Appendix B

Cramér–Rao Bound Tools

B.1 INTRODUCTION

In the text, we have kept the discussion of statistical aspects at a minimum for conciseness reasons. However, we have presented certain statistical tools and analyses that we have found useful to the understanding of the spectral analysis material discussed. In this appendix, we introduce some basic facts on an important statistical tool: the Cramér–Rao bound (abbreviated as CRB). We begin our discussion by explaining the importance of the CRB for *parametric spectral analysis*.

Let $\phi(\omega, \theta)$ denote a parametric spectral model, depending on a *real-valued* vector θ, and let $\phi(\omega, \hat{\theta})$ denote the spectral density estimated from N data samples. Assume that the estimate $\hat{\theta}$ of θ is *consistent*, so that the estimation error is small for large values of N. Then, by making use of a Taylor series expansion technique, we can write the estimation error $[\phi(\omega, \hat{\theta}) - \phi(\omega, \theta)]$ approximately as a linear function of $\hat{\theta} - \theta$, namely,

$$[\phi(\omega, \hat{\theta}) - \phi(\omega, \theta)] \simeq \psi^T(\omega, \theta)(\hat{\theta} - \theta) \tag{B.1.1}$$

where the symbol $\simeq$ denotes an asymptotically (in N) valid approximation and $\psi(\omega, \theta)$ is the gradient of $\phi(\omega, \theta)$ with respect to θ (evaluated at the true parameter values):

$$\psi(\omega, \theta) = \frac{\partial \phi(\omega, \theta)}{\partial \theta} \tag{B.1.2}$$

It follows from (B.1.1) that the mean squared error (MSE) of $\phi(\omega, \hat{\theta})$ is approximately given by

$$\text{MSE}[\phi(\omega, \hat{\theta})] \simeq \psi^T(\omega, \theta) P \psi(\omega, \theta) \qquad (\text{for } N \gg 1) \tag{B.1.3}$$

where

$$P = \text{MSE}[\hat{\theta}] = E\left\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\right\} \tag{B.1.4}$$

We see from (B.1.3) that the variance (or MSE) of the estimation errors in the spectral domain is linearly related to the variance (or MSE) of the parameter vector estimate $\hat{\theta}$, and so we can get an accurate spectral estimate only if we use an accurate parameter estimator. We start from this simple observation, which reduces the statistical analysis of $\phi(\omega, \hat{\theta})$ to the analysis of $\hat{\theta}$, to explain the importance of the CRB for the performance study of spectral analysis. Toward that end, we discuss several facts in the paragraphs that follow.

Assume that $\hat{\theta}$ is some *unbiased estimate* of θ (i.e., $E\{\hat{\theta}\} = \theta$), and let P denote the covariance matrix of $\hat{\theta}$ (*cf.* (B.1.4)):

$$P = E\left\{(\hat{\theta} - \theta)(\hat{\theta} - \theta)^T\right\} \tag{B.1.5}$$

(Note that here we do not require that N be large.) Then, under quite general conditions, there is a matrix (which we denote by P_{cr}) such that

$$P \geq P_{cr} \tag{B.1.6}$$

in the sense that the difference $(P - P_{cr})$ is a positive semidefinite matrix. This is basically the celebrated Cramér–Rao bound result [Cramér 1946; Rao 1945]. We will derive the inequality (B.1.6) along with an expression for the CRB in the next section.

In view of (B.1.6), we may think of assessing the performance of a given estimation method by comparing its covariance matrix P with the CRB. Such a comparison would make perfect sense whenever the CRB is *achievable*—that is, whenever there exists an estimation method such that its P equals the CRB. Unfortunately, this is rarely the case for finite N. Additionally, *biased* estimators with MSEs smaller than the CRB can exist. (See, for example, [Stoica and Moses 1990; Stoica and Ottersten 1996].) Hence, in the *finite sample case* (particularly for small samples), comparing with the CRB does not really make much sense, because

(i) there might be no unbiased estimator that attains the CRB and, consequently, a large difference $(P - P_{cr})$ would not necessarily mean bad accuracy; and
(ii) the equality $P = P_{cr}$ does not necessarily mean that we have achieved the ultimate possible performance, because there might be biased estimators with lower MSE than the CRB.

In the *large sample case*, on the other hand, the utility of the CRB result for the type of parameter estimation problems addressed in the text is significant, as explained next.

Let $y \in \mathbf{R}^{N\times 1}$ denote the sample of available observations. Any estimate $\hat{\theta}$ of θ will be a function of y. We assume that both θ and y are *real valued*. Working with real θ and y vectors appears to be the most convenient way when discussing the CRB theory, even when the original parameters and measurements are complex-valued. (If the parameters and measurements are complex-valued, θ and y are obtained by concatenating the real and imaginary parts of the

complex parameter and data vectors, respectively.) We also assume that the probability density of y, which we denote by $p(y, \theta)$, is a differentiable function of θ. An important general method for parameter estimation consists of maximizing $p(y, \theta)$ with respect to θ:

$$\hat{\theta} = \arg \max_{\theta} p(y, \theta) \tag{B.1.7}$$

The $p(y, \theta)$ in (B.1.7) with y fixed and θ variable is called the *likelihood function*, and $\hat{\theta}$ is called the *maximum likelihood* (ML) *estimate* of θ. Under regularity conditions, the ML estimate (MLE) is *consistent* (i.e., $\lim_{N \to \infty} \hat{\theta} = \theta$ stochastically), and its covariance matrix approaches the CRB as N increases:

$$P \simeq P_{cr} \qquad \text{for a MLE with } N \gg 1 \tag{B.1.8}$$

The aforementioned regularity conditions basically amount to requiring that the number of free parameters not increase with N, which is true for all but one of the parametric spectral estimation problems discussed in the text. The array processing problem of Chapter 6 does not satisfy the previous requirement when the signal snapshots are assumed to be unknown deterministic variables; in such a case, the number of unknown parameters grows without bound as N increases, and the equality in (B.1.8) does not hold; see [Stoica and Nehorai 1989a; Stoica and Nehorai 1990] and also Section B.6.

In summary, then, in *large samples*, the ML method attains the ultimate performance corresponding to the CRB, under rather general conditions. Furthermore, there are no other known *practical methods* that can provide consistent estimates of θ with lower variance than the CRB.[1] Hence, the ML method can be said to be asymptotically a *statistically efficient practical estimation approach*. The accuracy achieved by any other estimation method can therefore be assessed *by comparing the (large-sample) covariance matrix of that method with the CRB*, which approximately equals the covariance matrix of the MLE in large samples (*cf.* (B.1.8)). This performance-comparison ability is one of the most important uses of the CRB.

With reference to the spectral estimation problem, it follows from (B.1.3) and the previous observation that we can assess the performance of a given spectral estimator by comparing its large sample MSE values with

$$\boxed{\psi^T(\omega, \theta)[P_{cr}]\psi(\omega, \theta)} \tag{B.1.9}$$

The MSE values can be obtained either by the Monte Carlo simulation of a typical scenario representative of the problem of interest or by using analytical MSE formulas whenever they are available. In this book, we have emphasized the former, more pragmatic way of finding the MSE of a given spectral estimator.

[1] Consistent estimation methods whose asymptotic variance is lower than the CRB, at certain points in the parameter set, do exist! However, such methods (which are called "asymptotically statistically superefficient") have little practical relevance (they are mainly of a theoretical interest); see, for example, [Stoica and Ottersten 1996].

Remark: The CRB formula (B.1.9) for parametric (or model-based) spectral analysis holds in the case where the model order (i.e., the dimension of θ) is equal to the "true order." Of course, in any practical spectral analysis exercise using the parametric approach, we will have to estimate n, the model order, in addition to θ, the (real-valued) model parameters. The need for order estimation is a distinctive feature and an additional complication of parametric spectral analysis, as compared with nonparametric spectral analysis.

