other hand, we first postulate a model for the process of interest, where the model is described by a small number of parameters. Based on the model, the PSD of the process can be expressed in terms of the model parameters. Then the PSD estimate is obtained by substituting the estimated parameters of the model in the expression for the PSD. For example, if a random process $\{\tilde{x}[n]\}$ can be modeled by

$$\tilde{x}[n] = -a\tilde{x}[n] + \tilde{w}[n] \tag{14.30}$$

where a is an unknown parameter and $\{\tilde{w}[n]\}$ is a zero mean wide-sense stationary random process whose random variables are uncorrelated and with the same variance σ^2, it can be shown that the PSD of $\{\tilde{x}[n]\}$ is

$$P(f) = \frac{\sigma^2}{|1 + ae^{-j2\pi f}|^2}. \tag{14.31}$$

Thus, to find $P(f)$ it is sufficient to estimate a and σ^2.

The performance of a PSD estimator is evaluated by several measures of goodness. One is the bias of the estimator defined by

$$b(f) = E\left(\hat{P}(f) - P(f)\right) \tag{14.32}$$

where $\hat{P}(f)$ and $P(f)$ are the estimated and true PSD, respectively. If the bias $b(f)$ is identically equal to zero for all f, the estimator is said to be unbiased, which means that on average it yields the true PSD. Among the unbiased estimators, we search for the one that has minimal variability. The variability is measured by the variance of the estimator

$$v(f) = E\left([\hat{P}(f) - E(P(\hat{f}))]^2\right). \tag{14.33}$$

A measure that combines the bias and the variance is the relative mean square error given by [15]

$$\nu(f) = \frac{v(f) + b(f)^2}{P(f)}. \tag{14.34}$$

The variability of a PSD estimator is also measured by the normalized variance [8]

$$\psi(f) = \frac{v(f)}{E^2(\hat{P}(f))} \,. \tag{14.35}$$

Finally, another important metric for comparison is the resolution of the PSD estimators. It corresponds to the ability of the estimator to provide the fine details of the PSD of the random process. For example if the PSD of the random process has two peaks at frequencies f_1 and f_2, then the resolution of the estimator would be measured by the minimum separation of f_1 and f_2 for which the estimator still reproduces two peaks at f_1 and f_2.

14.4 Nonparametric Spectrum Estimation

When the method for PSD estimation is not based on any assumptions about the generation of the observed samples other than wide-sense stationarity, then it is termed a nonparametric estimator. According to Eq. (14.28), $P(f)$ can be obtained by first estimating the autocorrelation sequence from the observed samples $x[0], x[1], \cdots, x[N-1]$, and then applying the DTFT to these estimates. One estimator of the autocorrelation is given by

$$\hat{r}[k] = \frac{1}{N} \sum_{n=0}^{N-1-k} x^*[n]x[n+k], \quad 0 \le k \le N-1 \,. \tag{14.36}$$

The estimates of $\hat{r}[k]$ for $-N < k < 0$ are obtained from the identity

$$\hat{r}[-k] = \hat{r}^*[k] \tag{14.37}$$

and those for $|k| \ge N$ are set equal to zero. This estimator, although biased, has been preferred over others. An important reason for favoring it is that it always yields nonnegative estimates of the PSD, which is not the case with the unbiased estimator.

Many nonparametric estimators rely on using Eq. (14.36) and then transform the obtained autocorrelation sequence to estimate the PSD. Other nonparametric methods, however, operate directly on the observed data.

14.4.1 Periodogram

The periodogram was introduced by Schuster in 1898 when he was searching for hidden periodicities while studying sunspot data [19]. To find the periodogram of the data $\{x[n]\}_0^{N-1}$, first we determine the autocorrelation sequence $r[k]$ for $-(N-1) \le k \le N-1$ and then take the DTFT, i.e.,

$$\hat{P}_{\text{PER}}(f) = \sum_{k=-N+1}^{N-1} \hat{r}[k]e^{-j2\pi fk} \,. \tag{14.38}$$

It is more convenient to write the periodogram directly in terms of the observed samples $x[n]$. It is then defined as

$$\hat{P}_{\text{PER}}(f) = \frac{1}{N} \left| \sum_{n=0}^{N-1} x[n]e^{-j2\pi fn} \right|^2 \,. \tag{14.39}$$

Thus, the periodogram is proportional to the squared magnitude of the DTFT of the observed data. In practice, the periodogram is calculated by applying the FFT, which computes it at a discrete set of

frequencies $\mathcal{D}_f = \{f_k : f_k = k/N,\ k = 0, 1, 2, \cdots, (N-1)\}$. The periodogram is then expressed by

$$\hat{P}_{\text{PER}}(f_k) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]e^{-j2\pi kn/N}\right|^2, \qquad f_k \in \mathcal{D}_f\ . \tag{14.40}$$

To allow for finer frequency spacing in the computed periodogram, we define a zero padded sequence according to

$$x'[n] = \begin{cases} x[n], & n = 0, 1, \cdots, N-1 \\ 0, & n = N, N+1, \cdots, N' \end{cases}\ . \tag{14.41}$$

Then we specify the new set of frequencies $\mathcal{D}'_f = \{f_k : f_k = k/N',\ k \in \{0, 1, 2, \cdots, (N'-1)\}\}$, and obtain

$$\hat{P}_{\text{PER}}(f_k) = \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]e^{-j2\pi kn/N'}\right|^2, \quad f_k \in \mathcal{D}'_f\ . \tag{14.42}$$

A general property of good estimators is that they yield better estimates when the number of observed data samples increases. Theoretically, if the number of data samples tends to infinity, the estimates should converge to the true values of the estimated parameters. So, in the case of a PSD estimator, as we get more and more data samples, it is desirable that the estimated PSD tends to the true value of the PSD. In other words, if for finite number of data samples the estimator is biased, the bias should tend to zero as $N \to \infty$ as should the variance of the estimate. If this is indeed the case, the estimator is called consistent. Although the periodogram is asymptotically unbiased, it can be shown that it is not a consistent estimator. For example, if $\{\tilde{x}[n]\}$ is real zero-mean white Gaussian noise, which is a process whose random variables are independent, Gaussian, and identically distributed with variance σ^2, the variance of $\hat{P}_{\text{PER}}(f)$ is equal to σ^4 regardless of the length N of the observed data sequence [12]. The performance of the periodogram does not improve as N gets larger because as N increases, so does the number of parameters that are estimated, $P(f_0)$, $P(f_1)$, ..., $P(f_{N-1})$. In general, for the variance of the periodogram, we can write [12]

$$\text{var}(\hat{P}_{\text{PER}}(f)) \simeq P^2(f) \tag{14.43}$$

where $P(f)$ is the true PSD.

