

Chapter 8

Parameter Estimation I: Maximum Likelihood

8.1 Introduction

In Chapters 8 and 9, we consider the problem in which the parameter (or parameters) enters into the received signal in a nonlinear manner.

An important problem in array processing where the parameter is embedding the received waveform in a nonlinear manner is the case of a plane wave with an unknown wavenumber arriving at the array. For example, in the narrowband snapshot model with a linear array, the received snapshots are

$$\mathbf{x}(k) = \mathbf{v}(\psi_s)f(k) + \mathbf{n}(k), \quad k = 1, 2, \dots, K, \quad (8.1)$$

and the parameter ψ_s is unknown. For D plane waves,

$$\mathbf{x}(k) = \mathbf{V}(\psi_s)\mathbf{f}(k) + \mathbf{n}(k), \quad k = 1, 2, \dots, K, \quad (8.2)$$

where

$$\psi_s = \begin{bmatrix} \psi_1 & \psi_2 & \cdots & \psi_D \end{bmatrix}^T, \quad (8.3)$$

is an unknown D -dimensional vector. In our initial discussion, we assume D is known. In Section 7.8, techniques for estimating D were discussed. The topic is revisited in Section 8.8 in the context of parameter estimation.

For an arbitrary array with a single plane-wave input, the narrowband snapshot model is

$$\mathbf{x}(k) = \mathbf{v}(\psi_s)f(k) + \mathbf{n}(k), \quad k = 1, 2, \dots, K, \quad (8.4)$$

where

$$\boldsymbol{\psi}_s = \begin{bmatrix} \psi_x & \psi_y \end{bmatrix}^T. \quad (8.5)$$

is a 2-D vector.

Alternatively, we could write (8.4) as

$$\mathbf{x}(k) = \mathbf{v}(\theta, \phi)f(k) + \mathbf{n}(k), \quad k = 1, 2, \dots, K, \quad (8.6)$$

and estimate θ and ϕ .

The problem of estimating the wavenumber or angle of arrival of a plane wave (or multiple plane waves) is commonly referred to as the direction finding (DF) or direction of arrival (DOA) estimation problem. It is important in radar, sonar, seismic systems, electronic surveillance, medical diagnosis and treatment, and radio astronomy. Because of its widespread application and the difficulty of obtaining the optimum estimator, the topic has received a significant amount of attention over the last several decades. Many of our results and examples will focus on the direction finding problem.

However, it is important to note that there are many other nonlinear parameter estimation problems of interest in the array processing area. A representative list includes:

- (i) Estimating the position of the sensors in array and/or their gain and phase characteristics (the calibration problem);
- (ii) Using an AR or ARMA model for the signals' spatial (and/or temporal) characteristics and estimating the model parameters. Alternatively, we may need to estimate the noise/interference parameters;
- (iii) Estimating the range of a target in the near field.

Other parameter estimation problems will arise in the course of our development.

Our discussion considers parameters that are constant during the observation period. The extension to the problem of tracking a time-varying parameter (i.e., a moving target) is of obvious interest, but would take us too far afield.

Our discussion of the parameter estimation problem is divided into two chapters. In Chapter 8, we focus on maximum likelihood (ML) and maximum *a posteriori* probability (MAP) estimators and on bounds on the performance of any estimator. In Chapter 9, we develop a number of other estimation procedures that are computationally simpler than the ML estimator and, in many cases, provide adequate performance.

In Section 8.2, we review several classical estimation results that are used in subsequent sections of the chapter. In Section 8.3, we describe the parameter estimation model that we used in the subsequent discussion.

In Section 8.4, we derive the Cramér-Rao bound (CRB) for the multiple-parameter DOA estimation problem. We recall that the ML estimate approaches this bound under certain conditions. However, we observed in the scalar case that, as the *SNR* and/or the number of snapshots is decreased, the estimators exhibit a threshold phenomenon and the variance (or mean square error) increases rapidly above the Cramér-Rao bound.

In Section 8.5, we consider the problem of estimating the direction-of-arrivals (DOAs) or wavenumbers of D plane waves in the presence of additive Gaussian noise. We first develop two maximum likelihood estimates. The first estimate assumes the source signals are sample functions from Gaussian random processes. We refer to this estimate as the unconditional (or stochastic) maximum likelihood (UML) estimate. We show that its performance approaches the Cramér-Rao bound asymptotically. The second estimate assumes the source signals are unknown, but nonrandom, signals. We develop an estimator that is referred to as the conditional (or deterministic) maximum likelihood (CML) estimate, analyze its performance, and compare its performance to that of the UML estimator and the Cramér-Rao bound. The solution for the maximum likelihood estimator is computationally complex. We develop several other multidimensional estimators that have similar asymptotic performance, but are easier to implement.

In Section 8.6, we develop various computational algorithms that enable us to find the estimators in an efficient manner. The techniques in this section are valid for arbitrary array geometries.

In Section 8.7, we restrict our attention to standard linear arrays and develop a polynomial parameterization of the estimation problem. We then develop efficient estimation procedures using this polynomial representation.

The discussion in Sections 8.2 through 8.7 assumes that the dimension of the parameter vector that we are estimating is known. In Section 8.8, we review techniques for estimating the dimension of the parameter vector. The detection algorithms that we developed in Section 7.8 are directly applicable to the estimation problem. We also discuss algorithms that jointly detect the number of signals and estimate their DOAs.

In Section 8.9, we consider spatially spread signals, which can be characterized by a parameter vector of a reasonable dimension. We first consider the model in Section 5.3 in which the source is characterized by a spectral distribution $S_o(\omega_0 : \theta, \phi)$ on a large sphere. We utilize a parametric model for $S_o(\omega_0 : \theta, \phi)$ and find the maximum likelihood estimate of the

parameters. We next consider the parametric wavenumber models (AR and ARMA models) of Section 5.6 and discuss maximum likelihood estimates of the parameters.

In Section 8.10, we study parameter estimation in beamspace. Just as in the adaptive beamforming, we find that operating in beamspace provides a reduction in computational complexity and certain performance improvements.

In Section 8.11, we consider the impact on the estimation performance when there are sensor gain and phase errors or errors in the location of the sensors. This is the sensitivity problem that we encountered previously in our classical array discussion in Chapter 2 and our beamformer discussion in Chapter 7. We first develop a hybrid Cramér-Rao bound that indicates how the perturbations affect the variance of the DOA estimates. We then analyze the behavior of the ML estimates in the presence of perturbation. Finally, we develop techniques to jointly estimate the DOA parameters and the array parameters (such as sensor location, gain, and phase). The estimation of the array parameters is referred to as the calibration problem.

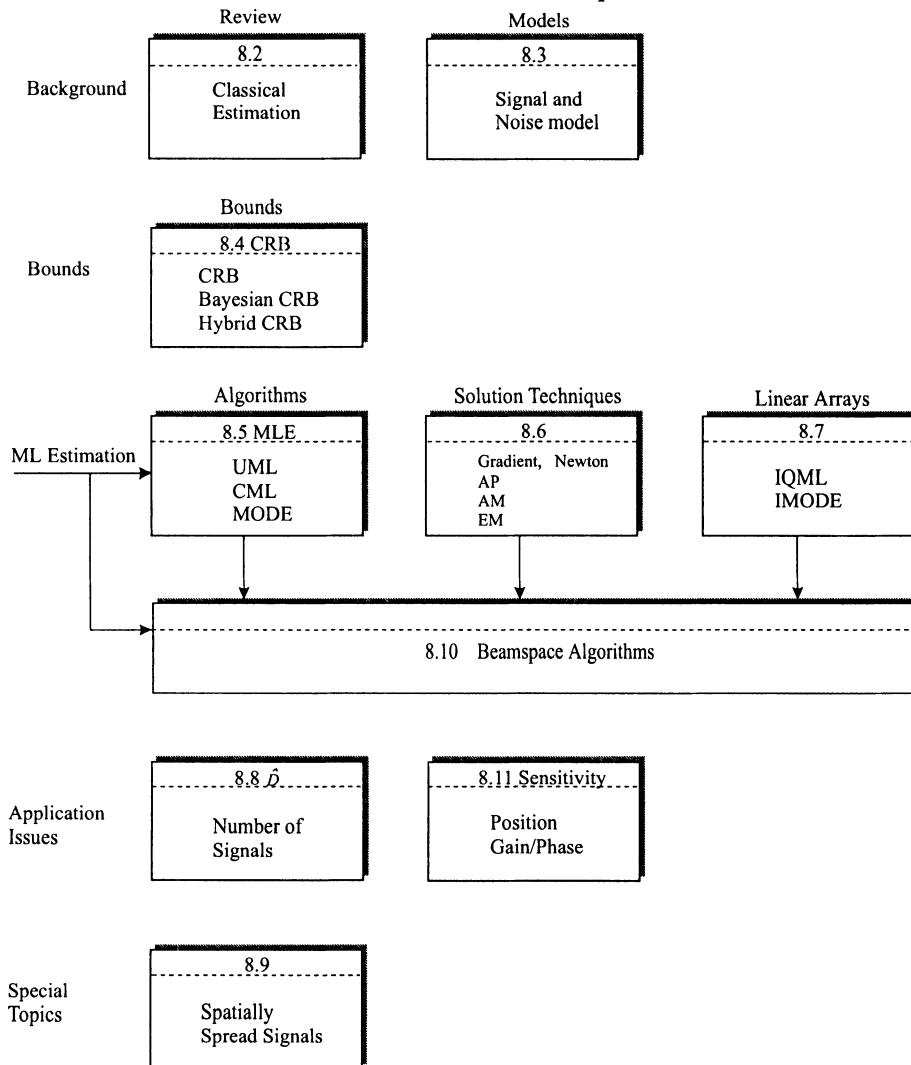
In Section 8.12, we summarize our results and discuss some related topics. In Table 8.1, we show the structure of the chapter.

8.2 Maximum Likelihood and Maximum *a posteriori* Estimators

One of the estimators that we will use in many applications is the maximum likelihood estimate of a vector parameter θ . In the discussion in this section, θ is an arbitrary vector parameter and the notation θ does not denote an angle. We recall from our discussion in Section 2.4 (pp. 52–86) of DEMT I [VT68] that we can bound the variance of any unbiased estimator by the Cramér-Rao bound. We can also show that, under conditions that are often encountered in practice, the ML estimator is unbiased and, asymptotically (as $K \rightarrow \infty$) its variance approaches the Cramér-Rao bound. Thus, the ML estimator is an efficient estimator. This asymptotic behavior is one of the motivations for the use of ML estimator. Moreover, we find that the ML estimator also exhibits good performance in the non-asymptotic region.

When θ is a random variable we use maximum *a posteriori* probability (MAP) estimators. We can bound the MSE performance of any estimator using the Bayesian version of the Cramér-Rao Bound (see p. 84 of DEMT I [VT68], [VT01a]).

Table 8.1 Structure of Chapter 8



In many of the applications that we consider, the model reduces to the case in which the observation is a complex Gaussian random vector \mathbf{x} whose mean $\mathbf{m}_x(\boldsymbol{\theta})$ and covariance $\mathbf{K}_x(\boldsymbol{\theta})$ depend on the vector parameter $\boldsymbol{\theta}$ that we want to estimate. In view of its widespread usage in subsequent sections, it is worthwhile to review some results from classical estimation theory before proceeding to the physical problems of interest.¹

¹The reader may want to review pp. 52–86 of DEMT I [VT68], [VT01a] as background

In Section 8.2.1, we derive the ML estimator for the Gaussian observation problem. In Section 8.2.2, we derive the MAP estimator for the Gaussian observation problem.

8.2.1 Maximum Likelihood (ML) Estimator

For an N -element array, the probability density for a single snapshot is

$$p_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}) = \frac{1}{\det[\pi \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})]} \exp \left\{ -(\mathbf{x}^H - \mathbf{m}_{\mathbf{x}}^H(\boldsymbol{\theta})) \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) (\mathbf{x} - \mathbf{m}_{\mathbf{x}}(\boldsymbol{\theta})) \right\}, \quad (8.7)$$

where \mathbf{x} is an $N \times 1$ complex Gaussian random variable and $\boldsymbol{\theta}$ is a $D \times 1$ nonrandom unknown vector that we want to estimate.

Assuming that successive snapshots are statistically independent, the joint probability density for K snapshots is

$$\begin{aligned} p_{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K | \boldsymbol{\theta}}(\mathbf{x}) &= \prod_{k=1}^K \frac{1}{\det[\pi \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})]} \\ &\times \exp \left\{ - \left(\mathbf{x}_k^H - \mathbf{m}_{\mathbf{x}}^H(\boldsymbol{\theta}) \right) \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) (\mathbf{x}_k - \mathbf{m}_{\mathbf{x}}(\boldsymbol{\theta})) \right\}. \end{aligned} \quad (8.8)$$

The log-likelihood function is

$$\begin{aligned} L_{\mathbf{x}}(\boldsymbol{\theta}) &= \ln p_{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_K | \boldsymbol{\theta}}(\mathbf{x}) \\ &= -K \ln \det[\mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})] - \sum_{k=1}^K \left(\mathbf{x}_k^H - \mathbf{m}_{\mathbf{x}}^H(\boldsymbol{\theta}) \right) \\ &\quad \times \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) (\mathbf{x}_k - \mathbf{m}_{\mathbf{x}}(\boldsymbol{\theta})) - KN \ln \pi. \end{aligned} \quad (8.9)$$

In most cases of interest to us, either $\mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})$ or $\mathbf{m}_{\mathbf{x}}(\boldsymbol{\theta})$ (but not both) are functions of $\boldsymbol{\theta}$. For notational simplicity in Sections 8.2.1 and 8.2.2, we will assume $\mathbf{m}_{\mathbf{x}}(\boldsymbol{\theta})$ is not a function of $\boldsymbol{\theta}$. Then we can let $\mathbf{m}_{\mathbf{x}} = \mathbf{0}$ without loss of generality.

Dropping the last term because it does not depend on $\boldsymbol{\theta}$ and dividing through by K , we have

$$L(\boldsymbol{\theta}) = - \left[\ln \det[\mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})] + \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k^H \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \mathbf{x}_k \right]. \quad (8.10)$$

(We omit the subscript on $L(\boldsymbol{\theta})$ because we have dropped the constant terms.) We can write the second term as

for this discussion.

$$\begin{aligned}
\frac{1}{K} \sum_{k=1}^K \mathbf{x}_k^H \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \mathbf{x}_k &= \text{tr} \left[\frac{1}{K} \sum_{k=1}^K \mathbf{x}_k^H \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \mathbf{x}_k \right] \\
&= \text{tr} \left[\mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \cdot \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k \mathbf{x}_k^H \right] \\
&= \text{tr} \left[\mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \mathbf{C}_{\mathbf{x}} \right], \tag{8.11}
\end{aligned}$$

where $\mathbf{C}_{\mathbf{x}}$ is the sample correlation matrix defined in Section 7.2.1.

$$\mathbf{C}_{\mathbf{x}} \triangleq \frac{1}{K} \sum_{k=1}^K \mathbf{x}_k \mathbf{x}_k^H. \tag{8.12}$$

Then (8.10) can be written as

$$L(\boldsymbol{\theta}) = - \left\{ \ln \det[\mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})] + \text{tr} \left[\mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \mathbf{C}_{\mathbf{x}} \right] \right\}. \tag{8.13}$$

The ML estimate is given by the value of $\boldsymbol{\theta}$ that maximizes $L(\boldsymbol{\theta})$. In general, we must conduct a search procedure to find that value. One of the topics that we study in Section 8.7 is efficient procedures for finding the maximum in the D -dimensional space.

A necessary, but not sufficient, condition is that

$$\frac{\partial}{\partial \theta_i} [L(\boldsymbol{\theta})]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ml}} = 0, \quad i = 1, 2, \dots, D. \tag{8.14}$$

This can be written in more compact form as

$$\nabla_{\boldsymbol{\theta}} [L(\boldsymbol{\theta})]_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ml}} = \mathbf{0}, \tag{8.15}$$

where $\nabla_{\boldsymbol{\theta}}$ is the $D \times 1$ derivative matrix defined in (A.373),

$$\nabla_{\boldsymbol{\theta}} = \begin{bmatrix} \frac{\partial}{\partial \theta_1} \\ \frac{\partial}{\partial \theta_2} \\ \vdots \\ \frac{\partial}{\partial \theta_D} \end{bmatrix}. \tag{8.16}$$

Using the derivative formulas in Section A.7 ((A.400) and (A.393)),

$$\begin{aligned}
\frac{\partial L(\boldsymbol{\theta})}{\partial \theta_i} &= -\text{tr} \left[\mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} \right] + \\
&\quad + \text{tr} \left\{ \left[\mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \mathbf{C}_{\mathbf{x}} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \right] \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} \right\}, \quad i = 1, 2, \dots, D. \tag{8.17}
\end{aligned}$$

Setting (8.17) equal to zero gives a necessary condition on $\hat{\theta}_{ml}$. Thus,

$$\left\{ \text{tr} \left[\left[\mathbf{K}_x^{-1}(\boldsymbol{\theta}) \mathbf{C}_x \mathbf{K}_x^{-1}(\boldsymbol{\theta}) \right] \frac{\partial \mathbf{K}_x(\boldsymbol{\theta})}{\partial \theta_i} \right] - \text{tr} \left[\mathbf{K}_x^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_x(\boldsymbol{\theta})}{\partial \theta_i} \right] \right\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ml}} = 0, \quad i = 1, 2, \dots, D, \quad (8.18)$$

or

$$\left\{ \text{tr} \left[\left[\mathbf{K}_x^{-1}(\boldsymbol{\theta}) \mathbf{C}_x \mathbf{K}_x^{-1}(\boldsymbol{\theta}) - \mathbf{K}_x^{-1}(\boldsymbol{\theta}) \right] \frac{\partial \mathbf{K}_x(\boldsymbol{\theta})}{\partial \theta_i} \right] \right\}_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{ml}} = 0, \quad i = 1, 2, \dots, D. \quad (8.19)$$

provides the necessary conditions on $\hat{\theta}_{ml}$.

When D is small (e.g., $D \leq 3$), we can use a grid search to find the approximate location of the peak of $L(\boldsymbol{\theta})$ and then use a gradient search technique to solve (8.18) in order to find the exact maximum. When D is large, we have to find more efficient techniques.

8.2.2 Maximum a posteriori (MAP) Estimator

In this case we assume that the vector parameter $\boldsymbol{\theta}$ is a random variable with a known probability density $p_{\boldsymbol{\theta}}(\boldsymbol{\theta})$. The MAP estimator is the value of $\boldsymbol{\theta}$ that maximizes the *a posteriori* density. This is equivalent to finding the value of $\boldsymbol{\theta}$ that maximizes

$$L_r(\boldsymbol{\theta}) = - \left\{ \ln \det[\mathbf{K}_x(\boldsymbol{\theta})] + \text{tr} \left[\mathbf{K}_x^{-1}(\boldsymbol{\theta}) \mathbf{C}_x \right] \right\} + \ln p_{\boldsymbol{\theta}}(\boldsymbol{\theta}), \quad (8.20)$$

where the subscript “*r*” denotes we are dealing with a random variable.

In many cases of interest we will model $\boldsymbol{\theta}$ as a zero-mean real Gaussian random vector whose probability density is

$$p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{D}{2}} |\mathbf{K}_{\boldsymbol{\theta}}|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} \boldsymbol{\theta}^T \mathbf{K}_{\boldsymbol{\theta}}^{-1} \boldsymbol{\theta} \right]. \quad (8.21)$$

The matrix $\mathbf{K}_{\boldsymbol{\theta}}$ is the covariance matrix of the probability density. It is *not* a function of $\boldsymbol{\theta}$.

To find a necessary condition on the MAP estimate, we differentiate (8.20) with respect to $\boldsymbol{\theta}$ and set the result equal to zero. The derivative of the first term is the left sides of (8.15), (8.16), and (8.17). The derivative of the second term is given by (A.380). Adding the two terms and solving the resulting equation gives a necessary condition on the MAP estimate,

$$\hat{\boldsymbol{\theta}}_{map} = \mathbf{K}_{\boldsymbol{\theta}} \nabla_{\boldsymbol{\theta}} [L(\boldsymbol{\theta})]|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_{map}}, \quad (8.22)$$

where the i th element of $\nabla_{\boldsymbol{\theta}} [L(\boldsymbol{\theta})]$ is given by (8.17).

We will look at numerous examples of ML and MAP estimates after we develop our estimation model.

8.2.3 Cramér-Rao Bounds

In order to understand the potential performance of parameter estimation algorithms we develop a set of bounds on their performance. In this section we develop the classical CRB and the Bayesian version of the bound.

These bounds are discussed in Section 2.4 of [VT68] and [VT01a]. The original development of the classic CRB is discussed in [Fis25], [Cra46], and [Rao46]. The Bayesian version of the bound was introduced in [VT68] and [VT01a].

We derive the classic CRB, the Bayesian version of the bound, and a hybrid bound for the case of a Gaussian observation model.

8.2.3.1 Classic Cramér-Rao bound

In this case the log-likelihood function is given by (8.9). The snapshots are assumed to be statistically independent, so we can deal with a single snapshot and combine the results at the end of the discussion. The log-likelihood function for single snapshot is

$$\begin{aligned} L_{\mathbf{x}}(\boldsymbol{\theta}) &\stackrel{\Delta}{=} \ln p_{\mathbf{x}|\boldsymbol{\theta}}(\mathbf{x}) \\ &= -\ln \det[\pi \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})] - \left\{ (\mathbf{x}^H - \mathbf{m}^H(\boldsymbol{\theta})) \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) (\mathbf{x} - \mathbf{m}(\boldsymbol{\theta})) \right\}. \end{aligned} \quad (8.23)$$

The CRB provides a bound on the covariance matrix of any *unbiased* estimate of $\boldsymbol{\theta}$. We denote the covariance matrix of the estimation errors by $\mathbf{C}(\boldsymbol{\theta})$. Then,

$$\mathbf{C}(\boldsymbol{\theta}) \stackrel{\Delta}{=} E \left[[\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}] [\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}]^T \right]. \quad (8.24)$$

The multiple-parameter CRB states that

$\mathbf{C}(\boldsymbol{\theta}) \geq \mathbf{C}_{CR}(\boldsymbol{\theta}) \stackrel{\Delta}{=} \mathbf{J}^{-1},$

(8.25)

for any unbiased estimate of $\boldsymbol{\theta}$. The matrix inequality means that $\mathbf{C}(\boldsymbol{\theta}) - \mathbf{C}_{CR}(\boldsymbol{\theta})$ is a non-negative definite matrix. The \mathbf{J} matrix is commonly referred to as Fisher's information matrix (or FIM).²

²The reader may want to review Section 2.4.3 of DEMT I [VT68], [VT01a] at this point.

The elements in \mathbf{J} are

$$\begin{aligned} J_{ij} &\triangleq E \left[\frac{\partial L_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} \cdot \frac{\partial L_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_j} \right] \\ &= -E \left[\frac{\partial^2 L_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right], \end{aligned} \quad (8.26)$$

or

$$\mathbf{J} = -E \left[\nabla_{\boldsymbol{\theta}} (L_{\mathbf{x}}(\boldsymbol{\theta}))^T \right]. \quad (8.27)$$

The result in (8.25) also provides a bound on the variance of any unbiased estimate of θ_i ,

$$\text{var} [\hat{\theta}_i - \theta_i] \geq [\mathbf{C}_{CR}(\boldsymbol{\theta})]_{ii} = [\mathbf{J}^{-1}]_{ii}. \quad (8.28)$$

The result in (8.28) can also be written as a Hadamard product,

$$\text{var} [\hat{\theta}_i - \theta_i] \geq [\mathbf{I} \odot \mathbf{C}_{CR}(\boldsymbol{\theta})]_{ii}. \quad (8.29)$$

For the problem of interest, J_{ij} is obtained by differentiating (8.23) and taking the expectation. The second derivative indicated in (8.26) is

$$\begin{aligned} \frac{\partial^2}{\partial \theta_i \partial \theta_j} [L_{\mathbf{x}}(\boldsymbol{\theta})] &= \text{tr} \left[-\mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} + \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial^2 \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right] \\ &- \left[(\mathbf{x}^H - \mathbf{m}^H(\boldsymbol{\theta})) \left(-\mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \right. \right. \\ &\quad \left. \left. + \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial^2 \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \right. \right. \\ &\quad \left. \left. + \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \right) \right. \\ &\quad \cdot (\mathbf{x} - \mathbf{m}(\boldsymbol{\theta})) \left. \right] - 2\text{Re} \left\{ -\frac{\partial^2 \mathbf{m}^H(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) [\mathbf{x} - \mathbf{m}(\boldsymbol{\theta})] \right. \\ &\quad + \frac{\partial \mathbf{m}^H(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) [\mathbf{x} - \mathbf{m}(\boldsymbol{\theta})] \\ &\quad + \frac{\partial \mathbf{m}^H(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) [\mathbf{x} - \mathbf{m}(\boldsymbol{\theta})] \\ &\quad \left. + \frac{\partial \mathbf{m}^H(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{m}(\boldsymbol{\theta})}{\partial \theta_j} \right\}. \end{aligned} \quad (8.30)$$

To take the expectation, we rewrite the second term as

$$E \left\{ -[\mathbf{x}^H - \mathbf{m}(\boldsymbol{\theta})](\cdots)[\mathbf{x} - \mathbf{m}(\boldsymbol{\theta})] \right\} =$$

$$E \left\{ -\text{tr} \left\{ (\cdots)[\mathbf{x} - \mathbf{m}(\boldsymbol{\theta})][\mathbf{x}^H - \mathbf{m}(\boldsymbol{\theta})] \right\} \right\} = -\text{tr} \left\{ (\cdots) \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta}) \right\}. \quad (8.31)$$

Taking the expectation of (8.30), using (8.31), and observing that the expectation of the first three terms in the brackets in (8.30) are zero gives

$$\begin{aligned} E \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} [L_{\mathbf{x}}(\boldsymbol{\theta})] \right] &= \text{tr} \left\{ \left[-\mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} + \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial^2 \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right] \right. \\ &\quad + \left[\mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_j} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} - \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial^2 \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right. \\ &\quad \left. \left. - \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_j} \right] \right\} \\ &\quad - 2\text{Re} \left\{ \frac{\partial \mathbf{m}^H(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{m}(\boldsymbol{\theta})}{\partial \theta_j} \right\}. \end{aligned} \quad (8.32)$$

The first four terms sum to zero. Thus,

$$J_{ij} = -E \left[\frac{\partial^2}{\partial \theta_i \partial \theta_j} [L_{\mathbf{x}}(\boldsymbol{\theta})] \right], \quad (8.33)$$

is

$$\boxed{J_{ij} = \text{tr} \left[\mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_j} \right] + 2\text{Re} \left[\frac{\partial \mathbf{m}^H(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{m}(\boldsymbol{\theta})}{\partial \theta_j} \right].} \quad (8.34)$$

In some cases, the mean is either zero or not a function of $\boldsymbol{\theta}$ so that (8.34) reduces to³

$$\boxed{J_{ij} = \text{tr} \left[\mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_j} \right].} \quad (8.35)$$

Using (A.401), we obtain a second form of (8.34) that is useful in many cases,

$$\begin{aligned} J_{ij} &= -\text{tr} \left[\frac{\partial \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta})}{\partial \theta_i} \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_j} \right] + \\ &\quad + 2\text{Re} \left[\frac{\partial \mathbf{m}^H(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{m}(\boldsymbol{\theta})}{\partial \theta_j} \right]. \end{aligned} \quad (8.36)$$

³This result was first published by Bangs [Ban71].

For the special case in which $\boldsymbol{\theta}$ is a scalar θ ,

$$\begin{aligned} \text{var}[\hat{\theta} - \theta] &\geq \left\{ \left[\text{tr} \left(\left[\mathbf{K}_{\mathbf{x}}^{-1}(\theta) \frac{\partial \mathbf{K}_{\mathbf{x}}(\theta)}{\partial \theta} \right]^2 \right) \right] + \right. \\ &\quad \left. + 2\text{Re} \left[\frac{\partial \mathbf{m}^H(\theta)}{\partial \theta} \mathbf{K}_{\mathbf{x}}^{-1}(\theta) \frac{\partial \mathbf{m}(\theta)}{\partial \theta} \right] \right\}^{-1}, \end{aligned} \quad (8.37)$$

or equivalently,

$$\begin{aligned} \text{var}[\hat{\theta} - \theta] &\geq \left\{ \left[-\text{tr} \left(\frac{\partial \mathbf{K}_{\mathbf{x}}^{-1}(\theta)}{\partial \theta} \frac{\partial \mathbf{K}_{\mathbf{x}}(\theta)}{\partial \theta} \right) \right] + \right. \\ &\quad \left. + 2\text{Re} \left[\frac{\partial \mathbf{m}^H(\theta)}{\partial \theta} \mathbf{K}_{\mathbf{x}}^{-1}(\theta) \frac{\partial \mathbf{m}(\theta)}{\partial \theta} \right] \right\}^{-1}. \end{aligned} \quad (8.38)$$

Note that the results in (8.34)–(8.38) are quite general and provide a starting point for many subsequent derivations. They apply whenever the observation is a complex Gaussian random process whose mean and covariance matrix are functions of the parameters of interest.

In many applications, the parameter vector $\boldsymbol{\theta}$ contains the parameters of interest to us as well as other unwanted parameters. A typical problem of this type is the case where we observe a plane-wave signal in additive spatially white Gaussian noise. We want to estimate the DOA but the signal power and the noise variance are unknown. In this case, the DOA, is the desired (or wanted) parameter, and signal power and the noise variance are unwanted parameters.

We partition $\boldsymbol{\theta}$ into two vectors,

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_w \\ \boldsymbol{\theta}_u \end{bmatrix}, \quad (8.39)$$

where $\boldsymbol{\theta}_w$ is a $D_1 \times 1$ nonrandom real vector containing the wanted parameters and $\boldsymbol{\theta}_u$ is a $D_2 \times 1$ nonrandom real vector containing the unwanted parameters.

We write \mathbf{J} as a partitioned matrix,

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\theta}_w \boldsymbol{\theta}_w} & \mathbf{J}_{\boldsymbol{\theta}_w \boldsymbol{\theta}_u} \\ \hline \mathbf{J}_{\boldsymbol{\theta}_u \boldsymbol{\theta}_w} & \mathbf{J}_{\boldsymbol{\theta}_u \boldsymbol{\theta}_u} \end{bmatrix}, \quad (8.40)$$

where $\mathbf{J}_{\boldsymbol{\theta}_w \boldsymbol{\theta}_u}$ is a $D_1 \times D_1$ matrix whose elements are given by (8.34) with $i, j = 1, \dots, D_1$. $\mathbf{J}_{\boldsymbol{\theta}_u \boldsymbol{\theta}_u}$ is a $D_1 \times D_2$ matrix whose elements are given by (8.34) with $i = 1, \dots, D_1; j = D_1 + 1, \dots, D$:

$$\mathbf{J}_{\boldsymbol{\theta}_u \boldsymbol{\theta}_u} = \mathbf{J}_{\boldsymbol{\theta}_w \boldsymbol{\theta}_u}^H. \quad (8.41)$$

$\mathbf{J}_{\boldsymbol{\theta}_u \boldsymbol{\theta}_u}$ is a $D_2 \times D_2$ matrix whose elements are given by (8.34) with $i, j = D_1 + 1, \dots, D$.

We partition the CRB in a similar manner.

$$\mathbf{C}_{CR}(\boldsymbol{\theta}) = \begin{bmatrix} \mathbf{C}_{CR}(\boldsymbol{\theta}_w) & \mathbf{C}_{CR}(\boldsymbol{\theta}_w, \boldsymbol{\theta}_u) \\ \hline \mathbf{C}_{CR}(\boldsymbol{\theta}_u, \boldsymbol{\theta}_w) & \mathbf{C}_{CR}(\boldsymbol{\theta}_u) \end{bmatrix} \quad (8.42)$$

We use the formula for the inverse of a block partitioned matrix to obtain

$$\boxed{\mathbf{C}_{CR}(\boldsymbol{\theta}_w) = \left[\mathbf{J}_{\boldsymbol{\theta}_w \boldsymbol{\theta}_w} - \mathbf{J}_{\boldsymbol{\theta}_w \boldsymbol{\theta}_u} \mathbf{J}_{\boldsymbol{\theta}_u \boldsymbol{\theta}_u}^{-1} \mathbf{J}_{\boldsymbol{\theta}_u \boldsymbol{\theta}_w} \right]^{-1}.} \quad (8.43)$$

The second term represents the effect of the unwanted parameters on the estimation error of the wanted parameters. It is always non-negative and will only be zero when $\mathbf{J}_{\boldsymbol{\theta}_w \boldsymbol{\theta}_u}$ is zero. This corresponds to uncoupled parameters. If $\mathbf{J}_{\boldsymbol{\theta}_w \boldsymbol{\theta}_u}$ is non-zero, the minus sign before the second term causes $C_{CR}(\boldsymbol{\theta}_w)$ to increase.

In array processing, we generally analyze the problem in ψ -space or u -space. In order to translate these results into angle-space, we need to consider the Fisher information matrix for functions of a variable.

We define a new $D_\gamma \times 1$ vector $\boldsymbol{\gamma}$, which is related to $\boldsymbol{\theta}$ by the functional relationship,

$$\boldsymbol{\gamma} = \mathbf{f}(\boldsymbol{\theta}). \quad (8.44)$$

Then,⁴

$$[\mathbf{J}(\boldsymbol{\gamma})]_{ij} = E \left\{ \sum_{p=1}^{D_\gamma} \frac{\partial \theta_p}{\partial \gamma_i} \frac{\partial L_x(\boldsymbol{\theta})}{\partial \theta_p} \sum_{q=1}^{D_\gamma} \frac{\partial L_x}{\partial \theta_q} \frac{\partial \theta_q}{\partial \gamma_j} \right\}. \quad (8.45)$$

We define the $D_\gamma \times D_\gamma$ matrix, \mathbf{G} , as

$$[\mathbf{G}]_{ip} = \frac{\partial \theta_p}{\partial \gamma_i}. \quad (8.46)$$

⁴See discussion on p. 83 of DEMT 1 [VT68] [VT01a].

Taking the expectation, (8.45) reduces to

$$\mathbf{J}(\boldsymbol{\gamma}) = \mathbf{G}\mathbf{J}(\boldsymbol{\theta})\mathbf{G}^T, \quad (8.47)$$

and the corresponding Cramér-Rao bound is

$$\mathbf{C}_{CR}(\boldsymbol{\gamma}) = \mathbf{G}^{-T}\mathbf{C}_{CR}(\boldsymbol{\theta})\mathbf{G}^{-1}. \quad (8.48)$$

In most applications that consider, γ_i , is a function of only θ_i so that the \mathbf{G} matrix is diagonal, (see Problems 8.2.1 and 8.2.2).

8.2.3.2 Bayesian Cramér-Rao bounds

In Section 2.4.3 of DEMT I [VT68] [VT01a], we derived a bound on the mean-square error in estimating a random vector parameter $\boldsymbol{\theta}$. The classic Cramér-Rao bound depends on the actual value of the parameter and can be described as a “local bound”. The Bayesian bound utilizes the *a priori* probability density of the parameter and provides a “global bound” that does not depend on the value of the parameter on a specific trial.

We denote the information matrix by \mathbf{J}_B . The subscript B denotes the Bayesian version of the Fisher information matrix. The \mathbf{J}_B matrix consists of two parts,

$$\mathbf{J}_B = \mathbf{J}_D + \mathbf{J}_P, \quad (8.49)$$

where the subscript “ D ” denotes the information due to the data and the subscript “ P ” denotes the information due to prior knowledge.

$$[\mathbf{J}_D]_{ij} \triangleq -E \left[\frac{\partial^2 L_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right], \quad (8.50)$$

where the expectation is over both \mathbf{x} and $\boldsymbol{\theta}$, and

$$[\mathbf{J}_P]_{ij} \triangleq -E \left[\frac{\partial^2 \ln p_{\boldsymbol{\theta}}(\boldsymbol{\theta})}{\partial \theta_i \partial \theta_j} \right], \quad (8.51)$$

where the expectation is over $\boldsymbol{\theta}$.

The correlation matrix of errors is

$$\mathbf{R}_{\boldsymbol{\epsilon}} = E \left[(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})^T \right]. \quad (8.52)$$

Note that $\mathbf{R}_{\boldsymbol{\epsilon}}$ is not a function of $\boldsymbol{\theta}$ because it is a random vector and we have taken the expected value (unlike $\mathbf{C}(\boldsymbol{\theta})$ in the nonrandom parameter case). From Property 2 on p. 84 of DEMT I [VT68] [VT01a],

$$\boxed{\mathbf{R}_{\boldsymbol{\epsilon}} \geq \mathbf{J}_B^{-1}}. \quad (8.53)$$

The inequality means that the matrix $[\mathbf{R}_\epsilon - \mathbf{J}_B^{-1}]$ is non-negative definite. Note that this is a bound on the mean-square errors.

For the mean-square error of the i th component of $\boldsymbol{\theta}$,

$$E \left[|\hat{\theta}_i - \theta_i|^2 \right] \geq [\mathbf{J}_B^{-1}]_{ii}. \quad (8.54)$$

For the Gaussian observation model, we use (8.34) in (8.50). Then,

$$\begin{aligned} [\mathbf{J}_D]_{ij} &= E_{\boldsymbol{\theta}} \left[\text{tr} \left[\mathbf{K}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}(\boldsymbol{\theta})}{\partial \theta_j} \right] \right. \\ &\quad \left. + 2\text{Re} \left[\frac{\partial \mathbf{m}^H(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{m}(\boldsymbol{\theta})}{\partial \theta_j} \right] \right], \end{aligned} \quad (8.55)$$

where the subscript “ $\boldsymbol{\theta}$ ” indicates that the expectation is with respect to the random parameter $\boldsymbol{\theta}$.

For the special case of a real parameter vector $\boldsymbol{\theta}$, whose *a priori* density is a multivariate Gaussian density with zero mean,

$$p_{\boldsymbol{\theta}}(\boldsymbol{\theta}) = \frac{1}{(2\pi)^{\frac{D}{2}} |\mathbf{K}_{\boldsymbol{\theta}}|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} \boldsymbol{\theta}^T \mathbf{K}_{\boldsymbol{\theta}}^{-1} \boldsymbol{\theta} \right], \quad (8.56)$$

and

$$\boxed{\mathbf{J}_P = \mathbf{K}_{\boldsymbol{\theta}}^{-1}.} \quad (8.57)$$

8.2.3.3 Hybrid Cramér-Rao-type bounds

In our study of array model perturbations and array calibration we find it useful to introduce a hybrid version of classic CRB and the Bayesian CRB.

We divide the parameter vector into a nonrandom and random components,

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \vdots \\ \boldsymbol{\theta}_2 \end{bmatrix}, \quad (8.58)$$

and assume $\boldsymbol{\theta}_1$ is a $D_1 \times 1$ nonrandom vector and $\boldsymbol{\theta}_2$ is a $D_2 \times 1$ random vector. Then \mathbf{J}_B is given by (8.49).

If $\boldsymbol{\theta}_2$ is a zero-mean random variable with a Gaussian probability density, then, \mathbf{J}_P is

$$\mathbf{J}_P = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{K}_{\boldsymbol{\theta}_2}^{-1} \end{bmatrix}. \quad (8.59)$$

The \mathbf{J}_D matrix contains an expectation over $\boldsymbol{\theta}_2$. For example, in the Gaussian observation model,

$$\begin{aligned} [\mathbf{J}_D]_{ij} &= E_{\boldsymbol{\theta}_2} \left[\text{tr} \left[\mathbf{K}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}(\boldsymbol{\theta})}{\partial \theta_j} \right] \right. \\ &\quad \left. + 2\text{Re} \left[\frac{\partial \mathbf{m}^H(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{m}(\boldsymbol{\theta})}{\partial \theta_j} \right] \right], \end{aligned} \quad (8.60)$$

where the subscript “ $\boldsymbol{\theta}_2$ ” indicates that the expectation is with respect to the random parameter $\boldsymbol{\theta}_2$.

We write the inverse of the \mathbf{J}_B as a partitioned matrix,

$$\mathbf{J}_B^{-1} \triangleq \left[\begin{array}{c|c} \mathbf{C}_{CR}(\boldsymbol{\theta}_1) & \mathbf{C}_1 \\ \hline \mathbf{C}_1^H & \mathbf{C}_{BCR}(\boldsymbol{\theta}_2) \end{array} \right]. \quad (8.61)$$

The $D_1 \times D_1$ matrix, $\mathbf{C}_{CR}(\boldsymbol{\theta}_1)$, provides a lower bound on the covariance of any unbiased estimate of $\boldsymbol{\theta}_1$. The $D_2 \times D_2$ matrix, $\mathbf{C}_{BCR}(\boldsymbol{\theta}_2)$, provides a lower bound of the mean-square error correlation matrix of $\boldsymbol{\theta}_2$. The proof of this assertion is a straightforward modification of the discussion in Section 2.4 of DEMT I [VT68] [VT01a]. The hybrid bound was introduced into the array processing literature by Rockah and Schultheiss [RS87a], [RS87b]. We will utilize it in Section 8.11.

8.2.3.4 Multiple snapshots

All of the expressions up to this point assumed a single snapshot (or sample). If we have K independent snapshots, then the log-likelihood function in (8.23) is multiplied by K and the FIM, \mathbf{J} , is multiplied by K . The classic CRB in (8.25) is divided by K .

In the Bayesian bound in Section 8.2.3.2, the data component \mathbf{J}_D is multiplied by K . Similarly, in the hybrid bound in Section 8.2.3.3, \mathbf{J}_D is multiplied by K .

8.2.3.5 Summary

The various forms of the CRB are important because, in many applications, the ML estimate (or the MAP estimate) approach the bound asymptotically as K goes to infinity.

In most of those applications, the value of K where the performance of estimator becomes close to the bound is a function of the signal to *SINR*.

We will quantify these ideas after we develop the model of interest.

8.3 Parameter Estimation Model

In this section, we develop the parameter estimation model that we utilize in our discussions in the remainder of this chapter and in Chapter 9. It is the same model we used in the beamformer problem in Chapters 6 and 7 except the parameters in the model such as the wavenumbers of the impinging plane waves are unknown.

In Section 8.3.1, we consider the case in which we have multiple plane waves impinging on the array. We want to estimate their directions of arrival (DOAs) and various signal parameters.

In Section 8.3.2, we consider the case in which the array configuration is perturbed from its nominal configuration. For example, the sensor positions may be perturbed from their nominal location, or the amplitude and phase of the sensor response may be different from their nominal values. We develop a model that incorporates these perturbations.

In Section 8.3.3, we consider spatially spread signals that can be modelled parametrically.

In Section 8.3.4, we summarize our results.

8.3.1 Multiple Plane Waves

In this section we develop the model that we use to estimate the parameters of multiple plane waves impinging on an array. We develop a frequency-domain model and a time-domain model. We then introduce notation to treat both models with a single generic notation.

The first model of interest is the frequency-domain snapshot model that we encountered previously in Section 5.2 and was used throughout Chapters 6 and 7.

The frequency-domain snapshot model is appropriate for either narrowband or wideband processes. For the narrowband case, we only use the snapshot corresponding to the carrier frequency.

For the case of a linear array, we can write the snapshots as

$$\mathbf{X}(k) = \mathbf{V}(\psi)\mathbf{F}(k) + \mathbf{N}(k), \quad k = 1, \dots, K, \quad (8.62)$$

where

$$\boldsymbol{\psi} = \begin{bmatrix} \psi_1 & \cdots & \psi_D \end{bmatrix}^T, \quad (8.63)$$

is a $D \times 1$ vector containing the wavenumbers of the D plane waves ($\psi_i = k_i d$) and $\mathbf{V}(\psi)$ is the array manifold matrix, $\mathbf{F}(k)$ is a $D \times 1$ vector of the source signals and $\mathbf{N}(k)$ is an $N \times 1$ vector of the noise.

For the arbitrary array case, ψ is a $D \times 2$ matrix,

$$\psi = \begin{bmatrix} \psi_1 & \cdots & \psi_D \end{bmatrix}, \quad (8.64)$$

where

$$\psi_i = \begin{bmatrix} \psi_{x_i} & \psi_{y_i} \end{bmatrix}^T \quad (8.65)$$

contains the two wavenumbers of each source.

The spatial spectral matrix is

$$\mathbf{S}_x = \mathbf{V}(\psi) \mathbf{S}_f \mathbf{V}^H(\psi) + \mathbf{S}_n. \quad (8.66)$$

For the narrowband case, the time-domain snapshot model discussed in Section 5.2 is also appropriate.

In this case,

$$\mathbf{x}(k) = \mathbf{V}(\psi) \mathbf{f}(k) + \mathbf{n}(k), \quad k = 1, 2, \dots, K. \quad (8.67)$$

where the argument “ k ” denotes the time-domain sample (or snapshot) at $t = k$. We assume the samples are from a stationary zero-mean Gaussian random process. Thus,

$$\begin{aligned} \mathbf{R}_x &= E \left[\mathbf{x}(k) \mathbf{x}^H(k) \right] \\ &= \mathbf{V}(\psi) \mathbf{R}_f \mathbf{V}^H(\psi) + \mathbf{R}_n. \end{aligned} \quad (8.68)$$

The majority of our discussion focuses on the narrowband case. We normally utilize the frequency-domain notation in (8.66), because the extension to the wideband case is straightforward. In all of our snapshot models, we assume that successive snapshots are statistically independent.

We emphasize four different models for the signal. In the narrowband case, these are:

Case S1: The source signals are sample functions from a zero-mean vector stationary complex Gaussian random process and \mathbf{S}_f is known.

Case S2: The source signals are sample functions from a zero-mean vector stationary complex Gaussian random process and \mathbf{S}_f is unknown.

In Case S2, the source signals may be uncorrelated, but we do not know that, so we must consider a general \mathbf{S}_f in our estimators and bounds. Case S2u considers the case where we know *a priori* that the source signals are uncorrelated.

Case S2u: The source signals are sample functions from a zero-mean vector stationary complex Gaussian random process. They are uncorrelated (and therefore, statistically independent). \mathbf{S}_f is a diagonal matrix denoted by Λ_f . Λ_f is unknown.

Case S3: The source signals are considered to be unknown complex sequences in order to estimate ψ . In this case, the performance will depend on the values in the complex sequences (just as in classical ML estimation). In order to compare the performance to Case S2, we can consider the unknown complex sequence to be a particular sample function from a zero-mean stationary complex random process. We can then use the statistics of the process to evaluate the performance. ML estimators for this model are referred to as conditional (on a particular sample function) ML estimators. The spectral matrix in (8.66) will have to be revised for this case.

There are corresponding signal models for the wideband case that we will discuss in various sections of the chapter.

In communication systems, we may want to impose more structure in the signal model. One model that we will consider is:

Case S4: The source signal is known. The received signal component from the i th source is $\alpha_i f_i(k)$, $k = 1, \dots, K$, where α_i is a complex constant. This corresponds to a communication system where a known training sequence is sent.

We could also consider cyclostationary signals (e.g., [Gaa88], [SG92], [XK92]) or [Sch94] and constant modulus (CM) signals (e.g., [GL86]). The reader is referred to these references for a discussion of these models.

There are three noise models for the narrowband case:

Case N1:

$$E [\mathbf{N}(k) \mathbf{N}^H(k)] = \sigma_w^2 \mathbf{I} \quad (8.69)$$

and σ_w^2 is known.

Case N2:

$$E [\mathbf{N}(k) \mathbf{N}^H(k)] = \sigma_w^2 \mathbf{I} \quad (8.70)$$

and σ_w^2 is unknown.

Case N3:

$$E [\mathbf{N}(k) \mathbf{N}^H(k)] = \mathbf{S}_n \quad (8.71)$$

and \mathbf{S}_n is unknown. Note that, if \mathbf{S}_n were known, Case N3 could be reduced to Case N1 by a spatial whitening filter.

There are corresponding noise models for the wideband case that we will discuss in later sections.

For each of these models we can identify the parameters that must be estimated. For the one-dimensional case, a representative set is summarized in Table 8.2. For the case of a complex parameter such as \mathbf{f}_k , we estimate both real and imaginary parts.

Table 8.2: Parameters to be Estimated

Signal Case	Noise Case	Parameters To Estimate	Dimension of $\boldsymbol{\theta}$
S1	N1	$\psi_i, i = 1, \dots, D$	D
S1	N2	$\psi_i, i = 1, \dots, D$ σ_w^2	$D + 1$
S1	N3	$\psi_i, i = 1, \dots, D$ \mathbf{S}_n	$D + N^2$
S2	N1	$\psi_i, i = 1, \dots, D$ \mathbf{S}_f	$D + D^2$
S2u	N1	$\psi_i, i = 1, \dots, D$ $\sigma_{s_i}^2, i = 1, \dots, D$	$2D$
S3	N1	$\psi_i, i = 1, \dots, D$ $\mathbf{f}_k, k = 1, \dots, K$	$D + 2DK$
S4	N1	$\psi_i, i = 1, \dots, D$ $\alpha_i, i = 1, \dots, D$	$3D$

In the (S1, N3) and (S2, N1) cases the dimension of $\boldsymbol{\theta}$ may be different if constraints are placed on \mathbf{S}_f or \mathbf{S}_n . The combinations (S2, N2), (S2, N3), (S3, N2), (S3, N3) are obvious modifications of earlier cases.

8.3.2 Model Perturbations

The nominal model given in (8.62) is

$$\mathbf{X}(k) = \mathbf{V}(\boldsymbol{\psi}) \mathbf{F}(k) + \mathbf{N}(k), \quad k = 1, 2, \dots, K. \quad (8.72)$$

In order to model the perturbations, we write (8.72) as

$$\mathbf{X}(k) = \mathbf{V}(\psi, \rho) \mathbf{F}(k) + \mathbf{N}(k), \quad k = 1, 2, \dots, K, \quad (8.73)$$

where ρ represents the parameters whose perturbations we will investigate. We refer to ρ as the **perturbation vector**.⁵

We rewrite the array manifold vector to include the gain and phase of each sensor explicitly

$$\mathbf{v}(\psi_i, \rho) = \left[a_0 e^{j\phi_0} e^{j\frac{2\pi}{\lambda} \mathbf{u}_i^T \mathbf{p}_0} \mid \dots \mid a_{N-1} e^{j\phi_{N-1}} e^{j\frac{2\pi}{\lambda} \mathbf{u}_i^T \mathbf{p}_{N-1}} \right]^T. \quad (8.74)$$

In the general case, the ρ vector is a $5N \times 1$ vector containing a gain, phase, and position vector. Define

$$[\mathbf{a}]_n = a_n^n (1 + \Delta a_n), \quad n = 0, \dots, N-1, \quad (8.75)$$

$$[\phi]_n = \phi_n^n + \Delta \phi_n, \quad n = 0, \dots, N-1, \quad (8.76)$$

$$[\mathbf{p}_x]_n = p_{xn}^n + \Delta p_{xn}, \quad n = 0, \dots, N-1, \quad (8.77)$$

$$[\mathbf{p}_y]_n = p_{yn}^n + \Delta p_{yn}, \quad n = 0, \dots, N-1, \quad (8.78)$$

$$[\mathbf{p}_z]_n = p_{zn}^n + \Delta p_{zn}, \quad n = 0, \dots, N-1. \quad (8.79)$$

The superscript “ n ” denotes the nominal value. This model is similar to the model in Section 2.6.3 and Section 6.6.3. Then, ρ is

$$\rho = [\mathbf{a}^T \ \phi^T \ \mathbf{p}_x^T \ \mathbf{p}_y^T \ \mathbf{p}_z^T]^T. \quad (8.80)$$

In most cases we study a subset of the possible variations and ρ has a smaller dimension. Normally we assume that ρ has a multivariate Gaussian density,

$$p_\rho(\rho) = \frac{1}{(2\pi)^{\frac{5N}{2}} |\Lambda_\rho|^{\frac{1}{2}}} \exp \left[-\frac{1}{2} [\rho - \rho_0]^T \Lambda_\rho^{-1} [\rho - \rho_0] \right], \quad (8.81)$$

where ρ_0 denotes the nominal value and Λ_ρ is the covariance matrix.

The second kind of variation is in the noise environment. Our nominal model assumes

$$E [\mathbf{N}(k) \mathbf{N}^H(k)] = \sigma_w^2 \mathbf{I}. \quad (8.82)$$

We can examine correlated noise by defining

$$E [\mathbf{N}(k) \mathbf{N}^H(k)] = \sigma_w^2 \mathbf{I} + \tilde{\Lambda}. \quad (8.83)$$

We look at the effect of $\tilde{\Lambda}$ on processing performance.

There are three ways to deal with model perturbations:

⁵Although we refer to ρ as the perturbation vector, it contains the nominal values.

- (i) We choose some nominal value for the array parameters and estimate the DOAs as if they were correct. We then investigate how the estimator performance degrades as the parameters vary from their nominal value.
- (ii) We can model the variations as random variables, treat them as nuisance parameters, and integrate them out of the likelihood function.
- (iii) We can jointly estimate the model parameters and the DOAs.

We will look at examples of these techniques in subsequent sections.

8.3.3 Parametric Spatially Spread Signals

In this case we assume that we have D spatially spread signals impinging on the array and the spatial characteristics of the i th signal can be characterized by an L_i -dimensional vector.

A simple example of this situation would be the case in which each of the D signals could be modeled as a complex AR(1) process (see Figures 5.24 and 5.25). For each signal we would estimate two parameters ϕ_a and $|a(1)|$ (or alternatively $Re[a(1)]$ and $Im[a(1)]$).

We will consider models of this type in Section 8.9.

8.3.4 Summary

There are other propagation models that may be appropriate in certain applications. These include multipath models and near-field models. The reader is referred to the literature for a discussion of these models. We will focus our attention on the three models described above.

8.4 Cramér-Rao Bounds

In Section 8.2.3, we developed the classical CRB for parameter estimation. In this section, we apply those results to the signal and noise models that we developed in Section 8.3.

In Section 8.4.1, we consider the case in which the signal vector is a sample function from a Gaussian random process with unknown signal spectral matrix and the noise is a sample function from a spatially white Gaussian random process. This case includes (S2, N1) and (S2, N2).

In Section 8.4.2, we consider the case in which the signal vector is a sample function from a Gaussian random process and we know *a priori* that the component signals are statistically independent with unknown Λ_f . The

noise is a sample function from a spatially white Gaussian random process. This case includes (S2u, N1) and (S2u, N2).

In Section 8.4.3, we consider the case in which the signal is a sample function from a Gaussian random process with a known signal spectrum and the noise is a sample function from a spatially white Gaussian random process.

In Section 8.4.4, we consider the case in which the signal is modelled as a nonrandom, unknown waveform and the noise is a sample function from a spatially white Gaussian random process.

In Section 8.4.5, we consider the case in which the signals are modelled as known waveforms with complex amplitude multipliers. The noise is a sample function from a spatially white complex Gaussian random process. This corresponds to case (S4, N1) and (S4, N2).

In Section 8.4.6, we summarize our results.

Before beginning the development, a brief commentary on the steps is useful:

- (i) The first step is to derive the Fisher information matrix \mathbf{J} (8.34) or \mathbf{J}_B (8.49). This step is usually straightforward, but may be tedious.
- (ii) In most cases, we are primarily interested in estimating ψ , the DOAs of the D signals. We partition \mathbf{J} into blocks corresponding to ψ and blocks corresponding to the other (unwanted) parameters and blocks corresponding to cross-terms. We partition \mathbf{C}_{CR} similarly and use the formula for the inverse of block matrices to find the block of \mathbf{C}_{CR} corresponding to ψ .

This step is straightforward, but may contain complicated expressions. In many of our examples, we evaluate the expression numerically and plot the result.

- (iii) The final step is to obtain a compact expression for $\mathbf{C}_{CR}(\psi)$. This step is normally a tour de force of linear algebra. When successful, it offers the ability to compare the bounds for different models. We will quote some of these results, but will normally not derive them.

8.4.1 Gaussian Model: Unknown Signal Spectrum

8.4.1.1 Cramér-Rao bound

The first model of interest corresponds to the signals being a sample function from a zero-mean stationary complex Gaussian random process whose

spectral matrix \mathbf{S}_f is unknown and the additive noise is a sample function from a spatially white Gaussian random process (this is Case (S2, N2) in Section 8.2).

We consider an array model in which the wavenumber is 1-D and assume that there are D signals impinging on the array. Thus,

$$\mathbf{X}(k) = \mathbf{V}(\psi)\mathbf{F}(k) + \mathbf{N}(k), \quad (8.84)$$

and

$$\mathbf{S}_n = \sigma_w^2 \mathbf{I}, \quad (8.85)$$

where σ_w^2 is unknown, and \mathbf{S}_f is the unknown signal spectral matrix.

The input spectral matrix is

$$\mathbf{S}_x = \mathbf{V}(\psi)\mathbf{S}_f\mathbf{V}^H(\psi) + \sigma_w^2 \mathbf{I}. \quad (8.86)$$

In this case, the parameter vector $\boldsymbol{\theta}$ can be written as,

$$\boldsymbol{\theta} \triangleq [\psi, \mu, \sigma_w^2]^T, \quad (8.87)$$

where ψ is a $D \times 1$ vector corresponding to signal DOAs ($\psi_i = \pi \cos \theta_i$), μ is a real vector corresponding to the elements in \mathbf{S}_f , and σ_w^2 is a scalar corresponding to the noise spectral height. The matrix \mathbf{S}_f is a complex Hermitian matrix.

$$\dim[\mu] = D + \sum_{k=1}^{D-1} 2k = D^2. \quad (8.88)$$

We construct the \mathbf{J} matrix in (8.35) in a partitioned form,

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{\psi\psi} & \mathbf{J}_{\psi\mu} & \mathbf{J}_{\psi\sigma_w^2} \\ \mathbf{J}_{\mu\psi} & \mathbf{J}_{\mu\mu} & \mathbf{J}_{\mu\sigma_w^2} \\ \mathbf{J}_{\sigma_w^2\psi} & \mathbf{J}_{\sigma_w^2\mu} & \mathbf{J}_{\sigma_w^2\sigma_w^2} \end{bmatrix}. \quad (8.89)$$

Each element is obtained by substituting \mathbf{S}_x from (8.86) for \mathbf{K}_x in (8.35). We then must find \mathbf{S}_x^{-1} , compute the derivative with respect to each parameter, and evaluate the trace in (8.35). Thus,

$$[\mathbf{J}_{\psi\psi}]_{ij} = \text{tr} \left[\mathbf{S}_x^{-1} \frac{\partial \mathbf{S}_x}{\partial \psi_i} \mathbf{S}_x^{-1} \frac{\partial \mathbf{S}_x}{\partial \psi_j} \right]. \quad (8.90)$$

Using the matrix inversion lemma (A.48) on (8.86) gives

$$\mathbf{S}_{\mathbf{x}}^{-1} = \frac{1}{\sigma_w^2} \left[\mathbf{I} - \mathbf{V} \left(\mathbf{V}^H \mathbf{V} + \sigma_w^2 \mathbf{S}_{\mathbf{f}}^{-1} \right) \mathbf{V}^H \right]. \quad (8.91)$$

The derivative needed for the $\mathbf{J}_{\psi\psi}$ matrix is

$$\frac{\partial \mathbf{S}_{\mathbf{x}}}{\partial \psi_i} = \frac{\partial \mathbf{V}(\psi)}{\partial \psi_i} \mathbf{S}_{\mathbf{f}} \mathbf{V}^H(\psi) + \mathbf{V}(\psi) \mathbf{S}_{\mathbf{f}} \frac{\partial \mathbf{V}^H(\psi)}{\partial \psi_i}. \quad (8.92)$$

The i th column of the derivative matrix is

$$\left[\frac{\partial \mathbf{V}(\psi)}{\partial \psi_i} \right]_i = \left. \frac{\partial \mathbf{v}(\psi)}{\partial \psi} \right|_{\psi=\psi_i} \triangleq \mathbf{d}(\psi_i), \quad (8.93)$$

and the remaining columns are zero. Using (8.91), (8.92), and (8.93) in (8.90) we obtain $[\mathbf{J}_{\psi\psi}]_{ij}$. We carry out the details of the derivation in Section 8.4.2.

The terms in the other submatrices follow in a similar manner. This procedure has been carried out by Weiss and Friedlander [WF93] (The evaluation of the \mathbf{J}_{ij} is carried out in [WF90] and summarized in [WF93].) The derivation is complicated and the interested reader should consult those two references.

We next partition the CRB matrix in a similar manner,

$$\mathbf{C}_{CR} = \begin{bmatrix} \mathbf{C}_{CR}(\psi) & \mathbf{C}_{CR}(\psi, \mu) & \mathbf{C}_{CR}(\psi, \sigma_w^2) \\ \mathbf{C}_{CR}(\mu, \psi) & \mathbf{C}_{CR}(\mu) & \mathbf{C}_{CR}(\mu, \sigma_w^2) \\ \mathbf{C}_{CR}(\sigma_w^2, \psi) & \mathbf{C}_{CR}(\sigma_w^2, \mu) & \mathbf{C}_{CR}(\sigma_w^2) \end{bmatrix}. \quad (8.94)$$

We then use the partitioned matrix inversion formula (A.67) to obtain $\mathbf{C}_{CR}(\psi)$.

The result is

$$\boxed{\mathbf{C}_{CR}(\psi) = \frac{\sigma_w^2}{2K} \left\{ Re \left[\left[\mathbf{S}_{\mathbf{f}} \mathbf{V}^H \mathbf{S}_{\mathbf{x}}^{-1} \mathbf{V} \mathbf{S}_{\mathbf{f}} \right] \odot \left[\mathbf{D}^H \mathbf{P}_{\mathbf{V}}^{\perp} \mathbf{D} \right]^T \right] \right\}^{-1}}, \quad (8.95)$$

where \odot denotes the Hadamard product (A.70). The result includes K to account for the K independent snapshots. Note that

$$\mathbf{P}_{\mathbf{V}}^{\perp} = \left[\mathbf{I} - \mathbf{V} \left(\mathbf{V}^H \mathbf{V} \right)^{-1} \mathbf{V}^H \right] \quad (8.96)$$

is the projection matrix onto the noise subspace, and

$$\mathbf{D} = \dot{\mathbf{V}}_{\psi} = \left[\frac{\partial \mathbf{v}(\psi_1)}{\partial \psi_1} : \frac{\partial \mathbf{v}(\psi_2)}{\partial \psi_2} : \dots : \frac{\partial \mathbf{v}(\psi_D)}{\partial \psi_D} \right] \quad (8.97)$$

is the derivative matrix. The result in (8.95) was first obtained by Stoica and Nehorai [SN89a] by analyzing the asymptotic behavior of the maximum likelihood estimate (e.g., Stoica and Nehori [SN89b], Viberg [Vib89], Otterson [Ott89], and Stoica and Nehorai [SN90a], [SN90b]). The first direct derivation is due to Weiss and Friedlander [WF93], [WF90], and [WF91a]. The most efficient derivation is contained in Stoica et al. [SLG01]. The reader is referred to these references for the details of the derivation.

There are two alternative forms of (8.95) that we will encounter in the sequel. The first is an obvious modification,

$$\mathbf{C}_{CR}(\psi) = \frac{\sigma_w^2}{2K} \left\{ \text{Re} \left[\left[\mathbf{D}^H \mathbf{P}_{\mathbf{V}}^{\perp} \mathbf{D} \right] \odot \left[\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f \right]^T \right] \right\}^{-1}, \quad (8.98)$$

and the second form specifies the ij element of \mathbf{C}_{CR}^{-1} ,

$$\left[\mathbf{C}_{CR}^{-1} \right]_{ij} = \frac{2K}{\sigma_w^2} \left\{ \text{Re} \left[\text{tr} \left[\left[\mathbf{D}_j^H \mathbf{P}_{\mathbf{V}}^{\perp} \mathbf{D}_i \right] \left[\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f \right] \right] \right] \right\}. \quad (8.99)$$

Note that if we want to find the CRB with respect to arrival angle θ , we substitute

$$\mathbf{D}_{\theta} = \dot{\mathbf{V}}_{\theta}. \quad (8.100)$$

We define

$$\mathbf{H} \stackrel{\Delta}{=} \mathbf{D}^H \mathbf{P}_{\mathbf{V}}^{\perp} \mathbf{D}, \quad (8.101)$$

and rewrite (8.95) as,

$$\boxed{\mathbf{C}_{CR}(\psi) = \frac{\sigma_w^2}{2K} \left\{ \text{Re} \left[\left(\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f \right) \odot \mathbf{H}^T \right] \right\}^{-1}.} \quad (8.102)$$

In [WF90], it is shown that the expressions in (8.95) and (8.102) are also valid for the case in which σ_w^2 is known.

We can simplify the first term in (8.102). We define

$$\Sigma = \frac{\mathbf{S}_f}{\sigma_w^2}, \quad (8.103)$$

and expand \mathbf{S}_x^{-1} using (8.91). Note that Σ is a matrix *SNR*:

$$\begin{aligned}\mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f &= \mathbf{V}^H \left[\mathbf{I} - \mathbf{V} \Sigma \left(\mathbf{V}^H \mathbf{V} \Sigma + \mathbf{I} \right)^{-1} \mathbf{V}^H \right] \mathbf{V} \Sigma \\ &= \mathbf{V}^H \mathbf{V} \Sigma - \mathbf{V}^H \mathbf{V} \Sigma \left(\mathbf{V}^H \mathbf{V} \Sigma + \mathbf{I} \right)^{-1} \mathbf{V}^H \mathbf{V} \Sigma \\ &= \left[\mathbf{I} - \mathbf{V}^H \mathbf{V} \Sigma \left(\mathbf{V}^H \mathbf{V} \Sigma + \mathbf{I} \right)^{-1} \right] \mathbf{V}^H \mathbf{V} \Sigma.\end{aligned}\quad (8.104)$$

We can rewrite the identity matrix in the bracket as

$$\begin{aligned}\mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f &= \left[\left(\mathbf{V}^H \mathbf{V} \Sigma + \mathbf{I} \right) \left(\mathbf{V}^H \mathbf{V} \Sigma + \mathbf{I} \right)^{-1} - \right. \\ &\quad \left. \mathbf{V}^H \mathbf{V} \Sigma \left(\mathbf{V}^H \mathbf{V} \Sigma + \mathbf{I} \right)^{-1} \right] \mathbf{V}^H \mathbf{V} \Sigma \\ &= \left(\mathbf{V}^H \mathbf{V} \Sigma + \mathbf{I} \right)^{-1} \mathbf{V}^H \mathbf{V} \Sigma.\end{aligned}\quad (8.105)$$

Using (8.105) in (8.102) we obtain

$$\mathbf{C}_{CR}(\psi) = \frac{\sigma_w^2}{2K} \left\{ Re \left\{ \left[\mathbf{S}_f \left[\left(\mathbf{I} + \mathbf{V}^H \mathbf{V} \frac{\mathbf{S}_f}{\sigma_w^2} \right)^{-1} \left(\mathbf{V}^H \mathbf{V} \frac{\mathbf{S}_f}{\sigma_w^2} \right) \right] \right] \odot \mathbf{H}^T \right\} \right\}^{-1}.$$

(8.106)

The result in (8.106) is the formula we shall use for most of our CRB calculations. As pointed out previously, the result is valid for both known and unknown noise variance.