There are several available rules for order selection; see Appendix C. For most of these rules, the probability of underestimating the true order approaches zero as N increases (if that is not the case, then the estimated spectrum could be heavily biased). The probability of overestimating the true order, on the other hand, may be nonzero even when $N \to \infty$. Let $\hat{n}$ denote the estimated order, n_0 the true order, and $p_n = \Pr(\hat{n} = n)$ for $N \to \infty$. Assume that $p_n = 0$ for $n < n_0$ and that the CRB formula (B.1.9) holds for any $n \geq n_0$ (which is a relatively mild restriction). Then it can be shown (see [Sando, Mitra, and Stoica 2002] and the references therein) that, whenever n is estimated along with θ, the formula (B.1.9) should be replaced with its average over the distribution of order estimates:

$$\sum_{n=n_0}^{n_{MAX}} p_n \psi_n^T(\omega, \theta_n)[P_{cr,n}]\psi_n(\omega, \theta_n) \tag{B.1.10}$$

Here we have emphasized by notation the dependence of ψ, θ, and P_{cr} on the model order n, and n_{MAX} denotes the maximum order value considered in the order-selection rule. The set of probabilities $\{p_n\}$ associated with various order-estimation rules is tabulated e.g., in [McQuarrie and Tsai 1998]. As expected, it can be proven that the spectral CRB in (B.1.10) increases (for each ω) with increasing n_{MAX} (see [Sando, Mitra, and Stoica 2002]). This increase of the spectral-estimation error is the price paid for not knowing the true model order. ■

B.2 THE CRB FOR GENERAL DISTRIBUTIONS

Result R36: *(Cramér–Rao Bound)* Consider the likelihood function $p(y, \theta)$, introduced in the previous section, and define

$$P_{cr} = \left(E\left\{ \left[\frac{\partial \ln p(y,\theta)}{\partial \theta} \right] \left[\frac{\partial \ln p(y,\theta)}{\partial \theta} \right]^T \right\} \right)^{-1} \tag{B.2.1}$$

where the inverse is assumed to exist. Then

$$P \geq P_{cr} \tag{B.2.2}$$

holds for any unbiased estimate of θ. Furthermore, the CRB matrix can alternatively be expressed as

$$\boxed{P_{cr} = -\left(E\left\{\frac{\partial^2 \ln p(y,\theta)}{\partial\theta\ \partial\theta^T}\right\}\right)^{-1}} \tag{B.2.3}$$

Proof: As $p(y,\theta)$ is a probability density function,

$$\int p(y,\theta)dy = 1 \tag{B.2.4}$$

where the integration is over $\mathbf{R}^N$. The assumption that $\hat{\theta}$ is an unbiased estimate implies

$$\int \hat{\theta} p(y,\theta)dy = \theta \tag{B.2.5}$$

Differentiation of (B.2.4) and (B.2.5) with respect to θ yields, under regularity conditions,

$$\int \frac{\partial p(y,\theta)}{\partial\theta}dy = \int \frac{\partial \ln p(y,\theta)}{\partial\theta} p(y,\theta)dy = E\left\{\frac{\partial \ln p(y,\theta)}{\partial\theta}\right\} = 0 \tag{B.2.6}$$

and

$$\int \hat{\theta}\frac{\partial p(y,\theta)}{\partial\theta}dy = \int \hat{\theta}\frac{\partial \ln p(y,\theta)}{\partial\theta} p(y,\theta)dy = E\left\{\hat{\theta}\frac{\partial \ln p(y,\theta)}{\partial\theta}\right\} = I \tag{B.2.7}$$

It follows from (B.2.6) and (B.2.7) that

$$E\left\{(\hat{\theta}-\theta)\frac{\partial \ln p(y,\theta)}{\partial\theta}\right\} = I \tag{B.2.8}$$

Next note that the matrix

$$E\left\{\begin{bmatrix} (\hat{\theta}-\theta) \\ \dfrac{\partial \ln p(y,\theta)}{\partial\theta} \end{bmatrix}\begin{bmatrix} (\hat{\theta}-\theta)^T & \left(\dfrac{\partial \ln p(y,\theta)}{\partial\theta}\right)^T \end{bmatrix}\right\} = \begin{bmatrix} P & I \\ I & P_{cr}^{-1} \end{bmatrix} \tag{B.2.9}$$

is, by construction, positive semidefinite. (To obtain the equality in (B.2.9), we used (B.2.8).) This observation implies (B.2.2) (see Result R20 in Appendix A).

Next, we prove the equality in (B.2.3). Differentiation of (B.2.6) gives

$$\int \frac{\partial^2 \ln p(y,\theta)}{\partial\theta\ \partial\theta^T} p(y,\theta)dy + \int \left[\frac{\partial \ln p(y,\theta)}{\partial\theta}\right]\left[\frac{\partial \ln p(y,\theta)}{\partial\theta}\right]^T p(y,\theta)dy = 0$$

or, equivalently,

$$E\left\{\left[\frac{\partial \ln p(y,\theta)}{\partial \theta}\right]\left[\frac{\partial \ln p(y,\theta)}{\partial \theta}\right]^T\right\} = -E\left\{\frac{\partial^2 \ln p(y,\theta)}{\partial \theta\, \partial \theta^T}\right\}$$

which is precisely what we had to prove. ■

The matrix

$$J = E\left\{\left[\frac{\partial \ln p(y,\theta)}{\partial \theta}\right]\left[\frac{\partial \ln p(y,\theta)}{\partial \theta}\right]^T\right\}$$

$$= -E\left\{\frac{\partial^2 \ln p(y,\theta)}{\partial \theta\, \partial \theta^T}\right\}, \tag{B.2.10}$$

the inverse of which appears in the CRB formula (B.2.1) (or (B.2.3)), is called the (Fisher) *information matrix* [Fisher 1922].

B.3 THE CRB FOR GAUSSIAN DISTRIBUTIONS

The CRB matrix in (B.2.1) depends implicitly on the data properties via the probability density function $p(y,\theta)$. To obtain a more explicit expression for the CRB, we should specify the data distribution. A particularly convenient CRB formula is obtained if the data vector is assumed to be Gaussian distributed—that is,

$$p(y,\theta) = \frac{1}{(2\pi)^{N/2}|C|^{1/2}} e^{-(y-\mu)^T C^{-1}(y-\mu)/2} \tag{B.3.1}$$

where μ and C are, respectively, the mean and the covariance matrix of y and C is assumed to be invertible. In the case of (B.3.1), the log-likelihood function that appears in (B.2.1) is given by

$$\ln p(y,\theta) = -\frac{N}{2}\ln 2\pi - \frac{1}{2}\ln|C| - \frac{1}{2}(y-\mu)^T C^{-1}(y-\mu) \tag{B.3.2}$$

Result R37: The CRB matrix corresponding to the Gaussian data distribution in (B.3.1) is given (elementwise) by

$$\boxed{[P_{cr}^{-1}]_{ij} = \frac{1}{2}\operatorname{tr}\left[C^{-1}C_i' C^{-1} C_j'\right] + \left[\mu_i'^T C^{-1}\mu_j'\right]} \tag{B.3.3}$$

where C_i' denotes the derivative of C with respect to the ith element of θ (and similarly for μ_i').

Proof: By using Result R21 and the notational foregoing convention for the first-order and second-order derivatives, we obtain

$$\begin{aligned}
2[\ln p(y,\theta)]''_{ij} &= \frac{\partial}{\partial\theta_i}\left\{-\operatorname{tr}\left[C^{-1}C'_j\right] + 2{\mu'_j}^T C^{-1}(y-\mu)\right. \\
&\qquad \left. +(y-\mu)^T C^{-1}C'_j C^{-1}(y-\mu)\right\} \\
&= \operatorname{tr}\left[C^{-1}C'_i C^{-1}C'_j\right] - \operatorname{tr}\left[C^{-1}C''_{ij}\right] \\
&\qquad +2\left\{\left[{\mu'_j}^T C^{-1}\right]'_i (y-\mu) - {\mu'_j}^T C^{-1}\mu'_i\right\} \\
&\qquad -2{\mu'_i}^T C^{-1}C'_j C^{-1}(y-\mu) \\
&\qquad +\operatorname{tr}\left\{(y-\mu)(y-\mu)^T\right. \\
&\qquad \left.\cdot\left[-C^{-1}C'_i C^{-1}C'_j C^{-1} + C^{-1}C''_{ij}C^{-1} - C^{-1}C'_j C^{-1}C'_i C^{-1}\right]\right\}
\end{aligned}$$

Taking the expectation of both sides of the preceding equation yields

$$\begin{aligned}
2\left[P_{cr}^{-1}\right]_{ij} &= -\operatorname{tr}\left[C^{-1}C'_i C^{-1}C'_j\right] + \operatorname{tr}\left[C^{-1}C''_{ij}\right] + 2{\mu'_i}^T C^{-1}\mu'_j \\
&\qquad +\operatorname{tr}\left[C^{-1}C'_i C^{-1}C'_j\right] - \operatorname{tr}\left[C^{-1}C''_{ij}\right] + \operatorname{tr}\left[C^{-1}C'_i C^{-1}C'_j\right] \\
&= \operatorname{tr}\left[C^{-1}C'_i C^{-1}C'_j\right] + 2{\mu'_i}^T C^{-1}\mu'_j
\end{aligned}$$

which concludes the proof. ■

The CRB expression in (B.3.3) is sometimes referred to as the *Slepian–Bangs formula*. (The second term in (B.3.3) is due to Slepian [Slepian 1954] and the first to Bangs [Bangs 1971].)

