# 4

# Parametric Methods for Line Spectra

## 4.1 INTRODUCTION

In several applications, particularly in communications, radar, sonar, and geophysical seismology, the signals dealt with can be described well by the *sinusoidal model*

$$y(t) = x(t) + e(t) \quad ; \quad x(t) = \sum_{k=1}^{n} \alpha_k e^{i(\omega_k t + \varphi_k)} \tag{4.1.1}$$

where $x(t)$ denotes the noise-free complex-valued sinusoidal signal; $\{\alpha_k\}$, $\{\omega_k\}$, $\{\varphi_k\}$ are its *amplitudes*, *(angular) frequencies*, and *initial phases*, respectively; and $e(t)$ is an additive observation noise. The complex-valued form (4.1.1), of course, is not encountered in practice as it stands; practical signals are real valued. However, as already mentioned in Chapter 1, in many applications both the *in-phase and quadrature components* of the studied signal are available. (See Chapter 6 for more details on this aspect.) In the case of a (real-valued) sinusoidal signal, this means that both the sine and the corresponding cosine components are available. These two components may be processed by arranging them in a two-dimensional vector signal or a complex-valued signal of the form of (4.1.1). Since the complex-valued description (4.1.1) of the in-phase and quadrature components of a sinusoidal signal is the most convenient one from a mathematical standpoint, we focus on it in this chapter.

The noise $\{e(t)\}$ in (4.1.1) is usually assumed to be (complex-valued) *circular white noise*, defined in (2.4.19). We also make the white-noise assumption in this chapter. We may argue in the following way that the white-noise assumption is not particularly restrictive. Let the continuous-time counterpart of the noise in (4.1.1) be correlated, but assume that the "correlation time" of the continuous-time noise is less than half of the shortest period of the sine-wave components in the continuous-time counterpart of $x(t)$ in (4.1.1). If this mild condition is satisfied, then choosing the sampling period larger than the noise correlation time (yet smaller than half the shortest sinusoidal signal period, to avoid aliasing) results in a white discrete-time noise sequence $\{e(t)\}$. If the correlation condition above is not satisfied, but we know the shape of the noise spectrum, we can filter $y(t)$ by a linear *whitening filter* that makes the noise component at the filter output white; the sinusoidal components remain sinusoidal with the same frequencies, but with amplitudes and phases altered in a known way.

If the noise process is not white and has unknown spectral shape, then accurate frequency estimates can still be found if we estimate the sinusoids, using the nonlinear least squares (NLS) method in Section 4.3. (See [Stoica and Nehorai 1989b], for example.) Indeed, the properties of the NLS estimates in the colored-noise and unknown-noise cases are quite similar to those for the white-noise case, only with the sinusoidal signal amplitudes "adjusted" to give corresponding local SNRs—the signal-to-noise power ratio at each frequency $\omega_k$. This amplitude adjustment is the same as that realized by the whitening filter approach. It is important to note that these comments apply only if the NLS method is used. The other estimation methods in this chapter (e.g., the subspace-based methods) depend on the assumption that the noise is white and can be affected adversely if the noise is not white (or is not prewhitened).

Concerning the signal in (4.1.1), we assume that $\omega_k \in [-\pi, \pi]$ and that $\alpha_k > 0$. We need to specify the sign of $\{\alpha_k\}$; otherwise we are left with a phase ambiguity. More precisely, without the condition $\alpha_k > 0$ in (4.1.1), both $\{\alpha_k, \omega_k, \varphi_k\}$ and $\{-\alpha_k, \omega_k, \varphi_k + \pi\}$ give the same signal $\{x(t)\}$, so the parameterization is not unique. As to the initial phases $\{\varphi_k\}$ in (4.1.1), one could assume that they are fixed (nonrandom) constants, which would result in $\{x(t)\}$ being a deterministic signal. In most applications, however, $\{\varphi_k\}$ are *nuisance parameters*, and it is more convenient to assume that they are random variables. Note that, if we try to mimic the conditions of a previous experiment as much as possible, we will usually be unable to ensure the same initial phases of the sine waves in the observed sinusoidal signal (this will be particularly true for received signals). Since there is usually no reason to believe that a specific set of initial phases is more likely than another one, or that two different initial phases are interrelated, we make the following assumption:

> The initial phases $\{\varphi_k\}$ are independent random variables uniformly distributed on $[-\pi, \pi]$. (4.1.2)

The covariance function and the PSD of the noisy sinusoidal signal $\{y(t)\}$ can be calculated in a straightforward manner under these assumptions. By using (4.1.2), we get

$$E\left\{e^{i\varphi_p} e^{-i\varphi_j}\right\} = 1 \quad \text{for} \quad p = j$$

and for $p \neq j$

$$E\left\{e^{i\varphi_p}e^{-i\varphi_j}\right\} = E\left\{e^{i\varphi_p}\right\}E\left\{e^{-i\varphi_j}\right\}$$
$$= \left[\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{i\varphi}d\varphi\right]\left[\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-i\varphi}d\varphi\right] = 0$$

Thus,

$$E\left\{e^{i\varphi_p}e^{-i\varphi_j}\right\} = \delta_{p,j} \tag{4.1.3}$$

Let

$$x_p(t) = \alpha_p e^{i(\omega_p t+\varphi_p)} \tag{4.1.4}$$

denote the $p$th sinusoid in (4.1.1). It follows from (4.1.3) that

$$E\left\{x_p(t)x_j^*(t-k)\right\} = \alpha_p^2 e^{i\omega_p k}\delta_{p,j} \tag{4.1.5}$$

which, in turn, gives

$$\boxed{r(k) = E\left\{y(t)y^*(t-k)\right\} = \sum_{p=1}^{n}\alpha_p^2 e^{i\omega_p k} + \sigma^2\delta_{k,0}} \tag{4.1.6}$$

where $\sigma^2$ is the variance of $e(t)$. The derivation of the covariance function of $y(t)$ is completed. The PSD of $y(t)$ is given by the DTFT of $\{r(k)\}$ in (4.1.6), which is

$$\boxed{\phi(\omega) = 2\pi\sum_{p=1}^{n}\alpha_p^2\delta(\omega-\omega_p) + \sigma^2} \tag{4.1.7}$$

where $\delta(\omega-\omega_p)$ is the Dirac impulse (or Dirac delta "function") which, by definition, has the property that

$$\int_{-\pi}^{\pi} F(\omega)\delta(\omega-\omega_p)\,d\omega = F(\omega_p) \tag{4.1.8}$$

for any function $F(\omega)$ that is continuous at $\omega_p$. The expression (4.1.7) for $\phi(\omega)$ may be verified by inserting it in the inverse transform formula (1.3.8) and checking that the result is the covariance function. Doing so, we obtain

$$\frac{1}{2\pi}\int_{-\pi}^{\pi}\left[2\pi\sum_{p=1}^{n}\alpha_p^2\delta(\omega-\omega_p)+\sigma^2\right]e^{i\omega k}\,d\omega = \sum_{p=1}^{n}\alpha_p^2 e^{i\omega_p k} + \sigma^2\delta_{k,o} = r(k) \tag{4.1.9}$$

which is the desired result.

The PSD (4.1.7) is depicted in Figure 4.1. It consists of a "floor" of constant level equal to the noise power $\sigma^2$, along with $n$ vertical lines (or impulses) located at the sinusoidal frequencies $\{\omega_k\}$ and having zero support but nonzero areas equal to $2\pi$ times the sine wave powers $\{\alpha_k^2\}$. Owing to its appearance, as exhibited in Figure 4.1, $\phi(\omega)$ in (4.1.7) is called a *line* or *discrete spectrum*.

It is evident from the previous discussion that a spectral analysis based on the parametric PSD model (4.1.7) reduces to the problem of estimating the parameters of the signal in (4.1.1). In most applications, such as those listed at the beginning of this chapter, the parameters of major interest are the locations of the spectral lines—namely, the sinusoidal frequencies. In the next sections, we present a number of methods for *spectral line analysis*. We focus on the problem of *frequency estimation*, meaning estimation of $\{\omega_k\}_{k=1}^n$ from a set of observations $\{y(t)\}_{t=1}^N$. Once the frequency estimates have been obtained, estimation of the other signal parameters (or PSD parameters) becomes a simple *linear regression problem*. More precisely, for given $\{\omega_k\}$, the observations $y(t)$ can be written as a linear regression function whose coefficients are equal to the remaining unknowns $\{\alpha_k e^{i\varphi_k} \triangleq \beta_k\}$:

$$y(t) = \sum_{k=1}^{n} \beta_k e^{i\omega_k t} + e(t) \tag{4.1.10}$$

If desired, $\{\beta_k\}$ (and hence $\{\alpha_k\}$, $\{\varphi_k\}$) in (4.1.10) can be obtained by a least-squares method (as in equation (4.3.8)). Alternatively, one may determine the signal powers $\{\alpha_k^2\}$—for given $\{\omega_k\}$—from the sample version of (4.1.6):

$$\hat{r}(k) = \sum_{p=1}^{n} \alpha_p^2 e^{i\omega_p k} + \text{residuals} \qquad \text{for } k \geq 1 \tag{4.1.11}$$

where the residuals arise from finite-sample estimation of $r(k)$; this is, once more, a linear regression with $\{\alpha_p^2\}$ as unknown coefficients. The solution to either linear regression problem is straightforward and is discussed in Section A.8 of Appendix A.

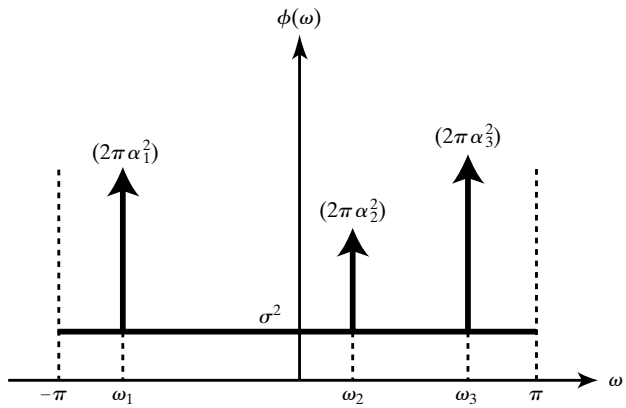


**Figure 4.1** The PSD of a complex sinusoidal signal in additive white noise.

The methods for frequency estimation that will be described in the following sections are sometimes called *high-resolution* (or, even, *superresolution*) techniques. This is due to their ability to resolve spectral lines separated in frequency $f = \omega/2\pi$ by less than $1/N$ cycles per sampling interval, which is the resolution limit for the classical periodogram-based methods. All of the high-resolution methods to be discussed in the following provide *consistent estimates* of $\{\omega_k\}$ under the assumptions we made. Their consistency will surface in the following discussion in an obvious manner; hence, we do not need to pay special attention to this aspect. Nor do we discuss in detail other statistical properties of the frequency estimates obtained by these high-resolution methods, though in Appendix B we review the Cramér–Rao bound and the best accuracy that can be achieved by such methods. For derivations and discussions of the statistical properties not addressed in this text, we refer the interested reader to [Stoica, Söderström, and Ti 1989; Stoica and Söderström 1991; Stoica, Moses, Friedlander, and Söderström 1989; Stoica and Nehorai 1989b]. Let us briefly summarize the conclusions of these analyses: All the high-resolution methods presented in the following provide very accurate frequency estimates, with only small differences in their statistical performances. Furthermore, the computational burdens associated with these methods are rather similar. Hence, selecting one of the high-resolution methods for frequency estimation is essentially a "matter of taste," even though we will identify some advantages of one of these methods, named ESPRIT, over the others.

We should point out that the comparison in the previous paragraph between the high-resolution methods and the periodogram-based techniques is unfair, in the sense that periodogram-based methods do not assume any knowledge about the data, whereas high-resolution methods exploit an exact description of the studied signal. The additional information assumed allows a parametric method to offer better resolution than the nonparametric method of the periodogram. On the other hand, when no two spectral lines in the spectrum are separated by less than $1/N$, the *unmodified periodogram* turns out to be an excellent frequency estimator which may outperform any of the high-resolution methods (as we shall see). One may ask why the *unmodified* periodogram is preferred over the many windowed or smoothed periodogram techniques to which we paid so much attention in Chapter 2. The explanation actually follows from the discussion in that chapter. The unmodified periodogram can be viewed as a Blackman–Tukey "windowed" estimator with a rectangular window of maximum length equal to $2N + 1$. Of all window sequences, this is exactly the one which has the narrowest main lobe and, hence, the one that affords the maximum spectral resolution, a desirable property for high-resolution spectral-line scenarios. It should be noted, however, that if the sinusoidal components in the signal are not very closely spaced in frequency, but their amplitudes differ significantly from one another, then a mildly windowed periodogram (to avoid leakage) might perform better than the unwindowed periodogram. In the unwindowed periodogram, the weaker sinusoids could be obscured by the leakage from the stronger ones, and hence they might not be visible in a plot of the estimated spectrum.

In order to simplify the discussion in this chapter, we assume that the number of sinusoidal components, $n$, in (4.1.1) is known. When $n$ is unknown, as could well be the case in many applications, it can be estimated from the available data, in a way, for example, described in [Fuchs 1988; Kay 1988; Marple 1987; Proakis, Rader, Ling, and Nikias 1992; Söderström and Stoica 1989] and in Appendix C.

## 4.2 MODELS OF SINUSOIDAL SIGNALS IN NOISE

The frequency estimation methods presented in this chapter rely on three different models for the noisy sinusoidal signal (4.1.1). This section introduces the three models of (4.1.1).

### 4.2.1 Nonlinear Regression Model

The nonlinear regression model is given by (4.1.1). Note that the $\{\omega_k\}$ enter in a nonlinear fashion in (4.1.1), hence the name "nonlinear regression" given to this type of model for $\{y(t)\}$. The other two models for $\{y(t)\}$, to be discussed in the following, are derived from (4.1.1); they are descriptions of the data that are not as complete as (4.1.1). However, they preserve the information required to extract the frequencies $\{\omega_k\}$ which, as already stated, are the parameters of major interest. Hence, in some sense, these two models are more appropriate for frequency estimation; they do not include some of the *nuisance parameters* that appear in (4.1.1).

### 4.2.2 ARMA Model

It can readily be verified that

$$(1 - e^{i\omega_k} z^{-1}) x_k(t) \equiv 0 \tag{4.2.1}$$

where $z^{-1}$ denotes the unit delay (or shift) operator introduced in Chapter 1. Hence, $(1 - e^{i\omega_k} z^{-1})$ is an *annihilating filter* for the $k$th component in $x(t)$. By using this simple observation, we obtain the *homogeneous AR* equation for $\{x(t)\}$, namely,

$$\boxed{A(z)x(t) = 0} \tag{4.2.2}$$

and the *ARMA model* for the noisy data $\{y(t)\}$—that is,

$$\boxed{\begin{aligned} A(z)y(t) &= A(z)e(t) \\ A(z) &= \prod_{k=1}^{n} (1 - e^{i\omega_k} z^{-1}) \end{aligned}} \tag{4.2.3}$$

It may be a useful exercise to derive equation (4.2.2) in a different way. The PSD of $x(t)$ consists of $n$ spectral lines located at $\{\omega_k\}_{k=1}^{n}$. It should then be clear, in view of the relation (1.4.9) governing the transfer of a PSD through a linear system, that any filter that has zeroes at frequencies $\{\omega_k\}$ is an annihilating filter for $x(t)$. The polynomial $A(z)$ in (4.2.3) is the simplest kind of such annihilating filter. This polynomial bears *complete information* about $\{\omega_k\}$; hence, the problem of estimating the frequencies can be reduced to that of determining $A(z)$.

We remark that the ARMA model (4.2.3) has a very special form (for which reason it is sometimes called a "degenerate" ARMA model). All its poles and zeroes are located exactly on the unit circle. Furthermore, its AR and MA parts are identical. It might be tempting to cancel the

common poles and zeroes in (4.2.3). However, such an operation leads to the wrong conclusion that $y(t) = e(t)$ and, therefore, should be invalid. Let us explain briefly why cancelation in (4.2.3) is not allowed. The ARMA equation description of a signal $y(t)$ is *asymptotically* equivalent to the associated transfer-function description (in the sense that both give the same signal sequence, for $t \to \infty$) if and only if the poles are situated strictly inside the unit circle. If there are poles on the unit circle, then the equivalence between these two descriptions ceases. In particular, the solution of an ARMA equation with poles on the unit circle strongly depends on the initial conditions, whereas the transfer-function description does not impose a dependence on initial values.

### 4.2.3 Covariance Matrix Model

A notation that will often be used in what follows is

$$\begin{aligned} a(\omega) &\triangleq [1 \quad e^{-i\omega} \dots e^{-i(m-1)\omega}]^T \qquad (m \times 1) \\ A &= [a(\omega_1) \dots a(\omega_n)] \qquad (m \times n) \end{aligned} \tag{4.2.4}$$

In (4.2.4), $m$ is a positive integer not yet specified. The matrix $A$ is a Vandermonde matrix, which enjoys the following rank property (see Result R24 in Appendix A):

$$\text{rank}(A) = n \quad \text{if} \quad m \geq n \quad \text{and} \quad \omega_k \neq \omega_p \quad \text{for} \quad k \neq p \tag{4.2.5}$$

By making use of the previous notation, along with (4.1.1) and (4.1.4), we can write

$$\begin{aligned} \tilde{y}(t) &\triangleq \begin{bmatrix} y(t) \\ y(t-1) \\ \vdots \\ y(t-m+1) \end{bmatrix} = A\tilde{x}(t) + \tilde{e}(t) \\ \tilde{x}(t) &= [x_1(t) \dots x_n(t)]^T \\ \tilde{e}(t) &= [e(t) \dots e(t-m+1)]^T \end{aligned} \tag{4.2.6}$$

The following expression for the covariance matrix of $\tilde{y}(t)$ can be readily derived from (4.1.5) and (4.2.6):

$$R \triangleq E\left\{\tilde{y}(t)\tilde{y}^*(t)\right\} = APA^* + \sigma^2 I \quad ; \quad P = \begin{bmatrix} \alpha_1^2 & & 0 \\ & \ddots & \\ 0 & & \alpha_n^2 \end{bmatrix} \tag{4.2.7}$$

This equation constitutes the covariance matrix model of the data. As we will show later, the *eigenstructure* of $R$ contains complete information on the frequencies $\{\omega_k\}$, and this is exactly where the usefulness of (4.2.7) lies.

From equations (4.2.6) and (4.1.5), we also derive the following result for later use:

$$\begin{aligned}\Gamma &\triangleq E\left\{\begin{bmatrix} y(t-L-1) \\ \vdots \\ y(t-L-M) \end{bmatrix} [y^*(t)\ldots y^*(t-L)]\right\} \\ &= E\left\{A_M \tilde{x}(t-L-1)\tilde{x}^*(t)A_{L+1}^*\right\} \\ &= A_M P_{L+1} A_{L+1}^* \qquad (L, M \geq 1)\end{aligned} \tag{4.2.8}$$

Here $A_K$ stands for $A$ in (4.2.4) with $m = K$, and

$$P_K = \begin{bmatrix} \alpha_1^2 e^{-i\omega_1 K} & & 0 \\ & \ddots & \\ 0 & & \alpha_n^2 e^{-i\omega_n K} \end{bmatrix}$$

As we explain in detail later, the *null space* of the matrix $\Gamma$ (with $L, M \geq n$) gives complete information on the frequencies $\{\omega_k\}$.

## 4.3 NONLINEAR LEAST-SQUARES METHOD

An intuitively appealing approach to spectral line analysis, based on the *nonlinear regression model* (4.1.1), consists of finding the unknown parameters as the minimizers of the criterion

$$f(\omega, \alpha, \varphi) = \sum_{t=1}^{N} \left| y(t) - \sum_{k=1}^{n} \alpha_k e^{i(\omega_k t + \varphi_k)} \right|^2 \tag{4.3.1}$$

where $\omega$ is the vector of frequencies $\omega_k$, and similarly for $\alpha$ and $\varphi$. The sinusoidal model determined as above has the smallest "sum of squares" distance to the observed data $\{y(t)\}_{t=1}^N$. Since $f$ is a nonlinear function of its arguments $\{\omega, \varphi, \alpha\}$, the method that obtains parameter estimates by minimizing (4.3.1) is called the *nonlinear least-squares (NLS) method*. When the (white) noise $e(t)$ is Gaussian distributed, the minimization of (4.3.1) can also be interpreted as the *method of maximum likelihood* (see Appendices B and C); in that case, minimization of (4.3.1) can be shown to provide the parameter values most likely to "explain" the observed data sequence. (See [Söderström and Stoica 1989; Kay 1988; Marple 1987].)

The criterion in (4.3.1) depends on $\{\alpha_k\}$, $\{\varphi_k\}$, and $\{\omega_k\}$. However, it can be *concentrated with respect to the nuisance parameters* $\{\alpha_k, \varphi_k\}$, as explained next. By making use of the notation,

$$\beta_k = \alpha_k e^{i\varphi_k} \tag{4.3.2}$$

$$\beta = [\beta_1 \dots \beta_n]^T \tag{4.3.3}$$

$$Y = [y(1) \dots y(N)]^T \tag{4.3.4}$$

$$B = \begin{bmatrix} e^{i\omega_1} & \dots & e^{i\omega_n} \\ \vdots & & \vdots \\ e^{iN\omega_1} & \dots & e^{iN\omega_n} \end{bmatrix} \tag{4.3.5}$$

we can write the function $f$ in (4.3.1) as

$$f = (Y - B\beta)^*(Y - B\beta) \tag{4.3.6}$$

The Vandermonde matrix $B$ in (4.3.5) (which resembles the matrix $A$ defined in (4.2.4)) has full column rank equal to $n$ under the weak condition that $N \geq n$; in this case, $(B^*B)^{-1}$ exists. By using this observation, we can put (4.3.6) in the more convenient form:

$$f = [\beta - (B^*B)^{-1}B^*Y]^*[B^*B][\beta - (B^*B)^{-1}B^*Y] + Y^*Y - Y^*B(B^*B)^{-1}B^*Y \tag{4.3.7}$$

For any choice of $\omega = [\omega_1, \dots, \omega_n]^T$ in $B$ (which is such that $\omega_k \neq \omega_p$ for $k \neq p$), we can choose $\beta$ to make the first term of $f$ zero; thus, we see that the vectors $\beta$ and $\omega$ that minimize $f$ are given by

$$\boxed{\begin{aligned} \hat{\omega} &= \arg\max_{\omega}[Y^*B(B^*B)^{-1}B^*Y] \\ \hat{\beta} &= (B^*B)^{-1}B^*Y|_{\omega=\hat{\omega}} \end{aligned}} \tag{4.3.8}$$

It can be shown that, as $N$ tends to infinity, $\hat{\omega}$ obtained as in the preceding discussion converges to $\omega$ (i.e., $\hat{\omega}$ is a consistent estimate) and, in addition, the estimation errors $\{\hat{\omega}_k - \omega_k\}$ have the following (asymptotic) covariance matrix:

$$\boxed{\mathrm{Cov}(\hat{\omega}) = \frac{6\sigma^2}{N^3}\begin{bmatrix} 1/\alpha_1^2 & & 0 \\ & \ddots & \\ 0 & & 1/\alpha_n^2 \end{bmatrix}} \tag{4.3.9}$$

(See [Stoica and Nehorai 1989a; Stoica, Moses, Friedlander, and Söderström 1989].) In the case of Gaussian noise, the matrix in (4.3.9) can also be shown to equal the *Cramér–Rao limit matrix*, which gives a lower bound on the covariance matrix of any unbiased estimator of $\omega$. (See Appendix B.) Hence, under the Gaussian hypothesis, the NLS method provides the most accurate (i.e., minimum variance) frequency estimates in a fairly general class of estimators. As a matter of fact, the variance of $\{\hat{\omega}_k\}$ (as given by (4.3.9)) often takes quite small values for reasonably large sample lengths $N$ and signal-to-noise ratios $\mathrm{SNR}_k = \alpha_k^2/\sigma^2$. For example, for $N = 300$ and $\mathrm{SNR}_k = 30\mathrm{dB}$, it follows from (4.3.9) that we may expect frequency estimation errors on the order of $10^{-5}$, which is comparable with the roundoff errors in a 32-bit fixed-point processor.