Interesting insight can be gained if one writes the periodogram as follows

$$\begin{aligned} \hat{P}_{\text{PER}}(f) &= \frac{1}{N}\left|\sum_{n=0}^{N-1} x[n]e^{-j2\pi fn}\right|^2 \\ &= \frac{1}{N}\left|\sum_{n=-\infty}^{\infty} x[n]w_R[n]e^{-j2\pi fn}\right|^2 \end{aligned} \tag{14.44}$$

where $w_R[n]$ is a rectangular window defined by

$$w_R[n] = \begin{cases} 1, & n \in \{0, 1, \cdots, N-1\} \\ 0, & \text{otherwise} \end{cases}\ . \tag{14.45}$$

Thus, we can regard the finite data record used for estimating the PSD as being obtained by multiplying the whole realization of the random process by a rectangular window. Then it is not difficult to show that the expected value of the periodogram is given by [8]

$$E\left\{\hat{P}_{\text{PER}}(f)\right\} = \frac{1}{N}\int_0^1 |W_R(f-\xi)|^2\, P(\xi)d\xi \tag{14.46}$$

where $W_R(f)$ is the DTFT of the rectangular window. Hence, the mean value of the periodogram is a smeared version of the true PSD. Since the implementation of the periodogram as defined in Eq. (14.44) implies the use of a rectangular window, a question arises as to whether we could use a window of different shape to reduce the variance of the periodogram. The answer is yes, and indeed many windows have been proposed which weight the data samples in the middle of the observed data more than those towards the ends of the observed data. Some frequently used alternatives to the rectangular window are the windows of Bartlett, Hanning, Hamming, and Blackman. The magnitude of the DTFT of a window provides two important characteristics about it. One is the width of the window's mainlobe and the other is the strength of its sidelobes. A narrow mainlobe allows for a better resolution, and low sidelobes improve the smoothing of the estimated spectrum. Unfortunately, the narrower its mainlobe, the higher the sidelobes, which is a typical trade-off in spectrum estimation. It turns out that the rectangular window allows for the best resolution but has the largest sidelobes.

14.4.2 The Bartlett Method

One approach to reduce the variance of the periodogram is to subdivide the observed data record into K nonoverlapping segments, find the periodogram of each segment, and finally evaluate the average of the so-obtained periodograms. This spectrum estimator, also known as the Bartlett's estimator, has variance that is smaller than the variance of the periodogram.

Suppose that the number of data samples N is equal to KL, where K is the number of segments and L is their length. If the i-th segment is denoted by $\{x_i[n]\}_0^{L-1}$, $i = 1, 2 \cdots, K$, where

$$x_i[n] = x[n + (i-1)L], \quad n \in \{0, 1, \cdots, L-1\} \tag{14.47}$$

and its periodogram by

$$\hat{P}_{\text{PER}}^{(i)}(f) = \frac{1}{L} \left| \sum_{n=0}^{L-1} x_i[n] e^{-j2\pi fn} \right|^2 \tag{14.48}$$

then the Bartlett spectrum estimator is

$$\hat{P}_{\text{B}}(f) = \frac{1}{K} \sum_{i=1}^{K} \hat{P}_{\text{PER}}^{(i)}(f) \, . \tag{14.49}$$

This estimator is consistent and its variance compared to the variance of the periodogram is reduced by a factor of K. This reduction, however, is paid by a decrease in resolution. The Bartlett estimator has a resolution K times less than that of the periodogram. Thus, this estimator allows for a straightforward trading of resolution for variance.

14.4.3 The Welch Method

The Welch method is another estimator that exploits the periodogram. It is based on the same idea as Bartlett's approach of splitting the data into segments and finding the average of their periodograms. The difference is that the segments are overlapped, where the overlaps are usually 50% or 75% large, and the data within a segment are windowed. Let the length of the segments be L, the i-th segment be denoted again by $\{x_i[n]\}_0^{L-1}$, and the offset of successive sequences by D samples. Then

$$N = L + D(K-1) \tag{14.50}$$

where N is the total number of observed samples and K the total number of sequences. Note that if there is no overlap, $K = N/L$, and if there is 50% overlap, $K = 2N/L - 1$. The i-th sequence is

defined by

$$x_i[n] = x[n + (i-1)D], \quad n \in \{0, 1, \cdots, L-1\} \tag{14.51}$$

where $i = 1, 2, \cdots, K$, and its periodogram by

$$\hat{P}_{\mathrm{M}}^{(i)}(f) = \frac{1}{L}\left|\sum_{n=0}^{L-1} w[n]x_i[n]e^{-j2\pi fn}\right|^2 . \tag{14.52}$$

Here $\hat{P}_{\mathrm{M}}^{(i)}(f)$ is the modified periodogram of the data because the samples $x[n]$ are weighted by a nonrectangular window $w[n]$. The Welch spectrum estimate is then given by

$$\hat{P}_{\mathrm{B}}(f) = \frac{1}{K}\sum_{i=1}^{K} \hat{P}_{\mathrm{M}}^{(i)}(f) . \tag{14.53}$$

By permitting overlap of sequences, we can form more segments than in the case of Bartlett's method. Also, if we keep the same number of segments, the overlap allows for longer segments. The increased number of segments reduces the variance of the estimator, and the longer segments improve its resolution. Thus, with the Welch method we can trade reduction in variance for improvement in resolution in many more ways than with the Bartlett method. It can be shown that if the overlap is 50%, the variance of the Welch estimator is approximately 9/16 of the variance of the Bartlett estimator [8].

14.4.4 Blackman-Tukey Method

The periodogram can be expressed in terms of the estimated autocorrelation lags as

$$\hat{P}_{\mathrm{PER}}(f) = \sum_{k=-(N-1)}^{N-1} \hat{r}[k]e^{-j2\pi fk} . \tag{14.54}$$

where

$$\hat{r}[k] = \begin{cases} \frac{1}{N}\sum_{n=0}^{N-1-k} x^*[n]x[n+k], & k = 0, 1, \cdots, N-1 \\ \hat{r}^*[-k], & k = -(N-1), -(N-2), \cdots, -1 \end{cases} . \tag{14.55}$$

From Eqs. (14.54) and (14.55) we see that the estimated autocorrelation lags are given the same weight in the periodogram regardless of the difference of their variances. From Eq. (14.55), however, it is obvious that the autocorrelations with smaller lags will be estimated more accurately than the ones with lags close to N because of the different number of terms that are used in the summation. For example, $\hat{r}[N-1]$ has only the term $x^*[0]x[n-1]$ compared to the N terms used in the computation of $\hat{r}[0]$. Therefore, the large variance of the periodogram can be ascribed to the large weight given to the poor autocorrelation estimates used in its evaluation.