For large values of $\mathbf{V}^H \mathbf{V} \mathbf{S}_f / \sigma_w^2$, the term in the innermost bracket approaches the identity matrix \mathbf{I} . To show this relationship, we expand $\mathbf{V}^H \mathbf{V} \Sigma$ in terms of its eigenvalues and eigenvectors,

$$\mathbf{V}^H \mathbf{V} \Sigma = \sum_{i=1}^N \lambda_i \Phi_i \Phi_i^H. \quad (8.107)$$

Then,

$$\begin{aligned}\left[\mathbf{I} + \mathbf{V}^H \mathbf{V} \Sigma \right]^{-1} \mathbf{V}^H \mathbf{V} \Sigma &= \sum_{i=1}^N \frac{\lambda_i}{1 + \lambda_i} \Phi_i \Phi_i^H \\ &= \sum_{i=1}^N \frac{1}{1 + \frac{1}{\lambda_i}} \Phi_i \Phi_i^H.\end{aligned}\quad (8.108)$$

Assuming that $\lambda_i > 1$, we can expand the denominator, giving

$$\begin{aligned} [\mathbf{I} + \mathbf{V}^H \mathbf{V} \boldsymbol{\Sigma}]^{-1} \mathbf{V}^H \mathbf{V} \boldsymbol{\Sigma} &= \mathbf{I} - \sum_{i=1}^N \frac{1}{\lambda_i} \boldsymbol{\Phi}_i \boldsymbol{\Phi}_i^H + \dots \\ &= \mathbf{I} - [\mathbf{V}^H \mathbf{V} \boldsymbol{\Sigma}]^{-1} + \dots \end{aligned} \quad (8.109)$$

The second and higher-order terms approach zero as λ_i approaches infinity.

This occurs when \mathbf{S}_f/σ_w^2 is large and:

- (i) \mathbf{S}_f is not close to singular. If the signals are correlated, \mathbf{S}_f approaches singularity as $|\rho_{ij}|$ approaches one for some ij pair.
- (ii) If the signals are uncorrelated, the smallest $[\mathbf{S}_f]_{ii}/\sigma_w^2$ must be large.
- (iii) The signals are not too closely spaced. Closely spaced signals result in $\mathbf{V}^H \mathbf{V}$ approaching a singular matrix.

When these conditions are satisfied, then we can approximate (8.106) as

$$\boxed{\mathbf{C}_{CR}(\psi) \cong \frac{\sigma_w^2}{2K} \left\{ Re \left[\mathbf{S}_f \odot \mathbf{H}^T \right] \right\}^{-1} \triangleq \mathbf{C}_{ACR}(\psi).} \quad (8.110)$$

where \mathbf{C}_{ACR} denotes the approximate CRB.

We consider several simple examples and then return to the CRB in (8.106) and derive some general properties.

8.4.1.2 Single signals

Example 8.4.1

Consider the case of a single signal. In this case J_ψ is a scalar. Then, (8.106) can be written as

$$\begin{aligned} J_\psi &= Re \left[\frac{2K\sigma_s^2}{\sigma_w^2} \cdot \left(1 + N \frac{\sigma_s^2}{\sigma_w^2} \right)^{-1} \left(N \frac{\sigma_s^2}{\sigma_w^2} \right) \right] \left[\mathbf{d}^H \left(1 - \frac{\mathbf{v}\mathbf{v}^H}{N} \right) \mathbf{d} \right] \\ &= \frac{2K}{N} \left[\frac{(ASNR)^2}{1 + ASNR} \right] \left[\mathbf{d}^H \left(1 - \frac{\mathbf{v}\mathbf{v}^H}{N} \right) \mathbf{d} \right], \end{aligned} \quad (8.111)$$

where the *ASNR* is

$$ASNR = N \frac{\sigma_s^2}{\sigma_w^2}. \quad (8.112)$$

The signal-to-noise behavior is contained in the term in the first bracket and the array geometry is contained in the second bracket.

If the array manifold is conjugate symmetric ($\mathbf{v} = \mathbf{J} \mathbf{v}^*$), then

$$\mathbf{d}^H \mathbf{v} = 0, \quad (8.113)$$

and

$$J_\psi = \frac{2K}{N} \left[\frac{(ASN R)^2}{1 + ASN R} \right] \| \mathbf{d} \|^2. \quad (8.114)$$

Then,

$$\begin{aligned} C_{CR}(\psi) &= \frac{N}{2K} \left[\frac{1 + ASN R}{(ASN R)^2} \right] \| \mathbf{d} \|^2 \\ &= \frac{N}{2K} \left[\frac{1}{ASN R} + \frac{1}{(ASN R)^2} \right] \| \mathbf{d} \|^2. \end{aligned} \quad (8.115)$$

We see that for small $ASN R$ the variance bound is proportional to $(ASN R)^{-2}$ and for large $ASN R$, it is proportional to $(ASN R)^{-1}$. We also observe that the effect of the array geometry is completely contained in the $\| \mathbf{d} \|^2$ term.

We recall that the conventional beam pattern is

$$B_C(\psi_1, \psi_2) = \mathbf{v}^H(\psi_1) \mathbf{v}(\psi_2), \quad (8.116)$$

where ψ_1 is steering direction and ψ_2 is direction of the incoming plane wave. Then,

$$\frac{\partial^2 B_C(\psi_1, \psi_2)}{\partial \psi_1 \partial \psi_2} = \mathbf{d}^H(\psi_1) \mathbf{d}(\psi_2), \quad (8.117)$$

and

$$\| \mathbf{d} \|^2 = \frac{\partial^2 B_C(\psi_1, \psi_2)}{\partial \psi_1 \partial \psi_2} \Big|_{\psi_1=\psi_2}. \quad (8.118)$$

The beam pattern in wavenumber space is only a function of $\psi_1 - \psi_2$, so we can also write (8.118) as

$$\| \mathbf{d} \|^2 = - \frac{\partial^2 B_C(\psi_1, \psi_2)}{\partial^2 \psi_2} \Big|_{\psi_2=0}. \quad (8.119)$$

Thus, the effect of the array geometry on the CRB for a single signal is completely specified by the second derivative (or the radius of curvature) of the conventional beam pattern.

If we use the approximation in (8.110), we obtain

$$C_{ACR}(\psi) = \frac{N}{2K} \left[\frac{1}{ASN R} \right] \| \mathbf{d} \|^2. \quad (8.120)$$

We see that $C_{CR}(\psi)$ approaches $C_{ACR}(\psi)$ for large $ASN R$. However, $C_{ACR}(\psi)$ loses the quadratic dependence at low $ASN R$.

We can obtain an explicit expression for the term in the second bracket in (8.111). We do not need to assume linearity or conjugate symmetry.

The array manifold vector is

$$\mathbf{v}(\psi) = \begin{bmatrix} e^{j \frac{2d_1}{\lambda} \psi} & e^{j \frac{2d_2}{\lambda} \psi} & \dots & e^{j \frac{2d_N}{\lambda} \psi} \end{bmatrix}^T \quad (8.121)$$

Then,

$$\mathbf{d}^H \mathbf{d} = \left(\frac{2}{\lambda} \right)^2 \sum_{i=1}^N d_i^2, \quad (8.122)$$

and the second bracket in (8.111) can be written as

$$\mathbf{d}^H \mathbf{d} - \frac{|\mathbf{v}^H(\psi) \mathbf{d}|^2}{N} = 4 \left\{ \sum_{i=1}^N \left(\frac{d_i}{\lambda} \right)^2 - \frac{1}{N} \left(\sum_{i=1}^N \frac{d_i}{\lambda} \right)^2 \right\}. \quad (8.123)$$

Defining,

$$\mathbf{p} = \begin{bmatrix} \frac{d_1}{\lambda} & \frac{d_2}{\lambda} & \dots & \frac{d_N}{\lambda} \end{bmatrix}^T, \quad (8.124)$$

we can write (8.123) as

$$\mathbf{d}^H \mathbf{d} - \frac{|\mathbf{v}^H(\psi) \mathbf{d}|^2}{N} = 4 \mathbf{p}^T \left[\mathbf{I} - \frac{1}{N} \mathbf{1} \mathbf{1}^T \right] \mathbf{p}. \quad (8.125)$$

We now apply this result to a linear array (similar simplifications follow for conjugate symmetric arrays).

Example 8.4.2 (continuation)

Consider a standard linear array and first assume N is odd for notational convenience. Then

$$\mathbf{p} = \begin{bmatrix} -\frac{N-1}{2} & \dots & \frac{N-1}{2} \end{bmatrix}, \quad (8.126)$$

where we have chosen the origin at the phase center of the array. Then,

$$\mathbf{p}^T \mathbf{1} = 0, \quad (8.127)$$

and

$$\mathbf{p}^T \mathbf{p} = 2 \sum_{i=1}^{\frac{N-1}{2}} i^2 = \frac{N(N^2 - 1)}{12}. \quad (8.128)$$

For N even, (8.127) is still valid and

$$\mathbf{p}^T \mathbf{p} = 2 \sum_{i=1}^{\frac{N}{2}} \left(i - \frac{1}{2} \right)^2 = \frac{N(N^2 - 1)}{12}. \quad (8.129)$$

Then, substituting into (8.115), we obtain

$$C_{CR}(\psi) = \frac{1}{K} \left[\frac{1}{ASNR} + \frac{1}{(ASNR)^2} \right] \frac{6}{(N^2 - 1)}. \quad (8.130)$$

For large N , it is proportional to N^{-2} . Note that the $ASNR$ is proportional to N , so that for a fixed element SNR , the leading term in (8.130) is proportional to N^{-3} .

A more general curve is obtained if we normalize the bound by the BW_{NN} (which is $4/N$ for a standard linear array with uniform weighting):

$$\begin{aligned} C_{CRN}(\psi) &\triangleq \frac{C_{CR}(\psi)}{BW_{NN}^2} \\ &= \frac{3}{8K} \left[\frac{1}{ASNR} + \frac{1}{(ASNR)^2} \right] \frac{1}{1 - \frac{1}{N^2}}. \end{aligned} \quad (8.131)$$

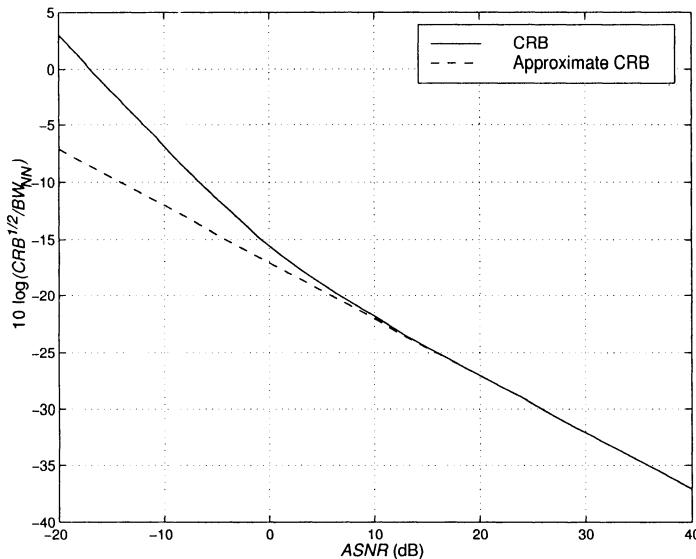


Figure 8.1 Single plane-wave signal: square root of $CRB(\psi)$ vs. $ASNR$, $N = 10$, $K = 100$.

The corresponding result for the normalized version of $C_{ACR}(\psi)$ in (8.120) is the same as in (8.131) except the $(ASNR)^2$ is omitted.

The result for two normalized bounds are plotted versus $ASNR$ for a standard 10-element array in Figure 8.1 and versus K for various $ASNR$ in Figure 8.2. The scale on the vertical axis is $10 \log \left(\sqrt{C_{CR}(\psi)} / BW_{NN} \right)$. Plotted versus $ASNR$, the two bounds are essentially equal above 10 dB. However, plotted versus K , there is a significant difference at lower $ASNR$. The value of plotting our results with this normalization is that the results do not vary significantly (≤ 0.01) with N for $N \geq 10$.

Another view of $C_{CRN}(\psi)$ that is useful is shown in Figure 8.3. Here, we plot the contours of constant $C_{CRN}(\psi)$ versus $ASNR$ and K . We will find this presentation useful when we study the performance of various estimators.

It is useful to introduce several ideas with respect to the $ASNR-K$ plane at this point. Some of ideas will become clearer when we examine specific estimators.

In Figure 8.3, we have identified intervals on each axis in the $ASNR-K$ plane. It is important to note that the exact boundaries vary with the particular application. On the K axis, the three intervals are:

(i) Low sample support

Here $K < 10N$. In this interval, particularly at the lower end, the stability of \mathbf{C}_x is an important issue. In many applications, we are forced

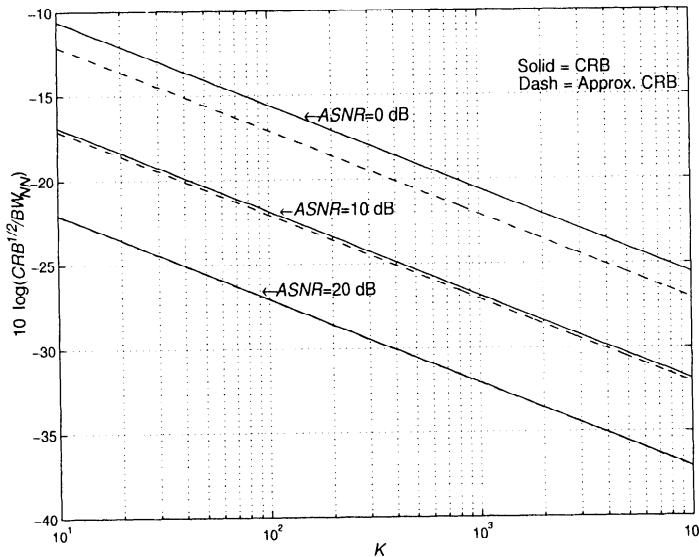


Figure 8.2 Single plane-wave signal: square root of $CRB(\psi)$ vs. K for various $ASNR$.

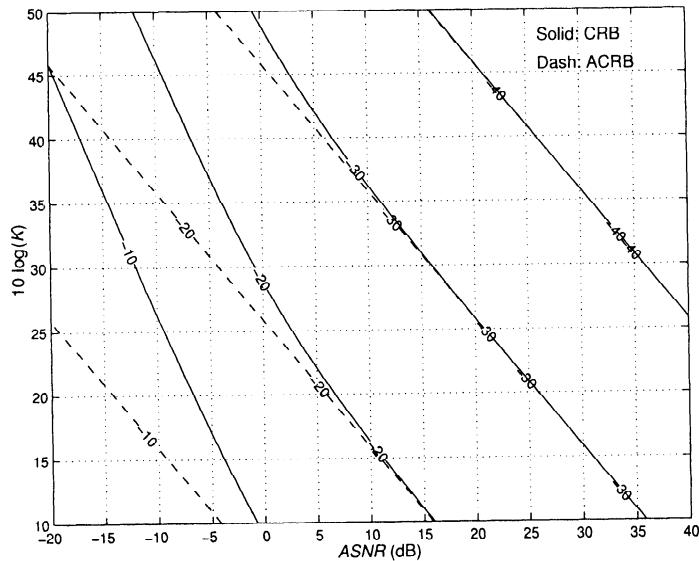


Figure 8.3 Single plane-wave signal: $CRB(\psi)$ in $ASNR-K$ plane.

to work in this interval because of a limited number of observations or because of the non-stationarity of the data.

(ii) Asymptotic

Here K is large and we can use analyses that neglect terms of $O(K^{-1})$. Specifically, the asymptotic behavior of the eigenvalues and eigenvectors that was given in Section 7.2.2 can be exploited. When we discuss the asymptotic properties of estimators (particularly ML estimators), we are referring to this region.

(iii) Transition

In this region we transition from one mode of behavior to another mode. Many analyses focus on identifying the boundaries of the transition region.

On the $ASNR$ axis, the three intervals are:

(i) Low $ASNR$

Here the $ASNR$ is so low that the data do not provide any useful information.

(ii) High $ASNR$

Here the $ASNR$ is high enough that, in many cases, the error is small enough that we can analyze it using a Taylor series expansion around the correct value.

(iii) Transition

In this region we transition from the small error behavior mode to a large error behavior mode.

These intervals provide one way to think about the problem that we will find to be useful.

It is important to note that in many cases the transition regions are bounded by iso-MSE curves in the $ASNR-K$ plane. We develop this behavior later in this section.

Note that we have considered the parameter in ψ -space. Recall that the corresponding angle is

$$\theta = \cos^{-1} \left(\frac{\psi}{\pi} \right). \quad (8.132)$$

Therefore the bound on angle estimate depends on the actual value of the angle. The \mathbf{G} matrix in (8.46) is derived in Problems 8.2.1 and 8.2.2. It is

a diagonal matrix so the mapping of the CRB from ψ -space to θ -space is straightforward.

For the case of two plane-wave signals, the algebraic expressions in the general case are complicated enough that they do not provide much insight into the bound. Most of our results will use numerical evaluation. We note that the expression in (8.130) always lower bounds the multiple-signal case.

In Section 8.4.1.3, we consider uncorrelated signals, and in Section 8.4.1.4 we consider correlated signals.

8.4.1.3 Uncorrelated signals

In this section, we consider the case of two uncorrelated plane-wave signals. However, we do not know *a priori* that they are uncorrelated. We evaluate $C_{CR}(\psi)$ using (8.102). The first example is for arbitrary array geometry.

Example 8.4.3

Consider the case of two plane-wave signals. We evaluate the \mathbf{H} matrix first. For two signals,

$$\mathbf{V} = \begin{bmatrix} \mathbf{v}(\psi_1) & | & \mathbf{v}(\psi_2) \end{bmatrix}. \quad (8.133)$$

To evaluate \mathbf{H} in (8.101), we first find $\mathbf{D}^H \mathbf{D}$,

$$\mathbf{D}^H \mathbf{D} = \begin{bmatrix} \mathbf{d}_1^H \mathbf{d}_1 & | & \mathbf{d}_1^H \mathbf{d}_2 \\ \mathbf{d}_2^H \mathbf{d}_1 & | & \mathbf{d}_2^H \mathbf{d}_2 \end{bmatrix}. \quad (8.134)$$

Each of the terms in (8.134) can be written in terms of the conventional beam pattern

$$\begin{aligned} \mathbf{d}_i^H \mathbf{d}_j &= \frac{\partial \mathbf{v}_i^H}{\partial \psi_i} \frac{\partial \mathbf{v}_j}{\partial \psi_j} = \frac{\partial^2}{\partial \psi_i \partial \psi_j} B_C(\psi_i, \psi_j) \\ &\triangleq B_{\psi_i \psi_j}(\psi_i, \psi_j), \quad i = 1, 2; j = 1, 2. \end{aligned} \quad (8.135)$$

Therefore,

$$\mathbf{D}^H \mathbf{D} = \begin{bmatrix} B_{\psi_1 \psi_1}(\psi_1, \psi_1) & | & B_{\psi_1 \psi_2}(\psi_1, \psi_2) \\ \bar{B}_{\psi_1 \psi_2}^*(\psi_1, \psi_2) & | & \bar{B}_{\psi_2 \psi_2}(\psi_2, \psi_2) \end{bmatrix}. \quad (8.136)$$

The next term of interest is \mathbf{P}_V ,

$$\mathbf{V}^H \mathbf{V} = N \begin{bmatrix} 1 & B(\psi_1, \psi_2) \\ B^*(\psi_1, \psi_2) & 1 \end{bmatrix}. \quad (8.137)$$

and

$$[\mathbf{V}^H \mathbf{V}]^{-1} = \frac{1}{N [1 - |B(\psi_1, \psi_2)|^2]} \begin{bmatrix} 1 & -B(\psi_1, \psi_2) \\ -B^*(\psi_1, \psi_2) & 1 \end{bmatrix}. \quad (8.138)$$

We use B_{12} as abbreviated notation for $B(\psi_1, \psi_2)$. Then, we can write

$$\begin{aligned}\mathbf{Pv} &= \mathbf{V} [\mathbf{V}^H \mathbf{V}]^{-1} \mathbf{V}^H \\ &= \frac{1}{N[1 - |B_{12}|^2]} \left[\begin{array}{c|c} \mathbf{v}_1 & \mathbf{v}_2 \end{array} \right] \left[\begin{array}{cc} 1 & -B_{12} \\ -B_{12}^* & 1 \end{array} \right] \left[\begin{array}{c} \mathbf{v}_1^H \\ \mathbf{v}_2^H \end{array} \right] \\ &= \frac{1}{N[1 - |B_{12}|^2]} \{ \mathbf{v}_1 \mathbf{v}_1^H - B_{12} \mathbf{v}_1 \mathbf{v}_2^H - B_{12}^* \mathbf{v}_2 \mathbf{v}_1^H + \mathbf{v}_2 \mathbf{v}_2^H \}.\end{aligned}\quad (8.139)$$

The ij th element of the second term in \mathbf{H} is,

$$\begin{aligned}[\mathbf{D}^H \mathbf{Pv} \mathbf{D}]_{ij} &= \frac{1}{N[1 - |B_{12}|^2]} \cdot \{ \mathbf{d}_i^H [\mathbf{v}_1 \mathbf{v}_1^H - B_{12} \mathbf{v}_1 \mathbf{v}_2^H \\ &\quad - B_{21}^* \mathbf{v}_2 \mathbf{v}_1^H + \mathbf{v}_2 \mathbf{v}_2^H] \mathbf{d}_j \}, \quad i, j = 1, 2.\end{aligned}\quad (8.140)$$

For conjugate symmetric \mathbf{v} ,

$$\mathbf{d}_i^H \mathbf{v}_i = \mathbf{v}_i^H \mathbf{d}_i = 0, \quad i = 1, 2. \quad (8.141)$$

Thus, (8.140) reduces to

$$\begin{aligned}\mathbf{D}^H \mathbf{Pv} \mathbf{D} &= \frac{1}{N[1 - |B_{12}|^2]} \\ &\times \left[\begin{array}{c|c} |\mathbf{d}_1^H \mathbf{v}_2|^2 & -B_{21}^* (\mathbf{d}_1^H \mathbf{v}_2) (\mathbf{v}_1^H \mathbf{d}_2) \\ -B_{12} (\mathbf{d}_2^H \mathbf{v}_1) (\mathbf{v}_2^H \mathbf{d}_1) & |\mathbf{d}_2^H \mathbf{v}_1|^2 \end{array} \right].\end{aligned}\quad (8.142)$$

From (8.116)

$$\mathbf{d}_1^H \mathbf{v}_2 = \frac{\partial B_c(\psi_1, \psi_2)}{\partial \psi_1} = \frac{\partial B_c(\psi_1 - \psi_2)}{\partial \psi_1} = -\frac{\partial B_c(\psi_1 - \psi_2)}{\partial \psi_2} = -B_{\psi_2}(\Delta\psi), \quad (8.143)$$

$$\mathbf{d}_2^H \mathbf{v}_1 = \frac{\partial B_c(\psi_1 - \psi_2)}{\partial \psi_2} = B_{\psi_2}(\Delta\psi), \quad (8.144)$$

where

$$\Delta\psi = \psi_1 - \psi_2. \quad (8.145)$$

Then,

$$H_{11} = B_{\psi_1 \psi_1}(\psi_1, \psi_1) - \frac{1}{N[1 - |B_{12}|^2]} |B_{\psi_2}(\Delta\psi)|^2, \quad (8.146)$$

$$H_{22} = B_{\psi_2 \psi_2}(\psi_2, \psi_2) - \frac{1}{N[1 - |B_{12}|^2]} |B_{\psi_2}(\Delta\psi)|^2, \quad (8.147)$$

$$H_{12} = B_{\psi_1 \psi_2}(\psi_1, \psi_2) - \frac{B_{12}}{N[1 - |B_{12}|^2]} |B_{\psi_2}(\Delta\psi)|^2, \quad (8.148)$$

and

$$H_{21} = H_{12}^*. \quad (8.149)$$

The final step is to evaluate the Hadamard product. For the special case of uncorrelated sources,

$$[\mathbf{S}_f]_{12} = [\mathbf{S}_f]_{21} = 0. \quad (8.150)$$

and

$$\mathbf{S}_f = \text{diag} [\sigma_{s1}^2, \sigma_{s2}^2] = \Lambda_f, \quad (8.151)$$

and

$$\Lambda'_f = \frac{\Lambda_f}{\sigma_w^2}. \quad (8.152)$$

Then (8.106) can be written as

$$\mathbf{C}_{CR}(\psi) = \frac{1}{2K} \left\{ \text{Re} \left\{ \left[\Lambda'_f [\mathbf{I} + \mathbf{V}^H \mathbf{V} \Lambda'_f]^{-1} \mathbf{V}^H \mathbf{V} \Lambda'_f \right] \odot \mathbf{H}^T \right\} \right\}^{-1}. \quad (8.153)$$

Note that, even with uncorrelated signals, the matrix in the first bracket will not be diagonal unless the conditions following (8.109) are satisfied. If these conditions are satisfied, then we can use the approximation in (8.110). The matrix in the first bracket in (8.153) will be diagonal and the resulting Hadamard product will be diagonal. Thus,

$$C_{ACR}(\psi_i) = \frac{1}{2K} \frac{H_{ii}^{-1}}{SNR_i}, i = 1, 2, \quad (8.154)$$

where H_{ii} is given by (8.146) and (8.147).

In the next example, we specialize the results to the case of a standard linear array.

Example 8.4.4 (continuation)

For a standard linear array the expression in (8.154) reduces to

$$C_{ACR}(\psi_1) = \frac{6}{K \cdot ASNR(N^2 - 1)} \left\{ 1 - \frac{12 |B_{\psi_2}(\Delta\psi)|^2}{N^2(N^2 - 1) [1 - |B_{12}|^2]} \right\}^{-1}. \quad (8.155)$$

The term in braces represents the increase in the bound due to the presence of the second signal. Note that the bound does not depend on the power of the second signal.

This is because the result is asymptotic and as the error becomes small, we find that the ML estimator places a null on the second signal (recall the discussion in Section 6.3). Later we shall see that the result is not true in the non-asymptotic area.

In Figure 8.4, we plot the square root of the normalized CRB versus $ASNR$ for $N = 10$, $K = 100$, and various values of $\Delta\psi/BW_{NN}$. We show the bound given by (8.110) and the exact bound given by (8.106). We also show the single-signal bound given by (8.130).

We see that, as Δ decreases, $C_{CR}(\psi)$ separates from $C_{ACR}(\psi)$ at higher $ASNR$. For $\Delta > 0.5BW_{NN}$, the two-signal bound is essentially the same as the single-signal bound for $ASNR > 0$ dB.

In Figure 8.5, we plot the exact CRB (8.106) versus normalized signal separation for various SNR . We see that the bound is proportional to $(\Delta u)^{-2}$ over the region $\Delta u/BW_{NN} \leq 1.0$ and is essentially constant for $\Delta u/BW_{NN} > 1.0$. We indicate the HPBW, $0.5BW_{NN}$, and BW_{NN} . The separation, $\Delta u = 0.5BW_{NN}$, is referred to as the Rayleigh resolution value (recall the discussion in Chapter 2).

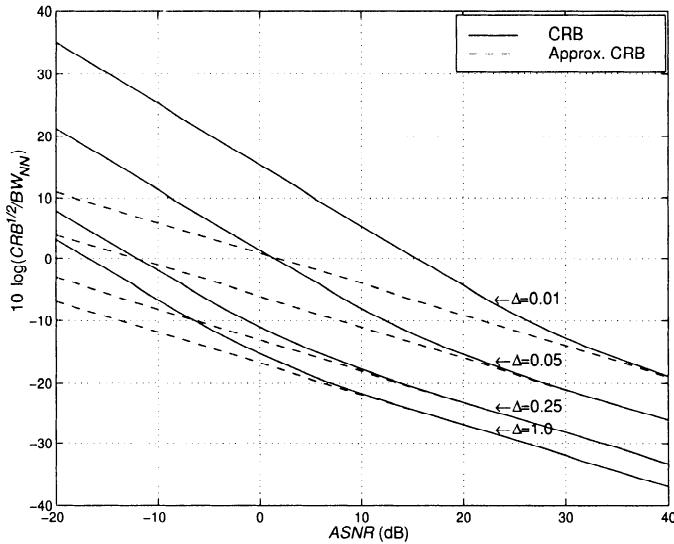


Figure 8.4 Square root of normalized CRB versus $ASNR : N = 10, K = 100$, two signals, various $\Delta\psi/BW_{NN}$.

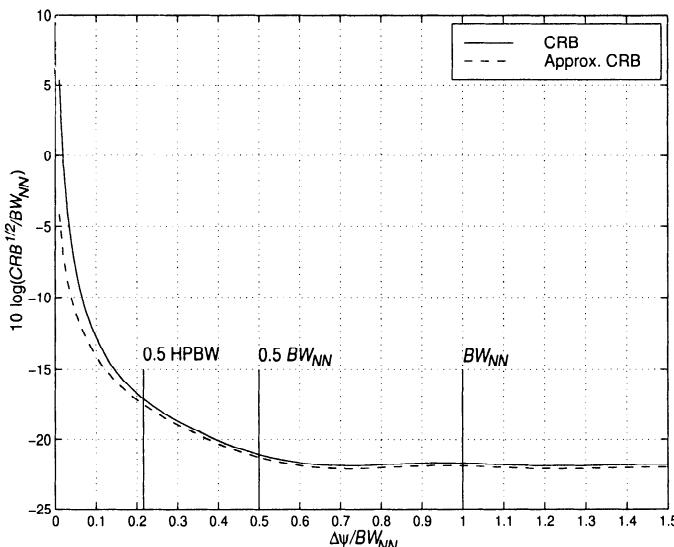


Figure 8.5 Square root of normalized CRB versus $\Delta\psi/BW_{NN}$ (normalized signal separation in ψ -space): two signals, $ASNR = 10$ dB, $N = 10, K = 100$.

8.4.1.4 Correlated signals

When the signals are correlated or coherent, we evaluate (8.106) numerically without any intermediate manipulation. One of the key results that comes out of the discussion is the importance of the phase of the correlation coefficient. This result was pointed out in [WF90] and [WF93].

In [YB92], Yau and Bresler developed conditions for the worst case asymptotic conditional CRB. For two signals, they obtained an analytic expression for the phase angle of the worst case CRB. They show that the worst case asymptotic conditional CRB (ACRB) is given by fully coherent signals with

$$S_{12} = \sqrt{S_{11} S_{22}} \exp[j \arg(H_{12})], \quad (8.156)$$

where \mathbf{H} is defined in (8.101).

However, whenever \mathbf{v} is conjugate symmetric, \mathbf{H} is real. Therefore, the worst case phase is 0° or 180° for standard linear arrays (and any conjugate symmetric array).

We illustrate the importance of the phase with a simple example.

Example 8.4.5

Consider a standard 10-element linear array. The $ASNR$ is 20 dB and $K = 100$. In Figure 8.6, we plot the normalized CRB given by (8.106) and the ACRB given by (8.110) versus the phase angle of ρ for various values of $|\rho|$. In Figure 8.6(a), the signal separation is $0.05BW_{NN}$. We see that the CRB has peaks at

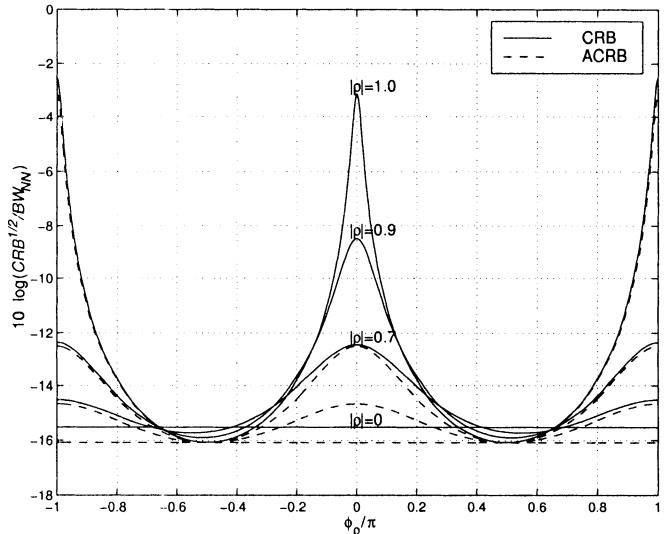
$$\phi_\rho = 0^\circ, \quad 180^\circ, \quad (8.157)$$

and that for larger values of $|\rho|$, the difference between the peak and the minimum value can be significant. We observe that, in the vicinity of $\phi_\rho = 90^\circ$, the CRB is lower for correlated signals than for uncorrelated signals. In Figure 8.6(b), the normalized CRB is plotted versus $|\rho|$ for various Δu . The phase angle, ϕ_ρ , equals 90° . For very small separation, there is a significant decrease as $|\rho|$ increases. For $\Delta u \geq 0.05BW_{NN}$, the decrease is small.

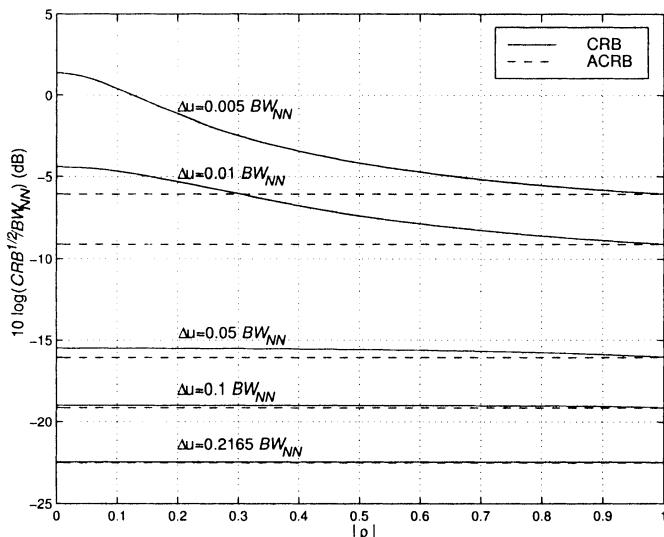
Note that our example uses both the CRB and the ACRB, whereas the analysis in [YB92] was for the ACRB.

This example illustrates the importance of considering the phase of the correlation coefficient when one models correlated or coherent signals. We note that it is important to have the phase reference point at the center of the array when studying coherent signals. If one uses the end element for the phase reference point, artifacts are introduced in the result. This issue can be verified by redoing Example 8.4.5 with the phase reference at the end element.

In Figure 8.7, we show the normalized bounds versus N for a small separation $\Delta\psi = 0.05BW_{NN}$ and an $ASNR = 20$ dB. The phase of ρ is zero.



(a)



(b)

Figure 8.6 Square root of normalized CRB and ACRB for various $|\rho|$: two signals, $N = 10, K = 100$: (a) normalized CRB and ACRB versus $\phi_\rho/\pi, \Delta u = 0.05 BW_{NN}$; (b) normalized CRB and ACRB versus $|\rho|, \phi_\rho = \pi/2$, various Δu .

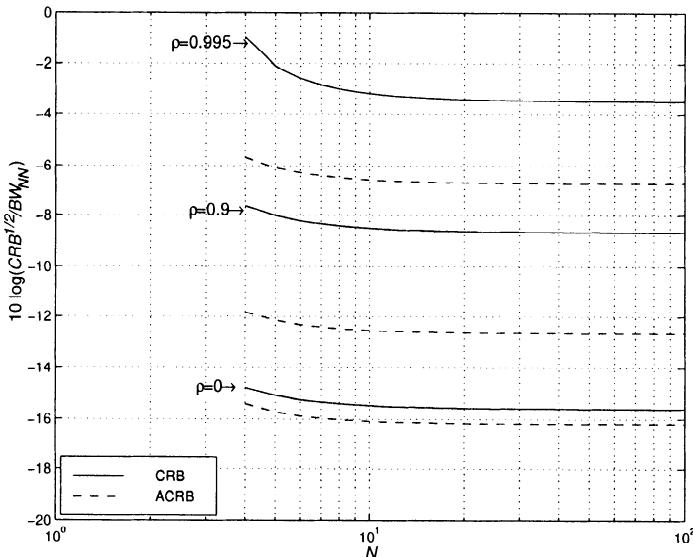


Figure 8.7 Square root of normalized CRB and ACRB versus number of sensors: two signals with $\Delta\psi = 0.5BW_{NN}$, $ASNR = 20$ dB, $K = 100$, various ρ .

We see that for $N \geq 5$, the normalized error is independent of the number of sensors. We also see that the bound increases significantly as $|\rho|$ approaches one.

In Figure 8.8, we show the CRB versus $ASNR$ for various ρ . In this case, $\phi_\rho = 0$. In Figure 8.8(a), $\Delta\psi = 0.05BW_{NN}$. In Figure 8.8(b), $\Delta\psi = 0.2165BW_{NN}$.

For more than two signals, it is not clear what the worst and best case phase relations are. However, for conjugate symmetric \mathbf{v} , \mathbf{H} is real and we would anticipate that the 0° or 180° relationship would continue to be the worst case. Several of the problems analyze the correlated signal case in more detail.

In some cases, having additional *a priori* information decreases the CRB. For the case of coherent signals, Stoica et al. [SOVM96] have shown that knowing that the signals are coherent does not decrease the CRB.

8.4.1.5 Closely spaced signals

As we have seen from our examples, as the signal separation decreases, the CRB increases. This corresponds to the intuitive result that the parame-

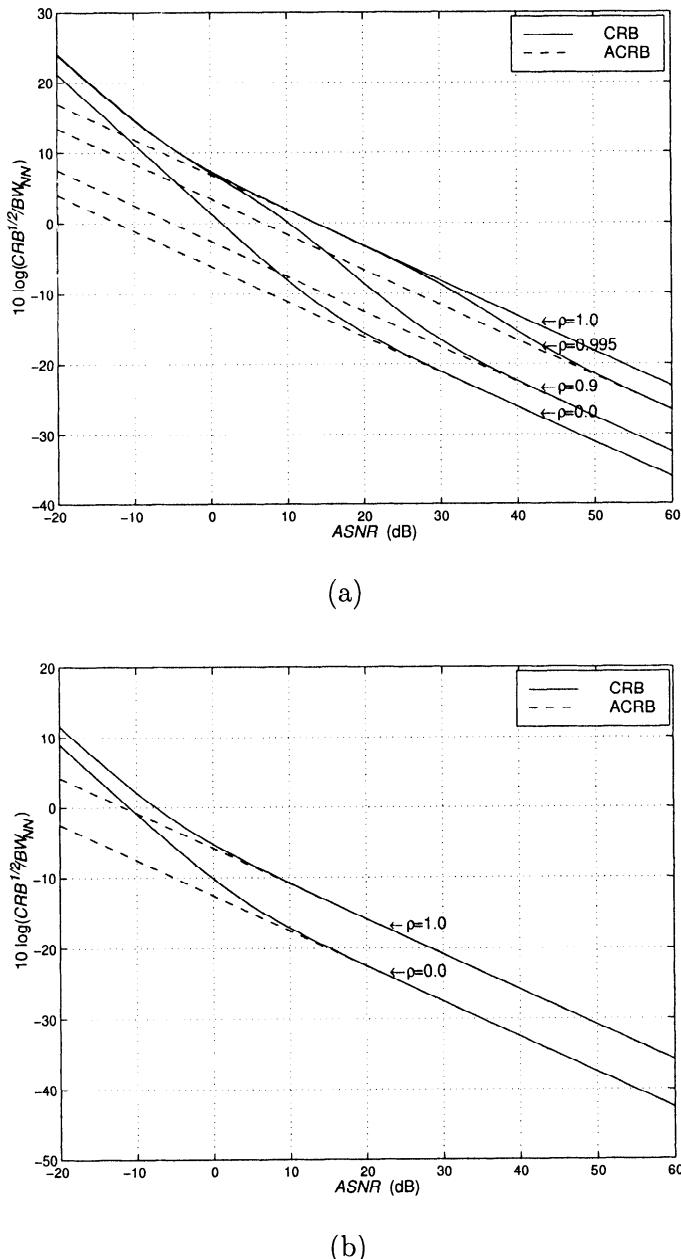


Figure 8.8 Square root of normalized CRB and ACRB versus $ASNR$: various values of ρ , phase of ρ is zero: two signals, $N = 10$, $K = 100$: (a) $\Delta\psi = 0.05BW_{NN}$; (b) $\Delta\psi = 0.2165BW_{NN}$.

ters of closely spaced signals will be harder to estimate. Swingler [Swi93] developed approximate expressions for the two-signal case.

For the case of D ($D > 2$), Lee ([Lee92], [Lee94]) has developed compact analytic expressions for the CRB in the case of small signal separation. The reader is referred to these references for a detailed discussion. The key result is that, for closely spaced signals,

$$CRB \propto (\Delta\psi)^{-2(D-1)}, \quad (8.158)$$

where D is the the number of signals.

8.4.1.6 Summary

In this section, we have derived the CRB for the case of an unknown source signal spectral matrix and unknown noise level. In a large number of applications, particularly passive sonar, surveillance, and radar applications, this is the appropriate model to use. Several points are worth reemphasizing:

- (i) For two uncorrelated signals, the bound is proportional to $ASNR^{-1}$ for large $ASNR$.
- (ii) For correlated signals, the phase of the correlation coefficient is very important.
- (iii) For closely spaced signals,

$$CRB \propto (\Delta\psi)^{-2(D-1)},$$

where D is the number of signals.

- (iv) The CRB is proportional to K^{-1} .
- (v) For standard linear arrays with $N > 10$, the normalized bound plotted versus $ASNR$ does not depend on N .

8.4.2 Gaussian Model: Uncorrelated Signals with Unknown Power

In this section, we consider the case of the Gaussian model in which we know *a priori* that the signals are uncorrelated. As part of the derivation we also find the CRB for the case of known source spectrum. The CRB for uncorrelated signals was used in many of the early studies of DOA estimation (e.g., Schmidt [Sch81] or Appendix A of Barabell et al. [BCD+84]). Porat and

Friedlander [PF88] use a high SNR approximation to it in their discussion of the relative efficiency of MUSIC. Jansson et al. [JGO99] have derived an expression for the uncorrelated CRB.

The spatial spectral matrix is

$$\mathbf{S}_x = \mathbf{V}\mathbf{S}_f\mathbf{V}^H + \sigma_w^2\mathbf{I}. \quad (8.159)$$

We can write \mathbf{S}_x as

$$\mathbf{S}_x = \sum_{i=1}^D \sigma_i^2 \mathbf{v}(\psi_i) \mathbf{v}^H(\psi_i) + \sigma_w^2 \mathbf{I}. \quad (8.160)$$

There are $2D + 1$ unknown parameters,

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\psi}^T & \boldsymbol{\sigma}^{2T} & \sigma_w^2 \end{bmatrix}^T, \quad (8.161)$$

where

$$\boldsymbol{\sigma}^2 \triangleq \begin{bmatrix} \sigma_1^2 & \sigma_2^2 & \cdots & \sigma_D^2 \end{bmatrix}^T. \quad (8.162)$$

We divide $\boldsymbol{\theta}$ into two parts as in (8.39),

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\psi}^T & \boldsymbol{\theta}_u^T \end{bmatrix}^T, \quad (8.163)$$

where

$$\boldsymbol{\theta}_u = \begin{bmatrix} \boldsymbol{\sigma}^{2T} & \sigma_w^2 \end{bmatrix}^T \quad (8.164)$$

is a $D + 1$ vector containing the unwanted parameters.

The Fisher information can be partitioned as

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi}} & \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\theta}_u} \\ \mathbf{J}_{\boldsymbol{\theta}_u\boldsymbol{\psi}} & \mathbf{J}_{\boldsymbol{\theta}_u\boldsymbol{\theta}_u} \end{bmatrix}. \quad (8.165)$$

Then,

$$\mathbf{J}^{-1} = \begin{bmatrix} (\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi}} - \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\theta}_u} \mathbf{J}_{\boldsymbol{\theta}_u\boldsymbol{\theta}_u}^{-1} \mathbf{J}_{\boldsymbol{\theta}_u\boldsymbol{\psi}})^{-1} & \cdot \\ \cdot & \cdot \end{bmatrix}, \quad (8.166)$$

where only the upper left block matrix is of interest, and

$$\mathbf{C}_{CR}(\boldsymbol{\psi}) = \left[\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi}} - \mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\theta}_u} \mathbf{J}_{\boldsymbol{\theta}_u\boldsymbol{\theta}_u}^{-1} \mathbf{J}_{\boldsymbol{\theta}_u\boldsymbol{\psi}} \right]^{-1}. \quad (8.167)$$

We next derive the entries in \mathbf{J} . In the derivation of $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi}}$, we will not assume that \mathbf{S}_f is diagonal, so the resulting $\mathbf{J}_{\boldsymbol{\psi}\boldsymbol{\psi}}$ will provide a bound for the general known signal spectrum case.

The ij term in $\mathbf{J}_{\psi\psi}$ is

$$[\mathbf{J}_{\psi\psi}]_{ij} = K \operatorname{tr} \left[\mathbf{S}_x^{-1} \frac{\partial \mathbf{S}_x}{\partial \psi_i} \mathbf{S}_x^{-1} \frac{\partial \mathbf{S}_x}{\partial \psi_j} \right]. \quad (8.168)$$

The derivative with respect to ψ_i is

$$\frac{\partial \mathbf{S}_x}{\partial \psi_i} = \frac{\partial \mathbf{V}(\psi)}{\partial \psi_i} \mathbf{S}_f \mathbf{V}^H(\psi) + \mathbf{V} \mathbf{S}_f \frac{\partial \mathbf{V}^H(\psi)}{\partial \psi_i}. \quad (8.169)$$

Now define

$$\mathbf{D} = \sum_{i=1}^D \frac{\partial \mathbf{V}(\psi)}{\partial \psi_i} = \begin{bmatrix} \frac{\partial \mathbf{v}(\psi_1)}{\partial \psi_1} & \dots & \frac{\partial \mathbf{v}(\psi_D)}{\partial \psi_D} \end{bmatrix}. \quad (8.170)$$

Then, we can write

$$\frac{\partial \mathbf{V}(\psi)}{\partial \psi_i} = \mathbf{D} \mathbf{e}_i \mathbf{e}_i^T, \quad (8.171)$$

where \mathbf{e}_i is a $D \times 1$ vector whose i th element is unity and whose remaining elements are zero (see A.101):

$$[\mathbf{J}_{\psi\psi}]_{ij} = K \operatorname{tr} \left[\mathbf{S}_x^{-1} \left[\mathbf{D} \mathbf{e}_i \mathbf{e}_i^T \mathbf{S}_f \mathbf{V}^H + \mathbf{V} \mathbf{S}_f \mathbf{e}_i \mathbf{e}_i^T \mathbf{D}^H \right] \right. \\ \left. \times \mathbf{S}_x^{-1} \left[\mathbf{D} \mathbf{e}_j \mathbf{e}_j^T \mathbf{S}_f \mathbf{V}^H + \mathbf{V} \mathbf{S}_f \mathbf{e}_j \mathbf{e}_j^T \mathbf{D}^H \right] \right], \quad (8.172)$$

or

$$[\mathbf{J}_{\psi\psi}]_{ij} = K \operatorname{tr} \left[\mathbf{S}_x^{-1} \mathbf{D} \mathbf{e}_i \mathbf{e}_i^T \mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{D} \mathbf{e}_j \mathbf{e}_j^T \mathbf{S}_f \mathbf{V}^H \right. \\ \left. + \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f \mathbf{e}_i \mathbf{e}_i^T \mathbf{D}^H \mathbf{S}_x^{-1} \mathbf{D} \mathbf{e}_j \mathbf{e}_j^T \mathbf{S}_f \mathbf{V}^H \right. \\ \left. + \mathbf{S}_x^{-1} \mathbf{D} \mathbf{e}_i \mathbf{e}_i^T \mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V}^H \mathbf{S}_f \mathbf{e}_j \mathbf{e}_j^T \mathbf{D}^H \right. \\ \left. + \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f \mathbf{e}_i \mathbf{e}_i^T \mathbf{D}^H \mathbf{S}_x^{-1} \mathbf{V}^H \mathbf{S}_f \mathbf{e}_j \mathbf{e}_j^T \mathbf{D}^H \right]. \quad (8.173)$$

Using the trace property in (A.27),

$$[\mathbf{J}_{\psi\psi}]_{ij} = K \operatorname{tr} \left[\left(\mathbf{e}_j^T \mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{D} \mathbf{e}_i \right) \left(\mathbf{e}_i^T \mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{D} \mathbf{e}_j \right) \right. \\ \left. + \left(\mathbf{e}_j^T \mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f \mathbf{e}_i \right) \left(\mathbf{e}_i^T \mathbf{D}^H \mathbf{S}_x^{-1} \mathbf{D} \mathbf{e}_j \right) \right. \\ \left. + \left(\mathbf{e}_j^T \mathbf{D}^H \mathbf{S}_x^{-1} \mathbf{D} \mathbf{e}_i \right) \left(\mathbf{e}_i^T \mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f \mathbf{e}_j \right) \right. \\ \left. + \left(\mathbf{e}_j^T \mathbf{D}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f \mathbf{e}_i \right) \left(\mathbf{e}_i^T \mathbf{D}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f \mathbf{e}_j \right) \right]. \quad (8.174)$$

Now all the terms are scalars and the trace can be removed. The fourth term is the conjugate of the first term and the third term is the conjugate of the second term. Therefore, (8.174) can be written as

$$\begin{aligned} [\mathbf{J}_{\psi\psi}]_{ij} &= 2K \operatorname{Re} \left\{ \left(\mathbf{e}_i^T \mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{D} \mathbf{e}_j \right) \left(\mathbf{e}_j^T \mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{D} \mathbf{e}_i \right) \right. \\ &\quad \left. + \left(\mathbf{e}_i^T \mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f \mathbf{e}_j \right) \left(\mathbf{e}_j^T \mathbf{D}^H \mathbf{S}_x^{-1} \mathbf{D} \mathbf{e}_i \right) \right\}, \end{aligned} \quad (8.175)$$

or, in matrix form,

$$\begin{aligned} \mathbf{J}_{\psi\psi} &= 2K \operatorname{Re} \left\{ \left[\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{D} \right] \odot \left[\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{D} \right]^T \right. \\ &\quad \left. + \left[\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f \right] \odot \left[\mathbf{D}^H \mathbf{S}_x^{-1} \mathbf{D} \right]^T \right\}, \end{aligned} \quad (8.176)$$

which is the desired result. The expression in (8.176) can also be written as

$$\begin{aligned} \mathbf{J}_{\psi\psi} &= 2K \operatorname{Re} \left\{ \left[\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{D} \right]^T \odot \left[\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{D} \right] \right. \\ &\quad \left. + \left[\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f \right]^T \odot \left[\mathbf{D}^H \mathbf{S}_x^{-1} \mathbf{D} \right] \right\}. \end{aligned} \quad (8.177)$$

The expression in (8.176) is for general \mathbf{S}_f . In order to derive the other terms in the \mathbf{J} matrix, we assume that (8.160) is satisfied.

We partition the $\boldsymbol{\theta}_u$ vector as in (8.164), so the cross-matrices can be written as

$$\mathbf{J}_{\psi\theta_u} = \begin{bmatrix} \mathbf{J}_{\psi s} \\ \mathbf{J}_{\psi n} \end{bmatrix}, \quad (8.178)$$

and

$$\mathbf{J}_{\theta_u\theta_u} = \begin{bmatrix} \mathbf{J}_{ss} & \mathbf{J}_{ns} \\ \mathbf{J}_{sn} & \mathbf{J}_{nn} \end{bmatrix}, \quad (8.179)$$

The first cross-matrix is

$$\begin{aligned} J_{\psi_i\sigma_j^2} &= K \operatorname{tr} \left\{ \mathbf{S}_x^{-1} \sigma_i^2 (\mathbf{d}_i \mathbf{v}_i^H + \mathbf{v}_i \mathbf{d}_i^H) \mathbf{S}_x^{-1} \mathbf{v}_j \mathbf{v}_j^H \right\} \\ &= K \operatorname{tr} \left\{ \mathbf{S}_x^{-1} \sigma_i^2 \mathbf{d}_i \mathbf{v}_i^H \mathbf{S}_x^{-1} \mathbf{v}_j \mathbf{v}_j^H + \mathbf{S}_x^{-1} \sigma_i^2 \mathbf{v}_i \mathbf{d}_i^H \mathbf{S}_x^{-1} \mathbf{v}_j \mathbf{v}_j^H \right\} \\ &= K \left\{ (\mathbf{v}_j^H \mathbf{S}_x^{-1} \mathbf{d}_i) (\sigma_i^2 \mathbf{v}_i^H \mathbf{S}_x^{-1} \mathbf{v}_j) + (\mathbf{v}_j^H \mathbf{S}_x^{-1} \mathbf{v}_i \sigma_i^2) (\mathbf{d}_i^H \mathbf{S}_x^{-1} \mathbf{v}_j) \right\}, \end{aligned} \quad (8.180)$$

or

$$J_{\psi_i \sigma_j^2} = 2K \operatorname{Re} \left\{ (\sigma_i^2 \mathbf{v}_i^H \mathbf{S}_x^{-1} \mathbf{v}_j) (\mathbf{v}_j^H \mathbf{S}_x^{-1} \mathbf{d}_i) \right\}, \quad (8.181)$$

or

$$J_{\psi_i \sigma_j^2} = 2K \operatorname{Re} \left\{ [\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V}]_{ij} [\mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{D}]_{ji} \right\}. \quad (8.182)$$

In matrix form,

$$\boxed{\mathbf{J}_{\psi_s} = 2K \operatorname{Re} \left\{ [\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V}] \odot [\mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{D}]^T \right\}}. \quad (8.183)$$

The next cross-matrix is

$$\begin{aligned} J_{\psi_i \sigma_w^2} &= K \operatorname{tr} \left\{ \mathbf{S}_x^{-1} \sigma_i^2 (\mathbf{d}_i \mathbf{v}_i^H + \mathbf{v}_i \mathbf{d}_i^H) \mathbf{S}_x^{-1} \right\} \\ &= K \operatorname{tr} \left\{ \mathbf{S}_x^{-1} \sigma_i^2 \mathbf{d}_i \mathbf{v}_i^H \mathbf{S}_x^{-1} + \mathbf{S}_x^{-1} \sigma_i^2 \mathbf{v}_i \mathbf{d}_i^H \mathbf{S}_x^{-1} \right\} \\ &= K \left\{ \sigma_i^2 \mathbf{v}_i^H \mathbf{S}_x^{-2} \mathbf{d}_i + \mathbf{d}_i^H \mathbf{S}_x^{-2} \mathbf{v}_i \sigma_i^2 \right\}, \end{aligned} \quad (8.184)$$

which reduces to

$$\begin{aligned} J_{\psi_i \sigma_w^2} &= 2K \operatorname{Re} \left\{ \sigma_i^2 \mathbf{v}_i^H \mathbf{S}_x^{-2} \mathbf{d}_i \right\} \\ &= 2K \operatorname{Re} \left\{ [\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-2} \mathbf{D}]_{ii} \right\}. \end{aligned} \quad (8.185)$$

In matrix form

$$\boxed{\mathbf{J}_{\psi_n} = 2K \operatorname{diag} \left[\operatorname{Re}(\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-2} \mathbf{D}) \right]}. \quad (8.186)$$

The signal power matrix is

$$J_{\sigma_i^2 \sigma_j^2} = K \operatorname{tr} \left\{ \mathbf{S}_x^{-1} \mathbf{v}_i \mathbf{v}_i^H \mathbf{S}_x^{-1} \mathbf{v}_j \mathbf{v}_j^H \right\}. \quad (8.187)$$

or

$$\begin{aligned} J_{\sigma_i^2 \sigma_j^2} &= K \left\{ (\mathbf{v}_i^H \mathbf{S}_x^{-1} \mathbf{v}_j) (\mathbf{v}_j^H \mathbf{S}_x^{-1} \mathbf{v}_i) \right\} \\ &= K \left\{ [\mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V}]_{ij} [\mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V}]_{ji} \right\}. \end{aligned} \quad (8.188)$$

In matrix form

$$\boxed{\mathbf{J}_{ss} = K \left\{ [\mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V}] \odot [\mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V}]^T \right\}}. \quad (8.189)$$

The noise-related matrices are

$$J_{\sigma_i^2 \sigma_w^2} = K \operatorname{tr} \left\{ \mathbf{S}_x^{-1} \mathbf{v}_i \mathbf{v}_i^H \mathbf{S}_x^{-1} \right\}, \quad (8.190)$$

or

$$\begin{aligned} J_{\sigma_i^2 \sigma_w^2} &= K \mathbf{v}_i^H \mathbf{S}_{\mathbf{x}}^{-2} \mathbf{v}_i \\ &= K \left[\mathbf{V}^H \mathbf{S}_{\mathbf{x}}^{-2} \mathbf{V} \right]_{ii}. \end{aligned} \quad (8.191)$$

In matrix form

$$\boxed{\mathbf{J}_{\mathbf{s}\mathbf{n}} = K \text{diag} \left[\mathbf{V}^H \mathbf{S}_{\mathbf{x}}^{-2} \mathbf{V} \right].} \quad (8.192)$$

Similarly,

$$J_{\sigma_w^2 \sigma_w^2} = K \text{tr} \left\{ \mathbf{S}_{\mathbf{x}}^{-1} \mathbf{S}_{\mathbf{x}}^{-1} \right\} = K \text{tr} \left\{ \mathbf{S}_{\mathbf{x}}^{-2} \right\}, \quad (8.193)$$

and

$$\boxed{\mathbf{J}_{nn} = K \text{tr} \left\{ \mathbf{S}_{\mathbf{x}}^{-2} \right\}.} \quad (8.194)$$

In addition to the six matrices in boxes ((8.176), (8.183), (8.186), (8.189), (8.192), and (8.194)) there are three Hermitian transposes,

$$\mathbf{J}_{s\psi} = \mathbf{J}_{\psi s}^H, \quad (8.195)$$

$$\mathbf{J}_{n\psi} = \mathbf{J}_{\psi n}^H, \quad (8.196)$$

and

$$\mathbf{J}_{ns} = \mathbf{J}_{sn}^H. \quad (8.197)$$

We substitute these eight equations into (8.178) and (8.179) and the result along with (8.176) into (8.167). All of the terms in (8.167) are explicitly defined and we can evaluate the uncorrelated CRB. We consider a simple example to illustrate the behavior.

Example 8.4.6

Consider a standard 10-element linear array. We plot the uncorrelated CRB given by (8.167) and the standard CRB given by (8.106) versus $ASNR_1$ for several signal separations; $0.01BW_{NN}$, $0.05BW_{NN}$, $0.25BW_{NN}$, and $0.75BW_{NN}$. We consider three signal models:

- (i) Figure 8.9: $ASNR_2 = ASNR_1$
- (ii) Figure 8.10: $ASNR_2 = ASNR_1 + 10$ dB
- (iii) Figure 8.11: $ASNR_2 = ASNR_1 + 20$ dB

In each figure, the (a) plot corresponds to signal 1 and the (b) plot corresponds to signal 2. The value of K is 100.

We see that:

- (i) For unequal power signals, the bound for the lower power signal is almost the same whether we use the standard or uncorrelated CRB.

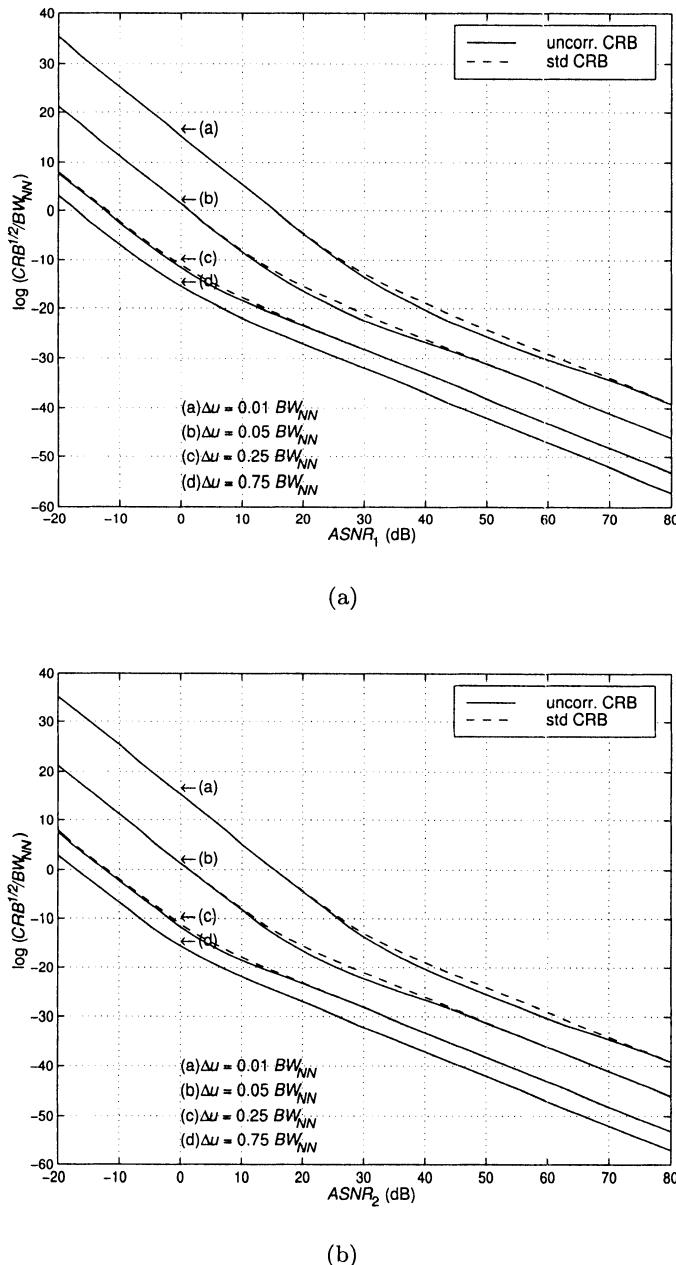


Figure 8.9 Normalized CRBs versus $ASNR$: two signals, $N = 10, K = 100$; $ASNR_2 = ASNR_1$: (a) CRBs for signal 1 versus $ASNR_1$; (b) CRBs for signal 2 versus $ASNR_2$.

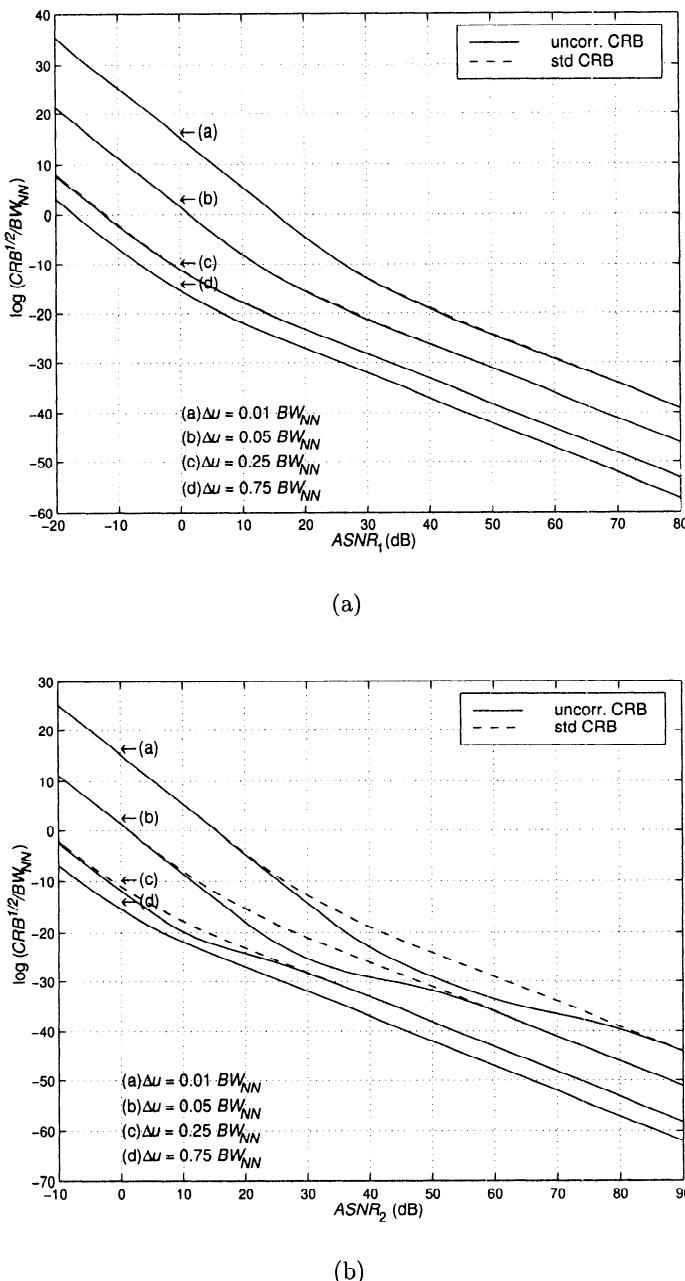


Figure 8.10 Normalized CRBs versus ASNR: two signals, $N = 10, K = 100$; $ASNR_2 = ASNR_1 + 10$ dB: (a) CRBs for signal 1 versus $ASNR_1$; (b) CRBs for signal 2 versus $ASNR_2$.

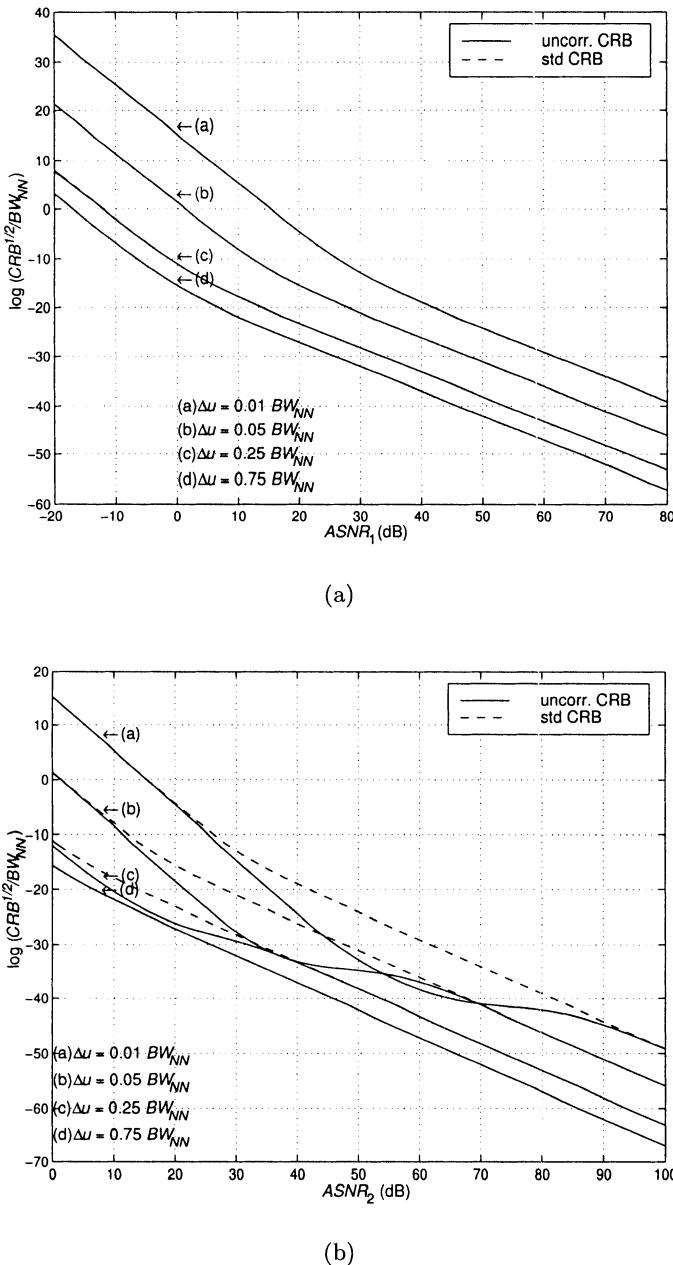


Figure 8.11 Normalized CRBs versus $ASNR$: two signals, $N = 10, K = 100$; $ASNR_2 = ASNR_1 + 20$ dB: (a) CRBs for signal 1 versus $ASNR_1$; (b) CRBs for signal 2 versus $ASNR_2$.