Next, we specialize the CRB formula (B.3.3) to a particular type of Gaussian distribution. Let $N = 2\bar{N}$ (hence, N is assumed to be even). Partition the vector y as

$$y = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \begin{matrix} \}\bar{N} \\ \}\bar{N} \end{matrix} \tag{B.3.4}$$

Accordingly, partition μ and C as

$$\mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \tag{B.3.5}$$

and

$$C = \begin{bmatrix} C_{11} & C_{12} \\ C_{12}^T & C_{22} \end{bmatrix} \tag{B.3.6}$$

The vector y is said to have a *circular* (or *circularly symmetric*) *Gaussian distribution* if

$$C_{11} = C_{22} \tag{B.3.7}$$

$$C_{12}^T = -C_{12} \tag{B.3.8}$$

Let

$$\mathbf{y} \stackrel{\triangle}{=} y_1 + iy_2 \tag{B.3.9}$$

and

$$\boldsymbol{\mu} = \mu_1 + i\,\mu_2 \tag{B.3.10}$$

We also say that the *complex-valued random vector* $\mathbf{y}$ *has a circular Gaussian distribution* whenever the conditions (B.3.7) and (B.3.8) are satisfied. It is a straightforward exercise to verify that the aforementioned conditions can be more compactly written as:

$$E\left\{(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^T\right\} = 0 \tag{B.3.11}$$

Both the Fourier transform (see Chapter 2) and the complex demodulation operation (see Chapter 6) often lead to signals satisfying (B.3.11) (see, e.g., [Brillinger 1981]). Hence, the *circularity* is a relatively common property of Gaussian random signals encountered in spectral analysis problems.

Remark: If a random vector $\mathbf{y}$ satisfies the "circularity condition" (B.3.11), then it is readily verified that $\mathbf{y}$ and $\mathbf{y}e^{iz}$ have the same second-order properties for every constant z in $[-\pi, \pi]$. Hence, the second-order properties of $\mathbf{y}$ do not change if its generic element $\mathbf{y}_k$ is replaced by any other value, $\mathbf{y}_k e^{iz}$, on the *circle* with radius $|\mathbf{y}_k|$ (recall that z is nonrandom and it does not depend on k). This observation provides a motivation for the name "circularly symmetric" given to such a random vector $\mathbf{y}$. ■

Let

$$\Gamma = E\left\{(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^*\right\} \tag{B.3.12}$$

For circular Gaussian random vectors y (or $\mathbf{y}$), the CRB formula (B.3.3) can be rewritten in a compact form as a function of Γ and $\boldsymbol{\mu}$. (Note that the dimensions of Γ and $\boldsymbol{\mu}$ are half the dimensions of C and μ appearing in (B.3.3).) In order to show how this can be done, we need some preparations.

Let

$$\bar{C} = C_{11} = C_{22} \tag{B.3.13}$$

$$\tilde{C} = C_{12}^T = -C_{12} \tag{B.3.14}$$

Hence,

$$C = \begin{bmatrix} \bar{C} & -\tilde{C} \\ \tilde{C} & \bar{C} \end{bmatrix} \tag{B.3.15}$$

and

$$\Gamma = 2(\bar{C} + i\tilde{C}) \tag{B.3.16}$$

To any complex-valued matrix $\mathcal{C} = \bar{C} + i\tilde{C}$ we associate a real-valued matrix C as defined in (B.3.15), and vice versa. It is a simple exercise to verify that, if

$$\mathcal{A} = \mathcal{B}\mathcal{C} \Longleftrightarrow \bar{A} + i\tilde{A} = (\bar{B} + i\tilde{B})(\bar{C} + i\tilde{C}) \tag{B.3.17}$$

then the real-valued matrix associated with $\mathcal{A}$ is given by

$$A = BC \Longleftrightarrow \begin{bmatrix} \bar{A} & -\tilde{A} \\ \tilde{A} & \bar{A} \end{bmatrix} = \begin{bmatrix} \bar{B} & -\tilde{B} \\ \tilde{B} & \bar{B} \end{bmatrix} \begin{bmatrix} \bar{C} & -\tilde{C} \\ \tilde{C} & \bar{C} \end{bmatrix} \tag{B.3.18}$$

In particular, it follows from (B.3.17) and (B.3.18) with $A = I$ (and hence $\mathcal{A} = I$) that the matrices C^{-1} and $\mathcal{C}^{-1}$ form a real-complex pair as just defined.

We deduce from the results previously derived that the matrix in the first term of (B.3.3),

$$D = C^{-1}C_i'C^{-1}C_j' \tag{B.3.19}$$

is associated with

$$\mathcal{D} = \mathcal{C}^{-1}\mathcal{C}_i'\mathcal{C}^{-1}\mathcal{C}_j' = \Gamma^{-1}\Gamma_i'\Gamma^{-1}\Gamma_j' \tag{B.3.20}$$

Furthermore, we have

$$\frac{1}{2}\operatorname{tr}(D) = \operatorname{tr}(\bar{D}) = \operatorname{tr}(\mathcal{D}) \tag{B.3.21}$$

The second equality in (B.3.21) follows from the fact that $\mathcal{C}$ is Hermitian, and hence

$$\operatorname{tr}(\mathcal{D}^*) = \operatorname{tr}(\mathcal{C}_j'\mathcal{C}^{-1}\mathcal{C}_i'\mathcal{C}^{-1}) = \operatorname{tr}(\mathcal{C}^{-1}\mathcal{C}_i'\mathcal{C}^{-1}\mathcal{C}_j') = \operatorname{tr}(\mathcal{D})$$

which in turn implies that $\operatorname{tr}(\tilde{D}) = 0$ and therefore that $\operatorname{tr}(\mathcal{D}) = \operatorname{tr}(\bar{D})$. Combining (B.3.20) and (B.3.21) shows that the first term in (B.3.3) can be rewritten as

$$\operatorname{tr}(\Gamma^{-1}\Gamma_i'\Gamma^{-1}\Gamma_j') \tag{B.3.22}$$

Next, we consider the second term in (B.3.3). Let

$$x = \begin{bmatrix} x_1 \\ x_2 \end{bmatrix} \qquad \text{and} \qquad z = \begin{bmatrix} z_1 \\ z_2 \end{bmatrix}$$

be two arbitrary vectors partitioned similarly to μ, and let $\mathbf{x} = x_1 + ix_2$ and $\mathbf{z} = z_1 + iz_2$. A straightforward calculation shows that

$$\begin{aligned} x^T A z &= x_1^T \bar{A} z_1 + x_2^T \bar{A} z_2 + x_2^T \tilde{A} z_1 - x_1^T \tilde{A} z_2 \\ &= \mathrm{Re}\left\{\mathbf{x}^* \mathcal{A} \mathbf{z}\right\} \end{aligned} \tag{B.3.23}$$

Hence,

$$\begin{aligned} \mu_i'^T C^{-1} \mu_j' &= \mathrm{Re}\left\{\boldsymbol{\mu}_i'^* \mathcal{C}^{-1} \boldsymbol{\mu}_j'\right\} \\ &= 2\,\mathrm{Re}\left\{\boldsymbol{\mu}_i'^* \Gamma^{-1} \boldsymbol{\mu}_j'\right\} \end{aligned} \tag{B.3.24}$$

Insertion of (B.3.22) and (B.3.24) into (B.3.3) yields the following CRB formula, which holds in the case of *circularly Gaussian-distributed data vectors* y (or $\mathbf{y}$):

$$\boxed{[P_{cr}^{-1}]_{ij} = \mathrm{tr}\left[\Gamma^{-1}\Gamma_i'\Gamma^{-1}\Gamma_j'\right] + 2\,\mathrm{Re}\left[\boldsymbol{\mu}_i'^* \Gamma^{-1} \boldsymbol{\mu}_j'\right]} \tag{B.3.25}$$

The importance of the Gaussian CRB formulas lies not only in the fact that Gaussian data are rather frequently encountered in applications, but also in a more subtle aspect, explained in what follows. Briefly stated, the second reason for the importance of the CRB formulas derived in this section is that

> Under rather general conditions and (at least) in large samples, the Gaussian CRB is the largest of all CRB matrices corresponding to different congruous distributions of the data sample.[2] (B.3.26)

To motivate the previous assertion, consider the ML estimate of θ derived under the Gaussian data hypothesis, which we denote by $\hat{\theta}_G$. According to the discussion around equation (B.1.8), the large sample covariance matrix of $\hat{\theta}$ equals P_{cr}^G—as with $\hat{\theta}_G$, we use an index G to denote the CRB matrix in the Gaussian-hypothesis case. Now, under rather general conditions, the large sample properties of the Gaussian ML estimator are independent of the data distribution. (See, for example, [Söderström and Stoica 1989].) In other words, the large sample covariance matrix of $\hat{\theta}_G$ is equal to P_{cr}^G for many data distributions other than the Gaussian one. This observation, along with the general CRB inequality, implies that

$$P_{cr}^G \geq P_{cr} \tag{B.3.27}$$

where the right-hand side is the CRB matrix corresponding to the data distribution at hand.

[2] A meaningful comparison of the CRBs under two different data distributions requires that the hypothesized distributional models not contain conflicting assumptions. In particular, when one of the two distributions is the Gaussian, the mean and covariance matrix should be the same for both distributions.