Blackman and Tukey proposed to weight the autocorrelation sequence so that the autocorrelations with higher lags are weighted less [3]. Their estimator is given by

$$\hat{P}_{\mathrm{BT}}(f) = \sum_{k=-(N-1)}^{N-1} w[k]\hat{r}[k]e^{-j2\pi fk} \tag{14.56}$$

where the window $w[k]$ is real nonnegative, symmetric, and nonincreasing with $|k|$, that is

$$\begin{aligned} &1. \quad 0 \le w[k] \le w[0] = 1 , \\ &2. \quad w[-k] = w[k], \ \text{and} \\ &3. \quad w[k] = 0, \quad M < |k|, \ M \le N-1 . \end{aligned} \tag{14.57}$$

Note that the symmetry property of $w[k]$ ensures that the spectrum is real.

The Blackman-Tukey estimator can be expressed in the frequency domain by the convolution

$$\hat{P}_{\mathrm{BT}}(f) = \int_0^1 W(f-\xi)\hat{P}_{\mathrm{PER}}(\xi)d\xi \ . \tag{14.58}$$

From Eq. (14.58) we deduce that the window's DTFT should satisfy

$$W(f) \geq 0, \quad f \in [0,1) \tag{14.59}$$

so that the spectrum is guaranteed to be a nonnegative function, i.e.,

$$\hat{P}_{\mathrm{BT}}(f) \geq 0, \qquad |f| \leq \frac{1}{2} \ . \tag{14.60}$$

The bias, the variance, and the resolution of the Blackman-Tukey method depend on the applied window. For example, if the window is triangular (Bartlett),

$$w_B[k] = \begin{cases} \frac{M-|k|}{M}, & |k| \leq M \\ 0, & \text{otherwise} \end{cases} \tag{14.61}$$

and if $N >> M >> 1$, the variance of the Blackman-Tukey estimator is [12]

$$\mathrm{var}(\hat{P}_{\mathrm{BT}}(f)) \simeq \frac{2M}{3N}P^2(f) \tag{14.62}$$

where $P(f)$ is the true spectrum of the process. Compared to Eq. (14.43), it is clear that the variance of this estimator may be significantly smaller than the variance of the periodogram. However, as M decreases, so does the resolution of the Blackman-Tukey estimator.

14.4.5 Minimum Variance Spectrum Estimator

The periodogram [Eq. (14.44)] can also be written as

$$\begin{aligned} \hat{P}_{\mathrm{PER}}(f) &= \frac{1}{N}\left|\mathbf{e}^H(f)\mathbf{x}\right|^2 \\ &= N\left|\mathbf{h}^H(f)\mathbf{x}\right|^2 \end{aligned} \tag{14.63}$$

where $\mathbf{e}(f)$ is an $N \times 1$ vector defined by

$$\mathbf{e}(f) = [1 \ e^{j2\pi f} \ e^{j4\pi f} \ \cdots \ e^{j2(N-1)\pi f}]^T \tag{14.64}$$

and $\mathbf{h}(f) = \mathbf{e}(f)/N$ with H denoting complex conjugate transpose. We could interpret $\mathbf{h}(f)$ as the impulse response of a finite impulse response (FIR) filter. It is easy to show that $\mathbf{h}(f)$ is a bandpass filter centered at f with a bandwidth of approximately $1/N$. Then starting with Eq. (14.63) we can show that the value of the periodogram at frequency f can be obtained by squaring the magnitude of the filter output at $N-1$. Such filters inherently exist for all the frequencies where the periodogram is evaluated, and they all have the same bandwidth. Thus, the periodogram may be viewed as a bank of FIR filters with equal bandwidths.

Capon proposed a spectrum estimator for processing large seismic arrays which, like the periodogram, can be interpreted as a bank of filters [5]. The width of these filters, however, is data

dependent and optimized to minimize their response to components outside the band of interest. If the impulse response of the filter centered at f_0 is $\mathbf{h}(f_0)$, then it is desired to minimize

$$\rho = \int_0^1 |H(f)|^2 P(f) df \tag{14.65}$$

subject to the constraint

$$H(f_0) = 1 \tag{14.66}$$

where $H(f)$ is the DTFT of $\mathbf{h}(f_0)$. This is a constrained minimization problem, and the solution provides the optimal impulse response. When the solutions are used to determine the PSD of the observed data, we obtain the minimum variance (MV) spectrum estimator

$$\hat{P}_{\text{MV}}(f) = \frac{N}{\mathbf{e}^H(f)\hat{\mathbf{R}}^{-1}\mathbf{e}(f)} \tag{14.67}$$

where $\hat{\mathbf{R}}^{-1}$ is the inverse matrix of the $N \times N$ autocorrelation matrix $\hat{\mathbf{R}}$ defined by

$$\hat{\mathbf{R}} = \begin{bmatrix} \hat{r}[0] & \hat{r}[-1] & \hat{r}[-2] & \cdots & \hat{r}[-N+1] \\ \hat{r}[1] & \hat{r}[0] & \hat{r}[-1] & \cdots & \hat{r}[-N+2] \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{r}[N-1] & \hat{r}[N-2] & \hat{r}[N-3] & \cdots & \hat{r}[0] \end{bmatrix}. \tag{14.68}$$

The length of the FIR filter does not have to be N, especially if we want to avoid the use of the unreliable estimates of $r[k]$. If the length of the filter's response is $p < N$, then the vector $\mathbf{e}(f)$, the autocorrelation matrix $\hat{\mathbf{R}}$, and the spectrum estimate $\hat{P}_{\text{MV}}(f)$ are defined by Eqs. (14.64), (14.68), and (14.67), respectively, with N replaced by p [12].

The MV estimator has better resolution than the periodogram and the Blackman-Tukey estimator. The resolution and the variance of the MV estimator depend on the choice of the filter length p. If p is large, the bandwidth of the filter is small, which allows for better resolution. A larger p, however, requires more autocorrelation lags in the autocorrelation matrix $\hat{\mathbf{R}}$, which increases the variance of the estimated spectrum. Again, we have a trade-off between resolution and variance.