The NLS method has another advantage that sets it apart from the subspace-based approaches that are discussed in the remainder of the chapter. The NLS method does not depend critically on the assumption that the noise process is white. If the noise process is not white, the NLS still gives consistent frequency estimates. In fact, the asymptotic covariance of the frequency estimates is diagonal, and $\mathrm{var}(\hat{\omega}_k) = 6/(N^3 \mathrm{SNR}_k)$, where $\mathrm{SNR}_k = \alpha_k^2/\phi_n(\omega_k)$ (here, $\phi_n(\omega)$ is the noise PSD) is the "local" signal-to-noise ratio of the sinusoid at frequency $\omega_k$; see [Stoica and Nehorai 1989b], for example. Interestingly enough, the NLS method remains the most accurate method (if the data length is large) even in those cases where the (Gaussian) noise is colored [Stoica and Nehorai 1989b]. This fact spurred a renewed interest in the NLS approach and in reliable algorithms for performing the minimization required in (4.3.1); see, for example, [Hwang and Chen 1993; Ying, Potter, and Moses 1994; Li and Stoica 1996b; Umesh and Tufts 1996] and Complement 4.9.5.

Unfortunately, the good statistical performance associated with the NLS method of frequency estimation is difficult to achieve, for the following reason. The function (4.3.8) has a *complicated multimodal shape* with a *very sharp global maximum* corresponding to $\hat{\omega}$ [Stoica, Moses, Friedlander, and Söderström 1989]. Hence, finding $\hat{\omega}$ by a search algorithm requires very accurate initialization. Initialization procedures that provide fairly accurate approximations of the maximizer of (4.3.8) have been proposed in [Kumaresan, Scharf, and Shaw 1986], [Bresler and Macovski 1986], and [Ziskind and Wax 1988]. However, there is no available method which is guaranteed to provide frequency estimates within the attraction domain of the global maximum $\hat{\omega}$ of (4.3.8). As a consequence, a search algorithm could fail to converge to $\hat{\omega}$, or might even diverge.

The kinds of difficulties indicated above, which must be faced when using the NLS method in applications, limit the practical interest in this approach to frequency estimation. There are, however, some instances when the NLS approach may be turned into a practical frequency estimation method. Consider, first, the case of a single sinusoid ($n = 1$). A straightforward calculation shows that, in such a case, the first equation in (4.3.8) can be rewritten as

$$\hat{\omega} = \arg\max_{\omega} \hat{\phi}_p(\omega) \tag{4.3.10}$$

where $\hat{\phi}_p(\omega)$ is the periodogram (see (2.2.1))

$$\hat{\phi}_p(\omega) = \frac{1}{N}\left|\sum_{t=1}^{N} y(t)e^{-i\omega t}\right|^2 \tag{4.3.11}$$

Hence, the NLS estimate of the frequency of a single sine wave buried in observation noise is given precisely by the highest peak of the unmodified periodogram.

Note that the above result is only approximately true (for $N \gg 1$) in the case of *real-valued* sinusoidal signals, a fact that lends additional support to the claim made in Chapter 1 that the analysis of the case of real-valued signals faces additional complications not encountered in the complex-valued case. Each real-valued sinusoid can be written as a sum of two complex exponentials, and the treatment of the real case with $n = 1$ is similar to that of the complex case with $n > 1$, presented next.

Next, consider the case of multiple sine waves ($n > 1$). The key condition that makes it possible to treat this case in a manner similar to the previous one is that the minimum frequency separation between the sine waves in the studied signal is larger than the periodogram's resolution limit:

$$\Delta\omega = \inf_{k \neq p} |\omega_k - \omega_p| > 2\pi/N \tag{4.3.12}$$

Since the estimation errors $\{\hat{\omega}_k - \omega_k\}$ from the NLS estimates are of order $\mathcal{O}(1/N^{3/2})$ (because $\text{cov}(\hat{\omega}) = \mathcal{O}(1/N^3)$; see (4.3.9)), equation (4.3.12) implies a similar inequality for the NLS frequency estimates $\{\hat{\omega}_k\}$: $\Delta\hat{\omega} > 2\pi/N$. It should then be possible to *resolve all n sine waves* in the noisy signal and to obtain *reasonable approximations* $\{\tilde{\omega}_k\}$ to $\{\hat{\omega}_k\}$ by evaluating the function in (4.3.8) at the points of a grid corresponding to the sampling of each frequency variable, as in the FFT:

$$\omega_k = \frac{2\pi}{N} j \qquad j = 0, \ldots, N-1 \qquad (k = 1, \ldots, n) \tag{4.3.13}$$

Of course, a direct application of such a grid method for the approximate maximization of (4.3.8) would be computationally burdensome for large values of $n$ or $N$. However, it can be greatly simplified, as described next.

The $p, k$ element of the matrix $B^*B$ occurring in (4.3.8), when evaluated *at the points of the grid* (4.3.13), is given by

$$[B^*B]_{p,k} = N \qquad \text{for } p = k \tag{4.3.14}$$

and

$$\begin{aligned} [B^*B]_{p,k} &= \sum_{t=1}^{N} e^{i(\omega_k - \omega_p)t} = e^{i(\omega_k - \omega_p)} \frac{e^{iN(\omega_k - \omega_p)} - 1}{e^{i(\omega_k - \omega_p)} - 1} \\ &= 0 \qquad \text{for } p \neq k \end{aligned} \tag{4.3.15}$$

which implies that the function to be minimized in (4.3.8) has, in such a case, the following form:

$$\sum_{k=1}^{n} \frac{1}{N} \left| \sum_{t=1}^{N} y(t) e^{-i\omega_k t} \right|^2 \tag{4.3.16}$$

The previous additive decomposition in $n$ functions of $\omega_1, \ldots, \omega_n$ (respectively) leads to the conclusion that $\{\tilde{\omega}_k\}$ (which, by definition, maximize (4.3.16) at the points of the grid (4.3.13)) are given by the $n$ largest peaks of the periodogram. To show this, let us write the function in (4.3.16) as

$$g(\omega_1, \ldots, \omega_n) = \sum_{k=1}^{n} \hat{\phi}_p(\omega_k)$$

where $\hat{\phi}_p(\omega)$ is once again the periodogram. Observe that

$$\frac{\partial g(\omega_1, \ldots, \omega_n)}{\partial \omega_k} = \hat{\phi}_p'(\omega_k)$$

and

$$\frac{\partial^2 g(\omega_1, \ldots, \omega_n)}{\partial \omega_k \partial \omega_j} = \hat{\phi}_p''(\omega_k)\delta_{k,j}$$

Hence, the maximum points of (4.3.16) satisfy

$$\hat{\phi}_p'(\omega_k) = 0 \quad \text{and} \quad \hat{\phi}_p''(\omega_k) < 0 \quad \text{for } k = 1, \ldots, n$$

It follows that the set of maximizers of (4.3.16) is given by all possible combinations of $n$ elements from the periodogram's peak locations. Now, recall the assumption made that $\{\omega_k\}$, and hence their estimates $\{\hat{\omega}_k\}$, are *distinct*. Under this assumption the highest maximum of $g(\omega_1, \ldots, \omega_n)$ is given by the locations of the $n$ largest peaks of $\hat{\phi}_p(\omega)$, which is the desired result.

These findings are summarized as follows:

> Under the condition (4.3.12), the unmodified periodogram resolves all the $n$ sine waves present in the noisy signal. Furthermore, the locations $\{\tilde{\omega}_k\}$ of the $n$ largest peaks in the periodogram provide $\mathcal{O}(1/N)$ approximations to the NLS frequency estimates $\{\hat{\omega}_k\}$. In the case of $n = 1$, we have $\tilde{\omega}_1 = \hat{\omega}_1$ exactly. (4.3.17)

The fact that the differences $\{\tilde{\omega}_k - \hat{\omega}_k\}$ are $\mathcal{O}(1/N)$ means, of course, that the computationally convenient estimates $\{\tilde{\omega}_k\}$ (derived from the periodogram) will generally have an inflated variance compared to $\{\hat{\omega}_k\}$. However, $\{\tilde{\omega}_k\}$ can at least be used as initial values in a numerical implementation of the NLS estimator. In any case, this discussion indicates that, under (4.3.12), the periodogram performs quite well as a frequency estimator (which actually is the task for which it was introduced by Schuster more than a century ago!).

In the next sections, we present several "high-resolution" methods for frequency estimation, which exploit the *covariance matrix models*. More precisely, all of these methods derive frequency estimates by exploiting the properties of the eigendecomposition of data covariance matrices and, in particular, the subspaces associated with those matrices. For this reason, these methods are

sometimes referred to by the generic name of *subspace methods*. However, in spite of their common subspace theme, the methods are quite different, and we will treat them in separate sections. The main features of these methods can be summarized as follows: (i) Their statistical performance is close to the ultimate performance corresponding to the NLS method (and given by the Cramér–Rao lower bound, (4.3.9)); (ii) unlike the NLS method, these methods are not based on multidimensional search procedures; and (iii) they do not depend on a "resolution condition," such as (4.3.12); thus, they could generally have a resolution threshold lower than that of the periodogram. The chief drawback of these methods, as compared with the NLS method, is that their performance significantly degrades if the measurement noise in (4.1.1) cannot be assumed to be white.

## 4.4 HIGH-ORDER YULE–WALKER METHOD

The high-order Yule–Walker (HOYW) method of frequency estimation can be derived from the *ARMA model* of the sinusoidal data, (4.2.3), much as can its counterpart in the rational PSD case. (See Section 3.7 and [Cadzow 1982; Stoica, Söderström, and Ti 1989; Stoica, Moses, Söderström, and Li 1991].) Actually, the HOYW method is based on an ARMA model of an order $L$ *higher* than the minimal order $n$, for a reason that will be explained shortly.

If the polynomial $A(z)$ in (4.2.3) is multiplied by any other polynomial $\bar{A}(z)$, say of degree equal to $L-n$, then we obtain a higher order ARMA representation of our sinusoidal data, given by

$$y(t)+b_1y(t-1)+\ldots+b_Ly(t-L)=e(t)+b_1e(t-1)+\ldots+b_Le(t-L) \tag{4.4.1}$$

or

$$B(z)y(t)=B(z)e(t)$$

where

$$B(z)=1+\sum_{k=1}^{L}b_kz^{-k}\triangleq A(z)\bar{A}(z) \tag{4.4.2}$$

Equation (4.4.1) can be rewritten in the following more condensed form (with obvious notation):

$$[y(t)\quad y(t-1)\ldots y(t-L)]\begin{bmatrix}1\\ b\end{bmatrix}=e(t)+\ldots+b_Le(t-L) \tag{4.4.3}$$

Premultiplying (4.4.3) by $[y^*(t-L-1)\ldots y^*(t-L-M)]^T$ and taking the expectation leads to

$$\Gamma^c\begin{bmatrix}1\\ b\end{bmatrix}=0 \tag{4.4.4}$$

where the matrix $\Gamma$ is defined in (4.2.8) and $M$ is a positive integer that is yet to be specified. In order to obtain (4.4.4) as indicated previously, we made use of the fact that $E\{y^*(t-k)e(t)\} = 0$ for $k > 0$.

The similarity of (4.4.4) to the Yule–Walker system of equations encountered in Chapter 3 (see equation (3.7.1)) is more readily seen if (4.4.4) is rewritten in the following more detailed form:

$$\begin{bmatrix} r(L) & \dots & r(1) \\ \vdots & & \vdots \\ r(L+M-1) & \dots & r(M) \end{bmatrix} b = - \begin{bmatrix} r(L+1) \\ \vdots \\ r(L+M) \end{bmatrix} \tag{4.4.5}$$

Owing to this analogy, the set of equations (4.4.5) associated with the noisy sinusoidal signal $\{y(t)\}$ is said to form a HOYW system.

The HOYW matrix equation (4.4.4) can also be obtained directly from (4.2.8). For any $L \geq n$ and any polynomial $\bar{A}(z)$ (used in the defining equation, (4.4.2), for $b$), the elements of the vector

$$A_{L+1}^T \begin{bmatrix} 1 \\ b \end{bmatrix} \tag{4.4.6}$$

are equal to zero. Indeed, the $k$th row of (4.4.6) is

$$\begin{aligned} [1 \quad e^{-i\omega_k} \dots e^{-iL\omega_k}] \begin{bmatrix} 1 \\ b \end{bmatrix} &= 1 + \sum_{p=1}^{L} b_p e^{-i\omega_k p} \\ &= A(\omega_k)\bar{A}(\omega_k) = 0, \quad k = 1, \dots, n \end{aligned} \tag{4.4.7}$$

(since $A(\omega_k) = 0$, *cf.* (4.2.3)). It follows from (4.2.8) and (4.4.7) that the vector $\begin{bmatrix} 1 \\ b \end{bmatrix}$ lies in the *null space* of $\Gamma^c$ (see Definition D2 in Appendix A),

$$\Gamma^c \begin{bmatrix} 1 \\ b \end{bmatrix} = 0$$

which is the desired result, (4.4.4).

The HOYW system of equations just derived can be used for frequency estimation in the following way: By replacing the unavailable theoretical covariances $\{r(k)\}$ in (4.4.5) by the sample covariances $\{\hat{r}(k)\}$, we obtain

$$\boxed{\begin{bmatrix} \hat{r}(L) & \dots & \hat{r}(1) \\ \vdots & & \vdots \\ \hat{r}(L+M-1) & \dots & \hat{r}(M) \end{bmatrix} \hat{b} \simeq - \begin{bmatrix} \hat{r}(L+1) \\ \vdots \\ \hat{r}(L+M) \end{bmatrix}} \tag{4.4.8}$$

Owing to the estimation errors in $\{\hat{r}(k)\}$, the matrix equation (4.4.8) cannot hold exactly in the general case, for any vector $\hat{b}$, as is indicated by the use of the "approximate equality" symbol $\simeq$.

We can solve (4.4.8) for $\hat{b}$ in a least-squares sense that is detailed in what follows, then form the polynomial

$$1 + \sum_{k=1}^{L} \hat{b}_k z^{-k} \tag{4.4.9}$$

and, finally (in view of (4.2.3) and (4.4.2)), obtain frequency estimates $\{\hat{\omega}_k\}$ as the angular positions of the $n$ roots of (4.4.9) that are located nearest the unit circle.

It can be expected that increasing the values of $M$ and $L$ results in improved frequency estimates. Indeed, by increasing $M$ and $L$ we use higher lag covariances in (4.4.8), which could bear "additional information" on the data at hand. Increasing $M$ and $L$ also has a second, more subtle, effect that is explained next.

Let $\Omega$ denote the $M \times L$ covariance matrix in (4.4.5) and, similarly, let $\hat{\Omega}$ denote the sample covariance matrix in (4.4.8). It can be seen from (4.2.8) that

$$\text{rank}(\Omega) = n \qquad \text{for} \quad M, L \geq n \tag{4.4.10}$$

On the other hand, the matrix $\hat{\Omega}$ has full rank (almost surely)

$$\text{rank}(\hat{\Omega}) = \min(M, L) \tag{4.4.11}$$

owing to the random errors in $\{\hat{r}(k)\}$. However, for reasonably large values of $N$, the matrix $\hat{\Omega}$ is close to the rank-$n$ matrix $\Omega$, since the sample covariances $\{\hat{r}(k)\}$ converge to $\{r(k)\}$ as $N$ increases (as is shown in Complement 4.9.1). Hence, we may expect the linear system (4.4.8) to be *ill conditioned from a numerical standpoint*. (See the discussion in Section A.8.1 in Appendix A.) In fact, there is compelling empirical evidence that any LS procedure that estimates $\hat{b}$ directly from (4.4.8) has very poor accuracy. In order to overcome the previously described difficulty, we can make use of the *a priori rank information* (4.4.10). However, some preparations are required before we shall be able to do so. Let

$$\hat{\Omega} = U \Sigma V^* \triangleq [\underbrace{U_1}_{n} \; \underbrace{U_2}_{M-n}] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} V_1^* \\ V_2^* \end{bmatrix} \begin{matrix} \} n \\ \} L-n \end{matrix} \tag{4.4.12}$$

denote the singular value decomposition (SVD) of the matrix $\hat{\Omega}$. (See Section A.4 in Appendix A; also [Söderström and Stoica 1989; Van Huffel and Vandewalle 1991] for general discussions on the SVD.) In (4.4.12), $U$ is an $M \times M$ unitary matrix, $V$ is an $L \times L$ unitary matrix, and $\Sigma$ is an $M \times L$ diagonal matrix. $\hat{\Omega}$ is close to a rank-$n$ matrix, so $\Sigma_2$ in (4.4.12) should be close to zero, which implies that

$$\hat{\Omega}_n \triangleq U_1 \Sigma_1 V_1^* \tag{4.4.13}$$

should be a good approximation for $\hat{\Omega}$. In fact, it can be proven that this $\hat{\Omega}_n$ is the *best* (in the Frobenius-norm sense) *rank-n approximation* of $\hat{\Omega}$ (Result R18 in Appendix A). Hence, in accordance with the rank information (4.4.10), we can use $\hat{\Omega}_n$ in (4.4.8) in lieu of $\hat{\Omega}$. The so-obtained *rank-truncated HOYW system of equations*

$$\hat{\Omega}_n \hat{b} \simeq - \begin{bmatrix} \hat{r}(L+1) \\ \vdots \\ \hat{r}(L+M) \end{bmatrix} \tag{4.4.14}$$

can be solved in a numerically sound way by using a simple LS procedure. It is readily verified that

$$\hat{\Omega}_n^{\dagger} = V_1 \Sigma_1^{-1} U_1^* \tag{4.4.15}$$

is the pseudoinverse of $\hat{\Omega}_n$. (See Definition D15 and Result R32.) Hence, the LS solution to (4.4.14) is given by

$$\boxed{\hat{b} = -V_1 \Sigma_1^{-1} U_1^* \begin{bmatrix} \hat{r}(L+1) \\ \vdots \\ \hat{r}(L+M) \end{bmatrix}} \tag{4.4.16}$$

The additional bonus for using $\hat{\Omega}_n$ instead of $\hat{\Omega}$ in (4.4.8) is an improvement in the statistical accuracy of the frequency estimates obtained from (4.4.16). This improved accuracy is explained by the fact that $\hat{\Omega}_n$ should be closer to $\Omega$ than $\hat{\Omega}$ is; the improved covariance matrix estimate $\hat{\Omega}_n$ obtained by exploitation of the rank information (4.4.10), when used in the HOYW system of equations, should lead to refined frequency estimates.

We remark that a *total least-squares* (TLS) solution for $\hat{b}$ can also be obtained from (4.4.8). (See Definition D17 and Result R33 in Appendix A.) A TLS solution makes sense, because we have errors in both $\hat{\Omega}$ and the right-hand-side vector in equation (4.4.8). In fact the TLS-based estimate of $b$ is often slightly better than the estimate discussed above, which is obtained as the LS solution to the *rank-truncated* system of linear equations in (4.4.14).

We next return to the selection of $L$ and $M$. As $M$ and $L$ increase, the information brought into the estimation problem under study by the rank condition (4.4.10) is more and more important, and hence the corresponding increase of accuracy is more and more pronounced. (For instance, the information that a $10 \times 10$ noisy matrix has rank one in the noise-free case leads to more relations between the matrix elements, and hence to more "noise cleaning," than if the matrix were $2 \times 2$.) In fact, for $M = n$ or $L = n$, the rank condition is inactive; $\hat{\Omega}_n = \hat{\Omega}$ in such a case. The previous discussion gives another explanation of why the accuracy of the frequency estimates obtained from (4.4.16) may be expected to increase with increasing $M$ and $L$.

The next box summarizes the *HOYW frequency estimation method*. It should be noted that the operation in Step 3 of the HOYW method is implicitly based on the assumption that the estimated "*signal roots*" (i.e., the roots of $A(z)$ in (4.4.2)) are always closer to the unit circle than the

estimated "*noise roots*" (i.e., the roots of $\bar{A}(z)$ in (4.4.2)). It can be shown that as $N \to \infty$, all roots of $\bar{A}(z)$ are strictly inside the unit circle; see, for example, Complement 6.5.1 and [KUMARESAN AND TUFTS 1983]. This property cannot be guaranteed in finite samples, but there is empirical evidence that it holds quite often. In those rare cases where it fails to hold, the HOYW method produces *spurious* (or *false*) *frequency estimates*. The risk of producing spurious estimates is the price paid for the improved accuracy obtained by increasing $L$. (Note that, for $L = n$, there is no "noise root," and hence no spurious estimate can occur in such a case.) The risk of false frequency estimation is a problem that is common to all methods that estimate the frequencies from the roots of a polynomial of degree larger than $n$, such as the MUSIC and Min–Norm methods, to be discussed in the next two sections.

**The HOYW Frequency Estimation Method**

**Step 1.** Compute the sample covariances $\{\hat{r}(k)\}_{k=1}^{L+M}$. We may set $L \simeq M$ and select the values of these integers so that $L + M$ is a fraction of the sample length (such as $N/3$). Note that, if $L + M$ is set to a value too close to $N$, then the higher lag covariances required in (4.4.8) cannot be estimated in a reliable way.

**Step 2.** Compute the SVD of $\hat{\Omega}$, (4.4.12), and compute $\hat{b}$ by using (4.4.16).

**Step 3.** Isolate the $n$ roots of the polynomial (4.4.9) that are closest to the unit circle and obtain the frequency estimates as the angular positions of these roots.