The inequality (B.3.27) (or, equivalently, the assertion (B.3.26)) shows that a method whose covariance matrix is much larger than P_{cr}^G cannot be a good estimation method. As a matter of fact, the "asymptotic properties" of most existing parameter estimation methods do not depend on the data distribution. This means that P_{cr}^G is a lower bound for the covariance matrices of a large class of estimation methods, regardless of the data distribution. On the other hand, the inequality (B.3.27) also shows that, for non-Gaussian data, it should be possible to beat the Gaussian CRB (for instance, by exploiting higher order moments of the data, beyond the first- and second-order moments used in the Gaussian ML method). However, general estimation methods with covariance matrices uniformly smaller than P_{cr}^G are yet to be discovered. In summary, comparing against the P_{cr}^G makes sense in most parameter estimation exercises.

In what follows, we briefly consider the application of the general Gaussian CRB formulas derived in this section to the three main parameter estimation problems treated in the text.

B.4 THE CRB FOR LINE SPECTRA

As explained in Chapter 4, the estimation of line spectra is basically a parameter estimation problem. The corresponding parameter vector is

$$\theta = \left[\alpha_1 \ \ldots \ \alpha_n,\ \varphi_1 \ \ldots \ \varphi_n,\ \omega_1 \ \ldots \ \omega_n,\ \sigma^2\right]^T \tag{B.4.1}$$

and the data vector is

$$\mathbf{y} = \left[y(1) \cdots y(N)\right]^T \tag{B.4.2}$$

or, in real valued form,

$$y = \left[\ \mathrm{Re}[y(1)] \cdots \mathrm{Re}[y(N)]\ \ \mathrm{Im}[y(1)] \cdots \mathrm{Im}[y(N)]\ \right]^T \tag{B.4.3}$$

When $\{\varphi_k\}$ are assumed to be random variables uniformly distributed on $[0, 2\pi]$ (whereas $\{\alpha_k\}$ and $\{\omega_k\}$ are deterministic constants), the distribution of $\mathbf{y}$ is *not* Gaussian and, hence, neither of the CRB formulas of the previous section is usable. To overcome this difficulty, it is customary to consider the distribution of $\mathbf{y}$ *conditioned on* $\{\varphi_k\}$ (i.e., for $\{\varphi_k\}$ fixed). This distribution is circular Gaussian, under the assumption that the (white) noise is circularly Gaussian distributed, with the following mean and covariance matrix:

$$\boldsymbol{\mu} = E\{\mathbf{y}\} = \begin{bmatrix} 1 & \cdots & 1 \\ e^{i\omega_1} & \cdots & e^{i\omega_n} \\ \vdots & & \vdots \\ e^{i(N-1)\omega_1} & \cdots & e^{i(N-1)\omega_n} \end{bmatrix} \begin{bmatrix} \alpha_1 e^{i\varphi_1} \\ \vdots \\ \alpha_n e^{i\varphi_n} \end{bmatrix} \tag{B.4.4}$$

$$\Gamma = E\left\{(\mathbf{y} - \boldsymbol{\mu})(\mathbf{y} - \boldsymbol{\mu})^*\right\} = \sigma^2 I \tag{B.4.5}$$

The differentiation of (B.4.4) and (B.4.5) with respect to the elements of the parameter vector θ can be done easily; we leave the details of this differentiation operation as an exercise to the reader. Hence, we can readily obtain all ingredients required to evaluate the CRB matrix in equation (B.3.25). If the distribution of $\mathbf{y}$ (or y) is Gaussian but not circular, we need additional parameters, besides σ^2, to characterize the matrix $E\left\{(\mathbf{y}-\boldsymbol{\mu})(\mathbf{y}-\boldsymbol{\mu})^T\right\}$. Once these parameters are introduced, the use of formula (B.3.3) to obtain the CRB is straightforward.

In Section 4.3, we gave a simple formula for the block of the CRB matrix corresponding to the frequency estimates $\{\hat{\omega}_k\}$. That formula holds asymptotically, as N increases. For finite values of N, it is a good approximation of the exact CRB whenever the minimum frequency separation is larger than $1/N$ [Stoica, Moses, Friedlander, and Söderström 1989]. In any case, the approximate (large-sample) CRB formula given in Section 4.3 is computationally much simpler to implement than the exact CRB.

The computation and properties of the CRB for line-spectral models are discussed in great detail in [Ghogho and Swami 1999]. In particular, a modified lower bound on the variance of any unbiased estimates of $\{\alpha_k\}$ and $\{\omega_k\}$ is derived for the case in which $\{\varphi_k\}$ are independent random variables uniformly distributed on $[0, 2\pi]$. That bound, which was obtained by using the so-called posterior CRB introduced in [Van Trees 1968] (as indicated above, the standard CRB does not apply to such a case), has an expression that is quite similar to the large-sample CRB given in [Stoica, Moses, Friedlander, and Söderström 1989] (see Section 4.3 for the large-sample CRB for $\{\hat{\omega}_k\}$). The paper [Ghogho and Swami 1999] also discusses the derivation of the CRB in the case of non-Gaussian noise distributions. The extension of the asymptotic CRB formula in Section 4.3 to the case of colored noise can be found in [Stoica, Jakobsson, and Li 1997].

B.5 THE CRB FOR RATIONAL SPECTRA

For rational (or ARMA) spectra, the Cramér–Rao lower bound on the variance of any consistently estimated spectrum is given by (B.1.9). The CRB matrix for the parameter-vector estimate, which appears in (B.1.9), can be evaluated as outlined in what follows.

In the case of ARMA spectral models, the parameter vector consists of the white-noise power σ^2 and the polynomial coefficients $\{a_k, b_k\}$. We arrange the ARMA coefficients in the following real-valued vector:

$$\theta = [\mathrm{Re}(a_1)\cdots\mathrm{Re}(a_n)\ \ \mathrm{Re}(b_1)\cdots\mathrm{Re}(b_m)\,\mathrm{Im}(a_1)\cdots\mathrm{Im}(a_n)\ \ \mathrm{Im}(b_1)\cdots\mathrm{Im}(b_m)]^T$$

The data vector is defined as in equations (B.4.2) or (B.4.3) and has zero mean ($\mu = 0$). The calculation of the covariance matrix of the data vector reduces to the calculation of ARMA covariances—that is,

$$r(k) = \sigma^2 E\left\{\left[\frac{B(z)}{A(z)}w(t)\right]\left[\frac{B(z)}{A(z)}w(t-k)\right]^*\right\}$$

where the white-noise sequence $\{w(t)\}$ is normalized in such a way that its variance is 1. Methods for computation of $\{r_k\}$ (for given values of σ^2 and θ) were outlined in Exercises C1.12 and 3.2.

The method in Exercise C1.12 should perform reasonably well as long as the zeroes of $A(z)$ are not too close to the unit circle. If the zeroes of $A(z)$ are close to the unit circle, it is advisable to use the method in Exercise 3.2 or in [KINKEL, PERL, SCHARF, AND STUBBERUD 1979; DEMEURE AND MULLIS 1989].

The calculation of the derivatives of $\{r(k)\}$ with respect to σ^2 and the elements of θ, which appear in the CRB formulas (B.3.3) or (B.3.25), can also be reduced to ARMA (cross)covariance computation. To see this, let α and γ be the real parts of a_p and b_p, respectively. Then

$$\frac{\partial r(k)}{\partial \alpha} = -\sigma^2 E\left\{\left[\frac{B(z)}{A^2(z)}w(t-p)\right]\left[\frac{B(z)}{A(z)}w(t-k)\right]^* + \left[\frac{B(z)}{A(z)}w(t)\right]\left[\frac{B(z)}{A^2(z)}w(t-k-p)\right]^*\right\}$$

and

$$\frac{\partial r(k)}{\partial \gamma} = \sigma^2 E\left\{\left[\frac{1}{A(z)}w(t-p)\right]\left[\frac{B(z)}{A(z)}w(t-k)\right]^* + \left[\frac{B(z)}{A(z)}w(t)\right]\left[\frac{1}{A(z)}w(t-k-p)\right]^*\right\}$$

The derivatives of $r(k)$ with respect to the imaginary parts of a_p and b_p can be similarly obtained. The differentiation of $r(k)$ with respect to σ^2 is immediate. Hence, by making use of an algorithm for ARMA cross-covariance calculation (similar to the ones for autocovariance calculation in Exercises C1.12 and 3.2) we can readily obtain all the ingredients needed to evaluate the CRB matrix in equation (B.3.3) or (B.3.25).

As in the case of line spectra, for relatively large values of N (e.g., on the order of hundreds), the use of the exact CRB formula for rational spectra could be burdensome computationally (given the need to multiply and invert matrices of large dimensions). In such large-sample cases, we might want to use an asymptotically valid approximation of the exact CRB, such as the one developed in [SÖDERSTRÖM AND STOICA 1989]. Below we present such an approximate (large-sample) CRB formula for ARMA parameter estimates.