14.4.6 Multiwindow Spectrum Estimator

Many efforts have been made to improve the performance of the periodogram by multiplying the data with a nonrectangular window. The introduction of such windows has been more or less ad hoc, although they have been constructed to have narrow mainlobes and low sidelobes. By contrast, Thomson has proposed a spectrum estimation method that also involves use of windows but is derived from fundamental principles. The method is based on the approximate solution of a Fredholm equation using an eigenexpansion [21]. The method amounts to applying multiple windows to the data, where the windows are discrete prolate spheroidal (Slepian) sequences. These sequences are orthogonal and their Fourier transforms have the maximum energy concentration in a given bandwidth W.

The multiwindow (MW) spectrum estimator is given by [21]

$$\hat{P}_{\text{MW}}(f) = \frac{1}{m} \sum_{i=0}^{m-1} \hat{P}_i(f) \tag{14.69}$$

where the $\hat{P}_i(f)$ is the i-th eigenspectrum defined by

$$\hat{P}_i(f) = \frac{1}{\lambda_i} \left| \sum_{n=0}^{N-1} x[n] w_i[n] e^{-j2\pi fn} \right|^2 \tag{14.70}$$

with $w_i[n]$ being the i-th Slepian sequence, λ_i the i-th Slepian eigenvalue, and W the analysis bandwidth. The steps for obtaining $\hat{P}_{\mathrm{MW}}(f)$ are [22]

1. Selection of the analysis bandwidth W whose typical values are between $1.5/N$ and $20/N$. The number of windows m depends on the selected W, and is given by $\lfloor 2NW \rfloor$, where $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. The spectrum estimator has a resolution equal to W.
2. Evaluation of the m eigenspectra according to Eq. (14.70), where the Slepian sequences and eigenvalues satisfy

$$\mathbf{C}\mathbf{w}_i = \lambda_i \mathbf{w}_i \ . \tag{14.71}$$

   with the elements of the matrix $\mathbf{C}$ being given by

$$c_{mn} = \frac{\sin(2\pi W(m-n))}{\pi(m-n)}, \quad m, n = 1, 2, \cdots, N \ . \tag{14.72}$$

   In the evaluation of the eigenspectra, only the Slepian sequences that correspond to the m largest eigenvalues of $\mathbf{C}$ are used.
3. Computation of the average spectrum according to Eq. (14.69). If the spectrum is mixed, that is, the observed data contain harmonics, the MW method uses a likelihood ratio test to determine if harmonics are present. If the test shows that there is a harmonic around the frequency f_0, the spectrum is reshaped by adding an impulse at f_0 followed by correction of the "local" spectrum for the inclusion of the impulse. For details, see [9, 21].

The MW method is consistent, and its variance for fixed W tends to zero as $1/N$ when $N \to \infty$. The variance, however, as well as the bias and the resolution, depend on the bandwidth W.

14.5 Parametric Spectrum Estimation

A philosophically different approach to spectrum estimation of a random process is the parametric one, which is based on the assumption that the process can be described by a parametric model. Based on the model, the spectrum of the process can then be expressed in terms of the parameters of the model. The approach thus consists of three steps: (1) selection of an appropriate parametric model (usually based on *a priori* knowledge about the process), (2) estimation of the model parameters, and (3) computation of the spectrum using the so-obtained parameters. In the literature the parametric spectrum estimation methods are known as high-resolution methods because they can achieve better resolution than the nonparametric methods.

The most frequently used models in the literature are the autoregressive (AR), the moving average (MA), the autoregressive moving average (ARMA), and the sum of harmonics (complex sinusoids) embedded in noise. With the AR model we assume that the observed data have been generated by a system whose input-output difference equation is given by

$$x[n] = -\sum_{k=1}^{p} a_k x[n-k] + e[n] \tag{14.73}$$

where $x[n]$ is the observed output of the system, $e[n]$ is the unobserved input of the system, and the a_k's are its coefficients. The input $e[n]$ is a zero mean white noise process with unknown variance σ^2, and p is the order of the system. This model is usually abbreviated as AR(p). The MA model is

given by

$$x[n] = \sum_{k=0}^{q} b_k e[n-k] \tag{14.74}$$

where the b_k's denote the MA parameters, $e[n]$ is a zero mean white noise process with unknown variance σ^2, and q is the order of the model. The first MA coefficient b_0 is set usually to be $b_0 = 1$, and the model is denoted by MA(q). The ARMA model combines the AR and MA models and is described by

$$x[n] = -\sum_{k=1}^{p} a_k x[n-k] + \sum_{k=0}^{q} b_k e[n-k]\,. \tag{14.75}$$

Since the AR and MA orders are p and q, respectively, the model in Eq. (14.75) is referred to as ARMA (p, q). Finally, the model of complex sinusoids in noise is

$$x[n] = \sum_{i=1}^{m} A_i e^{j2\pi f_i n} + e[n], \quad n = 0, 1, \cdots, N-1 \tag{14.76}$$

where m is the number of complex sinusoids, A_i and f_i are the complex amplitude and frequency of the i-th complex sinusoid, respectively, and $e[n]$ is a sample of a noise process, which is not necessarily white. Frequently, we assume that the samples $e[n]$ are generated by a certain parametric probability distribution whose parameters are unknown, or $e[n]$ itself is modeled as an AR, MA, or ARMA process.

14.5.1 Spectrum Estimation Based on Autoregressive Models

When the model of $x[n]$ is AR(p), the PSD of the process is given by

$$P_{\mathrm{AR}}(f) = \frac{\sigma^2}{|1 + \sum_{k=1}^{p} a_k e^{-j2\pi fk}|^2}\,. \tag{14.77}$$

Thus, to find $P_{\mathrm{AR}}(f)$ we need the estimates of the AR coefficients a_k and the noise variance σ^2.