(ii) The bound for the higher power signal can be significantly lower using the uncorrelated signal model rather than the general correlation model, indicating a possible improvement in DOA estimation for schemes that use this knowledge. The difference is most pronounced:

(a) as Δu decreases

(b) as the difference in $ASNR$ increases.

These results do not change significantly for $N \geq 6$. There are similar trends for $N = 3, 4$, and 5 .

Jansson et al. [JGO99] give several comparisons of the uncorrelated CRB and the general CRB. They also derive an estimator that achieves the uncorrelated CRB asymptotically.

8.4.3 Gaussian Model: Known Signal Spectrum

In this section we consider the case of the Gaussian model with a known signal spectrum and known σ_w^2 . The result was derived in Section 8.4.2 (8.176). Now the parameter vector θ is

$$\theta \stackrel{\triangle}{=} \psi. \quad (8.198)$$

The Fisher information matrix is the term in (8.176). Thus the CRB is⁶

$$C_{CR}(\psi) = \mathbf{J}_{\psi\psi}^{-1} = \frac{1}{2K} \left\{ Re \left\{ \left(\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f \right) \odot \left(\mathbf{D}^H \mathbf{S}_x^{-1} \mathbf{D} \right)^T \right. \right. \\ \left. \left. + \left(\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{D} \right) \odot \left(\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{D} \right)^T \right\} \right\}^{-1}, \quad (8.199)$$

where \mathbf{D} was defined in (8.170) and \odot denotes the Hadamard product. The second term in the upper left partition in (8.166) is always a non-negative definite matrix. Therefore, if \mathbf{S}_x is diagonal, then the known signal spectrum bound will be less than or equal to the uncorrelated signal bound in (8.167).

Note that the bound in (8.199) applies to correlated signals also. The expression in (8.199) is also the upper left matrix in (8.89), so a similar statement applies to correlated and coherent signals. A good discussion of the coherent signal case is given Stoica et al. [SOVM96].

⁶The result in (8.199) was published by Weiss and Friedlander [WF91a] and derived in [WF90]. It is also derived in Appendix A of [BCD+84]. [Wax85] derived the CRB for essentially this problem, but the form of the result was less compact.

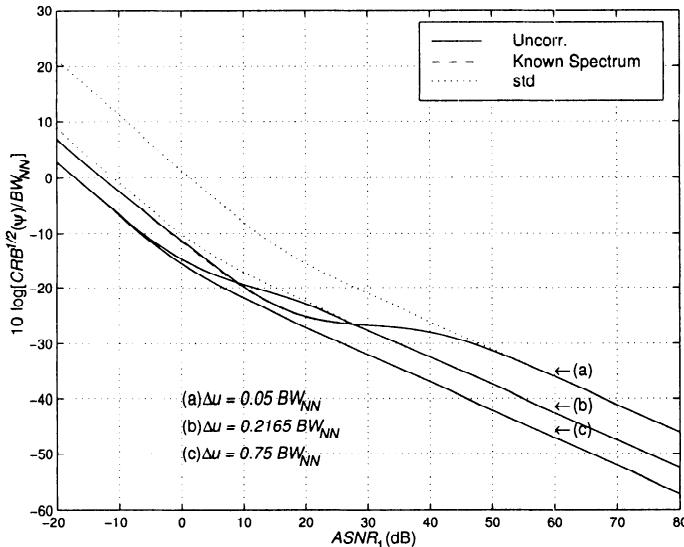


Figure 8.12 Normalized CRBs versus $ASNR_1$: $ASNR_2 = ASNR_1$.

We consider a simple example to show how knowledge of the spectrum affects the CRB.

Example 8.4.7 (continuation)

We consider a standard 10-element linear array. We first consider the case of two uncorrelated signals. We consider the same parameter set as in Example 8.4.6. The results are shown in Figures 8.12, 8.13, and 8.14. In each figure, we plot the normalized CRB using (8.106), (8.167), and (8.199) versus $ASNR$.

We see that the uncorrelated signal bound and the known spectrum bound appear to be the same. There is actually a very small difference. For $\Delta u = 0.2165BW_{NN}$ and $ASNR_1 = 0$ dB, the difference is 0.02 dB.

Example 8.4.8(continuation)

Consider the same model as in Example 8.4.7 except the two signals are coherent, with $\rho = 1$. Note that the phases of ρ is 0° , so forward-backward averaging cannot decorrelate the signals. In Figures 8.15 and 8.16, we plot the normalized CRB for the unknown signal spectral matrix case from (8.106) and for the known spectral matrix case from (8.199).

For a closely spaced signals there is a significant difference in the two bounds. For wider spacings, they coincide.

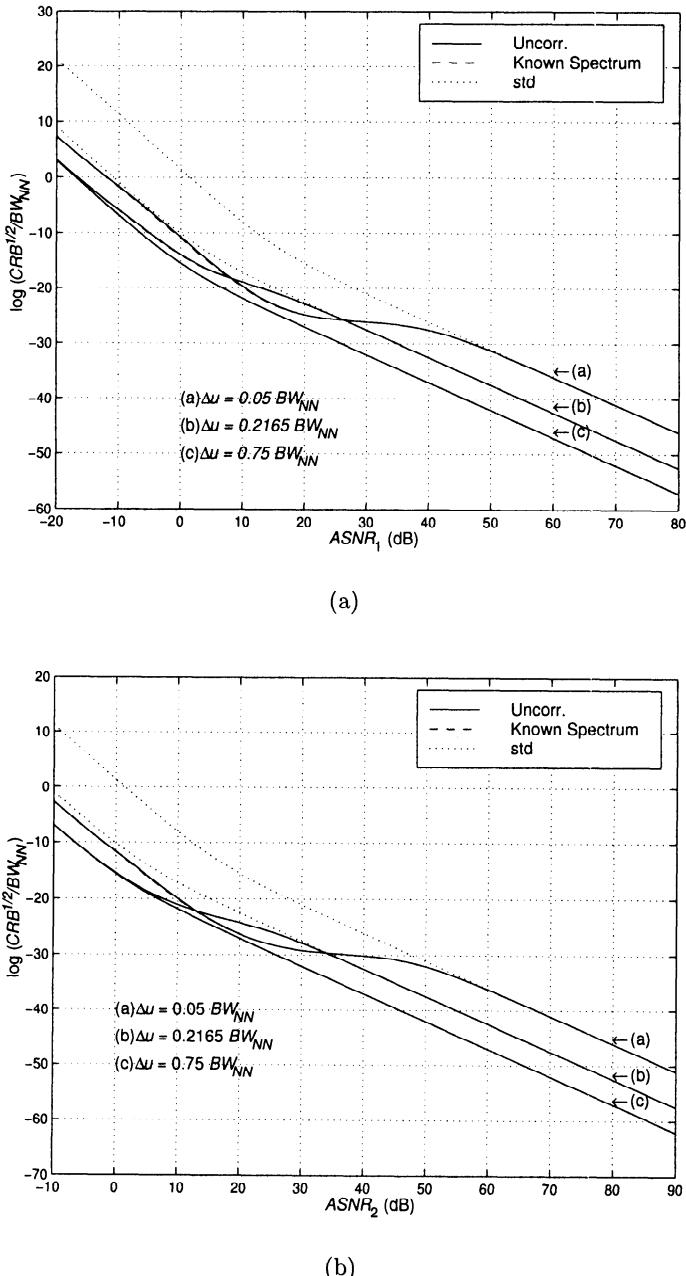
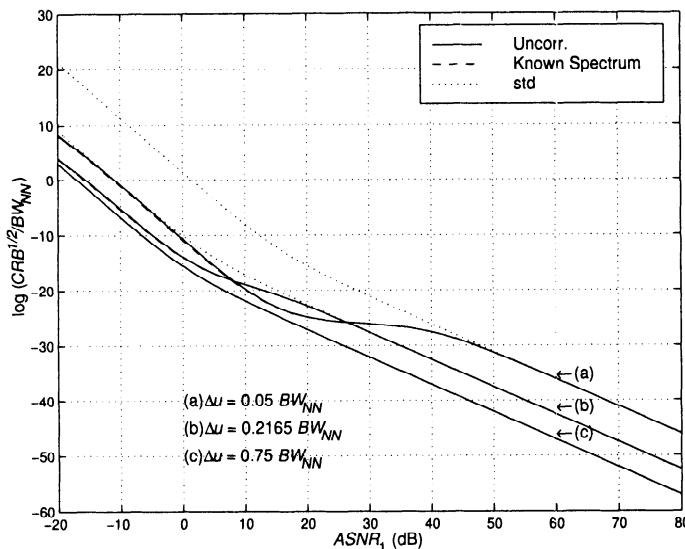
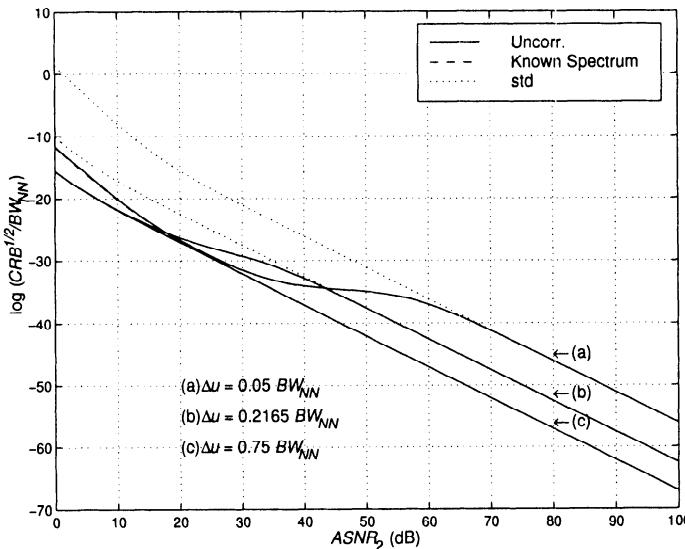


Figure 8.13 Normalized CRBs versus $ASNR$: $ASNR_2 = ASNR_1 + 10$ dB: (a) CRBs for signal 1 versus $ASNR_1$; (b) CRBs for signal 2 versus $ASNR_2$.



(a)



(b)

Figure 8.14 Normalized CRBs versus $ASNR$: $ASNR_2 = ASNR_1 + 20$ dB: (a) CRBs for signal 1 versus $ASNR_1$; (b) CRBs for signal 2 versus $ASNR_2$.

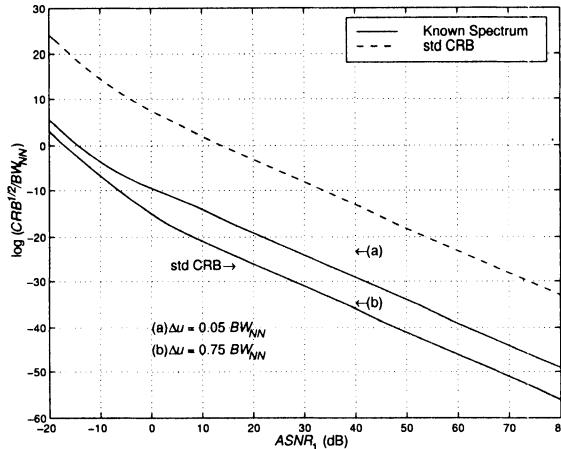


Figure 8.15 Normalized CRBs versus $ASNR_1$: two signals, $N = 10$, $K = 100$, $ASNR_2 = ASNR_1$, $\rho = 1$.

8.4.4 Nonrandom (Conditional) Signal Model

The nonrandom signal model was introduced in Section 8.3.1. For a narrow-band model, the snapshot model is (from (8.67))

$$\mathbf{X}(k) = \mathbf{V}(\psi) \mathbf{F}(k) + \mathbf{N}(k), \quad k = 1, 2, \dots, K. \quad (8.200)$$

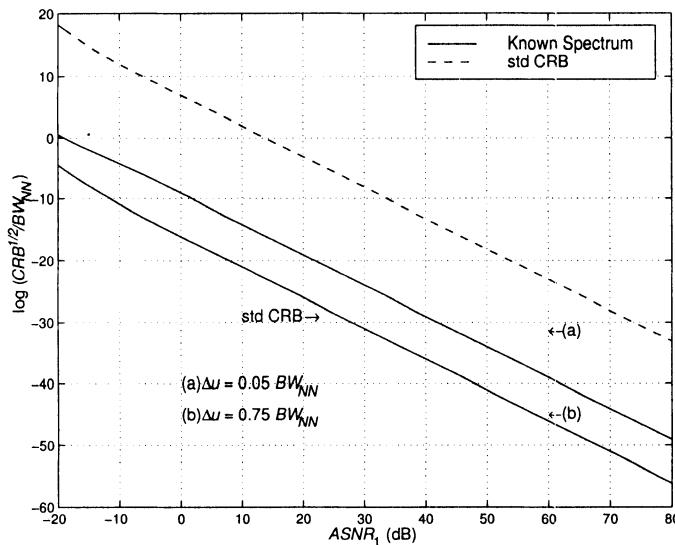
We assume that $\mathbf{F}(k)$ is an unknown nonrandom complex source-signal vector ($D \times 1$). Alternatively, we can view $\mathbf{F}(k)$ as a specific sample function of a complex random process, and we derive an estimator for this sample function and bound its performance. This second viewpoint leads to the description as a “conditional” (conditioned on $\mathbf{F}(k)$, $k = 1, 2, \dots, K$) estimator and a conditional Cramér-Rao bound. The two viewpoints lead to identical results.

The nonrandom (conditional) signal model was described in Section 8.3. To evaluate the CRB we utilize (8.35). The unknown parameter vector is

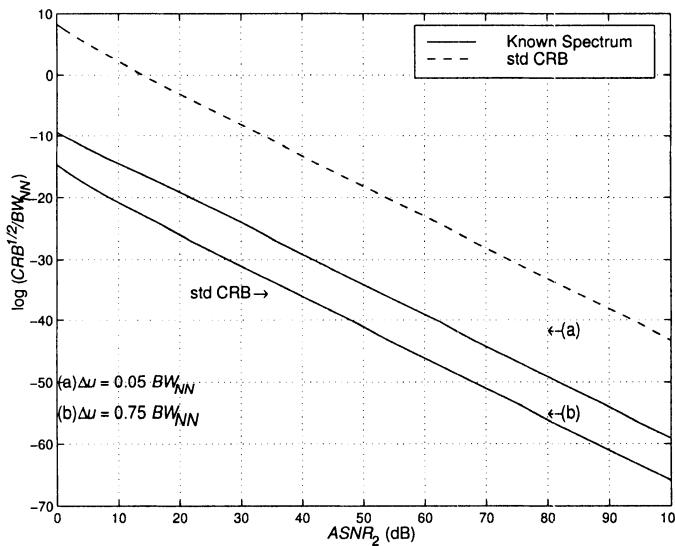
$$\boldsymbol{\theta} \triangleq [\psi, \quad \mathbf{F}, \quad \sigma_w^2]. \quad (8.201)$$

The vector ψ and the scalar σ_w^2 have been defined previously (8.87). \mathbf{F} is a real vector that contains the signal values at snapshot times $k = 1, 2, \dots, K$. Thus,

$$\mathbf{F} \triangleq [Re[\mathbf{F}(1)]^T \quad Im[\mathbf{F}(1)]^T \quad Re[\mathbf{F}(2)]^T \quad \cdots \quad Im[\mathbf{F}(K)]^T]^T. \quad (8.202)$$



(a)



(b)

Figure 8.16 Normalized CRBs versus $ASNR_1$: $ASNR_2 = ASNR_1 + 20$ dB: (a) CRBs for signal 1 versus $ASNR_1$; (b) CRBs for signal 2 versus $ASNR_2$.

It is a $2DK$ dimension vector.

The Fisher information matrix has the same structure as in (8.89).

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{\psi\psi} & \mathbf{J}_{\psi\mathbf{F}} & \mathbf{J}_{\psi\sigma_w^2} \\ \mathbf{J}_{\mathbf{F}\psi} & \mathbf{J}_{\mathbf{FF}} & \mathbf{J}_{\mathbf{F}\sigma_w^2} \\ \mathbf{J}_{\sigma_w^2\psi} & \mathbf{J}_{\sigma_w^2\psi} & \mathbf{J}_{\sigma_w^2\sigma_w^2} \end{bmatrix}. \quad (8.203)$$

The three principal submatrices are of dimension $D \times D$, $2KD \times 2KD$, and 1×1 , respectively.

To evaluate the terms, we use (8.34) with

$$\mathbf{K}_x(\boldsymbol{\theta}) = \sigma_w^2 \mathbf{I}, \quad (8.204)$$

and

$$\mathbf{m}(\boldsymbol{\theta}, k) = \mathbf{V}(\psi)\mathbf{F}(k). \quad (8.205)$$

We go through the steps in the derivation of the bound because it is representative of the technique for deriving a multiple-parameter CRB. Our derivation uses (8.204) and (8.205) and borrows techniques from Appendices C and G of Stoica and Nehorai [SN89a].

Derivation of CCRB⁷

The elements in the upper left sub-matrix are

$$\begin{aligned} [\mathbf{J}_{\psi\psi}]_{ij} &= \frac{2}{\sigma_w^2} \operatorname{Re} \sum_{k=1}^K \left[\mathbf{F}^H(k) \frac{\partial \mathbf{V}^H(\psi)}{\partial \psi_i} \frac{\partial \mathbf{V}(\psi)}{\partial \psi_j} \mathbf{F}(k) \right] \\ &= \frac{2}{\sigma_w^2} \operatorname{Re} \sum_{k=1}^K \left[F_i^*(k) \frac{\partial \mathbf{v}^H(\psi_i)}{\partial \psi_i} \frac{\partial \mathbf{v}(\psi_j)}{\partial \psi_j} F_j(k) \right] \\ &= \frac{2}{\sigma_w^2} \operatorname{Re} \sum_{k=1}^K \left[F_i^*(k) \mathbf{d}^H(\psi_i) \mathbf{d}(\psi_j) F_j(k) \right]. \end{aligned} \quad (8.206)$$

Defining

$$\mathbf{D} = \begin{bmatrix} \mathbf{d}(\psi_1) & \mathbf{d}(\psi_2) & \cdots & \mathbf{d}(\psi_D) \end{bmatrix}, \quad (8.207)$$

and

$$\mathbf{F}(k) = \operatorname{diag} \{F_1(k), F_2(k), \dots, F_D(k)\}, \quad (8.208)$$

we can write $\mathbf{J}_{\psi\psi}$ as

$$\mathbf{J}_{\psi\psi} = \frac{2}{\sigma_w^2} \operatorname{Re} \sum_{k=1}^K \mathbf{F}^H(k) \mathbf{D}^H \mathbf{D} \mathbf{F}(k). \quad (8.209)$$

⁷A simple derivation of the CCRB is given in Stoica and Larsson [SL01]. It was motivated by the results in Gu [Gu00]. It reparameterizes the original model to obtain a block-diagonal Fisher information matrix.

The second principal submatrix $\mathbf{J}_{\mathbf{FF}}(k)$ is $2KD \times 2KD$ and is block diagonal. Each block is $2D \times 2D$ and has an identical structure. The k th block corresponds to the k th snapshot. It has a structure

$$[\mathbf{J}_{\mathbf{FF}}(k)] = \frac{2}{\sigma_w^2} \begin{bmatrix} \mathbf{A}^R & -\mathbf{A}^I \\ \mathbf{A}^I & \mathbf{A}^R \end{bmatrix}, \quad (8.210)$$

where the superscripts “ R ” and “ I ” denote the real and imaginary parts, respectively, of \mathbf{A} . We now derive \mathbf{A}^R and \mathbf{A}^I :

$$\begin{aligned} \frac{2}{\sigma_w^2} [\mathbf{A}^R]_{ij} &= [[\mathbf{J}_{\mathbf{FF}}(k)]]_{ij} = \frac{2}{\sigma_w^2} \sum_{k=1}^K \operatorname{Re} \frac{\partial \mathbf{m}^H}{\partial F_i^R(k)} \frac{\partial \mathbf{m}}{\partial F_j^R(k)} \\ &= \frac{2}{\sigma_w^2} \operatorname{Re} \{ \mathbf{v}^H(\psi_i) \mathbf{v}(\psi_j) \}, \quad i, j = 1, 2, \dots, D \end{aligned} \quad (8.211)$$

Therefore, we can define

$$\mathbf{A} = [\mathbf{V}^H \mathbf{V}], \quad (8.212)$$

and

$$\mathbf{A}^R = \operatorname{Re} [\mathbf{V}^H \mathbf{V}]. \quad (8.213)$$

To find the matrix in the upper right corner we use

$$\begin{aligned} [[\mathbf{J}_{\mathbf{FF}}(k)]]_{ij} &= \frac{2}{\sigma_w^2} \sum_{k=1}^K \operatorname{Re} \frac{\partial \mathbf{m}^H}{\partial F_i^R(k)} \frac{\partial \mathbf{m}}{\partial F_j^I(k)} \\ &= \frac{2}{\sigma_w^2} \operatorname{Re} \{ j \mathbf{v}^H(\psi_i) \mathbf{v}(\psi_j) \}, \quad i = 1, 2, \dots, D; \\ &\quad j = D+1, \dots, 2D \\ &= -\frac{2}{\sigma_w^2} \operatorname{Im} \{ \mathbf{v}^H(\psi_i) \mathbf{v}(\psi_j) \}, \quad i = 1, 2, \dots, D; \\ &\quad j = D+1, \dots, 2D. \end{aligned} \quad (8.214)$$

Therefore, the matrix in the upper right-hand corner can be written as $-\mathbf{A}^I$. Note that the matrices do not depend on k , so each block is identical. The two other sub-matrices follow in a similar manner.

The third principal sub-matrix is a scalar. Using (8.34)

$$\mathbf{J}_{\sigma_w^2 \sigma_w^2} = \sum_{k=1}^K \operatorname{tr} \left[\frac{1}{\sigma_w^2} \mathbf{I} \cdot \frac{1}{\sigma_w^2} \mathbf{I} \right] = \frac{KN}{\sigma_w^4}. \quad (8.215)$$

We next evaluate the off-diagonal matrices. From (8.34), (8.204), and (8.205), we have

$$\mathbf{J}_{\psi \sigma_w^2} = \mathbf{J}_{\mathbf{F} \sigma_w^2} = \mathbf{0}. \quad (8.216)$$

We partition the $2KD \times D$ matrix, $\mathbf{J}_{\mathbf{F}\psi}$, into K $2D \times D$ matrices:

$$\begin{aligned} [\mathbf{J}_{\mathbf{F}\psi}(k)]_{ij} &= \frac{2}{\sigma_w^2} \sum_{k=1}^K \operatorname{Re} \left\{ \left[\frac{\partial}{\partial F_i^R(k)} \mathbf{F}^H(k) \mathbf{V}^H(\psi) \right] \left[\frac{\partial}{\partial \psi_j} \mathbf{V}(\psi) \mathbf{F}(k) \right] \right\} \\ &= \frac{2}{\sigma_w^2} \operatorname{Re} \left\{ \mathbf{V}^H(\psi) \frac{\partial \mathbf{v}(\psi_j)}{\partial \psi_j} \mathbf{F}(k) \right\}, \quad i = 1, 2, \dots, D; \\ &\quad j = 1, 2, \dots, D. \end{aligned} \quad (8.217)$$

Similarly,

$$\begin{aligned} [\mathbf{J}_{\mathbf{F}\psi}(k)]_{ij} &= \frac{2}{\sigma_w^2} \sum_{k=1}^K \operatorname{Re} \left\{ \left[\frac{\partial}{\partial F_i^T(k)} \mathbf{F}^H(k) \mathbf{V}^H(\psi) \right] \left[\frac{\partial}{\partial \psi_j} \mathbf{V}(\psi) \mathbf{F}(k) \right] \right\} \\ &= \frac{2}{\sigma_w^2} \operatorname{Im} \left\{ \mathbf{V}^H(\psi) \frac{\partial \mathbf{v}(\psi_j)}{\partial \psi_j} \mathbf{F}(k) \right\}, \quad i = D+1, \dots, 2D; \quad j = 1, 2, \dots, D. \end{aligned} \quad (8.218)$$

Thus,

$$[\mathbf{J}_{\mathbf{F}\psi}(k)] = \begin{bmatrix} \Delta^R(k) \\ \Delta^I(k) \end{bmatrix}, \quad (8.219)$$

where

$$\Delta(k) \triangleq \frac{2}{\sigma_w^2} [\mathbf{V}^H(\psi) \mathbf{D}(\psi) \mathbf{F}(k)]. \quad (8.220)$$

Similarly,

$$[\mathbf{J}_{\psi\mathbf{F}}(k)] = [\Delta^{RT}(k) \quad \Delta^{IT}(k)]. \quad (8.221)$$

We also define

$$\Delta^T \triangleq [\Delta^T(1) \mid \Delta^T(2) \mid \dots \mid \Delta^T(K)]. \quad (8.222)$$

The information matrix is,

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{\psi\psi} & \Delta^T(1) & \cdots & \Delta^T(K) & \mathbf{0} \\ \Delta(1) & \mathbf{A}^R & -\mathbf{A}^I & & \mathbf{0} \\ & \mathbf{A}^I & \mathbf{A}^R & & \mathbf{0} \\ & & & \ddots & \\ \Delta(K) & \mathbf{0} & & \mathbf{A}^R & -\mathbf{A}^I \\ & & & \mathbf{A}^I & \mathbf{A}^R \\ \mathbf{0} & \mathbf{0} & & & \frac{KN}{\sigma_w^4} \end{bmatrix}. \quad (8.223)$$

The bound on σ_w^2 is not coupled. Thus,

$$CRB(\sigma_w^2) = \frac{\sigma_w^4}{KN}. \quad (8.224)$$

Using the formula for the inverse of a partitioned matrix on the $(D+2KD) \times (D+2KD)$ matrix in the upper left-hand corner (A.62), we obtain the conditional CRB (CCR) on ψ ,

$$C_{CCR}^{-1}(\psi) = \mathbf{J}_{\psi\psi} - \Delta^T \mathbf{J}_{\mathbf{F}\mathbf{F}}^{-1} \Delta. \quad (8.225)$$

To evaluate the second term, we write

$$\mathbf{J}_{\mathbf{FF}}^{-1} = \begin{bmatrix} \mathbf{B}^R & -\mathbf{B}^I \\ \mathbf{B}^I & \mathbf{B}^R \end{bmatrix}, \quad (8.226)$$

where

$$\mathbf{B} \triangleq \mathbf{A}^{-1}, \quad (8.227)$$

(see (A.68)). Then, one can show by direct substitution that

$$\Delta^T(k)\mathbf{B}\Delta(k) = Re [\Delta^H(k)\mathbf{B}\Delta(k)]. \quad (8.228)$$

Using (8.228) in (8.225) gives

$$C_{CCR}^{-1}(\psi) = \mathbf{J}_\psi \psi - \sum_{k=1}^K Re [\Delta^H(k)\mathbf{B}\Delta(k)]. \quad (8.229)$$

Using (8.220) and (8.209) in (8.229) gives

$$\begin{aligned} C_{CCR}^{-1}(\psi) = & \frac{2}{\sigma_w^2} \sum_{k=1}^K Re \left\{ \mathbf{F}^H(k)\mathbf{D}^H\mathbf{D}\mathbf{F}(k) \right. \\ & \left. - \mathbf{F}^H(k)\mathbf{D}^H\mathbf{V} (\mathbf{V}^H\mathbf{V})^{-1} \mathbf{V}^H\mathbf{D}\mathbf{F}(k) \right\}. \end{aligned} \quad (8.230)$$

This result can be written as

$$\begin{aligned} C_{CCR}^{-1}(\psi) &= \frac{2}{\sigma_w^2} \sum_{k=1}^K Re \left\{ \mathbf{F}^H(k)\mathbf{D}^H [\mathbf{I} - \mathbf{P}_\mathbf{V}] \mathbf{D}\mathbf{F}(k) \right\} \\ &= \frac{2}{\sigma_w^2} \sum_{k=1}^K Re \left\{ \mathbf{F}^H(k)\mathbf{D}^H \mathbf{P}_\mathbf{V}^\perp \mathbf{D}\mathbf{F}(k) \right\}, \end{aligned} \quad (8.231)$$

which is the desired result.

We can rewrite (8.231) more compactly by defining

$$\mathbf{H} \triangleq \mathbf{D}^H \mathbf{P}_\mathbf{V}^\perp \mathbf{D}. \quad (8.232)$$

Then,

$$\begin{aligned} [\mathbf{C}_{CCR}^{-1}(\psi)]_{ij} &= \frac{2}{\sigma_w^2} Re \left\{ \mathbf{d}^H(\psi_i) \mathbf{P}_\mathbf{V}^\perp \mathbf{d}(\psi_j) \sum_{k=1}^K F_i^*(k) F_j(k) \right\} \\ &= \frac{2K}{\sigma_w^2} Re \left\{ \mathbf{H}_{ij} \cdot \frac{1}{K} \sum_{k=1}^K F_j(k) F_i^*(k) \right\} \\ &= \frac{2K}{\sigma_w^2} Re \left\{ \mathbf{H}_{ij} \cdot [\hat{\mathbf{S}}_\mathbf{f}]_{ji} \right\}, \end{aligned} \quad (8.233)$$

where

$$\hat{\mathbf{S}}_{\mathbf{f}} = \frac{1}{K} \sum_{k=1}^K \mathbf{F}(k) \mathbf{F}^H(k). \quad (8.234)$$

The result in (8.233) can be written using the Hadamard product as

$$\mathbf{C}_{CCR}^{-1}(\psi) = \frac{2K}{\sigma_w^2} \operatorname{Re} [\mathbf{H} \odot \hat{\mathbf{S}}_{\mathbf{f}}^T]. \quad (8.235)$$

As we would expect, the bound depends on the actual value of the vector parameter ψ and the actual signal waveform (through $\hat{\mathbf{S}}_{\mathbf{f}}$). However, if we now assume that $\mathbf{x}(k)$ is a sample function from an ergodic random process, then as K goes to infinity, $\hat{\mathbf{S}}_{\mathbf{f}}$ will approach the actual spectral matrix $\mathbf{S}_{\mathbf{f}}$ and we can write the asymptotic conditional Cramér-Rao bound (ACCR) as

$$\boxed{\mathbf{C}_{ACCR}(\psi) = \frac{\sigma_w^2}{2K} \left[\operatorname{Re} [\mathbf{H} \odot \mathbf{S}_{\mathbf{f}}^T] \right]^{-1}.} \quad (8.236)$$

This is the same bound that we encountered in Section 8.4.2.1 as a high *SNR* approximation to the standard CRB which we also refer to as the stochastic or unconditional CRB. We saw several examples of how the two bounds compared in that section. In all of the examples, the $\mathbf{C}_{CR}(\psi)$ was above $\mathbf{C}_{ACCR}(\psi)$ and they converged as the *SNR* increased. In other words, the stochastic CRB was a better (tighter) bound than the asymptotic conditional CRB (ACCR). We now show that

$$\mathbf{C}_{CR}(\psi) \geq \mathbf{C}_{ACCR}(\psi). \quad (8.237)$$

First, consider $\mathbf{C}_{CR}(\psi)$ as given by (8.102). Using the matrix inversion lemma we can write⁸

$$\begin{aligned} \mathbf{S}_{\mathbf{f}} \mathbf{V}^H \mathbf{S}_{\mathbf{x}}^{-1} \mathbf{V} \mathbf{S}_{\mathbf{f}} &= \mathbf{S}_{\mathbf{f}} - \mathbf{S}_{\mathbf{f}} \left[\mathbf{I} - \mathbf{V}^H \left[\mathbf{V} \mathbf{S}_{\mathbf{f}} \mathbf{V}^H + \sigma_w^2 \mathbf{I} \right]^{-1} \mathbf{V} \mathbf{S}_{\mathbf{f}} \right] \\ &= \mathbf{S}_{\mathbf{f}} - \mathbf{S}_{\mathbf{f}} \left[\mathbf{I} + \mathbf{V}^H \sigma_w^{-2} \mathbf{V} \mathbf{S}_{\mathbf{f}} \right]^{-1}. \end{aligned} \quad (8.238)$$

Now the second matrix on the right side of (8.238) is Hermitian and non-negative definite. Thus,

$$\mathbf{S}_{\mathbf{f}} \mathbf{V}^H \mathbf{S}_{\mathbf{x}}^{-1} \mathbf{V} \mathbf{S}_{\mathbf{f}} \leq \mathbf{S}_{\mathbf{f}}. \quad (8.239)$$

⁸This result is from Ottersten et al. [OVK92]. See also Stoica and Nehorai [SN90b]. Several useful order relationships are discussed in Stoica and Sharman [SS90a] and Stoica and Nehorai [SN89a].

Then, using the properties of Hadamard product (see (A.76)–(A.78)),

$$\left[\operatorname{Re} \left\{ \mathbf{H} \odot \left(\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f \right) \right\} \right]^{-1} \geq \left[\operatorname{Re} \left\{ \mathbf{H} \odot \mathbf{S}_f^T \right\} \right]^{-1}, \quad (8.240)$$

so that

$$\mathbf{C}_{CR}(\psi) \geq \mathbf{C}_{ACCR}(\psi). \quad (8.241)$$

If \mathbf{H} and \mathbf{S}_f are both positive definite, then the inequality is strict.

We recall from Examples 8.4.1 and 8.4.2 that even though the inequality is strict, the difference is negligible when the eigenvalues of $\mathbf{V}^H \mathbf{V} \mathbf{S}_f / \sigma_w^2$ become large.

8.4.5 Known Signal Waveforms

In this section we consider the case in which the signal waveforms are known. One application of this model is in communication systems in which a training sequence is sent to enable the receiver to synchronize. The CRB for this case is derived in Li and Compton [LC93].

The samples of the signal from ψ_i are denoted by

$$f_i(k) = \alpha_i p_i(k), \quad k = 1, 2, \dots, K, \quad (8.242)$$

where $p_i(k)$ are the samples from the known signal and α_i is a complex constant. We consider both known and unknown α_i .

The time-domain samples are

$$\begin{aligned} \mathbf{x}(k) &= \mathbf{V}(\psi) \mathbf{f}(k) + \mathbf{n}(k) \\ &= \mathbf{V}(\psi) \mathbf{p}(k) \boldsymbol{\alpha} + \mathbf{n}(k), \quad k = 1, 2, \dots, K, \end{aligned} \quad (8.243)$$

where

$$\mathbf{p}(k) \triangleq \operatorname{diag} \{p_1(k), p_2(k), \dots, p_D(k)\}, \quad (8.244)$$

and

$$\boldsymbol{\alpha} \triangleq \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_D \end{bmatrix}^T. \quad (8.245)$$

The case for known $\boldsymbol{\alpha}$ is a degenerate version of the model in Section 8.4.4. The CRB is given by the inverse of (8.209)

$$CRB(\psi) = \frac{\sigma_w^2}{2} \left\{ \operatorname{Re} \sum_{k=1}^K \mathbf{F}^H(k) \mathbf{D}^H \mathbf{D} \mathbf{F}(k) \right\}^{-1}, \quad (8.246)$$

where

$$\mathbf{F}(k) = \text{diag} \{ \alpha_1 p_1(k), \alpha_2 p_2(k), \dots, \alpha_D p_D(k) \}, \quad (8.247)$$

For the case of unknown $\boldsymbol{\alpha}$, we replace the Fisher information matrix in (8.203) with

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{\psi\psi} & \mathbf{J}_{\psi\alpha_R} & \mathbf{J}_{\psi\alpha_I} \\ \mathbf{J}_{\alpha_R\psi} & \mathbf{J}_{\alpha_R\alpha_R} & \mathbf{J}_{\alpha_R\alpha_I} \\ \mathbf{J}_{\alpha_I\psi} & \mathbf{J}_{\alpha_I\alpha_R} & \mathbf{J}_{\alpha_I\alpha_I} \end{bmatrix}, \quad (8.248)$$

where α_R and α_I are the real and imaginary parts, respectively, of $\boldsymbol{\alpha}$. Note that we omit the σ_w^2 term because there is no cross-coupling. The steps for deriving the new submatrices are analogous to those in (8.210)–(8.214). Carrying out those steps, substituting the results into (8.248), inverting \mathbf{J} , and retaining the matrix in the upper corner, we obtain

$$CRB(\psi) = \frac{\sigma_w^2}{2} \left[\frac{2}{\sigma_w^2} \mathbf{J}_{\psi\psi} - \text{Re} \left[\mathbf{A}^H \mathbf{B}^{-1} \mathbf{A} \right] \right]^{-1}, \quad (8.249)$$

where

$$\mathbf{A} \triangleq \sum_{k=1}^K \mathbf{F}^H(k) \mathbf{V}^H \mathbf{D} \mathbf{F}(k), \quad (8.250)$$

and

$$\mathbf{B} \triangleq \sum_{k=1}^K \mathbf{F}^H(k) \mathbf{V}^H \mathbf{V} \mathbf{F}(k). \quad (8.251)$$

Using the same technique as in (8.233)–(8.235), we can write,

$$\frac{2}{\sigma_w^2} \mathbf{J}_{\psi\psi} = K \text{Re} \left\{ \mathbf{D}^H \mathbf{D} \odot \widehat{\mathbf{S}}_f^T \right\}, \quad (8.252)$$

$$\mathbf{A}_K \triangleq \frac{1}{K} \mathbf{A} = \left\{ \mathbf{V}^H \mathbf{D} \odot \widehat{\mathbf{S}}_f^T \right\}, \quad (8.253)$$

$$\mathbf{B}_K \triangleq \frac{1}{K} \mathbf{B} = \left\{ \mathbf{V}^H \mathbf{V} \odot \widehat{\mathbf{S}}_f^T \right\}, \quad (8.254)$$

where

$$\widehat{\mathbf{S}}_f \triangleq \frac{1}{K} \sum_{k=1}^K \mathbf{f}(k) \mathbf{f}^H(k). \quad (8.255)$$

Using (8.252)–(8.254) in (8.249) gives

$$CRB(\psi) = \frac{\sigma_w^2}{2K} \left[Re \left\{ \mathbf{D}^H \mathbf{D} \odot \widehat{\mathbf{S}}_f^T - \mathbf{A}_K^H \mathbf{B}_K^{-1} \mathbf{A}_K \right\} \right]^{-1}. \quad (8.256)$$

For the special case in which $\widehat{\mathbf{S}}_f$ is diagonal, the $CRB(\psi)$ is diagonal.

The result in (8.256) is for an arbitrary array geometry. If we consider the case of a standard linear array and a diagonal $\widehat{\mathbf{S}}_f$, then the result can be simplified.

For unknown α ,

$$CRB(\psi) = \frac{6\sigma_w^2}{K(N^2 - 1)N} \text{diag} \left\{ P_1^{-1}, P_2^{-1}, \dots, P_D^{-1} \right\}, \quad (8.257)$$

where

$$P_i = \frac{1}{K} \sum_{k=1}^K |f_i(k)|^2. \quad (8.258)$$

is the average power in the i th signal and contains the effect of α_i .

The result in (8.257) can also be written as

$$CRB(\psi) = \frac{6}{K(N^2 - 1)} \text{diag} \left\{ ASNR_1^{-1}, ASNR_2^{-1}, \dots, ASNR_D^{-1} \right\}. \quad (8.259)$$

Comparing (8.259) with (8.115) and (8.131), we see that the CRB for any signal in a known multiple-signal environment is the same as the CRB for a single signal in the unknown signal environment. This is a logical result because we can use the temporal characteristics of the known signals to eliminate all of the signals except the desired signals. For two signals, we can obtain a diagonal $\widehat{\mathbf{S}}_f$ for arbitrary K . However, for $D > 2$, there are residual off-diagonal terms whose magnitude decreases as K increases.

We observe that the CRB is not a function of the angle between the plane waves. We also observe that there is no requirement that D , the number of signals, be less than N , the number of sensors.

8.4.6 Summary

In this section, we have developed CRBs for several signal and noise models that we will study in detail in the remainder of Chapter 8 and Chapter 9. The fundamental result that is the starting point for all of the derivation is

(8.34), which gives the expression for the ij element in the Fisher information matrix,

$$\begin{aligned} J_{ij} &= \text{tr} \left[\mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})}{\partial \theta_j} \right] \\ &\quad + 2\text{Re} \left[\frac{\partial \mathbf{m}^H(\boldsymbol{\theta})}{\partial \theta_i} \mathbf{K}_{\mathbf{x}}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{m}(\boldsymbol{\theta})}{\partial \theta_j} \right]. \end{aligned} \quad (8.260)$$

This result is valid whenever $\mathbf{X}(k)$ is a complex Gaussian vector with mean $\mathbf{m}_{\mathbf{x}}(\boldsymbol{\theta})$ and covariance matrix $\mathbf{K}_{\mathbf{x}}(\boldsymbol{\theta})$. The vector $\boldsymbol{\theta}$ includes all parameters that are unknown (both wanted parameters and unwanted, or nuisance, parameters).

In the cases we have developed in this section, $\mathbf{X}(k)$ is the sensor output. Later we look at cases where we pre-process $\mathbf{X}(k)$ with a linear transformation prior to doing the parameter estimation. In those cases, we can use (8.260) on the output of the transformation to compute the CRB. Ideally, we would like the transformation to generate a sufficient statistic for the desired parameter estimation problem so that the CRB would remain the same. In practice, we often have to use transformations that only increase the CRB slightly.

There are a number of other models where we utilize the CRB:

(i) Planar arrays

In this case, (8.200) can be written as

$$\mathbf{X}(k) = \mathbf{V}(\psi) \mathbf{F}(k) + \mathbf{N}(k), \quad (8.261)$$

where ψ is defined as

$$\psi = \begin{bmatrix} \psi_1 & \psi_2 & \cdots & \psi_D \end{bmatrix}^T. \quad (8.262)$$

Each component vector is a 2×1 real vector. We will use

$$\psi_i = \begin{bmatrix} \psi_{x_i} \\ \psi_{y_i} \end{bmatrix} \quad i = 1, \dots, D. \quad (8.263)$$

in the subsequent discussion but the components could also be (θ_i, ϕ_i) if desired.

Yau and Bresler [YB92] have derived the asymptotic conditional Cramér-Rao bound for the case in which ψ_i is a $M \times 1$ vector. We quote their result for the $M = 2$ case.

We define

$$\mathbf{D}_i = \begin{bmatrix} \frac{\partial \mathbf{v}(\psi_i)}{\partial \psi_{x_i}} & \frac{\partial \mathbf{v}(\psi_i)}{\partial \psi_{y_i}} \end{bmatrix} \quad i = 1, \dots, D, \quad (8.264)$$

and

$$\mathbf{D} = \begin{bmatrix} \mathbf{D}_1 & \mathbf{D}_2 & \cdots & \mathbf{D}_D \end{bmatrix}, \quad (8.265)$$

which is a $N \times 2D$ matrix. As in (8.232), we define

$$\mathbf{H}_2 \triangleq \mathbf{D}^H \mathbf{P}_{\mathbf{V}}^{\perp} \mathbf{D}. \quad (8.266)$$

We use the subscript 2 to denote that ψ_i has two components.

We define a 2×2 matrix of ones as,

$$\mathbf{1}_{2 \times 2} = \begin{bmatrix} 1 & 1 \\ 1 & 1 \end{bmatrix}. \quad (8.267)$$

Then, the ACCR is

$$\begin{aligned} C_{ACCR}(\psi) &= C_{ACCR}(\psi_1, \psi_2, \dots, \psi_D) \\ &= \frac{\sigma_w^2}{2K} \left[\operatorname{Re} \left[\mathbf{H}_2 \odot \left[\mathbf{S}_{\mathbf{f}}^T \otimes \mathbf{1}_{2 \times 2} \right] \right] \right]^{-1}, \end{aligned} \quad (8.268)$$

where \odot is a Hadamard product (A.70) and \otimes is a Kronecker product (A.79). Yau and Bresler [YB92] derived the conditional CRB, but the modification to obtain the asymptotic bound is clear.

We consider some examples in the problems. Note that this result also can be used for multiple polarization signals.

(ii) Broadband signals

We discuss broadband signals in Section 8.5.6 and discuss the appropriate CRB.

(iii) Spatially spread signals

We discuss parametric models for spatially spread signals in Section 8.9 and derive appropriate CRBs.

(iv) Beamspace processing

In Sections 6.9 and 7.10 we discussed beamspace beamformers. We will also use beamspace processing for parameter estimation. The beamspace matrix is \mathbf{B}_{bs}^H and we require

$$\mathbf{B}_{bs}^H \mathbf{B}_{bs} = \mathbf{I}. \quad (8.269)$$

The beamspace steering vector is

$$\mathbf{v}_{bs} = \mathbf{B}_{bs}^H \mathbf{v}. \quad (8.270)$$

We define a projection matrix,

$$\mathbf{P}_{\mathbf{V}_{bs}} = \mathbf{B}_{bs}^H \mathbf{V} [\mathbf{V}^H \mathbf{P}_{\mathbf{B}_{bs}} \mathbf{V}]^{-1} \mathbf{V}^H \mathbf{B}_{bs}. \quad (8.271)$$

The beamspace spectral matrix is

$$\begin{aligned} \mathbf{S}_{\mathbf{x}_{bs}} &= \mathbf{B}_{bs}^H \mathbf{S}_{\mathbf{x}} \mathbf{B}_{bs} \\ &= \mathbf{B}_{bs}^H \mathbf{V} \mathbf{S}_f \mathbf{V}^H \mathbf{B}_{bs} + \sigma_w^2 \mathbf{I}. \end{aligned} \quad (8.272)$$

The beamspace noise is white due to (8.269).

Thus, the CRB for beamspace estimation is

$$\begin{aligned} CRB_{bs}(\psi) &= \frac{\sigma_w^2}{2K} \left[Re \left\{ \left[\mathbf{D}^H \mathbf{B}_{bs} \mathbf{P}_{\mathbf{V}_{bs}}^\perp \mathbf{B}_{bs}^H \mathbf{D} \right] \right. \right. \\ &\quad \left. \left. \odot \left[\mathbf{S}_f \mathbf{V}^H \mathbf{B}_{bs} \mathbf{S}_{\mathbf{x}_{bs}}^{-1} \mathbf{B}_{bs}^H \mathbf{V} \mathbf{S}_f \right]^T \right\} \right]^{-1}. \end{aligned} \quad (8.273)$$

We discuss several examples in the problems and study beamspace estimators in more detail in Sections 8.11 and 9.7.

(v) Range and bearing estimation

In the case of near-field sources, the parameter vector includes both range and bearing. Several references discuss this model and derive CRB (e.g., Rockah and Schultheiss [RS87a] [RS87b] or Huang and Barkat [HB91]).

(vi) Multipath models

In many applications, multipath is an important factor. Several references discuss parameter estimation in radar and sonar systems. Rendas and Moura [RM91] derive the CRB for a model that is appropriate for the sonar environment.

(vii) Minimally redundant arrays

The CRB results applied to arbitrary array geometries, but our examples discussed uniform linear arrays. Another interesting class of arrays is the minimally redundant arrays discussed in Section 3.9. The CRB for this case are discussed in Chambers et al. [CTSD96] and Abramovich et al. [AGGS98]. They also discuss estimation techniques. We develop several examples in the problems.

(viii) Cyclostationary signals

In many communication systems the modulated signals exhibit a cyclostationary (periodic correlation) property that can be exploited to improve the DOA estimation performance. Schell [Sch94] derives the CRB for this model.

(ix) Computation

Computation of the inverse of the Fisher information matrix may be difficult for large parameter sets. Hero et al. [HUSF97] develop a recursive algorithm to compute the bound.

We encounter other examples as we proceed through Chapters 8 and 9. We now consider maximum likelihood estimation procedures. In many cases of interest, the maximum likelihood estimates achieve the Cramér-Rao bound as K , the number of snapshots, goes to infinity.

8.5 Maximum Likelihood Estimation

In this section, we derive the maximum likelihood estimator for the DOAs of D plane-wave signals. In Section 8.5.1, we consider the model in which the source signals are sample functions from a Gaussian random process with an unknown \mathbf{S}_f .

The maximum likelihood estimator that we consider in Section 8.5.1 is sometimes referred to in the literature as the unconditional ML estimate or the stochastic ML estimate (e.g., [SN90b]).

In Section 8.5.2, we consider a model in which we treat the source signals as nonrandom but unknown and derive an ML estimate that is referred to in the literature as the deterministic (or conditional) ML estimate. We compare the performance of the two estimates.

In Section 8.5.3, we discuss the asymptotic performance of the maximum likelihood estimators. In Section 8.5.4, we extend the results to wideband signals. In Section 8.5.5, we summarize our results.

8.5.1 Maximum Likelihood Estimation

In this section, we derive a family of unconditional ML (UML) estimators. The reason that we have several UML estimators rather than a single estimator will become apparent as we proceed.

We use the frequency-domain snapshot model from Section 8.3.1 (8.62). We assume that we have D plane-wave signals arriving at the array from

directions $\psi_1, \psi_2, \dots, \psi_D$. The signals are sample functions from Gaussian random processes whose source spectral matrix is unknown. The signals are corrupted by additive spatially uncorrelated Gaussian noise with spectral height $S_n(\omega)$. Thus,

$$\mathbf{X}(k) = \mathbf{V}(\psi)\mathbf{F}(k) + \mathbf{N}(k), \quad k = 1, 2, \dots, K. \quad (8.274)$$

$$\mathbf{S}_f = E [\mathbf{F}(k)\mathbf{F}^H(k)]. \quad (8.275)$$

We assume that

$$\mathbf{S}_n = \sigma_w^2 \mathbf{I}, \quad (8.276)$$

and that σ_w^2 is known. We consider the unknown σ_w^2 case subsequently. The component vector ψ_i summarizes the parameters from the D signals that we want to estimate. The component vector ψ_i is the 1- or 2-D wavenumber. Alternatively, we could estimate the azimuth and elevation angles (θ_i, ϕ_i) .

The first estimator will be referred to as the asymptotic ML (AML) estimator.⁹

8.5.1.1 AML estimators

The ML estimators for this model have been derived by Böhme [Boh86] and Jaffer [Jaf88]. Our discussion follows the latter reference. To simplify the notation, we indicate the snapshot number by a subscript.

The likelihood function is

$$L(\psi, \mathbf{S}_f) = -\ln \det \mathbf{S}_x - \frac{1}{K} \sum_{k=1}^K \mathbf{X}_k^H \mathbf{S}_x^{-1} \mathbf{X}_k, \quad (8.277)$$

and

$$\mathbf{S}_x = \mathbf{V}(\psi)\mathbf{S}_f\mathbf{V}^H(\psi) + \sigma_w^2 \mathbf{I}, \quad (8.278)$$

where we have dropped unnecessary constants. To avoid confusion between the sample covariance matrix and the ML estimate of \mathbf{S}_x , we use \mathbf{C}_x to designate the sample covariance matrix. Recall that the sample covariance matrix is

$$\mathbf{C}_x = \frac{1}{K} \sum_{k=1}^K \mathbf{X}_k \mathbf{X}_k^H. \quad (8.279)$$

⁹The abbreviation “AML” is sometimes used to denote approximate ML estimate in the literature. We do not use it in that manner.

Using (8.279), we can write (8.277) as

$$\begin{aligned} L(\psi, \mathbf{S}_f) &= - \left[\ln \det \mathbf{S}_x + \text{tr} \left[\frac{1}{K} \sum_{k=1}^K \mathbf{X}_k^H \mathbf{S}_x^{-1} \mathbf{X}_k \right] \right] \\ &= - \left[\ln \det \mathbf{S}_x + \text{tr} \left[\mathbf{S}_x^{-1} \cdot \frac{1}{K} \sum_{k=1}^K \mathbf{X}_k \mathbf{X}_k^H \right] \right] \\ &= - \left[\ln \det \mathbf{S}_x + \text{tr} \left[\mathbf{S}_x^{-1} \mathbf{C}_x \right] \right]. \end{aligned} \quad (8.280)$$

Fortunately, the solution is separable so that we can maximize over \mathbf{S}_x to obtain an explicit function of ψ and then maximize over ψ to get the total solution.

We denote the ij th element of \mathbf{S}_f as S_{ij} . Then \mathbf{S}_x can be written as

$$\mathbf{S}_x = \sum_{i=1}^D \sum_{j=1}^D S_{ij} \mathbf{v}(\psi_i) \mathbf{v}^H(\psi_j) + \sigma_w^2 \mathbf{I}. \quad (8.281)$$

Differentiating the first term in (8.280) using (A.400), we have

$$\frac{\partial \ln |\mathbf{S}_x|}{\partial S_{ij}} = \text{tr} \left[\left[\frac{\partial \ln |\mathbf{S}_x|}{\partial \mathbf{S}_x} \right]^T \frac{\partial \mathbf{S}_x}{\partial S_{ij}} \right], \quad (8.282)$$

and, from (A.397),

$$\frac{\partial \ln |\mathbf{S}_x|}{\partial \mathbf{S}_x} = [\mathbf{S}_x^{-1}]^T. \quad (8.283)$$

From (8.281),

$$\frac{\partial \mathbf{S}_x}{\partial S_{ij}} = \mathbf{v}(\psi_i) \mathbf{v}^H(\psi_j). \quad (8.284)$$

Using (8.283) and (8.284) in (8.282) gives

$$\frac{\partial \ln |\mathbf{S}_x|}{\partial S_{ij}} = \text{tr} \left[\mathbf{S}_x^{-1} \mathbf{v}(\psi_i) \mathbf{v}^H(\psi_j) \right] = \text{tr} \left[\mathbf{v}^H(\psi_j) \mathbf{S}_x^{-1} \mathbf{v}(\psi_i) \right]. \quad (8.285)$$

Differentiating the second term in (8.280) using (A.401) and (A.393), we have

$$\begin{aligned} \frac{\partial \text{tr} [\mathbf{S}_x^{-1} \mathbf{C}_x]}{\partial S_{ij}} &= \text{tr} \left\{ \left[\frac{\partial \text{tr} [\mathbf{S}_x^{-1} \mathbf{C}_x]}{\partial \mathbf{S}_x} \right]^T \frac{\partial \mathbf{S}_x}{\partial S_{ij}} \right\} \\ &= \text{tr} \left\{ -\mathbf{S}_x^{-1} \mathbf{C}_x \mathbf{S}_x^{-1} \mathbf{v}(\psi_i) \mathbf{v}^H(\psi_j) \right\} \\ &= \text{tr} \left\{ -\mathbf{v}^H(\psi_j) \mathbf{S}_x^{-1} \mathbf{C}_x \mathbf{S}_x^{-1} \mathbf{v}(\psi_i) \right\}. \end{aligned} \quad (8.286)$$

A necessary condition is that

$$\frac{\partial L(\psi, \mathbf{S}_f)}{\partial S_{ij}} = 0, \quad i, j = 1, 2, \dots, D. \quad (8.287)$$

Substituting (8.285) and (8.286) into (8.287) and noting that both terms are scalars,

$$\mathbf{v}^H(\psi_j) [\mathbf{S}_x^{-1} \mathbf{C}_x \mathbf{S}_x^{-1} - \mathbf{S}_x^{-1}] \mathbf{v}(\psi_i) = \mathbf{0}, \quad i, j = 1, 2, \dots, D, \quad (8.288)$$

or

$$\mathbf{V}^H(\psi) [\mathbf{S}_x^{-1} \mathbf{C}_x \mathbf{S}_x^{-1} - \mathbf{S}_x^{-1}] \mathbf{V}(\psi) = \mathbf{0}. \quad (8.289)$$

To get (8.289) into a more usable form, we write,

$$\mathbf{S}_x^{-1} = \frac{1}{\sigma_w^2} \left[\mathbf{I} - \mathbf{V} \left[\mathbf{S}_f \mathbf{V}^H \mathbf{V} + \sigma_w^2 \mathbf{I} \right]^{-1} \mathbf{S}_f \mathbf{V}^H \right], \quad (8.290)$$

where we have suppressed the ψ dependence of $\mathbf{V}(\psi)$.

$$\begin{aligned} \mathbf{S}_x^{-1} \mathbf{V} &= \frac{\mathbf{V}}{\sigma_w^2} \left[\mathbf{I} - \left[\mathbf{S}_f \mathbf{V}^H \mathbf{V} + \sigma_w^2 \mathbf{I} \right]^{-1} \mathbf{S}_f \mathbf{V}^H \mathbf{V} \right] \\ &= \frac{\mathbf{V}}{\sigma_w^2} \left[\left[\mathbf{S}_f \mathbf{V}^H \mathbf{V} + \sigma_w^2 \mathbf{I} \right]^{-1} \left(\left[\mathbf{S}_f \mathbf{V}^H \mathbf{V} + \sigma_w^2 \mathbf{I} \right] - \mathbf{S}_f \mathbf{V}^H \mathbf{V} \right) \right] \\ &= \mathbf{V} \left[\mathbf{S}_f \mathbf{V}^H \mathbf{V} + \sigma_w^2 \mathbf{I} \right]^{-1} = \mathbf{V} \mathbf{S}_x^{-1}. \end{aligned} \quad (8.291)$$

Using (8.291) and its conjugate transpose in (8.289) gives

$$\left[\mathbf{S}_f \mathbf{V}^H \mathbf{V} + \sigma_w^2 \mathbf{I} \right]^{-1} \mathbf{V}^H [\mathbf{C}_x - \mathbf{S}_x] \mathbf{V} \left[\mathbf{S}_f \mathbf{V}^H \mathbf{V} + \sigma_w^2 \mathbf{I} \right]^{-1} = \mathbf{0}. \quad (8.292)$$

Note that one cannot let $\widehat{\mathbf{S}}_{x,ml} = \mathbf{C}_x$ in order to solve (8.292) because $\widehat{\mathbf{S}}_{x,ml}$ must have the structure in (8.281) and, with probability 1, \mathbf{C}_x will not have that structure.

The condition in (8.292) implies

$$[\mathbf{V}^H [\mathbf{C}_x - \mathbf{S}_x] \mathbf{V}]_{\mathbf{S}_x = \widehat{\mathbf{S}}_{x,ml}} = \mathbf{0}. \quad (8.293)$$

Substituting

$$\mathbf{S}_x = \mathbf{V} \mathbf{S}_f \mathbf{V}^H + \sigma_w^2 \mathbf{I}, \quad (8.294)$$

into (8.293) gives

$$\mathbf{V}^H \mathbf{C}_x \mathbf{V} = \mathbf{V}^H \mathbf{V} \mathbf{S}_f \mathbf{V}^H \mathbf{V} + \sigma_w^2 \mathbf{V}^H \mathbf{V}. \quad (8.295)$$

The solution to (8.295) is denoted by $\widehat{\mathbf{S}}_{\mathbf{f},ml}(\psi)$:

$$\boxed{\widehat{\mathbf{S}}_{\mathbf{f},ml}(\psi) = [\mathbf{V}^H \mathbf{V}]^{-1} \mathbf{V}^H [\mathbf{C}_x - \sigma_w^2 \mathbf{I}] \mathbf{V} [\mathbf{V}^H \mathbf{V}]^{-1}.} \quad (8.296)$$

The result in (8.296) can also be written as

$$\widehat{\mathbf{S}}_{\mathbf{f},ml}(\psi) = \mathbf{V}^\dagger [\mathbf{C}_x - \sigma_w^2 \mathbf{I}] [\mathbf{V}^\dagger]^H, \quad (8.297)$$

where

$$\mathbf{V}^\dagger = [\mathbf{V}^H \mathbf{V}^{-1}] \mathbf{V}^H, \quad (8.298)$$

is the Moore-Penrose pseudoinverse. This result was previously given in [Sch79] and [Jaf85].

The result in (8.296) does not guarantee that $\widehat{\mathbf{S}}_{\mathbf{f},ml}(\psi)$ is non-negative definite because the maximization with respect to $\mathbf{S}_\mathbf{f}$ was over the set of Hermitian matrices and not over the set of non-negative definite matrices. We discuss the implication of this result after we complete the derivation.

We define

$$\widehat{\mathbf{S}}_{\mathbf{x},ml}(\psi) = \mathbf{V}(\psi) \widehat{\mathbf{S}}_{\mathbf{f},ml}(\psi) \mathbf{V}^H(\psi) + \sigma_w^2 \mathbf{I}, \quad (8.299)$$

where $\widehat{\mathbf{S}}_{\mathbf{f},ml}(\psi)$ is given by (8.296).

The relation in (8.299) can also be written as

$$\widehat{\mathbf{S}}_{\mathbf{x},ml}(\psi) = \mathbf{P}_{\mathbf{V}} [\mathbf{C}_x - \sigma_w^2 \mathbf{I}] \mathbf{P}_{\mathbf{V}} + \sigma_w^2 \mathbf{I}, \quad (8.300)$$

where $\mathbf{P}_{\mathbf{V}}$ is the projection matrix onto the range of $\mathbf{V}(\psi)$.

To find $\hat{\psi}_{ml}$, we maximize (8.280) with $\widehat{\mathbf{S}}_{\mathbf{x},ml}(\psi)$, as given by (8.300), substituted for $\mathbf{S}_\mathbf{x}$. Thus,

$$\hat{\psi}_{ml} = \arg \max_{\psi} \left\{ - \left[\ln \det \widehat{\mathbf{S}}_{\mathbf{x}}(\psi) + \text{tr} [\widehat{\mathbf{S}}_{\mathbf{x}}^{-1}(\psi) \mathbf{C}_x] \right] \right\}, \quad (8.301)$$

and

$$\widehat{\mathbf{S}}_{\mathbf{f}} \triangleq \widehat{\mathbf{S}}_{\mathbf{f},ml} = \widehat{\mathbf{S}}_{\mathbf{f}}(\hat{\psi}_{ml}), \quad (8.302)$$

where we drop the “*ml*” subscript on $\widehat{\mathbf{S}}_{\mathbf{f}}$ for simplicity.

The last step is to get the right side of (8.301) into a more usable form. Using (8.290),

$$\widehat{\mathbf{S}}_{\mathbf{x}}^{-1}(\psi) \mathbf{C}_x = \frac{1}{\sigma_w^2} \mathbf{C}_x - \frac{1}{\sigma_w^2} \mathbf{V} [\widehat{\mathbf{S}}_{\mathbf{f}} \mathbf{V}^H \mathbf{V} + \sigma_w^2 \mathbf{I}]^{-1} \widehat{\mathbf{S}}_{\mathbf{f}} \mathbf{V}^H \mathbf{C}_x. \quad (8.303)$$

From (8.297),

$$\begin{aligned}\hat{\mathbf{S}}_f \mathbf{V}^H \mathbf{V} + \sigma_w^2 \mathbf{I} &= [\mathbf{V}^H \mathbf{V}]^{-1} \mathbf{V}^H [\mathbf{C}_x - \sigma_w^2 \mathbf{I}] \mathbf{V} [\mathbf{V}^H \mathbf{V}]^{-1} \mathbf{V}^H \mathbf{V} + \sigma_w^2 \mathbf{I} \\ &= [\mathbf{V}^H \mathbf{V}]^{-1} \mathbf{V}^H \mathbf{C}_x \mathbf{V},\end{aligned}\quad (8.304)$$

so

$$[\hat{\mathbf{S}}_f \mathbf{V}^H \mathbf{V} + \sigma_w^2 \mathbf{I}]^{-1} = [\mathbf{V}^H \mathbf{C}_x \mathbf{V}]^{-1} \mathbf{V}^H \mathbf{V}. \quad (8.305)$$

Assuming \mathbf{C}_x is positive definite, the inverse exists.¹⁰

Then, from (8.303),

$$\begin{aligned}\text{tr} [\hat{\mathbf{S}}_x^{-1}(\psi) \mathbf{C}_x] &= \frac{\text{tr} [\mathbf{C}_x]}{\sigma_w^2} - \frac{1}{\sigma_w^2} \text{tr} [\mathbf{V} [\mathbf{V}^H \mathbf{C}_x \mathbf{V}]^{-1} \mathbf{V}^H \mathbf{V} \hat{\mathbf{S}}_f \mathbf{V}^H \mathbf{C}_x] \\ &= \frac{\text{tr} [\mathbf{C}_x]}{\sigma_w^2} - \frac{1}{\sigma_w^2} \text{tr} [\hat{\mathbf{S}}_f \mathbf{V}^H \mathbf{V}].\end{aligned}\quad (8.306)$$

Using (8.304) we have

$$\begin{aligned}\text{tr} [\hat{\mathbf{S}}_f \mathbf{V}^H \mathbf{V}] &= \text{tr} [[\mathbf{V}^H \mathbf{V}]^{-1} \mathbf{V}^H \mathbf{C}_x \mathbf{V} - \sigma_w^2 \mathbf{I}] \\ &= \text{tr} [\mathbf{P}_V \mathbf{C}_x] - D \sigma_w^2.\end{aligned}\quad (8.307)$$

Using (8.307) in (8.306) gives,

$$\begin{aligned}\text{tr} [\hat{\mathbf{S}}_x^{-1}(\psi) \mathbf{C}_x] &= \frac{1}{\sigma_w^2} \{ \text{tr} [(\mathbf{I} - \mathbf{P}_V) \mathbf{C}_x] \} - D \\ &= \frac{1}{\sigma_w^2} \{ \text{tr} [\mathbf{P}_V^\perp \mathbf{C}_x] \} - D.\end{aligned}\quad (8.308)$$

From (8.300),

$$\begin{aligned}\hat{\mathbf{S}}_x(\psi) &= \mathbf{P}_V \mathbf{C}_x \mathbf{P}_V + \sigma_w^2 [\mathbf{I} - \mathbf{P}_V] \\ &= \mathbf{P}_V \mathbf{C}_x \mathbf{P}_V + \sigma_w^2 \mathbf{P}_V^\perp.\end{aligned}\quad (8.309)$$

Using (8.308) and (8.309) in (8.301) and dropping terms that do not depend on ψ gives

$$\begin{aligned}\hat{\psi}_{aml} &= \arg \max_{\psi} \left\{ -\ln \det [\mathbf{P}_V \mathbf{C}_x \mathbf{P}_V + \sigma_w^2 \mathbf{P}_V^\perp] \right. \\ &\quad \left. - \frac{1}{\sigma_w^2} \text{tr} [\mathbf{P}_V^\perp \mathbf{C}_x] \right\}.\end{aligned}\quad (8.310)$$

¹⁰ \mathbf{C}_x is positive definite if $K \geq N$.