Let

$$\Lambda = E\left\{\begin{bmatrix}\mathrm{Re}[e(t)]\\ \mathrm{Im}[e(t)]\end{bmatrix}\begin{bmatrix}\mathrm{Re}[e(t)] & \mathrm{Im}[e(t)]\end{bmatrix}\right\} \tag{B.5.1}$$

Typically, the real and imaginary parts of the complex-valued white-noise sequence $\{e(t)\}$ are assumed to be mutually uncorrelated and have the same variance $\sigma^2/2$. In such a case, we have $\Lambda = (\sigma^2/2)I$. However, this assumption is not necessary for the result discussed below to hold; hence, we do not impose it. (In other words, Λ in (B.5.1) is constrained only to be a positive definite matrix.) We should also remark that, for the sake of simplicity, we assumed that the ARMA signal under discussion is scalar. Nevertheless, the extension of the discussion that

follows to multivariate ARMA signals is immediate. Finally, note that, for real-valued signals, the imaginary parts in (B.5.1) (and in equation (B.5.2)) should be omitted.

The real-valued white noise vector in (B.5.1) satisfies the equation

$$\underbrace{\begin{bmatrix} \mathrm{Re}[e(t)] \\ \mathrm{Im}[e(t)] \end{bmatrix}}_{\varepsilon(t)} = \underbrace{\begin{bmatrix} \mathrm{Re}\left[\frac{A(z)}{B(z)}\right] & -\mathrm{Im}\left[\frac{A(z)}{B(z)}\right] \\ \mathrm{Im}\left[\frac{A(z)}{B(z)}\right] & \mathrm{Re}\left[\frac{A(z)}{B(z)}\right] \end{bmatrix}}_{H(z)} \underbrace{\begin{bmatrix} \mathrm{Re}[y(t)] \\ \mathrm{Im}[y(t)] \end{bmatrix}}_{v(t)} \tag{B.5.2}$$

where z^{-1} is to be treated as the unit delay operator (not as a complex variable). As the coefficients of the polynomials $A(z)$ and $B(z)$ in $H(z)$ above are the unknowns in our estimation problem, we can rewrite (B.5.2) in the following form to stress the dependence of $\varepsilon(t)$ on θ:

$$\varepsilon(t, \theta) = H(z, \theta)v(t) \tag{B.5.3}$$

Because the polynomials of the ARMA model are monic by assumption, we have

$$H(z, \theta)|_{z^{-1}=0} = I \qquad \text{(for any } \theta\text{)} \tag{B.5.4}$$

This observation, along with the fact that $\varepsilon(t)$ is white and the "whitening filter" $H(z)$ is stable and causal (which follows from the fact that the complex-valued (equivalent) counterpart of (B.5.2), $e(t) = \frac{A(z)}{B(z)}y(t)$, is stable and causal), implies that (B.5.3) is a standard *prediction error* model, to which the CRB result of [Söderström and Stoica 1989] applies.

Let

$$\Delta(t) = \frac{\partial \varepsilon^T(t, \theta)}{\partial \theta} \tag{B.5.5}$$

($\varepsilon(t, \theta)$ depends on θ via $H(z, \theta)$ only; see (B.5.2)). Then, an asymptotically valid expression for the CRB block corresponding to the parameters in θ is given by

$$\boxed{P_{cr,\theta} = \left(E\left\{\Delta(t)\Lambda^{-1}\Delta^T(t)\right\}\right)^{-1}} \tag{B.5.6}$$

The calculation of the derivative matrix in (B.5.5) is straightforward. The evaluation of the statistical expectation in (B.5.6) can be reduced to ARMA cross-covariance calculations. Equation (B.5.6) does not require handling matrices of large dimensions (on the order of N), so its implementation is much simpler than that of the exact CRB formula.

For some recent results on the CRB for rational spectral analysis, see [Ninness 2003].

B.6 THE CRB FOR SPATIAL SPECTRA

Consider the model (6.2.21) for the output sequence $\{y(t)\}_{t=1}^{N}$ of an array that receives the signals emitted by n narrowband point sources:

$$\begin{aligned} y(t) &= As(t) + e(t) \\ A &= [a(\theta_1), \ldots, a(\theta_n)] \end{aligned} \tag{B.6.1}$$

The noise term, $e(t)$, in (B.6.1) is assumed to be circularly Gaussian distributed, with mean zero and the following covariances:

$$E\left\{e(t)e^*(\tau)\right\} = \sigma^2 I \delta_{t,\tau} \tag{B.6.2}$$

Regarding the signal vector, $s(t)$, in the equation (B.6.1), we can assume that either

Det: $\{s(t)\}$ is a deterministic, unknown sequence

or

Sto: $\{s(t)\}$ is a random sequence that is circularly Gaussian distributed with mean zero and covariances

$$E\left\{s(t)s^*(\tau)\right\} = P\delta_{t,\tau} \tag{B.6.3}$$

Hereafter, the acronyms Det and Sto are used to designate the case of deterministic or stochastic signals, respectively. Note that making one of these two assumptions on $\{s(t)\}$ is similar to assuming in the line-spectral analysis problem that the initial phases $\{\varphi_k\}$ are deterministic or random. (See Section B.4.) As we will see shortly, both the CRB analysis and the resulting CRB formulas depend heavily on which of the two assumptions we make on $\{s(t)\}$. The reader may already wonder which assumption should then be used in a given application. This is not a simple question, and we will be better prepared to answer it after deriving the corresponding CRB formulas.

In Chapter 6, we used the symbol θ to denote the DOA vector. To conform with the notation used in this appendix (and by a slight abuse of notation), we will here let θ denote the *entire* parameter vector.

As explained in Chapter 6, the use of array processing for spatial spectral analysis leads essentially to a parameter estimation problem. Under *the Det assumption* the parameter vector to be estimated is given by

$$\theta = \left[\theta_1, \ldots, \theta_n\,;\, \bar{s}^T(1), \ldots, \bar{s}^T(N)\,;\ldots;\, \tilde{s}^T(1), \ldots, \tilde{s}^T(N)\,;\, \sigma^2\right]^T \tag{B.6.4}$$

whereas under *the Sto assumption*

$$\theta = \left[\theta_1, \ldots, \theta_n\,;\, P_{11}, \bar{P}_{12}, \tilde{P}_{12}, \ldots, \bar{P}_{1n}, \tilde{P}_{1n}, P_{22}, \bar{P}_{23}, \tilde{P}_{23}, \ldots, P_{nn},\,;\, \sigma^2\right]^T \tag{B.6.5}$$

Hereafter, $\bar{s}(t)$ and $\tilde{s}(t)$ denote the real and imaginary parts of $s(t)$, and P_{ij} denotes the (i,j)th element of the matrix P. Furthermore, under both Det and Sto assumptions, the observed array output sample,

$$y(t) = \left[y^T(1), \ldots, y^T(N)\right]^T \tag{B.6.6}$$

is circularly Gaussian distributed with the following mean μ and covariance Γ:

Under Det:

$$\mu = \begin{bmatrix} As(1) \\ \vdots \\ As(N) \end{bmatrix}, \qquad \Gamma = \begin{bmatrix} \sigma^2 I & & 0 \\ & \ddots & \\ 0 & & \sigma^2 I \end{bmatrix} \tag{B.6.7}$$

Under Sto:

$$\mu = 0, \qquad \Gamma = \begin{bmatrix} R & & 0 \\ & \ddots & \\ 0 & & R \end{bmatrix} \tag{B.6.8}$$

where R is given by (see (6.4.3)

$$R = APA^* + \sigma^2 I \tag{B.6.9}$$

The differentiation of either (B.6.7) or (B.6.8) with respect to the elements of the parameter vector θ is straightforward. Use of the so-obtained derivatives of μ and Γ in the general CRB formula in (B.3.25) provides a simple means of computing $\mathrm{CRB}_{\mathrm{Det}}$ and $\mathrm{CRB}_{\mathrm{Sto}}$ for the entire parameter vector θ as defined in (B.6.4) or (B.6.5).

Computing the CRB as just described may be sufficient for many applications. However, sometimes we may need more than just that. For example, we may be interested in using the CRB for the design of array geometry or for getting insights into the various features of a specific spatial spectral analysis scenario. In such cases, we might want to have a closed-form (or analytical) expression for the CRB. More precisely, as the DOAs are usually the parameters of major interest, we often will want a closed-form expression for CRB(DOA) (i.e., the block of the CRB matrix that corresponds to the DOA parameters). Next, we consider the problem of obtaining such a closed-form CRB expression under both the Det and Sto assumptions just made.