If we multiply the two sides of Eq. (14.73) by $x^*[n-k]$, $k \geq 0$, and take their expectations, we obtain

$$E(x[n]x^*[n-k]) = -\sum_{l=1}^{p} a_l E(x[n-l]x^*[n-k]) + E(e[n]x^*[n-k]) \tag{14.78}$$

or

$$r[k] = \begin{cases} -\sum_{l=1}^{p} a_l r[k-l], & k > 0 \\ -\sum_{l=1}^{p} a_l r[k-l] + \sigma^2, & k = 0 \end{cases}\,. \tag{14.79}$$

The expressions in Eq. (14.79) are known as the Yule-Walker equations. To estimate the p unknown AR coefficients from Eq. (14.79), we need at least p equations as well as the estimates of the appropriate autocorrelations. The set of equations that requires the estimation of the minimum number of correlation lags is

$$\hat{\mathbf{R}}\mathbf{a} = -\hat{\mathbf{r}} \tag{14.80}$$

where $\hat{\mathbf{R}}$ is the $p \times p$ matrix

$$\hat{\mathbf{R}} = \begin{bmatrix} \hat{r}[0] & \hat{r}[-1] & \hat{r}[-2] & \cdots & \hat{r}[-p+1] \\ \hat{r}[1] & \hat{r}[0] & \hat{r}[-1] & \cdots & \hat{r}[-p+2] \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{r}[p-1] & \hat{r}[p-2] & \hat{r}[p-3] & \cdots & \hat{r}[0] \end{bmatrix} \tag{14.81}$$

and

$$\hat{\mathbf{r}} = [\hat{r}[1]\ \hat{r}[2]\ \cdots\ \hat{r}[p]]^T\ . \tag{14.82}$$

The parameters **a** are estimated by

$$\hat{\mathbf{a}} = -\hat{\mathbf{R}}^{-1}\hat{\mathbf{r}} \tag{14.83}$$

and the noise variance is found from

$$\hat{\sigma}^2 = \hat{r}[0] + \sum_{k=1}^{p} a_k \hat{r}^*[k]. \tag{14.84}$$

The PSD estimate is obtained when $\hat{\mathbf{a}}$ and $\hat{\sigma}^2$ are substituted in Eq. (14.77). This approach for estimating the AR parameters is known in the literature as the autocorrelation method.

Many other AR estimation procedures have been proposed including the maximum likelihood method, the covariance method, and the Burg method [12]. Burg's work in the late sixties has a special place in the history of spectrum estimation because it kindled the interest in this field. Burg showed that the AR model provides an extrapolation of a known autocorrelation sequence $r[k]$, $|k| \leq p$, for $|k|$ beyond p so that the spectrum corresponding to the extrapolated sequence is the flattest of all spectra consistent with the $2p + 1$ known autocorrelations [4].

An important issue in finding the AR PSD is the order of the assumed AR model. There exist several model order selection procedures, but the most widely used are the Information Criterion A (AIC) due to Akaike [2] and the Information Criterion B (BIC), also known as the Minimum Description Length (MDL) principle, of Rissanen [16] and Schwarz [20]. According to the AIC criterion, the best model is the one that minimizes the function $AIC(k)$ over k defined by

$$AIC(k) = N \log \hat{\sigma}_k^2 + 2k \tag{14.85}$$

where k is the model order, and $\hat{\sigma}_k^2$ is the estimated noise variance of that model. Similarly, the MDL criterion chooses the order which minimizes the function $MDL(k)$ defined by

$$MDL(k) = N \log \hat{\sigma}_k^2 + k \log N \tag{14.86}$$

where N is the number of observed data samples. It is important to emphasize that the MDL rule can be derived if, as a criterion for model selection, we use the maximum *a posteriori* principle. It has been found that the AIC is an inconsistent criterion whereas the MDL rule is consistent. Consistency here means that the probability of choosing the correct model order tends to one as $N \to \infty$.

The AR-based spectrum estimation methods show very good performance if the processes are narrowband and have sharp peaks in their spectra. Also, many good results have been reported when they are applied to short data records.

14.5.2 Spectrum Estimation Based on Moving Average Models

The PSD of a moving average process is given by

$$P_{\mathrm{MA}}(f) = \sigma^2 |1 + \sum_{k=1}^{q} b_k e^{-j2\pi fk}|^2\ . \tag{14.87}$$

It is not difficult to show that the $r[k]$'s for $|k| > q$ of an MA(q) process are identically equal to zero, and that Eq. (14.87) can be expressed also as

$$P_{\mathrm{MA}}(f) = \sum_{k=-q}^{q} r[k] e^{-j2\pi fk}\ . \tag{14.88}$$

Thus, to find $\hat{P}_{\mathrm{MA}}(f)$ it would be sufficient to estimate the autocorrelations $r[k]$ and use the found estimates in Eq. (14.88). Obviously, this estimate would be identical to $\hat{P}_{\mathrm{BT}}(f)$ when the applied window is rectangular and of length $2q+1$.

A different approach is to find the estimates of the unknown MA coefficients and σ^2 and use them in Eq. (14.87). The equations of the MA coefficients are nonlinear, which makes their estimation difficult. Durbin has proposed an approximate procedure that is based on a high order AR approximation of the MA process. First the data are modeled by an AR model of order L, where $L >> q$. Its coefficients are estimated from Eq. (14.83) and $\hat{\sigma}^2$ according to Eq. (14.84). Then the sequence $1, \hat{a}_1, \hat{a}_2, \cdots, \hat{a}_L$ is fitted with an AR(q) model, whose parameters are also estimated using the autocorrelation method. The estimated coefficients $\hat{b}_1, \hat{b}_2, \cdots, \hat{b}_q$ are subsequently substituted in Eq. (14.87) together with $\hat{\sigma}^2$.

Good results with MA models are obtained when the PSD of the process is characterized by broad peaks and sharp nulls. The MA models should not be used for processes with narrowband features.

14.5.3 Spectrum Estimation Based on Autoregressive Moving Average Models

The PSD of a process that is represented by the ARMA model is given by

$$P_{\mathrm{ARMA}}(f) = \sigma^2 \frac{|1 + \sum_{k=1}^{q} b_k e^{-j2\pi fk}|^2}{|1 + \sum_{k=1}^{p} a_k e^{-j2\pi fk}|^2} \,. \tag{14.89}$$