## 4.5 PISARENKO AND MUSIC METHODS

The *MUltiple SIgnal Classification* (or *MUltiple SIgnal Characterization*) (MUSIC) method [SCHMIDT 1979; BIENVENU 1979] and Pisarenko's method [PISARENKO 1973] (a special case of MUSIC, as is explained next) are derived from the covariance model (4.2.7) with $m > n$. Let $\lambda_1 \geq \lambda_2 \geq \ldots \geq \lambda_m$ denote the *eigenvalues* of $R$ in (4.2.7), arranged in nonincreasing order, and let $\{s_1, \ldots, s_n\}$ be the *orthonormal eigenvectors* associated with $\{\lambda_1, \ldots, \lambda_n\}$, and $\{g_1, \ldots, g_{m-n}\}$ a set of *orthonormal eigenvectors* corresponding to $\{\lambda_{n+1}, \ldots, \lambda_m\}$. (See Appendix A.) Since

$$\text{rank}(APA^*) = n \tag{4.5.1}$$

it follows that $APA^*$ has $n$ strictly positive eigenvalues, the remaining $(m - n)$ eigenvalues all being equal to zero. Combining this observation with the fact that (see Result R5 in Appendix A)

$$\lambda_k = \tilde{\lambda}_k + \sigma^2 \qquad (k = 1, \ldots, m) \tag{4.5.2}$$

where $\{\tilde{\lambda}_k\}_{k=1}^m$ are the eigenvalues of $APA^*$ arranged in nonincreasing order, leads to the following result:

$$\begin{cases} \lambda_k > \sigma^2 & \text{for } k = 1, \ldots, n \\ \lambda_k = \sigma^2 & \text{for } k = n+1, \ldots, m \end{cases} \tag{4.5.3}$$

The set of eigenvalues of $R$ can hence be split into two subsets. Next, we show that the eigenvectors associated with each of these subsets, as introduced previously, possess some interesting properties that can be used for frequency estimation.

Let

$$S = [s_1, \ldots, s_n] \quad (m \times n), \qquad G = [g_1, \ldots, g_{m-n}] \quad (m \times (m-n)) \tag{4.5.4}$$

From (4.2.7) and (4.5.3), we get at once

$$RG = G \begin{bmatrix} \lambda_{n+1} & & 0 \\ & \ddots & \\ 0 & & \lambda_m \end{bmatrix} = \sigma^2 G = APA^*G + \sigma^2 G \tag{4.5.5}$$

The first equality in (4.5.5) follows from the definition of $G$ and $\{\lambda_k\}_{k=n+1}^m$, the second equality follows from (4.5.3), and the third from (4.2.7). The last equality in equation (4.5.5) implies that $APA^*G = 0$, or (as the matrix $AP$ has full column rank)

$$\boxed{A^*G = 0} \tag{4.5.6}$$

In other words, the columns $\{g_k\}$ of $G$ belong to the *null space* of $A^*$, a fact which is denoted by $g_k \in \mathcal{N}(A^*)$. Since $\text{rank}(A) = n$, the dimension of $\mathcal{N}(A^*)$ is equal to $m - n$, which is also the dimension of the *range space* of $G$, $\mathcal{R}(G)$. It follows from this observation and (4.5.6) that

$$\boxed{\mathcal{R}(G) = \mathcal{N}(A^*)} \tag{4.5.7}$$

In words, (4.5.7) says that the vectors $\{g_k\}$ span both $\mathcal{R}(G)$ and $\mathcal{N}(A^*)$. Now, by definition,

$$S^*G = 0 \tag{4.5.8}$$

so we also have $\mathcal{R}(G) = \mathcal{N}(S^*)$; hence, $\mathcal{N}(S^*) = \mathcal{N}(A^*)$. Since $\mathcal{R}(S)$ and $\mathcal{R}(A)$ are the orthogonal complements to $\mathcal{N}(S^*)$ and $\mathcal{N}(A^*)$, it follows that

$$\boxed{\mathcal{R}(S) = \mathcal{R}(A)} \tag{4.5.9}$$

We can also derive the equality (4.5.9) directly from (4.2.7). Set

$$\mathring{\Lambda} = \begin{bmatrix} \lambda_1 - \sigma^2 & & 0 \\ & \ddots & \\ 0 & & \lambda_n - \sigma^2 \end{bmatrix} \tag{4.5.10}$$

From

$$RS = S \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} = APA^*S + \sigma^2 S \tag{4.5.11}$$

we obtain

$$S = A\left(PA^*S\mathring{\Lambda}^{-1}\right) \tag{4.5.12}$$

which shows that $\mathcal{R}(S) \subset \mathcal{R}(A)$. However, $\mathcal{R}(S)$ and $\mathcal{R}(A)$ have the same dimension (equal to $n$); hence, (4.5.9) follows. Owing to (4.5.9) and (4.5.8), the subspaces $\mathcal{R}(S)$ and $\mathcal{R}(G)$ are sometimes called the *signal subspace* and *noise subspace*, respectively.

The following key result is obtained from (4.5.6):

> The true frequency values $\{\omega_k\}_{k=1}^n$ are the only solutions of the equation $a^*(\omega)GG^*a(\omega) = 0$ for any $m > n$. (4.5.13)

The fact that $\{\omega_k\}$ satisfy this equation follows from (4.5.6). It only remains to prove that $\{\omega_k\}_{k=1}^n$ are the only solutions to (4.5.13). Let $\tilde{\omega}$ denote another possible solution, with $\tilde{\omega} \neq \omega_k$ $(k = 1, \ldots, n)$. In (4.5.13), $GG^*$ is the *orthogonal projector* onto $\mathcal{R}(G)$. (See Section A.4.) Hence, (4.5.13) implies that $a(\tilde{\omega})$ is orthogonal to $\mathcal{R}(G)$, which means that $a(\tilde{\omega}) \in \mathcal{N}(G^*)$. However, the Vandermonde vector $a(\tilde{\omega})$ is linearly independent of $\{a(\omega_k)\}_{k=1}^n$. Since $n + 1$ linearly independent vectors cannot belong to an $n$-dimensional subspace, which is $\mathcal{N}(G^*)$ in the present case, we conclude that no other solution $\tilde{\omega}$ to (4.5.13) can exist; with this, the proof is finished.

The *MUSIC algorithm* uses the previous result to derive frequency estimates in the following steps:

**Step 1.** Compute the sample covariance matrix

$$\hat{R} = \frac{1}{N}\sum_{t=m}^{N} \tilde{y}(t)\tilde{y}^*(t) \tag{4.5.14}$$

and its eigendecomposition. Let $\hat{S}$ and $\hat{G}$ denote the matrices defined similarly to $S$ and $G$, but made from the eigenvectors $\{\hat{s}_1, \ldots, \hat{s}_n\}$ and $\{\hat{g}_1, \ldots, \hat{g}_{m-n}\}$ of $\hat{R}$.

**Step 2a.** *(Spectral MUSIC)* [Schmidt 1979; Bienvenu 1979]. Determine frequency estimates as the locations of the $n$ highest peaks of the function

$$\frac{1}{a^*(\omega)\hat{G}\hat{G}^*a(\omega)}, \qquad \omega \in [-\pi, \pi] \tag{4.5.15}$$

(Sometimes (4.5.15) is called a "*pseudospectrum*," since it indicates the presence of sinusoidal components in the studied signal, but it is not a true PSD. This fact may explain the attribute "spectral" attached to this variant of MUSIC.)

**OR**

**Step 2b.** *(Root MUSIC)* [BARABELL 1983]. Determine frequency estimates as the angular positions of the $n$ (pairs of reciprocal) roots of the equation

$$a^T(z^{-1})\hat{G}\hat{G}^* a(z) = 0 \tag{4.5.16}$$

which are located nearest the unit circle. In (4.5.16), $a(z)$ stands for the vector $a(\omega)$, (4.2.4), with $e^{i\omega}$ replaced by $z$, so

$$a(z) = [1, z^{-1}, \ldots, z^{-(m-1)}]^T$$

For $m = n + 1$ (which is the *minimum* possible value), the MUSIC algorithm reduces to the Pisarenko method, which was the earliest proposal for an eigenanalysis-based (or subspace-based) method of frequency estimation [PISARENKO 1973]:

| The Pisarenko method is MUSIC with $m = n + 1$. | (4.5.17) |
|---|---|

In the Pisarenko method, the estimated frequencies are computed from (4.5.16). For $m = n + 1$, this $2(m - 1)$-degree equation can be reduced to the following equation of degree $m - 1 = n$:

$$a^T(z^{-1})\hat{g}_1 = 0 \tag{4.5.18}$$

The Pisarenko frequency estimates are obtained as the angular positions of the roots of (4.5.18). The Pisarenko method is the simplest version of MUSIC from a computational standpoint. In addition, unlike MUSIC with $m > n + 1$, the Pisarenko procedure does not have the problem of separating the "signal roots" from the "noise roots." (See the discussion on this point at the end of Section 4.4.) However, it can be shown that *the accuracy of the MUSIC frequency estimates increases significantly with increasing* $m$. Hence, the price paid for the computational simplicity of the Pisarenko method could be a relatively poor statistical accuracy.

Regarding the *selection of a value for* $m$, this parameter may be chosen as large as possible, but not too close to $N$, in order to still allow a reliable estimation of the covariance matrix (for example, as in (4.5.14)). In some applications, the largest possible value that may be selected for $m$ may also be limited by computational complexity considerations.

Whenever the *tradeoff between statistical accuracy and computational complexity* is an important issue, the following simple ideas can be valuable.

The *finite-sample statistical accuracy* of MUSIC frequency estimates may be improved by modifying the covariance estimator (4.5.14). For instance, $\hat{R}$ is not Toeplitz, whereas the true covariance matrix $R$ is. We may correct this situation by replacing the elements in each diagonal of $\hat{R}$ with their average. The so-corrected sample covariance matrix can be shown to be the best (in the Frobenius-norm sense) Toeplitz approximation of $\hat{R}$. Another modification of $\hat{R}$, with the same purpose of improving the finite-sample statistical accuracy, is described in Section 4.8.

The *computational complexity* of MUSIC, for a given $m$, can be reduced in various ways. Quite often, $m$ is such that $m - n > n$. Then, the computational burdens associated with both Spectral and Root MUSIC can be reduced by using $I - \hat{S}\hat{S}^*$ in (4.5.15) or (4.5.16) in lieu of $\hat{G}\hat{G}^*$. (Note that $\hat{S}\hat{S}^* + \hat{G}\hat{G}^* = I$ by the very definition of the eigenvector matrices.) The computational burden of Root MUSIC can be further reduced as explained next. The polynomial in (4.5.16) is a self-reciprocal (or symmetric) one: its roots appear in reciprocal pairs $(\rho e^{i\varphi}, \frac{1}{\rho}e^{i\varphi})$. On the unit circle $z = e^{i\omega}$, (4.5.16) is nonnegative and, hence, may be interpreted as a PSD. These properties mentioned imply that (4.5.16) can be factored as

$$a^T(z^{-1})\hat{G}\hat{G}^*a(z) = \alpha(z)\alpha^*(1/z^*) \tag{4.5.19}$$

where $\alpha(z)$ is a polynomial of degree $(m-1)$ with all its zeroes located within or on the unit circle. We can then find the frequency estimates from the $n$ roots of $\alpha(z)$ that are closest to the unit circle. Since there are efficient numerical procedures for spectral factorization, determining $\alpha(z)$, as in (4.5.19), and then computing its zeroes is usually computationally more efficient than finding the (reciprocal) roots of the $2(m-1)$-degree polynomial (4.5.16).

Finally, we address the issue of *spurious frequency estimates*. As implied by the result (4.5.13), for $N \to \infty$ there is no risk of obtaining false frequency estimates. However, in finite samples, such a risk always exists. Usually, this risk is quite small but it could become a real problem if $m$ takes on large values. The key result on which the standard MUSIC algorithm, (4.5.15), is based can be used to derive a *modified MUSIC* that does not suffer from the spurious-estimate problem. In what follows, we explain only the basic ideas leading to the modified MUSIC method, without going into details of its implementation. (For such details, the interested reader may consult [Stoica and Sharman 1990].) Let $\{c_k\}_{k=1}^n$ denote the coefficients of the polynomial $A(z)$ defined in (4.2.3); that is,

$$A(z) = 1 + c_1 z^{-1} + \ldots + c_n z^{-n} = \prod_{k=1}^{n}(1 - e^{i\omega_k} z^{-1}) \tag{4.5.20}$$

Introduce the following matrix made from $\{c_k\}$:

$$C^* = \begin{bmatrix} 1 & c_1 & \ldots & c_n & & 0 \\ & \ddots & \ddots & & \ddots & \\ 0 & & 1 & c_1 & \ldots & c_n \end{bmatrix}, \qquad (m-n) \times m \tag{4.5.21}$$

It is readily verified that

$$C^*A = 0, \qquad (m-n) \times n \tag{4.5.22}$$

where $A$ is defined in (4.2.4). Combining (4.5.9) and (4.5.22) gives

$$C^*S = 0, \qquad (m-n) \times n \tag{4.5.23}$$

which is the key property here. The matrix equation (4.5.23) can be rewritten in the form

$$\phi c = \mu \tag{4.5.24}$$

where the $(m-n)n \times n$ matrix $\phi$ and the $(m-n)n \times 1$ vector $\mu$ are entirely determined from the elements of $S$, and where

$$c = [c_1 \ldots c_n]^T \tag{4.5.25}$$

By replacing the elements of $S$ in $\phi$ and $\mu$ by the corresponding entries of $\hat{S}$, we obtain the sample version of (4.5.24),

$$\hat{\phi}\hat{c} \simeq \hat{\mu} \tag{4.5.26}$$

from which an estimate $\hat{c}$ of $c$ may be obtained by an LS or TLS algorithm; see Section A.8 for details. The frequency estimates can then be derived from the roots of the estimated polynomial (4.5.20) corresponding to $\hat{c}$. Since this polynomial has a (minimal) degree equal to $n$, there is *no risk* for false frequency estimation.

## 4.6 MIN–NORM METHOD

MUSIC uses $(m-n)$ linearly independent vectors in $\mathcal{R}(\hat{G})$ to obtain the frequency estimates. Since any vector in $\mathcal{R}(\hat{G})$ is (asymptotically) orthogonal to $\{a(\omega_k)\}_{k=1}^n$ (*cf.* (4.5.7)), we may think of using *only one* such vector for frequency estimation. By doing so, we might achieve some computational saving, hopefully without sacrificing too much accuracy.

The Min–Norm method proceeds to estimate the frequencies along these lines [KUMARESAN AND TUFTS 1983]. Let

$$\begin{bmatrix} 1 \\ \hat{g} \end{bmatrix} = \text{the vector in } \mathcal{R}(\hat{G}), \text{ with first element equal to one, that has minimum Euclidean norm.} \tag{4.6.1}$$

Then, the *Min–Norm frequency estimates* are determined as

(*Spectral Min–Norm*). The locations of the $n$ highest peaks in the pseudospectrum

$$\frac{1}{\left| a^*(\omega) \begin{bmatrix} 1 \\ \hat{g} \end{bmatrix} \right|^2} \tag{4.6.2}$$

or, alternatively,

(*Root Min–Norm*). The angular positions of the $n$ roots of the polynomial

$$a^T(z^{-1})\begin{bmatrix} 1 \\ \hat{g} \end{bmatrix} \tag{4.6.3}$$

that are located nearest to the unit circle.

It remains to find the vector in (4.6.1) and, in particular, to show that its first element can always be normalized to 1. We will later comment on the reason behind the specific selection (4.6.1) of a vector in $\mathcal{R}(\hat{G})$. In the following, the Euclidean norm of a vector is denoted by $\|\cdot\|$.

Partition the matrix $\hat{S}$ as

$$\hat{S} = \begin{bmatrix} \alpha^* \\ \bar{S} \end{bmatrix} \begin{matrix} \} 1 \\ \} m-1 \end{matrix} \tag{4.6.4}$$

As $\begin{bmatrix} 1 \\ \hat{g} \end{bmatrix} \in \mathcal{R}(\hat{G})$, it must satisfy the equation

$$\hat{S}^* \begin{bmatrix} 1 \\ \hat{g} \end{bmatrix} = 0 \tag{4.6.5}$$

which, by using (4.6.4), can be rewritten as

$$\bar{S}^* \hat{g} = -\alpha \tag{4.6.6}$$

The minimum–norm solution to (4.6.6) is given (see Result R31 in Appendix A) by

$$\hat{g} = -\bar{S}(\bar{S}^*\bar{S})^{-1}\alpha \tag{4.6.7}$$

assuming that the inverse exists. Note that

$$I = \hat{S}^*\hat{S} = \alpha\alpha^* + \bar{S}^*\bar{S} \tag{4.6.8}$$

and also that one eigenvalue of $I - \alpha\alpha^*$ is equal to $1 - \|\alpha\|^2$ and the remaining $(n-1)$ eigenvalues of $I - \alpha\alpha^*$ are equal to 1; it follows that the inverse in (4.6.7) exists if and only if

$$\|\alpha\|^2 \neq 1 \tag{4.6.9}$$

If this condition is not satisfied, there will be no vector of the form of (4.6.1) in $\mathcal{R}(\hat{G})$. We postpone the study of (4.6.9) until we obtain a final-form expression for $\hat{g}$.

Under the condition (4.6.9), a simple calculation shows that

$$(\bar{S}^*\bar{S})^{-1}\alpha = (I - \alpha\alpha^*)^{-1}\alpha = \alpha/(1 - \|\alpha\|^2) \tag{4.6.10}$$

Inserting (4.6.10) in (4.6.7) gives

$$\boxed{\hat{g} = -\bar{S}\alpha/(1 - \|\alpha\|^2)} \tag{4.6.11}$$

which expresses $\hat{g}$ as a function of the elements of $\hat{S}$.

We can also obtain $\hat{g}$ as a function of the entries in $\hat{G}$. To do so, partition $\hat{G}$ as

$$\hat{G} = \begin{bmatrix} \beta^* \\ \bar{G} \end{bmatrix} \tag{4.6.12}$$

Since $\hat{S}\hat{S}^* = I - \hat{G}\hat{G}^*$ by the definition of the matrices $\hat{S}$ and $\hat{G}$, it follows that

$$\begin{bmatrix} \|\alpha\|^2 & (\bar{S}\alpha)^* \\ \bar{S}\alpha & \bar{S}\bar{S}^* \end{bmatrix} = \begin{bmatrix} 1 - \|\beta\|^2 & -(\bar{G}\beta)^* \\ -\bar{G}\beta & I - \bar{G}\bar{G}^* \end{bmatrix} \tag{4.6.13}$$

Comparing the blocks in (4.6.13) makes it possible to express $\|\alpha\|^2$ and $\bar{S}\alpha$ as functions of $\bar{G}$ and $\beta$, which leads to the following equivalent expression for $\hat{g}$:

$$\boxed{\hat{g} = \bar{G}\beta/\|\beta\|^2} \tag{4.6.14}$$

*If* $m - n > n$, then it is computationally more advantageous to obtain $\hat{g}$ from (4.6.11); *otherwise*, (4.6.14) should be used.

Next, we return to the condition (4.6.9), which is implicitly assumed to hold in the previous derivations. As already mentioned, this condition is equivalent to $\text{rank}(\bar{S}^*\bar{S}) = n$ which, in turn, holds if and only if

$$\text{rank}(\bar{S}) = n \tag{4.6.15}$$

Now, it follows from (4.5.9) that any block of $S$ made from more than $n$ consecutive rows should have rank equal to $n$. Hence, (4.6.15) must hold at least for $N$ sufficiently large. With this observation, the derivation of the Min–Norm frequency estimator is complete.

The statistical accuracy of the Min–Norm method is similar to that corresponding to MUSIC. Hence, Min–Norm achieves MUSIC's performance at a reduced computational cost. It should be noted that the selection (4.6.1) of the vector in $\mathcal{R}(\hat{G})$, used in the Min–Norm algorithm, is critical in obtaining frequency estimates with satisfactory statistical accuracy. Other choices of vectors in $\mathcal{R}(\hat{G})$ could give rather poor accuracy. In addition, there is empirical evidence that the use of the minimum–norm vector in $\mathcal{R}(\hat{G})$, as in (4.6.1), can decrease the risk of spurious frequency estimates, as compared with the use of other vectors in $\mathcal{R}(\hat{G})$ or even with MUSIC. See Complement 6.5.1 for details on this aspect.

## 4.7 ESPRIT METHOD

Let

$$A_1 = [I_{m-1} \quad 0]A \qquad (m-1) \times n \tag{4.7.1}$$

and

$$A_2 = [0 \quad I_{m-1}]A \qquad (m-1) \times n \tag{4.7.2}$$

where $I_{m-1}$ is the identity matrix of dimension $(m-1) \times (m-1)$ and $[I_{m-1} \quad 0]$ and $[0 \quad I_{m-1}]$ are $(m-1) \times m$. It is readily verified that

$$A_2 = A_1 D \tag{4.7.3}$$

where

$$D = \begin{bmatrix} e^{-i\omega_1} & & 0 \\ & \ddots & \\ 0 & & e^{-i\omega_n} \end{bmatrix} \tag{4.7.4}$$

$D$ is a unitary matrix, so the transformation in (4.7.3) is a *rotation*. ESPRIT (*Estimation of Signal Parameters by Rotational Invariance Techniques*: [Paulraj, Roy, and Kailath 1986; Roy and Kailath 1989]; see also [Kung, Arun, and Rao 1983]), relies on the rotational transformation (4.7.3), as we detail next.

Similarly to (4.7.1) and (4.7.2), define

$$S_1 = [I_{m-1} \quad 0]S \tag{4.7.5}$$

$$S_2 = [0 \quad I_{m-1}]S \tag{4.7.6}$$

From (4.5.12), we have that

$$S = AC \tag{4.7.7}$$

where $C$ is the $n \times n$ nonsingular matrix given by

$$C = PA^* S \mathring{\Lambda}^{-1} \tag{4.7.8}$$

(Observe that both $S$ and $A$ in (4.7.7) have full column rank, and hence, $C$ must be nonsingular; see Result R2 in Appendix A.) The foregoing explicit expression for $C$ actually has no relevance to the present discussion. It is only (4.7.7) and the fact that $C$ is nonsingular that count.

By using (4.7.1)–(4.7.3) and (4.7.7), we can write

$$S_2 = A_2 C = A_1 DC = S_1 C^{-1} DC = S_1 \phi \tag{4.7.9}$$

where

$$\phi \triangleq C^{-1}DC \tag{4.7.10}$$

The Vandermonde structure of $A$, implies that the matrices $A_1$ and $A_2$ have full column rank (equal to $n$). In view of (4.7.7), $S_1$ and $S_2$ must also have full column rank. It then follows from (4.7.9) that the matrix $\phi$ is given uniquely by

$$\phi = (S_1^* S_1)^{-1} S_1^* S_2 \tag{4.7.11}$$

This formula expresses $\phi$ as a function of some quantities that can be estimated from the available sample. The importance of being able to estimate $\phi$ stems from the fact that $\phi$ and $D$ have the same eigenvalues. (This can be seen from equation (4.7.10), which is a *similarity transformation* relating $\phi$ and $D$, along with Result R6 in Appendix A.)