The estimator defined by (8.310) is referred to in the literature as the stochastic or unconditional ML estimate. We use the subscript “*aml*” for asymptotic maximum likelihood. A brief discussion of the reason for this description is useful.¹¹ If \mathbf{S}_f is strictly positive definite, then, since the ML estimates are consistent, $\hat{\mathbf{S}}_{f,ml}(\hat{\psi})$ tends to \mathbf{S}_f as $K \rightarrow \infty$. Therefore $\hat{\mathbf{S}}_{f,ml}(\hat{\psi})$ must be positive definite and therefore is a valid ML estimate. Hence, (8.310) provides a large-sample realization of the ML estimator in the case of a non-singular \mathbf{S}_f (non-coherent sources). The case of coherent sources is discussed in detail in Stoica et al. [SOVM96], and the reader is referred to that reference.¹² The key result (8.310) still provides a large-sample realization of $\hat{\psi}_{ml}$.

In order to find $\hat{\psi}_{aml}$, we need to perform a maximization over a mD -dimensional space where m is the number of parameters to be estimated in each plane wave (normally one or two). In Section 8.7, we discuss various implementation techniques to perform this maximization. All of these techniques require a significant amount of computation.

Before considering some examples, we indicate the effect of unknown noise variance on the ML estimate.

If the noise variance σ_w^2 is unknown, then the likelihood function in (8.280) becomes,

$$L(\psi, \mathbf{S}_f, \sigma_w^2) = -\left[\ln \det \mathbf{S}_x + \text{tr}\left[\mathbf{S}_x^{-1} \mathbf{C}_x\right]\right], \quad (8.311)$$

where

$$\mathbf{S}_x = \mathbf{V} \mathbf{S}_f \mathbf{V}^H + \sigma_w^2 \mathbf{I}. \quad (8.312)$$

Proceeding in exactly the same manner as above, we find

$$\hat{\sigma}_w^2 = \frac{\text{tr}\left[\mathbf{P}_V^\perp \mathbf{C}_x\right]}{N - D}. \quad (8.313)$$

Substituting (8.313) into (8.310), we obtain

$$\hat{\psi}_{aml} = \arg \max_{\psi} \left\{ -\ln \det \left[\mathbf{P}_V \mathbf{C}_x \mathbf{P}_V + \frac{\text{tr}\left[\mathbf{P}_V^\perp \mathbf{C}_x\right] \mathbf{P}_V^\perp}{N - D} \right] \right\}, \quad (8.314)$$

because the second term in (8.310) is no longer a function of ψ . We can

¹¹This discussion follows [SOVM96].

¹²M. Viberg directed me to this reference (private communication).

maximize the term in braces by minimizing the determinant. Thus,

$$\hat{\psi}_{aml} = \arg \min_{\psi} \left\{ \det \left[\mathbf{P}_V \mathbf{C}_x \mathbf{P}_V + \frac{\text{tr}[\mathbf{P}_V^\perp \mathbf{C}_x] \mathbf{P}_V^\perp}{N-D} \right] \right\}. \quad (8.315)$$

for the case of unknown noise variance.

When one carries out the derivation leading to (8.314), there is an intermediate step containing

$$\widehat{\mathbf{S}}_{f,ml}(\psi) = [\mathbf{V}^H \mathbf{V}]^{-1} \mathbf{V}^H [\mathbf{C}_x - \hat{\sigma}_w^2 \mathbf{I}] \mathbf{V} [\mathbf{V}^H \mathbf{V}]^{-1}, \quad (8.316)$$

with $\hat{\sigma}_w^2$ given by (8.313). The resulting $\widehat{\mathbf{S}}_{f,ml}(\psi)$ is not necessarily positive definite.

We consider a simple example to demonstrate the behavior of the ML estimator.

Example 8.5.1

We consider a standard 10-element linear array. The signal is a single plane wave arriving from $u_s = 0$. We assume the noise spectrum height σ_w^2 is known. We find \hat{u}_{ml} by calculating the term in brackets in (8.310) over a dense grid of points ($-1 \leq u \leq 1$). We then use a local minimization routine to find the exact minimum.

In Figure 8.17, we plot the normalized root mean-square error (RMSE) versus *ASNR*. In Figure 8.18, we plot the normalized RMSE versus *K* for several *ASNR*. In both figures we also plot the CRB.

We see that for $ASNR \geq -2$ dB, the RMSE of AML estimator coincides with the CRB. However, as the *ASNR* decreases, the RMSE increases sharply. This behavior is referred as the threshold phenomenon and is a characteristic of most nonlinear parameter estimation problems (e.g., Section 2.4 of DEMT I [VT68], [VT01a] or Wozencraft and Jacobs [WJ65]). In order to understand the behavior, we show a scatter plot of the estimates for various *ASNR* in Figure 8.19 and the corresponding histogram in Figure 8.20. We see that, above threshold, the estimates are clustered around the correct value.

At $ASNR = -3$ dB, most of the estimates (93/100) are clustered around the correct value, but the others correspond to a subsidiary peak that is not close to the correct value. These errors are sometimes referred to as anomalous (or global) errors. Note that even a few anomalous errors cause a dramatic increase in the RMSE, because they are so large compared to the local errors (the errors clustered around the correct value).

As the *ASNR* decreases further, the number of local errors decreases. At an $ASNR = -12$ dB, the histogram shows that the errors are spread across *u*-space in an almost uniform manner.

It is convenient to write total MSE as

$$MSE = p_{lo}(MSE_{lo}) + (1 - p_{lo})(MSE_{gl}), \quad (8.317)$$

where p_{lo} denotes the probability that the estimation error is local. The local MSE, “ MSE_{lo} ” is usually characterized by the CRB. It is usually difficult to

calculate p_{lo} except in the asymptotic (large K) region. We should observe that, in this example, the threshold occurs just as p_{lo} decreases from unity. We find that this characteristic is true in most cases. Thus, the region of primary interest is usually

$$0.95 \leq p_{lo} \leq 1.0. \quad (8.318)$$

The global MSE, “ MSE_{gl} ” is usually hard to calculate as we enter the threshold region. However, further below threshold, it generally approaches the variance of a uniform random variable (one-third in u -space for $u_s = 0$). If we have *a priori* information about the portion of u -space where the signal can originate, the MSE below threshold may be lower.

We will discuss performance issues further after we consider the multiple plane-wave problem.

Example 8.5.2

Consider a standard 10-element linear array with two equal-power plane-wave signals impinging on it. The signal separation is $\Delta\psi_R = 0.2165BW_{NN}$. The signals are uncorrelated. In Figure 8.21, we plot the normalized RMSE versus $ASNR$ for $K = 100$ snapshots. We see that for $ASNR \geq 5$ dB, the RMSE is equal to the CRB. At an $ASNR = 5$ dB, a threshold occurs and the RMSE increases rapidly.

In order to understand this behavior we show a sequence of scatter plots for various $ASNR$. In Figure 8.22, the $ASNR = 6$ dB and the errors are local. The likelihood function has a single peak that is close to the correct value. Note that we always assign $\hat{u}_2 \geq \hat{u}_1$ so that only the upper left triangle is needed.

In Figure 8.23, the $ASNR = 0$ dB. We see that the majority of the errors are local. However, there are a number of points in which \hat{u}_1 is at the midpoint of the two signal locations,

$$\hat{u}_1 = \frac{u_1 + u_2}{2} = 0, \quad (8.319)$$

and the second estimate \hat{u}_2 is scattered from $u = 0$ to $u = 1$. There is similar behavior for \hat{u}_2 .

When this result occurs we say that the ML estimator cannot resolve the two signals. In order to compute the probability of resolution, we say that the signals are resolved if

$$|\hat{u}_1 - u_1| \leq \min\left(\frac{u_2 - u_1}{2}, \frac{2}{N}\right), \quad (8.320)$$

and

$$|\hat{u}_2 - u_2| \leq \min\left(\frac{u_2 - u_1}{2}, \frac{2}{N}\right). \quad (8.321)$$

Each of the conditions in (8.320) and (8.321) correspond to a local error for the respective parameter. The probability of resolution, P_R , is the probability that both (8.320) and (8.321) are satisfied. Then, we can write the total MSE as

$$MSE = P_R(MSE_{lo}) + (1 - P_R)(MSE_{gl}). \quad (8.322)$$

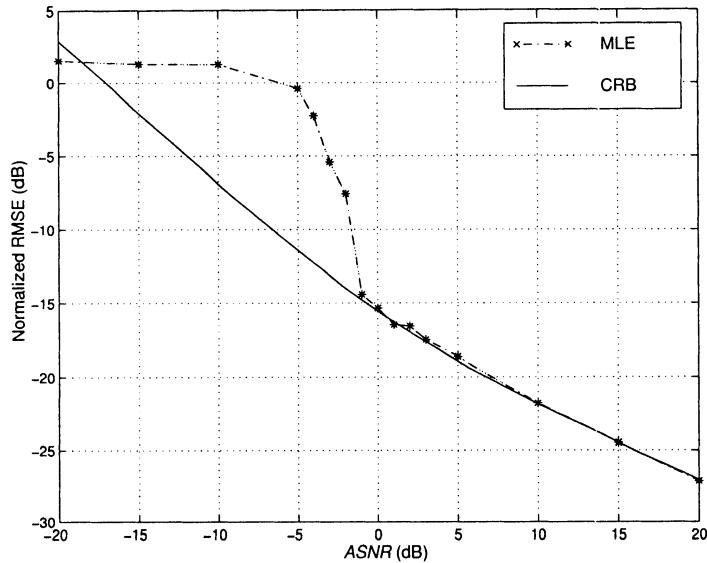


Figure 8.17 AML estimator, single signal, $u_s = 0$, $N = 10$, $K = 100$: normalized RMSE versus $ASNR$.

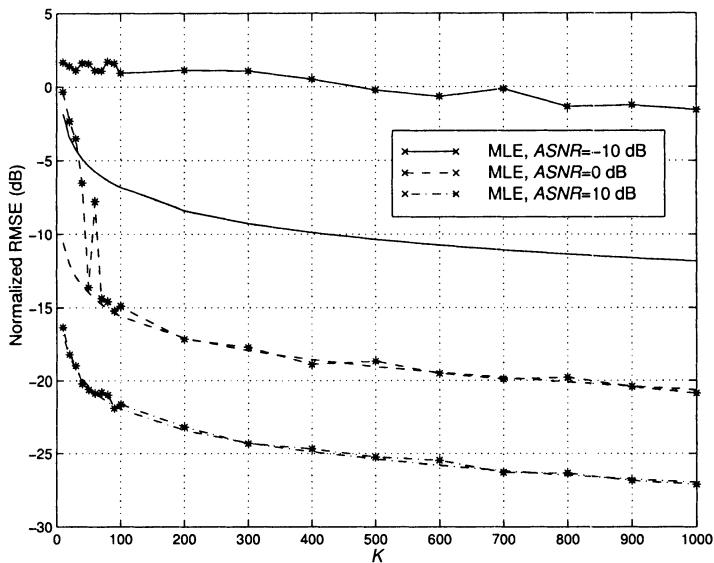


Figure 8.18 AML estimator, single signal, $u_s = 0$, $ASNR = -10$ dB, 0 dB, 10 dB: normalized RMSE versus K .

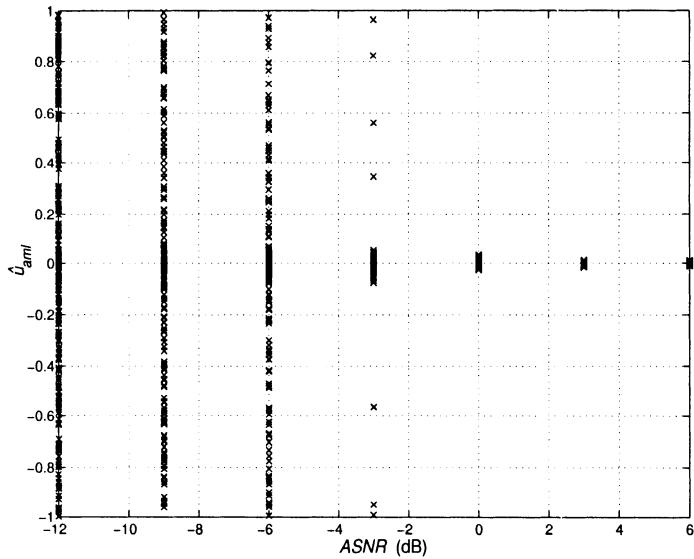


Figure 8.19 AML estimator, single signal, $u_s = 0$, $N = 10$, $K = 100$, 100 trials: scatter plot versus $ASNR$.

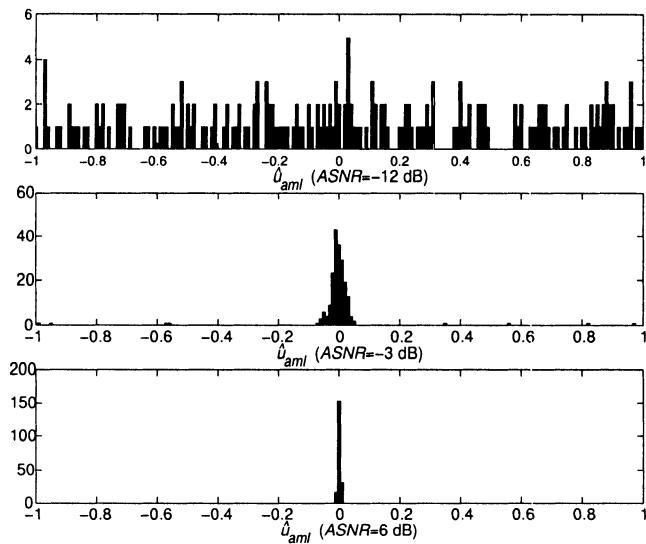


Figure 8.20 AML estimator, single signal, $u_s = 0$, $N = 10$, $K = 100$, 100 trials: histogram for $ASNR = -12$ dB, -3 dB, 6 dB.

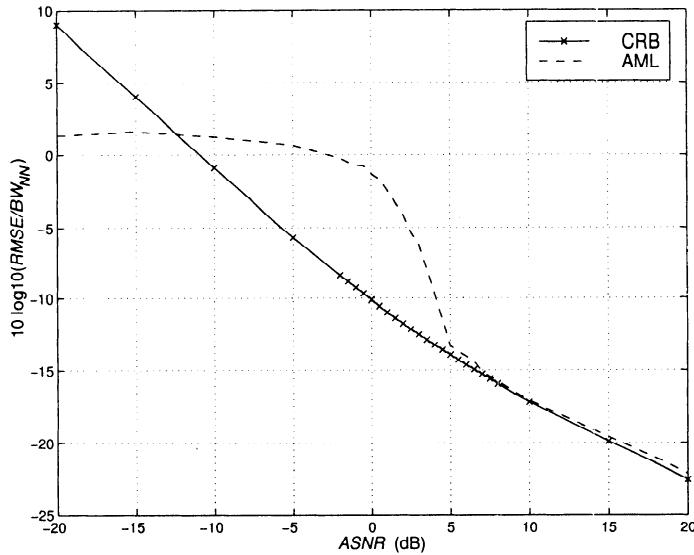


Figure 8.21 AML estimator: two equal-power uncorrelated plane-wave signals at $\pm\Delta u_R/2(0.2165BW_{NN}/2)$, $K = 100$: normalized RMSE versus ASNR.

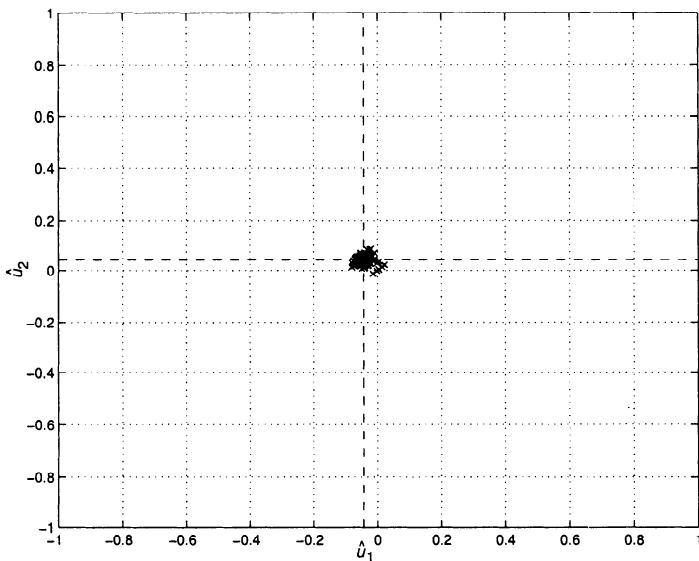


Figure 8.22 AML estimator: two equal-power uncorrelated plane-wave signals at $\pm\Delta u_R/2(0.2165BW_{NN}/2)$, $K = 100$, $ASNR = 6$ dB: scatter plot of AML estimates.

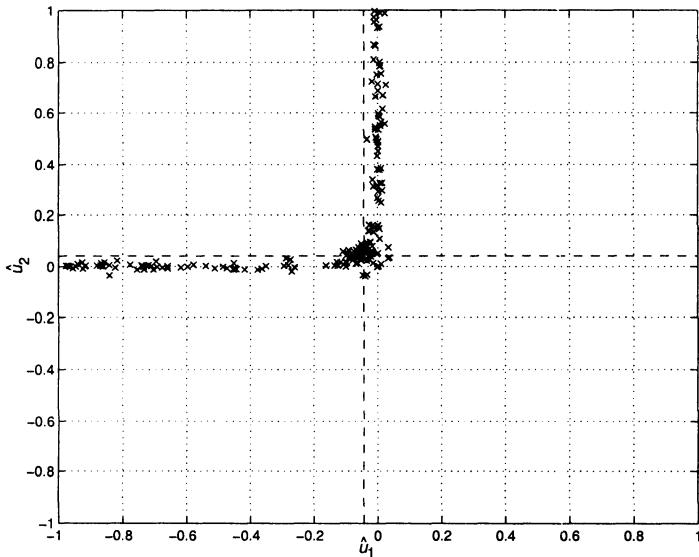


Figure 8.23 AML estimator: two equal-power uncorrelated plane-wave signals at $\pm\Delta u_R/2(0.2165BW_{NN}/2)$, $K = 100$, $ASNR = 0$ dB: scatter plot of AML estimates.

In Figure 8.24, we plot the probability of resolution versus $ASNR$. We see that the threshold in the RMSE occurs as soon as P_R decreases from unity. Therefore in analyzing the P_R behavior, the important region is approximately,

$$0.95 \leq P_R \leq 1.0. \quad (8.323)$$

In Figure 8.25, we show a scatter plot for an $ASNR = -6$ dB. The estimator is far below threshold and the estimates are spread over u -space.

This discussion is appropriate for the parameters in this example, $\Delta u \simeq 0.5HPBW$ and $K = 100$. If Δu becomes too small (e.g., $\Delta u = 0.1HPBW$), then the inequalities in (8.320) and (8.321) will be violated by local errors (and, in some cases, the CRB). In these cases, the threshold occurs at P_R values lower than predicted by (8.323).

In Example 8.5.2, the signal separation is approximately $0.5HPBW$. Another interesting case is a signal separation of $0.5BW_{NN}$. This separation is the classical resolution separation. This case is simulated in Problem 8.5.5. The result is that, for $K = 100$, the threshold occurs at $ASNR = 0$ dB. This result provides a quantitative basis for the classical resolution definition.

In Figure 8.17, the threshold for a single-signal occurs at $ASNR = -1$ dB. Thus, for uncorrelated signals, when the separation exceeds the classical

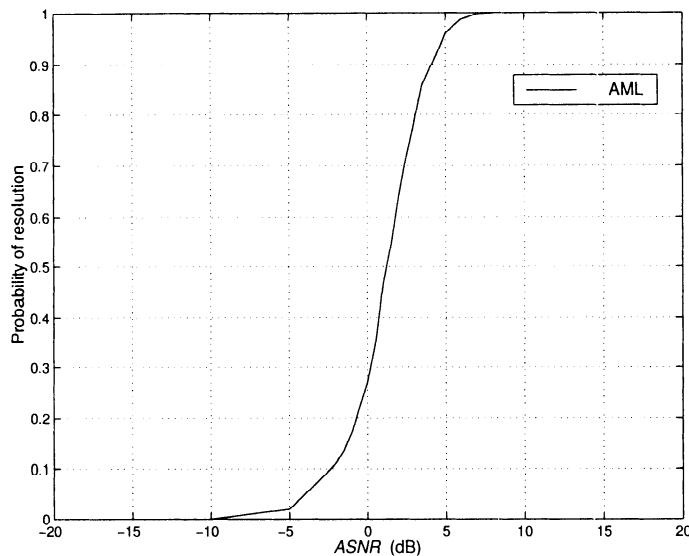


Figure 8.24 AML estimator: two equal-power uncorrelated plane-wave signals at $\pm\Delta u_R/2(0.2165BW_{NN}/2)$, $K = 100$: probability of resolution versus ASNR.

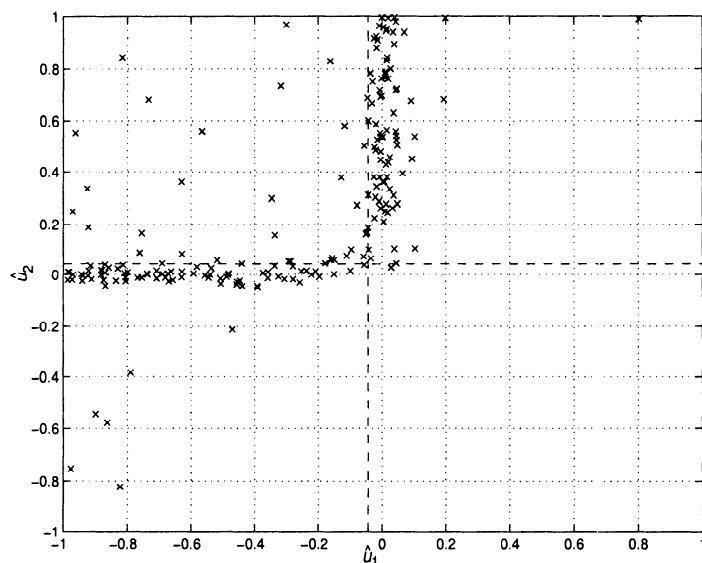


Figure 8.25 AML estimator: two equal-power uncorrelated plane-wave signals at $\pm\Delta u_R/2(0.2165BW_{NN}/2)$, $K = 100$, $ASNR = -6$ dB: scatter plot of AML estimates.

resolution separation, the threshold behavior will be similar to the single-signal model. The CRB behavior was shown in Figure 8.5.

The behavior in Example 8.5.2 is characteristic of most DOA estimation problems. The exact behavior will be a function of the number of signals, the separation of the signals, the source- signal spectral matrix (this includes the signal strengths and their correlation), the noise level, the number of sensor elements, and the number of snapshots.

In this section, we have derived the AML estimator and investigated the performance for two simple examples. The second example illustrates the performance issues that are of interest in all of the estimators that we study. They are:

- (i) The MSE in the high SNR region. We would like the estimator to become unbiased and have a variance that approaches the CRB.
- (ii) The MSE in the asymptotic region. As $K \rightarrow \infty$, we would like the estimator to become unbiased and have a variance that approaches the CRB.
- (iii) For a given K , as the SNR decreases, we reach a point in which the estimator starts to make global errors on some trials. These large errors dominate the small errors of other trials and the MSE rises sharply. This is the threshold behavior that is characteristic of nonlinear parameter estimation problems. We would like the threshold to occur at the lowest possible SNR .
- (iv) For the multiple-signal case, the threshold behavior can be described in terms of a probability of resolution, P_R . As soon as P_R decreases from unity, we move into the threshold region. Typically, this threshold behavior occurs in the range,

$$0.95 \leq P_R \leq 1.0.$$

It is possible to obtain useful analytic results for the asymptotic (high K) and the high SNR region. We discuss them briefly in Section 8.5.5. Some analytic results are available for the probability of resolution and threshold behavior, but we usually have to resort to simulation.

The above four factors relate to performance. The fifth issue is the computational complexity of the estimator. After we discuss the performance capabilities, we look for estimators that are less complex but perform close to optimum.

We will use a set of test scenarios to study the behavior of the parameter estimation algorithms that we develop in this chapter and in Chapter 9. Several of them are designed to stress the estimators. We study other scenarios that exhibit interesting properties, but the test scenarios in Table 8.3 allow a consistent comparison across algorithms. We use a 10-element standard linear array (SLA10) throughout the comparisons. We consider a number of other array geometries in examples and problems, but only do a complete comparison for the SLA. We normalize the RMSE and standard deviation by the BW_{NN} and plot the results versus $ASNR$. Thus the results are valid for any SRA with $N \geq 10$. For smaller arrays, we need to check the result for applicability. The ten test scenarios are shown in Table 8.3.

Table 8.3: Test Scenarios

Test scenario	No. of Signals	Power	Correlation	Separation	K
1	2	Equal	$\rho = 0$	$\Delta\psi = \Delta\psi_R$	$\geq 10N$
2	2	Equal	$\rho = 0$	$\Delta\psi = 0.05BW_{NN}$	$\geq 10N$
3	2	$ASNR_2 = 10ASNR_1$ $ASNR_2 = 100ASNR_1$	$\rho = 0$	$\Delta\psi = \Delta\psi_R$	$\geq 10N$
4	2	Equal	$0 < \rho < 1.0$	$\Delta\psi = \Delta\psi_R$	$\geq 10N$
5	2	Equal	$ \rho = 1$	$\Delta\psi = \Delta\psi_R$	$\geq 10N$
6	2	Equal	$\rho = 0$	$\Delta\psi = \Delta\psi_R$	$\geq 2N$
7	3	Equal	$\rho = \mathbf{I}$	$\psi_1 = -\Delta\psi_R$ $\psi_2 = 0$ $\psi_3 = \Delta\psi_R$	$\geq 10N$
8	3	Equal	$\rho \neq \mathbf{I}$	$\psi_1 = -\Delta\psi_R$ $\psi_2 = 0$ $\psi_3 = \Delta\psi_R$	$\geq 10N$
9	3	$ASNR_1 = ASNR_3$ $= 10 ASNR_2$ $ASNR_1 = ASNR_3$ $= 100 ASNR_2$	$\rho = \mathbf{I}$	$\psi_1 = -\Delta\psi_R$ $\psi_2 = 0$ $\psi_3 = \Delta\psi_R$	$\geq 10N$
10	5	Equal	$\rho = \mathbf{I}$	$\psi_1 = -2\Delta\psi_R$ $\psi_2 = -\Delta\psi_R$ $\psi_3 = 0$ $\psi_4 = \Delta\psi_R$ $\psi_5 = 2\Delta\psi_R$	$\geq 10N$

Notes: Table 8.3

- Recall that $\Delta\psi_R \triangleq 0.2165BW_{NN}$. This is approximately $0.5HPBW$. The subscript “R” denotes reference.
- Because we are working in ψ -space (or u -space), the estimation results for element space algorithms depend on signal separation, not absolute location. For beamspace algorithms, the location with respect to the beam fan must be specified.
- Various ρ matrices will be considered in Test Scenario 8.

In the text, we only show representative test scenarios. Before studying the behavior of AML in more detail, we look at other stochastic estimators.

8.5.1.2 Unconditional ML estimators

We now return to the issue that was pointed out after (8.296). The estimate of $\widehat{\mathbf{S}}_{\mathbf{f},ml}$ given by (8.296) may not be non-negative definite and, therefore, it is not necessarily a proper estimate to use in the subsequent part of the algorithm.

Bresler [Bre88] has analyzed this issue in detail and devised an ML algorithm that guarantees a non-negative definite estimate of $\widehat{\mathbf{S}}_{\mathbf{f}}$. Bresler's algorithm assumed that ψ was known, but it is straightforward to develop an iterative version of his algorithm. We implemented this iterative version of the Bresler algorithm for the test scenarios in Table 8.3 and found that the performance in estimating ψ was essentially the same as the AML algorithm over the entire range of *ASNR* (both above and below threshold).

There are two other estimates of $\widehat{\mathbf{S}}_{\mathbf{f}}$ that lead to different estimators. The first approach uses

$$\tilde{\sigma}_w^2 = \frac{1}{N-D} \sum_{i=D+1}^N \hat{\lambda}_i, \quad (8.324)$$

where the $\hat{\lambda}_i$ are the $N - D$ smallest eigenvalues of $\mathbf{C}_{\mathbf{x}}$, as the estimate of σ_w^2 . Then, we let

$$\widehat{\mathbf{S}}_{\mathbf{f},I} = \mathbf{V}^\dagger \left(\mathbf{C}_{\mathbf{x}} - \tilde{\sigma}_w^2 \mathbf{I} \right) \left(\mathbf{V}^\dagger \right)^H, \quad (8.325)$$

where \mathbf{V}^\dagger is the Moore-Penrose pseudoinverse of \mathbf{V} ,

$$\mathbf{V}^\dagger \triangleq [\mathbf{V}^H \mathbf{V}]^{-1} \mathbf{V}^H. \quad (8.326)$$

Then, by analogy with (8.300)

$$\widehat{\mathbf{S}}_{\mathbf{x},I}(\psi) = \mathbf{P}_{\mathbf{V}} \left[\mathbf{C}_{\mathbf{x}} - \tilde{\sigma}_w^2 \mathbf{I} \right] \mathbf{P}_{\mathbf{V}} + \tilde{\sigma}_w^2 \mathbf{I}. \quad (8.327)$$

Using (8.327) in (8.301) and proceeding as before, we obtain,

$$\hat{\psi}_{wml} = \arg \min_{\psi} \left\{ \left[\ln \det \widehat{\mathbf{S}}_{\mathbf{x},I}(\psi) + \text{tr} \left[\widehat{\mathbf{S}}_{\mathbf{x},I}^{-1}(\psi) \mathbf{C}_{\mathbf{x}} \right] \right] \right\}, \quad (8.328)$$

This estimate is due to Wax et al. [WSK82].

Schmidt [Sch79] suggested using

$$\widehat{\mathbf{S}}_{\mathbf{f},II}(\psi) = \mathbf{V}^\dagger \widehat{\mathbf{U}}_S \widehat{\Lambda}_S \widehat{\mathbf{U}}_S^H \left(\mathbf{V}^\dagger \right)^H. \quad (8.329)$$

This estimate is unbiased and consistent. Then,

$$\hat{\mathbf{S}}_{\mathbf{x},II}(\psi) = \mathbf{P}_{\mathbf{V}} \hat{\mathbf{U}}_S \hat{\Lambda}_S \hat{\mathbf{U}}_S^H \mathbf{P}_{\mathbf{V}} + \tilde{\sigma}_w^2 \mathbf{I}. \quad (8.330)$$

where $\tilde{\sigma}_w^2$ is given by (8.324) and

$$\hat{\psi}_{scml} = \arg \min_{\psi} \left\{ \ln \det \hat{\mathbf{S}}_{\mathbf{x},II}(\psi) + \text{tr} \left[\hat{\mathbf{S}}_{\mathbf{x},II}^{-1}(\psi) \mathbf{C}_{\mathbf{x}} \right] \right\}, \quad (8.331)$$

where $\hat{\mathbf{S}}_{\mathbf{x},II}(\psi)$ is given by (8.330).

We implemented the two algorithms for the test scenarios in Table 8.3 and found that their performance was essentially the same as the AML estimator over the entire range of $ASNR$. Stoica et al. [SOVM96] derive a ML estimator for coherent signals that guarantees that $\hat{\mathbf{S}}_{\mathbf{x}}$ is a positive semi-definite matrix of a given rank. They show that the AML estimate in (8.315) is the large-sample realization of the ML estimate even when the signals are coherent.

We focus our subsequent work on the AML estimator. In the next section, we simulate the AML estimator for some of the test scenarios in Table 8.3.

8.5.1.3 Performance of AML estimators

We consider a sequence of six examples to illustrate the behavior of the AML algorithms. All of the examples utilize a standard 10-element linear array with two signals impinging on it.

Example 8.5.3 (continuation, Example 8.5.2)

We use the same model as in Example 8.5.2, except the separation is reduced to $0.05BW_{NN}$. The signals are uncorrelated ($\rho = 0$). The results are shown in Figure 8.26. We see that the threshold has moved about 13 dB to the right. Later, we explore the threshold behavior as a function of $\Delta\psi$.

Example 8.5.4 (continuation)

We use the same model as in Example 8.5.2, except $|\rho| = 0.95$. The separation $\Delta\psi$ equals $\Delta\psi_R$. We consider two phase angles for ρ . In the first case, $\phi_{\rho} = 0$, and in the second case $\phi_{\rho} = \pi/4$. The results for $\phi_{\rho} = 0$ are shown in Figure 8.27. We see that the threshold is 5 dB higher and occurs at $ASNR = 10$ dB. Above the threshold the AML estimator converges to the CRB. We recall from Figure 8.6 that the CRB is higher for the correlated signal case. The results for $\phi_{\rho} = \pi/4$ are shown in Figure 8.28. We see that the behavior for this phase angle is very similar to the uncorrelated case. This result illustrates the importance of considering various phase angles when studying the correlated signal model.

We also simulated the AML estimator for $|\rho| = 1$ with $\phi_{\rho} = 0$ and $\pi/4$. The results are very similar to the $|\rho| = 0.95$ case.

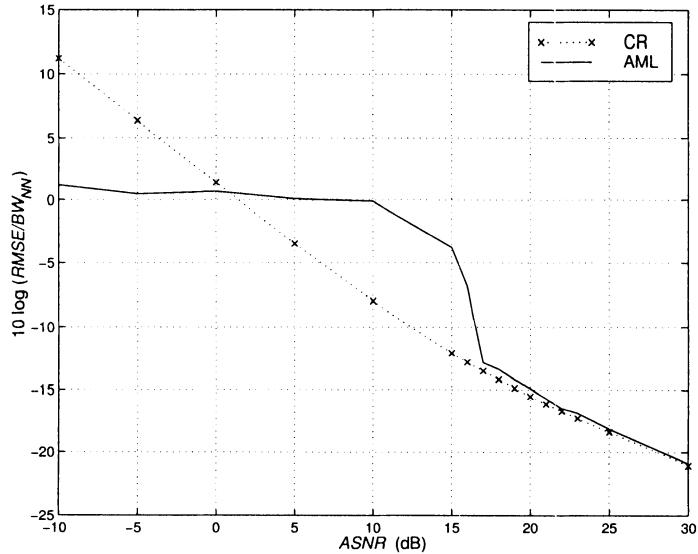


Figure 8.26 AML estimator: two equal-power uncorrelated plane-wave signals, $\Delta u = 0.05BW_{NN}$, $K = 10N$; normalized RMSE versus $ASNR$.

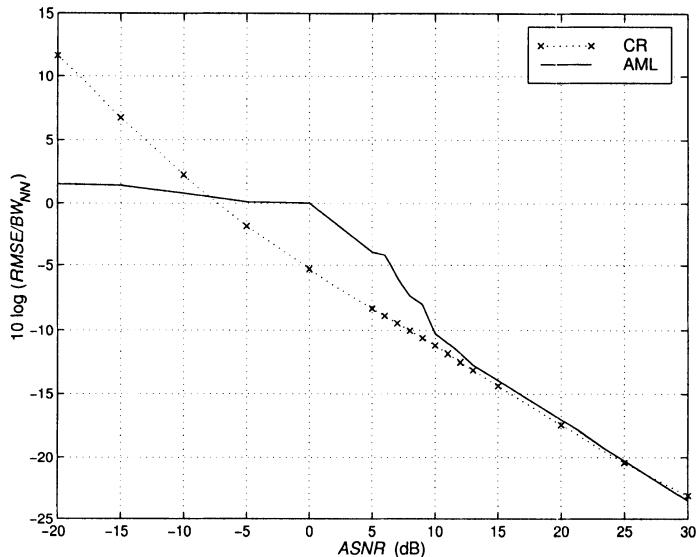


Figure 8.27 AML estimator: two equal-power correlated signals, $\Delta u = 0.2165BW_{NN}$, $K = 10N$, $\rho = 0.95$, $\phi_\rho = 0$; normalized RMSE versus $ASNR$.

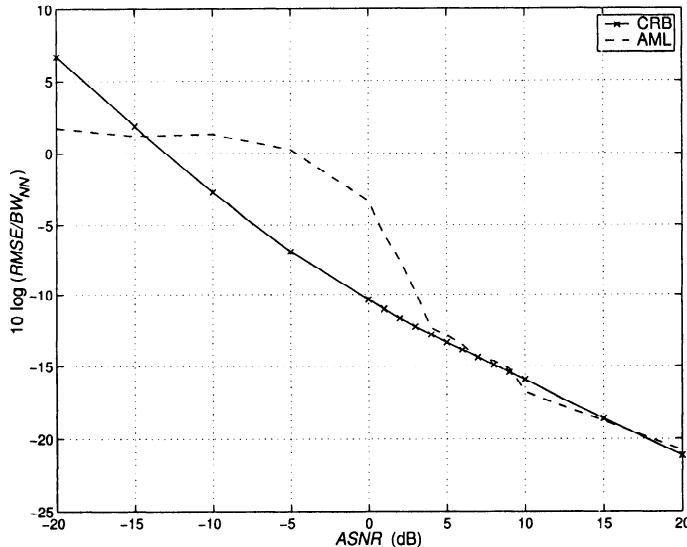


Figure 8.28 AML estimator: two equal-power correlated signals, $\Delta u = 0.2165BW_{NN}$, $K = 10N$, $\rho = 0.95$, $\phi_\rho = \pi/4$; normalized RMSE versus ASNR. (Note that the vertical scale is different from Figure 8.27.)

Example 8.5.5: Unequal signal powers

Consider a standard 10-element linear array with two uncorrelated plane-wave signals impinging on it. The signal separation is $\Delta\psi_R$. The *SNR* of signal 2 is 20 dB higher than the *SNR* of signal 1. The results are shown in Figures 8.29 and 8.30. The RMSE of the weaker signal moves away from the CRB at $ASNR_1 = 1$ dB. The RMSE of the stronger signal moves away from the CRB at $ASNR_2 = 24$ dB.

Example 8.5.6: (Low sample support)

We consider a standard 10-element linear array. We use the same signal model as in Example 8.5.3 except $K = 20$ ($2N$). The results are plotted in Figure 8.31. We see that the threshold occurs at an $ASNR = 11$ dB. This value is 6 dB higher than the $K = 10N$ case.

We have considered four scenarios in this section. Above threshold, the AML estimator approaches the CRB. However, the location of the threshold varies with the parameters in the scenario.

In our discussion of the CRB, we found that, if we knew that the signals were uncorrelated, the CRB was lower in certain scenarios. The next logical step is to develop a ML estimator that exploits that *a priori* knowledge. An uncorrelated ML has been developed by Bell [Bell99] and shows improved performance in the scenarios where the CRBs differed. The reader is referred

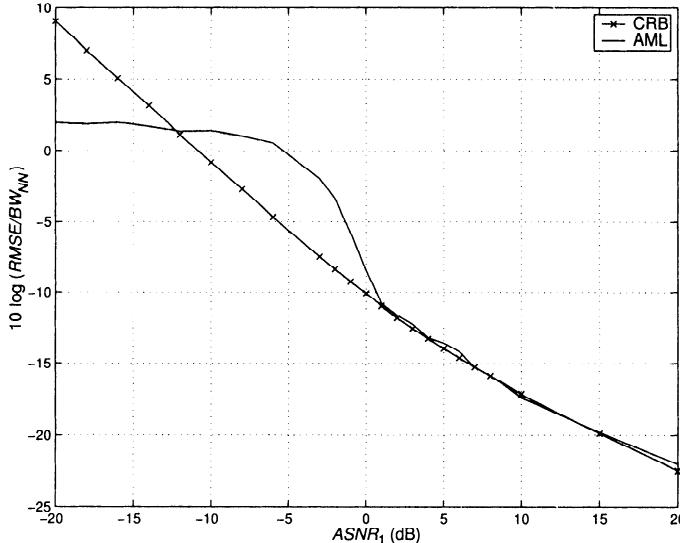


Figure 8.29 AML estimator: two uncorrelated plane-wave signals with unequal power, $\Delta u = 0.2165BW_{NN}$, $ASN R_2 = ASN R_1 + 20$ dB; normalized RMSE of signal 1 versus $ASN R_1$.

to this reference for further discussion.

8.5.2 Conditional Maximum Likelihood Estimators

In this section, we develop a ML estimator that is referred to in the literature as a deterministic or conditional maximum likelihood (CML) estimator. The reason for the name will be clear when we formulate the model.

In this section, we derive the estimator and investigate its performance. We compare its performance to the stochastic maximum likelihood estimator of Section 8.5.1 and to the CRB.

We consider the same snapshot model as in Section 8.5.1,

$$\mathbf{X}(k) = \mathbf{V}\mathbf{F}(k) + \mathbf{N}(k), \quad k = 1, 2, \dots, K. \quad (8.332)$$

However, we model the source signals as unknown nonrandom signals. Thus, $\mathbf{F}(k)$ is a $D \times 1$ vector

$$\mathbf{F}(k) = [F_1(k) \mid F_2(k) \mid \cdots \mid F_D(k)]^T, \quad (8.333)$$

whose elements are unknown nonrandom complex numbers.

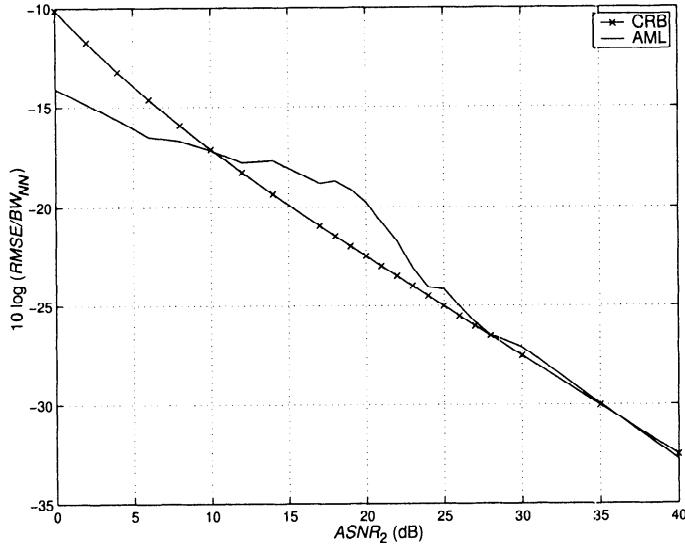


Figure 8.30 AML estimator: two uncorrelated plane-wave signals with unequal power, $\Delta u = 0.2165 BW_{NN}$, $ASNR_2 = ASNR_1 + 20$ dB; normalized RMSE of signal 2 versus $ASNR_2$.

Alternatively, we can consider $\mathbf{F}(k)$ as a specific sample function from a random process and design an estimator based on that sample function. This is sometimes referred to as *conditional maximum likelihood* estimation.

The noise process is a sample function from a Gaussian random process. We assume that it is spatially uncorrelated so that

$$E \left[\mathbf{N}(k) \mathbf{N}^H(k) \right] = \sigma_w^2 \mathbf{I}. \quad (8.334)$$

The noise in successive snapshots is assumed to be statistically independent. The spectral height σ_w^2 is known. Later we consider the case of unknown σ_w^2 . In that case we find the ML estimate of σ_w^2 as part of the estimation process. We assume the number of signals, D , is known.

The joint probability density is

$$p_{\mathbf{x}}(\mathbf{X}) = \prod_{k=1}^K \frac{1}{\pi \sigma_w^2 \mathbf{I}} \exp \left\{ -\frac{1}{\sigma_w^2} |\mathbf{X}(k) - \mathbf{V}(\psi) \mathbf{F}(k)|^2 \right\}, \quad (8.335)$$

and the ln likelihood function is

$$L(\psi, \mathbf{F}) = -KN \ln \sigma_w^2 - \frac{1}{\sigma_w^2} \sum_{k=1}^K |\mathbf{X}(k) - \mathbf{V}(\psi) \mathbf{F}(k)|^2 \quad (8.336)$$

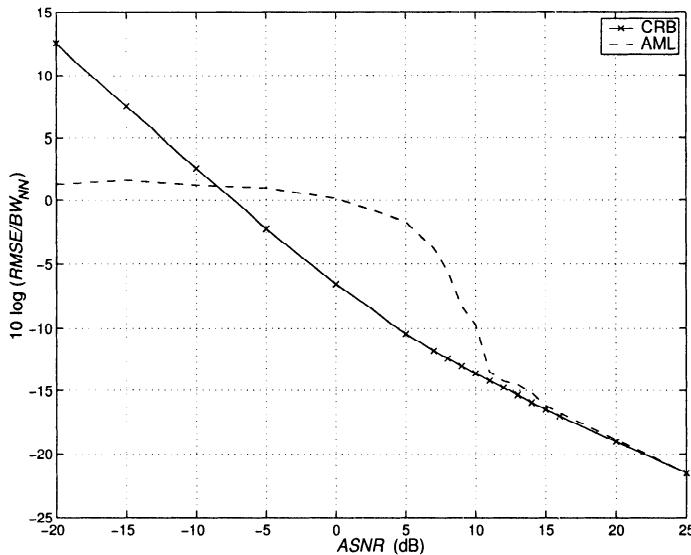


Figure 8.31 AML estimator: two equal-power uncorrelated plane-wave signals with low sample support, $\Delta u = 0.2165BW_{NN}$, $K = 2N$; normalized RMSE versus $ASNR$.

(we have dropped the term containing π), where ψ is an $D \times 1$ vector and \mathbf{F} is a $D \times K$ matrix whose k th column is $\mathbf{F}(k)$. We want to maximize this function over the unknown parameters (ψ, \mathbf{F}) . This is equivalent to minimizing¹³

$$L_2(\psi, \mathbf{F}) \triangleq \sum_{k=1}^K |\mathbf{X}(k) - \mathbf{V}(\psi)\mathbf{F}(k)|^2. \quad (8.337)$$

If we fix ψ and minimize over $\mathbf{F}(k)$, the result is just the matrix MVDR filter of Section 6.5 (see (6.200)). Thus,

$$\hat{\mathbf{F}}(k) = \mathbf{W}_o^H(\psi)\mathbf{X}(k), \quad (8.338)$$

where

$$\mathbf{W}_o^H(\psi) = [\mathbf{V}^H(\psi)\mathbf{V}(\psi)]^{-1}\mathbf{V}^H(\psi) = \mathbf{V}^\dagger(\psi). \quad (8.339)$$

The result in (8.338) and (8.339) is exactly what we derived in Section 6.5 where ψ was known. Substituting (8.338) and (8.339) into (8.337), we obtain

¹³The subscripts signify that the functions are different although they have the same minimum (or maximum).

the function

$$L_3(\psi) = \sum_{k=1}^K \left| \mathbf{X}(k) - \mathbf{V}(\psi) [\mathbf{V}^H(\psi) \mathbf{V}(\psi)]^{-1} \mathbf{V}^H(\psi) \mathbf{X}(k) \right|^2, \quad (8.340)$$

which is to be minimized over ψ . We recognize the coefficient of the second term as the projection matrix onto the columns of $\mathbf{V}(\psi)$

$$\begin{aligned} L_3(\psi) &= \sum_{k=1}^K \left| \mathbf{X}(k) - \mathbf{P}_{\mathbf{V}(\psi)} \mathbf{X}(k) \right|^2 \\ &= \sum_{k=1}^K \left| \mathbf{P}_{\mathbf{V}(\psi)}^\perp \mathbf{X}(k) \right|^2. \end{aligned} \quad (8.341)$$

Thus we can either minimize $L_3(\psi)$ over ψ or, equivalently, maximize

$$L_4(\psi) = \sum_{k=1}^K \left| \mathbf{P}_{\mathbf{V}(\psi)} \mathbf{X}(k) \right|^2. \quad (8.342)$$

To get a useful geometric interpretation of (8.342), recall our discussion of signal subspaces in Section 5.5. The signal component of $\mathbf{X}(k)$ always lies in the signal subspace defined by the columns of $\mathbf{V}(\psi)$. The noise has two effects; it adds noise into the signal subspace and it adds noise in an orthogonal space that causes $\mathbf{X}(k)$ to lie outside the signal subspace. The CML estimator finds the D steering vectors that form a signal subspace that is as close as possible to the $\mathbf{X}(k)$, $k = 1, 2, \dots, K$. Closeness is measured by the magnitude of the projection of $\mathbf{X}(k)$, $k = 1, 2, \dots, K$ onto the estimated signal subspace.¹⁴

A second interpretation of (8.342) is also useful. We can rewrite (8.342) as

$$\begin{aligned} L_5(\psi) &= \frac{1}{K} \text{tr} \left[\mathbf{P}_{\mathbf{V}} \sum_{k=1}^K \mathbf{X}(k) \mathbf{X}(k)^H \mathbf{P}_{\mathbf{V}} \right] \\ &= \text{tr} [\mathbf{P}_{\mathbf{V}} \mathbf{P}_{\mathbf{V}} \mathbf{C}_{\mathbf{x}}] = \text{tr} [\mathbf{P}_{\mathbf{V}} \mathbf{C}_{\mathbf{x}}], \end{aligned} \quad (8.343)$$

where

$$\mathbf{C}_{\mathbf{x}} \triangleq \frac{1}{K} \sum_{k=1}^K \mathbf{X}(k) \mathbf{X}(k)^H. \quad (8.344)$$

¹⁴This result in (8.342) is equivalent to that first obtained by Schweppe [Sch68]; the geometric interpretation is due to Ziskind and Wax [ZW88a], [ZW88b].

Thus, the CML estimate is

$$\hat{\psi}_{cml} = \arg \max_{\psi} \{ \text{tr} [\mathbf{P}_V \mathbf{C}_x] \}, \quad (8.345)$$

or equivalently,

$$\hat{\psi}_{cml} = \arg \min_{\psi} \left\{ \text{tr} [\mathbf{P}_V^\perp \mathbf{C}_x] \right\}. \quad (8.346)$$

Now decompose \mathbf{C}_x using an eigenvector expansion,

$$\mathbf{C}_x = \sum_{i=1}^N \hat{\lambda}_i \hat{\Phi}_i \hat{\Phi}_i^H. \quad (8.347)$$

Then we can write

$$L_5(\psi) = \sum_{i=1}^N \hat{\lambda}_i \left| \mathbf{P}_V \hat{\Phi}_i \right|^2. \quad (8.348)$$

Thus, the CML estimator projects each of the estimated eigenvectors onto the signal subspace, weights the magnitude squared of that projection with the estimated eigenvalue, and sums over all eigenvectors. It does this for each value of ψ and chooses the value of ψ that gives the maximum value.

If we compare the conditional likelihood function in (8.345) with the asymptotic stochastic likelihood function in (8.310), we see that the difference is the first term in (8.310):

$$L_{aml}(\psi) - L_{cml}(\psi) = -\ln \det [\mathbf{P}_V \mathbf{C}_x \mathbf{P}_V + \sigma_w^2 \mathbf{P}_V^\perp]. \quad (8.349)$$

Thus, the two estimates will be similar when the term in (8.349) is negligible compared to $\text{tr} [\mathbf{P}_V^\perp \mathbf{C}_x] / \sigma_w^2$. Conversely, the two estimates will be different when the term in (8.349) is significant compared to $\text{tr} [\mathbf{P}_V^\perp \mathbf{C}_x] / \sigma_w^2$ and has its maximum at a different ψ .

By considering several examples, we observe that the estimates are more likely to be different in the following cases:

- (i) Small signal separation relative to BW_{NN} .
- (ii) Sources with high correlation and coherent sources.
- (iii) A mixture of high SNR and low SNR signals.

If the noise variance σ_w^2 is unknown, then we estimate it. Returning to (8.336), we must retain the first term. Using (8.336) and (8.343), we have

$$L(\psi, \hat{\mathbf{F}}, \sigma_w^2) = -N \ln \sigma_w^2 - \frac{1}{\sigma_w^2} \text{tr} [\mathbf{P}_V^\perp \mathbf{C}_x]. \quad (8.350)$$

Differentiating with respect to σ_w^2 and setting the result equal to zero, we obtain

$$\hat{\sigma}_w^2 = \frac{\text{tr} [\mathbf{P}_{\mathbf{V}}^\perp \mathbf{C}_{\mathbf{x}}]}{N}. \quad (8.351)$$

Note that this estimate is different than the estimate in (8.313). Substituting (8.351) into (8.350) gives

$$L(\psi) = -N \ln (\text{tr} [\mathbf{P}_{\mathbf{V}}^\perp \mathbf{C}_{\mathbf{x}}]). \quad (8.352)$$

Maximizing (8.352) is equivalent to minimizing the argument of the logarithm, so

$$\hat{\psi}_{cml} = \arg \min_{\psi} \left\{ \text{tr} [\mathbf{P}_{\mathbf{V}}^\perp \mathbf{C}_{\mathbf{x}}] \right\}, \quad (8.353)$$

which is identical to (8.346). Thus, knowing the variance of the noise does not affect the CML estimate.

Before doing a set of examples, we derive a closely related estimator.

8.5.3 Weighted Subspace Fitting

We can interpret the CML estimator as an algorithm that fits the subspace spanned by $\mathbf{V}(\psi)$ to the measurements,

$$\tilde{\mathbf{X}}_K \triangleq \begin{bmatrix} \mathbf{X}(1) & \mathbf{X}(2) & \cdots & \mathbf{X}(K) \end{bmatrix}, \quad (8.354)$$

in a least squares sense. To obtain this interpretation we rewrite $L_2(\psi, \mathbf{F})$ in (8.337) as a Frobenius norm,

$$\begin{aligned} L_2(\psi, \mathbf{F}) &= \left\| \tilde{\mathbf{X}}_K - \mathbf{V}(\psi) \mathbf{F} \right\|_F^2 \\ &= \text{tr} \left\{ [\tilde{\mathbf{X}}_K - \mathbf{V}(\psi) \mathbf{F}]^H [\tilde{\mathbf{X}}_K - \mathbf{V}(\psi) \mathbf{F}] \right\}, \end{aligned} \quad (8.355)$$

and

$$(\hat{\psi}, \hat{\mathbf{F}}) = \arg \min_{\psi, \mathbf{F}} \{L_2(\psi, \mathbf{F})\}. \quad (8.356)$$

From (8.338)

$$\hat{\mathbf{F}} = \mathbf{V}^\dagger \tilde{\mathbf{X}}_K, \quad (8.357)$$

where \mathbf{V}^\dagger is the Moore-Penrose pseudoinverse,

$$\mathbf{V}^\dagger = [\mathbf{V}^H \mathbf{V}]^{-1} \mathbf{V}^H. \quad (8.358)$$

Then (8.356) reduces to

$$\hat{\psi}_{cml} = \arg \min_{\psi} \left\{ \text{tr} \left[\mathbf{P}_V^\perp \tilde{\mathbf{X}}_K \right]^H \left[\mathbf{P}_V^\perp \tilde{\mathbf{X}}_K \right] \right\}, \quad (8.359)$$

which can be written as

$$\begin{aligned} \hat{\psi}_{cml} &= \arg \min_{\psi} \left\{ \text{tr} \left[\mathbf{P}_V^\perp \tilde{\mathbf{X}}_K \tilde{\mathbf{X}}_K^H \right] \right\} \\ &= \arg \min_{\psi} \left\{ \text{tr} \left[\mathbf{P}_V^\perp \mathbf{C}_x \right] \right\}. \end{aligned} \quad (8.360)$$

We now consider more general subspace fitting techniques. The goal is to find an estimator that approaches the CRB asymptotically and has good performance in the threshold region. In addition, we would like the estimator to be computationally simpler than $\hat{\psi}_{cml}$.

The notion of weighted subspace fitting was introduced by Viberg and Ottersten [VO91].

We let \mathbf{M} be an $N \times D'$ matrix representing the data, D' is the rank of $\mathbf{V}(\psi)\mathbf{F}$ (unless the signals are coherent, $D' = D$). An example of an \mathbf{M} matrix that we use is obtained by an eigendecomposition of \mathbf{C}_x ,

$$\mathbf{C}_x = \widehat{\mathbf{U}}_S \widehat{\Lambda}_S \widehat{\mathbf{U}}_S^H + \widehat{\mathbf{U}}_N \widehat{\Lambda}_N \widehat{\mathbf{U}}_N^H. \quad (8.361)$$

If we only utilize the data in the estimated signal subspace, we can define \mathbf{M} as,

$$\mathbf{M} = \widehat{\mathbf{U}}_S \widehat{\Lambda}_S^{\frac{1}{2}}. \quad (8.362)$$

More generally we can write

$$\mathbf{M} = \widehat{\mathbf{U}}_S \mathbf{W}^{\frac{1}{2}}, \quad (8.363)$$

where \mathbf{W} is a $D' \times D'$ diagonal matrix. This definition leads to the name, **weighted (signal) subspace fitting** (WSF). We define the cost function,

$$\begin{aligned} L(\psi, \mathbf{T}) &\triangleq \|\mathbf{M} - \mathbf{V}(\psi)\mathbf{T}\|_F^2 \\ &= \text{tr} \left\{ [\mathbf{M} - \mathbf{V}(\psi)\mathbf{T}]^H [\mathbf{M} - \mathbf{V}(\psi)\mathbf{T}] \right\}, \end{aligned} \quad (8.364)$$

where \mathbf{T} is a $D \times D'$ matrix. \mathbf{M} is taking the place of $\tilde{\mathbf{X}}_k$ and incorporates the data input through (8.361). \mathbf{T} is taking the place of \mathbf{F} .

Minimizing $L(\psi, \mathbf{T})$ with respect to \mathbf{T} gives,

$$\widehat{\mathbf{T}} = \mathbf{V}^\dagger \mathbf{M}. \quad (8.365)$$

Substituting (8.365) into (8.364) and using the same steps as in (8.343)–(8.346), we obtain

$$\hat{\psi} = \arg \min_{\psi} \left\{ \text{tr} \left[\mathbf{P}_V^\perp \mathbf{M} \mathbf{M}^H \right] \right\}. \quad (8.366)$$

Using (8.363) in (8.366) gives

$$\hat{\psi} = \arg \min_{\psi} \left\{ \text{tr} \left[\mathbf{P}_V^\perp \hat{\mathbf{U}}_S \mathbf{W} \hat{\mathbf{U}}_S^H \right] \right\}. \quad (8.367)$$

The next step is to determine an optimum weighting \mathbf{W} . This is difficult to do for arbitrary K . Viberg and Ottersten [VO91] consider the asymptotic case and show that the weighting

$$\mathbf{W}_{ao} \triangleq \hat{\tilde{\Lambda}}_S^{-1}, \quad (8.368)$$

where

$$\hat{\tilde{\Lambda}} \triangleq \hat{\Lambda}_S - \hat{\sigma}_w^2 \mathbf{I}, \quad (8.369)$$

with

$$\hat{\sigma}_w^2 \triangleq \frac{1}{N-D} \sum_{i=D+1}^N \hat{\lambda}_i, \quad (8.370)$$

gives an estimator that asymptotically achieves the stochastic CRB. We investigate its performance in the non-asymptotic region by simulation. We refer to the estimator described by (8.367)–(8.370) as the asymptotically optimal WSF estimator (AOWSF) or more compactly, WSF_{ao} . When $D' = D$, the AOWSF cost function is equivalent to the method of direction estimation (MODE) algorithm derived by Stoica and Sharman [SS90a]. The two algorithms of interest are summarized in Table 8.4.

Table 8.4

CML	$\hat{\psi} = \arg \min_{\psi} \left\{ \text{tr} \left[\mathbf{P}_V^\perp \mathbf{C}_x \right] \right\}.$
$\text{WSF}_{ao}/\text{MODE}$	$\hat{\psi} = \arg \min_{\psi} \left\{ \text{tr} \left[\mathbf{P}_V^\perp \hat{\mathbf{U}}_S \hat{\tilde{\Lambda}}^2 \hat{\Lambda}_S^{-1} \hat{\mathbf{U}}_S^H \right] \right\}.$

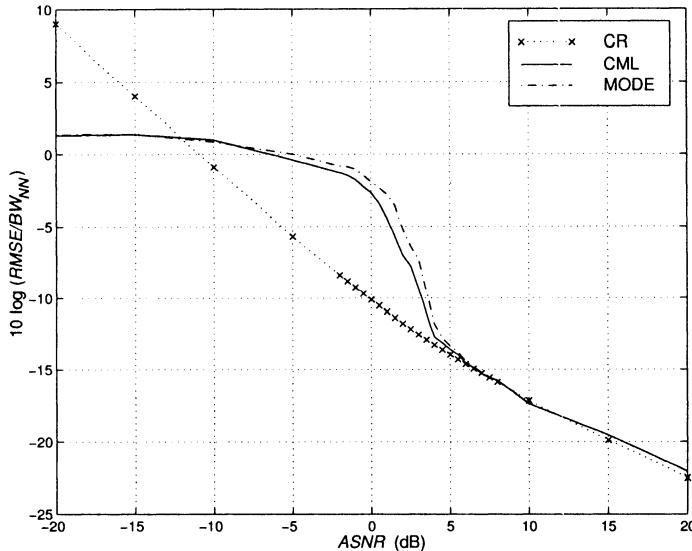


Figure 8.32 Normalized RMSE versus $ASNR$: CML and MODE estimators; $\Delta\psi = \Delta\psi_R$, $N = 10$, $K = 100$, $\rho = 0$, 100 trials.

We consider several of the examples that we discussed in the stochastic maximum likelihood case.

Example 8.5.7 (continuation, Example 8.5.2)

We consider the same model as in Example 8.5.2. The results are shown in Figure 8.32. We have also shown the AML estimator for comparison. We see that the estimators are close to the CRB for $ASNR$ s greater than 10 dB. Below the threshold, CML and WSF (MODE) behave in a similar manner. We see that the threshold of all of conditional estimators are slightly to the left of the threshold of the AML estimator.

In Example 8.5.3, we reduced the separation to $\Delta u = 0.05BW_{NN}$. We implemented the CML and MODE algorithms for this case and found that their threshold was the same as the AML estimator ($\simeq 17$ dB) and that their RMSE approached the CRB above threshold.

Example 8.5.8 (continuation, Example 8.5.4)

We consider the same model as in Example 8.5.4. $\Delta\psi = \Delta\psi_R$ and $\rho = 0.95e^{j\phi_\rho}$. In Figure 8.33, we show the results for $\phi_\rho = 0$. We see that MODE performs better than CML. The MODE threshold is slightly higher than the AML threshold in Figure 8.27. Above threshold, the approach to the CRB is slower than the uncorrelated case for both estimators.

In Figure 8.34, we plot the same results for $\phi_\rho = \pi/4$. The difference in performance between MODE and CML is larger. The MODE threshold is about 1 dB higher than the

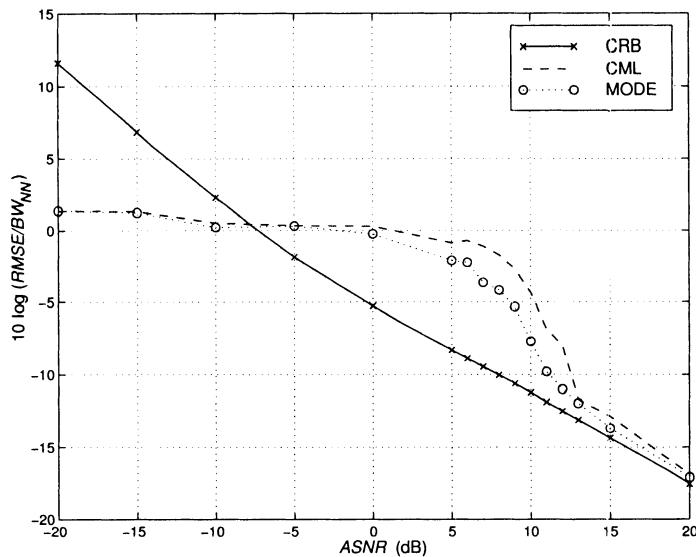


Figure 8.33 Normalized RMSE versus ASNR: MODE and CML estimators; $\Delta\psi = \Delta\psi_R$, $N = 10$, $K = 100$, $\rho = 0.95$, 100 trials.

AML threshold.

In Example 8.5.5, we considered the case of unequal signal powers. We implemented the CML and MODE algorithms for this scenario. The CML and MODE algorithms have very similar behavior and their performance is almost the same as the AML algorithm.

In Example 8.5.6, we considered the case of low sample support ($K = 2N$). We implemented the CML and MODE algorithms for this scenario. The CML and MODE algorithms have very similar performance and their performance is almost the same as the AML algorithm.

It is risky to draw general conclusions from a limited set of examples. However, it appears that AML, CML, and MODE all have similar performance in several interesting scenarios. In the correlated signal case, MODE is better than CML and similar to AML.

In the next section we discuss the asymptotic performance of the estimators that we have derived.

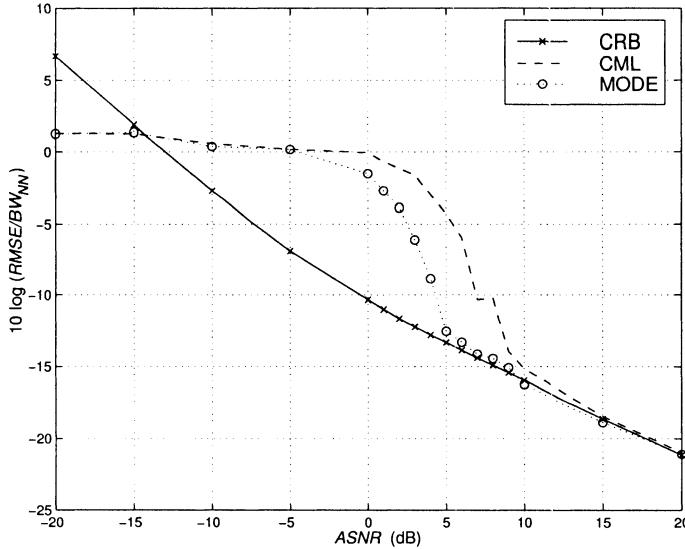


Figure 8.34 Normalized RMSE versus ASNR; MODE and CML estimators, $\Delta\psi = \Delta\psi_R$, $K = 10N$, $\rho = 0.95e^{j\pi/4}$, 100 trials.

8.5.4 Asymptotic Performance

The performance of estimators as the number of snapshots, K , goes to infinity is referred as the asymptotic performance and has been extensively studied in classical statistical theory (see, e.g., [And84] or [KS61]). An estimate is consistent if

$$\lim_{K \rightarrow \infty} E[\hat{\theta}] = \theta, \quad (8.371)$$

and is efficient if the variance of θ_i approaches the CRB.

The asymptotic performance of the UML and CML estimators has been analyzed by Stoica and Nehorai ([SN89a]), [SN90a], and [SN90b]). Their approach to the UML (AML) estimator utilizes a series expansion around the correct parameter value and standard statistical theory of ML estimators (e.g., [KS61]). They show that the AML estimator is consistent and that the variance of each component of ψ approaches the CRB. They reference an earlier paper by Ottersten and Ljung [OL89] that gives a direct derivation.