First, consider the Det assumption. Let us write the corresponding μ vector in (B.6.7) as

$$\mu = Gs \tag{B.6.10}$$

where

$$G = \begin{bmatrix} A & & 0 \\ & \ddots & \\ 0 & & A \end{bmatrix}, \qquad s = \begin{bmatrix} s(1) \\ \vdots \\ s(N) \end{bmatrix} \tag{B.6.11}$$

Then, a straightforward calculation yields

$$\frac{\partial \mu}{\partial \bar{s}^T} = G, \qquad \frac{\partial \mu}{\partial \bar{s}^T} = iG; \tag{B.6.12}$$

and

$$\frac{\partial \mu}{\partial \theta_k} = \begin{bmatrix} \frac{\partial A}{\partial \theta_k} s(1) \\ \vdots \\ \frac{\partial A}{\partial \theta_k} s(N) \end{bmatrix} = \begin{bmatrix} d_k s_k(1) \\ \vdots \\ d_k s_k(N) \end{bmatrix}, \qquad k = 1, \ldots, n \tag{B.6.13}$$

where $s_k(t)$ is the kth element of $s(t)$ and

$$d_k = \left. \frac{\partial a(\theta)}{\partial \theta} \right|_{\theta=\theta_k} \tag{B.6.14}$$

Using the notation

$$\Delta = \begin{bmatrix} d_1 s_1(1) & \cdots & d_n s_n(1) \\ \vdots & & \vdots \\ d_1 s_1(N) & \cdots & d_n s_n(N) \end{bmatrix}, \qquad (N \times n) \tag{B.6.15}$$

we can then write

$$\frac{d\mu}{d\theta^T} = \begin{bmatrix} \Delta, G, iG, 0 \end{bmatrix} \tag{B.6.16}$$

which gives the following expression for the second term in the general CRB formula in (B.3.25):

$$2 \operatorname{Re} \left\{ \frac{d\mu^*}{d\theta} \Gamma^{-1} \frac{d\mu}{d\theta^T} \right\} = \begin{bmatrix} J & 0 \\ 0 & 0 \end{bmatrix} \tag{B.6.17}$$

In this equation

$$J \triangleq \frac{2}{\sigma^2} \operatorname{Re} \left\{ \begin{bmatrix} \Delta^* \\ G^* \\ -iG^* \end{bmatrix} \begin{bmatrix} \Delta & G & iG \end{bmatrix} \right\} \tag{B.6.18}$$

Furthermore, Γ depends only on σ^2, and

$$\frac{d\Gamma}{d\sigma^2} = \begin{bmatrix} I & & 0 \\ & \ddots & \\ 0 & & I \end{bmatrix}$$

so we can easily verify that the matrix corresponding to the first term in the general CRB formula, (B.3.25), is given by

$$\operatorname{tr}\left[\Gamma^{-1}\Gamma_i'\Gamma^{-1}\Gamma_j'\right] = \begin{bmatrix} 0 & 0 \\ 0 & \frac{mN}{\sigma^4} \end{bmatrix}, \qquad i,j = 1,2,\ldots \tag{B.6.19}$$

Combining (B.6.17) and (B.6.19) yields the following CRB formula for the parameter vector θ in (B.6.4), under the Det assumption:

$$\mathrm{CRB}_{\mathrm{Det}} = \begin{bmatrix} J^{-1} & 0 \\ 0 & \frac{\sigma^4}{mN} \end{bmatrix} \tag{B.6.20}$$

Hence, to obtain the CRB for the DOA subvector of θ, we need to extract the corresponding block of J^{-1}. One convenient way of doing this is by suitably block-diagonalizing the matrix J. To this end, let us introduce the matrix

$$B = (G^*G)^{-1}G^*\Delta \tag{B.6.21}$$

Note that the inverse in (B.6.21) exists, because A^*A is nonsingular by assumption. Also, let

$$F = \begin{bmatrix} I & 0 & 0 \\ -\bar{B} & I & 0 \\ -\tilde{B} & 0 & I \end{bmatrix} \tag{B.6.22}$$

where $\bar{B} = \mathrm{Re}\{B\}$ and $\tilde{B} = \mathrm{Im}\{B\}$. It can be verified that

$$\begin{bmatrix}\Delta & G & iG\end{bmatrix}F = \begin{bmatrix}(\Delta - GB) & G & iG\end{bmatrix} = \begin{bmatrix}\Pi_G^{\perp}\Delta & G & iG\end{bmatrix} \tag{B.6.23}$$

where

$$\Pi_G^{\perp} = I - G(G^*G)^{-1}G^*$$

is the orthogonal projector onto the null space of G^* (see Result R17 in Appendix A); in particular, observe that $G^*\Pi_G^{\perp} = 0$. It follows from (B.6.18) and (B.6.23) that

$$\begin{aligned} F^TJF &= \frac{2}{\sigma^2}\,\mathrm{Re}\left\{F^*\begin{bmatrix}\Delta^* \\ G^* \\ -iG^*\end{bmatrix}\begin{bmatrix}\Delta & G & iG\end{bmatrix}F\right\} \\ &= \frac{2}{\sigma^2}\,\mathrm{Re}\left\{\begin{bmatrix}\Delta^*\Pi_G^{\perp} \\ G^* \\ -iG^*\end{bmatrix}\begin{bmatrix}\Pi_G^{\perp}\Delta & G & iG\end{bmatrix}\right\} \\ &= \frac{2}{\sigma^2}\,\mathrm{Re}\left\{\begin{bmatrix}\Delta^*\Pi_G^{\perp}\Delta & 0 & 0 \\ 0 & G^*G & iG^*G \\ 0 & -iG^*G & G^*G\end{bmatrix}\right\} \end{aligned} \tag{B.6.24}$$

and hence, that the CRB matrix for the DOAs and the signal sequence is given by

$$\begin{aligned} J^{-1} &= F\left(F^T J F\right)^{-1} F^T \\ &= \frac{\sigma^2}{2}\begin{bmatrix} I & 0 & 0 \\ -\bar{B} & I & 0 \\ -\tilde{B} & 0 & I \end{bmatrix}\begin{bmatrix} \left[\mathrm{Re}(\Delta^* \Pi_G^\perp \Delta)\right]^{-1} & 0 & 0 \\ 0 & x & x \\ 0 & x & x \end{bmatrix}\begin{bmatrix} I & -\bar{B}^T & -\tilde{B}^T \\ 0 & I & 0 \\ 0 & 0 & I \end{bmatrix} \\ &= \begin{bmatrix} \frac{\sigma^2}{2}\left[\mathrm{Re}(\Delta^* \Pi_G^\perp \Delta)\right]^{-1} & x & x \\ x & x & x \\ x & x & x \end{bmatrix} \end{aligned} \tag{B.6.25}$$

where we used the symbol x to denote a block of no interest in the derivation. From (B.6.4) and (B.6.25), we can immediately see that the CRB matrix for the DOAs is given by

$$\mathrm{CRB}_{\mathrm{Det}}(\mathrm{DOA}) = \frac{\sigma^2}{2}\left[\mathrm{Re}(\Delta^* \Pi_G^\perp \Delta)\right]^{-1} \tag{B.6.26}$$

It is possible to rewrite (B.6.26) in a more convenient form. To do so, we note that

$$\Pi_G^\perp = \begin{bmatrix} I & & 0 \\ & \ddots & \\ 0 & & I \end{bmatrix} - \begin{bmatrix} \Pi_A & & 0 \\ & \ddots & \\ 0 & & \Pi_A \end{bmatrix} = \begin{bmatrix} \Pi_A^\perp & & 0 \\ & \ddots & \\ 0 & & \Pi_A^\perp \end{bmatrix} \tag{B.6.27}$$

and, hence, that

$$\begin{aligned} \left[\Delta^* \Pi_G^\perp \Delta\right]_{kp} &= \sum_{t=1}^{N} d_k^* s_k^*(t) \Pi_A^\perp d_p s_p(t) \\ &= N\left[d_k^* \Pi_A^\perp d_p\right]\left[\frac{1}{N}\sum_{t=1}^{N} s_p(t) s_k^*(t)\right] \\ &= N\left[D^* \Pi_A^\perp D\right]_{kp}\left[\hat{P}^T\right]_{kp} \end{aligned} \tag{B.6.28}$$

where

$$D = \begin{bmatrix} d_1 & \dots & d_n \end{bmatrix} \tag{B.6.29}$$

$$\hat{P} = \frac{1}{N}\sum_{t=1}^{N} s(t) s^*(t) \tag{B.6.30}$$

It follows from (B.6.28) that

$$\Delta^* \Pi_G^\perp \Delta = N\ \left(D^* \Pi_A^\perp D\right) \odot \hat{P}^T \tag{B.6.31}$$

where $\odot$ denotes the Hadamard (or elementwise) matrix product, defined in Result R19 in Appendix A. Inserting (B.6.31) in (B.6.26) yields the following analytical expression for *the CRB matrix associated with the DOA vector under the Det assumption:*

$$\boxed{\mathrm{CRB}_{\mathrm{Det}}(\mathrm{DOA}) = \frac{\sigma^2}{2N}\left\{\mathrm{Re}\left[\left(D^*\Pi_A^{\perp}D\right)\odot\hat{P}^T\right]\right\}^{-1}} \tag{B.6.32}$$

We refer the reader to [Stoica and Nehorai 1989a] for more details about (B.6.32) and its possible uses in array processing. The presented derivation of (B.6.32) has been adapted from [Stoica and Larsson 2001]. Note that (B.6.32) can be applied directly to the temporal line-spectral model in Section B.4 (see equations (B.4.4) and (B.4.5)) to obtain an analytical CRB formula for the sinusoidal frequencies.