The ML estimates of the ARMA coefficients are difficult to obtain, so we usually resort to methods that yield suboptimal estimates. For example, we can first estimate the AR coefficients based on the equation,

$$\begin{bmatrix} \hat{r}[q] & \hat{r}[q-1] & \cdots & \hat{r}[q-p+1] \\ \hat{r}[q+1] & \hat{r}[q] & \cdots & \hat{r}[q-p+2] \\ \vdots & \vdots & \vdots & \vdots \\ \hat{r}[M-1] & \hat{r}[M-2] & \cdots & \hat{r}[M-p] \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_p \end{bmatrix} + \begin{bmatrix} \epsilon_{q+1} \\ \epsilon_{q+2} \\ \vdots \\ \epsilon_M \end{bmatrix} = - \begin{bmatrix} \hat{r}[q+1] \\ \hat{r}[q+2] \\ \vdots \\ \hat{r}[M] \end{bmatrix} \tag{14.90}$$

or

$$\hat{\mathbf{R}}\mathbf{a} + \boldsymbol{\epsilon} = -\hat{\mathbf{r}} \tag{14.91}$$

where ϵ_i is a term that models the errors in the Yule-Walker equations due to the estimation errors of the autocorrelation lags, and $M \geq p+q$. From Eq. (14.91), we can find the least squares estimates of $\mathbf{a}$ by

$$\hat{\mathbf{a}} = -\left(\hat{\mathbf{R}}^H \hat{\mathbf{R}}\right)^{-1} \hat{\mathbf{R}}^H \hat{\mathbf{r}} \,. \tag{14.92}$$

This procedure is known as the least-squares modified Yule-Walker equation method. Once the AR coefficients are estimated, we can filter the observed data

$$y[n] = x[n] + \sum_{k=1}^{p} \hat{a}_k x[n-k] \tag{14.93}$$

and obtain a sequence that is approximately modeled by an MA(q) model. From the data $y[n]$ we can estimate the MA PSD by Eq. (14.88) and obtain the PSD estimate of the data $x[n]$

$$\hat{P}_{\mathrm{ARMA}}(f) = \frac{\hat{P}_{\mathrm{MA}}(f)}{|1 + \sum_{k=1}^{p} \hat{a}_k e^{-j2\pi fk}|^2} \tag{14.94}$$

or estimate the parameters $b_1, b_2, ..., b_q$ and σ^2 by Durbin's method, for example, and then use

$$\hat{P}_{\text{ARMA}}(f) = \hat{\sigma}^2 \frac{|1 + \sum_{k=1}^{q} \hat{b}_k e^{-j2\pi fk}|^2}{|1 + \sum_{k=1}^{p} \hat{a}_k e^{-j2\pi fk}|^2} . \tag{14.95}$$

The ARMA model has an advantage over the AR and MA models because it can better fit spectra with nulls and peaks. Its disadvantage is that it is more difficult to estimate its parameters than the parameters of the AR and MA models.

14.5.4 Pisarenko Harmonic Decomposition Method

Let the observed data represent m complex sinusoids in noise, i.e.,

$$x[n] = \sum_{i=1}^{m} A_i e^{j2\pi f_i n} + e[n], \quad n = 0, 1, \cdots, N-1 \tag{14.96}$$

where f_i is the frequency of the i-th complex sinusoid, A_i is the complex amplitude of the i-th sinusoid,

$$A_i = |A_i| e^{j\phi_i} \tag{14.97}$$

with ϕ_i being a random phase of the i-th complex sinusoid, and $e[n]$ is a sample of a zero mean white noise. The PSD of the process is a sum of the continuous spectrum of the noise and a set of impulses with area $|A_i|^2$ at the frequencies f_i, or

$$P(f) = \sum_{i=1}^{m} |A_i|^2 \delta(f - f_i) + P_e(f) \tag{14.98}$$

where $P_e(f)$ is the PSD of the noise process.

Pisarenko studied the model in Eq. (14.96) and found that the frequencies of the sinusoids can be obtained from the eigenvector corresponding to the smallest eigenvalue of the autocorrelation matrix. His method, known as Pisarenko harmonic decomposition (PHD), led to important insights and stimulated further work which resulted in many new procedures known today as "signal and noise subspace" methods.

When the noise $\{\tilde{e}[n]\}$ is zero mean white with variance σ^2, the autocorrelation of $\{\tilde{x}[n]\}$ can be written as

$$r[k] = \sum_{i=1}^{m} |A_i|^2 e^{j2\pi f_i k} + \sigma^2 \delta[k] \tag{14.99}$$

or the autocorrelation matrix can be represented by

$$\mathbf{R} = \sum_{i=1}^{m} |A_i|^2 \mathbf{e}_i \mathbf{e}_i^H + \sigma^2 \mathbf{I} \tag{14.100}$$

where

$$\mathbf{e}_i = \left[1 \; e^{j2\pi f_i} \; e^{j4\pi f_i} \; e^{j2\pi(N-1)f_i}\right]^T \tag{14.101}$$

and $\mathbf{I}$ is the identity matrix. It is seen that the autocorrelation matrix $\mathbf{R}$ is composed of the sum of signal and noise autocorrelation matrices

$$\mathbf{R} = \mathbf{R}_s + \sigma^2 \mathbf{I} \tag{14.102}$$

where

$$\mathbf{R}_s = \mathbf{E}\mathbf{P}\mathbf{E}^H \tag{14.103}$$

for

$$\mathbf{E} = [\mathbf{e}_1\ \mathbf{e}_2\ \cdots\ \mathbf{e}_m] \tag{14.104}$$

and $\mathbf{P}$ a diagonal matrix

$$\mathbf{P} = \text{diag}\left\{|A_1|^2,\ |A_2|^2,\ \cdots,\ |A_m|^2\right\}\ . \tag{14.105}$$

If the matrix $\mathbf{R}_s$ is $M \times M$, where $M \geq m$, its rank will be equal to the number of complex sinusoids m. Another important representation of the autocorrelation matrix $\mathbf{R}$ is via its eigenvalues and eigenvectors, i.e.,

$$\mathbf{R} = \sum_{i=1}^{m}(\lambda_i + \sigma^2)\mathbf{v}_i\mathbf{v}_i^H + \sum_{i=m+1}^{M}\sigma^2\mathbf{v}_i\mathbf{v}_i^H \tag{14.106}$$

where the λ_i's, $i = 1, 2, \cdots, m$, are the nonzero eigenvalues of $\mathbf{R}_s$. Let the eigenvalues of $\mathbf{R}$ be arranged in decreasing order so that $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M$, and let $\mathbf{v}_i$ be the eigenvector corresponding to λ_i. The space spanned by the eigenvectors $\mathbf{v}_i$, $i = 1, 2, \cdots, m$, is called the signal subspace, and the space spanned by $\mathbf{v}_i$, $i = m+1, m+2, \cdots, M$, the noise subspace. Since the set of eigenvectors are orthonormal, that is

$$\mathbf{v}_i^H\mathbf{v}_l = \begin{cases} 1, & i = l \\ 0, & i \neq l \end{cases} \tag{14.107}$$

the two subspaces are orthogonal. In other words if $\mathbf{s}$ is in the signal subspace, and $\mathbf{z}$ is in the noise subspace, then $\mathbf{s}^H\mathbf{z} = 0$.