ESPRIT uses the previous observations to compute frequency estimates as described here:

ESPRIT estimates the frequencies $\{\omega_k\}_{k=1}^n$ as $-\arg(\hat{\nu}_k)$, where $\{\hat{\nu}_k\}_{k=1}^n$ are the eigenvalues of the following (consistent) estimate of the matrix $\phi$:

$$\hat{\phi} = (\hat{S}_1^* \hat{S}_1)^{-1} \hat{S}_1^* \hat{S}_2 \tag{4.7.12}$$

It should be noted that this estimate of $\phi$ is implicitly obtained by solving the linear system of equations

$$\hat{S}_1 \hat{\phi} \simeq \hat{S}_2 \tag{4.7.13}$$

by an *LS method*. It has been empirically observed that better finite-sample accuracy might be achieved if (4.7.13) is solved for $\hat{\phi}$ by a *Total LS method*. (See Section A.8 and [Van Huffel and Vandewalle 1991] for discussions on the TLS approach.)

The *statistical accuracy* of ESPRIT is similar to that of the previously described methods: HOYW, MUSIC, and Min–Norm. In fact, in most cases, ESPRIT may provide slightly more accurate frequency estimates than do the other methods mentioned, yet at similar computational cost. In addition, unlike these other methods, ESPRIT has *no problem* with separating the "signal roots" from the "noise roots," as can be seen from (4.7.12). Note that this property is shared by the modified MUSIC method (discussed in Section 4.5); however, in many cases, ESPRIT outperforms modified MUSIC in terms of statistical accuracy. All these considerations recommend ESPRIT as the first choice in a frequency estimation application.

## 4.8 FORWARD–BACKWARD APPROACH

The previously described eigenanalysis-based methods (MUSIC, Min–Norm, and ESPRIT) derive their frequency estimates from the eigenvectors of the sample covariance matrix $\hat{R}$, (4.5.14), which

is restated here for easy reference:

$$\hat{R} = \frac{1}{N}\sum_{t=m}^{N}\begin{bmatrix} y(t) \\ \vdots \\ y(t-m+1) \end{bmatrix}[y^*(t)\ldots y^*(t-m+1)] \tag{4.8.1}$$

$\hat{R}$ is recognized to be the matrix that appears in the least-squares (LS) estimation of the coefficients $\{\alpha_k\}$ of an $m$th-order *forward* linear predictor of $y^*(t+1)$:

$$\hat{y}^*(t+1) = \alpha_1 y^*(t) + \ldots + \alpha_m y^*(t-m+1) \tag{4.8.2}$$

For this reason, the methods that obtain frequency estimates from $\hat{R}$ are named *forward (F) approaches*.

Extensive numerical experience with the aforementioned methods has shown that the corresponding frequency-estimation accuracy can be enhanced by using, in lieu of $\hat{R}$, the modified sample covariance matrix

$$\tilde{R} = \frac{1}{2}(\hat{R} + J\hat{R}^T J) \tag{4.8.3}$$

where

$$J = \begin{bmatrix} 0 & & 1 \\ & \iddots & \\ 1 & & 0 \end{bmatrix} \tag{4.8.4}$$

is the so-called "*exchange*" (or "*reversal*") *matrix*. The second term in (4.8.3) has the following detailed form:

$$J\hat{R}^T J = \frac{1}{N}\sum_{t=m}^{N}\begin{bmatrix} y^*(t-m+1) \\ \vdots \\ y^*(t) \end{bmatrix}[y(t-m+1)\ldots y(t)] \tag{4.8.5}$$

The matrix (4.8.5) is the one that appears in the LS estimate of the coefficients of an $m$th-order *backward* linear predictor of $y(t-m)$:

$$\hat{y}(t-m) = \mu_1 y(t-m+1) + \ldots + \mu_m y(t) \tag{4.8.6}$$

This observation, along with the previous remark made about $\hat{R}$, suggests the name *forward–backward (FB) approaches* for methods that obtain frequency estimates from $\tilde{R}$ in (4.8.3).

The $(i,j)$ element of $\tilde{R}$ is given by

$$\tilde{R}_{i,j} = \frac{1}{2N}\sum_{t=m}^{N}[y(t-i)y^*(t-j) + y^*(t-m+1+i)y(t-m+1+j)]$$
$$\triangleq T_1 + T_2 \qquad (i,j = 0,\ldots,m-1) \tag{4.8.7}$$

Assume that $i \le j$ (the other case $i \ge j$ can be similarly treated). Let $\hat{r}(j-i)$ denote the usual sample covariance:

$$\hat{r}(j-i) = \frac{1}{N}\sum_{t=(j-i)+1}^{N} y(t)y^*(t-(j-i)) \tag{4.8.8}$$

A straightforward calculation shows that the two terms $T_1$ and $T_2$ in (4.8.7) can be written as

$$T_1 = \frac{1}{2N}\sum_{p=m-i}^{N-i} y(p)y^*(p-(j-i)) = \frac{1}{2}\hat{r}(j-i) + \mathcal{O}(1/N) \tag{4.8.9}$$

and

$$T_2 = \frac{1}{2N}\sum_{p=j+1}^{N-m+j+1} y(p)y^*(p-(j-i)) = \frac{1}{2}\hat{r}(j-i) + \mathcal{O}(1/N) \tag{4.8.10}$$

where $\mathcal{O}(1/N)$ denotes a term that tends to zero as $1/N$ when $N$ increases (it is here assumed that $m \ll N$). It follows from (4.8.7)–(4.8.10) that, for large $N$, the difference between $\tilde{R}_{i,j}$ or $\hat{R}_{i,j}$ and the sample covariance lag $\hat{r}(j-i)$ is "small." Hence, the frequency estimation methods based on $\hat{R}$ or $\tilde{R}$ (or on $[\hat{r}(j-i)]$) may be expected to have similar performances in large samples.

In summary, it follows from the previous discussion that the empirically observed performance superiority of the forward–backward approach over the forward-only approach should only be manifest in samples with relatively small lengths. As such, this superiority cannot easily be established by theoretical means. Let us then argue heuristically.

First, note that the transformation $J(\cdot)^T J$ is such that the following equalities hold:

$$(\hat{R})_{i,j} = (J\hat{R}J)_{m-i,m-j} = (J\hat{R}^T J)_{m-j,m-i} \tag{4.8.11}$$

and

$$(\hat{R})_{m-j,m-i} = (J\hat{R}^T J)_{i,j} \tag{4.8.12}$$

This implies that the $(i,j)$ and $(m-j,m-i)$ elements of $\tilde{R}$ are both given by

$$\tilde{R}_{i,j} = \tilde{R}_{m-j,m-i} = \frac{1}{2}(\hat{R}_{i,j} + \hat{R}_{m-j,m-i}) \tag{4.8.13}$$

Equations (4.8.11)–(4.8.12) imply that $\tilde{R}$ is invariant to the transformation $J(\cdot)^T J$:

$$J\tilde{R}^T J = \tilde{R} \tag{4.8.14}$$

Such a matrix is said to be *persymmetric* (or *centrosymmetric*). In order to see the reason for this name, note that $\tilde{R}$ is Hermitian (symmetric in the real-valued case) with respect to its main diagonal; *in addition*, $\tilde{R}$ is symmetric about its main antidiagonal. Indeed, the equal elements $\tilde{R}_{i,j}$ and $\tilde{R}_{m-j,m-i}$ of $\tilde{R}$ belong to the same diagonal as $i-j=(m-j)-(m-i)$. They are also symmetrically placed with respect to the main antidiagonal; $\tilde{R}_{i,j}$ lies on antidiagonal $(i+j)$, $\tilde{R}_{m-j,m-i}$ on the $[2m-(j+i)]$th one, and the main antidiagonal is the $m$th one (and $m=[(i+j)+2m-(i+j)]/2$).

The theoretical (and unknown) covariance matrix $R$ is Toeplitz and hence persymmetric. Since $\tilde{R}$ is persymmetric like $R$, whereas $\hat{R}$ is not, we may expect $\tilde{R}$ to be a better estimate of $R$ than $\hat{R}$. In turn, this means that the frequency estimates derived from $\tilde{R}$ are likely to be more accurate than those obtained from $\hat{R}$.

The impact of enforcing the persymmetric property can be seen by examining, say, the $(1,1)$ and $(m,m)$ elements of $\hat{R}$ and $\tilde{R}$. Both the $(1,1)$ and $(m,m)$ elements of $\hat{R}$ are estimates of $r(0)$; however, the $(1,1)$ element does not use the first $(m-1)$ lag products $|y(1)|^2,\ldots,|y(m-1)|^2$, and the $(m,m)$ element does not use the last $(m-1)$ lag products $|y(N-m+2)|^2,\ldots,|y(N)|^2$. If $N \gg m$, the omission of these lag products is negligible; for small $N$, however, this omission can be significant. On the other hand, all lag products of $y(t)$ are used to form the $(1,1)$ and $(m,m)$ elements of $\tilde{R}$, and, in general, the $(i,j)$ element of $\tilde{R}$ uses more lag products of $y(t)$ than does the corresponding element of $\hat{R}$. (For more details on the FB approach, we refer the reader to, e.g., [Rao and Hari 1993; Pillai 1989]; see also Complement 6.5.8.)

Finally, the reader might wonder why we do not replace $\hat{R}$ by a Toeplitz estimate, obtained (for example) by averaging the elements along each diagonal of $\hat{R}$. This Toeplitz estimate would at first seem to be a better approximation of $R$ than either $\hat{R}$ or $\tilde{R}$. The reason why we do not "Toeplitz-ize" $\hat{R}$ or $\tilde{R}$ is that, for finite $N$ and infinite signal-to-noise ratio ($\sigma^2 \to 0$), the use of either $\hat{R}$ or $\tilde{R}$ gives exact frequency estimates, whereas the Toeplitz-averaged approximation of $R$ does not. As $\sigma^2 \to 0$, both $\hat{R}$ and $\tilde{R}$ have rank $n$, but the Toeplitz-averaged approximation of $R$ has full rank in general.

## 4.9 COMPLEMENTS

### 4.9.1 Mean-Square Convergence of Sample Covariances for Line Spectral Processes

In this complement, we prove that

$$\lim_{N\to\infty} \hat{r}(k) = r(k) \quad \text{(in the mean-square sense)} \tag{4.9.1}$$

(i.e., $\lim_{N\to\infty} E\left\{|\hat{r}(k) - r(k)|^2\right\} = 0$). The above result has already been referred to in Section 4.4, in the discussion on the rank properties of $\hat{\Omega}$ and $\Omega$. It is also the basic result from which the consistency of all covariance-based frequency estimators discussed in this chapter can be readily concluded. Note that a signal $\{y(t)\}$ satisfying (4.9.1) is said to be *second-order ergodic*. (See [Söderström and Stoica 1989; Brockwell and Davis 1991] for a more detailed discussion of the ergodicity property.)

A straightforward calculation gives

$$\begin{aligned}\hat{r}(k) &= \frac{1}{N}\sum_{t=k+1}^{N}[x(t)+e(t)][x^*(t-k)+e^*(t-k)] \\ &= \frac{1}{N}\sum_{t=k+1}^{N}[x(t)x^*(t-k)+x(t)e^*(t-k)+e(t)x^*(t-k) \\ &\quad + e(t)e^*(t-k)] \triangleq T_1+T_2+T_3+T_4 \end{aligned} \tag{4.9.2}$$

The limit of $T_1$ is found as follows. First note that

$$\begin{aligned}\lim_{N\to\infty} E\left\{|T_1 - r_x(k)|^2\right\} &= \lim_{N\to\infty}\left\{\frac{1}{N^2}\sum_{t=k+1}^{N}\sum_{s=k+1}^{N} E\left\{x(t)x^*(t-k)x^*(s)x(s-k)\right\}\right. \\ &\quad \left. -\left(\frac{2}{N}\sum_{t=k+1}^{N}|r_x(k)|^2\right)+|r_x(k)|^2\right\} \\ &= \lim_{N\to\infty}\left\{\frac{1}{N^2}\sum_{t=k+1}^{N}\sum_{s=k+1}^{N} E\left\{x(t)x^*(t-k)x^*(s)x(s-k)\right\}\right\} \\ &\quad -|r_x(k)|^2\end{aligned}$$

Now,

$$\begin{aligned}E\left\{x(t)x^*(t-k)x^*(s)x(s-k)\right\} &= \sum_{p=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n} a_p a_j a_l a_m e^{i(\omega_p-\omega_j)t} e^{i(\omega_m-\omega_l)s} \\ &\quad \cdot e^{i(\omega_j-\omega_m)k} E\left\{e^{i\varphi_p} e^{-i\varphi_j} e^{i\varphi_m} e^{-i\varphi_l}\right\} \\ &= \sum_{p=1}^{n}\sum_{j=1}^{n}\sum_{l=1}^{n}\sum_{m=1}^{n} a_p a_j a_l a_m e^{i(\omega_p-\omega_j)t} e^{i(\omega_m-\omega_l)s} \\ &\quad \cdot e^{i(\omega_j-\omega_m)k}\left(\delta_{p,j}\delta_{m,l}+\delta_{p,l}\delta_{m,j}-\delta_{p,j}\delta_{m,l}\delta_{p,m}\right)\end{aligned}$$

where the last equality follows from the assumed independence of the initial phases $\{\varphi_k\}$. Combining the results of the above two calculations yields

$$\begin{aligned}\lim_{N\to\infty} E\left\{|T_1 - r_x(k)|^2\right\} &= \lim_{N\to\infty} \frac{1}{N^2} \sum_{t=k+1}^{N} \sum_{s=k+1}^{N} \left\{ \sum_{p=1}^{n} \sum_{m=1}^{n} a_p^2 a_m^2 e^{i(\omega_p-\omega_m)k} \right. \\ &\qquad \left. + \sum_{p=1}^{n} \sum_{m=1}^{n} a_p^2 a_m^2 e^{i(\omega_p-\omega_m)(t-s)} - \sum_{p=1}^{n} a_p^4 \right\} - |r_x(k)|^2 \\ &= \sum_{p=1}^{n} \sum_{\substack{m=1 \\ m\neq p}}^{n} a_p^2 a_m^2 \lim_{N\to\infty} \frac{1}{N^2} \sum_{\tau=-N}^{N} (N - |\tau|) e^{i(\omega_p-\omega_m)\tau} \\ &= 0 \end{aligned} \tag{4.9.3}$$

It follows that $T_1$ converges to $r(k)$ (in the mean-square sense) as $N$ tends to infinity.

The limits of $T_2$ and $T_3$ are equal to zero, as shown next for $T_2$; the proof for $T_3$ is similar. Using the fact that $\{x(t)\}$ and $\{e(t)\}$ are, by assumption, independent random signals, we get

$$\begin{aligned} E\left\{|T_2|^2\right\} &= \frac{1}{N^2} \sum_{t=k+1}^{N} \sum_{s=k+1}^{N} E\left\{x(t)e^*(t-k)x^*(s)e(s-k)\right\} \\ &= \frac{\sigma^2}{N^2} \sum_{t=k+1}^{N} \sum_{s=k+1}^{N} E\left\{x(t)x^*(s)\right\} \delta_{t,s} \\ &= \frac{\sigma^2}{N^2} \sum_{t=k+1}^{N} E\left\{|x(t)|^2\right\} = \frac{(N-k)\sigma^2}{N^2} E\left\{|x(t)|^2\right\} \end{aligned} \tag{4.9.4}$$

which tends to zero as $N \to \infty$. Hence, $T_2$ (and, similarly, $T_3$) converges to zero in the mean-square sense.

The last term, $T_4$, in (4.9.2), converges to $\sigma^2\delta_{k,0}$ by the "law of large numbers" (as shown in [Söderström and Stoica 1989; Brockwell and Davis 1991]). In fact, it is readily verified, at least under the Gaussian hypothesis, that

$$\begin{aligned} E\left\{|T_4 - \sigma^2\delta_{k,0}|^2\right\} &= \frac{1}{N^2} \sum_{t=k+1}^{N} \sum_{s=k+1}^{N} E\left\{e(t)e^*(t-k)e^*(s)e(s-k)\right\} \\ &\quad - \sigma^2\delta_{k,0} \left\{ \frac{1}{N} \sum_{t=k+1}^{N} E\left\{e(t)e^*(t-k) + e^*(t)e(t-k)\right\} \right\} \\ &\quad + \sigma^4\delta_{k,0} \end{aligned}$$

$$= \frac{1}{N^2} \sum_{t=k+1}^{N} \sum_{s=k+1}^{N} [\sigma^4 \delta_{k,0} + \sigma^4 \delta_{t,s}]$$

$$- 2\sigma^4 \delta_{k,0} \frac{1}{N} \sum_{t=k+1}^{N} (\delta_{k,0}) + \sigma^4 \delta_{k,0}$$

$$\to \sigma^4 \delta_{k,0} - 2\sigma^4 \delta_{k,0} + \sigma^4 \delta_{k,0} = 0 \tag{4.9.5}$$

Hence, $T_4$ converges to $\sigma^2 \delta_{k,0}$ in the mean-square sense if $e(t)$ is Gaussian. It can be shown by using the law of large numbers that $T_4 \to \sigma^2 \delta_{k,0}$ in the mean-square sense even if $e(t)$ is non-Gaussian, as long as the fourth-order moment of $e(t)$ is finite.

Next, observe that since, for example, $E\{|T_2|^2\}$ and $E\{|T_3|^2\}$ converge to zero, then $E\{T_2 T_3^*\}$ also converges to zero (as $N \to \infty$); this is so because

$$\left| E\left\{T_2 T_3^*\right\} \right| \leq \left[ E\left\{|T_2|^2\right\} E\left\{|T_3|^2\right\} \right]^{1/2}$$

With this observation, the proof of (4.9.1) is complete.

### 4.9.2 The Carathéodory Parameterization of a Covariance Matrix

The covariance matrix model in (4.2.7) is more general than it might appear at first sight. We show that for *any* given covariance matrix $R = \{r(i-j)\}_{i,j=1}^{m}$, there exist $n \leq m$, $\sigma^2$ and $\{\omega_k,\ \alpha_k\}_{k=1}^{n}$ such that $R$ can be written as in (4.2.7). Equation (4.2.7), associated with an arbitrary given covariance matrix $R$, is named the *Carathéodory parameterization* of $R$.

Let $\sigma^2$ denote the minimum eigenvalue of $R$. Because $\sigma^2$ is not necessarily unique, let $\bar{n}$ denote its multiplicity and set $n = m - \bar{n}$. Define

$$\Gamma = R - \sigma^2 I$$

The matrix $\Gamma$ is positive semidefinite and Toeplitz and, hence, must be the covariance matrix associated with a stationary signal, say $y(t)$:

$$\Gamma = E\left\{ \begin{bmatrix} y(t) \\ \vdots \\ y(t-m+1) \end{bmatrix} [y^*(t) \ldots y^*(t-m+1)] \right\}$$

By definition,

$$\operatorname{rank}(\Gamma) = n \tag{4.9.6}$$

which implies that there must exist a linear combination between $\{y(t), \ldots, y(t-n)\}$ for all $t$. Moreover, both $y(t)$ and $y(t-n)$ must appear with nonzero coefficients in that linear combination (otherwise either $\{y(t) \ldots y(t-n+1)\}$ or $\{y(t-1) \ldots y(t-n)\}$ would be linearly related, and

rank($\Gamma$) would be less than $n$, which would contradict (4.9.6)). Hence, $y(t)$ obeys the homogeneous AR equation

$$B(z)y(t) = 0 \tag{4.9.7}$$

where $z^{-1}$ is the unit delay operator, and

$$B(z) = 1 + b_1 z^{-1} + \cdots + b_n z^{-n}$$

with $b_n \neq 0$. Let $\phi(\omega)$ denote the PSD of $y(t)$. Then we have the following equivalences:

$$\begin{aligned} B(z)y(t) = 0 &\Longleftrightarrow \int_{-\pi}^{\pi} |B(\omega)|^2 \, \phi(\omega) \, d\omega = 0 \\ &\Longleftrightarrow |B(\omega)|^2 \, \phi(\omega) = 0 \\ &\Longleftrightarrow \{\text{If } \phi(\omega) > 0 \text{ then } B(\omega) = 0\} \\ &\Longleftrightarrow \{\phi(\omega) > 0 \text{ for at most } n \text{ values of } \omega\} \end{aligned}$$

Furthermore,

$$\begin{aligned} &\{y(t), \ldots y(t-n+1) \text{ are linearly independent}\} \\ &\quad\Longleftrightarrow \left\{E\left\{|g_0 y(t) + \ldots + g_{n-1} y(t-n+1)|^2\right\} > 0 \text{ for every } [g_0 \ldots g_{n-1}]^T \neq 0\right\} \\ &\quad\Longleftrightarrow \left\{\int_{-\pi}^{\pi} |G(\omega)|^2 \, \phi(\omega) \, d\omega > 0 \text{ for every } G(z) = \sum_{k=0}^{n-1} g_k z^{-k} \neq 0\right\} \\ &\quad\Longleftrightarrow \{\phi(\omega) > 0 \text{ for at least } n \text{ distinct values of } \omega\} \end{aligned}$$

It follows from these two results that $\phi(\omega) > 0$ for exactly $n$ distinct values of $\omega$. Furthermore, the values of $\omega$ for which $\phi(\omega) > 0$ are given by the $n$ roots of the equation $B(\omega) = 0$. A signal $y(t)$ with such a PSD consists of a sum of $n$ sinusoidal components with an $m \times m$ covariance matrix given by

$$\Gamma = APA^* \tag{4.9.8}$$

(*cf.* (4.2.7)). In (4.9.8), the frequencies $\{\omega_k\}_{k=1}^n$ are defined as previously indicated and can be found from $\Gamma$ by using any of the subspace-based frequency-estimation methods in this chapter. Once $\{\omega_k\}$ are available, $\{\alpha_i^2\}$ can be determined from $\Gamma$. (Show that.) By combining the additive decomposition $R = \Gamma + \sigma^2 I$ and (4.9.8), we obtain (4.2.7). With this observation, the derivation of the Carathéodory parameterization is complete.

It is interesting to note that the sinusoids-in-noise signal that "realizes" a given covariance sequence $\{r(0), \ldots, r(m)\}$ also provides a *positive definite extension* of that sequence. More precisely, the covariance lags $\{r(m+1),\ r(m+2), \ldots\}$ derived from the sinusoidal signal equation, when appended to $\{r(0), \ldots, r(m)\}$, provide a positive definite covariance sequence of infinite length. The AR covariance realization (see Complement 3.9.2) is the other well-known method for obtaining a positive definite extension of a given covariance sequence of finite length.

### 4.9.3 Using the Unwindowed Periodogram for Sine Wave Detection in White Noise

As shown in Section 4.3, the unwindowed periodogram is an accurate frequency estimation method whenever the minimum frequency separation is larger than $1/N$. A simple intuitive explanation as to why the unwindowed periodogram is a better frequency estimator than the windowed periodogram(s) follows. The principal effect of a window is to remove the tails of the sample covariance sequence from the periodogram formula; this is appropriate for signals whose covariance sequence "rapidly" goes to zero, but inappropriate for sinusoidal signals, whose covariance sequence never dies out. (For sinusoidal signals, the use of a window is expected to introduce a significant bias in the estimated spectrum.) Note, however, that, if the data contains sinusoidal components with significantly different amplitudes, then it could be advisable to use a (mildly) windowed periodogram. This will induce bias in the frequency estimates, but, on the other hand, will reduce the leakage and hence make it possible to detect the low-amplitude components.