The derivation of an analytical expression for *the CRB matrix associated with the DOAs under the Sto assumption* is more intricate, and we give only the final formula here (see [Stoica, Larsson, and Gershman 2001] and its references for a derivation):

$$\boxed{\mathrm{CRB}_{\mathrm{Sto}}(\mathrm{DOA}) = \frac{\sigma^2}{2N}\left\{\mathrm{Re}\left[\left(D^*\Pi_A^{\perp}D\right)\odot\left(PA^*R^{-1}AP\right)^T\right)\right]\right\}^{-1}} \tag{B.6.33}$$

At this point, we should emphasize the fact that the two CRBs, $\mathrm{CRB}_{\mathrm{Det}}$ and $\mathrm{CRB}_{\mathrm{Sto}}$, correspond to two different models of the data vector y (see (B.6.7) and (B.6.8)); hence, they are *not* directly comparable. On the other hand, the CRBs for the DOA parameters can be compared with one another. To make this comparison possible, let us introduce the assumption that the sample covariance matrix $\hat{P}$ in (B.6.30) converges to the P matrix in (B.6.3), as $N \to \infty$. Let $\overline{\mathrm{CRB}}_{\mathrm{Det}}(\mathrm{DOA})$ denote the CRB matrix in (B.6.32) with $\hat{P}$ replaced by P. Then, the following interesting order relation holds true:

$$\boxed{\mathrm{CRB}_{\mathrm{Sto}}(\mathrm{DOA}) \geq \overline{\mathrm{CRB}}_{\mathrm{Det}}(\mathrm{DOA})} \tag{B.6.34}$$

To prove (B.6.34), we need to show (see (B.6.32) and (B.6.33)) that

$$\left\{\mathrm{Re}\left[\left(D^*\Pi_A^{\perp}D\right)\odot\left(PA^*R^{-1}AP\right)^T\right)\right]\right\}^{-1} \geq \left\{\mathrm{Re}\left[\left(D^*\Pi_A^{\perp}D\right)\odot P^T\right]\right\}^{-1}$$

or, equivalently, that

$$\mathrm{Re}\left[\left(D^*\Pi_A^{\perp}D\right)\odot\left(P - PA^*R^{-1}AP\right)^T\right)\right] \geq 0 \tag{B.6.35}$$

The real part of a positive semidefinite matrix is positive semidefinite itself:

$$H \geq 0 \implies \mathrm{Re}[H] \geq 0 \tag{B.6.36}$$

(Indeed, for any real-valued vector h we have: $h^* \operatorname{Re}[H]h = \operatorname{Re}[h^*Hh] \geq 0$ for $H \geq 0$.) Combining this observation with Result R19 in Appendix A shows that, to prove (B.6.35), it is sufficient to verify that

$$P \geq PA^*R^{-1}AP \tag{B.6.37}$$

or, equivalently,

$$I \geq P^{1/2}A^*R^{-1}AP^{1/2} \tag{B.6.38}$$

where $P^{1/2}$ denotes the Hermitian square root of P; see Definition D12 in Appendix A. Let

$$Z = AP^{1/2}$$

Then (B.6.38) can be rewritten as

$$I - Z^*\left(ZZ^* + \sigma^2 I\right)^{-1} Z \geq 0 \tag{B.6.39}$$

To prove (B.6.39), we use the fact that the following matrix is evidently positive semidefinite:

$$\begin{bmatrix} I & Z^* \\ Z & ZZ^* + \sigma^2 I \end{bmatrix} = \begin{bmatrix} I \\ Z \end{bmatrix} \begin{bmatrix} I & Z^* \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 0 & \sigma^2 I \end{bmatrix} \geq 0 \tag{B.6.40}$$

and therefore

$$\begin{bmatrix} I & -Z^*\left(ZZ^* + \sigma^2 I\right)^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} I & Z^* \\ Z & ZZ^* + \sigma^2 I \end{bmatrix} \begin{bmatrix} I & 0 \\ -\left(ZZ^* + \sigma^2 I\right)^{-1} Z & I \end{bmatrix}$$

$$= \begin{bmatrix} I - Z^*\left(ZZ^* + \sigma^2 I\right)^{-1} Z & 0 \\ 0 & ZZ^* + \sigma^2 I \end{bmatrix} \geq 0 \tag{B.6.41}$$

The inequality in (B.6.39) is a simple consequence of (B.6.41), and so the proof of (B.6.34) is concluded.

To understand (B.6.34) at an *intuitive level*, we note that the ML method for DOA estimation under the Sto assumption, $\mathrm{ML}_{\mathrm{Sto}}$, can be shown to achieve $\mathrm{CRB}_{\mathrm{Sto}}(\mathrm{DOA})$ (for sufficiently large values of N). (See, e.g., [Stoica and Nehorai 1990] and [Ottersten, Viberg, Stoica, and Nehorai 1993].) This result should in fact be no surprise, because the general ML method of parameter estimation is known to be asymptotically statistically efficient (i.e., it achieves the CRB as $N \to \infty$) under some regularity conditions that are satisfied in the Sto assumption case. Specifically, the regularity conditions require that the number of unknown parameters not increase as N increases, as is indeed true for the Sto model (see (B.6.5)). Let $\mathrm{CML}_{\mathrm{Sto}}(\mathrm{DOA})$ denote the asymptotic covariance matrix of the $\mathrm{ML}_{\mathrm{Sto}}$ estimate of the DOA parameter vector. According to the preceding discussion, we have that

$$\mathrm{CML}_{\mathrm{Sto}}(\mathrm{DOA}) = \mathrm{CRB}_{\mathrm{Sto}}(\mathrm{DOA}) \tag{B.6.42}$$

At the same time, under the Det assumption, the $\mathrm{ML}_{\mathrm{Sto}}$ can be viewed as *some* method for DOA estimation, and hence its asymptotic covariance matrix must satisfy the CRB inequality (corresponding to the Det assumption):

$$\mathrm{CML}_{\mathrm{Sto}}(\mathrm{DOA}) \geq \overline{\mathrm{CRB}}_{\mathrm{Det}}(\mathrm{DOA}) \tag{B.6.43}$$

(Note that the asymptotic covariance matrix of $\mathrm{ML}_{\mathrm{Sto}}$ can be shown to be the same under either the Sto or Det assumption.) This equation, along with (B.6.42), provides a heuristic motivation for the relationship between $\mathrm{CRB}_{\mathrm{Sto}}(\mathrm{DOA})$ and $\overline{\mathrm{CRB}}_{\mathrm{Det}}(\mathrm{DOA})$ in (B.6.34). Note that the inequality in (B.6.34) is, in general, *strict*, but the relative difference between $\mathrm{CRB}_{\mathrm{Sto}}(\mathrm{DOA})$ and $\overline{\mathrm{CRB}}_{\mathrm{Det}}(\mathrm{DOA})$ is usually fairly small. (See, e.g., [Ottersten, Viberg, Stoica, and Nehorai 1993].)

A remark similar to the one in the previous paragraph can be made on the ML method for DOA estimation under the Det assumption, which we abbreviate as $\mathrm{ML}_{\mathrm{Det}}$. Note that $\mathrm{ML}_{\mathrm{Det}}$ can be readily seen to coincide with the *NLS method* discussed in Section 6.4.1. Under the Sto assumption, $\mathrm{ML}_{\mathrm{Det}}$ (i.e., the NLS method) can be viewed as just *some* method for DOA estimation. Hence, its (asymptotic) covariance matrix must be bounded below by the CRB corresponding to the Sto assumption:

$$\mathrm{CML}_{\mathrm{Det}}(\mathrm{DOA}) \geq \mathrm{CRB}_{\mathrm{Sto}}(\mathrm{DOA}) \tag{B.6.44}$$

Like $\mathrm{ML}_{\mathrm{Sto}}$, the asymptotic covariance matrix of $\mathrm{ML}_{\mathrm{Det}}$ can also be shown to be the same under either the Sto or Det assumption. Hence, we can infer from (B.6.34) and (B.6.44) that *ML_{Det} does not attain $\overline{CRB}_{Det}(DOA)$*, as is indeed the case (as is shown in, e.g., [Stoica and Nehorai 1989a]). To understand why this happens, note that the Det model contains $(2N + 1)n + 1$ real-valued parameters (see (B.6.4)), which must be estimated from $2mN$ data samples. Hence, for large N, the ratio between the number of unknown parameters and the available data samples approaches a constant (equal to n/m), which violates one of the aforementioned regularity conditions for the statistical efficiency of the ML method.