Now suppose that the matrix $\mathbf{R}$ is $(m+1) \times (m+1)$. Pisarenko observed that the noise variance corresponds to the smallest eigenvalue of $\mathbf{R}$ and that the frequencies of the complex sinusoids can be estimated by using the orthogonality of the signal and noise subspaces, that is,

$$\mathbf{e}_i^H\mathbf{v}_{m+1} = 0, \quad i = 1, 2, \cdots, m\ . \tag{14.108}$$

We can estimate the f_i's by forming the pseudospectrum

$$\hat{P}_{\text{PHD}}(f) = \frac{1}{|\mathbf{e}^H(f)\mathbf{v}_{m+1}|^2} \tag{14.109}$$

which should theoretically be infinite at the frequencies f_i. In practice, however, the pseudospectrum does not exhibit peaks exactly at these frequencies because $\mathbf{R}$ is not known and, instead, is estimated from finite data records.

The PSD estimate in Eq. (14.109) does not include information about the power of the noise and the complex sinusoids. The powers, however, can easily be obtained by using Eq. (14.98). First note that $P_e(f) = \sigma^2$, and $\hat{\sigma}^2 = \lambda_{m+1}$. Second, the frequencies f_i are determined from the pseudospectrum Eq. (14.109), so it remains to find the powers of the complex sinusoids $P_i = |A_i|^2$. This can readily be accomplished by using the set of m linear equations

$$\begin{bmatrix} |\hat{\mathbf{e}}_1^H\mathbf{v}_1|^2 & |\hat{\mathbf{e}}_2^H\mathbf{v}_1|^2 & \cdots & |\hat{\mathbf{e}}_m^H\mathbf{v}_1|^2 \\ |\hat{\mathbf{e}}_1^H\mathbf{v}_2|^2 & |\hat{\mathbf{e}}_2^H\mathbf{v}_2|^2 & \cdots & |\hat{\mathbf{e}}_m^H\mathbf{v}_2|^2 \\ \vdots & \vdots & \vdots & \vdots \\ |\hat{\mathbf{e}}_1^H\mathbf{v}_m|^2 & |\hat{\mathbf{e}}_2^H\mathbf{v}_m|^2 & \cdots & |\hat{\mathbf{e}}_m^H\mathbf{v}_m|^2 \end{bmatrix} \begin{bmatrix} P_1 \\ P_2 \\ \vdots \\ P_m \end{bmatrix} = \begin{bmatrix} \lambda_1 - \hat{\sigma}^2 \\ \lambda_2 - \hat{\sigma}^2 \\ \vdots \\ \lambda_m - \hat{\sigma}^2 \end{bmatrix} \tag{14.110}$$

where

$$\hat{\mathbf{e}}_i = \left[1\ e^{j2\pi\hat{f}_i}\ e^{j4\pi\hat{f}_i}\ \cdots\ e^{j2\pi(N-1)\hat{f}_i}\right]^T\ . \tag{14.111}$$

In summary, Pisarenko's method consists of four steps:

1. Estimate the $(m+1) \times (m+1)$ autocorrelation matrix $\mathbf{R}$ (provided it is known that the number of complex sinusoids is m).
2. Evaluate the minimum eigenvalue λ_{m+1} and the eigenvectors of $\hat{\mathbf{R}}$.
3. Set the white noise power to $\hat{\sigma^2} = \lambda_{m+1}$, estimate the frequencies of the complex sinusoids from the peak locations of $\hat{P}_{\text{PHD}}(f)$ in Eq. (14.109), and compute their powers from Eq. (14.110).
4. Substitute the estimated parameters in Eq. (14.98).

Pisarenko's method is not used frequently in practice because its performance is much poorer than the performance of some other signal and noise subspace based methods developed later.

14.5.5 Multiple Signal Classification (MUSIC)

A procedure very similar to Pisarenko's is the MUltiple SIgnal Classification (MUSIC) method, which was proposed in the late 1970's by Schmidt [18]. Suppose again that the process $\{\tilde{x}[n]\}$ is described by m complex sinusoids in white noise. If we form an $M \times M$ autocorrelation matrix $\mathbf{R}$, find its eigenvalues and eigenvectors and rank them as before, then as mentioned in the previous subsection, its m eigenvectors corresponding to the m largest eigenvalues span the signal subspace, and the remaining eigenvectors, the noise subspace. According to MUSIC, we estimate the noise variance from the $M - m$ smallest eigenvalues of $\hat{\mathbf{R}}$

$$\hat{\sigma}^2 = \frac{1}{M-m} \sum_{i=m+1}^{M} \lambda_i \tag{14.112}$$

and the frequencies from the peak locations of the pseudospectrum

$$\hat{P}_{\text{MU}}(f) = \frac{1}{\sum_{i=m+1}^{M} |\mathbf{e}(f)^H \mathbf{v}_i|^2} \,. \tag{14.113}$$

It should be noted that there are other ways of estimating the f_i's. Finally the powers of the complex sinusoids are determined from Eq. (14.110), and all the estimated parameters substituted in Eq. (14.98).

MUSIC has better performance than Pisarenko's method because of the introduced averaging via the extra noise eigenvectors. The averaging reduces the statistical fluctuations present in Pisarenko's pseudospectrum, which arise due to the errors in estimating the autocorrelation matrix. These fluctuations can further be reduced by applying the Eigenvector method [11], which is a modification of MUSIC and whose pseudospectrum is given by

$$\hat{P}_{\text{EV}}(f) = \frac{1}{\sum_{i=m+1}^{M} |\frac{1}{\lambda_i} \mathbf{e}(f)^H \mathbf{v}_i|^2} \,. \tag{14.114}$$

Pisarenko's method, MUSIC, and its variants exploit the noise subspace to estimate the unknown parameters of the random process. There are, however, approaches that estimate the unknown parameters from vectors that lie in the signal subspace. The main idea there is to form a reduced rank autocorrelation matrix which is an estimate of the signal autocorrelation matrix. Since this estimate is formed from the m principal eigenvectors and eigenvalues, the methods based on them are called principal component spectrum estimation methods [8, 12]. Once the signal autocorrelation matrix is obtained, the frequencies of the complex sinusoids are found, followed by estimation of the remaining unknown parameters of the model.