When using the (unwindowed) periodogram for frequency estimation, an important problem is to infer whether any of the many peaks of the erratic periodogram plot can really be associated with the existence of a sinusoidal component in the data. In order to be more precise, consider the following two hypotheses:

$H_0$: The data consists of (complex circular Gaussian) white noise only (with unknown variance $\sigma^2$).

$H_1$: The data consists of a sum of sinusoidal components and noise.

Deciding between $H_0$ and $H_1$ constitutes the so-called *(signal) detection problem*. A solution to the detection problem can be obtained as follows: From the calculations leading to the result (2.4.21), one can see that the normalized periodogram values in (4.9.15) are independent random variables (under $H_0$). It remains to derive their distribution. Let

$$\epsilon_r(\omega) = \frac{\sqrt{2}}{\sigma\sqrt{N}} \sum_{t=1}^{N} \mathrm{Re}[e(t)e^{-i\omega t}]$$

$$\epsilon_i(\omega) = \frac{\sqrt{2}}{\sigma\sqrt{N}} \sum_{t=1}^{N} \mathrm{Im}[e(t)e^{-i\omega t}]$$

With this notation, and under the null hypothesis $H_0$,

$$2\hat{\phi}_p(\omega)/\sigma^2 = \epsilon_r^2(\omega) + \epsilon_i^2(\omega) \tag{4.9.9}$$

For any two complex scalars, $z_1$ and $z_2$, we have

$$\mathrm{Re}(z_1)\,\mathrm{Im}(z_2) = \frac{z_1 + z_1^*}{2}\,\frac{z_2 - z_2^*}{2i} = \frac{1}{2}\,\mathrm{Im}\,(z_1 z_2 + z_1^* z_2) \tag{4.9.10}$$

and, similarly,

$$\mathrm{Re}(z_1)\,\mathrm{Re}(z_2) = \frac{1}{2}\,\mathrm{Re}(z_1 z_2 + z_1^* z_2) \tag{4.9.11}$$

$$\mathrm{Im}(z_1)\,\mathrm{Im}(z_2) = \frac{1}{2}\,\mathrm{Re}(-z_1 z_2 + z_1^* z_2) \tag{4.9.12}$$

By making use of (4.9.10)–(4.9.12), we can write

$$\begin{aligned} E\{\epsilon_r(\omega)\epsilon_i(\omega)\} &= \frac{1}{\sigma^2 N}\,\mathrm{Im}\left\{\sum_{t=1}^{N}\sum_{s=1}^{N} E\left\{e(t)e(s)e^{-i\omega(t+s)} + e^*(t)e(s)e^{i\omega(t-s)}\right\}\right\} \\ &= \mathrm{Im}\{1\} = 0 \\ E\left\{\epsilon_r^2(\omega)\right\} &= \frac{1}{\sigma^2 N}\,\mathrm{Re}\left\{\sum_{t=1}^{N}\sum_{s=1}^{N} E\left\{e(t)e(s)e^{-i\omega(t+s)} + e^*(t)e(s)e^{i\omega(t-s)}\right\}\right\} \\ &= \mathrm{Re}\{1\} = 1 \qquad (4.9.13) \\ E\left\{\epsilon_i^2(\omega)\right\} &= \frac{1}{\sigma^2 N}\,\mathrm{Re}\left\{\sum_{t=1}^{N}\sum_{s=1}^{N} E\left\{-e(t)e(s)e^{-i\omega(t+s)} + e^*(t)e(s)e^{i\omega(t-s)}\right\}\right\} \\ &= \mathrm{Re}\{1\} = 1 \qquad (4.9.14) \end{aligned}$$

In addition, note that the random variables $\epsilon_r(\omega)$ and $\epsilon_i(\omega)$ are zero-mean Gaussian distributed, because they are linear transformations of the Gaussian white-noise sequence. Then, it follows that, *under* $H_0$,

> The random variables
>
> $$\{2\hat{\phi}_p(\omega_k)/\sigma^2\}_{k=1}^{N},$$
>
> with $\min_{k\neq j}\,|\omega_k - \omega_j| \geq 2\pi/N$, are asymptotically independent and $\chi^2$ distributed with 2 degrees of freedom. (4.9.15)

(See, e.g., [Priestley 1981] and [Söderström and Stoica 1989] for the definition and properties of the $\chi^2$ distribution.) It is worth noting that, if $\{\omega_k\}$ are equal to the Fourier frequencies $\{2\pi k/N\}_{k=0}^{N-1}$, then the previous distributional result is *exactly valid* (i.e., it holds in samples of finite length; see, for example, equation (2.4.26)). However, this observation is not as important as it might at first seem, because $\sigma^2$ in (4.9.15) is unknown. When the noise power in (4.9.15) is replaced by a consistent estimate $\hat{\sigma}^2$, the normalized periodogram values so obtained,

$$\{2\hat{\phi}_p(\omega_k)/\hat{\sigma}^2\} \tag{4.9.16}$$

are $\chi^2(2)$ distributed only asymptotically (for $N \gg 1$). A consistent estimate of $\sigma^2$ can be obtained as follows: From (4.9.9), (4.9.13), and (4.9.14) we have that, under $H_0$,

$$E\left\{\hat{\phi}_p(\omega_k)\right\} = \sigma^2 \qquad \text{for } k = 1, 2, \ldots, N$$

Since $\{\hat{\phi}_p(\omega_k)\}_{k=1}^N$ are independent random variables, a consistent estimate of $\sigma^2$ is given by

$$\hat{\sigma}^2 = \frac{1}{N} \sum_{k=1}^{N} \hat{\phi}_p\ (\omega_k)$$

Inserting this expression for $\hat{\sigma}^2$ into (4.9.16) leads to the following "test statistic":

$$\mu_k = \frac{2N\hat{\phi}_p(\omega_k)}{\sum_{k=1}^{N} \hat{\phi}_p(\omega_k)}$$

In accordance with the (asymptotic) $\chi^2$ distribution of $\{\mu_k\}$, we have (for any given $c \geq 0$; see, for example, [PRIESTLEY 1981])

$$\Pr(\mu_k \leq c) = \int_0^c \frac{1}{2} e^{-x/2}\, dx = 1 - e^{-c/2}. \tag{4.9.17}$$

Let

$$\mu = \max_k [\mu_k]$$

Using (4.9.17) (and the fact that $\{\mu_k\}$ are independent random variables) we find that (for any $c \geq 0$)

$$\begin{aligned} \Pr(\mu > c) &= 1 - \Pr(\mu \leq c) \\ &= 1 - \Pr(\mu_k \leq c \text{ for all } k) \\ &= 1 - (1 - e^{-c/2})^N \qquad (\text{under } H_0) \end{aligned}$$

This result can be used to set a bound on $\mu$ that, under $H_0$, holds with a (high) preassigned probability, say $1 - \alpha$. More precisely, *let $\alpha$ be given* (e.g., $\alpha = 0.05$), and solve for $c$ from the equation

$$(1 - e^{-c/2})^N = 1 - \alpha$$

Then

- If $\mu \leq c$, accept $H_0$ with an unknown risk. (That risk depends on the signal-to-noise ratio (SNR). The lower the SNR, the larger the risk of accepting $H_0$ when it does not hold.)
- If $\mu > c$, reject $H_0$ with a risk equal to $\alpha$.

It should be noted that, whenever $H_0$ is rejected by the above test, what we can really infer is that the periodogram peak in question is significant enough to make the existence of a sinusoidal component in the studied data highly probable. However, the previous test does not tell us the number of sinusoidal components in the data. In order to determine that number, the test should be continued by looking at the second-highest peak in the periodogram. For a test of the significance of the second-highest value of the periodogram, and so on, we refer to [Priestley 1981].

Finally, we note that, in addition to the test presented in this complement, there are several other tests to decide between the hypotheses $H_0$ and $H_1$; see [Priestley 1997] for a review.

### 4.9.4 NLS Frequency Estimation for a Sinusoidal Signal with Time-Varying Amplitude

Consider the sinusoidal data model in (4.1.1) for the case of a single component ($n = 1$), but with a time-varying amplitude

$$y(t) = \alpha(t)e^{i(\omega t+\varphi)} + e(t), \quad t = 1, \ldots, N \tag{4.9.18}$$

where $\alpha(t) \in \mathbf{R}$ is an arbitrary unknown envelope modulating the sinusoidal signal. The NLS estimates of $\alpha(t)$, $\omega$, and $\varphi$ are obtained by minimizing the criterion

$$f = \sum_{t=1}^{N} \left| y(t) - \alpha(t)e^{i(\omega t+\varphi)} \right|^2$$

(*cf.* (4.3.1)). In this complement, we show that this seemingly complicated minimization problem has, in fact, a simple solution. We also discuss briefly an FFT-based algorithm for computing that solution. The reader interested in more details on the topic of this complement can consult [Besson and Stoica 1999; Stoica, Besson, and Gershman 2001] and references therein.

A straightforward calculation shows that

$$f = \sum_{t=1}^{N} \left\{ \left| y(t) \right|^2 + \left[ \alpha(t) - \mathrm{Re}\left( e^{-i(\omega t+\varphi)} y(t) \right) \right]^2 - \left[ \mathrm{Re}\left( e^{-i(\omega t+\varphi)} y(t) \right) \right]^2 \right\} \tag{4.9.19}$$

The minimization of (4.9.19) with respect to $\alpha(t)$ is immediate:

$$\hat{\alpha}(t) = \mathrm{Re}\left( e^{-i(\hat{\omega} t+\hat{\varphi})} y(t) \right) \tag{4.9.20}$$

Note that the NLS estimates $\hat{\omega}$ and $\hat{\varphi}$ are yet to be determined. Inserting (4.9.20) into (4.9.19) shows that the NLS estimates of $\varphi$ and $\omega$ are obtained by maximizing the function

$$g = 2\sum_{t=1}^{N}\left[\mathrm{Re}\left(e^{-i(\omega t+\varphi)}y(t)\right)\right]^2$$

where the factor 2 has been introduced for the sake of convenience. For any complex number $c$ we have

$$[\mathrm{Re}(c)]^2 = \frac{1}{4}\left(c+c^*\right)^2 = \frac{1}{2}\left[|c|^2 + \mathrm{Re}\left(c^2\right)\right]$$

It follows that

$$\begin{aligned} g &= \sum_{t=1}^{N}\left\{|y(t)|^2 + \mathrm{Re}\left[e^{-2i(\omega t+\varphi)}y^2(t)\right]\right\} \\ &= \text{constant} + \left|\sum_{t=1}^{N} y^2(t)e^{-i2\omega t}\right| \cdot \cos\left[\arg\left(\sum_{t=1}^{N} y^2(t)e^{-i2\omega t}\right) - 2\varphi\right] \end{aligned} \tag{4.9.21}$$

Clearly, the maximizing $\varphi$ is given by

$$\hat{\varphi} = \frac{1}{2}\arg\left(\sum_{t=1}^{N} y^2(t)e^{-i2\hat{\omega}t}\right)$$

with the NLS estimate of $\omega$ given by

$$\boxed{\hat{\omega} = \arg\max_{\omega}\left|\sum_{t=1}^{N} y^2(t)e^{-i2\omega t}\right|} \tag{4.9.22}$$

It is important to note that the maximization in (4.9.22) should be conducted over $[0, \pi]$ instead of over $[0, 2\pi]$; indeed, the function in (4.9.22) is periodic with a period equal to $\pi$. The restriction of $\omega$ to $[0, \pi]$ is not a peculiar feature of the NLS approach; rather, it is a consequence of the generality of the problem considered in this complement. This is easily seen by making the substitution $\omega \to \omega + \pi$ in (4.9.18), which yields

$$y(t) = \tilde{\alpha}(t)e^{i(\omega t+\varphi)} + e(t), \quad t = 1, \ldots, N$$

where $\tilde{\alpha}(t) = (-1)^t\alpha(t)$ is another valid (i.e., real-valued) envelope. This simple calculation confirms the fact that $\omega$ is uniquely identifiable only in the interval $[0, \pi]$. In applications, the frequency can be made to belong to $[0, \pi]$ by using a sufficiently small sampling period.

The previous estimate of $\omega$ should be contrasted with the NLS estimate of $\omega$ in the constant-amplitude case (see (4.3.11), (4.3.17)):

$$\hat{\omega} = \arg\max_{\omega} \left| \sum_{t=1}^{N} y(t) e^{-i\omega t} \right| \quad (\text{for } \alpha(t) = \text{constant}) \tag{4.9.23}$$

There is a striking similarity between (4.9.22) and (4.9.23); the only difference between these equations is the squaring of the terms in (4.9.22). As a consequence, we can apply the FFT to the squared data sequence $\{y^2(t)\}$ to obtain the $\hat{\omega}$ in (4.9.22).

The reader perhaps wonders whether there is an *intuitive* reason for the occurrence of the squared data in (4.9.22). A possible way to explain this occurrence goes as follows: Assume that $\alpha(t)$ has zero average value. Then the DFT of $\{\alpha(t)\}$, denoted $A(\bar{\omega})$, takes on small values (theoretically zero) at $\bar{\omega} = 0$. But the DFT of $\alpha(t)e^{i\omega t}$ is $A(\bar{\omega} - \omega)$, so it follows that the modulus of this DFT has a valley instead of a peak at $\bar{\omega} = \omega$; hence, the standard periodogram (see (4.9.23)) should not be used to estimate $\omega$. On the other hand, $\alpha^2(t)$ always has a nonzero average value (or DC component); hence, the modulus of the DFT of $\alpha^2(t)e^{i2\omega t}$ will typically have a peak at $\bar{\omega} = 2\omega$. This observation provides an heuristic reason for the squaring operation in (4.9.22).

### 4.9.5 Monotonically Descending Techniques for Function Minimization

As was explained in Section 4.3, minimizing the NLS criterion with respect to the unknown frequencies is made rather difficult by existence of possibly many local minima and by the sharpness of the global minimum. In this complement (based on [Stoica and Selén 2004a]), we will discuss a number of methods that can be used to solve such a minimization problem. Our discussion is quite general and applies to many other functions, not to just the NLS criterion that is used as an illustrating example in what follows.

We will denote the function to be minimized by $f(\theta)$, where $\theta$ is a vector. Sometimes we will write this function as $f(x, y)$ where $[x^T, y^T]^T = \theta$. The algorithms for minimizing $f(\theta)$ discussed in this complement are iterative. We let $\theta^i$ denote the value taken by $\theta$ at the $i$th iteration (and similarly for $x$ and $y$). The *common feature* of the algorithms included in this complement is that *they all monotonically decrease the function at each iteration*:

$$\boxed{f(\theta^{i+1}) \leq f(\theta^i) \quad \text{for } i = 0, 1, 2, \ldots} \tag{4.9.24}$$

Hereafter, $\theta^0$ denotes the initial value (or estimate) of $\theta$ used by the minimization algorithm in question. Clearly, (4.9.24) is an appealing property, which is the main reason for the interest in the algorithms discussed here. However, we should note that usually (4.9.24) can do no more than guarantee the convergence to a *local minimum* of $f(\theta)$. The goodness of the initial estimate $\theta^0$ will often determine whether the algorithm will converge to the global minimum. In fact, for some of the algorithms to be discussed, not even the convergence to a local minimum is guaranteed. For

example, the EM algorithm (discussed later in this complement) can converge to saddle points or local maxima. (See, for example, [McLachlan and Krishnan 1997].) However, such a behavior is rare in applications, provided that some regularity conditions are satisfied.

### Cyclic Minimizer

To describe the main idea of this type of algorithm in its simplest form, let us partition $\theta$ into two subvectors:

$$\theta = \begin{bmatrix} x \\ y \end{bmatrix}$$

Then the *generic iteration of a cyclic algorithm* for minimizing $f(x, y)$ will have the following form:

$$\begin{aligned} &y^0 = \text{ given} \\ &\text{For } i = 1, 2, \ldots \text{ compute:} \\ &\qquad x^i = \arg\min_x f(x, y^{i-1}) \\ &\qquad y^i = \arg\min_y f(x^i, y) \end{aligned} \tag{4.9.25}$$

Note that (4.9.25) alternates (or cycles) between the minimization of $f(x, y)$ with respect to $x$ for given $y$ and the minimization of $f(x, y)$ with respect to $y$ for given $x$; hence, the name "cyclic" given to this type of algorithm. An obvious modification of (4.9.25) allows us to start with $x^0$, if so desired. It is readily verified that the cyclic minimizer (4.9.25) possesses the property (4.9.24)—that is,

$$f(x^i, y^i) \leq f(x^i, y^{i-1}) \leq f(x^{i-1}, y^{i-1})$$

where the first inequality follows from the definition of $y^i$ and the second from the definition of $x^i$.

*The partitioning of $\theta$ into subvectors is usually done in such a way that the minimization operations in* (4.9.25) *(or at least one of them) are "easy"* (in any case, easier than the minimization of $f$ jointly with respect to $x$ and $y$). Quite often, to achieve this desired property, we need to partition $\theta$ into more than two subvectors. The extension of (4.9.25) to such a case is straightforward and will not be discussed here. However, there is one point about this extension that we would like to make briefly: whenever $\theta$ is partitioned into three or more subvectors, we can choose the way in which the various minimization subproblems are iterated. For instance, if $\theta = [x^T, y^T, z^T]^T$ then we may iterate the minimization steps with respect to $x$ and with respect to $y$ a number of times (with $z$ being fixed), before reestimating $z$, and so forth.

With reference to the NLS problem in Section 4.3, we can apply the preceding ideas to the following natural partitioning of the parameter vector:

$$\theta = \begin{bmatrix} \gamma_1 \\ \gamma_2 \\ \vdots \\ \gamma_n \end{bmatrix}, \qquad \gamma_k = \begin{bmatrix} \omega_k \\ \varphi_k \\ \alpha_k \end{bmatrix} \tag{4.9.26}$$

The main virtue of this partitioning of $\theta$ is that the problem of minimizing the NLS criterion with respect to $\gamma_k$, for given $\{\gamma_j\}$ $(j = 1, \ldots, n; j \neq k)$, can be solved via the FFT (see (4.3.10), (4.3.11)). Furthermore, the cyclic minimizer corresponding to (4.9.26) can be initialized, with $\gamma_2 = \cdots = \gamma_n = 0$, in which case the $\gamma_1$ minimizing the NLS criterion is obtained from the highest peak of the periodogram (which should give a reasonably accurate estimate of $\gamma_1$), and so on.

An elaborated cyclic algorithm, called RELAX, for the minimization of the NLS criterion based on the preceding ideas (see (4.9.26)), was proposed in [Li and Stoica 1996b]. Note that cyclic minimizers are sometimes called *relaxation algorithms*, which provide a motivation for the name given to the algorithm in [Li and Stoica 1996b].

### Majorization Technique

The main idea of this type of iterative technique for minimizing a given function $f(\theta)$ is quite simple. (See, for example, [Heiser 1995] and the references therein.) Assume that, at the $i$th iteration, we can find a function $g_i(\theta)$ (the subindex $i$ indicates the dependence of this function on $\theta^i$) that possesses the following three properties:

$$g_i(\theta^i) = f(\theta^i) \tag{4.9.27}$$

$$g_i(\theta) \geq f(\theta) \tag{4.9.28}$$

and

$$\text{the minimization of } g_i(\theta) \text{ with respect to } \theta \text{ is "easy" (or, in any case, easier than the minimization of } f(\theta)). \tag{4.9.29}$$

Owing to (4.9.28), $g_i(\theta)$ is called a *majorizing function* for $f(\theta)$ at the $i$th iteration. In the majorization technique, the parameter vector at iteration $(i + 1)$ is obtained from the minimization of $g_i(\theta)$:

$$\boxed{\theta^{i+1} = \arg\min_{\theta} g_i(\theta)} \tag{4.9.30}$$

The key property (4.9.24) is satisfied for (4.9.30), since

$$f(\theta^i) = g_i(\theta^i) \geq g_i(\theta^{i+1}) \geq f(\theta^{i+1}) \tag{4.9.31}$$

The first inequality in (4.9.31) follows from the definition of $\theta^{i+1}$ in (4.9.30), the second from (4.9.28). Note that, in fact, from (4.9.31), we get

$$f(\theta^i) - f(\theta^{i+1}) \geq g_i(\theta^i) - g_i(\theta^{i+1}) \geq 0$$

which shows not only that $f(\theta)$ is monotonically decreased at each iteration but also that the decrease in $f(\theta)$ is not smaller than the corresponding decrease of the majorizing function $g_i(\theta)$.

Note that any parameter vector $\theta^{i+1}$ that gives a smaller value of $g_i(\theta)$ than does $g_i(\theta^i)$ will satisfy (4.9.31). Consequently, whenever the minimum point of $g_i(\theta)$ (see (4.9.30)) cannot be derived in closed form, we can think of computing $\theta^{i+1}$ by, for example, performing a few iterations with a gradient-based algorithm initialized at $\theta^i$ and using a line search (to guarantee that $g_i(\theta^{i+1}) \leq g_i(\theta^i)$). A similar observation could be made on the cyclic minimizer in (4.9.25) when the minimization of either $f(x, y^{i-1})$ or $f(x^i, y)$ cannot be done in closed form. The modification of either (4.9.30) or (4.9.25) in this way usually simplifies the computational effort of each iteration, but could slow down the convergence speed of the algorithm by increasing the number of iterations needed to achieve convergence.

An interesting question regarding the two algorithms discussed so far is whether we could obtain the cyclic minimizer by using the majorization principle on a certain majorizing function. In general, it appears difficult or impossible to do so; nor can the majorization technique be obtained as a special case of a cyclic minimizer. Hence, these two iterative minimization techniques appear to have "independent lives."

To draw more parallels between the cyclic minimizer and the majorization technique, we remark on the fact that, in the former, the user has to choose the partitioning of $\theta$ that makes the minimization in, for example, (4.9.25) "easy," whereas in the latter a function $g_i(\theta)$ has to be found that is not only "easy" to minimize but also possesses the essential property (4.9.28). Fortunately for the majorization approach, finding such functions $g_i(\theta)$ is not as hard as it might at first seem. In what follows, we will develop a method for constructing a function $g_i(\theta)$ possessing the desired properties (4.9.27) and (4.9.28) for a *general class* of functions $f(\theta)$ (including the NLS criterion) that are commonly encountered in parameter estimation applications.

### EM Algorithm

The NLS criterion (see (4.3.1)),

$$f(\theta) = \sum_{t=1}^{N} \left| y(t) - \sum_{k=1}^{n} \alpha_k e^{i(\omega_k t + \varphi_k)} \right|^2 \tag{4.9.32}$$

where $\theta$ is defined in (4.9.26), is obtained from the data equation (4.1.1) in which the noise $\{e(t)\}$ is assumed to be circular and white with mean zero and variance $\sigma^2$. Let us also assume that $\{e(t)\}$ is Gaussian distributed; then the probability density function of the data vector $y = [y(1), \ldots, y(N)]^T$, for given $\theta$, is

$$p(y, \theta) = \frac{1}{(\pi\sigma^2)^N} e^{-\frac{f(\theta)}{\sigma^2}} \tag{4.9.33}$$

where $f(\theta)$ is as defined in (4.9.32). The *method of maximum likelihood* (ML) obtains an estimate of $\theta$ by maximizing (4.9.33) (see (B.1.7) in Appendix B) or, equivalently, by minimizing the so-called *negative log-likelihood function*:

$$-\ln p(y,\theta) = \text{constant} + N\ln\sigma^2 + \frac{f(\theta)}{\sigma^2} \tag{4.9.34}$$

Minimizing (4.9.34) with respect to $\theta$ is equivalent to minimizing (4.9.32), which shows that the NLS method is identical to the ML method under the assumption that $\{e(t)\}$ is Gaussian white noise.