Remark: $\mathrm{CRB}_{\mathrm{Det}}(\mathrm{DOA})$ depends on the signal sequence $\{s(t)\}_{t=1}^{N}$. However, neither $\overline{\mathrm{CRB}}_{\mathrm{Det}}(\mathrm{DOA})$ nor the asymptotic covariance matrix of $\mathrm{ML}_{\mathrm{Sto}}$, of $\mathrm{ML}_{\mathrm{Det}}$, or, in fact, of many other DOA estimation methods depends on this sequence. We will use the symbol C to denote the (asymptotic) covariance matrix of such a DOA estimation method for which C is independent of the signal sequence.

From $\mathrm{CRB}_{\mathrm{Det}}(\mathrm{DOA})$ we can obtain a matrix, different from $\overline{\mathrm{CRB}}_{\mathrm{Det}}(\mathrm{DOA})$, which is independent of the signal sequence, in the following manner:

$$\mathrm{ACRB}_{\mathrm{Det}}(\mathrm{DOA}) = \tilde{E}\left\{\mathrm{CRB}_{\mathrm{Det}}(\mathrm{DOA})\right\} \tag{B.6.45}$$

Here $\tilde{E}$ is an averaging operator and $\mathrm{ACRB}_{\mathrm{Det}}$ stands for Averaged $\mathrm{CRB}_{\mathrm{Det}}$. For example, $\tilde{E}\{\cdot\}$ in (B.6.45) can be a simple arithmetic averaging of $\mathrm{CRB}_{\mathrm{Det}}(\mathrm{DOA})$ over a set of signal sequences. Using the fact that $\tilde{E}\{C\} = C$ (because C does not depend on the sequence $\{s(t)\}_{t=1}^{N}$), we can apply the operator $\tilde{E}\{\cdot\}$ to both sides of the CRB inequality

$$C \geq \mathrm{CRB}_{\mathrm{Det}}(\mathrm{DOA}) \tag{B.6.46}$$

to obtain

$$C \geq \mathrm{ACRB}_{\mathrm{Det}}(\mathrm{DOA}) \tag{B.6.47}$$

(Note that the inequality in (B.6.47) and, hence, that in (B.6.47), hold at least for sufficiently large values of N.) It follows from (B.6.47) that $\mathrm{ACRB}_{\mathrm{Det}}(\mathrm{DOA})$ can also be used as a lower bound on the DOA estimation error covariance. Furthermore, it can be shown that $\mathrm{ACRB}_{\mathrm{Det}}(\mathrm{DOA})$ is *tighter* than $\overline{\mathrm{CRB}}_{\mathrm{Det}}(\mathrm{DOA})$:

$$\boxed{\mathrm{ACRB}_{\mathrm{Det}}(\mathrm{DOA}) \geq \overline{\mathrm{CRB}}_{\mathrm{Det}}(\mathrm{DOA})} \tag{B.6.48}$$

To prove (B.6.48), we introduce the matrix

$$X = \frac{2N}{\sigma^2} \mathrm{Re}\left[(D^* \Pi_A^{\perp} D) \odot \hat{P}^T \right] \tag{B.6.49}$$

Using this notation, along with the fact that $\tilde{E}\{\hat{P}\} = P$ (which holds under mild conditions), we can rewrite (B.6.48) as follows:

$$\boxed{\tilde{E}\left\{X^{-1}\right\} \geq \left[\tilde{E}\{X\}\right]^{-1}} \tag{B.6.50}$$

To prove (B.6.50), we note that the matrix

$$\tilde{E}\left\{ \begin{bmatrix} X^{-1} & I \\ I & X \end{bmatrix} \right\} = \tilde{E}\left\{ \begin{bmatrix} X^{-1/2} \\ X^{1/2} \end{bmatrix} \begin{bmatrix} X^{-1/2} & X^{1/2} \end{bmatrix} \right\}$$

(where $X^{1/2}$ and $X^{-1/2}$ denote the Hermitian square roots of X and X^{-1}, respectively) is clearly positive semidefinite, and therefore so must be the following matrix:

$$\begin{aligned} &\begin{bmatrix} I & -\left[\tilde{E}\{X\}\right]^{-1} \\ 0 & I \end{bmatrix} \begin{bmatrix} \tilde{E}\left\{X^{-1}\right\} & I \\ I & \tilde{E}\{X\} \end{bmatrix} \begin{bmatrix} I & 0 \\ -\left[\tilde{E}\{X\}\right]^{-1} & I \end{bmatrix} \\ &\quad = \begin{bmatrix} \tilde{E}\left\{X^{-1}\right\} - \left[\tilde{E}\{X\}\right]^{-1} & 0 \\ 0 & \tilde{E}\{X\} \end{bmatrix} \geq 0 \end{aligned} \tag{B.6.51}$$

The matrix inequality in (B.6.50), which is somewhat similar to the scalar Jensen inequality (see, e.g., Complement 4.9.5) readily follows from (B.6.51).

The inequality (B.6.48) looks appealing. On the other hand, $\mathrm{ACRB}_{\mathrm{Det}}(\mathrm{DOA})$ should be *less tight* than $\mathrm{CRB}_{\mathrm{Sto}}(\mathrm{DOA})$, in view of the results in (B.6.42) and (B.6.47). Also, $\mathrm{CRB}_{\mathrm{Sto}}(\mathrm{DOA})$ has a simpler analytical form. Hence, we may have little reason to use $\mathrm{ACRB}_{\mathrm{Det}}(\mathrm{DOA})$ in lieu of $\mathrm{CRB}_{\mathrm{Sto}}(\mathrm{DOA})$. Despite these drawbacks of $\mathrm{ACRB}_{\mathrm{Det}}(\mathrm{DOA})$, we have included this discussion for the potential usefulness of the inequality in (B.6.50) and of the basic idea behind the introduction of $\mathrm{ACRB}_{\mathrm{Det}}(\mathrm{DOA})$. ■

In the remainder of this section, we rely on the previous results to compare the Det and Sto model assumptions, to discuss the consequences of making these assumptions, and to draw some conclusions.

First, consider the array output model in equation (B.6.1). To derive the ML estimates of the unknown parameters in (B.6.1), we must make some assumptions on the signal sequence $\{s(t)\}$. The $\mathrm{ML}_{\mathrm{Sto}}$ method for DOA estimation (derived under the Sto assumption) turns out to be more accurate than the $\mathrm{ML}_{\mathrm{Det}}$ method (obtained under the Det assumption), under quite general conditions on $\{s(t)\}$. However, the $\mathrm{ML}_{\mathrm{Sto}}$ method is somewhat more complicated computationally than the $\mathrm{ML}_{\mathrm{Det}}$ method; see, e.g., [Ottersten, Viberg, Stoica, and Nehorai 1993].

The previous discussion implies that the question about which assumption should be used (because "it is more likely to be true") is in fact irrelevant in this case. Indeed, we should see the two assumptions only as instruments for deriving the two corresponding ML methods. Once we have completed the derivations, the assumption issue is no longer important, and we can simply *choose the ML method that we prefer, regardless of the nature of* $\{s(t)\}$. The choice should be based on the facts that (a) $\mathrm{ML}_{\mathrm{Det}}$ is computationally simpler than $\mathrm{ML}_{\mathrm{Sto}}$, and (b) $\mathrm{ML}_{\mathrm{Sto}}$ is statistically more accurate than $\mathrm{ML}_{\mathrm{Det}}$ under quite general conditions on $\{s(t)\}$.

Second, regarding the two CRB matrices that correspond to the Det and Sto assumptions, respectively, we can argue as follows: Under the Sto assumption, $\mathrm{CRB}_{\mathrm{Sto}}(\mathrm{DOA})$ is *the* Cramér–Rao bound and, hence, the lower bound to use. Under the Det assumption, $\mathrm{CRB}_{\mathrm{Sto}}(\mathrm{DOA})$ is no longer the true CRB, but it is still a tight lower bound on the asymptotic covariance matrix of any known DOA estimation method. $\overline{\mathrm{CRB}}_{\mathrm{Det}}(\mathrm{DOA})$ is also a lower bound, but it is not tight. Hence, *CRB_{Sto}(DOA) should be the normal choice for a lower bound, regardless of the assumption (Det or Sto) that the signal sequence is likely to satisfy*. Note that, under the Det assumption, $\mathrm{ML}_{\mathrm{Sto}}$ can be seen as *some* DOA estimation method. Therefore, in principle, a better DOA estimation method than $\mathrm{ML}_{\mathrm{Sto}}$ could exist (where by "better" we mean that the covariance matrix of such an estimation method would be smaller than $\mathrm{CRB}_{\mathrm{Sto}}(\mathrm{DOA})$). However, no such DOA estimation method appears to be available, in spite of a significant literature on the so-called problem of "estimation in the presence of many nuisance parameters," of which the DOA estimation problem under the Det assumption is a special case.