14.6 Recent Developments

Spectrum estimation continues to attract the attention of many researchers. The answers to many interesting questions are still unknown, and many problems still need better solutions. The field of spectrum estimation is constantly enriched with new theoretical findings and a wide range of results obtained from examinations of various physical processes. In addition, new concepts are being introduced that provide tools for improved processing of the observed signals and that allow for a better understanding. Many new developments are driven by the need to solve specific problems that arise in applications, such as in sonar and communications.

Recently, for example, the notion of canonical autoregressive decomposition has been introduced [14]. It is a parametric approach for estimation of mixed spectra where the continuous part of the spectrum is modeled by an AR model. Another development is related to Bayesian spectrum estimation. Jaynes has introduced it in [10] and some interesting results for spectra of harmonics in white Gaussian noise have been reported in [7]. A Bayesian spectrum estimate is based on

$$\hat{P}_{\mathrm{BA}}(f) = \int_{\boldsymbol{\Theta}} P(f, \boldsymbol{\theta}) f(\boldsymbol{\theta} | \{x[n]\}_0^{N-1}) d\boldsymbol{\theta} \tag{14.115}$$

where $P(f, \boldsymbol{\theta})$ is the theoretical parametric spectrum, $\boldsymbol{\theta}$ denotes the parameters of the process, $\boldsymbol{\Theta}$ is the parameter space, and $f(\boldsymbol{\theta} | \{x[n]\}_0^{N-1})$ is the *a posteriori* probability density function of the process parameters. Therefore, the Bayesian spectrum estimate is defined as the expected value of the theoretical spectrum over the joint posterior density function of the model parameters.

The processes that we have addressed here are wide-sense stationary. The stationarity assumption, however, is often a mathematical abstraction and only an approximation in practice. Many physical processes are actually nonstationary and their spectra change with time. In biomedicine, speech analysis, and sonar, for example, it is typical to observe signals whose power during some time intervals is concentrated at high frequencies and, shortly thereafter, at low or middle frequencies. In such cases it is desirable to describe the PSD of the process at every instant of time, which is possible if we assume that the spectrum of the process changes smoothly over time. Such description requires a combination of the time- and frequency-domain concepts of signal processing into a single framework [6]. So there is an important distinction between the PSD estimation methods discussed here and the time-frequency representation approaches. The former provide the PSD of the process for all times, whereas the latter yield the local PSD's at every instant of time. This area of research is well developed but still far from complete. Although many theories have been proposed and developed, including evolutionary spectra [15], the Wigner-Wille method [13], and the kernel choice approach [1], time-varying spectrum analysis has remained a challenging and fascinating area of research.

References

[1] Amin, M.G., Time-frequency spectrum analysis and estimation for nonstationary random processes, in *Time-Frequency Signal Analysis*, B. Boashash, Ed., pp. 208–232, Longman Cheshire, 1992.

[2] Akaike, H., A new look at the statistical model identification, *IEEE Trans. Automatic Control*, Vol. AC-19, pp. 716–723, 1974.

[3] Blackman, R.B. and Tukey, J.W., *The Measurement of Power Spectra from the Point of View of Communications Engineering*, Dover Publications, New York, 1958.

[4] Burg, J.P., Maximum Entropy Spectral Analysis, Ph.D. dissertation, Stanford University, 1975.

[5] Capon, J., High-resolution frequency-wavenumber spectrum analysis, *Proc. IEEE*, Vol. 57, pp. 1408–1418, 1969.

[6] Cohen, L., *Time-Frequency Analysis*, Prentice Hall, Englewood Cliffs, NJ, 1995.

[7] Djurić, P.M. and Li, H.-T., Bayesian spectrum estimation of harmonic signals, *Signal Process. Lett.*, Vol. 2, pp. 213–215, 1995.
[8] Hayes, M.S., *Statistical Digital Signal Processing and Modeling*, John Wiley & Sons, New York, 1996.
[9] Haykin, S., *Advances in Spectrum Analysis and Array Processing*, Prentice Hall, Englewood Cliffs, NJ, 1991.
[10] Jaynes, E.T., Bayesian spectrum and chirp analysis, in *Maximum Entropy and Bayesian Spectral Analysis and Estimation Problems*, C. R. Smith and G. J. Erickson, Eds., pp. 1–37, D. Reidel, Dordrecht, Holland, 1987.
[11] Johnson, D.H. and DeGraaf, S.R., Improving the resolution of bearing in passive sonar arrays by eigenvalue analysis, *IEEE Trans. Acoustics, Speech, Signal Process.*, Vol. ASSP-30, pp. 638–647, 1982.
[12] Kay, S.M., *Modern Spectral Estimation*, Prentice Hall, Englewood Cliffs, NJ, 1988.
[13] Martin, W. and Flandrin, P., Wigner-Ville spectral analysis of nonstationary processes, *IEEE Trans. Acoustics, Speech, Signal Process.*, Vol. 33, pp. 1461–1470, 1985.
[14] Nagesha, V. and Kay, S.M., Spectral analysis based on the canonical autoregressive decomposition, *IEEE Trans. Signal Process.*, Vol. SP-44, pp. 1719–1733, 1996.
[15] Priestley, M.B., *Spectral Analysis and Time Series*, Academic Press, New York, 1981.
[16] Rissanen, J., Modeling by shortest data description, *Automatica*, Vol. 14, pp. 465–471, 1978.
[17] Robinson, E.A., A historical perspective of spectrum estimation, *Proc. IEEE*, Vol. 70, pp. 885–907, 1982.
[18] Schmidt, R., Multiple emitter location and signal parameter estimation, *Proc. RADC Spectrum Estimation Workshop*, pp. 243–258, 1979.
[19] Schuster, A., On the investigation on hidden periodicities with application to a supposed 26-day period of meteorological phenomena, *Terrestrial Magnetism*, Vol. 3, pp. 13–41, 1898.
[20] Schwarz, G., Estimating the dimension of the model, *Annals Statist.*, Vol. 6, pp. 461–464, 1978.
[21] Thomson, D.J., Spectrum estimation and harmonic analysis, *Proc. IEEE*, Vol. 70, pp. 1055–1096, 1982.
[22] Thomson, D.J., Quadratic-inverse spectrum estimates: applications to paleoclimatology, *Phil. Trans. R. Soc. London*, A, Vol. 332, pp. 539–597, 1990.