The ML is without a doubt the most widely studied method of parameter estimation. In what follows, we assume that this is the method used for parameter estimation and hence that the function we want to minimize with respect to $\theta$ is the negative log-likelihood:

$$\boxed{f(\theta) = -\ln p(y,\theta)} \tag{4.9.35}$$

Our main goal in this subsection is to show how to construct *a majorizing function for the estimation criterion* in (4.9.35) and how the use of *the corresponding majorization technique leads to the expectation-maximization (EM) algorithm* introduced in [Dempster, Laird, and Rubin 1977]. See also [McLachlan and Krishnan 1997] and [Moon 1996] for more recent and detailed accounts on the EM algorithm.

A notation that will be frequently used concerns the expectation with respect to the distribution of a certain random vector—say $z$—which we will denote by $E_z\{\cdot\}$. When the distribution concerned is conditioned on another random vector—say $y$—we will use the notation $E_{z|y}\{\cdot\}$. If we also want to stress the dependence of the distribution (with respect to which the expectation is taken) on a certain parameter vector $\theta$, then we write $E_{z|y,\theta}\{\cdot\}$.

The main result that we will use in the following is *Jensen's inequality*. It asserts that, for any *concave function* $h(x)$, where $x$ is a random vector, the following inequality holds:

$$\boxed{E\{h(x)\} \le h(E\{x\})} \tag{4.9.36}$$

The proof of (4.9.36) is simple. Let $d(x)$ denote the plane tangent to $h(x)$ at the point $E\{x\}$. Then

$$E\{h(x)\} \le E\{d(x)\} = d(E\{x\}) = h(E\{x\}) \tag{4.9.37}$$

which proves (4.9.36). The inequality in (4.9.37) follows from the concavity of $h(x)$, the first equality follows from the fact that $d(x)$ is a linear function of $x$, and the second equality from the fact that $d(x)$ is tangent (and hence equal) to $h(x)$ at the point $E\{x\}$.

**Remark:** We note in passing that, despite its simplicity, Jensen's inequality is a powerful analysis tool. As a simple illustration of this fact, consider a scalar random variable $x$ with a discrete probability distribution:

$$\Pr\{x = x_k\} = p_k, \quad k = 1,\ldots,M$$

Then, using (4.9.36) and the fact that the *logarithm is a concave function*, we obtain (assuming $x_k > 0$)

$$E\{\ln(x)\} = \sum_{k=1}^{M} p_k \ln(x_k) \leq \ln\left[E\{x\}\right] = \ln\left[\sum_{k=1}^{M} p_k x_k\right]$$

or, equivalently,

$$\sum_{k=1}^{M} p_k x_k \geq \prod_{k=1}^{M} x_k^{p_k} \quad (\text{for } x_k > 0 \text{ and } \sum_{k=1}^{M} p_k = 1) \tag{4.9.38}$$

For $p_k = 1/M$, (4.9.38) reduces to the well-known inequality between the arithmetic and geometric means—that is,

$$\frac{1}{M}\sum_{k=1}^{M} x_k \geq \left(\prod_{k=1}^{M} x_k\right)^{1/M}$$

which is so easily obtained in the present framework. ■

After these preparations, we turn our attention to the main question of finding a majorizing function for (4.9.35). Let $z$ be a random vector whose probability density function conditioned on $y$ is completely determined by $\theta$, and let

$$g_i(\theta) = f(\theta^i) - E_{z|y,\theta^i}\left\{\ln\left[\frac{p(y,z,\theta)}{p(y,z,\theta^i)}\right]\right\} \tag{4.9.39}$$

Clearly $g_i(\theta)$ satisfies

$$g_i(\theta^i) = f(\theta^i) \tag{4.9.40}$$

Furthermore, it follows from Jensen's inequality (4.9.36), the concavity of the function $\ln(\cdot)$, and Bayes' rule for conditional probabilities that

$$\begin{aligned} g_i(\theta) &\geq f(\theta^i) - \ln\left[E_{z|y,\theta^i}\left\{\frac{p(y,z,\theta)}{p(y,z,\theta^i)}\right\}\right] \\ &= f(\theta^i) - \ln\left[E_{z|y,\theta^i}\left\{\frac{p(y,z,\theta)}{p(z|y,\theta^i)p(y,\theta^i)}\right\}\right] \\ &= f(\theta^i) - \ln\left[\frac{1}{p(y,\theta^i)}\underbrace{\int p(y,z,\theta)\,dz}_{p(y,\theta)}\right] \\ &= f(\theta^i) - \ln\left[\frac{p(y,\theta)}{p(y,\theta^i)}\right] \\ &= f(\theta^i) + \left[f(\theta) - f(\theta^i)\right] = f(\theta) \end{aligned} \tag{4.9.41}$$

which shows that the function $g_i(\theta)$ in (4.9.39) also satisfies the key majorization condition (4.9.28). Usually, $z$ is called the *unobserved data* (to distinguish it from the observed data vector $y$), and the combination $(z, y)$ is called the *complete data*, while $y$ is called the *incomplete data*.

It follows from (4.9.40) and (4.9.41), along with the discussion in the previous subsection about the majorization approach, that the following algorithm *will monotonically reduce the negative log-likelihood function at each iteration*:

**The Expectation–Maximization (EM) Algorithm**

$\theta^0 =$ given

For $i = 0, 1, 2, \ldots$: (4.9.42)

*Expectation step:* Evaluate $E_{z|y,\theta^i}\{\ln p(y, z, \theta)\} \triangleq \overline{g}_i(\theta)$

*Maximization step:* Compute $\theta^{i+1} = \arg\max_{\theta} \overline{g}_i(\theta)$

This is the *EM algorithm* in a nutshell.

An important aspect of the EM algorithm, which must be considered in every application, is *the choice of the unobserved data vector* $z$. This choice should be done such that the maximization step of (4.9.42) is "easy" or, in any case, much easier than the maximization of the likelihood function. In general, doing so is not an easy task. In addition, the evaluation of the conditional expectation in (4.9.42) might also be rather challenging. Somewhat paradoxically, these difficulties associated with the EM algorithm have perhaps been a cause for its considerable popularity. Indeed, the detailed derivation of the EM algorithm for a particular application is a more challenging research problem (and hence more appealing to many researchers) than, for instance, the derivation of a cyclic minimizer (which also possesses the key property (4.9.24) of the EM algorithm).

### 4.9.6 Frequency-Selective ESPRIT-Based Method

In several applications of spectral analysis, the user is interested only in the components lying in a small frequency band of the spectrum. A *frequency-selective* method deals precisely with this kind of spectral analysis: It estimates the parameters of only those sinusoidal components in the data that lie in a prespecified band of the spectrum, with as little interference as possible from the out-of-band components, and in a computationally efficient way. To be more specific, let us consider the sinusoidal data model in (4.1.1):

$$y(t) = \sum_{k=1}^{\bar{n}} \beta_k e^{i\omega_k t} + e(t); \quad \beta_k = \alpha_k e^{i\varphi_k}, \quad t = 0, \ldots, N-1 \tag{4.9.43}$$

In some applications (see, e.g., [McKelvey and Viberg 2001; Stoica, Sandgren, Selén, Vanhamme, and Van Huffel 2003] and the references therein), it would be computationally too intensive to estimate the parameters of all components in (4.9.43). For instance, this is the case when $\bar{n}$ takes on values close to $N$ or when $\bar{n} \ll N$ but we have many sets of data to process.

In such applications, because of computational and other reasons (see points (i) and (ii) below for details), we focus on only those components of (4.9.43) that are of direct interest to us. Let us assume that the components of interest lie in a prespecified frequency band composed of the following Fourier frequencies:

$$\left\{\frac{2\pi}{N}k_1, \frac{2\pi}{N}k_2, \ldots, \frac{2\pi}{N}k_M\right\} \tag{4.9.44}$$

where $\{k_1, \ldots, k_M\}$ are $M$ given (typically consecutive) integers. We assume that the number of components of (4.9.43) lying in (4.9.44), which we denote by

$$n \leq \bar{n} \tag{4.9.45}$$

is given. If $n$ is *a priori* unknown, then it could be estimated from the data by the methods described in Appendix C.

Our problem is to estimate the parameters of the $n$ components of (4.9.43) that lie in the frequency band in (4.9.44). Furthermore, we want to find a solution to this frequency-selective estimation problem that has the following properties:

(i) *It is computationally efficient.* In particular, the computational complexity of such a solution should be comparable with that of a standard ESPRIT method for a sinusoidal model with $n$ components.

(ii) *It is statistically accurate.* To be more specific about this aspect, we will split the discussion into two parts. From a theoretical standpoint, estimating $n < \bar{n}$ components of (4.9.43) (in the presence of the remaining components and noise) cannot produce more accurate estimates than estimating all $\bar{n}$ components. However, for a good frequency-selective method, the degradation of theoretical statistical accuracy should not be significant. On the other hand, from a practical standpoint, a sound frequency-selective method could give better performance than a non-frequency-selective counterpart that deals with all $\bar{n}$ components of (4.9.43). This is so because some components of (4.9.43) that do not belong to (4.9.44) might not be well-described by a sinusoidal model; consequently, treating such components as interference and eliminating them from the model could improve the estimation accuracy of the components of interest.

In this complement, following [McKelvey and Viberg 2001] and [Stoica, Sandgren, Selén, Vanhamme, and Van Huffel 2003], we present a *frequency-selective ESPRIT-based* (FRES-ESPRIT) method that possesses the previous two desirable features. The following notation will be used frequently in what follows:

$$w_k = e^{i\frac{2\pi}{N}k} \quad k = 0, 1, \ldots, N-1 \tag{4.9.46}$$

$$u_k = [w_k, \ldots, w_k^m]^T \tag{4.9.47}$$

$$v_k = [1, w_k, \ldots, w_k^{N-1}]^T \tag{4.9.48}$$

$$y = [y(0), \ldots, y(N-1)]^T \tag{4.9.49}$$

$$Y_k = v_k^* y \qquad k = 0, 1, \ldots, N-1 \tag{4.9.50}$$

$$e = [e(0), \ldots, e(N-1)]^T \tag{4.9.51}$$

$$E_k = v_k^* e \qquad k = 0, 1, \ldots, N-1 \tag{4.9.52}$$

$$a(\omega_k) = \left[e^{i\omega_k}, \ldots, e^{im\omega_k}\right]^T \tag{4.9.53}$$

$$b(\omega_k) = \left[1, e^{i\omega_k}, \ldots, e^{i(N-1)\omega_k}\right]^T \tag{4.9.54}$$

Hereafter, $m$ is a *user parameter* whose choice will be discussed later on. Note that $\{Y_k\}$ is *the FFT of the data*.

First, we show that the following key equation involving the FFT sequence $\{Y_k\}$ holds true:

$$u_k Y_k = [a(\omega_1), \ldots, a(\omega_{\bar{n}})] \begin{bmatrix} \beta_1 v_k^* b(\omega_1) \\ \vdots \\ \beta_{\bar{n}} v_k^* b(\omega_{\bar{n}}) \end{bmatrix} + \Gamma u_k + u_k E_k \tag{4.9.55}$$

Here $\Gamma$ is an $m \times m$ matrix defined in equation (4.9.61). (It will become clear shortly that the definition of $\Gamma$ has no importance for what follows; hence, it is not repeated here.)

To prove (4.9.55), we first write the data vector $y$ as

$$y = \sum_{\ell=1}^{\bar{n}} \beta_\ell b(\omega_\ell) + e \tag{4.9.56}$$

Next, we note that (for $p = 1, \ldots, m$)

$$\begin{aligned} w_k^p \left[v_k^* b(\omega)\right] &= \sum_{t=0}^{N-1} e^{i\left(\omega - \frac{2\pi}{N}k\right)t} e^{i\frac{2\pi}{N}kp} \\ &= e^{i\omega p} \sum_{t=0}^{N-1} e^{i\left(\omega - \frac{2\pi}{N}k\right)(t-p)} \\ &= e^{i\omega p}\left[v_k^* b(\omega)\right] + e^{i\omega p}\left[\sum_{t=0}^{p-1} e^{i\omega(t-p)} e^{-i\frac{2\pi}{N}k(t-p)} \right. \\ &\qquad \left. - \sum_{t=N}^{N+p-1} e^{i\omega(t-p)} e^{-i\frac{2\pi}{N}k(t-p)}\right] \\ &= e^{i\omega p}\left[v_k^* b(\omega)\right] + e^{i\omega p} \sum_{\ell=1}^{p} \left[e^{-i\omega\ell} e^{i\frac{2\pi}{N}k\ell} - e^{i\omega(N-\ell)} e^{i\frac{2\pi}{N}k\ell}\right] \\ &= e^{i\omega p}\left[v_k^* b(\omega)\right] + \sum_{\ell=1}^{p} e^{i\omega(p-\ell)}\left(1 - e^{i\omega N}\right) w_k^\ell \end{aligned} \tag{4.9.57}$$

Let (for $p = 1, \ldots, m$)

$$\gamma_p^*(\omega) = \left(1 - e^{i\omega N}\right)\left[e^{i\omega(p-1)}, e^{i\omega(p-2)}, \ldots, e^{i\omega}, 1, 0, \ldots, 0\right] \quad (1 \times m) \tag{4.9.58}$$

Using (4.9.58), we can rewrite (4.9.57) in the following more compact form (for $p = 1, \ldots, m$):

$$w_k^p\left[v_k^* b(\omega)\right] = e^{i\omega p}\left[v_k^* b(\omega)\right] + \gamma_p^*(\omega) u_k \tag{4.9.59}$$

or, equivalently,

$$u_k\left[v_k^* b(\omega)\right] = a(\omega)\left[v_k^* b(\omega)\right] + \begin{bmatrix} \gamma_1^*(\omega) \\ \vdots \\ \gamma_m^*(\omega) \end{bmatrix} u_k \tag{4.9.60}$$

From (4.9.56) and (4.9.60), it follows that

$$\begin{aligned} u_k Y_k &= \sum_{\ell=1}^{\bar{n}} \beta_\ell u_k\left[v_k^* b(\omega_\ell)\right] + u_k E_k \\ &= [a(\omega_1), \ldots, a(\omega_{\bar{n}})] \begin{bmatrix} \beta_1 v_k^* b(\omega_1) \\ \vdots \\ \beta_{\bar{n}} v_k^* b(\omega_{\bar{n}}) \end{bmatrix} + \left\{ \sum_{\ell=1}^{\bar{n}} \beta_\ell \begin{bmatrix} \gamma_1^*(\omega_\ell) \\ \vdots \\ \gamma_m^*(\omega_\ell) \end{bmatrix} \right\} u_k + u_k E_k \end{aligned} \tag{4.9.61}$$

which proves (4.9.55).

Next, we let $\{\omega_k\}_{k=1}^n$ denote *the frequencies of interest* (i.e., those frequencies of (4.9.43) that lie in (4.9.44)). To separate the terms in (4.9.55) corresponding to the components of interest from those associated with the nuisance components, we use the notation

$$A = [a(\omega_1), \ldots, a(\omega_n)] \tag{4.9.62}$$

$$x_k = \begin{bmatrix} \beta_1 v_k^* b(\omega_1) \\ \vdots \\ \beta_n v_k^* b(\omega_n) \end{bmatrix} \tag{4.9.63}$$

for the components of interest, and similarly $\tilde{A}$ and $\tilde{x}_k$ for the other components. Finally, to write the equation (4.9.55) for $k = k_1, \ldots, k_M$ in a compact matrix form, we need the additional notation

$$Y = \left[u_{k_1} Y_{k_1}, \ldots, u_{k_M} Y_{k_M}\right], \qquad (m \times M) \tag{4.9.64}$$

$$E = \left[u_{k_1} E_{k_1}, \ldots, u_{k_M} E_{k_M}\right], \qquad (m \times M) \tag{4.9.65}$$

$$U = \left[u_{k_1}, \ldots, u_{k_M}\right], \qquad (m \times M) \tag{4.9.66}$$

$$X = \left[x_{k_1}, \ldots, x_{k_M}\right], \qquad (n \times M) \tag{4.9.67}$$

and similarly for $\tilde{X}$. Using this notation, we can write (4.9.55) (for $k = k_1, \ldots, k_M$) as follows:

$$Y = AX + \Gamma U + \tilde{A}\tilde{X} + E \tag{4.9.68}$$

Next, we assume that

$$\boxed{M \geq n + m} \tag{4.9.69}$$

which can be satisfied by choosing the user parameter $m$ appropriately. Under (4.9.69) (in fact only $M \geq m$ is required for this part), the orthogonal projection matrix onto the null space of $U$ is given by (see Appendix A)

$$\Pi_U^{\perp} = I - U^* \left(UU^*\right)^{-1} U \tag{4.9.70}$$

We will eliminate the second term in (4.9.68) by postmultiplying (4.9.68) with $\Pi_U^{\perp}$. However, before doing so, we make the following observations about the third and fourth terms in (4.9.68):

(a) The elements of the noise term $E$ in (4.9.68) are much smaller than the elements of $AX$. In effect, it can be shown that $E_k = \mathcal{O}\left(N^{1/2}\right)$ (stochastically), whereas the order of the elements of $X$ is typically $\mathcal{O}(N)$.

(b) Assuming that the out-of-band components are not much stronger than the components of interest, and that the frequencies of the former are not too close to the interval of interest in (4.9.44), the elements of $\tilde{X}$ are also much smaller than the elements of $X$.

(c) To understand what happens in the case that the assumption made in (b) does not hold, let us consider a generic out-of-band component $(\omega, \beta)$. The part of $y$ corresponding to this component can be written as $\beta b(\omega)$. Hence, the corresponding part in $u_k Y_k$ is given by $\beta u_k \left[v_k^* b(\omega)\right]$; consequently, the part of $Y$ due to this generic component is

$$\beta U \begin{bmatrix} v_{k_1}^* b(\omega) & & 0 \\ & \ddots & \\ 0 & & v_{k_M}^* b(\omega) \end{bmatrix} \tag{4.9.71}$$

Even if $\omega$ is relatively close to the band of interest, (4.9.44), we may expect that $v_k^* b(\omega)$ does not vary significantly for $k \in [k_1, k_M]$ (in other words, the "spectral tail" of the out-of-band component could well have a small dynamic range in the interval of interest). As a consequence, the matrix in (4.9.71) will be approximately proportional to $U$ and hence it will be attenuated via the postmultiplication of it by $\Pi_U^{\perp}$ (see below). A similar argument shows that the noise term in (4.9.68) is also attenuated by postmultiplying (4.9.68) with $\Pi_U^{\perp}$.

It follows from the previous discussion and (4.9.68) that

$$Y \Pi_U^{\perp} \simeq AX \Pi_U^{\perp} \tag{4.9.72}$$

This equation resembles equation (4.7.7), on which the standard ESPRIT method is based, provided that

$$\operatorname{rank}\left(X\Pi_U^{\perp}\right) = n \tag{4.9.73}$$

(similarly to $\operatorname{rank}(C) = n$ for (4.7.7)). In the following, we prove that (4.9.73) *holds under* (4.9.69) *and the regularity condition that* $e^{iN\omega_k} \neq 1$ (for $k = 1, \ldots, n$).

To prove (4.9.73), we first note that $\operatorname{rank}\left(\Pi_U^{\perp}\right) = M - m$, which implies that $M \geq m + n$ (i.e., (4.9.69)) is a necessary condition for (4.9.73) to hold.

Next we show that (4.9.73) is equivalent to

$$\operatorname{rank}\left(\begin{bmatrix} X \\ U \end{bmatrix}\right) = m + n \tag{4.9.74}$$

To verify this equivalence, let us decompose X additively as

$$X = X\Pi_U + X\Pi_U^{\perp} = XU^*\left(UU^*\right)^{-1}U + XV^*V \tag{4.9.75}$$

where the $M \times (M - m)$ matrix $V^*$ comprises a unitary basis of $\mathcal{N}(U)$; hence, $UV^* = 0$ and $VV^* = I$. Now, the matrix in (4.9.74) has the same rank as

$$\begin{bmatrix} I & -XU^*(UU^*)^{-1} \\ 0 & I \end{bmatrix}\begin{bmatrix} X \\ U \end{bmatrix} = \begin{bmatrix} XV^*V \\ U \end{bmatrix} \tag{4.9.76}$$

(we used (4.9.75) to obtain (4.9.76)), which, in turn, has the same rank as

$$\begin{bmatrix} XV^*V \\ U \end{bmatrix}\begin{bmatrix} V^*VX^* & U^* \end{bmatrix} = \begin{bmatrix} XV^*VX^* & 0 \\ 0 & UU^* \end{bmatrix} \tag{4.9.77}$$

However, $\operatorname{rank}(UU^*) = m$. Thus, (4.9.74) holds if and only if

$$\operatorname{rank}(XV^*VX^*) = n$$

As

$$\operatorname{rank}(XV^*VX^*) = \operatorname{rank}(X\Pi_U^{\perp}X^*) = \operatorname{rank}(X\Pi_U^{\perp})$$

the equivalence between (4.9.73) and (4.9.74) is proven.