The CML estimator does not obey the regularity conditions of ML estimator theory because the numbers of parameters approaches infinity as $K \rightarrow \infty$. Stoica and Nehorai [SN89a] show that the CML estimator is not consistent as $K \rightarrow \infty$ and that it is not efficient. For a single signal, they

show that

$$\frac{\sigma_{CML}^2}{\sigma_{CRB}^2} = 1 + \frac{1}{ASNR}. \quad (8.372)$$

In most applications, in order for the estimator to be above threshold, the $ASNR$ is such that the second term in (8.372) is very small. We saw this behavior in our simulation examples where, on the scale that we were plotting the MSE, the RMSE of the CML estimator appeared to approach the CRB bound.

Other references that discuss the asymptotic behavior of ML estimates include Sandkühler and Böhme [SB87] and Benitz [Ben93]. We discuss asymptotic performance of other estimators in Section 9.5.

Note that we are frequently interested in the performance of the estimators for a fixed K as the $ASNR$ goes to infinity. In many cases, the behavior is similar but a different analysis is required.

Asymptotic analyses play an important role in parameter estimation problems, and we quote results at various points in Chapters 8 and 9. It is important to remember two points when using the results of asymptotic analyses:

- (i) Without simulations, it is often difficult to determine the values of SNR and K where the asymptotic results are valid.
- (ii) It is desireable for an estimator to be consistent and efficient. However, we will encounter a number of useful estimators that are “almost consistent” or “almost efficient.” If the bias or excess variance above the CRB is very small, the performance may be satisfactory. In many cases, the error due to the model mismatch will be larger than these terms.

8.5.5 Wideband Signals

A detailed discussion of the extension of ML techniques to wideband signals is contained in Doron and Weiss [DW92], [DWM93]. Our discussion follows [DW92]. We consider the stochastic ML estimate first. We use the frequency-domain snapshot model,

$$\mathbf{X}_k(m\omega_0) = \mathbf{V}(m\omega_0, \psi) \mathbf{F}_k(m\omega_0) + \mathbf{N}_k(m\omega_0), \quad \begin{matrix} m = 1, 2, \dots, M \\ k = 1, 2, \dots, K \end{matrix} \quad (8.373)$$

There are three noise cases of interest:

1. $\eta_m \triangleq \sigma_w^2(m\omega_0)$, $m = 1, 2, \dots, M$, is known.

2. $\eta_m \stackrel{\Delta}{=} \sigma_w^2(m\omega_0)$, $m = 1, 2, \dots, M$, is unknown.
3. $\eta \stackrel{\Delta}{=} \sigma_w^2(m\omega_0) = \sigma_w^2$, $m = 1, 2, \dots, M$, and σ_w^2 is unknown.

We first consider the wideband version of the asymptotic ML estimate developed in Section 8.5.1 (see (8.310) and (8.315)).

The compressed likelihood functions follow in a manner analogous to the narrowband case. For case 1,

$$L_1(\psi) = \sum_{m=1}^M \left\{ -\frac{1}{\eta_m} \text{tr} [\mathbf{P}_{\mathbf{V}_m}^\perp \mathbf{C}_m] - \ln \det [\mathbf{P}_{\mathbf{V}_m} \mathbf{C}_m \mathbf{P}_{\mathbf{V}_m} + \eta_m \mathbf{P}_{\mathbf{V}_m}^\perp] \right\}, \quad (8.374)$$

where

$$\mathbf{C}_m \stackrel{\Delta}{=} \frac{1}{K} \sum_{k=1}^K \mathbf{X}_k(m\omega_0) \mathbf{X}_k^H(m\omega_0), \quad (8.375)$$

is the sample covariance matrix at $m\omega_0$. Then,

$$\begin{aligned} \hat{\psi}_{aml} = \arg \max_{\psi} & \sum_{m=1}^M \left\{ -\ln \det [\mathbf{P}_{\mathbf{V}_m} \mathbf{C}_m \mathbf{P}_{\mathbf{V}_m} + \eta_m \mathbf{P}_{\mathbf{V}_m}^\perp] \right. \\ & \left. - \frac{1}{\eta_m} \text{tr} [\mathbf{P}_{\mathbf{V}_m}^\perp \mathbf{C}_m] \right\}. \end{aligned} \quad (8.376)$$

Note that \mathbf{V}_m , the array manifold at $\omega = m\omega_0$ is a function of m and therefore $\mathbf{P}_{\mathbf{V}_m}$ is a function of m .

For case 2,

$$\hat{\eta}_m = \frac{\text{tr} [\mathbf{P}_{\mathbf{V}_m}^\perp \mathbf{C}_m]}{N - D}, \quad (8.377)$$

and

$$L_2(\psi) = -H_S - (N - D) \sum_{m=1}^M \ln \left\{ \text{tr} [\mathbf{P}_{\mathbf{V}_m}^\perp \mathbf{C}_m] \right\}, \quad (8.378)$$

where

$$H_S = \sum_{m=1}^M \ln \det [\mathbf{P}_{\mathbf{V}_m} \mathbf{C}_m \mathbf{P}_{\mathbf{V}_m} + \mathbf{P}_{\mathbf{V}_m}^\perp]. \quad (8.379)$$

For case 3,

$$\hat{\eta} = \frac{\sum_{m=1}^M \text{tr} [\mathbf{P}_{\mathbf{V}_m}^\perp \mathbf{C}_m]}{(N - D)M}, \quad (8.380)$$

and

$$L_3(\psi) = -H_S - M(N-D) \ln \sum_{m=1}^M \text{tr} [\mathbf{P}_{\mathbf{V}_m}^\perp \mathbf{C}_m]. \quad (8.381)$$

The likelihood functions for the CML estimator follow in a similar manner.

For case 1 (known noise variance case),

$$L_1^C(\psi) = \sum_{m=1}^M \frac{1}{\eta_m} \text{tr} [\mathbf{P}_{\mathbf{V}_m} \mathbf{C}_m], \quad (8.382)$$

and

$$\hat{\psi}_{cml} = \arg \max_{\psi} L_1^C(\psi). \quad (8.383)$$

For case 2,

$$L_2^C(\psi) = - \sum_{m=1}^M \ln (\text{tr} [\mathbf{P}_{\mathbf{V}_m}^\perp \mathbf{C}_m]), \quad (8.384)$$

where we have dropped unnecessary constants. Thus,

$$\hat{\psi}_{cml} = \arg \max_{\psi} \left\{ \sum_{m=1}^M \ln (\text{tr} [\mathbf{P}_{\mathbf{V}_m}^\perp \mathbf{C}_m]) \right\}. \quad (8.385)$$

For case 3,

$$L_3^C(\psi) = \sum_{m=1}^M \text{tr} [\mathbf{P}_{\mathbf{V}_m} \mathbf{C}_m]. \quad (8.386)$$

Thus,

$$\hat{\psi}_{cml} = \arg \max_{\psi} \sum_{m=1}^M \text{tr} [\mathbf{P}_{\mathbf{V}_m} \mathbf{C}_m]. \quad (8.387)$$

The noise estimate is

$$\hat{\eta}_n = \left. \frac{\sum_{m=1}^M \text{tr} [\mathbf{P}_{\mathbf{V}_m}^\perp \mathbf{C}_m]}{MN} \right|_{\psi=\psi_{cml}^{(3)}}. \quad (8.388)$$

The extension of the CRB to the wideband case follows in a straightforward manner. The other ML techniques that were developed for the narrowband case can also be extended to the wideband case.

Böhme and his colleagues have published a number of papers on wideband ML estimation (e.g., [Boh89], [KDB93], [KMB92], and [KB93]). We revisit the wideband case briefly in Section 9.10 and give a few more references. However, we do not study the case in a detailed manner.

8.5.6 Summary

In this section, we have studied ML estimators in detail. We have found the estimates using a grid search over a D -dimensional u -space followed by a local minimization. Due to the computational complexity, this search approach is generally not used in practice. In the next two sections, we develop computational algorithms that provide good results with reduced computational complexity.

8.6 Computational Algorithms

In this section, we discuss computational techniques for finding stochastic maximum likelihood estimators and CML estimators.

In Section 8.6.1, we discuss several optimization techniques. In Section 8.6.2, we develop the alternating projection (AP) and alternating maximization algorithm. In Section 8.6.3, we develop the expectation-maximization algorithm. In Section 8.6.4, we summarize our results and discuss other techniques that are available.

8.6.1 Optimization Techniques

In this section, we give a brief discussion of search techniques that can be used to solve the minimization problem,

$$\hat{\psi} = \arg \min_{\psi} [F(\psi)]. \quad (8.389)$$

We are interested in three expressions for $F(\psi)$ corresponding to the AML, CML, and WSF estimators.

From (8.315),

$$F_{aml} = \det \left[\mathbf{P}_V \mathbf{C}_x \mathbf{P}_V + \frac{\text{tr} [\mathbf{P}_V^\perp \mathbf{C}_x]}{N - D} \right]. \quad (8.390)$$

From Table 8.4,

$$F_{cml} = \text{tr} [\mathbf{P}_V^\perp \mathbf{C}_x], \quad (8.391)$$

and

$$F_{wsf} = \text{tr} \left[\mathbf{P}_V^\perp \hat{\mathbf{U}}_S \hat{\tilde{\Lambda}}^2 \hat{\Lambda}_S^{-1} \hat{\mathbf{U}}_S^H \right]. \quad (8.392)$$

Most of the techniques that we discuss in this section originated in the nonlinear optimization area and are adapted to solve our estimation problems. The reader needs to explore some the optimization sources, such as Dennis and Schnabel [DS96], Gill et al. [GMW81], or Nash and Sofer [NS96], to get a comprehensive discussion. These sources have been utilized for ML estimation by Ottersten et al. [OVSN93], and several of our results are from that reference.

We utilize a search technique to find the value of ψ that minimizes $F(\psi)$. The basic idea is to model $F(\psi)$ as a quadratic function in the vicinity of the minimum. We select a starting point $\hat{\psi}_0$. We then find a descent direction that will cause $F(\psi)$ to decrease and calculate a step size to determine how far to move in the descent direction. We discuss techniques for choosing $\hat{\psi}_0$ in Section 8.6.2.

At the first iteration,

$$\hat{\psi}_1 = \hat{\psi}_0 + \mathbf{a}_0 \left[\nabla_{\psi} F(\psi) \right] \Big|_{\psi=\hat{\psi}_0}. \quad (8.393)$$

In (8.393), $\hat{\psi}_0$ and $\hat{\psi}_1$ are $D \times 1$ vectors, $\nabla_{\psi} F(\psi)$ is the $D \times 1$ gradient vector,

$$\nabla_{\psi} F(\psi) \triangleq \left[\frac{\partial F(\psi)}{\partial \psi_1} \dots \frac{\partial F(\psi)}{\partial \psi_D} \right]^T \triangleq \mathbf{F}'(\psi), \quad (8.394)$$

and \mathbf{a}_0 is a $D \times D$ matrix. The gradient vector will equal $\mathbf{0}$ at a stationary point and

$$\nabla_{\psi} F(\psi) = \mathbf{0}, \quad (8.395)$$

is a necessary but not sufficient condition for a minimum. The matrix \mathbf{a}_0 determines both the descent direction and the step size. We will separate \mathbf{a}_0 into two terms in subsequent equations. At the k th iteration,

$$\hat{\psi}_{k+1} = \hat{\psi}_k + \mathbf{a}_k \left\{ \left[\nabla_{\psi} F(\psi) \right] \Big|_{\psi=\hat{\psi}_k} \right\}. \quad (8.396)$$

We continue the iteration until we satisfy a stopping rule of the form,

$$|F(\hat{\psi}_{k+1}) - F(\hat{\psi}_k)| < \delta_1, \quad (8.397)$$

or

$$\|\hat{\psi}_{k+1} - \hat{\psi}_k\| < \delta_2. \quad (8.398)$$

is satisfied.

In the classical Newton method,

$$\mathbf{a}_k = -[\nabla^2 F(\psi)]_{\psi=\hat{\psi}_k}^{-1}, \quad (8.399)$$

where $\nabla^2 F(\psi)$ is the Hessian matrix.

$$\mathbf{H}(\psi_k) \triangleq [\nabla^2 F(\psi)]_{ij} \triangleq \frac{\partial^2 F(\psi)}{\partial \psi_i \partial \psi_j}. \quad (8.400)$$

In order for $-\mathbf{H}^{-1}(\psi_k)\mathbf{F}'(\psi_k)$ to be a descent direction, $\mathbf{H}(\psi_k)$ must be positive definite. This may not be true if we are too far from the minimum, because $F(\psi)$ may not be quadratic. There are different techniques for modifying \mathbf{H} to make it positive definite. One approach is to use DL,

$$\tilde{\mathbf{H}}_k = \mathbf{H}_k + \sigma_k^2 \mathbf{I}, \quad (8.401)$$

where σ_k^2 is chosen at each step to make $\tilde{\mathbf{H}}_k$ both positive definite and well-conditioned. This technique is discussed in Chapter 5 of [DS96].

In practice, we want to adjust the step size, so we use a damped Newton algorithm

$$\hat{\psi}_{k+1} = \hat{\psi}_k - \mu_k \mathbf{H}_k^{-1} \mathbf{F}'_k, \quad (8.402)$$

where \mathbf{F}'_k is the gradient defined in (8.394) (e.g., [GMW81], [DS83], [DS96]).

In order to choose the step length μ_k , we choose a $\mu < 1$ and let

$$\mu_k = (\mu)^i, \quad i \geq 0. \quad (8.403)$$

At each step in the iteration, we try successive values of i starting at $i = 0$ and use the smallest i that causes an adequate decrease in $F(\psi_k)$. For example, if $\mu = 0.5$, we would try

$$\mu_k = 1, \frac{1}{2}, \left(\frac{1}{2}\right)^2, \left(\frac{1}{2}\right)^3, \dots \quad (8.404)$$

until we obtain a satisfactory decrease in $F(\psi_k)$. Quadratic convergence is obtained if the step length converges to unity.

The advantage of the Newton algorithm is that one can show quadratic convergence. However, in order to implement the Newton method, we must compute a matrix of second derivatives and invert a $D \times D$ matrix. In most applications, we try to find a computationally simpler approximation to the Newton method that still converges at an adequate rate.

One approach is to replace the Hessian by an approximate Hessian that has the same form as the asymptotic form of the Hessian matrix. This approach is referred to in the statistical literature as the **scoring** method. Ottersten [OVSN93] have derived the expression and we use their derivation.

Using (8.296), we can rewrite (8.390) as

$$F_{aml}(\psi) = \ln \det \left[\mathbf{V} \hat{\mathbf{S}}_f(\psi) \mathbf{V}^H + \hat{\sigma}_w^2 \mathbf{I} \right], \quad (8.405)$$

where, from (8.313),

$$\hat{\sigma}_w^2(\psi) = \frac{\text{tr} \left[\mathbf{P}_{\mathbf{V}}^{\perp} \mathbf{C}_{\mathbf{x}} \right]}{N - D}. \quad (8.406)$$

We included the \ln in (8.405) to be consistent with the derivation in [OVSN93]. Then, using $|\mathbf{I} + \mathbf{AB}| = |\mathbf{I} + \mathbf{BA}|$, we can write (8.405) as

$$\begin{aligned} F_{aml}(\psi) &= \ln \left\{ \hat{\sigma}_w^{2N}(\psi) \left| \hat{\sigma}_w^{-2}(\psi) \hat{\mathbf{S}}_f(\psi) \mathbf{V}^H \mathbf{V} + \mathbf{I} \right| \right\} \\ &= \ln \left\{ \hat{\sigma}_w^{2N}(\psi) \left| \hat{\sigma}_w^{-2}(\psi) \left[\mathbf{V}^{\dagger} \left[\mathbf{C}_{\mathbf{x}} - \hat{\sigma}_w^2(\psi) \mathbf{I} \right] \mathbf{V} + \mathbf{I} \right] \right| \right\} \\ &= \ln \left\{ \hat{\sigma}_w^{2N}(\psi) \left| \hat{\sigma}_w^{-2}(\psi) \left[\mathbf{V}^{\dagger} \mathbf{C}_{\mathbf{x}} \mathbf{V} \right] \right| \right\} \\ &= \ln \left\{ \hat{\sigma}_w^{2(N-D)}(\psi) \left| \mathbf{V}^{\dagger} \mathbf{C}_{\mathbf{x}} \mathbf{V} \right| \right\}. \end{aligned} \quad (8.407)$$

Now define a matrix

$$\mathbf{G} \triangleq \mathbf{V} \left[\left[\mathbf{V}^H \mathbf{C}_{\mathbf{x}} \mathbf{V} \right]^{-1} - \hat{\sigma}_w^{-2}(\psi) \left[\mathbf{V}^H \mathbf{V} \right]^{-1} \right]. \quad (8.408)$$

The gradient matrix is

$$\mathbf{F}'_{aml}(\psi) = 2 \text{Re}(\text{DIAG} \left[\mathbf{G}^H \mathbf{C}_{\mathbf{x}} \mathbf{P}_{\mathbf{V}}^{\perp} \mathbf{D} \right]), \quad (8.409)$$

where \mathbf{D} is defined in (8.97) and $\text{DIAG}[\mathbf{A}]$ is a column matrix formed from the diagonal elements of \mathbf{A} .

Because the AML estimate is efficient, the approximate Hessian is derived from the CRB in (8.102),

$$\mathbf{H}_{aml,a}(\psi) = \frac{2}{\sigma_w^2} \text{Re} \left\{ \left[\mathbf{D}^H \mathbf{P}_{\mathbf{V}}^{\perp} \mathbf{D} \right] \odot \left[\mathbf{S}_f \mathbf{V}^H \mathbf{S}_{\mathbf{x}}^{-1} \mathbf{V} \mathbf{S}_f \right]^T \right\}. \quad (8.410)$$

We replace the ensemble statistics \mathbf{S}_f and σ_w^2 by their AML estimates $\hat{\mathbf{S}}_f(\psi)$ and $\hat{\sigma}_w^2(\psi)$ and use (8.299) to replace $\mathbf{S}_{\mathbf{x}}$ by

$$\hat{\mathbf{S}}_{\mathbf{x}}(\psi) = \mathbf{V} \hat{\mathbf{S}}_f(\psi) \mathbf{V}^H + \hat{\sigma}_w^2(\psi) \mathbf{I}. \quad (8.411)$$

Then, the second matrix in (8.410) can be written as

$$\hat{\mathbf{S}}_{\mathbf{f}}(\psi) \mathbf{V}^H \left[\mathbf{V} \hat{\mathbf{S}}_{\mathbf{f}}(\psi) \mathbf{V}^H + \hat{\sigma}_w^2(\psi) \mathbf{I} \right]^{-1} \mathbf{V} \hat{\mathbf{S}}_{\mathbf{f}}(\psi) = \hat{\sigma}_w^4(\psi) \mathbf{G}^H \mathbf{C}_{\mathbf{x}} \mathbf{G}. \quad (8.412)$$

Using (8.412) in (8.410) gives the approximate Hessian matrix,

$$\mathbf{H}_{aml,a}(\psi) = 2\hat{\sigma}_w^2(\psi) \operatorname{Re} \left\{ \left[\mathbf{D}^H \mathbf{P}_{\mathbf{V}}^{\perp} \mathbf{D} \right] \odot \left[\mathbf{G}^H \mathbf{C}_{\mathbf{x}} \mathbf{G} \right]^T \right\}. \quad (8.413)$$

The terms in (8.402) are now defined. At each step, we compute $\hat{\sigma}_w^2(\psi_k)$ using (8.406) and use the result in (8.408), (8.409), and (8.413) to compute $\mathbf{F}'_{aml}(\psi_k)$ and $\mathbf{H}_{aml,a}(\psi_k)$. We substitute the results in (8.402) and use (8.403) and (8.404) to compute the step size. We then compute $\hat{\psi}_{k+1}$. We continue the iteration until the stopping rule is satisfied. We choose $\hat{\psi}_0$ using the AP or AM technique, which we derive in Section 8.6.2.

The approximate Hessians for the CML estimator and the WSF_{ao} estimator are also derived in [OVSN93]. These authors also consider several examples and discuss efficient computational techniques. The reader is referred to this reference for a more detailed discussion.

In the next section, we discuss a different approach to quasi-Newton methods.

8.6.1.1 Quasi-Newton methods

There are a number of other algorithms that attempt to retain some of the good properties of the Newton algorithm, but with a reduced computational cost. In this section, we present one of these algorithms that is effective in array processing problems.

The starting point for the methods in this section is the Newton algorithm.

Once again, we choose an initial estimate $\hat{\psi}_0$. The iteration is

$$\hat{\psi}_{k+1} = \hat{\psi}_k - \mu_k \mathbf{H}_k^{-1}(\psi_k) \mathbf{F}'(\psi). \quad (8.414)$$

We use a modified form of the Newton algorithm which is referred to as the **modified variable projection algorithm** (e.g., [Kau75], [RW80], [GP73]). The algorithm was applied to the array processing problem by Viberg et al. [VOK91], and our discussion follows that reference. We consider the CML and WSF_{ao} estimators. We write $F(\psi)$ as

$$F(\psi) = \operatorname{tr} \left\{ \mathbf{P}_{\mathbf{V}}^{\perp}(\psi) \mathbf{M} \mathbf{M}^H \right\} = \left\| \mathbf{P}_{\mathbf{V}}^{\perp}(\psi) \mathbf{M} \right\|_F^2. \quad (8.415)$$

The matrix \mathbf{M} represents the data. In the case of CML,

$$\mathbf{M}\mathbf{M}^H = \mathbf{C}_x, \quad (8.416)$$

so \mathbf{M} equals $\tilde{\mathbf{X}}_K$.

In the case of weighted subspace fitting,

$$\mathbf{M} = \widehat{\mathbf{U}}_s \mathbf{W}^{\frac{1}{2}} \quad (8.417)$$

(see (8.363)).

We now define a vector \mathbf{r} by stacking the columns of $\mathbf{P}_{\mathbf{V}}^{\perp}(\psi)\mathbf{M}$,

$$\mathbf{r} = \text{vec} \left(\mathbf{P}_{\mathbf{V}}^{\perp}(\psi)\mathbf{M} \right). \quad (8.418)$$

Using (8.418) in (8.415) gives

$$F(\psi) = |\mathbf{r}|^2. \quad (8.419)$$

The gradient of $F(\psi)$ with respect to ψ_i is

$$\frac{\partial}{\partial \psi_i} F = 2\text{Re} \left\{ \left(\frac{\partial \mathbf{r}}{\partial \psi_i} \right)^H \mathbf{r} \right\} = 2\text{Re} \left\{ \mathbf{r}_i^H \mathbf{r} \right\}, \quad (8.420)$$

where

$$\mathbf{r}_i \triangleq \frac{\partial \mathbf{r}}{\partial \psi_i} = \frac{\partial \mathbf{P}_{\mathbf{V}}^{\perp}(\psi)}{\partial \psi_i} \mathbf{M}. \quad (8.421)$$

The derivative of the projection matrix is,

$$\frac{\partial}{\partial \psi_i} \mathbf{P}_{\mathbf{V}}^{\perp} = -\frac{\partial}{\partial \psi_i} \mathbf{P}_{\mathbf{V}} = -\mathbf{P}_{\mathbf{V}}^{\perp} \mathbf{V}_i \mathbf{V}^{\dagger} - \left(-\mathbf{P}_{\mathbf{V}}^{\perp} \mathbf{V}_i \mathbf{V}^{\dagger} \right)^H, \quad (8.422)$$

where

$$\mathbf{V}_i \triangleq \frac{\partial \mathbf{V}}{\partial \psi_i}. \quad (8.423)$$

Using (8.422) in (8.420) gives

$$\mathbf{F}'(\psi) = -2\text{Re} \left\{ \text{DIAG} \left[\mathbf{V}^{\dagger} \mathbf{M} \mathbf{M}^H \mathbf{P}_{\mathbf{V}}^{\perp} \mathbf{D}(\psi) \right] \right\}, \quad (8.424)$$

where $\mathbf{D}(\psi)$ is defined in (8.97) and $\text{DIAG}[\mathbf{A}]$ is defined after (8.409).

The next step is to derive the Hessian,

$$\frac{\partial^2}{\partial \psi_i \partial \psi_j} F = 2\text{Re} \left\{ \mathbf{r}_i^H \mathbf{r}_j + \mathbf{r}_{ij}^H \mathbf{r} \right\}. \quad (8.425)$$

The Gaussian modification of the Newton method assumes that the second term in (8.425) is small and approximates the Hessian by

$$\mathbf{H}_G = 2\operatorname{Re} \left\{ \mathbf{r}_i^H \mathbf{r}_j \right\}. \quad (8.426)$$

Discarding the second derivative guarantees that (8.414) is a descent method because \mathbf{H}_G is nonnegative definite. The resulting algorithm is the variable projection algorithm of Golub and Pereyra [GP73]. Using (8.418) in (8.422) and observing that $\mathbf{V}^\dagger \mathbf{P}_V^\perp = \mathbf{0}$, we obtain

$$\mathbf{r}_i^H \mathbf{r}_j = \operatorname{tr} \left\{ \left[[\mathbf{V}^\dagger]^H \mathbf{V}_i^H \mathbf{P}_V^\perp \mathbf{V}_j \mathbf{V}^\dagger + \mathbf{P}_V^\perp \mathbf{V}_i \mathbf{V}^\dagger [\mathbf{V}^\dagger]^H \mathbf{V}_j^H \mathbf{P}_V^\perp \right] \mathbf{M} \mathbf{M}^H \right\}. \quad (8.427)$$

Kaufman [Kau75] modifies the Gauss-Newton algorithm by deleting the second term in (8.427). With this modification,

$$H_{ij} = 2\operatorname{Re} \left\{ \operatorname{tr} \left[[\mathbf{V}^\dagger]^H \mathbf{V}_i^H \mathbf{P}_V^\perp \mathbf{V}_j \mathbf{V}^\dagger \mathbf{M} \mathbf{M}^H \right] \right\}, \quad (8.428)$$

which can be expressed in matrix notation as,

$$\mathbf{H} = 2\operatorname{Re} \left\{ [\mathbf{D}^H \mathbf{P}_V^\perp \mathbf{D}] \odot \left[\mathbf{V}^\dagger \mathbf{M} \mathbf{M}^H [\mathbf{V}^\dagger]^H \right]^T \right\}. \quad (8.429)$$

The iteration algorithm is defined by (8.414), (8.424), and (8.429). It is pointed out in [OVSN93] that, for the WSF case, this algorithm is the same as the WSF version of the asymptotic Hessian (scoring) algorithm discussed in Section 8.6.1.1. (We did not derive the WSF version.)

In order to obtain convergence to a global minimum we must initialize the algorithm appropriately. In Section 8.7.2.1, we develop the alternating projection algorithm that we use to initialize the modified Gauss-Newton algorithm. In [OVSN93], examples are given to show the effect of initialization accuracy.

8.6.1.2 Summary

In this section, we have provided a brief discussion of gradient techniques for solving ML estimation problems. The reader is referred to the various references, particularly [OVSN93] for further discussion.

8.6.2 Alternating Maximization Algorithms

In this section, we develop techniques that are sometimes referred to in the optimization literature as the relaxation method. The basic idea is straightforward. We have a function of D variables, $F(\psi)$. We want to find the value of ψ that maximizes $F(\psi)$. We accomplish this by a sequence of 1-D maximization problems.

The technique was applied to the array processing problem by Ziskind and Wax [ZW88b]. For the CML cases, they utilize a property of projection matrices to simplify the algorithm, and the resulting algorithm is referred to in the array processing literature as the alternating projection (AP) algorithm. For the stochastic ML case, the algorithm is referred to as the alternating maximization (AM) algorithm. Both algorithms use the same idea.

8.6.2.1 Alternating projection algorithm

The AP algorithm was used by Ziskind and Wax [ZW88b] in order to replace the multidimensional maximization problem by a sequence of 1-D maximization problems. Our discussion follows [ZW88b].

In this subsection we consider the CML estimate. From (8.345), we must maximize

$$\hat{\psi}_{cml} = \arg \max_{\psi} \{ \text{tr} [\mathbf{P}_V \mathbf{C}_x] \}. \quad (8.430)$$

The basic idea is straightforward. It is an iterative technique in which, at each step of the iteration, we hold $D - 1$ parameter values constant and maximize over a single parameter.

Thus, the value of ψ_i at the $(k + 1)$ -th iteration is obtained by the following one-dimensional maximization problem,

$$\hat{\psi}_i^{(k+1)} = \arg \max_{\psi_i} \left\{ \text{tr} \left[\mathbf{P} \left[\mathbf{V}(\hat{\psi}_{(i)}^{(k)}, \mathbf{v}(\psi_i)) \right] \mathbf{C}_x \right] \right\}, \quad (8.431)$$

where $\hat{\psi}_{(1)}^{(k)}$ is the value of the estimated vector parameter ($D - 1 \times 1$) at the k th iteration with the first component removed,

$$\hat{\psi}_{(1)}^{(k)} = \left[\begin{array}{c} \hat{\psi}_2^{(k)} \\ \vdots \\ \hat{\psi}_3^{(k)} \\ \vdots \\ \hat{\psi}_D^{(k)} \end{array} \right]. \quad (8.432)$$

The algorithm maximizes the ψ_i in order, starting at $i = 1$. For $i > 1$, $\hat{\psi}_{(i)}^{(k)}$ is defined as

$$\hat{\psi}_{(i)}^{(k)} = \left[\begin{array}{c|c|c|c|c} \hat{\psi}_1^{(k+1)} & \cdots & \hat{\psi}_{i-1}^{(k+1)} & \hat{\psi}_{i+1}^{(k)} & \cdots & \hat{\psi}_D^{(k)} \end{array} \right]. \quad (8.433)$$

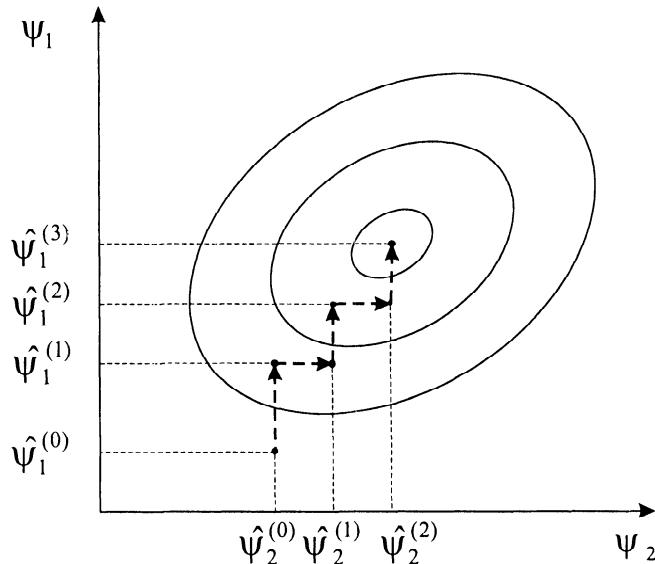


Figure 8.35 AP Algorithm: successive iterations.

The estimates of ψ_r for $r < i$ are the results of the $(k+1)$ th iteration, which was done previously.

Note that we have separated the space of the projection matrix into a $N \times (D-1)$ fixed component and a $N \times 1$ component that is allowed to vary.

The graphical behavior of the algorithm is shown in Figure 8.35. The algorithm moves to the peak in steps parallel to the axes. Since the value of $L(\psi)$ is maximized at each step, convergence to a local maximum is guaranteed. The initial condition is key to convergence to the global maximum.

Ziskind and Wax [ZW88b] use the following initialization procedure. First solve the problem for a single source ψ_1 . Thus

$$\hat{\psi}_1^{(0)} = \arg \max_{\psi_1} \left\{ \text{tr} \left[\mathbf{P}_{\mathbf{v}(\psi_1)} \mathbf{C}_x \right] \right\}. \quad (8.434)$$

This is equivalent to assuming there is a single source. Then solve for the second source assuming the first source is at $\hat{\psi}_1^{(0)}$. Thus,

$$\hat{\psi}_2^{(0)} = \arg \max_{\psi_2} \left\{ \text{tr} \left[\mathbf{P}_{[\mathbf{v}(\hat{\psi}_1^{(0)})], \mathbf{v}(\psi_2)} \mathbf{C}_x \right] \right\}. \quad (8.435)$$

We continue this procedure until we obtain the D initial estimated values,

$\hat{\psi}_1^{(0)}, \dots, \hat{\psi}_D^{(0)}$. At each step, we assume that all preceding initial values are known. With these initial values, we then carry out the iteration in (8.431).

There is a property of projection matrices referred to as the projection matrix update formula that simplifies the computation at each iteration.

Property

Let \mathbf{B} and \mathbf{C} be arbitrary matrices with the same number of rows and define the composite matrix $[\mathbf{B} : \mathbf{C}]$. If \mathbf{B} and \mathbf{C} were orthogonal, then

$$\mathbf{P}_{\mathbf{B}, \mathbf{C}} = \mathbf{P}_{\mathbf{B}} + \mathbf{P}_{\mathbf{C}}. \quad (8.436)$$

However, since they are not orthogonal, we define $\mathbf{P}_{\mathbf{C}(\mathbf{B})}$, where $\mathbf{C}(\mathbf{B})$ is the residual of the columns of \mathbf{C} , when \mathbf{C} is projected on \mathbf{B} ,

$$\mathbf{C}(\mathbf{B}) = [\mathbf{I} - \mathbf{P}_{\mathbf{B}}] \mathbf{C} = \mathbf{P}_{\mathbf{B}}^\perp \mathbf{C}. \quad (8.437)$$

Then, the projection matrix onto the column space of $[\mathbf{B} : \mathbf{C}]$ is

$$\mathbf{P}_{\mathbf{B}, \mathbf{C}} = \mathbf{P}_{\mathbf{B}} + \mathbf{P}_{\mathbf{C}(\mathbf{B})}. \quad (8.438)$$

To utilize this result, we let

$$\mathbf{B} = \mathbf{V}(\hat{\psi}_{(i)}^{(k)}), \quad (8.439)$$

and

$$\mathbf{C} = \mathbf{v}(\psi_i). \quad (8.440)$$

Using this result in (8.431), we can write

$$\mathbf{P}_{\mathbf{V}(\hat{\psi}_{(i)}^{(k)}), \mathbf{v}(\psi_i)} = \mathbf{P}_{\mathbf{V}(\hat{\psi}_{(i)}^{(k)})} + \mathbf{P}_{\mathbf{v}(\psi_i) \left(\mathbf{V}(\hat{\psi}_{(i)}^{(k)}) \right)}. \quad (8.441)$$

The vector in the subscript in the second term is generated by constructing:

(i) $\mathbf{P}_{\mathbf{V}(\hat{\psi}_{(i)}^{(k)})}$ (corresponds to $\mathbf{P}_{\mathbf{B}}$).

(ii) $\left[\mathbf{P}_{\mathbf{V}(\hat{\psi}_{(i)}^{(k)})}^\perp \right] \mathbf{v}(\psi_i)$ (corresponds to $\mathbf{P}_{\mathbf{B}}^\perp \mathbf{C}$).

(iii) $\mathbf{P}_{\mathbf{v}(\psi_i)(\mathbf{V}(\hat{\psi}_{(i)}^{(k)}))}$ (corresponds to $\mathbf{P}_{\mathbf{C}(\mathbf{B})}$).

Since the first term in (8.441) is not a function of ψ_i it can be dropped. Thus, (8.431) reduces to

$$\hat{\psi}_i^{(k+1)} = \arg \max_{\psi_i} \left\{ \text{tr} \left[\mathbf{P}_{\mathbf{v}(\psi_i)(\mathbf{V}(\hat{\psi}_{(i)}^{(k)}))} \mathbf{C}_{\mathbf{x}} \right] \right\}. \quad (8.442)$$

Now define a normalized vector

$$\mathbf{b}(\psi_i, \hat{\psi}_{(i)}^{(k)}) \triangleq \frac{\mathbf{C}(\mathbf{B})}{\|\mathbf{C}(\mathbf{B})\|} = \frac{\mathbf{P}_\mathbf{B}^\perp \mathbf{C}}{\|\mathbf{P}_\mathbf{B}^\perp \mathbf{C}\|}, \quad (8.443)$$

or

$$\mathbf{b}(\psi_i, \hat{\psi}_{(i)}^{(k)}) = \frac{\begin{bmatrix} \mathbf{P}^\perp \\ \mathbf{V}(\hat{\psi}_{(i)}^{(k)}) \end{bmatrix} \mathbf{v}(\psi_i)}{\left\| \begin{bmatrix} \mathbf{P}^\perp \\ \mathbf{V}(\hat{\psi}_{(i)}^{(k)}) \end{bmatrix} \mathbf{v}(\psi_i) \right\|}. \quad (8.444)$$

Then,

$$\mathbf{P}_{\mathbf{v}(\psi_i)(\mathbf{V}(\hat{\psi}_{(i)}^{(k)}))} = \mathbf{b}(\psi_i, \hat{\psi}_{(i)}^{(k)}) \mathbf{b}^H(\psi_i, \hat{\psi}_{(i)}^{(k)}), \quad (8.445)$$

and (8.442) can be written as,

$$\hat{\psi}_i^{(k+1)} = \arg \max_{\psi_i} \left\{ \mathbf{b}^H(\psi_i, \hat{\psi}_{(i)}^{(k)}) \mathbf{C}_x \mathbf{b}(\psi_i, \hat{\psi}_{(i)}^{(k)}) \right\}. \quad (8.446)$$

We continue the iteration across i and k until

$$\left| \hat{\psi}_i^{(k+1)} - \hat{\psi}_i^{(k)} \right| \leq \delta, \quad i = 1, \dots, D, \quad (8.447)$$

where δ is a function of the desired accuracy.

The AP algorithm can be summarized:

- (i) Initialize the algorithm using the procedure in (8.434) and (8.435) to obtain $\hat{\psi}_1^{(0)}, \hat{\psi}_2^{(0)}, \dots, \hat{\psi}_D^{(0)}$.
- (ii) For $i = 1, \dots, D$, and $k = 1$, use (8.446) to obtain $\hat{\psi}_1^{(1)}, \hat{\psi}_2^{(1)}, \hat{\psi}_3^{(1)}, \dots, \hat{\psi}_D^{(1)}$.
- (iii) Iterate (8.446) for $k = 2, \dots$.
- (iv) Repeat until (8.447) is satisfied for all $i = 1, \dots, D$.

The issues of interest with respect to the AP algorithms are:

- (i) What are the conditions on signal geometry, SNR_i , and K that will cause the initial conditions to be such that the AP algorithm converges to a global maximum?
- (ii) What is the rate of convergence?
- (iii) Is the rate of convergence improved by using the AP algorithm for the first several iterations and then switching to a gradient procedure?

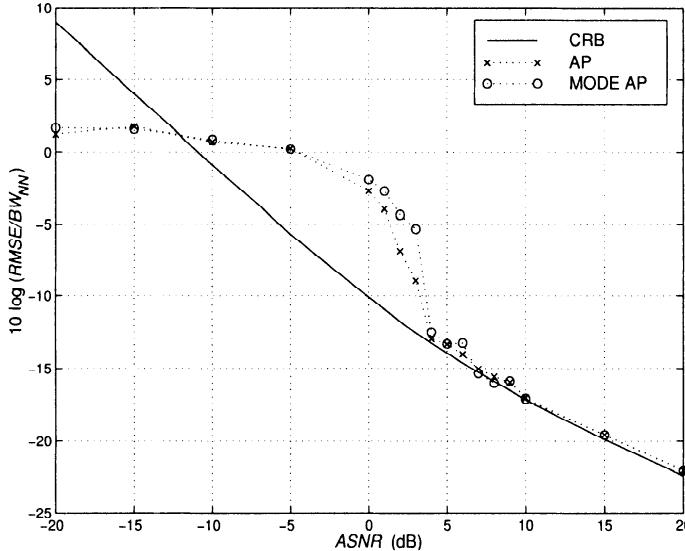


Figure 8.36 Normalized RMSE versus $ASNR$: $\Delta\psi = \Delta\psi_R$, $N = 10$, $K = 100$, $\rho = 0$.

We consider two of the test scenarios in Table 8.3 to illustrate the behavior of the AP algorithm.

Example 8.6.1 (continuation, Examples 8.5.2 and 8.5.9)

Consider the same model as in Examples 8.5.2 and 8.5.9. There are two equal-power uncorrelated signals impinging on a standard 10-element linear array. The signal separation is $\Delta\psi_R$. We use the AP algorithm to estimate the DOAs. We initialize the algorithm using (8.434) and (8.435).

The results are shown in Figure 8.36. Comparing the result to the CML and MODE curves in Figure 8.32, we see that the curves are the same.

Example 8.6.2 (continuation, Example 8.5.8)

Consider the same model as in Example 8.5.8. The result for $\phi_\rho = \pi/4$ is shown in Figure 8.37. We see that the curve is the same as the curve in Figure 8.34.

In all of the other test scenarios, the AP results are the same as the grid search results. There is a significant computational saving, so the AP algorithm is attractive. There are certain cases in which the initialization will cause the algorithm to converge to a local maximum rather than the global maximum, but we did not encounter this problem in our test scenarios.

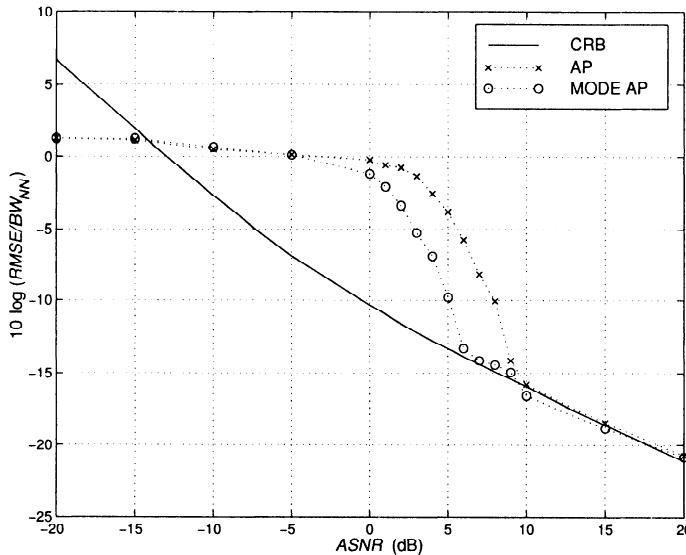


Figure 8.37 Normalized RMSE versus $ASNR$: $\Delta\psi = \Delta\psi_R$, $N = 10$, $K = 100$, $\rho = 0.95e^{j\pi/4}$.

8.6.2.2 Alternating maximization algorithm

In this section we consider the use of the AM technique to find the AML estimate given by (8.310). From (8.310), the function that we want to maximize is

$$\hat{\psi}_{aml} = \arg \max_{\psi} \left\{ -\ln \det \left[\mathbf{P}_V \mathbf{C}_x \mathbf{P}_V + \sigma_w^2 \mathbf{P}_V^\perp \right] - \frac{1}{\sigma_w^2} \text{tr} \left[\mathbf{P}_V^\perp \mathbf{C}_x \right] \right\}, \quad (8.448)$$

in the case of known noise variance and, from (8.315),

$$\hat{\psi}_{aml} = \arg \min_{\psi} \left\{ \det \left[\mathbf{P}_V \mathbf{C}_x \mathbf{P}_V + \frac{\text{tr} \left[\mathbf{P}_V^\perp \mathbf{C}_x \right] \mathbf{P}_V^\perp}{N - D} \right] \right\}, \quad (8.449)$$

in the case of unknown noise variance.

The vectors, $\hat{\psi}_{(1)}^{(k)}$ and $\hat{\psi}_{(i)}^{(k)}$ are defined in (8.432) and (8.433), respectively. We proceed in an iterative manner

$$\hat{\psi}_i^{(k+1)} = \arg \max_{\psi_i} \left\{ F \left(\psi : \left\{ \mathbf{V} \left(\hat{\psi}_{(i)}^{(k)} \right), \mathbf{v}(\psi_i) \right\} \right) \right\}, \quad (8.450)$$

where the term in the braces in (8.450) is the term inside the braces in (8.448) or (8.449) with $\mathbf{V}(\psi)$ replaced by $\left\{ \mathbf{V}\left(\hat{\psi}_{(i)}^{(k)}\right), \mathbf{v}(\psi_i) \right\}$ in the projection matrices. We initialize the algorithm in a manner similar to (8.434).

We implemented the AM algorithm for the scenarios in Table 8.3. We used the AP initialization technique in (8.434) and (8.435). In all cases, the results were the same as in the grid search AML algorithm.

8.6.3 Expectation Maximization Algorithm

In this section we develop the expectation maximization (EM) algorithm and apply it to the parameter estimation problem. The basis for the name will be clear when the algorithm is developed.

In Section 8.6.3.1, we derive a general form of the algorithm. In Section 8.6.3.2, we apply it to the CML estimation problem. In Section 8.6.3.3, we summarize our results.

8.6.3.1 EM Algorithm

The original derivation of the EM algorithm is due to Laird et al. [LDR77]. It has been applied to the array processing problem by Feder and Weinstein [FW88] and Miller and Fuhrmann [MF90]. A tutorial article by Moon [Moo96] discusses various other applications. Our discussion follows these references. We begin with a general discussion of the EM algorithm, then consider the case when the probability densities are Gaussian, and then apply the EM algorithm to the problem of interest.

The basic idea of the EM algorithm is straightforward. Recall from (8.2), that on the k th snapshot, we observe

$$\mathbf{X}_k = \mathbf{V}(\psi)\mathbf{F}_k + \mathbf{N}_k \quad k = 1, 2, \dots, K, \quad (8.451)$$

and want to find $\hat{\psi}_{ml}$. Now suppose that, instead of \mathbf{X}_k , we could observe

$$\mathbf{Y}_{ki} = \mathbf{v}(\theta_i)\mathbf{F}_{ki} + \mathbf{N}_{ki} \quad i = 1, 2, \dots, D. \quad (8.452)$$

Then we would have a set of D uncoupled 1-D maximization problems that are reasonably easy to solve. We refer to the transformation between the \mathbf{Y}_{ki} and \mathbf{X}_k as \mathbf{H} ,

$$\mathbf{X}_k = \mathbf{H}(\mathbf{Y}_{k1}, \mathbf{Y}_{k2}, \dots, \mathbf{Y}_{kD}), \quad (8.453)$$

and refer to the $\mathbf{Y}_{ki}, i = 1, \dots, D$ as the “complete data” and \mathbf{X}_k as the “incomplete data.”

The EM algorithm is a technique for inferring the complete data from the incomplete data and using this inferred data to find the ML estimate.

EM Algorithm¹⁵ We denote the observed (“incomplete”) data by the vector \mathbf{X} whose probability density, $p_{\mathbf{X}}(\mathbf{X} : \boldsymbol{\theta})$ ¹⁶ depends on the vector parameter $\boldsymbol{\theta}$. We denote the complete data by the vector \mathbf{Y} , which is related to \mathbf{X} by the many-to-one (non-invertible) transformation $\mathbf{H}(\cdot)$

$$\mathbf{X} = \mathbf{H}(\mathbf{Y}). \quad (8.454)$$

Note that the choice of \mathbf{Y} and therefore $\mathbf{H}(\cdot)$ is not unique, and one of the keys to successful application of the EM algorithm is an appropriate choice of \mathbf{Y} and $\mathbf{H}(\cdot)$.

We assume that the estimate of $\boldsymbol{\theta}$ at step n of the iteration is $\hat{\boldsymbol{\theta}}^{(n)}$. We denote the likelihood function of \mathbf{Y} as,¹⁷

$$L_{\mathbf{Y}}(\boldsymbol{\theta}) \triangleq \ln p_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y}) = \ln p_{\mathbf{Y}}(\mathbf{Y} : \boldsymbol{\theta}). \quad (8.455)$$

If \mathbf{Y} were available, we would maximize $L_{\mathbf{Y}}(\boldsymbol{\theta})$. As only \mathbf{X} is available, we find the expectation of $L_{\mathbf{Y}}(\boldsymbol{\theta})$, given that we have observed \mathbf{X} and that our current parameter estimate is $\hat{\boldsymbol{\theta}}^{(n)}$.

We define the resulting expectation as $U(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(n)})$:

$$U(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(n)}) \triangleq E \left\{ \ln p_{\mathbf{Y}}(\mathbf{Y} : \boldsymbol{\theta} | \mathbf{X} : \hat{\boldsymbol{\theta}}^{(n)}) \right\}. \quad (8.456)$$

This is the expectation step. Note that it is conditional expectation with respect to \mathbf{X} and $\hat{\boldsymbol{\theta}}^{(n)}$.

The next step is to maximize $U(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(n)})$ with respect to $\boldsymbol{\theta}$. The resulting estimate is $\hat{\boldsymbol{\theta}}^{(n+1)}$. Thus,

$$\hat{\boldsymbol{\theta}}^{(n+1)} = \arg \max_{\boldsymbol{\theta}} U(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(n)}). \quad (8.457)$$

The steps in (8.456) and (8.457) define the EM algorithm.

¹⁵This discussion follows [FW88] and [Moo96], which follow [LDR77].

¹⁶We use $p_{\mathbf{Y}}(\mathbf{Y} : \boldsymbol{\theta})$ instead of $p_{\mathbf{Y}|\boldsymbol{\theta}}(\mathbf{Y})$ to minimize the use of double subscripts in the derivation.

¹⁷We have used n to denote the steps in the iteration because we use k to denote the snapshot.

The algorithm starts with an initial estimate $\hat{\boldsymbol{\theta}}^{(0)}$. The procedure described in (8.434)–(8.435) can be used to obtain this initial estimate. The iteration proceeds until

$$\left| \hat{\boldsymbol{\theta}}^{(n+1)} - \hat{\boldsymbol{\theta}}^{(n)} \right| < \epsilon. \quad (8.458)$$

The maximization step ensures that $U(\boldsymbol{\theta}, \boldsymbol{\theta}^{(n)})$ increases on each iteration cycle.

Wu [Wu83] shows that if $U(\boldsymbol{\theta}, \boldsymbol{\theta}^{(n)})$ is continuous in both variables, then the algorithm converges to a stationary point. In the next section, we apply the algorithm to the CML estimation problem.

8.6.3.2 Conditional maximum likelihood estimate

We now apply the EM algorithm to solve for $\hat{\boldsymbol{\theta}}_{cml}$. From (8.451), the observation model is

$$\mathbf{X}_k = \mathbf{V}(\psi)\mathbf{F}_k + \mathbf{N}_k, \quad k = 1, 2, \dots, K. \quad (8.459)$$

We assume the white noise level σ_w^2 is known.

In this case, the unknown parameters are the $D \times 1$ vector ψ , corresponding to the DOAs, and the $2DK \times 1$ vector \mathbf{F} . Thus,

$$\boldsymbol{\theta} = [\psi, \mathbf{F}]. \quad (8.460)$$

The likelihood function is

$$L(\boldsymbol{\theta}) = - \sum_{k=1}^K |\mathbf{X}_k - \mathbf{V}(\psi)\mathbf{F}_k|^2, \quad (8.461)$$

where we have discarded constant factors. The incomplete data are $\mathbf{X}_k, k = 1, 2, \dots, K$. A logical choice for the complete data would be the observation of each plane-wave signal by itself in the presence of noise. Thus,

$$\mathbf{Y}_{kl} = \mathbf{v}(\psi_l)F_{kl} + \mathbf{N}_{kl}, \quad l = 1, 2, \dots, D, \quad (8.462)$$

where

$$E[\mathbf{N}_{kl}\mathbf{N}_{kl}^H] = \beta_l \sigma_w^2 \mathbf{I}, \quad (8.463)$$

and

$$\sum_{l=1}^D \beta_l = 1. \quad (8.464)$$

For simplicity, we will use

$$\beta_l = \frac{1}{D}. \quad (8.465)$$

Then,

$$\mathbf{X}_k = \sum_{l=1}^D \mathbf{Y}_{kl} \triangleq \mathbf{H} \begin{bmatrix} \mathbf{Y}_{k1} \\ \mathbf{Y}_{k2} \\ \vdots \\ \mathbf{Y}_{kD} \end{bmatrix}. \quad (8.466)$$

We see that \mathbf{H} is a linear transformation so that \mathbf{X}_k and $\mathbf{Y}_{kl}, l = 1, 2, \dots, D$ are jointly Gaussian.

Using (8.460) in (8.456), we can write

$$\begin{aligned} U(\boldsymbol{\theta}, \hat{\boldsymbol{\theta}}^{(n)}) &= U(\boldsymbol{\psi}, \mathbf{F}; \hat{\boldsymbol{\psi}}^{(n)}, \mathbf{F}^{(n)}) \\ &= E \left\{ \ln p_{\mathbf{Y}} \left(\mathbf{Y} : \boldsymbol{\psi}, \mathbf{F} | \mathbf{X} : \hat{\boldsymbol{\psi}}^{(n)}, \hat{\mathbf{F}}^{(n)} \right) \right\}. \end{aligned} \quad (8.467)$$

Neglecting constant terms, we can write the ln likelihood function as

$$L(\mathbf{Y} : \boldsymbol{\psi}, \mathbf{F}) = - \sum_{k=1}^K \sum_{l=1}^D |\mathbf{Y}_{kl} - \mathbf{v}(\psi_l) F_{kl}|^2. \quad (8.468)$$

Since \mathbf{Y}_{kl} is a sufficient statistic for estimating $\mathbf{v}(\psi_l) F_{kl}$, we only need to find the conditional mean,

$$\hat{\mathbf{Y}}_{kl}^{(n)} \triangleq E \left\{ \mathbf{Y}_{kl} | \mathbf{X}, \hat{\boldsymbol{\psi}}^{(n)}, \hat{\mathbf{F}}_k^{(n)} \right\}. \quad (8.469)$$

Because \mathbf{Y}_{kl} and \mathbf{X} are jointly Gaussian, this is a classical estimation result. (We use Bayes rule to obtain $p_{\mathbf{Y}_{kl}|\mathbf{X}}(\cdot)$ from $p_{\mathbf{X}|\mathbf{Y}_{kl}}(\cdot)$ and find the mean by inspection.)

$$\hat{\mathbf{Y}}_{kl}^{(n)} - \mathbf{v} \left(\hat{\boldsymbol{\psi}}_l^{(n)} \right) \hat{F}_{kl}^{(n)} = \frac{1}{D} \left[\mathbf{X}_k - \mathbf{V}(\hat{\boldsymbol{\psi}}^{(n)}) \hat{\mathbf{F}}_k^{(n)} \right], \quad (8.470)$$

or

$$\hat{\mathbf{Y}}_{kl}^{(n)} = \mathbf{v} \left(\hat{\boldsymbol{\psi}}_l^{(n)} \right) \hat{F}_{kl}^{(n)} + \frac{1}{D} \left[\mathbf{X}_k - \mathbf{V}(\hat{\boldsymbol{\psi}}^{(n)}) \hat{\mathbf{F}}_k^{(n)} \right]. \quad (8.471)$$

The conditional mean is the signal component from the n th iteration plus a portion $1/D$ of the component of the current observation vector, which is orthogonal to the estimated signal subspace. The result in (8.471) is the

expectation step. We observe that the expectation result is also an estimation result, so the EM algorithm is sometimes referred to as the estimation-maximization algorithm.

To define the maximization step, we recall from (8.345) and (8.338) that

$$\hat{\psi} = \arg \max_{\psi} \left\{ \text{tr} \left[\mathbf{V}(\psi) \left[\mathbf{V}^H(\psi) \mathbf{V}^H(\psi) \right]^{-1} \mathbf{V}^H(\psi) \mathbf{C}_X \right] \right\} \quad (8.472)$$

and

$$\hat{\mathbf{F}}_k = \left[\mathbf{V}^H(\psi) \mathbf{V}(\psi) \right]^{-1} \mathbf{V}^H(\psi) \mathbf{X}_k. \quad (8.473)$$

To get the corresponding relation for the complete data, we define

$$\hat{\mathbf{S}}_{\mathbf{Y}_l}^{(n)} \triangleq \frac{1}{K} \sum_{k=1}^K \hat{\mathbf{Y}}_{kl}^{(n)} \hat{\mathbf{Y}}_{kl}^{(n)H}. \quad (8.474)$$

Then, using (8.474) in the 1-D version of (8.472), we have

$$\hat{\psi}_l^{(n+1)} = \arg \max_{\psi_l} \left\{ \frac{\mathbf{v}^H(\psi_l) \hat{\mathbf{S}}_{\mathbf{Y}_l}^{(n)} \mathbf{v}(\psi_l)}{|\mathbf{v}(\psi_l)|^2} \right\}, \quad l = 1, 2, \dots, D, \quad (8.475)$$

and the 1-D version of (8.473) is

$$\hat{F}_{kl}^{(n+1)} = \frac{\mathbf{v}^H(\hat{\psi}_l^{(n+1)}) \hat{\mathbf{Y}}_{kl}^{(n)}}{|\mathbf{v}(\hat{\psi}_l^{(n+1)})|^2}, \quad l = 1, 2, \dots, D, \quad k = 1, 2, \dots, K. \quad (8.476)$$

Since $|\mathbf{v}(\psi_l)|^2 = N$, (8.475) and (8.476) can be rewritten as

$$\hat{\psi}_l^{(n+1)} = \arg \max_{\psi_l} \left\{ \mathbf{v}^H(\psi_l) \hat{\mathbf{S}}_{\mathbf{Y}_l}^{(n+1)} \mathbf{v}(\psi_l) \right\}, \quad l = 1, 2, \dots, D, \quad (8.477)$$

and

$$\hat{F}_{kl}^{(n+1)} = \frac{1}{N} \mathbf{v}^H(\hat{\psi}_l^{(n+1)}) \hat{\mathbf{Y}}_{kl}^{(n)}, \quad l = 1, 2, \dots, D, \quad k = 1, 2, \dots, K. \quad (8.478)$$

The EM algorithm is defined by (8.471), (8.477), and (8.478). We observe that (8.477) can also be written as

$$\hat{\psi}_l^{(n+1)} = \arg \max_{\psi_l} \left\{ \sum_{k=1}^K |\mathbf{v}^H(\psi_l) \hat{\mathbf{Y}}_{kl}^{(n)}|^2 \right\}. \quad (8.479)$$

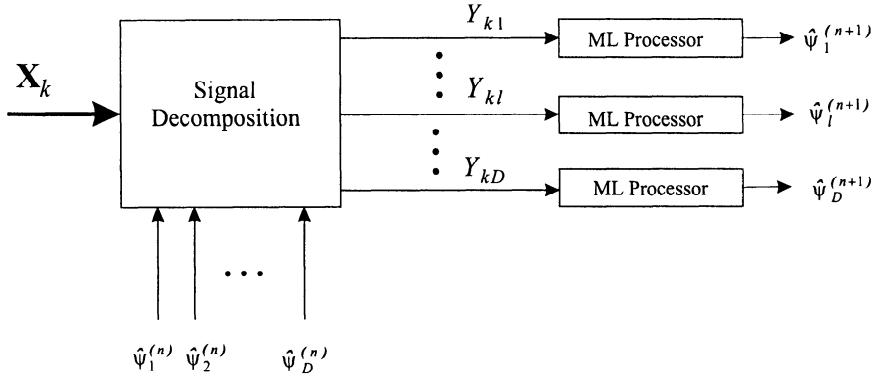


Figure 8.38 Implementation of EM algorithm.

Note that the term in the braces is just the power output of a conventional beamformer when the input is $\hat{\mathbf{Y}}_{kl}^{(n)}$; $k = 1, 2, \dots, K$. This leads to the block diagram of the iteration process shown in Figure 8.38.¹⁸ Note that all of the maximizations are done in parallel.

Summarizing, the EM algorithm for the CML estimate is:

- (i) Initialize the algorithm with $\boldsymbol{\psi}^{(0)}$ and $\hat{F}_{kl}^{(0)}$ (from (8.434) and (8.435)).
- (ii) Estimate $\hat{\mathbf{Y}}_{kl}^{(0)}$, $k = 1, \dots, K$; $l = 1, \dots, D$ using (8.471).
- (iii) Maximize

$$\hat{\boldsymbol{\psi}}_l^{(n+1)} = \arg \max_{\boldsymbol{\psi}_l} \left\{ \sum_{k=1}^K \left| \mathbf{v}^H(\boldsymbol{\psi}_l) \hat{\mathbf{Y}}_{kl}^{(n)} \right|^2 \right\}, \quad (8.480)$$

with $n = 0$.

- (iv) Compute

$$\hat{F}_{kl}^{(n+1)} = \frac{1}{N} \mathbf{v}^H \left(\hat{\boldsymbol{\psi}}_l^{(n+1)} \right) \hat{\mathbf{Y}}_{kl}^{(n)}. \quad (8.481)$$

- (v) Estimate $\hat{\mathbf{Y}}_{kl}^{(n+1)}$ using (8.471).

- (vi) Iterate through steps (ii)–(iv) until

$$|\boldsymbol{\psi}^{(n+1)} - \boldsymbol{\psi}^{(n)}| \leq \epsilon. \quad (8.482)$$

¹⁸This figure is similar to Figure 1 in [FW88], but their model assumed a known signal rather than an unknown nonrandom signal.

The EM algorithm provides an alternative procedure to the AP algorithm for solving the CML problem. A similar procedure is available for the UML algorithm. Miller and Fuhrmann [MF90] use the UML algorithm for uncorrelated signals and enforce the non-negative definite requirement at each iteration. Therefore, their solution is the UML estimator as contrasted to the AML estimator.

For our test scenarios, the EM algorithm converges to the correct solution, but is slower than the AP algorithm.

Fessler and Hero [FH94] develop a new space-alternating generalized EM algorithm (SAGE), and demonstrate that it converges significantly faster than the conventional EM algorithm. The reader is referred to that reference for discussion.

8.6.4 Summary

In this section, we have developed several techniques for solving the ML estimation problem. The most efficient approach in terms of both computation and speed of convergence appears to consist of the following steps:

- (i) Initialize the algorithm by using the procedure in (8.434) or by making an initial estimate using one of the simpler algorithms that we develop in Chapter 9.
- (ii) Use the relaxation methods, either AP or AM as appropriate to get closer to the minimum.
- (iii) Use a search technique such as a quasi-Newton algorithm to achieve the final convergence to the estimate.

All of the computational algorithms developed in this section are appropriate for arbitrary array geometries. In the next section, we consider an approach that is useful for standard linear arrays.

8.7 Polynomial Parameterization

In this section we develop two computationally efficient algorithms that are applicable to standard linear arrays.

In Section 8.7.1, we show how we can reparameterize the ML estimation in terms of a polynomial with its roots on the unit circle.

In Section 8.7.2, we utilize the polynomial parameterization for the CML estimator to develop an algorithm that is called the **iterative quadratic**

maximization likelihood (IQML) algorithm and provides a computationally efficient solution.

In Section 8.7.3, we utilize the polynomial parameterization for the WSF (or MODE) estimator to develop an iterative MODE (IMODE).

In Section 8.7.4, we summarize our results.

8.7.1 Polynomial Parameterization

The technique of reparameterizing the ML estimator problem in terms of a polynomial is due to Kumaresan et al. [KSS86], and Kumaresan and Shaw [KS85] (e.g., Evans and Fishl [EF73]).

In many of the estimators we have discussed the vector ψ enters into the cost function through the projection matrix \mathbf{P}_V (e.g., (8.315) or Table 8.4). We now want to develop an alternative expression for \mathbf{P}_V that will be easier to evaluate.

For a standard linear array, \mathbf{V} is an $N \times D$ matrix,

$$\mathbf{V}(\psi) = \begin{bmatrix} \mathbf{v}(\psi_1) & \mathbf{v}(\psi_2) & \cdots & \mathbf{v}(\psi_D) \end{bmatrix}, \quad (8.483)$$

where

$$\mathbf{v}(\psi_i) = \begin{bmatrix} e^{-j\left(\frac{N-1}{2}\right)\psi_i} & \dots & e^{j\left(\frac{N-1}{2}\right)\psi_i} \end{bmatrix}^T, \quad i = 1, \dots, D. \quad (8.484)$$

The first step in developing the parameterization is to recognize that, if we let

$$z = e^{j\psi}, \quad (8.485)$$

then the polynomial

$$b(z) = b_0 z^D + b_1 z^{D-1} + \cdots + b_D \quad (8.486)$$

describes the spatial characteristics of the signal component. It can be written as

$$b(z) = b_0 \prod_{i=1}^D (z - z_i), \quad (8.487)$$

and has roots at

$$z_i = e^{j\psi_i}, \quad i = 1, 2, \dots, D. \quad (8.488)$$

We define the coefficient vector of the polynomial as

$$\mathbf{b} = \begin{bmatrix} b_0 & b_1 & \cdots & b_D \end{bmatrix}^T. \quad (8.489)$$

If we find the ML estimate of \mathbf{b} , we can use it to obtain the ML estimate of $\psi_i, i = 1, 2, \dots, D$. To carry out this procedure we construct a $N \times (N - D)$ Toeplitz matrix, \mathbf{B} ,

$$\mathbf{B} = \begin{bmatrix} b_D^* & & & \\ b_{D-1}^* & b_D^* & & \mathbf{0} \\ \vdots & \ddots & \ddots & \\ b_0^* & \vdots & \ddots & b_D^* \\ 0 & b_0^* & & \\ 0 & 0 & & \\ \vdots & \vdots & \ddots & \\ 0 & 0 & \cdots & b_0^* \end{bmatrix}, \quad (8.490)$$

and show that we can write the noise subspace projection matrix as

$$\mathbf{P}_{\mathbf{V}}^\perp = \mathbf{P}_{\mathbf{B}} = \mathbf{B} (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H. \quad (8.491)$$

To verify (8.491), we consider the i th column of $\mathbf{V}(\psi)$, $\mathbf{v}(\psi_i)$. Denote the first column of \mathbf{B} as \mathbf{b}_1 . Then,

$$\begin{aligned} \mathbf{b}_1^H \mathbf{v}(\psi_i) &= b_D + b_{D-1} e^{j\psi_i} + \cdots + b_0 e^{jD\psi_i} \\ &= b(e^{j\psi_i}) = 0, \quad i = 1, 2, \dots, D, \end{aligned} \quad (8.492)$$

where the last equality follows from (8.488). Similarly,

$$\mathbf{b}_2^H \mathbf{v}(\psi_i) = b(e^{j\psi_i}) \cdot e^{j\psi_i} = 0, \quad i = 1, 2, \dots, D, \quad (8.493)$$

and so forth. Thus, the columns of \mathbf{B} are orthogonal to \mathbf{V} . Since for any $\mathbf{b} \neq \mathbf{0}$, \mathbf{B} has rank $N - D$, its columns span the orthogonal complement to \mathbf{V} and \mathbf{P}_B is equal to $\mathbf{P}_{\mathbf{V}}^\perp$.

The equivalence of the two projection matrices,

$$\mathbf{P}_{\mathbf{V}}^\perp = \mathbf{P}_{\mathbf{B}}, \quad (8.494)$$

is a key result that is used to solve ML problems.

8.7.2 Iterative Quadratic Maximum Likelihood (IQML)

In this section, we illustrate one application of polynomial parameterization. The specific case of interest is the CML estimate. The resulting algorithm is called the IQML algorithm.

The IQML algorithm was derived independently by Bresler and Macovski [BM86] and Kumaresan, Scharf, and Shaw (Kumaresan et al. [KSS86] and Kumaresan and Shaw [KS88][Sha87]). Our discussion is similar to that in [BM86] but we include a modification due to Nagesha and Kay [NK94]. We give a short discussion of other similar algorithms and other implementations of IQML at the end of this section. Although we are using the technique for CML, we see that it can be extended to any estimator that relies on \mathbf{P}_V or \mathbf{P}_V^\perp in the estimation algorithm.

The CML estimate is given by (8.346) as

$$\hat{\psi}_{cml} = \arg \min_{\psi} \left\{ \text{tr} [\mathbf{P}_V^\perp \mathbf{C}_x] \right\}. \quad (8.495)$$

We can write the required minimization in (8.495) as

$$\min_{\mathbf{b} \in \Omega_b} J(\mathbf{b}), \quad (8.496)$$

where

$$J(\mathbf{b}) = \text{tr} [\mathbf{P}_B \mathbf{C}_x] = \text{tr} \left[\mathbf{B} \left(\mathbf{B}^H \mathbf{B} \right)^{-1} \mathbf{B}^H \mathbf{C}_x \right], \quad (8.497)$$

and we must specify the constraints on \mathbf{b} by defining Ω_b .

We first derive an alternative form for $J(\mathbf{b})$ that is useful for computation. We define an $(N - D) \times (D + 1)$ data matrix, for the k th snapshot as

$$\mathbf{A}_k = \begin{bmatrix} X_D(k) & X_{D-1}(k) & \cdots & X_0(k) \\ X_{D+1}(k) & X_D(k) & \cdots & X_1(k) \\ \vdots & \vdots & \ddots & \vdots \\ X_{N-1}(k) & X_{N-2}(k) & \cdots & X_{N-1-D}(k) \end{bmatrix}, \quad (8.498)$$

where $X_n(k)$ is the frequency-domain snapshot of the output of the n th sensor at snapshot k . We use \mathbf{A}_k to avoid confusion with the snapshot vector $\mathbf{X}(k)$.

By direct multiplication, one can show that

$$\mathbf{B}^H \mathbf{X}(k) = \mathbf{A}_k \mathbf{b}. \quad (8.499)$$

Substituting (8.499) into (8.497) gives

$$\begin{aligned} K \cdot J(\mathbf{b}) &= \text{tr} \left[\sum_{k=1}^K \mathbf{B} \left(\mathbf{B}^H \mathbf{B} \right)^{-1} \mathbf{B}^H \mathbf{X}(k) \mathbf{X}^H(k) \right] \\ &= \mathbf{b}^H \left[\sum_{k=1}^K \mathbf{A}_k^H \left(\mathbf{B}^H \mathbf{B} \right)^{-1} \mathbf{A}_k \right] \mathbf{b}. \end{aligned} \quad (8.500)$$

The form of (8.500) suggests an iterative procedure in which we hold $\mathbf{B}^H \mathbf{B}$ fixed from the previous step and solve the quadratic minimization in \mathbf{b} . We must impose some constraints on \mathbf{b} to get a useful solution.

The first constraint is imposed to guarantee a non-zero \mathbf{b} . Breslev and Macouski [BM86] use

$$\operatorname{Re}[b_0] = 1, \quad (8.501)$$

to guarantee a non-zero \mathbf{b} . Nagesha and Kay [NK94] and others use the norm constraint

$$\|\mathbf{b}\|^2 = 1. \quad (8.502)$$

As pointed out in [NK94], the constraint in (8.501) is not useful for certain signal geometries. A simple example is the case of a single signal at $\psi = 0$. Here $\operatorname{Re}[b_0] = 0$ (see Problem 8.7.12). We use the quadratic constraint in (8.502) and refer to the algorithm as IQML-QC to avoid confusion with some of IQML algorithms in the literature.

We would like the roots of the \mathbf{b} polynomial to lie on the unit circle. We impose a conjugate symmetry constraint on \mathbf{b} . Conjugate symmetry is a necessary, but not sufficient condition.¹⁹ We find that with the conjugate symmetry constraint, when the algorithm is above its threshold, the roots lie on the unit circle and start to move off the circle as we transition into the threshold region. A discussion of this root behavior is contained in Stoica and Nehorai [SN89a].

Thus, we require

$$b_i = b_{D-i}^*, \quad i = 0, 1, \dots, D. \quad (8.503)$$

We introduce the constraint in (8.503) using a technique from Kumaresan and Shaw [KS88] and Shaw [Sha87]. We assume that D is odd.²⁰ We define a $(D + 1) \times 1$ vector, \mathbf{c} as

$$\mathbf{c} = \begin{bmatrix} \operatorname{Re}(b_0) & \operatorname{Im}(b_0) & \operatorname{Re}(b_1) & \dots & \operatorname{Re}(b_{(D-1)/2}) & \operatorname{Im}(b_{(D-1)/2}) \end{bmatrix}^T. \quad (8.504)$$

¹⁹The exception to this statement is the case of a single root, where the conjugate symmetry constraint is also sufficient. Shaw [Sha95] has derived a constrained MLE estimator based on this property.

²⁰The development for even D is done in Problems 8.7.9 and 8.7.10.

We define a $(D + 1) \times (D + 1)$ transformation matrix \mathbf{T} ,

$$\mathbf{T} \triangleq \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & j & 0 & 0 & \cdots & 0 & 0 \\ 0 & 0 & 1 & j & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & 0 & \cdots & 1 & j \\ 0 & 0 & 0 & 0 & \cdots & 1 & -j \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 1 & -j & \cdots & 0 & 0 \\ 1 & -j & 0 & 0 & \cdots & 0 & 0 \end{bmatrix}. \quad (8.505)$$

This can be written as

$$\mathbf{T} = \frac{1}{\sqrt{2}} \begin{bmatrix} \mathbf{I}_{(D+1)/2} \otimes \begin{bmatrix} 1 & j \end{bmatrix} \\ \mathbf{J}_{(D+1)/2} \otimes \begin{bmatrix} 1 & -j \end{bmatrix} \end{bmatrix}, \quad (8.506)$$

where \otimes is Kronecker product. Then,

$$\mathbf{b} = \mathbf{T}\mathbf{c}. \quad (8.507)$$

We define

$$\tilde{\mathbf{Q}}_{\mathbf{x}} \triangleq \mathbf{T}^H \left(\sum_{k=1}^K \mathbf{A}_k^H [\mathbf{B}^H \mathbf{B}]^{-1} \mathbf{A}_k \right) \mathbf{T}, \quad (8.508)$$

and

$$F(\mathbf{c}) = \mathbf{c}^T \tilde{\mathbf{Q}}_{\mathbf{x}} \mathbf{c}. \quad (8.509)$$

Because $Im\{\tilde{\mathbf{Q}}_{\mathbf{x}}\}$ is skew-symmetric, its quadratic form is zero. Thus, $Im[F(\mathbf{c})] = 0$, so we can use²¹

$$F(\mathbf{c}) = \mathbf{c}^T [Re\{\tilde{\mathbf{Q}}_{\mathbf{x}}\}] \mathbf{c}. \quad (8.510)$$

We assume \mathbf{B} is fixed from the previous iteration. We minimize $\mathbf{F}(\mathbf{c})$ subject to the norm constraint

$$\|\mathbf{c}\|^2 = 1. \quad (8.511)$$

The solution follows immediately if we expand $Re\{\tilde{\mathbf{Q}}_{\mathbf{x}}\}$ in terms of its eigenvalues and eigenvectors:

$$Re\{\tilde{\mathbf{Q}}_{\mathbf{x}}\} = \sum_{i=1}^{D+1} \lambda_i \phi_i \phi_i^H. \quad (8.512)$$

²¹An alternative derivation of (8.509) is developed in Problem 8.7.8.

The non-zero \mathbf{c} that minimizes $F(\mathbf{c})$ is the eigenvector corresponding to the smallest eigenvalue,

$$\hat{\mathbf{c}} = \phi_{min}, \quad (8.513)$$

where the notation emphasizes that ϕ_{min} corresponds to the smallest eigenvalue.

We can summarize the iterative algorithm using the quadratic constraint (IQML-QC):

- (i) Initialization: set $m = 0$ and $\mathbf{B}_{(0)} = \mathbf{B}_0$.
- (ii) Compute

$$\tilde{\mathbf{Q}}_{\mathbf{x}}^{(m)} = \mathbf{T}^H \left(\sum_{k=1}^K \mathbf{A}_k^H \left[\mathbf{B}_{(m)}^H \mathbf{B}_{(m)} \right]^{-1} \mathbf{A}_k \right) \mathbf{T}, \quad (8.514)$$

where $\mathbf{B}_{(m)}$ is given by (8.490), and \mathbf{A}_k is given by (8.498).

- (iii) Perform an eigendecomposition of $Re\{\tilde{\mathbf{Q}}_{\mathbf{x}}^{(m)}\}$ as in (8.512). Set

$$\hat{\mathbf{c}}^{(m)} = \phi_{min}^{(m)}. \quad (8.515)$$

- (iv) Find

$$\hat{\mathbf{b}}^{(m)} = \mathbf{T}\hat{\mathbf{c}}^{(m)}. \quad (8.516)$$

- (v) Find the roots of $\hat{\mathbf{b}}_{(m)}(z)$. Denote these roots as \hat{z}_i , $i = 1, 2, \dots, D$. Then

$$\hat{\psi}_i = \angle \arg(\hat{z}_i), \quad i = 1, 2, \dots, D. \quad (8.517)$$

- (vi) Set $m = m + 1$.

- (vii) Repeat steps (ii)–(v).

- (viii) Check convergence: Is

$$\|\hat{\mathbf{b}}(m+1) - \hat{\mathbf{b}}(m)\| < \epsilon, \quad (8.518)$$

where ϵ reflects the desired precision. If (8.518) is satisfied we terminate the iteration and use the $\hat{\psi}_i$ from step (v), otherwise return to step (ii).

We can initialize the algorithm with

$$\mathbf{B}_{(0)}^H \mathbf{B}_{(0)} = \mathbf{I}, \quad (8.519)$$

or we can perform an initial estimate of ψ using one of the simpler algorithms that we develop in Chapter 9.

As pointed out in [BM86], the matrices $(\mathbf{B}^H \mathbf{B})$ and $\tilde{\mathbf{Q}}_{\mathbf{x}}$ have considerable structure that can be exploited for computational efficiency. $\mathbf{B}^H \mathbf{B}$ is a banded Hermitian matrix, so its inverse can be efficiently computed (e.g., [Kum85], [KSS86]). The inverse of $\mathbf{B}^H \mathbf{B}$ can be computed as the inverse of a $D \times D$ matrix rather than a $(N - D) \times (N - D)$ matrix.

There are a number of references concerning IQML-type of algorithms. An early reference is Evans and Fischl [EF73]. Other iteration algorithms include Matausek et al. [MSR83], Kay [Kay84]. Algorithms using other approaches include Tufts and Kumaresan [TK82] and Kumaresan et al. [KT83]. In [BM86], the relationship of IQML to these algorithms is discussed.

There have been a number of papers discussing IQML since [BM86]. McClellan and Lee [Mcl91] show the equivalence of the Steiglitz-McBride algorithm [SM65] and IQML.

Clark and Scharf [CS92] discuss the complexity of the IQML algorithm and compare it to other algorithms including the Steiglitz-McBride algorithm (e.g., [SM65]). An efficient implementation of IQML is given by Hua [Hua94]. [KS88] also discuss an adaptive version to track slowly moving sources. Stoica et al. [SLS97] analyze the asymptotic behavior. Li et al. [LSL98] compare the performance of IQML and MODE estimators.

The issues of interest with respect to the IQML algorithm are:

- (i) Rate of convergence.
- (ii) Comparison of threshold behavior with CML algorithm in Section 10.5.2.
- (iii) Computational complexity.

We illustrate the performance of IQML-QC for several of the test cases that we have studied in Sections 8.5 and 8.6. However, before doing the examples we develop the polynomial parameterized version of MODE.

Before developing MODE, we should note that we have developed the IQML algorithm for solving the CML estimation problem. The polynomial parameterization can be used for other algorithms.

For the AML algorithm

$$F(\mathbf{c}) = \ln \det \left[\mathbf{P}_{\mathbf{B}}^{\perp} \mathbf{C}_{\mathbf{x}} \mathbf{P}_{\mathbf{B}}^{\perp} + \hat{\sigma}_w^2 \mathbf{P}_{\mathbf{B}} \right], \quad (8.520)$$

with

$$\hat{\sigma}_w^2 = \frac{1}{N - D} \text{tr} [\mathbf{P}_{\mathbf{B}} \mathbf{C}_{\mathbf{x}}]. \quad (8.521)$$

and \mathbf{c} is defined in (8.504). We use the same general iterative procedure as above.

In the next section we discuss the polynomial parameterized version of WSF (MODE).

8.7.3 Polynomial WSF (MODE)

The weighted subspace fitting (WSF) (or MODE) algorithm was described in Section 8.5.3 (see (8.367)–(8.370)). The algorithm is

$$\hat{\psi} = \arg \min_{\psi} \left\{ \text{tr} \left[\mathbf{P}_{\mathbf{V}}^{\perp} \hat{\mathbf{U}}_S \mathbf{W}_{ao} \hat{\mathbf{U}}_S^H \right] \right\}, \quad (8.522)$$

where $\hat{\mathbf{U}}_S$ is an $N \times D$ matrix whose columns are the estimated signal eigenvectors and \mathbf{W}_{ao} is a $D \times D$ diagonal matrix,

$$\mathbf{W}_{ao} = \hat{\tilde{\Lambda}}_S^{-1}, \quad (8.523)$$

$$\hat{\tilde{\Lambda}} = \hat{\Lambda}_S - \hat{\sigma}_w^2 \mathbf{I}, \quad (8.524)$$

$$\hat{\sigma}_w^2 = \frac{1}{N-D} \sum_{i=D+1}^N \hat{\lambda}_i. \quad (8.525)$$

The polynomial implementation of MODE was defined in Stoica and Sharman's original papers ([SS90a],[SS90b]). Li et al. [LSL98] give a more explicit definition of the algorithm. Both references emphasize a two-step algorithm. We focus on an iterative version of the algorithm, that we call IMODE. In the examples in [LSL98] they use IMODE for some of their results. Li et al. [LSL98] also contain a discussion of relative computational complexity of IQML and MODE.

Using a polynomial parameterization, we can rewrite (8.522) as

$$\hat{\mathbf{b}} = \arg \min_{\psi} \left\{ \text{tr} \left[\mathbf{P}_{\mathbf{B}} \hat{\mathbf{U}}_S \mathbf{W}_{ao} \hat{\mathbf{U}}_S^H \right] \right\}, \quad (8.526)$$

subject to a quadratic constraint,

$$\|\mathbf{b}\|^2 = 1, \quad (8.527)$$

and a conjugate symmetry constraint

$$b_i = b_{D-i}^*. \quad (8.528)$$

We can rewrite the right side of (8.526) as

$$J(\mathbf{b}) = \text{tr} \left[\mathbf{B} (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \hat{\mathbf{U}}_S \mathbf{W}_{ao} \hat{\mathbf{U}}_S^H \right]. \quad (8.529)$$

We use an approach similar to the IQML approach to get (8.529) into a more usable form. For notational simplicity we define

$$\tilde{\mathbf{U}}_S \triangleq \hat{\mathbf{U}}_S \mathbf{W}_{ao}^{\frac{1}{2}}, \quad (8.530)$$

and write

$$\hat{\mathbf{U}}_S \mathbf{W}_{ao} \hat{\mathbf{U}}_S^H = \tilde{\mathbf{U}}_S \tilde{\mathbf{U}}_S^H = \sum_{d=1}^D \tilde{\phi}_d \tilde{\phi}_d^H, \quad (8.531)$$

where $\tilde{\phi}_d$ are the columns of $\tilde{\mathbf{U}}_S$.

We can write (8.529) as

$$\begin{aligned} J(\mathbf{b}) &= \text{tr} \left[\sum_{d=1}^D \mathbf{B} (\mathbf{B}^H \mathbf{B}) \mathbf{B}^H \tilde{\phi}_d \tilde{\phi}_d^H \right] \\ &= \text{tr} \left[\sum_{d=1}^D \tilde{\phi}_d^H \mathbf{B} (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{B}^H \tilde{\phi}_d \right]. \end{aligned} \quad (8.532)$$

We define a matrix

$$\mathbf{A}_d = \begin{bmatrix} \tilde{\phi}_d(D+1) & \tilde{\phi}_d(D) & \cdots & \tilde{\phi}_d(1) \\ \tilde{\phi}_d(D+2) & \tilde{\phi}_d(D+1) & \cdots & \tilde{\phi}_d(2) \\ \vdots & \vdots & \ddots & \vdots \\ \tilde{\phi}_d(N) & \tilde{\phi}_d(N-1) & \cdots & \tilde{\phi}_d(N-D) \end{bmatrix}. \quad (8.533)$$

This matrix is analogous to the data matrix in (8.498) except its components come from the signal subspace matrix. Then

$$\mathbf{B}^H \tilde{\phi}_d = \mathbf{A}_d \mathbf{b}, \quad (8.534)$$

and

$$J(\mathbf{b}) = \mathbf{b}^H \left[\sum_{d=1}^D \mathbf{A}_d^H (\mathbf{B}^H \mathbf{B})^{-1} \mathbf{A}_d \right] \mathbf{b}. \quad (8.535)$$

This form is identical to (8.500), so the subsequent steps are identical to (8.502)–(8.518). The \mathbf{c} vector is defined in (8.504) and the \mathbf{T} matrix is defined in (8.505):

$$\tilde{\mathbf{Q}}_D \triangleq \mathbf{T}^H \left(\sum_{d=1}^D \mathbf{A}_d^H [\mathbf{B}^H \mathbf{B}]^{-1} \mathbf{A}_d \right) \mathbf{T}, \quad (8.536)$$

and

$$F(\mathbf{c}) = \mathbf{c}^T \tilde{\mathbf{Q}}_D \mathbf{c} = \mathbf{c}^T [\operatorname{Re}\{\tilde{\mathbf{Q}}_D\}] \mathbf{c}. \quad (8.537)$$

We minimize $F(\mathbf{c})$ subject to a unit norm constraint on \mathbf{c} . The resulting $\hat{\mathbf{c}}$ is the eigenvector of $\operatorname{Re}\{\tilde{\mathbf{Q}}_D\}$ corresponding to the smallest eigenvalue.

We can now define an iterative version of the MODE algorithm using the same steps as in the IQML. We refer to it as the IMODE algorithm.

The IMODE algorithm can be summarized:

- (i) Initialization: set $m = 0$ and choose $\mathbf{B}_{(0)}$ such that

$$\mathbf{B}_{(0)}^H \mathbf{B}_{(0)} = \mathbf{I}. \quad (8.538)$$

- (ii) Compute

$$\tilde{\mathbf{Q}}_D^{(m)} \triangleq \mathbf{T}^H \left(\sum_{d=1}^D \mathbf{A}_d^H [\mathbf{B}_{(m)}^H \mathbf{B}_{(m)}]^{-1} \mathbf{A}_d \right) \mathbf{T}. \quad (8.539)$$

In 8.539, the matrix $\mathbf{B}_{(m)}$ is given by (8.490). Compute $\tilde{\mathbf{U}}_S$ and $\tilde{\phi}_d$. Use $\tilde{\phi}_d$ in (8.533) to form \mathbf{A}_d .

- (iii) Perform an eigendecomposition of $\operatorname{Re}\{\tilde{\mathbf{Q}}_D^{(m)}\}$ and set

$$\hat{\mathbf{c}}^{(m)} = \hat{\phi}_{min}^{(m)}, \quad (8.540)$$

where $\hat{\phi}_{min}^{(m)}$ is the eigenvector corresponding to the smallest eigenvalue.

- (iv) Find

$$\hat{\mathbf{b}}^{(m)} = \mathbf{T} \hat{\mathbf{c}}^{(m)}. \quad (8.541)$$

- (v) Find the roots of $\hat{\mathbf{b}}^{(m)}(z)$. Denote these roots as $\hat{z}_i^{(m)}, i = 1, \dots, D$. Then

$$\hat{\psi}_i^{(m)} = \arg(\hat{z}_i^{(m)}), \quad i = 1, 2, \dots, D. \quad (8.542)$$

- (vi) Set $m = m + 1$.

- (vii) Repeat steps (ii)–(v).

- (viii) Check convergence

$$\|\hat{\mathbf{b}}(m+1) - \hat{\mathbf{b}}(m)\| < \epsilon. \quad (8.543)$$

If (8.543) is satisfied, terminate the iteration and set

$$\hat{\psi}_i = \psi_i^{(m+1)}. \quad (8.544)$$

Otherwise, continue the iteration.

Stoica and Sharman [SS90a] showed that, asymptotically, replacement of $[\mathbf{B}^H \mathbf{B}]^{-1}$ by a consistent estimate will have a negligible effect. Thus, a large sample realization of $\hat{\mathbf{b}}$ is given by

$$\hat{\mathbf{b}} = \min_{\mathbf{b}} \left\{ \text{tr} \left\{ [\tilde{\mathbf{U}}_s^H \mathbf{B}] [\hat{\mathbf{B}}_1^H \hat{\mathbf{B}}_1]^{-1} [\mathbf{B}^H \tilde{\mathbf{U}}_s] \right\} \right\}. \quad (8.545)$$

In (8.545), the matrix $\hat{\mathbf{B}}_1$ is the consistent estimate of \mathbf{B} obtained by minimizing $\text{tr} \left[(\tilde{\mathbf{U}}_s^H \mathbf{B}) (\mathbf{B}^H \tilde{\mathbf{U}}_s) \right]$. Therefore, the MODE algorithm developed by Stoica and Sharman is a two-step algorithm that corresponds to terminating the IMODE algorithm in (8.538)–(8.542) at the $m = 1$ step. When comparing IMODE and MODE, we refer to the latter as two-step MODE to emphasize the difference.

We now consider two of the test cases that we have analyzed previously.

Example 8.7.1 (continuation, Example 8.5.2)²²

Consider the same model as in Example 8.5.2. There are two equal-power uncorrelated plane-wave signals impinging on the array from $\psi = \pm \Delta\psi_R/2$. We implement three algorithms; IQML-QC²³, IMODE, and two-step MODE.

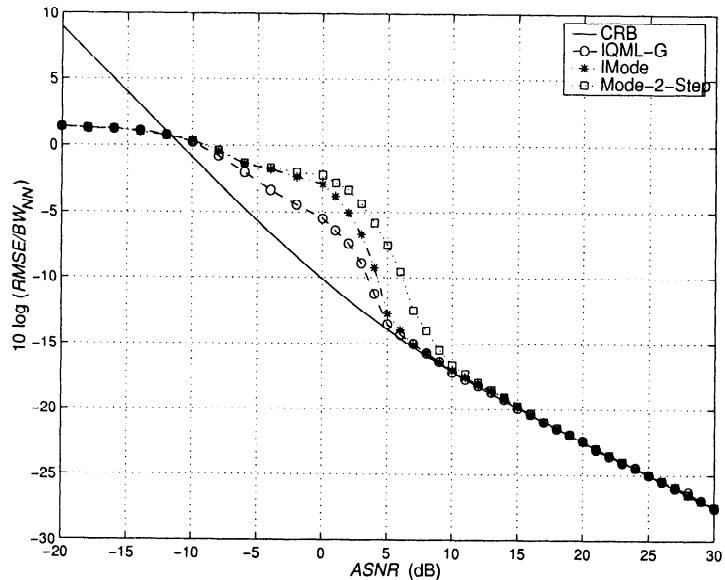
In Figure 8.39(a), the normalized RMSE is plotted versus $ASNR$ for the three algorithms. The thresholds for IMODE and IQML-QC are slightly higher than the threshold using grid search shown in Figure 8.32. The threshold for two-step MODE is about 3 dB higher. Above an $ASNR = 10$ dB, all three algorithms are essentially on the CRB.

In Figure 8.39(b), a histogram showing the number of iterations required when the $ASNR = 10$ dB and the error threshold, $\epsilon = 0.01$ is plotted. This $ASNR$ is above the threshold point. The most common number of iterations is four. However, from Figure 8.39(a), the last iterations do not improve the performance significantly, so two-step MODE is adequate above threshold if the number of iterations needs to be constrained.

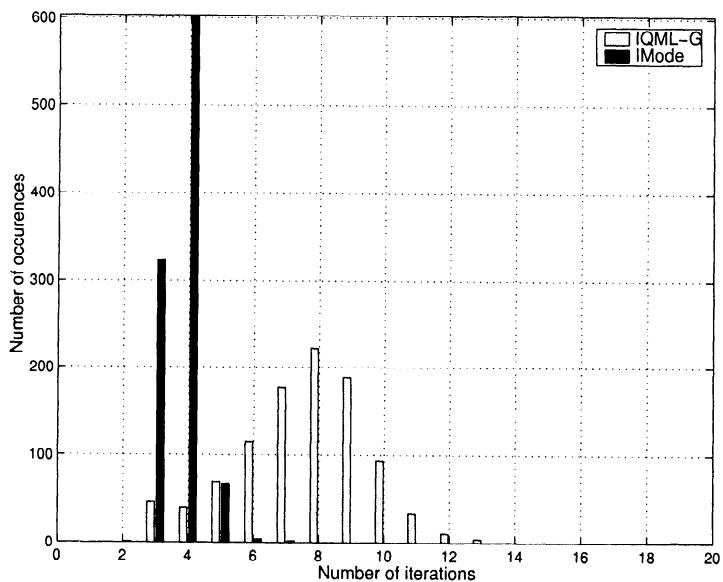
In Figure 8.39(c), the same results are shown for an $ASNR = 5$ dB which is just below the threshold. We see that, in this set of 1000 trials, three or more iterations were always used and 4–6 iterations were generally needed. In this case, the additional iterations provide a useful decrease in the RMSE, and IMODE should be used.

²²Figures 8.39 and 8.40 are due to J. Hiemstra (private communication).

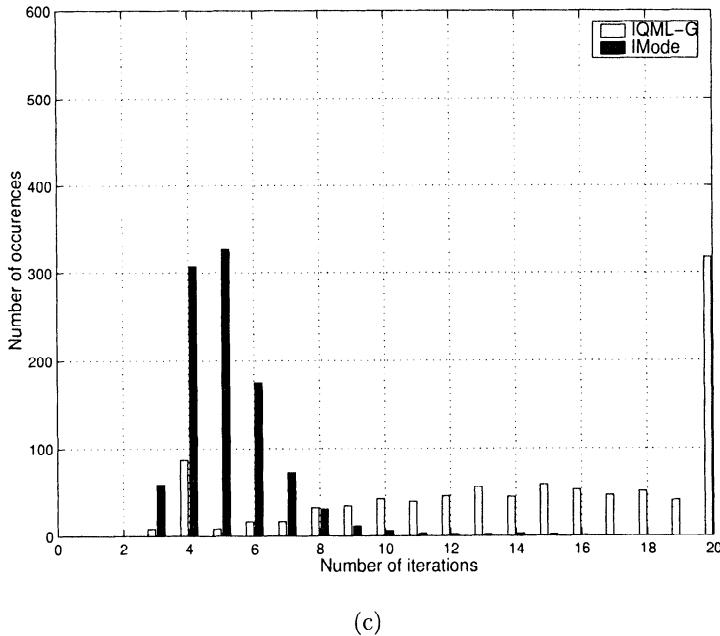
²³In implementing IQML-QC in MATLAB[®], we found that, after finding $\hat{\mathbf{c}}$ in (8.515) to an adequate precision, we could decrease the RMSE further by using the *fmins* function to minimize $\text{Re} [\mathbf{c}^T \tilde{\mathbf{Q}}_{\mathbf{x}} \mathbf{c}] + |\|\mathbf{c}\|^2 - 1|$. This is a minimization with a penalty function and can be viewed as a slight modification of a pure IQML algorithm. A gradient technique could also be used. The technique is most useful near the threshold. MATLAB[®] also has a constrained minimization that could be used. We denote this version as IQML-G in the legend.



(a)



(b)



(c)

Figure 8.39 IQML-QC, IMODE, two-step MODE: $N = 10$, $K = 100$, $\Delta\psi = \Delta\psi_R$, $\rho = 0$: (a) normalized RMSE versus $ASNR$; (b) histogram of number of iterations, $ASNR = 10$ dB; (c) histogram of number of iterations, $ASNR = 5$ dB.

Example 8.7.2 (continuation)

Consider the same model as in Example 8.7.1 except that the signals are correlated. Two values of ρ , $\rho = 0.95$ and $\rho = 1.0$, are considered. Note that $\phi_\rho = 0$, which is the most difficult case for the algorithms. The results are shown in Figure 8.40.

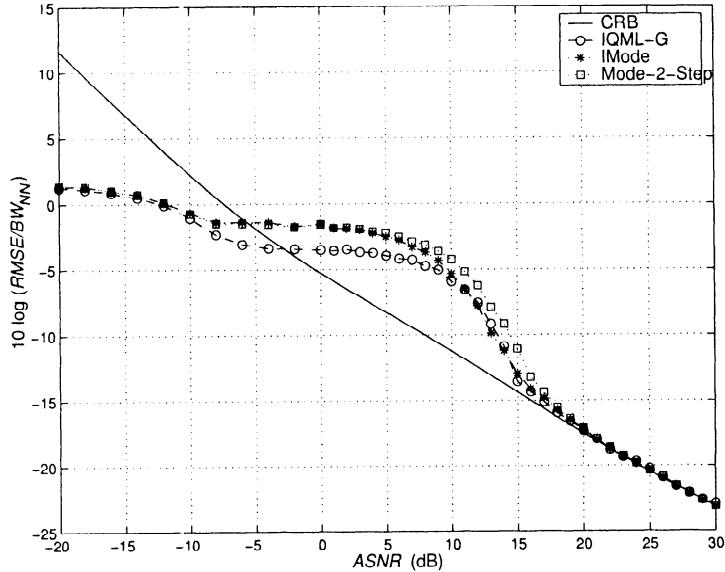
In Figure 8.40(a), the normalized RMSE is plotted versus $ASNR$ for the three algorithms for the $\rho = 0.95$ case. The threshold for IQML-QC and IMODE is 15 dB and two-step MODE is slightly higher. This value is 10 dB higher than the uncorrelated case. Above threshold, the RMSEs of all three algorithms are close to the CRB.

In Figure 8.40(b), the normalized RMSE is plotted versus $ASNR$ for the three algorithms for the $\rho = 1.0$ case. The approach to the CRB is more gradual. The threshold is in the 23–25 dB range. The IMODE and two-step MODE algorithms have smaller RMSEs than IQML-QC in the $ASNR$ regions of interest.

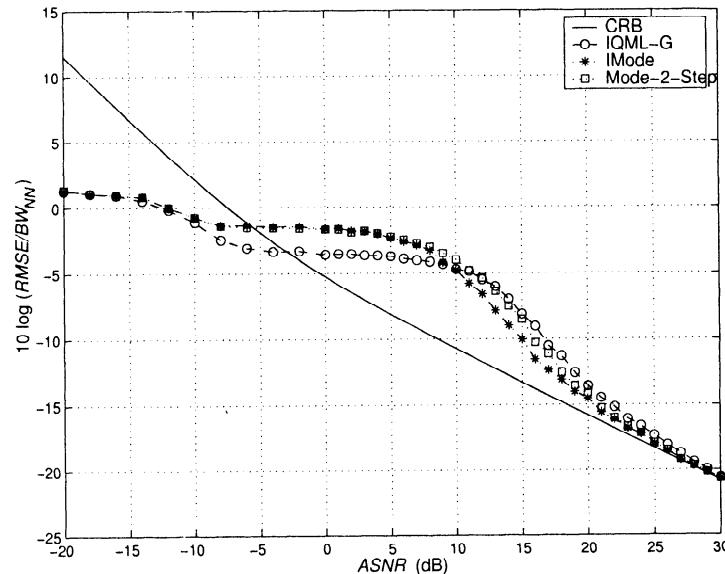
In Figure 8.40(c), a histogram for the $\rho = 1.0$ case showing the number of iterations required when the $ASNR = 25$ dB and the error threshold $\epsilon = 0.01$ is plotted. This $ASNR$ is above the threshold. About 95% of the trials used three or four iterations.

In Figure 8.40(d), the same information is plotted for the $ASNR = 20$ dB case. This $ASNR$ is in the threshold region; The number of iterations range from three to seven, but three and four iterations were used in almost 80% of the trials.

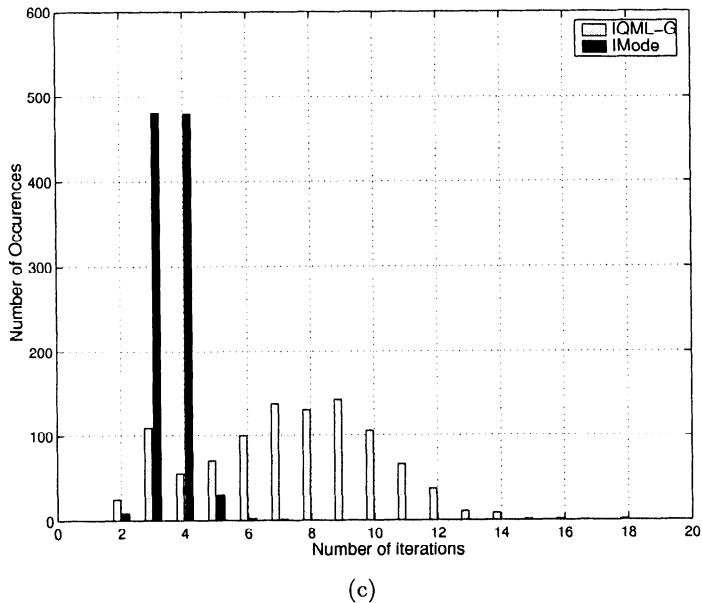
For the uncorrelated signals case in Example 8.7.1, the modified IQML-



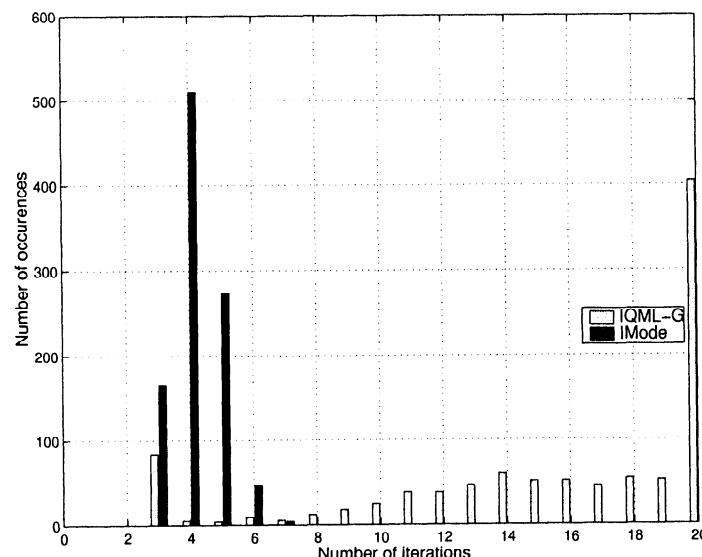
(a)



(b)



(c)



(d)

Figure 8.40 IQML-QC, IMODE, two-step MODE: $N = 10$, $K = 100$, $\Delta\psi = \Delta\psi_R$: (a) normalized RMSE versus ASNR, $\rho = 0.95$; (b) normalized RMSE versus ASNR, $\rho = 1.0$; (c) histogram of number of iterations, $\rho = 1.0$, ASNR = 25 dB; (d) histogram of number of iterations, $\rho = 1.0$, ASNR = 20 dB.

QC algorithm and IMODE had similar threshold behavior and approached the CRB above threshold. The two-step MODE algorithm had a higher threshold, but approached the CRB above threshold. The IQML-QC algorithm has a higher flop count, so IMODE is the preferred algorithm for this case. The behavior in this example appears to be characteristic of the general uncorrelated signal case. If the threshold behavior is an important factor in a particular application, then IMODE is useful (or perhaps, five-step MODE). If performance above threshold is the important factor, the two-step MODE appears to be adequate. Similar comments apply to the correlated and coherent signal case.

Li et al. [LSL98] provide a more general comparison of IQML, IMODE, and two-step MODE. They also provide a detailed discussion of an efficient implementation of IMODE and a comparison of flop counts. The interested reader should consult that reference.

For standard linear arrays, IMODE is the leading candidate of the algorithms developed up to this point in the text. There are still issues, such as the effect of an unknown number of signals and robustness to array perturbation that must be explored. In addition, other algorithms are developed in Chapter 9.

A unitary version of IMODE is developed in Gershman and Stoica [GS99]. We discuss it in Chapter 9 after other unitary algorithms have been developed. It is identical to FB-IMODE, which is standard IMODE using FB averaging in the sample spectral matrix (e.g., Stoica and Jansson [SJ97]). In [SJ97], it is shown that the RMSE of FB-MODE in the asymptotic region is greater than or equal to the RMSE of standard MODE. In Example 8.7.2, $\phi_\rho = 0^\circ$, so the two versions are identical.

8.7.4 Summary

In this section we have developed a polynomial parameterization and then showed how it could be used to develop efficient computational algorithms. Although we used CML and MODE as examples, the general technique is applicable to any algorithm that utilizes \mathbf{P}_V . Whenever the array geometry allows it, the polynomial parameterization approach should be considered.

We can also develop Newton or quasi-Newton algorithms that utilize the polynomial parameterization. The advantage of doing the Newton algorithm in polynomial space is that the number of computations at each iteration is reduced. Starer and Nehorai [SN92] have derived the algorithm and give examples.

8.8 Detection of Number of Signals

Up to this point in our discussion of parameter estimation we have assumed that the number of signals, D , is known. In practice, we would use one of the algorithms developed in Section 7.8 to estimate D . We use that estimate, denoted as \hat{D} , in our estimation algorithm. We refer to this technique as **separable detection**. All of the results in Section 7.8 apply to this problem.

There are a number of scenarios studied in the problems. In some cases, the threshold of the detection algorithms occurs at the same *ASNR* or higher *ASNR* than the threshold of an ML estimator operating with the correct number of signals. This result suggests that all estimation algorithms needed to be checked for robustness to errors in \hat{D} .

In most parameter estimation problems, if $\hat{D} \neq D$, it is better to overestimate rather than underestimate. In many cases, the output of the parameter estimator serves as the input to another processor such as a tracker. We can rely on the subsequent processor to eliminate the extra signals. This observation suggests that if we must operate in the vicinity of the threshold, we should use AIC or AIC-FB and require subsequent processing (after the estimator) to eliminate an extra signals introduced by an overestimation error from AIC. If we are operating above threshold, we should use MDL or MDL-FB, because it provides consistent estimates.

A second approach is to jointly detect the number of signals and estimate their location. The CML version of the joint detection-estimation algorithm was developed by Wax and Ziskind [WZ89]. The AML version of the joint detection-estimation algorithm was developed by Wax [Wax91]. One can show that the MDL version of resulting estimators is consistent (e.g., the approach by Zhao et al. [ZKB87] or Wax and Ziskind [WZ89]).

Cho and Djuric [CD94] approached the joint detection and estimation problem using a technique they called Bayesian predictive densities (e.g., [Dju90]). Their result contains the same data term as the AML estimator but has a different penalty term.

Ottersten et al. [OVSN93] utilize a generalized likelihood ratio test with a different model for the two hypotheses than was used in Section 7.8.1. They obtain a sequential hypothesis test that utilizes the AML estimates at each candidate value of d .

All of the joint detection and estimation techniques have significant computational complexity and do not appear to be widely used in practical applications.

We defer an analysis of this problem until Chapter 9, where we compare the performance of a number of algorithms for the unknown D case.

8.9 Spatially Spread Signals

In this section we discuss parameter estimation for spatially spread signals. In Section 8.9.1, we consider the model that we developed in Section 5.3. Here the source signal is characterized by a spectrum distributed on the surface of a large sphere whose distribution is specified by $S_0(\omega : \theta, \phi)$. We parameterize the spatial spectrum and find the ML estimate of the parameters. We also derive the corresponding CRB.

In Section 8.9.2, we consider linear arrays and assume that the spatial spectrum can be modeled as a spatial ARMA process. We discuss the ML estimate of the ARMA parameters and the corresponding CRB.

In Section 8.9.3, we summarize our discussion.

8.9.1 Parameterized $S(\theta, \phi)$

In this section we consider the model described in Section 5.3.4. For numerical simplicity, we first consider the case of a standard linear array along the z -axis²⁴. From (5.145), the 1-D signal spectrum is

$$S_s(\omega : \Delta p_z) = \int_0^\pi d\theta \frac{\sin \theta}{2} e^{-jk_0 \Delta p_z \cos \theta} \bar{S}_0(\omega : \theta), \quad (8.546)$$

where

$$\bar{S}_0(\omega : \theta) \triangleq \frac{1}{2\pi} \int_0^{2\pi} S_0(\omega : \theta, \phi) d\phi. \quad (8.547)$$

Letting

$$\psi = \pi \cos \theta, \quad (8.548)$$

$$\Delta p_z = (m - n) \frac{\lambda}{2}, \quad (8.549)$$

and

$$S_f(\psi) = \bar{S}_0(\omega : \theta)|_{\psi=\pi \cos \theta}, \quad (8.550)$$

we can write the elements in the array spectral matrix as

$$[\mathbf{S}_f]_{mn} = \frac{1}{2\pi} \int_{-\pi}^{\pi} d\psi e^{-j(m-n)\psi} S_f(\psi). \quad (8.551)$$

At this point we must specify $S_f(\psi)$ in order to proceed. We consider two simple examples²⁵ in which we can evaluate the integral in (8.551) analytically. In other cases, we can calculate \mathbf{S}_f numerically.

²⁴The ML part of this discussion follows Meng et al. [MWW93].

²⁵The two examples are taken from [MWW93].

Example 8.9.1

Assume

$$S_f(\psi) = \frac{\sigma_f^2}{\pi} \frac{\sigma_\psi}{\sigma_\psi^2 + (\psi - \psi_0)^2}. \quad (8.552)$$

The parameter ψ_0 is the mean value of the angle of arrival and σ_ψ^2 controls the width. Assuming $\sigma_\psi \ll \pi$, we use (8.552) in (8.551) and recognize a familiar Fourier transform pair, to obtain,

$$\mathbf{S}_f = \frac{\sigma_f^2}{2\pi^2} \begin{bmatrix} 1 & z_\psi^H & \cdots & (z_\psi^H)^{N-1} \\ z_\psi & 1 & \cdots & (z_\psi^H)^{N-2} \\ \vdots & \vdots & \ddots & \vdots \\ z_\psi^{N-1} & z_\psi^{N-2} & \cdots & 1 \end{bmatrix} \triangleq \frac{\sigma_f^2}{2\pi^2} \Sigma_S, \quad (8.553)$$

where

$$z_\psi \triangleq \exp[-\sigma_\psi + j\psi_0]. \quad (8.554)$$

In the second example, we assume that $S_f(\psi)$ has a Gaussian shape central at ψ_0 with standard deviation of σ_ψ .²⁶

Example 8.9.2

Assume

$$S_f(\psi) = \frac{\sigma_f^2}{\sqrt{2\pi}\sigma_\psi^2} \exp\left(-\frac{(\psi - \psi_0)^2}{2\sigma_\psi^2}\right), \quad (8.555)$$

where $\sigma_\psi \ll \pi$. Using (8.555) in (8.551) and recognizing the familiar Fourier transform pair, gives

$$[\mathbf{S}_f]_{mn} = \frac{\sigma_f^2}{2\pi} \exp\left\{-\left(m-n\right)^2 \frac{\sigma_\psi^2}{2} + j\psi_0(m-n)\right\}. \quad (8.556)$$

For D independent signals, we have

$$\mathbf{S}_f = \sum_{i=1}^D \mathbf{S}_{f_i}, \quad (8.557)$$

where the \mathbf{S}_{f_i} are given by (8.553) or (8.556).

A particular case that we will study is the two-signal case in which

$$S_{f_1} = \frac{\sigma_{f_1}^2}{\sqrt{2\pi}\sigma_1^2} \exp\left(-\frac{(\psi - \psi_{c1})^2}{2\sigma_1^2}\right), \quad (8.558)$$

and

²⁶This model is applicable in a number of wireless communication applications.

$$S_{f_2} = \frac{\sigma_{f_2}^2}{\sqrt{2\pi\sigma_2^2}} \exp\left(-\frac{(\psi - \psi_{c2})^2}{2\sigma_2^2}\right). \quad (8.559)$$

The spectral matrix of the total input is

$$\mathbf{S}_x = \mathbf{S}_f + \sigma_w^2 \mathbf{I}. \quad (8.560)$$

We can now use (8.560) directly in the ln likelihood function to find the ML estimate and the CRB.

We first consider the CRB. For simplicity, we write the result using the notation of our two examples.

We define ψ_0 as the $D \times 1$ parameter vector denoting the center of each $S_{f_i}(\psi)$,

$$\psi_0 \triangleq \begin{bmatrix} \psi_{c1} & \psi_{c2} & \cdots & \psi_{cD} \end{bmatrix}^T. \quad (8.561)$$

The vector σ_0^2 denotes the spread of the $S_{f_i}(\psi)$,

$$\sigma_0^2 \triangleq \begin{bmatrix} \sigma_{\psi 1}^2 & \sigma_{\psi 2}^2 & \cdots & \sigma_{\psi D}^2 \end{bmatrix}^T. \quad (8.562)$$

The σ_f^2 denotes the power in each $S_{f_i}(\psi)$,

$$\sigma_f^2 \triangleq \begin{bmatrix} \sigma_{f_1}^2 & \sigma_{f_2}^2 & \cdots & \sigma_{f_D}^2 \end{bmatrix}^T. \quad (8.563)$$

Assuming σ_w^2 is unknown, we have a $(3D+1)$ parameter estimation problem.

To evaluate the CRB, we write the information matrix in a partitioned form

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{\psi_0 \psi_0} & \mathbf{J}_{\psi_0 \sigma_\psi^2} & \mathbf{J}_{\psi_0 \sigma_s^2} & \mathbf{J}_{\psi_0 \sigma_w^2} \\ \hline \mathbf{J}_{\sigma_\psi^2 \psi_0} & \mathbf{J}_{\sigma_\psi^2 \sigma_\psi^2} & \mathbf{J}_{\sigma_\psi^2 \sigma_s^2} & \mathbf{J}_{\sigma_\psi^2 \sigma_w^2} \\ \hline \mathbf{J}_{\sigma_s^2 \psi_0} & \mathbf{J}_{\sigma_s^2 \sigma_\psi^2} & \mathbf{J}_{\sigma_s^2 \sigma_s^2} & \mathbf{J}_{\sigma_s^2 \sigma_w^2} \\ \hline \mathbf{J}_{\sigma_w^2 \psi_0} & \mathbf{J}_{\sigma_w^2 \sigma_\psi^2} & \mathbf{J}_{\sigma_w^2 \sigma_s^2} & \mathbf{J}_{\sigma_w^2 \sigma_w^2} \end{bmatrix}, \quad (8.564)$$

where each term is given by (8.35),

$$\mathbf{J}_{ij} = \text{tr} \left[\frac{\partial \mathbf{S}_x}{\partial \theta_i} \mathbf{S}_x^{-1} \frac{\partial \mathbf{S}_x}{\partial \theta_j} \mathbf{S}_x^{-1} \right], \quad (8.565)$$

where θ_i denotes the appropriate parameter. To evaluate (8.564), we utilize (8.553) or (8.556) in (8.560) and calculate the derivatives in (8.565).

We consider two simple examples to illustrate the technique.

Example 8.9.3 (continuation)

Consider a standard 10-element linear array. There is a single spread signal impinging on the array. The spectral matrix is given by (8.556). We first consider the case in which σ_ψ^2 , σ_f^2 , and σ_w^2 are known. This result provides a bound on more realistic models. In this case, \mathbf{J} is a scalar. To evaluate it, we differentiate (8.556). The result is

$$\frac{\partial}{\partial \psi_0} \{ [\mathbf{S}_r]_{mn} \} = j(m-n) \cdot \frac{\sigma_f^2}{2\pi} \exp \left\{ -(m-n)^2 \frac{\sigma_\psi^2}{2} + j\psi_0(m-n) \right\}. \quad (8.566)$$

We substitute (8.566) and the inverse of (8.556) into (8.555) to obtain $J_{\psi_0 \psi_0}$. Then, the CRB(ψ_0) is the reciprocal of it. We plot the normalized bound in Figure 8.41 for various values of

$$\sigma_u = \sigma_\psi / \pi. \quad (8.567)$$

We also show the conventional CRB ($\sigma_u = 0$). We see that the bound increases significantly as σ_u increases.

The next step is to consider the case when σ_ψ^2 , σ_s^2 , and σ_w^2 are unknown. The resulting cross-matrices are all zero, so the bound in Example 8.9.3 also applies to this case.

We next consider the case of two spatially spread signals.

Example 8.9.4 (continuation)

Consider the case in which there are two uncorrelated spread signals impinging on a standard 10-element linear array. We use the signal model in (8.555) for each signal. For the case of known signal and noise power, there are four parameters to estimate: ψ_{c1} , ψ_{c2} , $\sigma_{\psi_1}^2$, and $\sigma_{\psi_2}^2$. If the signal powers and noise power are unknown, then there are three additional parameters: $\sigma_{f_1}^2$, $\sigma_{f_2}^2$, and σ_w^2 .

We assume that the $\psi_{c2} = -\psi_{c1}$ and that $\Delta\psi = \Delta\psi_R$. We also assume that $\sigma_{\psi_2}^2 = \sigma_{\psi_1}^2$. In Figure 8.42, we plot the bound on $\text{var}[\psi_{c1}]$ versus ASNR for various σ_{ψ_1} for the four parameter case. In Figure 8.43, we plot the bound on $\text{var}[\psi_{c1}]$ for the seven-parameter case.

The parameters are coupled so the CRB for the seven parameter case is higher than in the four-parameter case.

In order to find the ML estimate, we must conduct a search over the parameter space.

Using (8.13), we can write the ln likelihood function as

$$L(\psi_0, \sigma_\psi^2, \sigma_f^2, \sigma_w^2) = -K \left\{ \ln |\mathbf{S}_x| + \text{tr} \left[\mathbf{S}_x^{-1} \mathbf{C}_x \right] \right\}, \quad (8.568)$$

where $\psi_0, \sigma_\psi^2, \sigma_f^2$ are $D \times 1$ vectors containing the parameters of each of the D spread signals.

Then,

$$\hat{\theta}_{ml} = \arg \min_{\boldsymbol{\theta}} \left\{ \ln |\mathbf{S}_x| + \text{tr} \left[\mathbf{S}_x^{-1} \mathbf{C}_x \right] \right\}, \quad (8.569)$$

where $\boldsymbol{\theta}$ is a $(3D + 1) \times 1$ parameter vector.

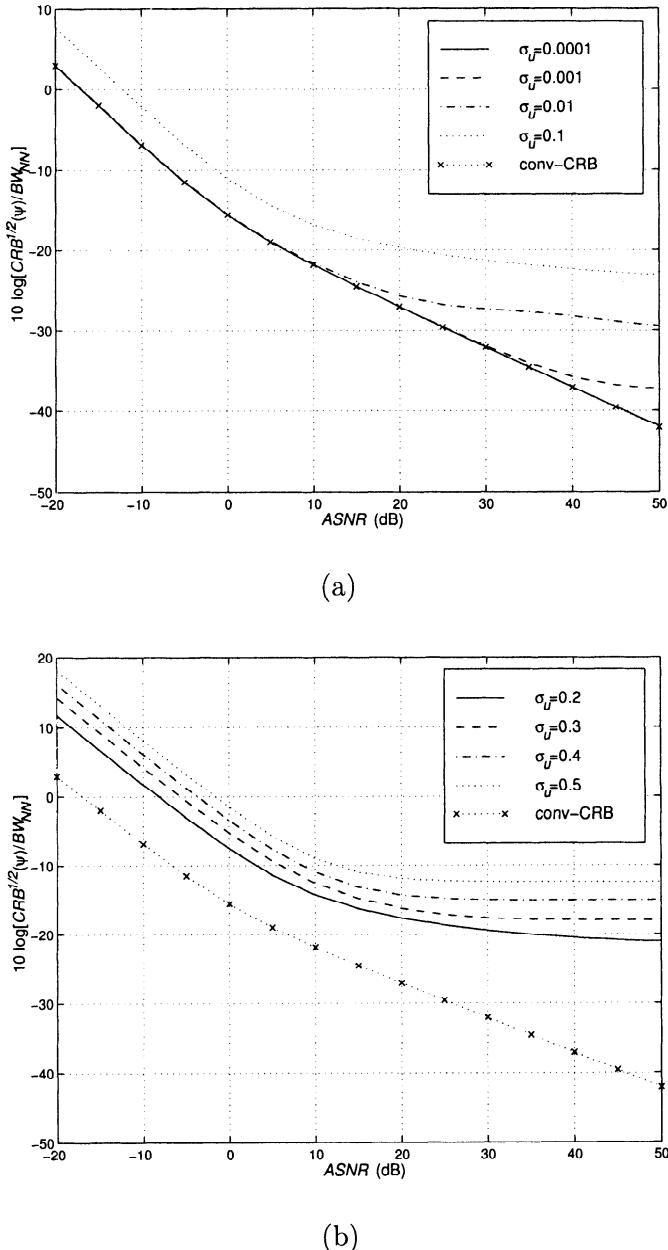


Figure 8.41 Normalized CRB on ψ for single spread signal versus ASNR: $N = 10, K = 100, \psi_s = 0$: (a) values of σ_u from 10^{-4} to 0.1; (b) values of σ_u for 0.2 to 0.5.

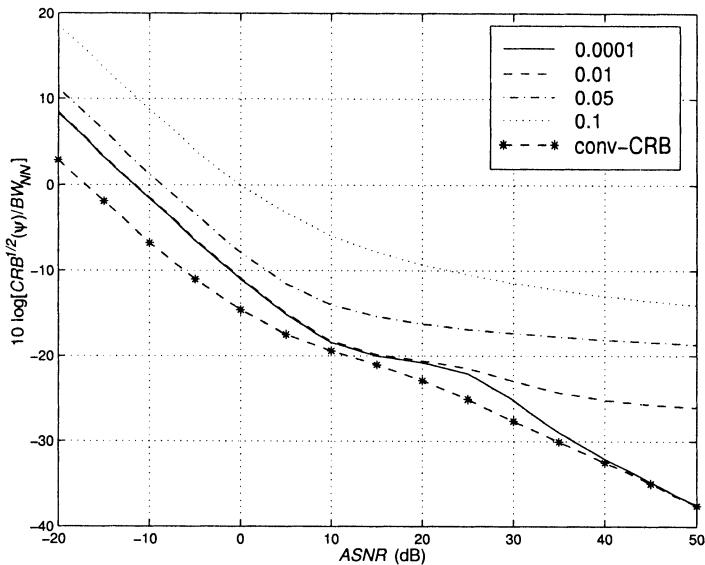


Figure 8.42 Normalized CRB on ψ_{c1} and ψ_{c2} for two spatially spread signals: $N = 10, K = 100, \psi_{c1} = 0.5\Delta_R, \psi_{c2} = -0.5\Delta_R, \sigma_{\psi_1}^2$ and $\sigma_{\psi_2}^2$ are unknown, $\sigma_{f1}^2, \sigma_{f2}^2$, and σ_w^2 are known.

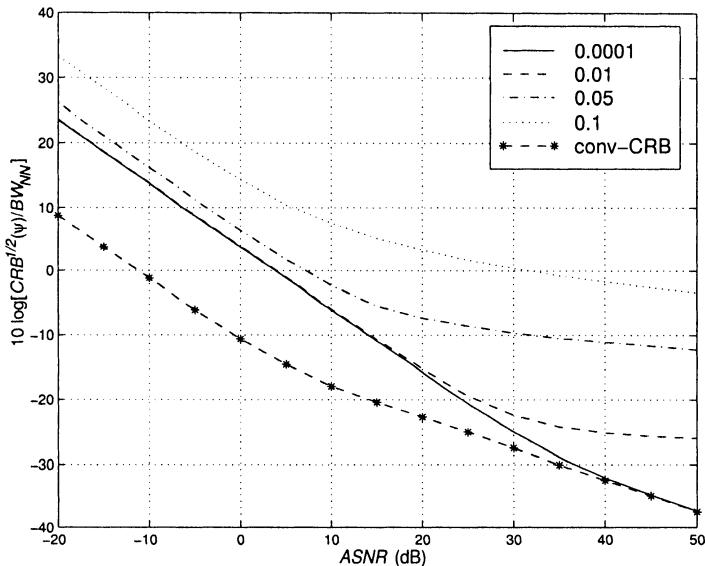


Figure 8.43 Normalized CRB on ψ_{c1} and ψ_{c2} for two spatially spread signals: $N = 10, K = 100, \psi_{c1} = 0.5\Delta_R, \psi_{c2} = -0.5\Delta_R$, all seven parameters are unknown.

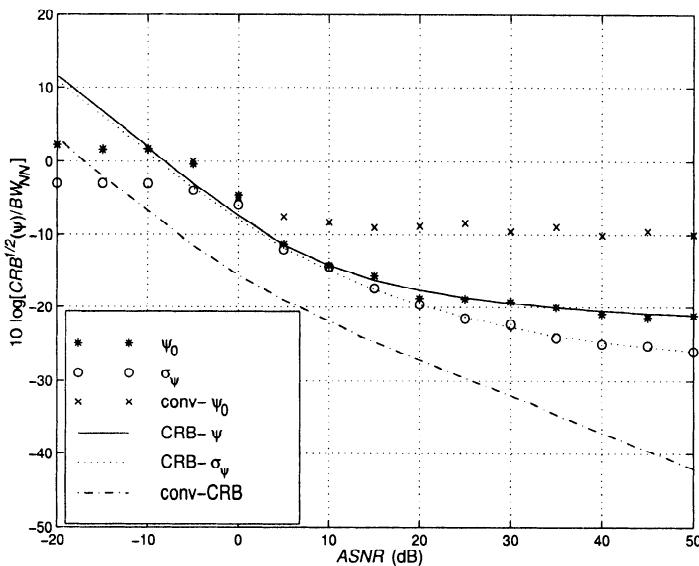


Figure 8.44 ML estimator of ψ_0 and σ_ψ : $N = 10, K = 100, \psi_0 = 0.0433\pi, \sigma_\psi = 0.2\pi, \sigma_s^2$ and σ_w^2 known; normalized RMSE versus ASNR.

In order to keep the computations manageable we consider the case of a single signal with ψ_0 and σ_u unknown, but where σ_f^2 and σ_w^2 are known.

Example 8.9.5 (continuation, Example 8.9.3)

Consider a single spatially spread signal impinging on a standard 10-element linear array. The signal component of the spatial spectral matrix is given by (8.556). We assume that σ_f^2 and σ_w^2 are known and find the ML estimate of ψ_0 and σ_ψ .

The normalized RMSE is plotted versus ASNR in Figure 8.44. We also plot the CRB, the normalized RMSE of the conventional ML algorithm (assumes a plane-wave signal), and the conventional CRB.

We see that RMSE of both estimates, $\hat{\psi}_0$ and $\hat{\sigma}_\psi$, coincide with the CRB above ASNR = 5 dB. This threshold is about 7 dB above the single plane-wave case in Figure 8.17. We see that the conventional ML estimator is not effective.

For multiple signals, the computational complexity makes ML impractical. In Chapter 9, we develop simpler estimation algorithms. Some of these algorithms can be modified to accommodate this spatially spread model, but we will not discuss the necessary modifications in the text.

8.9.2 Spatial ARMA Process

If we restrict our attention to an SLA, then we can model a spatially spread signal as an auto-regressive (AR) process and the total input including the white noise component can be modeled as an ARMA process.

We developed models for these processes in Section 5.7. We can estimate the parameters of the spread signal using various techniques. This approach is referred to in the literature as parametric spectral estimation and has been widely studied and applied. Our model is the spatial version, but the results carry over directly. We refer the reader to several references for a complete discussion.

The article by Kay and Marple [KM81] followed by their comprehensive books, Kay [Kay88] and Marple [Mar87], contains a discussion of ARMA processes. The book by Stoica and Moses [SM97] discusses ARMA models and the associated CRBs. There is a sequence of papers that compute CRBs for AR or ARMA models (cf. Friedlander and Porat [FP89], Anderson [And71], Porat and Friedlander [PF86], and Friedlander [Fri84]).

8.9.3 Summary

In this section we have introduced the problem of estimating the parameters of a spatially spread signal. The CRBs quantify the potential performance degradation caused by the spreading. The ML estimators approach the CRB in simple cases, but are computationally prohibitive in the multiple-signal environment.

In the single spread signal example, we saw that the conventional ML estimators that ignored the spreading did not perform well. The same result is true for the multiple-signal environment.

8.10 Beamspace algorithms

8.10.1 Introduction

In Sections 6.9 and 7.10, we saw the advantages of beamspace processing in the context of adaptive beamforming. In this section, we show its application in the parameter estimation problem.

A number of authors have discussed the advantages of beamspace processing for parameter estimation. References include Bienvenu and Kopp [BK84], Gray [Gra82], Forster and Vezzosi [FV87], Van Veen and Williams [VVW88a], [VVW88b], Lee and Wengrovitz [LW88], [LW90], Buckley and Xu [XB88], [XB89], [BX90], and Zoltowski [Zol88].

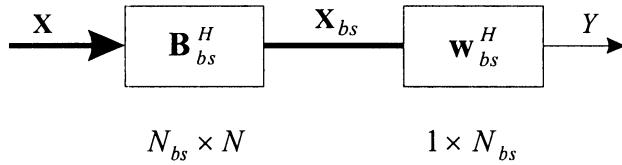


Figure 8.45 Beamspace processing.

The beamspace model was shown in Figure 6.81 and is repeated in Figure 8.45 for convenience. The output of the beamspace matrix is,

$$\mathbf{X}_{bs} = \mathbf{B}_{bs}^H \mathbf{X}, \quad (8.570)$$

where \mathbf{B}_{bs}^H is an $N_{bs} \times N$ matrix. We process \mathbf{X}_{bs} to estimate the parameters of interest.

As in the beamformer problem, there are several advantages:

- (i) The dimension of the beamspace N_{bs} is usually much smaller than N so the computational complexity is reduced and the statistical stability of the estimate of the spatial spectral matrix is improved.
- (ii) If there are strong interfering signals that are not in the same sector as the signals of interest, we may be able to null them out prior to implementing our parameter estimation algorithm.
- (iii) Most of our algorithms have assumed additional white Gaussian noise (or Gaussian noise with a known correlation function that we could pre-whiten). In Figure 8.46 we show a representative non-white spatial noise spectrum that we assume is unknown. We indicate how we have divided ψ -space into six sectors by using six parallel beamspace processors. In each of these sectors we can model the noise as white with an unknown variance and use the beamspace version of algorithms that we have developed.

On the other hand, the disadvantage is that the performance of the algorithm after pre-processing may be poorer. If we choose \mathbf{B}_{bs} improperly, then we may lose information that degrades the performance of the parameter estimation algorithm. We would like to show that the output of the beamspace processor is a sufficient statistic for the estimation problem of interest. A simpler approach is to compare the beamspace CRB to the element-space CRB. We find that the CRB on the asymptotic variance of the beamspace

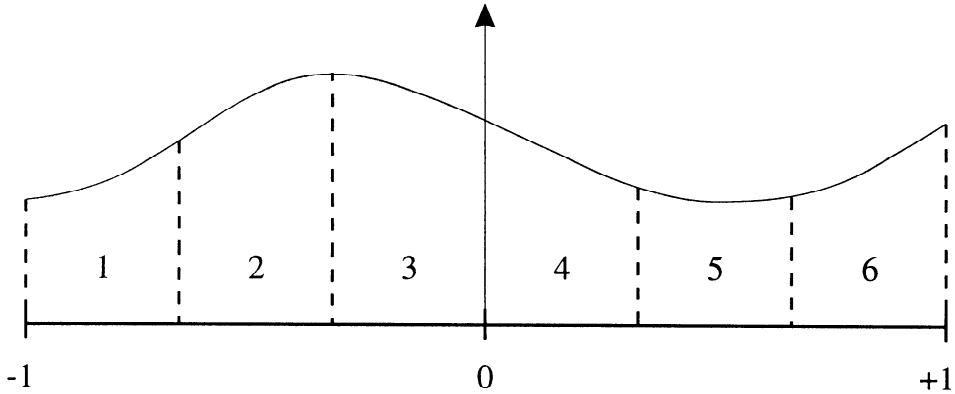


Figure 8.46 Beamspace processing for non-white spatial noise spectrum: six beam sectors.

estimator is always greater than or equal to the CRB bound in element space. However, we find that, with an appropriate choice of \mathbf{B}_{bs} , the probability of resolution can be improved.

The model of interest is the familiar snapshot model,

$$\mathbf{X}(k) = \mathbf{V} \mathbf{F}(k) + \mathbf{N}(k), \quad k = 1, 2, \dots, K, \quad (8.571)$$

where $\mathbf{X}(k)$ is an $N \times 1$ vector, $\mathbf{F}(k)$ is the $D \times 1$ signal in space vector, $\mathbf{N}(k)$ is an $N \times 1$ white Gaussian noise vector, and \mathbf{V} is the array manifold matrix,

$$\mathbf{V} = \left[\mathbf{v}_1 \mid \mathbf{v}_2 \mid \dots \mid \mathbf{v}_D \right], \quad (8.572)$$

where

$$\mathbf{v}_n \triangleq \mathbf{v}(\psi_n). \quad (8.573)$$

We use the Gaussian signal model. The signal vectors are independent samples of the stationary zero-mean complex Gaussian random process with unknown spectral matrix \mathbf{S}_f . The noise vectors are independent samples of a stationary zero-mean complex Gaussian random process with spectral matrix $\sigma_w^2 \mathbf{I}$ where σ_w^2 is unknown.

The spectral matrix of the sampled input vector $\mathbf{X}(k)$ is

$$\begin{aligned} \mathbf{S}_x &\triangleq E[\mathbf{X}(k) \mathbf{X}^H(k)] \\ &= \mathbf{V} \mathbf{S}_f \mathbf{V}^H + \sigma_w^2 \mathbf{I}, \end{aligned} \quad (8.574)$$

where $\psi_1, \psi_2, \dots, \psi_D, \mathbf{S}_f$, and σ_w^2 are unknown.