It follows from this equivalence and the definition of $X$ and $U$ that we want to prove that

$$\operatorname{rank}\left\{\underbrace{\begin{bmatrix} v_{k_1}^* b(\omega_1) & \cdots & v_{k_M}^* b(\omega_1) \\ \vdots & & \vdots \\ v_{k_1}^* b(\omega_n) & \cdots & v_{k_M}^* b(\omega_n) \\ u_{k_1} & \cdots & u_{k_M} \end{bmatrix}}_{(n+m)\times M}\right\} = n + m \tag{4.9.78}$$

Now,

$$v_k^* b(\omega) = \sum_{t=0}^{N-1} e^{i\left(\omega - \frac{2\pi}{N}k\right)t} = \frac{1 - e^{iN\left(\omega - \frac{2\pi}{N}k\right)}}{1 - e^{i\left(\omega - \frac{2\pi}{N}k\right)}} = \frac{1 - e^{iN\omega}}{w_k - e^{i\omega}} w_k$$

so we can rewrite the matrix in (4.9.78) as follows:

$$\begin{bmatrix} 1 - e^{iN\omega_1} & & & & 0 \\ & \ddots & & & \\ & & 1 - e^{iN\omega_n} & & \\ & & & 1 & \\ & & & & \ddots \\ 0 & & & & 1 \end{bmatrix} \begin{bmatrix} \frac{w_{k_1}}{w_{k_1} - e^{i\omega_1}} & \cdots & \frac{w_{k_M}}{w_{k_M} - e^{i\omega_1}} \\ \vdots & & \vdots \\ \frac{w_{k_1}}{w_{k_1} - e^{i\omega_n}} & \cdots & \frac{w_{k_M}}{w_{k_M} - e^{i\omega_n}} \\ w_{k_1} & \cdots & w_{k_M} \\ \vdots & & \vdots \\ w_{k_1}^m & \cdots & w_{k_M}^m \end{bmatrix} \tag{4.9.79}$$

Because, by assumption, $1 - e^{iN\omega_k} \neq 0$ (for $k = 1, \ldots, n$), it follows that (4.9.78) holds if and only if the second matrix in (4.9.79) has full row rank (under (4.9.69)), which holds true if and only if we cannot find some numbers $\{\rho_k\}_{k=1}^{m+n}$ (not all zero) such that

$$\begin{aligned} &\frac{\rho_1 z}{z - e^{i\omega_1}} + \cdots + \frac{\rho_n z}{z - e^{i\omega_n}} + \rho_{n+1} z + \cdots + \rho_{n+m} z^m \\ &\qquad = z\left(\frac{\rho_1}{z - e^{i\omega_1}} + \cdots + \frac{\rho_n}{z - e^{i\omega_n}} + \rho_{n+1} + \cdots + \rho_{n+m} z^{m-1}\right) \end{aligned} \tag{4.9.80}$$

is equal to zero at $z = w_{k_1}, \ldots, z = w_{k_M}$. However, (4.9.80) can have at most $m + n - 1$ zeroes of this form, and $m + n - 1 < M$ from (4.9.69). With this observation, the proof of (4.9.73) is concluded.

To make use of (4.9.72) and (4.9.73) in an ESPRIT-like approach, we also assume that

$$\boxed{m \geq n} \tag{4.9.81}$$

(which is an easily satisfied condition); then it follows from (4.9.72) and (4.9.73) that the effective rank of the "data" matrix $Y\Pi_U^{\perp}$ is $n$, and that

$$\boxed{\hat{S} \simeq A\hat{C}} \tag{4.9.82}$$

where $\hat{C}$ is an $n \times n$ nonsingular transformation matrix, and

$$\boxed{\hat{S} = \text{the } m \times n \text{ matrix whose columns are the left singular vectors of } Y\Pi_U^{\perp} \text{ associated with the } n \text{ largest singular values.}} \tag{4.9.83}$$

Equation (4.9.82) is very similar to (4.7.7); hence, it can be used in *an ESPRIT-like approach to estimate the frequencies* $\{\omega_k\}_{k=1}^n$. After the frequency-estimation step, the amplitudes $\{\beta_k\}_{k=1}^n$ can be estimated, for instance, as described in [McKelvey and Viberg 2001; Stoica, Sandgren, Selén, Vanhamme, and Van Huffel 2003].

An implementation detail that we would like to address, at least briefly, is the choice of $m$. We recommend choosing $m$ as the integer part of $M/2$; that is,

$$\boxed{m = \lfloor M/2 \rfloor} \tag{4.9.84}$$

provided that $\lfloor M/2 \rfloor \in [n, M-n]$, to satisfy the assumptions in (4.9.69) and (4.9.81). To motivate this choice of $m$, we refer to the matrix equation (4.9.72) that lies at the basis of the proposed estimation approach. Previous experience with ESPRIT, MUSIC, and other, similar approaches has shown that their accuracy increases as the number of *independent* equations in (4.9.72) (and its counterparts) increases. The matrix $Y\Pi_U^{\perp}$ in (4.9.72) is $m \times M$, and its rank is generically equal to

$$\min\{\text{rank}(Y),\ \text{rank}(\Pi_U^{\perp})\} = \min(m, M-m) \tag{4.9.85}$$

Evidently, this rank determines the aforementioned number of linearly independent equations in (4.9.72). Hence, for enhanced estimation accuracy, we should maximize (4.9.85) with respect to $m$: the solution is clearly given by (4.9.84).

To end this complement, we show that *the proposed FRES-ESPRIT method with* $M = N$ *is equivalent to the standard ESPRIT method*. For $M = N$, we have that

$$[b_1, \ldots, b_N] \triangleq \begin{bmatrix} w_1 & \cdots & w_N \\ w_1^2 & \cdots & w_N^2 \\ \vdots & & \vdots \\ w_1^N & \cdots & w_N^N \end{bmatrix} = \underbrace{\begin{bmatrix} U \\ \bar{U} \end{bmatrix}}_{N} \begin{matrix} \} m \\ \} N-m \end{matrix} \tag{4.9.86}$$

where $U$ is as defined before (with $M = N$) and $\bar{U}$ is defined via (4.9.86). Note that

$$UU^* = NI; \qquad \bar{U}\bar{U}^* = NI; \qquad U\bar{U}^* = 0; \qquad U^*U + \bar{U}^*\bar{U} = NI \tag{4.9.87}$$

Hence,

$$\Pi_U^{\perp} = I - \frac{1}{N} U^* U = \frac{1}{N} \bar{U}^* \bar{U} \tag{4.9.88}$$

Also, note that (for $p = 1, \ldots, m$)

$$\begin{aligned} w_k^p Y_k &= \sum_{t=0}^{N-1} y(t) e^{-i\frac{2\pi}{N}k(t-p)} \\ &= \sum_{t=0}^{p-1} y(t) w_k^{p-t} + \sum_{t=p}^{N-1} y(t) w_k^{N+p-t} \\ &= \left[ y(p-1), \ldots, y(0), 0, \ldots, 0 \right] \begin{bmatrix} w_k \\ \vdots \\ w_k^m \end{bmatrix} + \left[ 0, \ldots, 0, y(N-1), \ldots, y(p) \right] \begin{bmatrix} w_k \\ \vdots \\ w_k^N \end{bmatrix} \\ &\triangleq \mu_p^* u_k + \psi_p^* b_k \end{aligned} \tag{4.9.89}$$

where $u_k$ and $b_k$ are as defined before (see (4.9.47) and (4.9.86)). Consequently, for $M = N$, the "data" matrix $Y \Pi_U^{\perp}$ used in the FRES-ESPRIT method can be written as (*cf.* (4.9.86)–(4.9.89))

$$\begin{aligned} [u_1 Y_1, \ldots, u_N Y_N] \, \Pi_U^{\perp} &= \left\{ \begin{bmatrix} \mu_1^* \\ \vdots \\ \mu_m^* \end{bmatrix} [u_1, \ldots, u_N] + \begin{bmatrix} \psi_1^* \\ \vdots \\ \psi_m^* \end{bmatrix} [b_1, \ldots, b_N] \right\} \bar{U}^* \bar{U} \cdot \frac{1}{N} \\ &= \left\{ \begin{bmatrix} \mu_1^* \\ \vdots \\ \mu_m^* \end{bmatrix} U + \begin{bmatrix} \psi_1^* \\ \vdots \\ \psi_m^* \end{bmatrix} \begin{bmatrix} U \\ \bar{U} \end{bmatrix} \right\} \bar{U}^* \bar{U} \cdot \frac{1}{N} \\ &= \begin{bmatrix} \psi_1^* \\ \vdots \\ \psi_m^* \end{bmatrix} \begin{bmatrix} 0 \\ \bar{U} \end{bmatrix} = \begin{bmatrix} y(N-m) & \cdots & y(1) \\ y(N-m+1) & \cdots & y(2) \\ \vdots & & \vdots \\ y(N-1) & \cdots & y(m) \end{bmatrix} \bar{U} \end{aligned} \tag{4.9.90}$$

It follows from (4.9.90) that the $n$ principal (or dominant) left singular vectors of $Y \Pi_U^{\perp}$ are equal to the $n$ principal eigenvectors of the following matrix (obtained by postmultiplying the

right-hand side of (4.9.90) with its conjugate transpose and using the fact that $\bar{U}\bar{U}^* = NI$ from (4.9.87)):

$$\begin{bmatrix} y(N-m) & \cdots & y(1) \\ \vdots & & \vdots \\ y(N-1) & \cdots & y(m) \end{bmatrix} \begin{bmatrix} y^*(N-m) & \cdots & y^*(N-1) \\ \vdots & & \vdots \\ y^*(1) & \cdots & y^*(m) \end{bmatrix}$$

$$= \sum_{t=1}^{N-m} \begin{bmatrix} y(t) \\ \vdots \\ y(t+m-1) \end{bmatrix} \left[ y^*(t), \ldots, y^*(t+m-1) \right] \tag{4.9.91}$$

which is precisely the type of sample covariance matrix used in the standard ESPRIT method. (Compare with (4.5.14); the difference between (4.9.91) and (4.5.14) is due to some notational changes made in this complement, such as in the definition of the matrix $A$.)

### 4.9.7 A Useful Result for Two-Dimensional (2D) Sinusoidal Signals

For a noise-free 1D sinusoidal signal

$$y(t) = \sum_{k=1}^{n} \beta_k e^{i\omega_k t}, \qquad t = 0, 1, 2, \ldots \tag{4.9.92}$$

a data vector of length $m$ can be written as

$$\begin{bmatrix} y(0) \\ y(1) \\ \vdots \\ y(m-1) \end{bmatrix} = \begin{bmatrix} 1 & \cdots & 1 \\ e^{i\omega_1} & \cdots & e^{i\omega_n} \\ \vdots & & \vdots \\ e^{i(m-1)\omega_1} & \cdots & e^{i(m-1)\omega_n} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \vdots \\ \beta_n \end{bmatrix} \triangleq A\beta \tag{4.9.93}$$

The matrix $A$ just introduced is the complex conjugate of the one in (4.2.4). In this complement, we prefer to work with the type of $A$ matrix in (4.9.93), to simplify the notation, but note that the discussion which follows applies without change to the complex conjugate of the above $A$ as well (or, to its extension to 2D sinusoidal signals).

Let $\{c_k\}_{k=1}^n$ be defined uniquely via the equation

$$1 + c_1 z + \cdots + c_n z^n = \prod_{k=1}^{n} \left(1 - z e^{-i\omega_k}\right) \tag{4.9.94}$$

Then, it can be readily checked (see (4.5.21)) that the matrix

$$C^* = \begin{bmatrix} 1 & c_1 & \cdots & c_n & 0 \\ & \ddots & \ddots & & \ddots \\ 0 & & 1 & c_1 & \cdots & c_n \end{bmatrix}, \qquad (m-n) \times m \tag{4.9.95}$$

satisfies

$$C^*A = 0 \tag{4.9.96}$$

(To verify (4.9.96), it is enough to observe, from (4.9.94), that $1 + c_1 e^{i\omega_k} + \cdots + c_n e^{in\omega_k} = 0$ for $k = 1, \ldots, n$.) Furthermore, as $\text{rank}(C) = m - n$ and $\dim[\mathcal{N}(A^*)] = m - n$ too, it follows from (4.9.96) that

$$\boxed{C \text{ is a basis for the null space of } A^*,\ \mathcal{N}(A^*)} \tag{4.9.97}$$

The matrix $C$ plays an important role in the derivation and analysis of several frequency estimators. (See, e.g., Section 4.5, [Bresler and Macovski 1986], and [Stoica and Sharman 1990].)

In this complement, *we will extend the result* (4.9.97) *to 2D sinusoidal signals*. The derivation of a result similar to (4.9.97) for such signals is a rather more difficult problem than in the 1D case. The solution that we will present was introduced in [Clark and Scharf 1994]. (See also [Clark, Eldén, and Stoica 1997].) Using the extended result, we can derive parameter estimation methods for 2D sinusoidal signals in much the same manner as for 1D signals. (See the cited papers and Section 4.5.)

A noise-free 2D sinusoidal signal is described by the following equation (compare with (4.9.92)):

$$y(t, \bar{t}) = \sum_{k=1}^{n} \beta_k e^{i\omega_k t} e^{i\bar{\omega}_k \bar{t}}, \qquad t, \bar{t} = 0, 1, 2, \ldots \tag{4.9.98}$$

Let

$$\gamma_k = e^{i\omega_k}, \quad \lambda_k = e^{i\bar{\omega}_k} \tag{4.9.99}$$

Using this notation allows us to write (4.9.98) in the more compact form

$$y(t, \bar{t}) = \sum_{k=1}^{n} \beta_k \gamma_k^t \lambda_k^{\bar{t}} \tag{4.9.100}$$

Moreover, equation (4.9.100) (unlike (4.9.98)) also covers the case of *damped (2D) sinusoidal signals*, for which

$$\gamma_k = e^{\mu_k + i\omega_k}, \quad \lambda_k = e^{\bar{\mu}_k + i\bar{\omega}_k} \tag{4.9.101}$$

with $\{\mu_k, \bar{\mu}_k\}$ being the damping parameters ($\mu_k, \bar{\mu}_k \leq 0$).

The following notation will be used frequently in this complement:

$$g_t^* = \begin{bmatrix} \gamma_1^t & \cdots & \gamma_n^t \end{bmatrix} \tag{4.9.102}$$

$$\Gamma = \begin{bmatrix} \gamma_1 & & 0 \\ & \ddots & \\ 0 & & \gamma_n \end{bmatrix} \tag{4.9.103}$$

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \ddots & \\ 0 & & \lambda_n \end{bmatrix} \tag{4.9.104}$$

$$\beta = \begin{bmatrix} \beta_1 & \dots & \beta_n \end{bmatrix}^T \tag{4.9.105}$$

$$A_L = \begin{bmatrix} 1 & \dots & 1 \\ \lambda_1 & \dots & \lambda_n \\ \vdots & & \vdots \\ \lambda_1^{L-1} & \dots & \lambda_n^{L-1} \end{bmatrix} \quad \text{for } L \geq n \tag{4.9.106}$$

Using (4.9.102), (4.9.104), and (4.9.105), we can write

$$y(t, \bar{t}) = g_t^* \Lambda^{\bar{t}} \beta \tag{4.9.107}$$

Hence, similarly to (4.9.93), we can write the $m\bar{m} \times 1$ data vector obtained from (4.9.98) for $t = 0, \dots, m-1$ and $\bar{t} = 0, \dots, \bar{m} - 1$ as

$$\begin{bmatrix} y(0,0) \\ \vdots \\ y(0, \bar{m}-1) \\ \cdots\cdots\cdots\cdots \\ \vdots \\ \cdots\cdots\cdots\cdots \\ y(m-1, 0) \\ \vdots \\ y(m-1, \bar{m}-1) \end{bmatrix} = \begin{bmatrix} g_0^* \Lambda^0 \\ \vdots \\ g_0^* \Lambda^{\bar{m}-1} \\ \cdots\cdots\cdots \\ \vdots \\ \cdots\cdots\cdots \\ g_{m-1}^* \Lambda^0 \\ \vdots \\ g_{m-1}^* \Lambda^{\bar{m}-1} \end{bmatrix} \beta \triangleq \mathcal{A}\beta \tag{4.9.108}$$

The matrix $\mathcal{A}$ just defined,

$$\mathcal{A} = \begin{bmatrix} g_0^* \Lambda^0 \\ \vdots \\ g_0^* \Lambda^{\bar{m}-1} \\ \cdots\cdots\cdots \\ \vdots \\ \cdots\cdots\cdots \\ g_{m-1}^* \Lambda^0 \\ \vdots \\ g_{m-1}^* \Lambda^{\bar{m}-1} \end{bmatrix} \qquad (m\bar{m} \times n) \tag{4.9.109}$$

plays the same role for 2D sinusoidal signals as the matrix $A$ in (4.9.93) does for 1D signals. Therefore, it is the null space of (4.9.109) that we want to characterize. More precisely, we want to find a *linearly parameterized basis* for the null space of the matrix $\mathcal{A}^*$ in (4.9.109), similar to the basis $C$ for $A^*$ in (4.9.93). (See (4.9.97).)

Note that, using (4.9.103), we can also write $y(t, \bar{t})$ as

$$y(t, \bar{t}) = \begin{bmatrix} \lambda_1^{\bar{t}} & \dots & \lambda_n^{\bar{t}} \end{bmatrix} \Gamma^t \beta \tag{4.9.110}$$

This means that $\mathcal{A}$ can also be written as follows:

$$\mathcal{A} = \begin{bmatrix} A_{\bar{m}} \Gamma^0 \\ \cdots\cdots\cdots \\ \vdots \\ \cdots\cdots\cdots \\ A_{\bar{m}} \Gamma^{m-1} \end{bmatrix} \tag{4.9.111}$$

Similarly to (4.9.94), let us define the parameters $\{c_k\}_{k=1}^n$ uniquely via the equation

$$\boxed{1 + c_1 z + \cdots + c_n z^n = \prod_{k=1}^{n} \left(1 - \frac{z}{\lambda_k}\right)} \tag{4.9.112}$$

Note that there is a *one-to-one mapping between* $\{c_k\}$ *and* $\{\lambda_k\}$ ($\lambda_k \neq 0$). In particular, we can obtain $\{\lambda_k\}$ uniquely from $\{c_k\}$. (See [Stoica and Sharman 1990] for more details on this aspect in the case of $\{\lambda_k = e^{i\omega_k}\}$.) Consequently, we can see the introduction of $\{c_k\}$ as a new parameterization of the problem, which replaces the parameterization via $\{\lambda_k\}$. Using $\{c_k\}$, we build the following matrix, similarly to (4.9.95), assuming $\bar{m} > n$:

$$C^* = \begin{bmatrix} 1 & c_1 & \cdots & c_n & & 0 \\ & \ddots & \ddots & & \ddots & \\ 0 & & 1 & c_1 & \cdots & c_n \end{bmatrix}, \qquad (\bar{m} - n) \times \bar{m} \tag{4.9.113}$$

We note (*cf.* (4.9.96)) that

$$C^* A_{\bar{m}} = 0 \tag{4.9.114}$$

It follows from (4.9.111) and (4.9.114) that

$$\underbrace{\begin{bmatrix} C^* & & 0 \\ & \ddots & \\ 0 & & C^* \end{bmatrix}}_{[m(\bar{m}-n)] \times m\bar{m}} \mathcal{A} = 0 \tag{4.9.115}$$

Hence, we have found $(m\bar{m} - mn)$ vectors of the sought basis for $\mathcal{N}(\mathcal{A}^*)$. It remains to find $(m-1)n$ additional (linearly independent) vectors of this basis (note that $\dim[\mathcal{N}(\mathcal{A}^*)] = m\bar{m} - n$). To find the remaining vectors, we need an approach that is rather different from that used so far.

Let us assume that

$$\boxed{\lambda_k \neq \lambda_p \text{ for } k \neq p} \tag{4.9.116}$$

and let the vector

$$b^* = [b_1, \ldots, b_n]$$

be defined via the linear (interpolation) equation

$$\boxed{b^* A_n = \left[\gamma_1, \ldots, \gamma_n\right]} \tag{4.9.117}$$

(with $A_n$ as defined in (4.9.106)). *Under* (4.9.116) *and for given* $\{\lambda_k\}$*, there exists a one-to-one map between* $\{b_k\}$ *and* $\{\gamma_k\}$; hence, we can view the use of $\{b_k\}$ as a reparameterization of the problem. (Note that, if (4.9.116) does not hold, i.e., $\lambda_k = \lambda_p$, then, for identifiability reasons, we must have $\gamma_k \neq \gamma_p$, and therefore no vector $b$ that satisfies (4.9.117) can exist.) From (4.9.117), we easily obtain

$$b^* A_n \Gamma^t = \left[\gamma_1, \ldots, \gamma_n\right] \Gamma^t = g_{t+1}^*$$

and hence (see also (4.9.109) and (4.9.111))

$$b^* \begin{bmatrix} g_t^* \Lambda^0 \\ \vdots \\ g_t^* \Lambda^{n-1} \end{bmatrix} = b^* A_n \Gamma^t = g_{t+1}^* \Lambda^0 \tag{4.9.118}$$

Next, we assume that

$$\bar{m} \geq 2n - 1 \tag{4.9.119}$$

which is a weak condition (typically we have $m, \bar{m} \gg n$). Under (4.9.119), we can write (making use of (4.9.118)):

$$\underbrace{\begin{bmatrix} b^* & & 0 \\ & \ddots & \\ 0 & & b^* \end{bmatrix}}_{B^*} \begin{bmatrix} g_t^* \Lambda^0 \\ \vdots \\ g_t^* \Lambda^{\bar{m}-1} \end{bmatrix} - \begin{bmatrix} g_{t+1}^* \Lambda^0 \\ \vdots \\ g_{t+1}^* \Lambda^{n-1} \end{bmatrix} = 0 \tag{4.9.120}$$

where

$$B^* = \begin{bmatrix} b_1 & b_2 & \dots & b_n & 0 & \dots & 0 \\ & \ddots & & & \ddots & \ddots & \vdots \\ 0 & & b_1 & b_2 & \dots & b_n & 0 \end{bmatrix} \qquad (n \times \bar{m})$$

Note that, indeed, we need $\bar{m} \geq 2n - 1$ to be able to write (4.9.120) (if $\bar{m} > 2n - 1$, then the rightmost $\bar{m} - 2n - 1$ columns of $B^*$ are zeroes). Combining (4.9.115) and (4.9.120) yields the following matrix, whose rows lie in the left null space of $\mathcal{A}$:

$$\left.\begin{bmatrix} \mathcal{D} & \mathcal{I} & & & \\ & \mathcal{D} & \mathcal{I} & 0 & \\ & & \ddots & \ddots & \\ 0 & & & \mathcal{D} & \mathcal{I} \\ & & & & C^* \end{bmatrix}\right\} \begin{matrix} m \\ \text{block rows} \end{matrix} \qquad (4.9.121)$$

where

$$\mathcal{D} = \begin{bmatrix} C^* \\ B^* \end{bmatrix} = \left[\begin{array}{ccccccc} 1 & c_1 & \cdots & c_n & & 0 & \\ & \ddots & \ddots & & \ddots & & \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} \bar{m} - n \\ 0 & & 1 & c_1 & \cdots & c_n & \\ \hline b_1 & \dots & b_n & 0 & \dots & 0 & \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} n \\ & \ddots & & & \ddots & \ddots & \vdots \\ 0 & & b_1 & \dots & b_n & 0 & \end{array}\right] \qquad (\bar{m} \times \bar{m})$$

$$\mathcal{I} = \left[\begin{array}{ccccc} 0 & & \cdots & & 0 \\ \vdots & & & & \vdots \\ 0 & & \cdots & & 0 \\ \hline -1 & 0 & & \dots & 0 \\ & \ddots & \ddots & & \vdots \\ 0 & & -1 & 0 \dots & 0 \end{array}\right] \begin{matrix} \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} \bar{m} - n \\ \\ \left.\vphantom{\begin{matrix}a\\a\\a\end{matrix}}\right\} n \end{matrix} \qquad (\bar{m} \times \bar{m})$$

The matrix in (4.9.121) is of dimension $[(m-1)\bar{m} + (\bar{m} - n)] \times m\bar{m}$, that is $(m\bar{m} - n) \times m\bar{m}$, and its rank is equal to $m\bar{m} - n$ (i.e., it has full row rank, as $c_n \neq 0$). Consequently, *the rows of* (4.9.121) *form a linearly parameterized basis for the null space of* $\mathcal{A}$. We remind the reader that, under (4.9.116), there is a one-to-one map between $\{\lambda_k, \gamma_k\}$ and the basis parameters $\{c_k, b_k\}$. (See (4.9.112) and (4.9.117).) Hence, we can think of estimating $\{c_k, b_k\}$ in lieu of $\{\lambda_k, \gamma_k\}$, at least in a first stage, and, when we do so, the linear dependence of (4.9.121) on the unknown parameters will come in quite handy. As a simple example of such an estimation method based on (4.9.121), note that the modified MUSIC procedure outlined in Section 4.5 can easily be extended to the case of 2D signals by making use of (4.9.121).