We define an $N_{bs} \times N$ matrix \mathbf{B}_{bs}^H , which is used to pre-process the data,

$$\mathbf{X}_{bs}(k) = \mathbf{B}_{bs}^H \mathbf{X}(k). \quad (8.575)$$

We assume

$$N_{bs} \leq N \quad (8.576)$$

and

$$N_{bs} \geq D + 1. \quad (8.577)$$

We also assume that the columns of \mathbf{B} are orthonormal so that

$$\mathbf{B}_{bs}^H \mathbf{B}_{bs} = \mathbf{I}. \quad (8.578)$$

In some applications, we may start with a pre-processing matrix \mathbf{B}_{no} that does not satisfy (8.578). We then construct \mathbf{B}_{bs} using the transformation.

$$\mathbf{B}_{bs} = \mathbf{B}_{no} \left(\mathbf{B}_{no}^H \mathbf{B}_{no} \right)^{-\frac{1}{2}}. \quad (8.579)$$

The choice of $\left(\mathbf{B}_{no}^H \mathbf{B}_{no} \right)^{-\frac{1}{2}}$ is available. One choice is the inverse of the Cholesky factorization of $\mathbf{B}_{no}^H \mathbf{B}_{no}$. A second choice is the Cholesky factorization of the inverse of $(\mathbf{B}_{no}^H \mathbf{B}_{no})$. We find that different choices may result in different performance.

The spectral matrix of \mathbf{X}_{bs} is

$$\mathbf{S}_{\mathbf{X}_{bs}} = \mathbf{B}_{bs}^H \mathbf{V} \mathbf{S}_f \mathbf{V}^H \mathbf{B}_{bs} + \sigma_w^2 \mathbf{I}. \quad (8.580)$$

We operate on \mathbf{X}_{bs} to estimate the various parameters.

In Section 8.10.2, we discuss various beamspace matrices. In Section 8.10.3, we derive the beamspace CRB and derive conditions on \mathbf{B}_{bs} under which the beamspace CRB is equal to the element-space CRB. In Section 8.10.4, we derive the beamspace ML estimator and study its performance. In Section 8.10.5, we summarize our results.

8.10.2 Beamspace Matrices

In this section we discuss various beamspace matrices. We encountered most of these matrices in earlier discussions (e.g., Sections 3.10, 6.9, or 7.10), but we need to revisit them in the context of DOA estimation. For notational simplicity, we consider a standard linear array in the text.

8.10.2.1 Conventional beams (DFT)

The most common beamspace matrix has rows that consist of conventional beams whose pointing directions are spaced at $2\pi/N$ intervals in ψ -space. In Figure 8.47, we show a 7-beam fan, and in Figure 8.48, we show a 6-beam fan.

The rows of \mathbf{B}_{bs}^H are given by

$$[\mathbf{B}_{bs}^H]_m = \frac{1}{N} e^{j(\frac{N-1}{2})m\frac{2\pi}{N}} [1 \quad e^{-jm\frac{2\pi}{N}} \quad \dots \quad e^{-j(N-1)m\frac{2\pi}{N}}], m \in \Omega_{N_{bs}}, \quad (8.581)$$

where $\Omega_{N_{bs}}$ is the set containing the values of m included in the beamspace.

In most cases, it is convenient to denote the value of m corresponding to the beam closest to $\psi = -\pi$ as m_L (the subscript “L” denotes left). Then, the range of m is $m_L \leq m \leq m_L + N_{bs}$. If the beam sector is near $\psi = \pi$ (endfire) then $m_L + N_{bs}$ may exceed N and the beams wrap around.

The beam formed by (8.581) corresponds to a conventional beam pointed at $\psi = 2\pi m/N$.

If the beamspace sector is centered at $\psi = 0$ and N_{bs} is odd, there is a center beam aimed at $\psi = 0$. If the beamspace sector is centered at $\psi = 0$ and N_{bs} is even, then m in (8.581) is replaced by $m' = m - \pi/N$ and the indexing limits are modified. Then, the two inner beams are aimed at $\pm\pi/N$ in ψ -space. This beamspace matrix is referred to as the DFT beamformer in the signal processing literature. It is referred to as the Butler beamformer in the classic antenna literature. The m th beam in ψ -space is (2.127),

$$b_m(\psi) = [\mathbf{B}_{bs}^H]_m \mathbf{v}(\psi) = \frac{1}{N} \frac{\sin[\frac{N}{2}(\psi - m\frac{2\pi}{N})]}{\sin[\frac{1}{2}(\psi - m\frac{2\pi}{N})]}, \quad m_L \leq m \leq m_L + N_{bs}. \quad (8.582)$$

The resulting beamspace array manifold vector is a real $N_{bs} \times 1$ vector,

$$\mathbf{v}_{bs}(\psi) = \mathbf{B}_{bs}^H \mathbf{v}(\psi) = [b_1(\psi) \cdots b_{N_{bs}}(\psi)]^T. \quad (8.583)$$

We also consider the case in which the conventional beams are placed closer together. In this case, they are not orthogonal. We write

$$\begin{aligned} [\mathbf{b}_{no,i}^H]_n &= \frac{1}{N} \left[e^{j(n-\frac{N-1}{2})(\left(i-(\frac{N_{bs}+1}{2})\right)\frac{2\pi\alpha}{N})} \right], \\ n &= 0, 1, \dots, N-1, \\ i &= 1, 2, \dots, N_{bs}, \end{aligned} \quad (8.584)$$

where $1 \leq \alpha \leq 2$. The beamspace matrix is orthogonalized using (8.579).

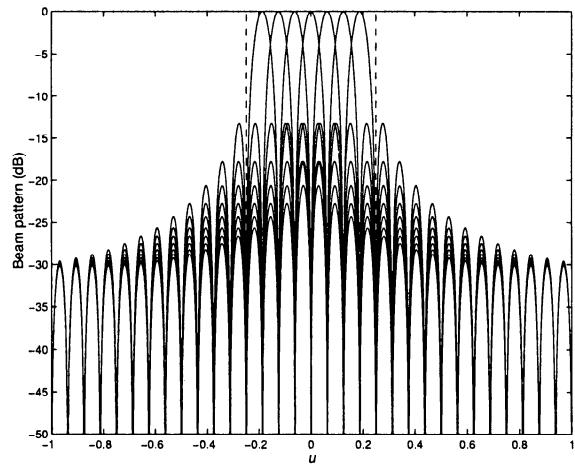


Figure 8.47 Seven-beam fan.

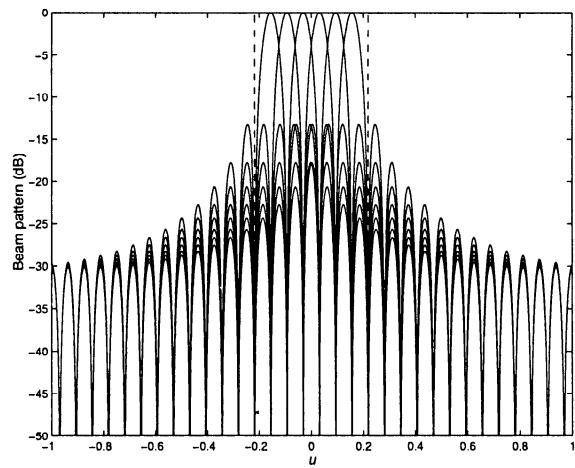


Figure 8.48 Six-beam fan.

We can also use beams with lower sidelobes such as Dolph-Chebychev or Hamming. This type of beamspace matrix was discussed in Section 3.10.

8.10.2.2 Taylor series beamspace

For the case of two plane-wave signals, we find that the beamspace matrix that provides good threshold performance is

$$\mathbf{B}_{no} = \left[\mathbf{v}(\psi_1) \mid \mathbf{v}(\psi_2) \mid \mathbf{v}(\psi_m) \right], \quad (8.585)$$

where $\mathbf{v}(\psi_i)$, $i = 1, 2$ is the steering vector corresponding to the true wavenumbers, ψ_1 and ψ_2 , of the two sources and $\psi_m = (\psi_1 + \psi_2)/2$ is the midpoint. However, we cannot construct the \mathbf{B}_{no} in (8.585) because ψ_1 and ψ_2 are the wavenumbers we are trying to estimate.

However, we can approximate the matrix in (8.585) by using a Taylor series expansion of the array manifold vector. We can do an initial beamscan on the sensor output and determine the areas in u -space that appear to contain one or more signals. We denote the midpoint of one of these areas (the peak in the beamscan) as $\tilde{\psi}_m$ and use a Taylor series expansion around $\tilde{\psi}_m$:

$$\begin{aligned} \mathbf{v}(\psi_i) &= \mathbf{v}(\tilde{\psi}_m) + \mathbf{v}^{(1)}(\tilde{\psi}_m)(\psi_i - \tilde{\psi}_m) \\ &\quad + \mathbf{v}^{(2)}(\tilde{\psi}_m) \frac{(\psi_i - \tilde{\psi}_m)^2}{2!} + \mathbf{v}^{(3)}(\tilde{\psi}_m) \frac{(\psi_i - \tilde{\psi}_m)^3}{3!} + \dots, \end{aligned} \quad (8.586)$$

where $i = 1, 2, \dots, m$ and

$$\mathbf{v}^{(k)}(\tilde{\psi}_m) = \left. \frac{\partial^k \mathbf{v}(\psi)}{\partial^k \psi} \right|_{\psi=\tilde{\psi}_m}, \quad (8.587)$$

is the k th derivative of the array manifold with respect to ψ evaluated at $\psi = \tilde{\psi}_m$. Then,

$$\mathbf{B}_{no} = \left[\mathbf{v}(\tilde{\psi}_m) \mid \mathbf{v}^{(1)}(\tilde{\psi}_m) \mid \mathbf{v}^{(2)}(\tilde{\psi}_m) \mid \dots \mid \mathbf{v}^{(m)}(\tilde{\psi}_m) \right], \quad (8.588)$$

and

$$\mathbf{B}_{bs} = \mathbf{B}_{no} \left[\mathbf{B}_{no}^H \mathbf{B}_{no} \right]^{-\frac{1}{2}}. \quad (8.589)$$

We consider an example to illustrate the behavior of this beamspace processor.

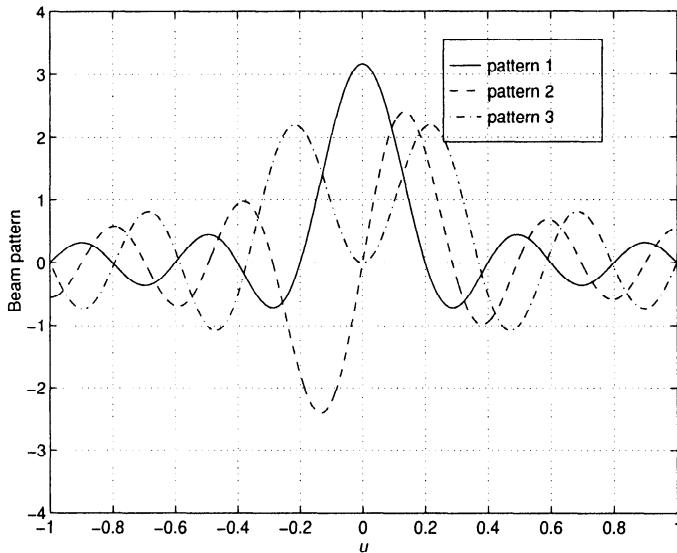


Figure 8.49 Beam patterns for Taylor series beamspace matrix, $N = 10$, $N_{bs} = 3$.

Example 8.10.1

Consider a standard 10-element linear array. Assume that the beamspace is centered at $\tilde{\psi}_m = 0$. If we use a 3-D beamspace, then \mathbf{B}_{no} is a 10×3 matrix with a Vandermonde row structure

$$\mathbf{B}_{no} = \begin{bmatrix} 1 & 0 & 0 \\ 1 & j & (j)^2 \\ \vdots & \vdots & \vdots \\ 1 & (nj) & (nj)^2 \\ \vdots & \vdots & \vdots \\ 1 & (9j) & (9j)^2 \end{bmatrix}. \quad (8.590)$$

We compute \mathbf{B}_{bs} using (8.589) and plot the corresponding beam patterns in Figure 8.49.

8.10.2.3 Discrete prolate spheroidal sequences (DPSS beamspace)

We have encountered DPSS in Section 3.1 (3.25)–(3.35). The application to beamspace processing is due to Forster and Vezzosi [FV87].

We define \mathbf{b}_i , $i = 1, 2, \dots, M$ to be the columns of the beamforming matrix \mathbf{B}_{bs} . We require the columns to be orthonormal. Then, as in (3.26),

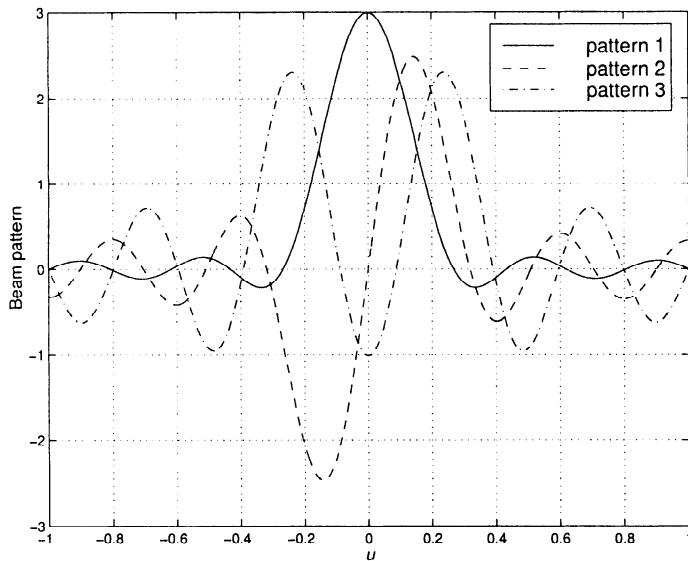


Figure 8.50 Beam patterns for DPSS beamspace matrix: $N = 10$, $\psi_0 = 0.2\pi$, $N_{bs} = 3$.

we define

$$\alpha_i = \frac{\int_{-\psi_0}^{\psi_0} |\mathbf{b}_i \mathbf{v}(\psi)|^2 d\psi}{\int_{-\pi}^{\pi} |\mathbf{b}_i \mathbf{v}(\psi)|^2 d\psi}, \quad i = 1, 2, \dots, M, \quad (8.591)$$

which is the ratio of the energy of the i th beam in $[-\psi_0, \psi_0]$ to the energy of the i th beam in $[-\pi, \pi]$. The numerator is

$$\alpha_{iN} = \mathbf{b}_i \mathbf{A} \mathbf{b}_i^H, \quad (8.592)$$

where

$$\mathbf{A} \triangleq \int_{-\psi_0}^{\psi_0} \mathbf{v}(\psi) \mathbf{v}^H(\psi) d\psi. \quad (8.593)$$

For a linear array, the mn element is

$$[\mathbf{A}]_{mn} = \frac{2 \sin [(m-n)\psi_0]}{(m-n)}, \quad m \neq n, \quad (8.594)$$

and

$$[\mathbf{A}]_{mn} = 2\psi_0, \quad m = n. \quad (8.595)$$

The denominator is

$$\alpha_{iD} = 2\pi \mathbf{b}_i^H \mathbf{b}_i. \quad (8.596)$$

Thus,

$$\alpha_i = \frac{\mathbf{b}_i^H \mathbf{A} \mathbf{b}_i^H}{2\pi \mathbf{b}_i^H \mathbf{b}_i}, \quad i = 1, 2, \dots, N_{se}. \quad (8.597)$$

We want to maximize $\alpha_i, i = 1, 2, \dots, N_{se}$ subject to the orthonormality constraint. This corresponds to finding the eigenvectors of the matrix \mathbf{A} , which corresponds to the M largest eigenvalues. Thus, we solve

$$2\pi\lambda \mathbf{b}_i = \mathbf{A} \mathbf{b}_i. \quad (8.598)$$

Using (8.594) in (8.598), this corresponds to

$$\sum_{n=1}^N \frac{\sin(m-n)\psi_0}{(m-n)} b_n = \pi\lambda b_m, \quad m = 1, 2, \dots, N. \quad (8.599)$$

For each of the M largest eigenvalues, we obtain a sequence that defines the column vector \mathbf{b}_i . These sequences are called DPSSs and are discussed in detail by Slepian [Sle78].

The number of significant eigenvalues is

$$N_{se} = \frac{\psi_0}{\pi} N + 1. \quad (8.600)$$

The corresponding spheroidal functions are

$$G_i(\psi) = c_i \sum_{n=1}^N b_n^{(i)} e^{j\psi n}, \quad (8.601)$$

where

$$c_i = \begin{cases} 1, & i \text{ odd}, \\ j, & i \text{ even}. \end{cases} \quad (8.602)$$

The spheroidal functions are real functions that are doubly orthonormal,

$$\int_{-\pi}^{\pi} G_i(\psi) G_j(\psi) d\psi = \delta_{ij}, \quad (8.603)$$

and

$$\int_{-\psi_0}^{\psi_0} G_i(\psi) G_j(\psi) d\psi = \lambda_i \delta_{ij}. \quad (8.604)$$

The spheroidal functions correspond to beam patterns in the following manner,

$$B_i(\psi) = \sum_{n=1}^N b_n^{(i)} e^{j\psi n}. \quad (8.605)$$

Thus,

$$G_i(\psi) = c_i B_i(\psi). \quad (8.606)$$

The odd-numbered beams are real and symmetric and can be considered sum beams. The even-numbered beams are odd and asymmetric and can be considered difference beams.

Thus, the beamspace processor determines ψ_0 from prior knowledge of the target environment or by a preliminary processing with one of the simpler algorithms that are derived in Chapter 9. The value of ψ_0 determines N_{se} from (8.600). We find the \mathbf{b}_i as the eigenvectors corresponding to the N_{se} largest eigenvalues of (8.598). We use these $\mathbf{b}_i, i = 1, 2, \dots, N_{se}$, as the columns of \mathbf{B} . In Figure 8.50, we plot the first three eigenbeams for $u_0 = 0.2$.

We find that this spheroidal function decomposition is an effective beamspace matrix.

8.10.2.4 Conjugate symmetric beamspace matrices

In conjugate symmetric arrays we can use FB averaging of the data to estimate \mathbf{S}_x . If, in addition, the columns in the \mathbf{B}_{bs} matrix are conjugate symmetric or conjugate asymmetric,

$$\mathbf{b}_{bs,i} = \mathbf{J} \mathbf{b}_{bs,i}^*, \quad i = 1, \dots, N_{bs}, \quad (8.607)$$

or can be made to satisfy by a suitable choice of the origin, additional computational implications occurs.

When (8.607) is satisfied and we use FB averaging of the data, the subsequent processing can be done using real matrices. Alternatively, we can process the data without FB averaging and use the real part of the resulting matrix and achieve the same result. Utilizing real computations provides significant computational saving (e.g., Linebarger et al. [LDD94] and Zoltowski et al. [ZKS93]).

Note that we can use FB averaging even if (8.607) is not satisfied. However, we need to use complex processing in beamspace.

8.10.3 Beamspace Cramér-Rao Bound

8.10.3.1 Introduction

As in the element-space algorithms, we can analyze the behavior by finding the probability of resolution, the bias, and the variance. In this section, we develop the beamspace CRB and compare them the element-space CRB.

From our earlier discussions, we know that we can find algorithms (e.g., the unconditional ML estimate) that will approach the bounds. Thus, the ratio of the beamspace CRB to the element-space CRB will indicate the potential increase in variance by operating in beamspace. There will be an additional increase in variance if we use an algorithm that does not approach the CRB.

In Section 8.4, we developed the CRB for the element-space model. We now want to develop the CRB for the beamspace model and develop necessary and sufficient conditions on \mathbf{B} in order for the two bounds to be equal. Our discussion follows Weiss and Friedlander [WF94].²⁷ We consider the stochastic signal model. The deterministic signal model is also discussed in [WF94]. We then explore the behavior of beamspace CRB for the various pre-processing matrices in Section 8.10.2.

The model of interest is given by (8.571)–(8.580) in Section 8.10.1. We first define the following projection matrices,

$$\mathbf{P}_V \triangleq V \left(V^H V \right)^{-1} V^H, \quad (8.608)$$

and

$$\mathbf{P}_V^\perp \triangleq I - \mathbf{P}_V. \quad (8.609)$$

The beamspace steering vectors are

$$\mathbf{v}_B = \mathbf{B}_{bs}^H \mathbf{v}. \quad (8.610)$$

Then,

$$\mathbf{P}_{V_B} = \mathbf{B}_{bs}^H V \left[V^H \mathbf{P}_B V \right]^{-1} V^H \mathbf{B}_{bs}, \quad (8.611)$$

and

$$\mathbf{P}_{V_B}^\perp \triangleq I - \mathbf{P}_{V_B}. \quad (8.612)$$

The last projection matrix is,

$$\mathbf{P}_B \triangleq \mathbf{B}_{bs} \mathbf{B}_{bs}^H, \quad (8.613)$$

²⁷An earlier discussion of beamspace pre-processing matrices is contained in Anderson [And91] (see [And91], [And93], and [AN95]).

and \mathbf{B}_{bs} satisfies (8.578).

The CRB for the Gaussian signal model was given in Section 8.4.2.1 (see (8.95)).

$$\mathbf{C}_{CR}(\psi) = \frac{\sigma_w^2}{2K} \left\{ Re \left[\left(\mathbf{D}^H \mathbf{P}_V^\perp \mathbf{D} \right) \odot \left(\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f \right)^T \right] \right\}^{-1}. \quad (8.614)$$

The output of the beamspace matrix, $\mathbf{X}_{bs}(k)$, $k = 1, 2, \dots, K$ satisfies the same conditions as the original data sequence. The beamspace spectral matrix is

$$\mathbf{S}_{\mathbf{x}_{bs}} = \mathbf{B}_{bs}^H \mathbf{S}_x \mathbf{B}_{bs} = \mathbf{B}_{bs}^H \mathbf{V} \mathbf{S}_f \mathbf{V}^H \mathbf{B}_{bs} + \sigma_w^2 \mathbf{I}. \quad (8.615)$$

The beamspace noise is white due to (8.578).

Thus the bound for beamspace estimation is,

$$\boxed{\mathbf{C}_{CR,bs}(\psi) = \frac{\sigma_w^2}{2K} \left[Re \left\{ \left(\mathbf{D}^H \mathbf{B}_{bs} \mathbf{P}_{V_B}^\perp \mathbf{B}_{bs}^H \mathbf{D} \right) \odot \left(\mathbf{S}_f \mathbf{V}^H \mathbf{B}_{bs} \mathbf{S}_{\mathbf{x}_{bs}}^{-1} \mathbf{B}_{bs}^H \mathbf{V} \mathbf{S}_f \right)^T \right\} \right]^{-1}}. \quad (8.616)$$

Note that this bound corresponds to the general CRB and includes uncorrelated signals as a special case. If we know *a priori* that the signals are uncorrelated, we would adapt the bound in Section 8.4.2.

8.10.3.2 Beamspace matrix conditions

We now want to find conditions on \mathbf{B}_{bs} such that $\mathbf{C}_{CR,bs}(\psi) = \mathbf{C}_{CR}(\psi)$. Weiss and Friedlander [WF94] derive the following result:

Assuming the \mathbf{S}_f is positive definite, then²⁸

$$\mathbf{C}_{CR}(\psi) \leq \mathbf{C}_{CR,bs}(\psi), \quad (8.617)$$

with equality if and only if

$$\mathbf{P}_B \mathbf{V} = \mathbf{V}, \quad (8.618)$$

and

$$\mathbf{P}_B \mathbf{P}_V^\perp \mathbf{D} = \mathbf{P}_V^\perp \mathbf{D}. \quad (8.619)$$

The first condition (8.618) implies that we point the beams at the sources. The second condition implies that the column space of \mathbf{B}_{bs} includes the projection of \mathbf{d}_n , $n = 1, 2, \dots, D$ on the noise subspace \mathbf{P}_V^\perp . If \mathbf{B}_{bs} spans

²⁸Recall that the matrix inequality means that $\mathbf{C}_{CR,bs}(\psi) - \mathbf{C}_{CR}(\psi)$ is a non-negative definite matrix.

both \mathbf{V} and \mathbf{D} , it will span the required projections. One approach is to have D columns of \mathbf{B}_{bs} corresponding to the signal steering vectors,

$$\mathbf{b}_i = \mathbf{v}(\psi_i), i = 1, 2, \dots, D, \quad (8.620)$$

and D columns corresponding to the derivatives of the steering vectors,

$$\mathbf{d}_i = \mathbf{d}(\psi_i), i = D + 1, \dots, 2D. \quad (8.621)$$

A second approach is to approximate the derivatives by differences. This also requires two beams for each source. For example, if there is a source at ψ_n , we point one beam at ψ_n and a second beam at $\psi_n + \epsilon$, where ϵ is small enough that

$$\mathbf{d}_n \simeq \frac{1}{\epsilon} [\mathbf{v}(\psi_n + \epsilon) - \mathbf{v}(\psi_n)] \quad (8.622)$$

The obvious practical problem with this result is that we are trying to estimate the source locations, so that we cannot point the beams exactly at the sources. Two solutions appear logical.

If we have some prior knowledge of the signal DOAs and they are closely spaced, one of the \mathbf{B}_{bs} matrices described in Section 8.10.2 may be an adequate approximation. We investigate the behavior for several cases in subsequent examples.

If the performance using this technique is not adequate, we can develop a simple iterative scheme to obtain the required beamspace matrix.

We first consider some beamspace matrices that are similar to (8.620) and then examine the family of beamspace matrices that are most commonly used in practice.

The first example uses the clairvoyant beamspace matrix from (8.585).

Example 8.10.2

Consider a standard 10-element linear array. We first consider the clairvoyant beamspace matrix from (8.585). We consider two equal-power uncorrelated sources with $SNR = 10$ dB and use $K = 100$ snapshots.

We define

$$\mathbf{B}_{no} \triangleq [\mathbf{v}(\psi_1) \mid \mathbf{v}(\psi_2) \mid \mathbf{v}(\psi_m)], \quad (8.623)$$

and

$$\mathbf{B}_{bs} = \mathbf{B}_{no} (\mathbf{B}_{no}^H \mathbf{B}_{no})^{-\frac{1}{2}}. \quad (8.624)$$

In Figure 8.51, we show the beamspace CRB using (8.624) in (8.616) and the element-space CRB. We see that, if the signal separation is less than $0.7BW_{NN}$ ($\Delta u \leq 0.28$) that the beamspace matrix given by (8.624) does not increase the CRB.

Example 8.10.3 (continuation)

Consider the same model as in Example 8.10.2.

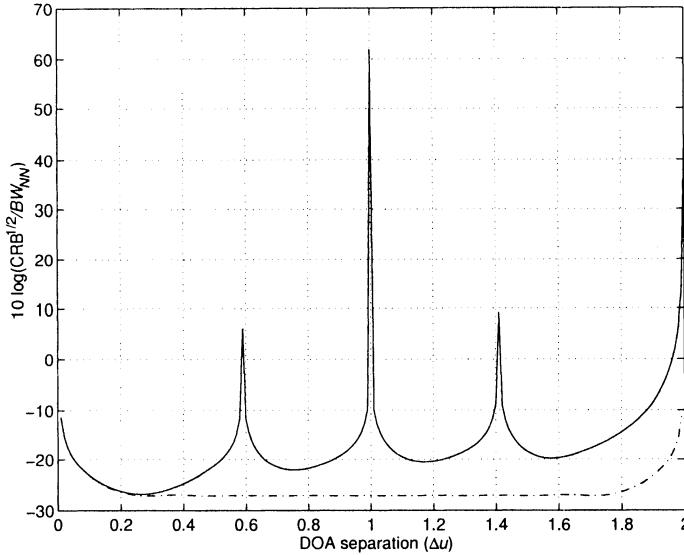


Figure 8.51 Normalized CRB for beamspace processing using clairvoyant beamspace matrix: $N = 10, N_{bs} = 3, K = 100, SNR = 20$ dB; element-space CRB shown as reference.

We next define an approximation to the \mathbf{B}_{no} in (8.623) by the first three terms

$$\mathbf{B}_{no} \triangleq [\mathbf{v}(\tilde{\psi}_m) \mid \dot{\mathbf{v}}(\tilde{\psi}_m) \mid \ddot{\mathbf{v}}(\tilde{\psi}_m)]. \quad (8.625)$$

Two cases are considered. In case 1, $\tilde{\psi}_m = \psi_m$. In case 2, $\tilde{\psi}_m = \psi_1$.

We substitute (8.625) into (8.624) to obtain \mathbf{B}_{bs} . The results from (8.616) are shown in Figure 8.52.

We see that the beamspace CRB coincides with the element-space CRB for $\Delta u \leq 0.20$ and is only slightly above it for $\Delta u \leq 0.30$.

Example 8.10.4 (continuation)

We consider the same model as in Example 8.10.2. We use \mathbf{B}_{no} as given by (8.588) which we repeat,

$$\mathbf{B}_{no} = [\mathbf{v}(\psi_0) \mid \mathbf{v}^{(1)}(\psi_0) \mid \mathbf{v}^{(2)}(\psi_0) \mid \cdots \mid \mathbf{v}^{(m)}(\psi_0)]. \quad (8.626)$$

We utilize \mathbf{B}_{bs} as given by (8.589) and evaluate the beamspace CRB as a function of m . The results are shown in Figure 8.53. Note that $N_{bs} = m + 1$, so there are $m + 1$ beams.

As we would expect, as m increases the beamspace CRB matches the element-space CRB for larger Δu . For $m = 5$, there is negligible difference for $\Delta u = 1$. Note, however, that this requires six beams from the 10-element array.

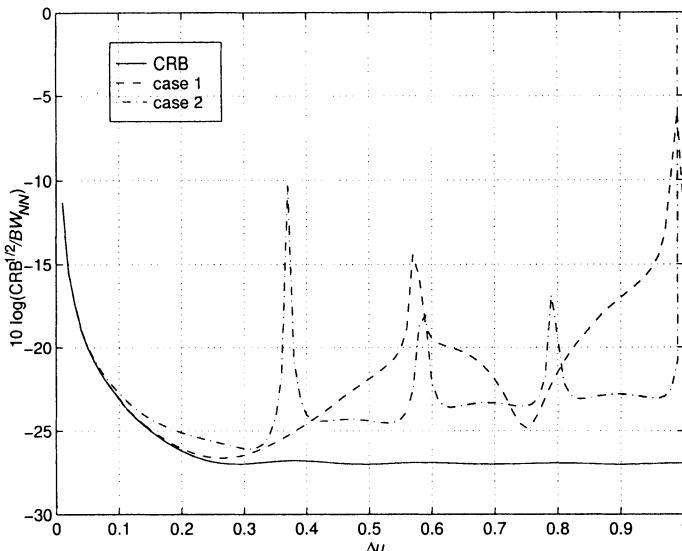


Figure 8.52 Normalized beamspace CRB versus Δu using a Taylor series beamspace matrix: $N = 10$, $N_{bs} = 3$, SNR = 10 dB, $K = 100$.

We next consider a beamspace matrix using the discrete prolate spheroidal sequences.

Example 8.10.5

We consider the same model as in Example 8.10.2. We utilize the DPSS functions derived in Section 8.10.2.1 to construct the beamspace matrix. We denote the width of the sector by $2\psi_0$. The \mathbf{B}_{no} matrix is given by (8.599). In Figure 8.54, we use three beams and plot the beamspace CRB versus source separation for various values of $2u_0$. For $\Delta u \leq 0.3$, the bounds are close to the element-space CRB for all of the four cases. In Figure 8.55, we use four beams. For $N_{bs} = 4$, the beamspace CRB remains close to the element-space CRB for $\Delta u \leq 0.5$. In Figure 8.56, we increase N_{bs} as $2u_0$ increases. In this case the bounds remain close out to $\Delta u = 1.0$.

8.10.3.3 DFT beamspace matrices

In most applications utilizing beamspace processing the beamspace matrix consists of conventional beams whose pointing directions are spaced by $2/N$ in u -space. It is referred to as the DFT beamspace. In this section we consider this model and a variation of it. We compute the beamspace CRB for two examples.

Example 8.10.6

Consider a standard 32-element linear array and a $N_{bs} \times N$ DFT beamspace matrix.

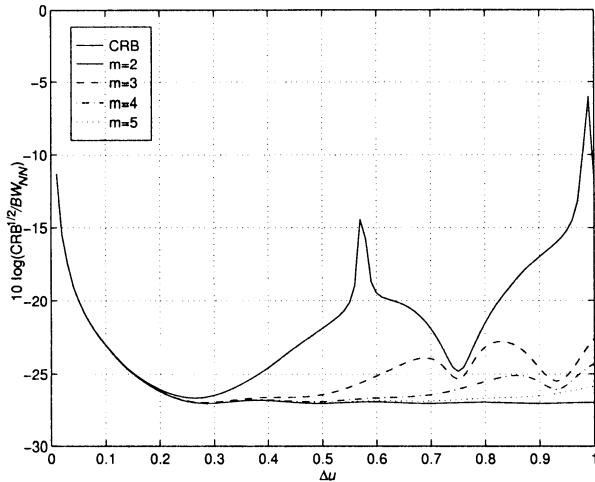


Figure 8.53 Normalized beamspace CRB versus signal separation for various m : Taylor series beamspace matrix, $N = 10$, $N_{bs} = m + 1$, $SNR = 10$ dB, $K = 100$.

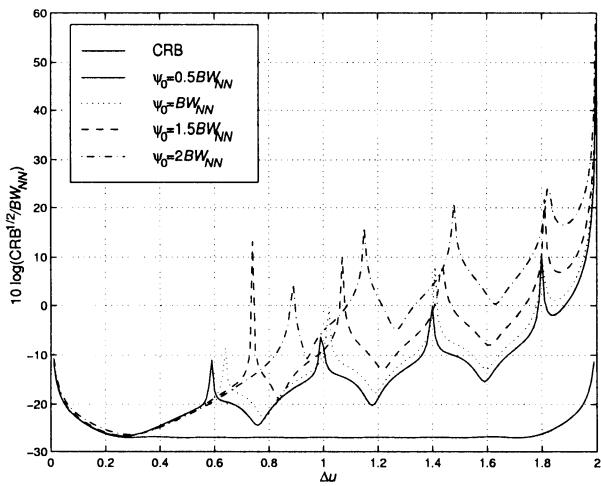


Figure 8.54 Normalized beamspace CRB using DPSS function beamspace matrix: $N = 10$, $N_{bs} = 3$, $SNR = 10$ dB, $K = 100$.

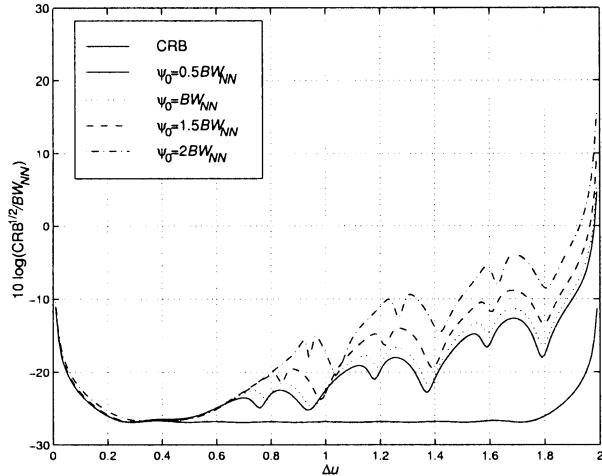


Figure 8.55 Normalized beamspace CRB for DPSS function beamspace matrix: $N = 10$, $N_{bs} = 4$, $SNR = 10$ dB, $K = 100$.

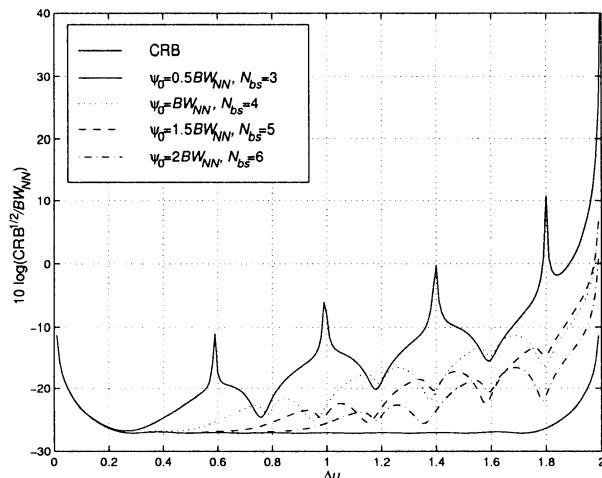


Figure 8.56 Normalized beamspace CRB for DPSS function beamspace matrix: $N = 10$, $N_{bs} = 3, 4, 5$, and 6 , $SNR = 10$ dB, $K = 100$.

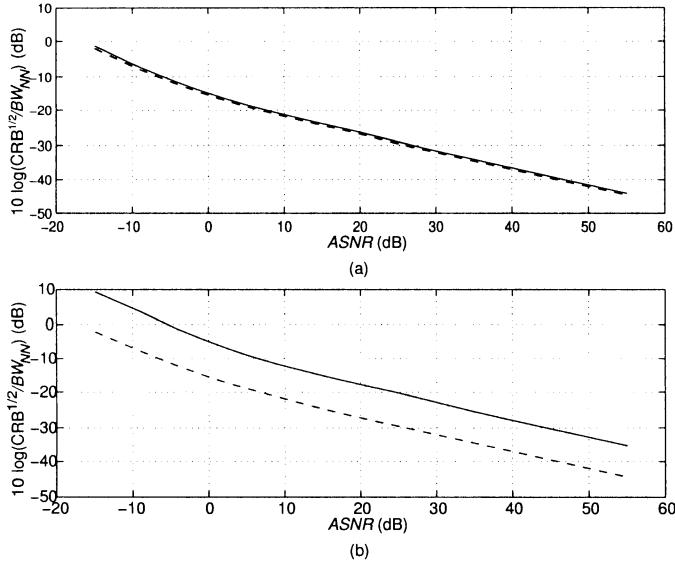


Figure 8.57 Normalized beamspace and element-space CRBs versus $ASNR$ for DFT beamspace matrix: $N = 32, N_{bs} = 3$, two plane-wave signals at $\pm\Delta u/2$: (a) $\Delta u = N_{bs}/N$; (b) $\Delta u = 2N_{bs}/N$.

There are two equal-power uncorrelated plane-wave signals impinging of the array from $u = \pm\Delta u/2$. We plot the CRB versus $ASNR$ for various Δu . We also plot the element-space CRB for comparison. In Figure 8.57, $N_{bs} = 3$. In Figure 8.58, $N_{bs} = 5$.

In both cases, the beamspace CRB and the element-space CRB are almost equal for $\Delta u = N_{bs}/N$. This spacing is well inside the beamspace sector. At $\Delta u = 2N_{bs}/N$, the signals are on the sidelobes of the outer beams and the Cramér-Rao differ by almost 10 dB.

Example 8.10.7 (continuation)

Consider the same 32-element standard linear array as in Example 8.10.6. We use Dolph-Chebychev (-40 dB SLL) beamformer spaced at $2/N$, for the rows in a 5×32 beamspace matrix. The motivation for this choice is to decrease the effect of any high-power out-of-sector interferers.

In Figure 8.59, we plot the beamspace CRB versus $ASNR$. We see that the beamspace CRB and the element-space CRB are essentially the same as in Example 8.10.6.

8.10.3.4 Summary

We see that, by suitably choosing the beamspace matrix, we can obtain a beamspace CRB that is very close to the element-space CRB for signals that are well inside the beamspace sector.

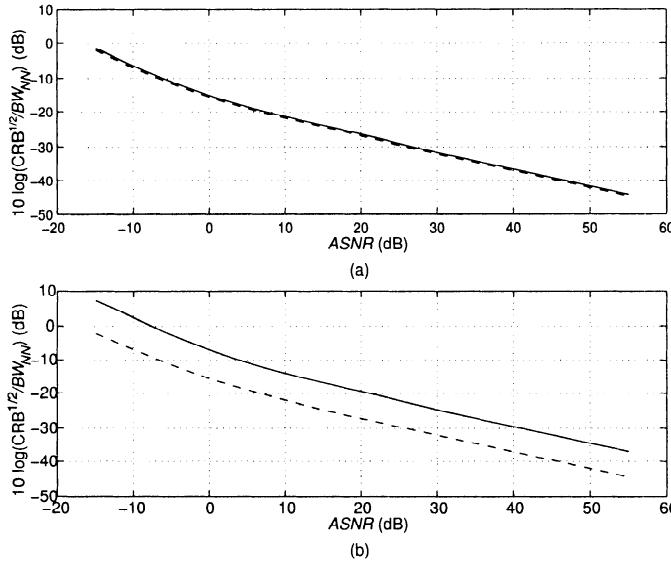


Figure 8.58 Normalized beamspace and element-space CRBs versus $ASNR$ for DFT beamspace matrix: $N = 32, N_{bs} = 5$, two plane-wave signals at $\pm \Delta u/2$: (a) $\Delta u = N_{bs}/N$; (b) $\Delta u = 2N_{bs}/N$.

8.10.4 Beamspace Maximum Likelihood

In this section, we derive the beamspace maximum likelihood estimator. We restrict our attention to the beamspace version of the CML estimator discussed in Section 8.5.2.

Using the model in Section 8.5.2 ((8.332)–(8.353)), the likelihood function can be written as,

$$L_{bs}(\psi, \mathbf{F}) = -KN_{bs} \ln \sigma_w^2 - \frac{1}{\sigma_w^2} \sum_{k=1}^K |\mathbf{X}_{bs}(k) - \mathbf{V}_{bs}(\psi)\mathbf{F}_k|^2, \quad (8.627)$$

where

$$\mathbf{V}_{bs}(\psi) = \begin{bmatrix} \mathbf{v}_{bs}(\psi_1) & \cdots & \mathbf{v}_{bs}(\psi_D) \end{bmatrix}, \quad (8.628)$$

and

$$\mathbf{v}_{bs}(\psi) = \mathbf{B}_{bs}^H \mathbf{v}(\psi). \quad (8.629)$$

Then, proceeding as in Section 8.5.2, we obtain

$$\hat{\psi}_{cml,bs} = \arg \min_{\psi} \left\{ \text{tr} \left[\mathbf{P}_{\mathbf{V}_{bs}}^\perp \mathbf{C}_{\mathbf{X}_{bs}} \right] \right\}, \quad (8.630)$$

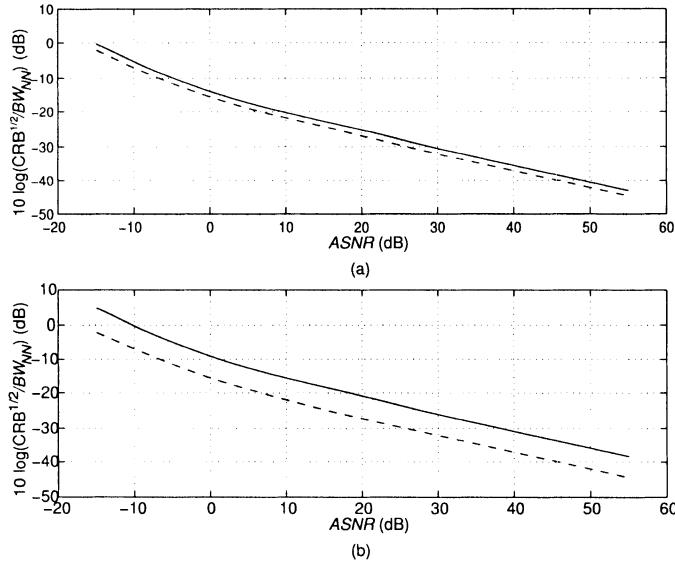


Figure 8.59 Normalized beamspace and element-space CRBs versus $ASNR$ for Dolph-Chebychev beamspace matrix: $N = 32$, $N_{bs} = 5$, two plane-wave signals at $\pm\Delta u/2$: (a) $\Delta u = N_{bs}/N$; (b) $\Delta u = 2N_{bs}/N$.

where

$$\mathbf{P}_{\mathbf{V}_{bs}} = \mathbf{V}_{bs} \left[\mathbf{V}_{bs}^H \mathbf{V}_{bs} \right]^{-1} \mathbf{V}_{bs}^H, \quad (8.631)$$

and

$$\mathbf{C}_{\mathbf{X}_{bs}} = \frac{1}{K} \sum_{k=1}^K \mathbf{X}_{bs}(k) \mathbf{X}_{bs}^H(k). \quad (8.632)$$

In Section 8.7, we saw that from a computational point of view, the IQML and IMODE algorithms were attractive in estimating direction of arrivals in the element space. In beamspace processing, Zoltowski and Lee proposed 2-D beamspace domain ML (BDML) and 3-D BDML [ZL91] schemes for estimating the DOAs in the cases of a single signal and the two signals, respectively. These two schemes are computationally simple. Both of them use DFT beamspace matrices, and choose the number of beams to be the number of impinging signals plus one.

Zoltowski also presented an IQML algorithm in which the beamspace transformation is based on subspace processing [Zol88].

Tian and Van Trees [TVT00] have derived a beamspace IQML estimator

using DFT matrix beamformers. It utilizes several key results from Zoltowski et al. [ZKS93]. We summarize the derivation in [TVT00].

In the IQML algorithm, the spatial characteristics of the signal components is parameterized by a coefficient vector $\mathbf{b} = [b_0 \cdots b_D]^T$. The vector \mathbf{b} is defined such that the polynomial

$$b(z) = b_0 z^D + b_1 z^{D-1} + \cdots + b_D, \quad (8.633)$$

has D roots at $z_i = e^{j\pi u_i}$, $i = 1, \dots, D$. The ML estimate of u_i , $i = 1, \dots, D$, can be obtained from the ML estimate of \mathbf{b} .

In order to extend the element-space IQML algorithm to the beamspace domain, it is necessary to find a linear parameterization of the null space of \mathbf{V}_{bs} . This is equivalent to finding a full-rank matrix $\mathbf{B}(\mathbf{b})$ such that

$$\mathcal{L}\{\mathbf{B}\} \mathbf{V}_{bs} = 0. \quad (8.634)$$

The operator $\mathcal{L}\{\cdot\}$ represents a linear transform operation, and \mathbf{B} is a $N_{bs} \times (N_{bs} - D)$ Toeplitz matrix given by

$$\mathbf{B}^H = \begin{bmatrix} b_D & \cdots & b_0 & 0 & 0 \\ 0 & \ddots & \ddots & \ddots & 0 \\ 0 & 0 & b_D & \ddots & b_0 \end{bmatrix}. \quad (8.635)$$

Polynomial parameterization by (8.634) is not always possible for any beamforming transformation. In element space, the linear parameterization of the noise space is made possible by the Vandermonde property of the element-space array response matrix. The Vandermonde structure of the array manifold may not be preserved after beamspace transformation. A crucial step in developing the BS-IMQL algorithm is to find the linear parameterization formula $\mathcal{L}\{\mathbf{B}\}$.

We use the DFT beamforming matrix in (8.581). Due to the common out-of-band nulls in DFT beams, a DFT beamforming matrix can be transformed into a banded Toeplitz matrix [ZKS93]. This transformation preserves the Vandermonde property of the beamspace array manifold and enables the polynomial parameterization in the beamspace. The beamspace matrix consisting of N_{bs} DFT beamformers can be rewritten as an $N \times N_{bs}$ matrix²⁹

$$\mathbf{W} = \frac{1}{N} [\mathbf{v}_N(m \frac{2}{N}) \cdots \mathbf{v}_N((m + N_{bs} - 1) \frac{2}{N})], \quad (8.636)$$

²⁹We have replaced \mathbf{B}_{bs} with \mathbf{W} to avoid confusion with the \mathbf{B} matrix in (8.635).

where the integer number m should be properly chosen so that the N_{bs} beams cover most of the signal energy, and

$$\mathbf{v}_N(u_i) = [e^{-j(\frac{N-1}{2})2\pi du_i} \dots e^{j(\frac{N-1}{2})2\pi du_i}]^T, \quad (8.637)$$

where u_i is the i th direction of arrival in u -space, and d is the element spacing measured in wavelengths.

It has been shown ([ZKS93]) that \mathbf{W} can be factored as

$$\mathbf{W} = \mathbf{CQ}, \quad (8.638)$$

where \mathbf{Q} is a $N_{bs} \times N_{bs}$ full-rank matrix and \mathbf{C} is a $N \times N_{bs}$ banded Toeplitz matrix

$$\mathbf{C} = \begin{bmatrix} c_0 & 0 & \cdots & 0 \\ c_1 & c_0 & & \\ \vdots & \vdots & \ddots & \vdots \\ c_{N-N_{bs}} & c_{N-N_{bs}-1} & c_0 & \\ 0 & c_{N-N_{bs}} & c_1 & \\ \vdots & & \ddots & \vdots \\ 0 & \cdots & & c_{N-N_{bs}} \end{bmatrix}. \quad (8.639)$$

The vector \mathbf{c} is an $(N - N_{bs} + 1) \times 1$ vector whose elements are the first $(N - N_{bs} + 1)$ elements in the first column of \mathbf{C} .

The following equality holds by arithmetic manipulations [ZKS93]:

$$\mathbf{C}^H \mathbf{v}_N(u) = \alpha(u) \mathbf{v}_{N_{bs}}(u), \quad (8.640)$$

where $\mathbf{v}_{N_{bs}}(u)$ is defined by (8.637) with N replaced by N_{bs} . Note that

$$\alpha(u) = \sum_{i=0}^{N-N_{bs}} c_i^* e^{j(-\frac{N-N_{bs}}{2}+i)2\pi du} = \mathbf{c}^H \mathbf{V}_{N-N_{bs}+1}(u) \quad (8.641)$$

is a scalar, which does not affect the structure in $\mathbf{v}_{N_{bs}}(u)$. The property in (8.640) is critical to the applicability of IQML in beamspace.

It follows from (8.633) that

$$\mathbf{B}^H \mathbf{v}_{N_{bs}}(u_i) = z_i^{-\frac{N_{bs}-1}{2}} \begin{bmatrix} z_i^0 \\ \vdots \\ z_i^{N_{bs}-D-1} \end{bmatrix} b(z_i) = \mathbf{0}, \quad (8.642)$$

for $i = 1, \dots, D$. Now, define a $N_{bs} \times (N_{bs} - D)$ matrix $\mathbf{F} = \mathbf{Q}^{-1}\mathbf{B}$. From (8.638), (8.640), and (8.642), we have

$$\begin{aligned}\mathbf{F}^H \mathbf{v}_{bs}(u_i) &= \mathbf{B}^H \mathbf{Q}^{-H} \mathbf{W}^H \mathbf{v}_N(u_i) = \mathbf{B}^H \mathbf{C}^H \mathbf{v}_N(u_i) \\ &= \alpha(u_i) \mathbf{B}^H \mathbf{v}_{N_{bs}}(u_i) = \mathbf{0}, \quad i = 1, \dots, D.\end{aligned}\quad (8.643)$$

Since \mathbf{F} has rank $N_{bs} - D$, its columns span the orthogonal complement to the signal subspace, that is,

$$\mathbf{P}_{\mathbf{V}_{bs}}^\perp = \mathbf{P}_{\mathbf{F}}, \quad (8.644)$$

where $\mathbf{P}_{\mathbf{F}} = \mathbf{F}(\mathbf{F}^H \mathbf{F})^{-1} \mathbf{F}^H$. Therefore, we have

$$\mathcal{L}\{\mathbf{B}\} = \mathbf{F}^H = \mathbf{B}^H \mathbf{Q}^{-H}. \quad (8.645)$$

The CML estimate is given in the beamspace by

$$\begin{aligned}\hat{\mathbf{u}} &= \arg \min_{\mathbf{u}} \text{tr} \left\{ \mathbf{P}_{\mathbf{V}_{bs}}^\perp(\mathbf{u}) \hat{\mathbf{S}}_{bs} \right\} \\ &= \arg \min_{\mathbf{b}(\mathbf{u})} \text{tr} \left\{ \mathbf{P}_{\mathbf{F}}(\mathbf{b}) \hat{\mathbf{S}}_{bs} \right\},\end{aligned}\quad (8.646)$$

where $\hat{\mathbf{S}}_{bs}$ is the beamspace spatial spectral matrix.

Define $\mathbf{S}_Q = \mathbf{Q}^{-H} \hat{\mathbf{S}}_{bs} \mathbf{Q}^{-1}$. We have

$$\text{tr}\{\mathbf{P}_{\mathbf{F}} \hat{\mathbf{S}}_{bs}\} = \text{tr} \left\{ \mathbf{F}(\mathbf{F}^H \mathbf{F})^{-1} \mathbf{F}^H \hat{\mathbf{S}}_{bs} \right\} \quad (8.647)$$

$$= \text{tr} \left\{ \mathbf{Q}^{-1} \mathbf{B}(\mathbf{F}^H \mathbf{F})^{-1} \mathbf{B}^H \mathbf{Q}^{-H} \hat{\mathbf{S}}_{bs} \right\} \quad (8.648)$$

$$= \text{tr} \left\{ \mathbf{B}(\mathbf{F}^H \mathbf{F})^{-1} \mathbf{B}^H \mathbf{S}_Q \right\}. \quad (8.649)$$

Minimization of the objective function in (8.649) can be readily solved by the IQML method discussed in Section 8.7.2.³⁰

We need to find \mathbf{Q}^{-1} in order to implement the IQML algorithm. The DFT beamforming matrix may be decomposed into the product of a banded Toeplitz matrix \mathbf{C} and a full-rank matrix \mathbf{Q} (8.638). Only the inverse of the matrix \mathbf{Q} needs to be computed for the IQML procedure. Zoltowski et al. [ZKS93] provide a closed-form expression for \mathbf{Q} without forming \mathbf{C} . Here we briefly explain an intuitive way to compute the matrix \mathbf{C} , and give an simple expression to compute \mathbf{Q}^{-1} directly.

The N_{bs} columns of \mathbf{W} are part of the $N \times N$ DFT matrix. Due to the orthogonal properties of DFT beams, the other $N - N_{bs}$ columns of the

³⁰We use IQML-QC. The QC descriptor is omitted for simplicity.

$N \times N$ DFT matrix that are not contained in \mathbf{W} are orthogonal to each of the N_{bs} columns of \mathbf{W} . Mathematically,

$$\mathbf{W}^H \mathbf{v}_N(u_n) = \mathbf{0}, \quad (8.650)$$

or

$$\mathbf{Q}^H \mathbf{C}^H \mathbf{v}_N(u_n) = \mathbf{0}, \quad (8.651)$$

for $u_n \in \{(m + N_{bs})\frac{2}{N}, \dots, (N + m - 1)\frac{2}{N}\}$. Define a polynomial

$$c(z) = c_{N-N_{bs}}^* z^{N-N_{bs}} + \dots + c_1^* z + c_0. \quad (8.652)$$

Equation (8.651) implies that $c(z)$ has $N - N_{bs}$ roots at $z_n = \exp\{j\pi u_n\}$. Therefore, the coefficients of $c(z)$ can be found by

$$[c_o \ \dots \ c_{N-N_{bs}}]^H = \text{poly}\{e^{j\pi u_n}\}, \quad (8.653)$$

where the operator $\text{poly}\{\cdot\}$ converts the roots to a polynomial.

Once the matrix \mathbf{C} is computed, the inverse of the matrix \mathbf{Q} can be found by pre-multiplying both sides of (8.638) by \mathbf{W}^H :

$$\mathbf{W}^H \mathbf{W} = \mathbf{W}^H \mathbf{C} \mathbf{Q} = \mathbf{I}. \quad (8.654)$$

Therefore, the inverse of \mathbf{Q} is given by $\mathbf{Q}^{-1} = \mathbf{W}^H \mathbf{C}$.

We consider an example to illustrate the performance of beamspace IQML.

Example 8.10.8

Consider a SLA 32. There are two equal-power uncorrelated signals impinging on the array from $\pm\Delta\psi/2$ where $\Delta\psi = \Delta\psi_R$. The beamspace dimension, N_{bs} , equals 8 and is centered at $\psi = 0$. We simulate beamspace IQML and element-space IQML for $K = 100$.

In Figure 8.60(a), we plot the normalized RMSE versus $ASNR$ for the two IQML algorithms. We also plot the element-space CRB and the beamspace CRB. In Figure 8.60(b), we plot the probability of resolution versus $ASNR$.

We see that both IQML algorithms approach their respective CRBs above threshold (the difference in the bounds is 0.61 dB). The probability of resolution performance is similar. The performance of ES-IQML below threshold is handicapped because we have not used the beam fan information.

The lower computational complexity of BS-IQML compensates for the slightly larger RMSE.

Other examples exhibit similar behavior. Beamspace IQML provides good performance with reasonable computational complexity. In Chapter 9, we discuss two other beamspace estimation algorithms and compare their performance to the IQML algorithm.

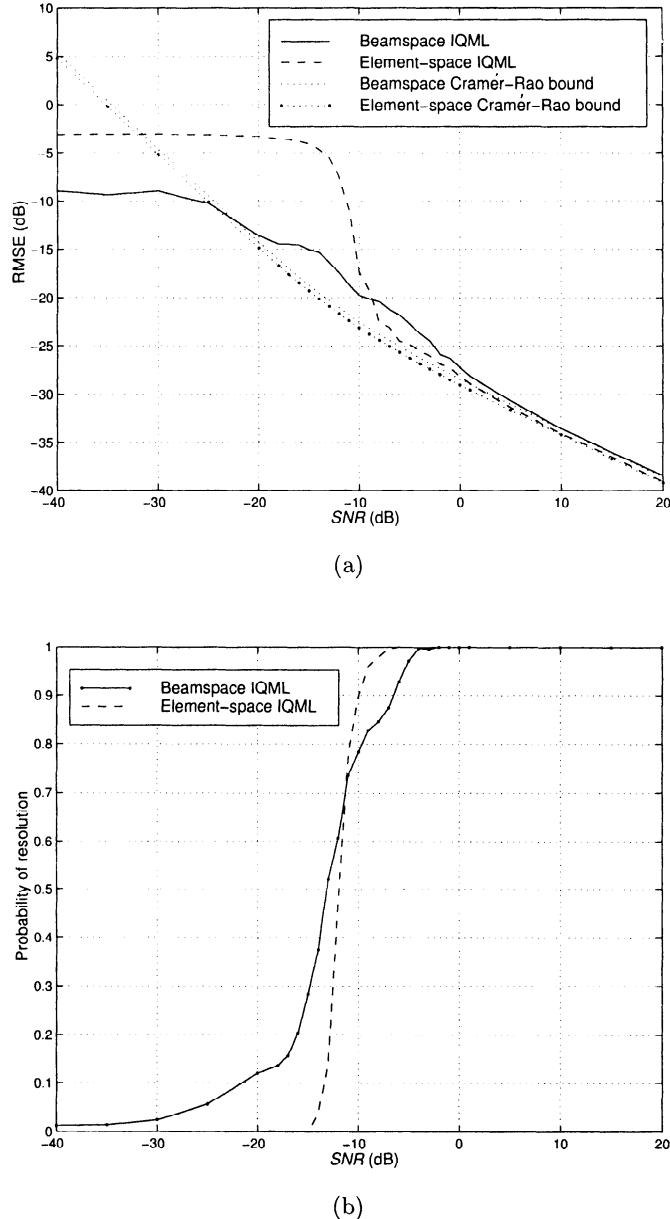


Figure 8.60 Beamspace and element-space IQML: $N = 32$, $N_{bs} = 8$, $K = 100$, $\Delta\psi = \Delta\psi_R$, $\rho = 0$, 500 trials: (a) RMSE versus SNR ; (b) probability of resolution versus SNR .

8.10.5 Summary

In this section, we have considered maximum likelihood estimation in beampspace. In most cases, we can select a beampspace matrix such that the beampspace CRB is very close to the element-space CRB. The beampspace maximum likelihood estimate approaches the beampspace CRB above threshold. The thresholds occur at about the same *ASNR* as in element-space estimation. However the increase in the RMSE is not as sharp, because we have assumed that the signals are inside the beam sector.

For uniform linear arrays, the beampspace IQML algorithm provides excellent performance with reasonable computational complexity. Using a similar technique, a beampspace IMODE algorithm has been developed (see [Tia01]). The performance for Example 8.10.8 is similar to IQML.

8.11 Sensitivity, Robustness, and Calibration

In our discussion of classical antenna processing and adaptive beamformers we saw the importance of array calibration in maintaining satisfactory performance. We would anticipate that the various high-resolution parameter estimation algorithms that we have developed would exhibit varying degrees of sensitivity to the model assumptions about the environment.

There are three related problems of interest:

- (i) **Sensitivity** We implement one of the parameter estimation algorithms assuming a nominal array manifold and nominal noise model. We analyze (or simulate) the performance under the perturbed conditions (e.g. sensor position perturbations) and measure the degradation in performance. For the perturbation model that we utilize, we can bound the resulting performance with the hybrid CRB derived in Section 8.4.1.3.
- (ii) **Robustness** We modify the parameter estimation algorithm to reduce the degradation due to the perturbations. We describe an algorithm as robust if it is reasonably insensitive to model perturbations.

One approach to designing a robust algorithm is to jointly estimate the DOAs and the model parameters. In the context of robust beamformers, we are not explicitly interested in the model parameters (they are “unwanted parameters”). We only estimate them to improve our DOA estimation performance.

There are other approaches to robustness that do not estimate the model parameters.

- (iii) **Calibration** In this problem we are explicitly interested in estimating the array parameters (e.g., sensor position, gain, and phase). This problem is referred to as the **calibration problem**.

In some cases, we try to calibrate the array using signals with unknown DOAs. This technique is referred to as the **blind calibration** problem.

A complete discussion of these three topics is beyond the scope of our discussion. We provide a brief introduction to the area and highlight the issues involved. We provide some references here and expand the reference list in Section 9.8.

In Section 8.11.1, we review the models for the perturbations in the array parameters and the noise environment that we introduced in Section 6.6.3. In Section 8.11.2, we develop hybrid Cramér-Rao bounds for the joint estimation of the signal parameters and the array parameters. In Section 8.11.3, we analyze how the array perturbations affect the performance of some of the maximum likelihood estimators that we derived in this chapter. We find that the performance can decline significantly under certain scenarios. In Section 8.11.4, we consider a joint estimation approach in order to improve robustness. The array parameters are estimated as part of the algorithm, but the emphasis is on improving the DOA estimation performance. In Section 8.11.5, we briefly discuss the calibration problem. In Section 8.11.6, we summarize our results.

8.11.1 Model Perturbations

In this section, we develop several characterizations to describe perturbations in our nominal observation model. We first consider the case of perturbations in the array manifold matrix. The models consider perturbations in the gain and phase of the sensors and in the position of the array elements.

The frequency-domain snapshot model can be written as

$$\mathbf{X}(k) = \mathbf{V}(\boldsymbol{\theta}, \boldsymbol{\phi}; \boldsymbol{\rho}) \mathbf{F}(k) + \mathbf{W}(k), \quad k = 1, 2, \dots, K, \quad (8.655)$$

where $\mathbf{F}(k)$ is a $D \times 1$ complex source signal with source spectral matrix \mathbf{S}_f . The vector $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ represent the elevation and azimuth angles of the D plane-wave signals. The vector $\boldsymbol{\rho}$ is a real $M \times 1$ vector that represents the perturbations in the array parameters. The additive white noise has spectral height σ_w^2 .

For notational simplicity, we will consider the case in which the array is in the x - y -plane, and the plane waves are arriving in the $x - y$ plane so we are only estimating ϕ . Then, (8.655) can be written as

$$\mathbf{X}(k) = \mathbf{V}(\phi, \rho) \mathbf{F}(k) + \mathbf{W}(k), \quad k = 1, 2, \dots, K, \quad (8.656)$$

where the i th element of ϕ is the angle between the i th plane wave and the x -axis (see Figure 2.1).

This model is the same model that has been used throughout the chapter except that $\mathbf{V}(\phi, \rho)$ is written as a function of ϕ rather than ψ . This change is made so that our results can be more easily compared to those in the literature. The two models can be related by (8.44)–(8.48). The array perturbation model in Section 6.3.3 is used with this change.

8.11.2 Cramér-Rao Bounds

In this section, we derive the Cramér-Rao bound on the variance of any unbiased estimator in the presence of array perturbations.

When we derive the CRB we are implicitly assuming that we are jointly estimating all of the parameters in the model (ϕ and ρ). However, the resulting bound also applies to the case in which we only estimate ϕ and treat ρ as an unwanted or nuisance parameter. Therefore, we can use it to bound the behavior of algorithms that assume some nominal array manifold matrix, $\mathbf{V}(\phi, \rho_0)$, in the presence of array perturbations. The CRB also bounds the behavior of autocalibration algorithms that jointly estimate ϕ and ρ .

Our approach utilizes the hybrid Cramér-Rao bound which was derived in Section 8.2.3.3. This approach was first used in the array context to study the array shape calibration problem by Rockah and Schultheiss [RS87a]. This paper contains several useful results on array calibration and provides useful background reading. Subsequent work by Weiss and Friedlander ([WF89] and Appendix B of Chapter 10 of [Hay91a]) derived compact expressions for the CRB in the presence of position displacements and gain and phase errors. They consider the case in which the source spectral matrix \mathbf{S}_f and the white noise level σ_w^2 are known. Although this case is generally not applicable in practice, it bounds the unknown spectrum case and the resulting CRB enables us to identify some of the key issues in the calibration problem.

In this section, we consider the more common case in which the spectral matrix \mathbf{S}_f and the noise level are unknown. After deriving a general expression we focus our attention on position displacements.

8.11.2.1 Cramér-Rao bound: Unknown spectral matrix

We consider the case in which there are D plane waves impinging on the array. The signals are sample functions from zero-mean Gaussian random processes with source spectral matrix \mathbf{S}_f . The additive noise is white with spectral height σ_w^2 . The parameters to be perturbed are denoted by ρ and are imbedded in the array manifold matrix \mathbf{V} , which we denote by $\mathbf{V}(\phi, \rho)$.

From (8.656),

$$\mathbf{X}(k) = \mathbf{V}(\phi, \rho) \mathbf{F}(k) + \mathbf{W}(k), \quad k = 1, 2, \dots, K. \quad (8.657)$$

We assume that \mathbf{S}_f and σ_w^2 are unknown.

The total parameter vector of interest is

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_w \\ \boldsymbol{\theta}_u \end{bmatrix}, \quad (8.658)$$

where

$$\boldsymbol{\theta}_w = \begin{bmatrix} \phi \\ \rho \end{bmatrix}, \quad (8.659)$$

is a $(D + M) \times 1$ vector that represents the wanted parameters and

$$\boldsymbol{\theta}_u = \begin{bmatrix} \mu \\ \sigma_w^2 \end{bmatrix}, \quad (8.660)$$

is a $(D^2 + 1) \times 1$ vector that represents the unwanted parameters. We first consider the Fisher information matrix associated with $\boldsymbol{\theta}_w$. The vector ϕ is a $D \times 1$ vector corresponding to source DOAs in angle space. The parameter perturbation vector ρ is M -dimensional. For example, in the case of position perturbations, $M = 2N$:

$$\rho = [p_{x_0} \ p_{y_0} \ p_{x_1} \ p_{y_1} \ \cdots \ p_{x_{N-1}} \ p_{y_{N-1}}]^T. \quad (8.661)$$

We assume that ρ is a real Gaussian random vector

$$p_\rho(\rho) = \frac{1}{(2\pi)^{\frac{M}{2}} |\Lambda_\rho|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\rho - \rho_0)^T \Lambda_\rho^{-1} (\rho - \rho_0) \right\}, \quad (8.662)$$

where ρ_0 represents the nominal value of the parameter vector ρ . The Gaussian assumption is a good model for many applications.