Compared with the basis matrix for the 1D case (see (4.9.95)), the null space basis (4.9.121) in the 2D case is apparently much more complicated. In addition, the given 2D basis result depends on the condition (4.9.116); if (4.9.116) is even approximately violated (i.e., if there exist $\lambda_k$ and $\lambda_p$ with $k \neq p$ such that $\lambda_k \simeq \lambda_p$), then the mapping $\{\gamma_k\} \leftrightarrow \{b_k\}$ could become ill-conditioned and so cause a deterioration of the estimation accuracy.

Finally, we remark on the fact that, for damped sinusoids, the parameterization via $\{b_k\}$ and $\{c_k\}$ is parsimonious. However, for undamped sinusoidal signals, the parameterization via $\{\omega_k, \bar{\omega}_k\}$ contains $2n$ real-valued unknowns, whereas the one based on $\{b_k, c_k\}$ has $4n$ unknowns, or $3n$ unknowns if a certain conjugate symmetry property of $\{b_k\}$ is exploited (see, e.g., [Stoica and Sharman 1990]); hence, in such a case the use of $\{b_k\}$ and, in particular, $\{c_k\}$ leads to an overparameterized problem, which might also result in a (slight) accuracy degradation. The previous criticism of the result (4.9.121) is, however, minor, and, in fact, (4.9.121) is the *only* known basis for $\mathcal{N}(\mathcal{A}^*)$.

## 4.10 EXERCISES

**Exercise 4.1: Speed Measurement by a Doppler Radar as a Frequency Estimation Problem** Assume that a radar system transmits a sinusoidal signal towards an object. For the sake of simplicity, further assume that the object moves along a trajectory parallel to the wave propagation direction, at a constant velocity $v$. Let $\alpha e^{i\omega t}$ denote the signal emitted by the radar. Show that the backscattered signal, measured by the radar system after reflection off the object, is given by

$$s(t) = \beta e^{i(\omega - \omega^D)t} + e(t) \tag{4.10.1}$$

where $e(t)$ is measurement noise, $\omega^D$ is the so-called *Doppler frequency*,

$$\omega^D \triangleq 2\omega v/c$$

and

$$\beta = \mu\alpha e^{-2i\omega r/c}$$

Here $c$ denotes the speed of wave propagation, $r$ is the object range, and $\mu$ is an attenuation coefficient.

Conclude from (4.10.1) that the problem of speed measurement can be reduced to one of frequency determination. The latter problem can be solved by using the methods of this chapter.

**Exercise 4.2: ACS of Sinusoids with Random Amplitudes or Nonuniform Phases** In some applications, it is not reasonable to assume that the amplitudes of the sinusoidal terms are fixed or that their phases are uniformly distributed. Examples are fast fading in mobile telecommunications (where the amplitudes vary), and sinusoids that have been tracked so that their phase is random, near zero, but not uniformly distributed. We derive the ACS for such cases.

Let $x(t) = \alpha e^{i(\omega_0 t + \varphi)}$, where $\alpha$ and $\varphi$ are statistically independent random variables and $\omega_0$ is a constant. Assume that $\alpha$ has mean $\bar{\alpha}$ and variance $\sigma_\alpha^2$.

**(a)** If $\varphi$ is uniformly distributed on $[-\pi, \pi]$, find $E\{x(t)\}$ and $r_x(k)$. Show also that, if $\alpha$ is constant, the expression for $r_x(k)$ reduces to equation (4.1.5).

**(b)** If $\varphi$ is not uniformly distributed on $[-\pi, \pi]$, express $E\{x(t)\}$ in terms of the probability density function $p(\varphi)$. Find sufficient conditions on $p(\varphi)$ such that $x(t)$ is zero mean, find $r_x(k)$ in this case, and give an example of such a $p(\varphi)$.

**Exercise 4.3: A Nonergodic Sinusoidal Signal**
As shown in Complement 4.9.1, the signal

$$x(t) = \alpha e^{i(\omega t + \varphi)}$$

with $\alpha$ and $\omega$ being nonrandom constants and $\varphi$ being uniformly distributed on $[0,\ 2\pi]$, is second-order ergodic in the sense that the mean and covariances determined from an (infinitely long) temporal realization of the signal coincide with the mean and covariances obtained from an ensemble of (infinitely many) realizations. In the present exercise, assume that $\alpha$ and $\omega$ are independent random variables, with $\omega$ being uniformly distributed on $[0,\ 2\pi]$; the initial-phase variable $\varphi$ may be arbitrarily distributed (in particular it can be nonrandom). Show that, in such a case,

$$E\left\{x(t)x^*(t-k)\right\} = \begin{cases} E\left\{\alpha^2\right\} & \text{for } k = 0 \\ 0 & \text{for } k \neq 0 \end{cases} \tag{4.10.2}$$

Also, show that the covariances obtained by "temporal averaging" differ from those given, and hence deduce that the signal is not ergodic. Comment on the behavior of such a signal over the ensemble of realizations and in each realization, respectively.

**Exercise 4.4: AR Model-Based Frequency Estimation**
Consider the noisy sinusoidal signal

$$y(t) = x(t) + e(t)$$

where $x(t) = \alpha e^{i(\omega_0 t + \varphi)}$ (with $\alpha > 0$ and $\varphi$ uniformly distributed on $[0,\ 2\pi]$) and $e(t)$ is white noise with zero mean and unit variance. An AR model of order $n \geq 1$ is fitted to $\{y(t)\}$ by using the Yule–Walker or LS method. In the limiting case of an infinitely long data sample, the AR coefficients are given by the solution to (3.4.4). Show that the PSD, corresponding to the AR model determined from (3.4.4), has a global peak at $\omega = \omega_0$. Conclude that AR modeling can be used in this case to find the sinusoidal frequency, in spite of the fact that $\{y(t)\}$ does not satisfy an AR equation of finite order. (In the case of multiple sinusoids, the AR frequency estimates are biased.) Regarding the estimation of the signal power, however, show that the height of the global peak of the AR spectrum does not directly provide an "estimate" of $\alpha^2$.

**Exercise 4.5: An ARMA Model-Based Derivation of the Pisarenko Method**
Let $R$ denote the covariance matrix (4.2.7) with $m = n + 1$, and let $g$ be the eigenvector of $R$ associated with its minimum eigenvalue. The Pisarenko method determines the signal frequencies by exploiting the fact that

$$a^*(\omega)g = 0 \qquad \text{for } \omega = \omega_k,\ k = 1, \ldots, n \tag{4.10.3}$$

(*cf.* (4.5.13) and (4.5.17)). Derive the property (4.10.3) directly from the ARMA model equation (4.2.3).

**Exercise 4.6: Frequency Estimation when Some Frequencies Are Known**
Assume that $y(t)$ is known to have $p$ sinusoidal components at known frequencies $\{\tilde{\omega}_k\}_{k=1}^{p}$ (but with unknown amplitudes and phases), plus $n-p$ other sinusoidal components whose frequencies are unknown. Develop a modification of the HOYW method to estimate the unknown frequencies from measurements $\{y(t)\}_{t=1}^{N}$ without estimating the known frequencies.

**Exercise 4.7: A Combined HOYW–ESPRIT Method for the MA Noise Case**
The HOYW method, presented in Section 4.4 for the white-noise case, is based on the matrix $\Gamma$ in (4.2.8). Let us assume that the noise sequence $\{e(t)\}$ in (4.1.1) is known to be an MA process of order $m$ and that $m$ is given. A simple way to handle such a colored noise in the HOYW method consists of modifying the expression (4.2.8) of $\Gamma$ as follows:

$$\tilde{\Gamma} = E\left\{\begin{bmatrix} y(t-L-1-m) \\ \vdots \\ y(t-L-M-m) \end{bmatrix} [y^*(t), \ldots, y^*(t-L)]\right\} \tag{4.10.4}$$

Derive an expression for $\tilde{\Gamma}$ similar to the one for $\Gamma$ in (4.2.8). Furthermore, make use of that expression in an ESPRIT-like method to estimate the frequencies $\{\omega_k\}$, instead of using it in an HOYW-like method (as in Section 4.4). Discuss the advantage of this so-called HOYW–ESPRIT method over the HOYW method based on $\tilde{\Gamma}$. Assuming that the noise is white (i.e., $m = 0$) and hence that ESPRIT is directly applicable, would you prefer using HOYW–ESPRIT (with $m = 0$) in lieu of ESPRIT? Why or why not?

**Exercise 4.8: Chebyshev Inequality and the Convergence of Sample Covariances**
Let $x$ be a random variable with finite mean $\mu$ and variance $\sigma^2$. Show that, for any positive constant $c$, the so-called Chebyshev inequality holds:

$$\boxed{\Pr(|x-\mu| \geq c\sigma) \leq 1/c^2} \tag{4.10.5}$$

Use (4.10.5) to show that, if a sample covariance lag $\hat{r}_N$ (estimated from $N$ data samples) converges to the true value $r$ *in the mean-square sense*

$$\lim_{N\to\infty} E\left\{|\hat{r}_N - r|^2\right\} = 0 \tag{4.10.6}$$

then $\hat{r}_N$ also converges to $r$ *in probability*:

$$\lim_{N\to\infty} \Pr(|\hat{r}_N - r| \neq 0) = 0 \tag{4.10.7}$$

For sinusoidal signals, the mean-square convergence of $\{\hat{r}_N(k)\}$ to $\{r(k)\}$, as $N \to \infty$, has been proven in Complement 4.9.1. (In this exercise, we omit the argument $k$ in $\hat{r}_N(k)$ and $r(k)$, for

notational simplicity.) Additionally, discuss the use of (4.10.5) to set *bounds* (which hold with a specified probability) on an arbitrary random variable with given mean and variance. Comment on the conservatism of the bounds obtained from (4.10.5) by comparing them with the bounds corresponding to a Gaussian random variable.

**Exercise 4.9: More about the Forward–Backward Approach**

The sample covariance matrix in (4.8.3), used by the forward–backward approach, is often a better estimate of the theoretical covariance matrix than $\hat{R}$ is (as argued in Section 4.8). Another advantage of (4.8.3) is that the forward–backward sample covariance is always numerically better conditioned than the usual (forward-only) sample covariance matrix $\hat{R}$. To understand this statement, let $R$ be a Hermitian matrix (not necessarily a Toeplitz one, like the $R$ in (4.2.7)). The "condition number" of $R$ is defined as

$$\text{cond}(R) = \lambda_{\max}(R)/\lambda_{\min}(R)$$

where $\lambda_{\max}(R)$ and $\lambda_{\min}(R)$ are the maximum and minimum eigenvalues of $R$, respectively. The numerical errors that affect many algebraic operations on $R$, such as inversion, eigendecomposition, and so on, are essentially proportional to $\text{cond}(R)$. Hence, the smaller $\text{cond}(R)$, the better. (See Appendix A for details on this aspect.)

Next, let $U$ be a unitary matrix (the $J$ in (4.8.3) being a special case of such a matrix). Observe that the forward–backward covariance in equation (4.8.3) is of the form $R + U^* R^T U$. Prove that

$$\boxed{\text{cond}(R) \geq \text{cond}(R + U^* R^T U)} \tag{4.10.8}$$

for any unitary matrix $U$. We note that the result (4.10.8) applies to any Hermitian matrix $R$ and unitary matrix $U$, and thus is valid in cases more general than the forward–backward approach in Section 4.8, in which $R$ is Toeplitz and $U = J$.

**Exercise 4.10: ESPRIT and Min–Norm Under the Same Umbrella**

ESPRIT and Min–Norm methods are seemingly quite different from one another; it might well seem unlikely that there is any strong relationship between them. It is the goal of this exercise to show that in fact ESPRIT and Min–Norm are quite related closely to each other. We will see that ESPRIT and Min–Norm are members of a well-defined class of frequency estimates.

Consider the equation

$$\hat{S}_2^* \hat{\Psi} = \hat{S}_1^* \tag{4.10.9}$$

where $\hat{S}_1$ and $\hat{S}_2$ are as defined in Section 4.7. The $(m-1) \times (m-1)$ matrix $\hat{\Psi}$ in (4.10.9) is the unknown. First, show that the asymptotic counterpart of (4.10.9),

$$S_2^* \Psi = S_1^* \tag{4.10.10}$$

has the property that any of its solutions $\Psi$ has $n$ eigenvalues equal to $\{e^{-i\omega_k}\}_{k=1}^n$. This property, along with the fact that there is an infinite number of matrices $\hat{\Psi}$ satisfying (4.10.9) (see

Section A.8 in Appendix A), implies that (4.10.9) generates a class of frequency estimators with an infinite number of members.

As a second task, show that ESPRIT and Min–Norm belong to this class of estimators. In other words, prove that there is a solution of (4.10.9) whose nonzero eigenvalues have exactly the same arguments as the eigenvalues of the ESPRIT matrix $\hat{\phi}$ in (4.7.12), and also that there is another solution of (4.10.9) whose eigenvalues are equal to the roots of the Min–Norm polynomial in (4.6.3). For more details on the topic of this exercise, see [Hua and Sarkar 1990].

**Exercise 4.11: Yet Another Relationship between ESPRIT and Min–Norm**
Let the vector $[\hat{\rho}^T, 1]^T$ be defined similarly to the Min–Norm vector $[1, \hat{g}^T]^T$ (see (4.6.1)), the only difference being that now we constrain the last element to be equal to one. Hence, $\hat{\rho}$ is the minimum-norm solution to (see (4.6.5))

$$\hat{S}^* \begin{bmatrix} \hat{\rho} \\ 1 \end{bmatrix} = 0$$

Use the Min–Norm vector $\hat{\rho}$ to build the following matrix

$$\tilde{\phi} = \hat{S}^* \left[ \begin{array}{c|c} 0 & \begin{matrix} I_{m-1} \\ \hline -\hat{\rho}^* \end{matrix} \end{array} \right] \hat{S} \quad (n \times n)$$

Prove the somewhat curious fact that $\tilde{\phi}$ above is equal to the ESPRIT matrix, $\hat{\phi}$, in (4.7.12).

## COMPUTER EXERCISES

**Tools for Frequency Estimation:**
The text website `www.prenhall.com/stoica` contains the following MATLAB functions for use in computing frequency estimates and estimating the number of sinusoidal terms. In the first four functions, `y` is the data vector and `n` is the desired number of frequency estimates. The remaining variables are described below.

- `w=hoyw(y,n,L,M)`
  The HOYW estimator given in the box on page 166; `L` and `M` are the matrix dimensions as in (4.4.8).
- `w=music(y,n,m)`
  The Root MUSIC estimator given by (4.5.12); `m` is the dimension of $a(\omega)$. This function also implements the Pisarenko method by setting $m = n + 1$.
- `w=minnorm(y,n,m)`
  The Root Min–Norm estimator given by (4.6.3); `m` is the dimension of $a(\omega)$.
- `w=esprit(y,n,m)`
  The ESPRIT estimator given by (4.7.12); `m` is the size of the square matrix $\hat{R}$ there, and $S_1$ and $S_2$ are chosen as in equations (4.7.5) and (4.7.6).

- `order=sinorder(mvec,sig2,N,nu)`
  Computes the AIC, $AIC_c$, GIC, and BIC model-order selections for sinusoidal parameter estimation problems. See Appendix C for details on the derivations of these methods. Here, `mvec` is a vector of candidate sinusoidal model orders, `sig2` is the vector of estimated residual variances corresponding to the model orders in `mvec`, `N` is the length of the observed data vector, and `nu` is a parameter in the GIC method. The 4-element output vector `order` contains the selected model orders obtained from AIC, $AIC_c$, GIC, and BIC, respectively.

**Exercise C4.12: Resolution Properties of Subspace Methods for Estimation of Line Spectra** In this exercise, we test and compare the resolution properties of four subspace methods: Min–Norm, MUSIC, ESPRIT, and HOYW.

Generate realizations of the sinusoidal signal

$$y(t) = 10\sin(0.24\pi t + \varphi_1) + 5\sin(0.26\pi t + \varphi_2) + e(t), \qquad t = 1, \ldots, N$$

where $N = 64$, $e(t)$ is Gaussian white noise with variance $\sigma^2$, and $\varphi_1$, $\varphi_2$ are independent random variables each uniformly distributed on $[-\pi, \pi]$.

Generate 50 Monte Carlo realizations of $y(t)$, and present the results from these experiments. The results of frequency estimation can be presented, comparing the sample means and variances of the frequency estimates from the various estimators.

**(a)** Find the exact ACS for $y(t)$. Compute the "true" frequency estimates from the four methods, for $n = 4$ and various choices of the order $m \geq 5$ (and corresponding choices of $M$ and $L$ for HOYW). Which method(s) are able to resolve the two sinusoids, and for what values of $m$ (or $M$ and $L$)?

**(b)** Consider now $N = 64$, and set $\sigma^2 = 0$; this corresponds to the case of finite data length but infinite SNR. Compute frequency estimates for the four techniques again, using $n = 4$ and various choices of $m$, $M$, and $L$. Which method(s) are reliably able to resolve the sinusoids? Explain why.

**(c)** Obtain frequency estimates from the four methods when $N = 64$ and $\sigma^2 = 1$. Use $n = 4$, and experiment with different choices of $m$, $M$, and $L$ to see the effect on estimation accuracy (e.g., try $m = 5$, 8, and 12 for MUSIC, Min–Norm, and ESPRIT, and try $L = M = 4$, 8, and 12 for HOYW). Which method(s) give reliable "superresolution" estimation of the sinusoids? Is it possible to resolve the two sinusoids in the signal? Discuss how the choices of $m$, $M$, and $L$ influence the resolution properties. Which method appears to have the best resolution?

You may want to experiment further by changing the SNR and the relative amplitudes of the sinusoids to gain a better understanding of the differences between the methods.

**(d)** Compare the estimation results with the AR and ARMA results obtained in Exercise C3.18 in Chapter 3. What are the major differences between the techniques? Which method(s) do you prefer for this problem?

**Exercise C4.13: Model Order Selection for Sinusoidal Signals**
In this exercise, we examine four methods for model order selection for sinusoidal signals. As discussed in Appendix C, several important model order selection rules have the general form (see (C.8.1)–(C.8.2))

$$-2\ln p_n(y,\hat{\theta}^n) + \eta(r,N)r \tag{4.10.11}$$

with different *penalty coefficients* $\eta(r,N)$ for the different methods:

$$\begin{aligned} \text{AIC}: &\quad \eta(r,N) = 2 \\ \text{AIC}_\text{c}: &\quad \eta(r,N) = 2\frac{N}{N-r-1} \\ \text{GIC}: &\quad \eta(r,N) = \nu \ (e.g.,\ \nu = 4) \\ \text{BIC}: &\quad \eta(r,N) = \ln N \end{aligned} \tag{4.10.12}$$

Here, $N$ is the length of the observed data vector $y$ and, for sinusoidal signals, $r$ is given (see Appendix C) by

$$\begin{aligned} r &= 3n+1 \quad \text{for AIC, AIC}_\text{c}\text{, and GIC} \\ r &= 5n+1 \quad \text{for BIC} \end{aligned}$$

where $n$ is the number of sinusoids in the model. The term $\ln p_n(y,\hat{\theta}^n)$ is the log-likelihood of the observed data vector $y$ given the maximum likelihood (ML) estimate of the parameter vector $\theta$ for a model order of $n$; it is given (*cf.* (C.2.7)–(C.2.8) in Appendix C) by

$$-2\ln p_n(y,\hat{\theta}^n) = N\hat{\sigma}_n^2 + \text{constant} \tag{4.10.13}$$

where

$$\hat{\sigma}_n^2 = \frac{1}{N}\sum_{t=1}^{N}\left|y(t) - \sum_{k=1}^{n}\hat{\alpha}_k e^{i(\hat{\omega}_k t+\hat{\varphi}_k)}\right|^2 \tag{4.10.14}$$

The selected model order is the value of $n$ that minimizes (4.10.11). The preceding order selection rules, although derived for ML estimates of $\theta$, can be used even with approximate ML estimates of $\theta$, albeit with some loss of performance.

### Well-Separated Sinusoids.

**(a)** Generate 100 realizations of

$$y(t) = 10\sin[2\pi f_0 t + \varphi_1] + 5\sin[2\pi(f_0+\Delta f)t + \varphi_2] + e(t), \qquad t = 1,\ldots,N$$

for $f_0 = 0.24$, $\Delta f = 3/N$, and $N = 128$. Here, $e(t)$ is real-valued white noise with variance $\sigma^2$. For each realization, generate $\varphi_1$ and $\varphi_2$ as random variables uniformly distributed on $[0, 2\pi]$.

**(b)** Set $\sigma^2 = 10$. For each realization, estimate the frequencies of $n = 1, \ldots, 10$ real-valued sinusoidal components, by using ESPRIT, and estimate the amplitudes and phases by using the second equation in (4.3.8), where $\hat{\omega}$ is the vector of ESPRIT frequency estimates. Note that you will need to use two complex exponentials to model each real-valued sinusoid, so the number of frequencies to estimate with ESPRIT will be $2, 4, \ldots, 20$; however, the frequency estimates will be in symmetric pairs. Use $m = 40$ as the covariance matrix size in ESPRIT.

**(c)** Find the model orders that minimize AIC, $\text{AIC}_c$, GIC (with $\nu = 4$), and BIC. For each of the four order-selection methods, plot a histogram of the selected orders for the 100 realizations. Comment on their relative performance.

**(d)** Repeat the preceding experiment, using $\sigma^2 = 1$ and $\sigma^2 = 0.1$, and comment on the performance of the order selection methods as a function of SNR.

**Closely Spaced Sinusoids.** Generate 100 realizations of $y(t)$ as in previous case, but this time using $\Delta f = 0.5/N$. Repeat the preceding experiments. In addition, compare the relative performance of the order selection methods for well-separated versus closely spaced sinusoidal signals.

**Exercise C4.14: Line Spectral Methods Applied to Measured Data**
Apply the Min–Norm, MUSIC, ESPRIT, and HOYW frequency estimators to the data in the files `sunspotdata.mat` and `lynxdata.mat` (using both the original `lynx` data and the logarithmically transformed data, as in Exercise C2.23). These files can be obtained from the text website `www.prenhall.com/stoica`. Try to answer the following questions:

**(a)** Is the sinusoidal model appropriate for the data sets under study?

**(b)** Suggest how to choose the number of sinusoids in the model. (See Exercise C4.13.)

**(c)** What periodicities can you find in the two data sets?

Compare the results you obtain here to the AR(MA) and nonparametric spectral estimation results you obtained in Exercises C2.23 and C3.20.