The Fisher information matrix for a single observation is given by,

$$\mathbf{J}_B(\boldsymbol{\theta}_w) = \mathbf{J}_D(\boldsymbol{\theta}_w) + \mathbf{J}_P(\boldsymbol{\theta}_w). \quad (8.663)$$

The data matrix \mathbf{J}_D is partitioned as,

$$\mathbf{J}_D(\boldsymbol{\theta}_w) = \begin{bmatrix} \mathbf{J}_{\phi\phi} & \mathbf{J}_{\phi\rho} \\ \mathbf{J}_{\rho\phi} & \mathbf{J}_{\rho\rho} \end{bmatrix}. \quad (8.664)$$

The prior matrix \mathbf{J}_P is given by (8.59) as

$$\mathbf{J}_P(\boldsymbol{\theta}_w) = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \Lambda_{\rho}^{-1} \end{bmatrix}, \quad (8.665)$$

because $\boldsymbol{\phi}$ is an unknown nonrandom vector.

To evaluate the terms in $\mathbf{J}_D(\boldsymbol{\theta}_w)$, we use

$$[\mathbf{J}_{\phi\phi}]_{ij} = -E_{\mathbf{x},\boldsymbol{\theta}} \left[\frac{\partial^2 \ln p_{\mathbf{x}}(\mathbf{x}|\boldsymbol{\theta})}{\partial \phi_i \partial \phi_j} \right], \quad i, j = 1, \dots, D. \quad (8.666)$$

$$[\mathbf{J}_{\rho\rho}]_{ij} = -E_{\mathbf{x},\boldsymbol{\theta}} \left[\frac{\partial^2 \ln p_{\mathbf{x}}(\mathbf{x}|\boldsymbol{\theta})}{\partial \rho_i \partial \rho_j} \right], \quad i, j = 1, \dots, M. \quad (8.667)$$

$$[\mathbf{J}_{\phi\rho}]_{ij} = -E_{\mathbf{x},\boldsymbol{\theta}} \left[\frac{\partial^2 \ln p_{\mathbf{x}}(\mathbf{x}|\boldsymbol{\theta})}{\partial \phi_i \partial \rho_j} \right], \quad i = 1, 2, \dots, D; \quad j = 1, 2, \dots, M. \quad (8.668)$$

We can use (8.35) to obtain

$$[\mathbf{J}_{\phi\phi}]_{ij} = E_{\boldsymbol{\theta}} \left\{ \text{tr} \left[\mathbf{K}_{\mathbf{x}}^{-1} \frac{\partial \mathbf{K}_{\mathbf{x}}}{\partial \phi_i} \mathbf{K}_{\mathbf{x}}^{-1} \frac{\partial \mathbf{K}_{\mathbf{x}}}{\partial \phi_j} \right] \right\}, \quad (8.669)$$

$$[\mathbf{J}_{\rho\rho}]_{ij} = -E_{\boldsymbol{\theta}} \left\{ \text{tr} \left[\mathbf{K}_{\mathbf{x}}^{-1} \frac{\partial \mathbf{K}_{\mathbf{x}}}{\partial \rho_i} \mathbf{K}_{\mathbf{x}}^{-1} \frac{\partial \mathbf{K}_{\mathbf{x}}}{\partial \rho_j} \right] \right\}, \quad (8.670)$$

and a similar relationship for the cross-term matrices. Note that, although an expectation over $\boldsymbol{\theta}$ is indicated, the only random component is $\boldsymbol{\rho}$ so it is really an expectation over $\boldsymbol{\rho}$.

The expectation over $\boldsymbol{\theta}$ is difficult to evaluate so we introduce an approximation that was originally proposed by Rockah and Schultheiss [RS87a]. We assume that the traces in (8.669) and (8.670) are reasonably smooth in the

vicinity of the nominal value of $\boldsymbol{\rho} = \boldsymbol{\rho}_0$ and if the variances of the array perturbations are small, then we can approximate (8.669) and (8.670) by

$$[\mathbf{J}_{\boldsymbol{\phi}\boldsymbol{\phi}}]_{ij} \cong \text{tr} \left[\mathbf{K}_{\mathbf{x}}^{-1} \frac{\partial \mathbf{K}_{\mathbf{x}}}{\partial \phi_i} \mathbf{K}_{\mathbf{x}}^{-1} \frac{\partial \mathbf{K}_{\mathbf{x}}}{\partial \phi_j} \right] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \quad (8.671)$$

$$[\mathbf{J}_{\boldsymbol{\rho}\boldsymbol{\rho}}]_{ij} \cong \text{tr} \left[\mathbf{K}_{\mathbf{x}}^{-1} \frac{\partial \mathbf{K}_{\mathbf{x}}}{\partial \rho_i} \mathbf{K}_{\mathbf{x}}^{-1} \frac{\partial \mathbf{K}_{\mathbf{x}}}{\partial \rho_j} \right] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}, \quad (8.672)$$

and

$$[\mathbf{J}_{\boldsymbol{\phi}\boldsymbol{\rho}}]_{ij} \cong \text{tr} \left[\mathbf{K}_{\mathbf{x}}^{-1} \frac{\partial \mathbf{K}_{\mathbf{x}}}{\partial \phi_i} \mathbf{K}_{\mathbf{x}}^{-1} \frac{\partial \mathbf{K}_{\mathbf{x}}}{\partial \rho_j} \right] \Big|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \quad (8.673)$$

The Fisher information matrix for the total vector $\boldsymbol{\theta}$ can be written as

$$\mathbf{J}(\boldsymbol{\theta}) = \mathbf{J}(\boldsymbol{\theta}_w, \boldsymbol{\theta}_u) = \begin{bmatrix} \mathbf{J}_B(\boldsymbol{\theta}_w) & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{bmatrix}, \quad (8.674)$$

where the \mathbf{X} , \mathbf{Y} , and \mathbf{Z} matrices correspond to the appropriate \mathbf{J} matrices.

Then, the upper left $(D + M) \times (D + M)$ matrix in $\mathbf{J}_B(\boldsymbol{\theta})$ is,

$$\mathbf{J}_B(\boldsymbol{\theta}_w) = \mathbf{J}_D(\boldsymbol{\theta}_w) + \mathbf{J}_P(\boldsymbol{\theta}_w), \quad (8.675)$$

and is given by (8.664) and (8.665).

We are interested in the CRB on $\boldsymbol{\theta}_w$. Thus, we write,

$$\begin{aligned} \mathbf{J}^{-1}(\boldsymbol{\theta}) &= \left[\begin{bmatrix} \mathbf{J}_D(\boldsymbol{\theta}_w) & \mathbf{X} \\ \mathbf{Y} & \mathbf{Z} \end{bmatrix} + \begin{bmatrix} \mathbf{J}_P(\boldsymbol{\theta}_w) & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \right]^{-1} \\ &= \begin{bmatrix} [\mathbf{J}_D(\boldsymbol{\theta}_w) + \mathbf{X}\mathbf{Z}^{-1}\mathbf{Y} + \mathbf{J}_P(\boldsymbol{\theta}_w)]^{-1} & \mathbf{X}' \\ \mathbf{Y}' & \mathbf{Z}' \end{bmatrix}. \end{aligned} \quad (8.676)$$

We observe that the first two terms in the matrix in the upper left corner correspond to the matrix in the upper left corner of (8.94). Thus, the hybrid CRB for the $\boldsymbol{\theta}_w$ block of the hybrid CRB is given by (8.99):

$$\begin{aligned} &\left[[\mathbf{C}_{HCR}(\boldsymbol{\phi}, \boldsymbol{\rho})]^{-1} \right]_{ij} \\ &= \left[\frac{2K}{\sigma_w^2} \text{Re} \left[\text{tr} \left[\left[\mathbf{D}_j^H \mathbf{P}_V^\perp \mathbf{D}_i \right] \left[\mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f \right] \right] \right] + \mathbf{J}_P(\boldsymbol{\theta}_w) \right]^{-1}, \end{aligned} \quad (8.677)$$

where $\mathbf{J}_P(\boldsymbol{\theta}_w)$ is given by (8.665).³¹

The matrix \mathbf{D}_i is a $N \times (D + M)$ matrix that can be divided into two component matrices,

$$\mathbf{D}_i = [\mathbf{D}_{\phi_i} \quad \mathbf{D}_{\rho_i}]. \quad (8.678)$$

The left matrix is an $N \times D$ matrix,

$$\mathbf{D}_{\phi_i} = \frac{\partial}{\partial \phi_i} [\mathbf{V}(\phi, \rho)]|_{\rho=\rho_0}, \quad i = 1, 2, \dots, D, \quad (8.679)$$

and we have used the same smoothness assumption as in (8.671)–(8.673). The right matrix is an $N \times M$ matrix,

$$\mathbf{D}_{\rho_i} = \frac{\partial}{\partial \rho_i} [\mathbf{V}(\phi, \rho)]|_{\rho=\rho_0}, \quad i = D + 1, \dots, D + M, \quad (8.680)$$

where M is the number of parameters that are perturbed. The formula in (8.677) with the definitions in (8.679) and (8.680) is an adequate starting place for numerical evaluation.

Once the specific parameterization is specified, the expression can be evaluated. We consider the case in which the sensor positions are displaced. We assume that each perturbation parameter is associated with a single sensor.

We define

$$\mathbf{D}_{\phi} \triangleq \left[\begin{array}{cccc} \frac{\partial \mathbf{v}}{\partial \phi_1} & \frac{\partial \mathbf{v}}{\partial \phi_2} & \cdots & \frac{\partial \mathbf{v}}{\partial \phi_D} \end{array} \right], \quad (8.681)$$

and

$$\mathbf{D}_{\mathbf{x}} \triangleq \left[\begin{array}{cccc} \frac{\partial \mathbf{V}_{1,:}^T}{\partial p_{x_0}} & \frac{\partial \mathbf{V}_{2,:}^T}{\partial p_{x_1}} & \cdots & \frac{\partial \mathbf{V}_{N,:}^T}{\partial p_{x_{N-1}}} \end{array} \right]^T, \quad (8.682)$$

where $\mathbf{V}_{k,:}$ denotes the k th row of the array manifold matrix \mathbf{V} .³²

We can also write $\mathbf{D}_{\mathbf{x}}$ as,

$$\mathbf{D}_{\mathbf{x}} = \sum_{n=0}^{N-1} \frac{\partial \mathbf{V}}{\partial p_{x_n}}. \quad (8.683)$$

Similarly, we can define

$$\mathbf{D}_{\mathbf{y}} \triangleq \left[\begin{array}{cccc} \frac{\partial \mathbf{V}_{1,:}^T}{\partial p_{y_0}} & \frac{\partial \mathbf{V}_{2,:}^T}{\partial p_{y_1}} & \cdots & \frac{\partial \mathbf{V}_{N,:}^T}{\partial p_{y_{N-1}}} \end{array} \right]^T. \quad (8.684)$$

³¹The result in (8.677) is due to Wahlberg et al. [WOV91].

³²This result is also in [WOV91].

We also define³³

$$\Sigma \triangleq \mathbf{S}_f \mathbf{V}^H \mathbf{S}_x^{-1} \mathbf{V} \mathbf{S}_f. \quad (8.685)$$

Then, (8.677) reduces to

$$\begin{aligned} [C_{HCR}(\phi, \rho)]^{-1} &= \frac{2K}{\sigma_w^2} \\ &\times \operatorname{Re} \left[\begin{array}{c|c|c} \mathbf{D}_\phi^H \mathbf{P}_V^\perp \mathbf{D}_\phi \odot \Sigma^T & \mathbf{D}_\phi^H \mathbf{P}_V^\perp \odot (\mathbf{D}_x \Sigma)^T & \mathbf{D}_\phi^H \mathbf{P}_V^\perp \odot (\mathbf{D}_y \Sigma)^T \\ \hline (\mathbf{D}_x \Sigma) \odot (\mathbf{D}_\phi^H \mathbf{P}_V^\perp)^T & (\mathbf{D}_x \Sigma \mathbf{D}_x^H) \odot (\mathbf{P}_V^\perp)^T & (\mathbf{D}_x \Sigma \mathbf{D}_y^H) \odot (\mathbf{P}_V^\perp)^T \\ \hline (\mathbf{D}_y \Sigma) \odot (\mathbf{D}_\phi^H \mathbf{P}_V^\perp)^T & (\mathbf{D}_y \Sigma \mathbf{D}_x^H) \odot (\mathbf{P}_V^\perp)^T & (\mathbf{D}_y \Sigma \mathbf{D}_y^H) \odot (\mathbf{P}_V^\perp)^T \end{array} \right] \\ &+ \left[\begin{array}{c|c|c} \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \hline \mathbf{0} & \Lambda_x^{-1} & \mathbf{0} \\ \hline \mathbf{0} & \mathbf{0} & \Lambda_y^{-1} \end{array} \right]. \end{aligned} \quad (8.686)$$

The result in (8.686) assumes that all p_{x_n} and p_{y_n} are perturbed. In Example 8.11.2, we indicate how to modify the formula when some p_{x_n} and p_{y_n} are not perturbed.

We observe that the matrix in the upper right corner is identical to the result in (8.98).³⁴ It represents the CRB on ϕ in the absence of array errors. To find the impact of the array perturbations on the DOA estimation accuracy bound, we partition the right side of (8.686) as,

$$[C_{HCR}(\phi)]^{-1} = \left[\begin{array}{c|c} \frac{2K}{\sigma_w^2} \left\{ \operatorname{Re} \left[\mathbf{D}_\phi^H \mathbf{P}_V^\perp \mathbf{D}_\phi \odot \Sigma^T \right] \right\} & \mathbf{F} \\ \hline \mathbf{F}^H & \mathbf{G} \end{array} \right], \quad (8.687)$$

where \mathbf{F} and \mathbf{G} are the partitions of (8.686). Using the formula for the inverse of a partitioned matrix gives

$$[C_{HCR}(\phi)] = \left[\left\{ \frac{2K}{\sigma_w^2} \operatorname{Re} \left[\mathbf{D}_\phi^H \mathbf{P}_V^\perp \mathbf{D}_\phi \odot \Sigma^T \right] - \mathbf{F} \mathbf{G}^{-1} \mathbf{F}^H \right\} \right]^{-1}. \quad (8.688)$$

³³This Σ is different from the Σ in (8.103).

³⁴The bound in (8.98) is in wavenumber space (ψ). Here, the bound is in angle space.

The second term in (8.688) represents the increase in the DOA estimation bound due to array perturbations. Note that it is a bound on the covariance matrix of any unbiased estimate of ϕ . In addition, recall that it is an approximation to the bound because of (8.671), (8.672), and (8.673).

In a similar manner we can obtain a bound on the mean-square calibration error.

The result in (8.686) can be extended to the case in which³⁵

$$\boldsymbol{\rho} = \begin{bmatrix} g_0 & g_1 & \cdots & g_{N-1} & \gamma_0 & \gamma_1 & \cdots & \gamma_{N-1} \end{bmatrix}^T. \quad (8.689)$$

A key question that was first investigated by Rockah and Schultheiss [RS87a] is what conditions are required on the array geometry such that the second term in (8.688) will go to zero as the *SNR* goes to infinity.

Rockah and Schultheiss [RS87a] investigated array geometry conditions under which the variance and mean-square error would approach zero as the SNR approached infinity.³⁶ They assume that there are no signals with known directions. This is sometimes referred to as the blind calibration problem. They found that:

- (i) If the location of one sensor and the direction to a second sensor is known, then, except for some pathological array configurations, three sources were adequate to guarantee that the bound would approach zero as the *SNR* approaches infinity.
- (ii) The most conspicuous example of a pathological geometry is a linear array. Thus, the nominal array configuration must be nonlinear in order for the second term to approach zero. This result arises from the inability of the calibration procedure to estimate displacements along the axis of a linear array. If we can impose constraints to force Δp_{x_n} to equal zero, then the result changes.

We now consider two examples to illustrate typical behavior. The first example is a linear array, and we will see the result of the observation in (ii).

Example 8.11.1

Consider the model in which the nominal array is a standard 10-element linear array along the x -axis.

The sensor patterns are isotropic. We assume there are two plane-wave sources located at $\pm\Delta\phi/2$. We assume that the Δp_{x_n} and Δp_{y_n} position perturbations are independently-distributed Gaussian random variables with variance σ_p^2 . We let $K = 100$. We plot the

³⁵We have used γ for the sensor phase to avoid confusion with the DOA in angle space.

³⁶Rockah and Schultheiss [RS87a] used a simpler version of the CRB to derive their result (i.e., the plane-wave signal arrivals were assumed to be disjoint in time).

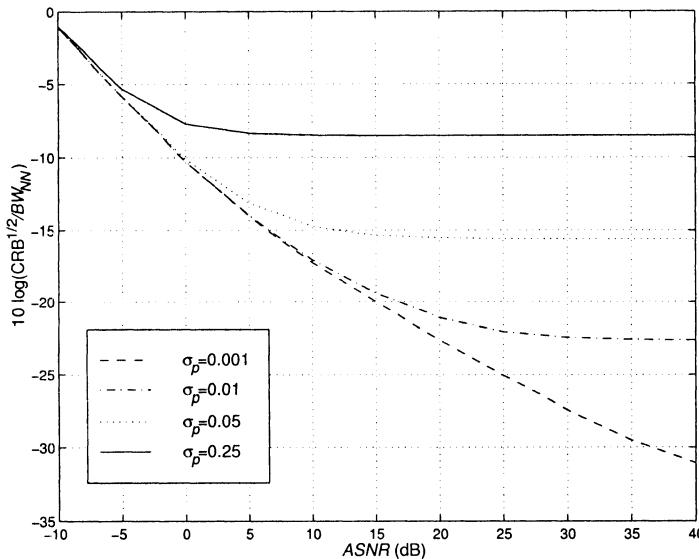


Figure 8.61 Normalized hybrid Cramér-Rao bound on DOA versus $ASNR$ for various σ_p : SLA 10 array, two plane-wave signals with $\Delta\phi/2 = \pm 2.55^\circ$, $K = 100$.

hybrid Cramér-Rao bound for the plane wave at $\Delta\phi/2$. We also plot CRB_0 , the CRB for perfect sensor location information. Note that these are angle bounds in contrast to the wavenumber bounds we used in earlier sections.

In Figure 8.61, we plot the CRB on DOA for $\Delta\phi = \Delta\psi_R$ (5.1°) versus $ASNR$ for various σ_p . We see that the bounds level off as the $ASNR$ increases. The departure from the zero calibration error bound occurs at an $ASNR$ that decreases as σ_p increases.

The second example of interest is a circular array that satisfies the constraints in (i).

Example 8.11.2

Consider a 33-element circular array located in the x - y -plane. The nominal locations correspond to a standard uniform circular array with interelement spacing of $\lambda/2$. The location of the first element is fixed on the x -axis. The location of the seventeenth element is fixed in the y -direction and allowed to vary in the x -direction. Thus, the model satisfies the identifiability criterion in (i). Three equal-power uncorrelated plane waves impinge on the array from $\phi_1 = -\Delta\phi$, $\phi_2 = 0$, and $\phi_3 = \Delta\phi$. For notational simplicity, we rewrite (8.686) as

$$[C_{HCR}(\phi, \rho)]^{-1} = \frac{2K}{\sigma_w^2} \cdot Re \left[\begin{bmatrix} \mathbf{F}_{\phi\phi} & \mathbf{F}_{\phi x} & \mathbf{F}_{\phi y} \\ \mathbf{F}_{\phi x}^H & \mathbf{F}_{xx} & \mathbf{F}_{xy} \\ \mathbf{F}_{\phi y}^H & \mathbf{F}_{xy}^H & \mathbf{F}_{yy} \end{bmatrix} + \begin{bmatrix} 0 & 0 & 0 \\ 0 & \Lambda_x^{-1} & 0 \\ 0 & 0 & \Lambda_y^{-1} \end{bmatrix} \right]. \quad (8.690)$$

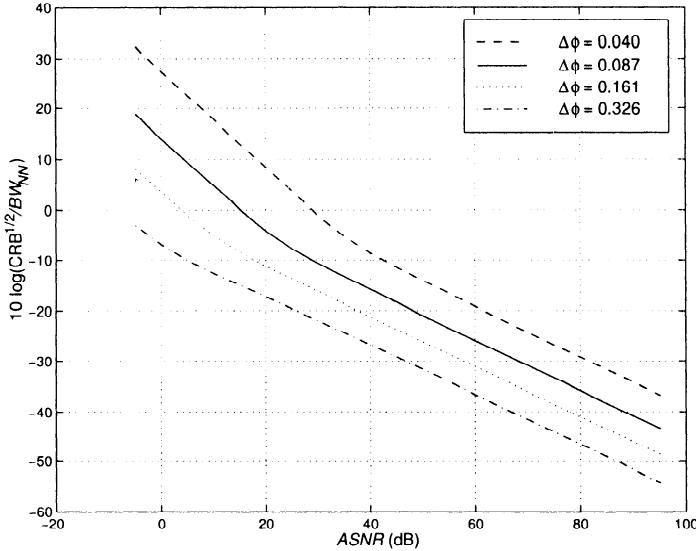


Figure 8.62 Normalized hybrid Cramér-Rao bound (average of three signals) versus $ASNR$: circular array with $N = 33$, three plane-wave signals at 0 and $\pm\Delta\phi$ for various $\Delta\phi$, $\sigma_p = 0.05\lambda$.

We delete the first column in \mathbf{F}_{ϕ_x} , \mathbf{F}_{xx} , \mathbf{F}_{xy} and the first and seventeenth columns in \mathbf{F}_{ϕ_y} , \mathbf{F}_{xy} , and \mathbf{F}_{yy} . We delete the first rows in \mathbf{F}_{xx} , \mathbf{F}_{xy} and the first and seventeenth row in \mathbf{F}_{yy} . We delete the first elements in $\mathbf{\Lambda}_x^{-1}$ and the first and seventeenth element in $\mathbf{\Lambda}_y^{-1}$.

In Figure 8.62, we plot the CRB (average of three signals) versus $ASNR$ for various $\Delta\phi$. The standard deviation of the perturbation is $\sigma_p = 0.05\lambda$. We see that the CRBs are parallel and continue to decrease as the $ASNR$ increases.

Similar bounds for gain and phase perturbations follow directly from (8.677) (e.g., [Fla00]).

8.11.3 Sensitivity of ML Estimators

In this section, we discuss the sensitivity of the ML algorithms of Sections 8.5–8.7 to array perturbations. Our approach requires a simulation of a particular algorithm in the presence of the array perturbations. We restrict our attention to a single example that conveys the type of behavior we can expect. We consider the CML-AP algorithm that we previously discussed in Example 8.6.1.

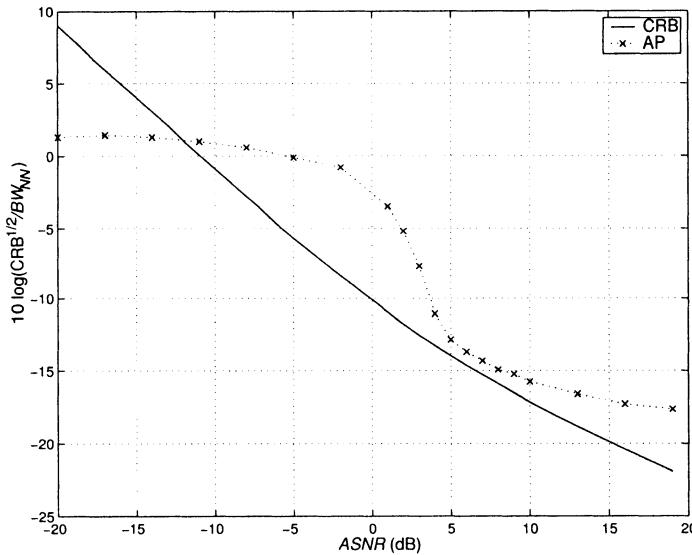


Figure 8.63 Normalized RMSE versus $ASNR$; nominal linear array with sensor position perturbations; $N = 10$, $K = 100$, two plane-wave signals with $\Delta u = \Delta\psi_R$, $\rho = 0$, $\sigma_p = 0.05\lambda$.

Example 8.11.3 (continuation, Example 8.6.1)

Consider the same model as in Example 8.6.1. There are two equal-power uncorrelated signals impinging on an array that is nominally a standard 10-element linear array along the x -axis. The signal separation is $\Delta\psi_R$. We use the AP algorithm to estimate the DOAs.

There are no gain or phase perturbations. The sensor positions are perturbed. The p_{x_n} and p_{y_n} , $n = 0, \dots, 9$, are statistically independent Gaussian random variables with standard deviation σ_p .

The algorithm does not use this information about the displacements. It implements the algorithm assuming the nominal array manifold.

In Figure 8.63, we plot the normalized RMSE versus $ASNR$ for $\sigma_p = 0.05\lambda$. We see that the RMSE is flattening out as the $ASNR$ increases.

In this example, the RMSE does not approach the CRB. However, it flattens out at a reasonably low value.

A discussion of the sensitivity of the ML algorithm is given in Friedlander [Fri90].

8.11.4 MAP Joint Estimation

The formulation of the joint MAP estimation problem follows naturally from our hybrid model. Wahlberg et al. [WOV91] has discussed this model and

Viberg and Swindlehurst [VS94] have extended their development. Our discussion follows [WOV91].

The cost function of interest is

$$F_{map}(\boldsymbol{\theta}, \boldsymbol{\rho}) = F_{ml}(\boldsymbol{\theta}, \boldsymbol{\rho}) + \frac{1}{2}(\boldsymbol{\rho} - \boldsymbol{\rho}_0)\Lambda_{\boldsymbol{\rho}}^{-1}(\boldsymbol{\rho} - \boldsymbol{\rho}_0). \quad (8.691)$$

For the stochastic ML model, we use the AML model. From (8.315),

$$\begin{aligned} F_{ml}(\boldsymbol{\theta}, \boldsymbol{\rho}) &= F_{aml}(\boldsymbol{\theta}, \boldsymbol{\rho}) \\ &= K \ln \det \left[\mathbf{P}_{\mathbf{V}} \mathbf{C}_{\mathbf{x}} \mathbf{P}_{\mathbf{V}} + \frac{\text{tr} [\mathbf{P}_{\mathbf{V}}^{\perp} \mathbf{C}_{\mathbf{x}}] \mathbf{P}_{\mathbf{V}}^{\perp}}{N - D} \right]. \end{aligned} \quad (8.692)$$

For the conditional model, we use a scaled version of the WSF_{ao} estimator (see (8.367)–(8.370)). Asymptotically,

$$F_{ml}(\boldsymbol{\theta}, \boldsymbol{\rho}) \simeq \frac{K}{\hat{\sigma}_w^2} F_{WSF_{ao}}(\boldsymbol{\theta}, \boldsymbol{\rho}), \quad (8.693)$$

where

$$\hat{\sigma}_w^2 = \frac{1}{N - D} \sum_{i=D+1}^N \hat{\lambda}_i, \quad (8.694)$$

and

$$F_{WSF_{ao}} = \text{tr} \left[\mathbf{P}_{\mathbf{V}}^{\perp}(\boldsymbol{\theta}) \widehat{\mathbf{U}}_S \left[\widehat{\Lambda}_S - \hat{\sigma}_w^2 \mathbf{I} \right]^2 \widehat{\Lambda}_S^{-1} \widehat{\mathbf{U}}_S \right]. \quad (8.695)$$

Thus, we define

$$F_{wmap} \triangleq \frac{K}{\hat{\sigma}_w^2} F_{WSF_{ao}}(\boldsymbol{\theta}, \boldsymbol{\rho}) + \frac{1}{2}(\boldsymbol{\rho} - \boldsymbol{\rho}_0)\Lambda_{\boldsymbol{\rho}}^{-1}(\boldsymbol{\rho} - \boldsymbol{\rho}_0). \quad (8.696)$$

We can maximize the expression in (8.696) by using one of the algorithms discussed in Section 8.7.

We can obtain a simpler expression if we assume the variations in $\boldsymbol{\rho}$ are small. Expanding (8.691) around $\boldsymbol{\rho}_0$, we have, for small variations around the nominal value,

$$\begin{aligned} F_{map}(\boldsymbol{\theta}, \boldsymbol{\rho}) &\simeq F_{ml}(\boldsymbol{\theta}, \boldsymbol{\rho}_0) + \left[\frac{\partial F_{ml}(\boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \boldsymbol{\rho}} \Big|_{\boldsymbol{\rho}=\boldsymbol{\rho}_0} \right]^T (\boldsymbol{\rho} - \boldsymbol{\rho}_0) \\ &\quad + \frac{1}{2}(\boldsymbol{\rho} - \boldsymbol{\rho}_0)^T \left\{ \left[\frac{\partial^2 F_{ml}(\boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}^T} \right]_{\boldsymbol{\rho}=\boldsymbol{\rho}_0} + \Lambda_{\boldsymbol{\rho}}^{-1} \right\} (\boldsymbol{\rho} - \boldsymbol{\rho}_0). \end{aligned} \quad (8.697)$$

We can maximize $F_{map}(\boldsymbol{\theta}, \boldsymbol{\rho})$ with respect to $\boldsymbol{\rho}$,

$$\hat{\boldsymbol{\rho}} = \boldsymbol{\rho}_0 - \left[\frac{\partial^2 F_{ml}(\boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}^T} \Big|_{\boldsymbol{\rho}=\boldsymbol{\rho}_0} + \boldsymbol{\Lambda}_{\boldsymbol{\rho}}^{-1} \right]^{-1} \frac{\partial F_{ml}(\boldsymbol{\theta}, \boldsymbol{\rho})}{\partial \boldsymbol{\rho}} \Big|_{\boldsymbol{\rho}=\boldsymbol{\rho}_0}. \quad (8.698)$$

Using (8.698) in (8.697) gives a concentrated cost function,

$$\begin{aligned} \tilde{V}_{map}(\boldsymbol{\theta}, \boldsymbol{\rho}) &= F_{ml}(\boldsymbol{\theta}, \boldsymbol{\rho}_0) \\ &- \frac{1}{2} \left[\frac{\partial F_{ml}(\boldsymbol{\theta}, \boldsymbol{\rho}_0)}{\partial \boldsymbol{\rho}} \right]^T \left[\frac{\partial^2 F_{ml}(\boldsymbol{\theta}, \boldsymbol{\rho}_0)}{\partial \boldsymbol{\rho} \partial \boldsymbol{\rho}} + \boldsymbol{\Lambda}_{\boldsymbol{\rho}}^{-1} \right]^{-1} \frac{\partial F_{ml}(\boldsymbol{\theta}, \boldsymbol{\rho}_0)}{\partial \boldsymbol{\rho}}. \end{aligned} \quad (8.699)$$

We now have reduced the problem to a D -dimensional minimization rather than $D + M$. Wahlberg et al. [WOV91] suggested using the approximation in (8.693) in (8.699) and refer to resulting algorithm as the MAPprox algorithm. Wahlberg et al. [WOV91] and Viberg and Swindlehurst [VS94] have studied the algorithm. Jansson et al. [JSO98] developed a generalized weighted subspace fitting (GWSF) algorithm that is a generalization of the MODE and WSF algorithms developed in Sections 8.5.3 and 8.7.3 (e.g. Stoica and Sharman [SS90a] and Viborg and Ottersten [VO91]).

The advantage of the GWSF algorithm is that it can be implemented using a two-step procedure when the nominal array is a uniform linear array (see Section 8.7.3). The algorithm is also consistent even if the signals are fully correlated. The paper also contains a large sample analysis of the MAPprox algorithm and a summary of earlier work on the model error problem. The reader is referred to these references ([WOV91], [VS94], and [JSO98]) for further discussion.

The other approach to joint estimation treats $\boldsymbol{\rho}$ as an unknown nonrandom vector. We discuss this approach briefly in the next section.

8.11.5 Self-Calibration Algorithms

The basic idea of self-calibration algorithms is to formulate a joint estimation problem that contains the parameters of interest. The goal is to calibrate an array using a set of plane-wave signals whose DOAs and powers are unknown. This leads us to a joint estimation problem.

Many of the self-calibration algorithms have the following characteristics:

- (i) They estimate the DOAs assuming a nominal array configuration. They then estimate the array parameters assuming the DOAs are correct. They repeat the iteration until the estimates converge. This can be

viewed as a joint estimation problem using “group alternating maximization.”

- (ii) They do not assume an *a priori* density of the array parameters. However, they do assume a nominal configuration to initialize the iteration in (i).
- (iii) They frequently use a simpler DOA estimator than ML. The reasons for this simplification will be easier to explain after we have discussed some of the simpler algorithms in Chapter 9.

There are discussions of self-calibration algorithms in Weiss and Friedlander (e.g., [FW88], [WF89], or Chapter 10 in [Hay91b], [WF91b]), Weiss et al. [WWL88], Rockah and Schultheiss [RS87a], [RS87b], and Porat and Friedlander [PF97].

8.11.6 Summary

In this section we have discussed how array perturbations affect the performance of DOA algorithms. After reviewing the perturbation model, we derived the CRB for the joint estimation problem of ϕ and ρ . We then studied the performance of the CML algorithm that we had derived in Section 8.6. In order to improve performance we introduced the technique of joint estimation (or autocalibration), but did not develop it in detail.

8.12 Summary

In Section 8.12.1, we summarize the major results in Chapter 8. In Section 8.12.2, we briefly discuss some related topics. In Section 8.12.3, we provide a brief introduction to Chapter 9.

8.12.1 Major Results

In this chapter we have studied algorithms for estimating the parameters of a signal arriving at an antenna array. Although the formulation was general, most of our discussion focussed on estimating the directions of arrival of a set of plane waves impinging on the array.

In Section 8.2, we reviewed classical ML estimation and MAP estimation. We also reviewed three versions of the Cramér-Rao bound: the classic CRB, the Bayesian CRB, and the hybrid CRB. These results form the background for most of later developments in this chapter.

In Section 8.3, we developed the models that we use in the parameter estimation model. The model of most interest consisted of multiple plane waves impinging on the array in the presence of additive noise. We emphasized two temporal models for the signals. In the first model, the signals were sample functions from a zero-mean vector stationary complex Gaussian random processes with an unknown source spectral matrix \mathbf{S}_f . In the second model, the signals were considered to be unknown nonrandom complex sequences. We modeled the additive noise as a zero-mean vector complex Gaussian random process that is temporally white and spatially uncorrelated. We also introduced array perturbation models and spatially spread models.

In Section 8.4, we developed CRBs for the various models. These CRBs are a key issue in any parameter estimation discussion for two reasons:

- (i) They provide a lower bound on the covariance matrix of any unbiased estimator.
- (ii) Above some value of K (the number of snapshots) and $ASNR$ (the array signal-to-noise ratio), in many cases, the ML estimator reaches the bound.

There were different bounds for the different signal models. One can show that the CRB using the conditional signal model was also greater than or equal to the CRB using the unconditional (or stochastic) CRB. However, in most cases, in the $ASNR-K$ region where the CRB is conveying useful information (where the performance of the ML estimator approaches the bound), the two bounds practically coincide.

In Section 8.5, we studied ML estimators in detail. For a fixed K , we found that the ML estimator exhibited a threshold phenomenon. Above a certain $ASNR$, the performance of the ML estimator approaches the CRB. However, as the $ASNR$ was decreased, a threshold point was reached where the estimator started making large errors, so the MSE increased quickly. Figures 8.17–8.25 helped explain this behavior. We argued that estimators should be judged according to two criteria:

- (i) Behavior above threshold. The estimator should approach the CRB. Generally, this behavior can be studied analytically using asymptotic or Taylor series techniques.
- (ii) The location of the threshold. We want the estimator to have its threshold at as low a value of $ASNR$ (for a given K) as possible.

We emphasized two types of ML estimators; the unconditional (or stochastic) ML estimator (AML) and the conditional (or deterministic) ML estimator (CML and the related MODE/WSF estimator). In the examples that we considered, there was very little difference in performance.

In the simulations in Section 8.5 we utilized a grid search to ensure that we found the global maximum. This approach is computationally demanding for two plane-wave signals and generally not feasible for a larger number of signals.

In Section 8.6, we developed computational algorithms for finding the ML estimators. We introduced several gradient algorithms, two alternating maximization (AM) algorithms, and the expectation-maximization (EM) algorithm. The AM algorithms, which maximize each variable separately in an iterative manner, were efficient when a suitable initialization was possible. The EM algorithm converged more slowly in most examples. We did not study the gradient algorithms in detail. All of the algorithms in this section were applicable to an arbitrary array geometry.

In Section 8.7, we developed polynomial parameterization techniques that were applicable to standard linear arrays. After introducing polynomial parameterization, we derived the IQML and IMODE algorithms. Both algorithms required less computation than our previous algorithms and provided similar threshold performance.

In Section 8.8, we revisited the problem of estimating the number of plane-wave signals that were present in the input to the array. We emphasized the separable detection algorithms (AIC, MDL) that we had studied earlier and assumed that the output of the detection step was used as an input to estimation algorithm. In some cases, the threshold of the detection algorithms occurred at the same *ASNR* as the threshold of the ML estimator operating with the correct number of signals. We suggested that, if we must operate in the vicinity of the threshold, we should use AIC or AIC-FB and require subsequent processing (after the estimator) to eliminate any extra signals introduced by an over-estimation error from AIC. If we are operating above threshold, we should use MDL or MDL-FB because it provides consistent results. We introduced the idea of joint detection and estimation, but did not pursue it because of the computational requirements.

In Section 8.9, we introduced the parameter estimation problem for spatially spread signals and discussed it briefly. Both models that we discussed were parametric models. After deriving the CRBs, we considered several simple examples. We found that for a single spatially spread signal, even a small amount of spread caused the CRB to level off as the *ASNR* increased. For the single-signal case, the multiple-parameter ML estimator approached

the CRB above threshold. The ML estimator for the multiple-signal problem did not appear to be practical because of the computational complexity.

In Section 8.10, we discussed parameter estimation in beamspace. We first developed conditions on the beamspace such that the beamspace CRB would equal the element-space CRB. Several beamspace matrices, such as the Taylor series preprocessor and the discrete prolate spheroidal functions had a CRB_{bs} that came very close to the CRB_{els} over a useful range of signal separation Δu . The conventional DFT beamspace processor also had a CRB_{bs} that was close to the CRB_{els} for signals in the center part of the beamspace sector. We derived the beamspace ML estimator and evaluated its performance. The MSE of the BS-ML estimator approached the CRB_{bs} above the threshold $ASNR$.

In Section 8.11, we discussed the effect of array perturbations on the performance of our parameter estimation algorithms. We modeled the problem as a combined estimation problem in which we estimated both the signal and noise parameters $(\psi_s, \mathbf{S}_f, \sigma_w^2)$ and the perturbation parameters. After reviewing the perturbation model, we derived the CRBs for the composite parameter set. The resulting CRB also applies to the case in which we only estimate the signal and noise parameters. For the case of position perturbations, we found that the CRB for a linear array leveled off as the $ASNR$ increased. We investigated the performance of the conventional ML estimator and found there was significant degradation. We introduced the idea of a joint estimator called the MAPprox algorithm, but did not pursue it in detail.

8.12.2 Related Topics

One area of interest that we did not explore can be referred to as **structured adaptive beamforming**. Recall that the purpose of the adaptive beamformer was to estimate the waveform of the desired signal. In the literature, this is sometimes referred to as the “signal copy” problem. In Chapters 6 and 7 we normally assumed that we knew the direction-of-arrival of the desired signal, but did not impose any additional structure on the environment. However, in many applications, we can impose the structure by assuming that

$$\mathbf{X}(k) = \mathbf{V}(\psi)\mathbf{F}(k) + \mathbf{W}(k), \quad k = 1, 2, \dots, K, \quad (8.700)$$

which is familiar as the model used in most of this chapter.

The basic idea is straightforward. If we are using the unconditional model (Section 8.5.1), we estimate ψ , \mathbf{S}_f , and σ_w^2 . We then use these estimates, as if they were correct, in one of the beamformers in Chapter 7. An early

reference that studied a similar approach was the Lincoln Laboratory Report by Barabell et al. [BCD+84]. They conducted extensive simulations of various DOA estimation algorithms and explored the calibration issue. One can refer to this approach as DF-based signal copy.

Various analyses of this problem have appeared in the literature (e.g., Ottersten et al. [ORK89], Friedlander and Weiss [WF93], Wax [Wax85], and Yang and Swindlehurst [YS95]). The paper by Yang and Swindlehurst [YS95] contains a good summary of results. Note that we are imposing a structure on the signal and noise model and not the array geometry. The reader is referred to these references for a discussion of this problem.

A second problem of interest that we did not explore is analytic expressions for the behavior of a ML estimator (or any other estimator) as the $ASNR$ or K decreased. We did not develop bounds on the performance. We refer to this as the **nonlinear behavior** region.

The covariance predicted by the CRB corresponds to small errors in the estimator performance and was directly related to the shape of the main lobe of the beam pattern. In our simulations, we observed that as the number of observations or SNR decreased we reached a point where the estimator was making large errors and the performance deviated significantly from the CRB.

We would like to develop bounds that enable us to predict this nonlinear behavior. The bounds can be used for several purposes.

The first purpose is to predict the value of SNR or K where the “threshold” occurs. The threshold is defined to be the point where the MSE of the estimator starts to deviate from the CRB. If we have control over the system parameters (e.g., an active radar or sonar) we will design the system so it operates above the threshold.

The second purpose of the bounds is to describe the behavior in the transition region below the threshold. In many systems, we will have to operate in this region and need to predict the behavior.

There are two models that are used in the nonlinear bound literature. In the first model, we treat the vector parameter as an unknown nonrandom quantity. We utilized this approach in DEMT I [VT68] [VT01a] (see pp. 71, 147, 284–286, 386) and developed the Bhattacharyya and Barankin bounds. We can extend these results to the array processing case and develop several hybrid bounds.

There are several approaches in the statistics literature for this model that have been successfully applied to parameter estimation problem by various researchers.

The first approach is the Barankin bound [Bar49]. The original version

is a greatest lower bound, but is computationally complex so that simplified versions (e.g., Chapman and Robbins [CR51], Hammersley [Ham50], and Kiefer [Kie52]) are normally used. Application of the Barankin bound to the array processing problem was first done by Baggeroer [Bag69] and later by Becker [Bec77] and Chow and Schultheiss [CS81].

The Barankin bound has been applied in the radar area by Swerling [Swe59], McAulay and Seidman [MS69], and McAulay and Hofstetter [MH71].

The second approach uses the Bhattacharyya bound [Bha46], which is an extension of the CRB utilizing higher derivatives of the likelihood function.

We can also derive a hybrid bound that contains the Cramér-Rao, Bhattacharyya, Barankin, and Hammersley-Chapman-Robbins bound as special cases. The hybrid bound³⁷ is due to Abel [Abe93], [Abe90], and is based on a derivation using the covariance inequality in Ibragimov and Has'minskii [IH81] and Lehmann [Leh83]. The reader is referred to these references for a discussion of these approaches.

In the second model we treat the parameter as a random variable with a known probability density and develop bounds on the MSE of the estimator. These bounds extend into the threshold and transition region.

The bounds using the first model are local bounds that depend on the actual value of the parameter θ . These bounds are useful in many applications. One of their disadvantages is that they fail to limit the MSE when the parameter space is finite. In many cases, bounds exceed the MSE of the ML estimator in the low SNR region. This behavior can be attributed both to the violation of the unbiased estimator assumption and to the lack of *a priori* information in the bounds. Similar observations were made in [Abe93]. Clearly, these bounds must be used cautiously for performance analysis in bearing estimation problems where the parameter space is compact and estimators are inherently biased at low SNR .

There are two general approaches to the global bound problems. The first approach is referred to as extended Ziv-Zakai (EZZ) bounds ([Bel95], [BEVT96]). This is a Bayesian bound that assumes that the parameter is a random variable with a known *a priori* distribution. They provide a bound on the global MSE averaged over the *a priori* probability density function (pdf). There are no restrictions on the class of estimators to which Bayesian bounds apply, and they incorporate knowledge of the *a priori* parameter space via the prior distribution. The basic approach in the Ziv-Zakai bound and its extension is to relate the MSE to the probability of error in a binary

³⁷Note that we are using “hybrid” in a different manner than in Section 8.2.3.3. In that section, hybrid denotes a mix of real unknown parameters and random parameters. Here, hybrid denotes a mixture of different bounds on real unknown parameters.

detection problem. In order for the bound to be useful we must be able to evaluate or bound the resulting probability of error.

There have been a number of results on global bounds over the last three decades. The original Ziv-Zakai bound [ZZ69] was improved by Chazan et al. [CZZ75] and Bellini and Tartara [BT74]. The disadvantage was that the bounds were restricted to scalar parameters with uniform *a priori* densities. In Bell [Bel95], the EZZ bound was developed that was applicable to vector parameters and arbitrary *a priori* densities. The reader is referred to the above references for a discussion of these bounds.

The second approach to global bounds was developed by Weiss and Weinstein [WW83], [WW84]. It is based on the $\mu(s)$ function that we discussed in Section 2.7 of DEMT I [VT68] [VT01a]. In some situations it provides a tighter bound than the extended Ziv-Zakai bound. In other situations, the EZZ bound provides better results. The reader is referred to the above references for a discussion of the Weiss-Weinstein bounds.

The above references provide a number of useful results. However, we were unable to apply these results to accurately estimate the threshold where we estimate the DOAs of two plane waves in additive noise.³⁸

The third related topic is the problem of DOA estimation in the presence of **unknown correlated** noise. There are a number of papers dealing with this problem. Both parametric and nonparametric noise models have been utilized. Parametric models have been utilized by Le Cadre [LeC89], Bohme and Kraus [BK88], and Nagesha and Kay [NK96]. A useful parametric noise model is the spatial ARMA model that we discussed in Section 5.7. Nonparametric approaches include the work of Wu and Wong ([WW94], [WW95]), Wong et al. [WRWQ92], Reilly and Wong [RW92], Stoica et al. [SVWW96], Stoica et al. [SOV92], Wax [Wax92], Prasad et al. [PWMS88], Paulraj and Kailath [PK86], Ye and DeGroat [YDG95], Rajagopal et al. [RKR94], and Harmanci et al. [HTK00]. Papers that utilize an instrument variable approach include Stoica et al. [SVO94] and Stoica et al. [SVWW98].

We discuss other related topics in Section 9.10.

8.12.3 Algorithm complexity

In this chapter, we have focused on ML estimators and the CRB. By utilizing AP or AM techniques and IQML and IMODE where applicable, we were able to reduce the computational complexity. Further reductions were obtained

³⁸This inability was the primary reason we did not include a more thorough discussion of non-linear bounds.

by utilizing beamspace processing. In spite of these improvements, the ML algorithms still required a fair amount of computation.

In Chapter 9, we develop parameter estimation algorithms that require less computation. We analyze their performance and compare their performance to the ML estimators developed in this chapter and the CRB.

8.13 Problems

P8.2 Maximum Likelihood and Maximum *A Posteriori* Estimators

Problem 8.2.1

Consider the problem of estimating the direction of arrival of D plane waves using a linear array. In our subsequent discussions, we usually compute the CRB on the vector ψ , where

$$[\psi]_i = \pi u_i = \pi \cos \theta_i, i = 1, \dots, D, \quad (8.701)$$

where θ_i is the angle of the i th plane wave with respect to the z-axis.

Show that

$$\mathbf{C}_{CR}(\boldsymbol{\theta}) = \mathbf{G}_{\boldsymbol{\theta}}^{-1} \mathbf{C}_{CR}(\psi) \mathbf{G}_{\boldsymbol{\theta}}^{-1}, \quad (8.702)$$

where

$$\mathbf{G}_{\boldsymbol{\theta}}^{-1} \triangleq \text{diag} \{1/\pi \sin \theta_1, \dots, 1/\pi \sin \theta_D\}. \quad (8.703)$$

Discuss your result.

Problem 8.2.2 (continuation)

Consider the converse problem of going from $\mathbf{C}_{CR}(\boldsymbol{\theta})$ to $\mathbf{C}_{CR}(\psi)$. Show that

$$\mathbf{C}_{CR}(\psi) = \mathbf{G}_{\psi}^{-1} \mathbf{C}_{CR}(\boldsymbol{\theta}) \mathbf{G}_{\psi}^{-1}, \quad (8.704)$$

and find \mathbf{G}_{ψ} . How are \mathbf{G}_{ψ} and $\mathbf{G}_{\boldsymbol{\theta}}$ related?

P8.3 Parameter Estimation Models

Problem Note 8.3.1: The ln likelihood function is the starting point for most of the development in this chapter. In all of the following problems, develop the ln likelihood function for the indicated signal and noise model and identify the unknown parameters. Unless noted, we are considering narrowband processes. We assume that there are D signals and that ψ_i is 1-D. There are K statistically independent samples.

Problem 8.3.1

Signal case S1 and noise case N1.

Problem 8.3.2

Signal case S1 and noise case N2.

Problem 8.3.3

Signal case S1 and noise case N3.

Problem 8.3.4

Signal case S1 and noise case N3. There are D signals. Noise consists of D_N uncorrelated plane waves whose DOA and power are unknown plus white noise with unknown σ_w^2 .

Problem 8.3.5

Signal case S1 and noise case N3. There are D signals. The array is an SLA. Noise consists of a complex spatial AR(1) process (Section 5.7) plus white noise with unknown σ_w^2 .

Problem 8.3.6 (continuation)

Repeat Problem 8.3.5 for the case in which the noise consists of the sum of D_N complex spatial AR(1) processes plus white noise with unknown σ_w^2 .

Problem 8.3.7 (continuation, Problem 8.3.5)

Repeat Problem 8.3.5 for the case in which the noise consists of a complex spatial AR(p) process plus white noise with unknown σ_w^2 .

Problem 8.3.8

Repeat Problem 8.3.1 for signal case S2.

Problem 8.3.9

Repeat Problem 8.3.2 for signal case S2.

Problem 8.3.10

Repeat Problem 8.3.3 for signal case S2.

Problem 8.3.11

Repeat Problem 8.3.4 for signal case S2.

Problem 8.3.12

Repeat Problem 8.3.5 for signal case S2.

Problem 8.3.13

Repeat Problem 8.3.6 for signal case S2.

Problem 8.3.14

Repeat Problem 8.3.7 for signal case S2.

Problem 8.3.15

Repeat Problem 8.3.1 for signal case S2u.

Problem 8.3.16

Repeat Problem 8.3.2 for signal case S2u.

Problem 8.3.17

Repeat Problem 8.3.3 for signal case S2u.

Problem 8.3.18

Repeat Problem 8.3.4 for signal case S2u.

Problem 8.3.19

Repeat Problem 8.3.5 for signal case S2u.

Problem 8.3.20

Repeat Problem 8.3.6 for signal case S2u.

Problem 8.3.21

Repeat Problem 8.3.7 for signal case S2u.

Problem 8.3.22

Repeat Problem 8.3.1 for signal case S3.

Problem 8.3.23

Repeat Problem 8.3.2 for signal case S3.

Problem 8.3.24

Repeat Problem 8.3.3 for signal case S3.

Problem 8.3.25

Repeat Problem 8.3.4 for signal case S3.

Problem 8.3.26

Repeat Problem 8.3.5 for signal case S3.

Problem 8.3.27

Repeat Problem 8.3.6 for signal case S3.

Problem 8.3.28

Repeat Problem 8.3.7 for signal case S3.

Problem Note 8.3.2: The next problems consider model perturbations.

The parameter perturbation vector ρ is M -dimensional. For example, in the case of position perturbations, $M = 2N$.

$$\rho = [p_{x_0} \quad p_{y_0} \quad p_{x_1} \quad \cdots \quad p_{y_{N-1}}]^T. \quad (8.705)$$

We assume that ρ is a real Gaussian random vector

$$p\rho(\rho) = \frac{1}{(2\pi)^{\frac{M}{2}} |\Lambda\rho|^{\frac{1}{2}}} \exp \left\{ -\frac{1}{2} (\rho - \rho_0)^T \Lambda^{-1} \rho + \frac{1}{2} \rho_0^T \Lambda^{-1} \rho_0 \right\}, \quad (8.706)$$

where ρ_0 represents the nominal value of the parameter vector ρ . The Gaussian assumption is a good model for many applications.

Problems 8.3.29–8.3.31 consider position perturbations. Problems 8.3.32–8.3.34 consider gain and phase perturbations.

Problem 8.3.29

Consider a SLA10 along the x -axis. The position perturbation model is given by (8.73)–(8.80). We assume $\Lambda\rho$ is diagonal. Note that p_{x_n} and p_{y_n} are constant over the entire observation period. Repeat Problem 8.3.1 for this model.

Problem 8.3.30 (continuation)

Repeat Problem 8.3.2 for the perturbation model in Problem 8.3.29.

Problem 8.3.31 (continuation, Problem 8.3.29)

Repeat Problem 8.3.3 for the perturbation model in Problem 8.3.29.

Problem 8.3.32

Consider a SLA10 along the x -axis. The gain and phase perturbation model is given by (8.73)–(8.80). We assume $\Lambda\rho$ is diagonal. Repeat Problem 8.3.1 for this model.

Problem 8.3.33 (continuation)

Repeat Problem 8.3.2 for the perturbation model in Problem 8.3.32.

Problem 8.3.34 (continuation, Problem 8.3.32)

Repeat Problem 8.3.3 for the perturbation model in Problem 8.3.32.

P8.4 Cramér-Rao Bounds

Problem Note 8.4.1: The first set of problems considers the test scenarios in Table 8.3 at the end of Section 8.5.1.1. We also use these test scenarios in Sections P.8.5, P.8.6, P.8.7, P.8.10, and P.8.11. Table 8.5 shows where the various test scenarios are used.

Table 8.5 Problem Structure for SLA 10.

Test Scenarios are described in Table 8.3

	TS1	TS2	TS3	TS4	TS5
8.4			P.8.4.1	P.8.4.2	
8.5		Ex.8.5.3 P.8.5.1	Ex.8.5.5 P.8.5.3	Ex.8.5.4	P.8.5.2
8.6	Ex.8.6.1 P.8.6.1 P.8.6.8		Ex.8.6.2 P.8.6.3		
8.7		P.8.7.1	P.8.7.3		P.8.7.2
8.10	P.8.10.1		P.8.10.2	P.8.10.3	
8.11	P.8.11.1 P.8.11.7		P.8.11.2 P.8.11.8	P.8.11.3 P.8.11.9	
	TS6	TS7	TS8	TS9	TS10
8.4		P.8.4.3	P.8.4.5	P.8.4.4	P.8.4.6
8.5	Ex.8.5.6 P.8.5.4				
8.6		P.8.6.3	P.8.6.5	P.8.6.4	P.8.6.6
8.7	P.8.7.4	P.8.7.5	P.8.7.6		P.8.7.7
8.10		P.8.10.4		P.8.10.5	P.8.10.6
8.11		P.8.11.4 P.8.11.10		P.8.11.5 P.8.11.11	

Problem Note 8.4.2: We derive the CRB for a number of array geometries and signal scenarios. In most cases, we find an estimator in a subsequent problem and compare its performance.

Problem 8.4.1

Consider a SLA10. There are two uncorrelated plane-wave signals impinging on the array from $\pm\Delta\psi/2$, where $\Delta\psi = \Delta\psi_R$. $SNR_1 = 100ASNR_2$.

Plot the normalized CRB($10 \log(\text{CRB}^{\frac{1}{2}}/BW_{NN})$) versus $ASNR$.

Problem 8.4.2

Consider a SLA10. There are two equal-power correlated signals impinging on the array from $\pm\Delta\psi/2$, where $\Delta\psi = \Delta\psi_R$. The correlation coefficient is $|\rho| \exp(j\phi_\rho)$. Consider three values of $|\rho|$: 0.95, 0.99, and 1.0. Consider three values of ϕ_ρ : 0, $\pi/4$, and $\pi/2$. Plot the normalized CRB versus $ASNR$.

Problem 8.4.3

Consider a SLA10. There are three equal-power uncorrelated plane-wave signals impinging on the array from $\psi = 0$, $\psi = \Delta\psi$, and $\psi = -\Delta\psi$, where $\Delta\psi = \Delta\psi_R$. Plot the normalized CRB versus $ASNR$.

Problem 8.4.4 (continuation, Problem 8.4.3)

Repeat Problem 8.4.3 for the case in which the two plane waves at $\pm\Delta\psi$ have an SNR that is 10 dB higher than the plane wave at $\psi = 0$.

Problem 8.4.5 (continuation, Problem 8.4.3)

Consider the same model as in Problem 8.4.3 except the signals are correlated with unequal-power. Denote the signals as: No.1, $\psi = -\Delta\psi$; No.2, $\psi = 0$; No.3, $\psi = \Delta\psi$, where $\Delta\psi = \Delta\psi_R$:

$$\begin{aligned} SNR_1 &= SNR_3 = 0.5SNR_2 \\ \rho_{12} &= 0.9, \quad \rho_{23} = 0.9, \quad \rho_{13} = 0.5 \exp(j\pi/2) \end{aligned}$$

Plot the normalized CRB versus $ASNR$.

Problem 8.4.6 (continuation, Problem 8.4.3)

Consider the model in Problem 8.4.3. Repeat the problem for the case of five equal-power uncorrelated plane waves located at $\psi = 0, \pm\Delta\psi, \pm2\Delta\psi$, where $\Delta\psi = \Delta\psi_R$. Plot the normalized CRB versus $ASNR$.

Problem Note 8.4.3: The next set of problems considers a 32-element standard linear array (SLA32). The problems with two and three signals are similar to the earlier problems. The larger array size allows us to consider more complicated scenarios.

Problem 8.4.7

Consider a SRA32. There are two equal-power uncorrelated signals impinging on the array at $\pm\Delta u/2$, where $\Delta u = 1/32$.

Problem 8.4.8 (continuation)

Repeat Problem 8.4.7 with $\Delta u = 1/128$.

Problem 8.4.9 (continuation, Problem 8.4.7)

Repeat Problem 8.4.7 with $\Delta u = 1/32$ and $ASNR_1 = 100ASNR_2$.

Problem 8.4.10 (continuation, Problem 8.4.7)

Repeat Problem 8.4.7 with $\Delta u = 1/32$ and $\rho = 0.95$, $\rho = 0.95 \exp(j\pi/4)$, and $\rho = 0.95 \exp(j\pi/2)$.

Problem 8.4.11 (continuation, Problem 8.4.7)

- (a) Repeat Problem 8.4.7 with three equal-power uncorrelated signals located at $-1/32, 0, 1/32$.
- (b) Repeat part(a) with $ASNR_1 = ASNR_3 = 100ASNR_2$.

Problem 8.4.12 (continuation, Problem 8.4.7)

Repeat Problem 8.4.7 for the following cases:

- (a) Five equal-power uncorrelated signals located at $\pm 2/32, \pm 1/32, 0$.
- (b) Seven equal-power uncorrelated signals located at $\pm 3/32, \pm 2/32, \pm 1/32, 0$.
- (c) Fifteen equal-power uncorrelated signals located at $\pm m/32, m = 0, \dots, 7$.

Problem Note 8.4.4:

The next several examples consider low redundancy arrays that were introduced in Section 3.9.2.

Problem 8.4.13

Consider the 5-element linear array in Table 3.9. The sensor spacing is

$$1 \cdot 3 \cdot 5 \cdot 2.$$

Two equal-power uncorrelated signals impinge on the array from $\pm \Delta u/2$, where $\Delta u = 0.0866$.

- (a) Plot the square root of the CRB ($10 \log \text{CRB}^{\frac{1}{2}}$) versus SNR . We do not normalize because of the comparison in parts (b) and (c).
- (b) Compare your result to those for an SLA with the same number of elements.
- (c) Compare your result to those for an SLA with the same aperture.
- (d) Discuss your results.

Problem 8.4.14 (continuation, Problem 8.4.13)

Repeat Problem 8.4.13 for the 7-element linear array in Table 3.9. The sensor spacing is

$$1 \cdot 3 \cdot 6 \cdot 8 \cdot 5 \cdot 2.$$

Problem 8.4.15 (continuation, Problem 8.4.13)

Repeat Problem 8.4.13 for the 9-element linear array in Table 3.9. The sensor spacing is

$$1 \cdot 4 \cdot 7 \cdot 13 \cdot 2 \cdot 8 \cdot 6 \cdot 3.$$

Problem 8.4.16 (continuation, Problem 8.4.13)

Repeat Problem 8.4.13 for the 10-element linear array in Table 3.9. The sensor spacing is

$$1 \cdot 5 \cdot 4 \cdot 13 \cdot 3 \cdot 8 \cdot 7 \cdot 12 \cdot 2.$$

Problem Note 8.4.5: The next several examples consider planar arrays. We utilize (8.268) and obtain the asymptotic conditional CRB (ACCR).

Problem 8.4.17

Consider a standard 10×10 rectangular array. There are two equal-power uncorrelated plane waves impinging on the array. Their (θ, ϕ) directions are $(34^\circ, 45^\circ)$ and $(26^\circ, 45^\circ)$, respectively. Plot the ACCR ($10 \log \text{ACCR}^{\frac{1}{2}}$) versus ASNR .

Problem 8.4.18 (continuation)

Repeat Problem 8.4.17 for the case in which $\text{ASNR}_1 = 100 \text{ASNR}_2$.

Problem 8.4.19 (continuation, Problem 8.4.17)

Repeat Problem 8.4.17 for the case in which the sources are correlated. Consider $\rho = 0.95$ and $0.95 \exp(j\pi/2)$.

Problem 8.4.20 (continuation, Problem 8.4.17)

Consider a standard 10×10 rectangular array. There are three equal-power uncorrelated plane waves impinging on the array. Their (θ, ϕ) directions are $(45^\circ, 50^\circ)$, $(45^\circ, 45^\circ)$, $(45^\circ, 40^\circ)$ respectively. Repeat Problem 8.4.17.

Problem Note 8.4.6: The next several problems consider hexagonal arrays. In order to get some comparison with a 10×10 rectangular array, we use the standard 91-element hexagonal array in Example 4.4.1.

Problem 8.4.21

Repeat Problem 8.4.17. Compare your results to those in Problem 8.4.17.

Problem 8.4.22

Repeat Problem 8.4.18. Compare your results to those in Problem 8.4.18.

Problem 8.4.23

Repeat Problem 8.4.19. Compare your results to those in Problem 8.4.19.

Problem 8.4.24

Repeat Problem 8.4.20. Compare your results to those in Problem 8.4.20.

Problem 8.4.25

Consider a cross array consisting of an N -element SLA along the x -axis and an M -element SLA along the y -axis. The linear arrays are symmetric about the origin. N and M are even, so there is no element at the origin.

Use the signal model in Problem 8.4.17 and assume $N = M = 10$. Plot the ACCR ($10 \log \text{ACCR}^{\frac{1}{2}}$) versus ASNR . Compare your results to those in Problem 8.4.17.

Problem 8.4.26

Derive the stochastic CRB for the planar array model. Compare your result to the ACCR in (8.268).

P8.5 Maximum Likelihood Estimation

Problem 8.5.1 (continuation, Example 8.5.3)

Simulate the CML algorithm for the model in Example 8.5.3. Compare your results to the AML results. (Test scenario 2.)

Problem 8.5.2 (continuation, Example 8.5.4)

Consider the model in Example 8.5.4 except $|\rho| = 1$ and $\phi_\rho = \pi/2$. (Test scenario 5.) Simulate the AML, CML, and MODE algorithms. Compare the results of three algorithms.

Problem 8.5.3 (continuation, Example 8.5.5)

- (a) Repeat Example 8.5.5 for $ASNR_1 = 10ASNR_2$.
- (b) Repeat Example 8.5.5 for the CML and MODE algorithms. Compare your results. (Test scenario 3.)

Problem 8.5.4 (continuation, Example 8.5.6)

Repeat Example 8.5.6 for the CML and MODE algorithm. (Test scenario 6.) Compare your results.

Problem 8.5.5 (continuation, Example 8.5.2)

Repeat Example 8.5.2 with $\Delta\psi = 0.5BW_{NN}$. Note that this is the classical resolution separation. Plot the same results as in Figures 8.21 and 8.24. Discuss your results.

Problem 8.5.6 (continuation, Example 8.5.4)

Repeat Example 8.5.4 with $\Delta\psi = 0.5BW_{NN}$. Plot the same results as in Figures 8.27 and 8.28. Also plot the case when $\rho = 0.95 \exp(j\pi/2)$. Consider preprocessing the data using FB averaging.

Problem 8.5.7 (continuation, Example 8.5.5)

Repeat Example 8.5.5 with $\Delta\psi = 0.5BW_{NN}$. Plot the same results as in Figures 8.29 and 8.30.

P8.6 Computational Algorithms

Problem Note 8.6.1: All of the algorithms in this section require initialization. A bad initialization can result in convergence to local maximum (or minimum) instead of the global maximum (or minimum). We use the initialization procedure in (8.434) and (8.435).

Problem 8.6.1 (continuation, Example 8.6.1)

Consider the same model as in Example 8.6.1. Solve for the CML estimate using the quasi-Newton technique in Section 8.6.1. Discuss your results. (Test scenario 1.)

Problem 8.6.2 (continuation, Example 8.5.5)

Consider the model in Example 8.5.5. Use the AP technique to find the CML estimate. Use the AM technique to find the AML estimate. Compare your results. (Test scenario 3.)

Problem 8.6.3 (continuation, Problem 8.4.3)

Consider the same model as in Problem 8.4.3. Use the AP technique to find the CML estimate. Use the AM technique to find the AML estimate. Compare your results to the CRB. Discuss your results. (Test scenario 7.)

Problem 8.6.4 (continuation, Problem 8.4.4)

Repeat Problem 8.6.3 for the model in Problem 8.4.4. (Test scenario 9.)

Problem 8.6.5 (continuation, Problem 8.4.5)

Repeat Problem 8.6.3 for the model in Problem 8.4.5. (Test scenario 8.)

Problem 8.6.6 (continuation, Problem 8.4.6)

Repeat Problem 8.6.3 for the model in Problem 8.4.6. (Test scenario 10.)

Problem 8.6.7 (continuation, Problem 8.4.17)

Repeat Problem 8.6.3 for the model in Problem 8.4.17.

Problem 8.6.8

Consider a standard 10-element linear array. There are two equal-power uncorrelated plane-wave signals impinging on the array.

- (a) Use the EM algorithm in Section 8.6.3.2 to find $\hat{\psi}_{cml}$. Compare your results to the AP and discuss the computational requirements.
- (b) Read [FH94]. Implement their algorithm for this problem.

Problem Note 8.6.2: The SLA is a useful example. However, in practice, we will normally use the algorithms that is developed in Section 8.7 for SLAs. The next set of problems considers linear arrays with non-uniform spacing where the techniques in Section 8.7 do not apply.

Problem 8.6.9 (continuation, Problem 8.4.13)

Consider the 5-element linear array in Problem 8.4.13 and the same signal model. Find the CML estimate using the AP algorithm. Plot the normalized RMSE versus $ASNR$ and compare it to the CRB derived in Problem 8.4.13.

Problem 8.6.10 (continuation, Problem 8.4.14)

Consider the 7-element linear array in Problem 8.4.14 and the same signal model. Find the CML estimate using the AP algorithm. Plot the normalized RMSE versus $ASNR$ and compare it to the CRB derived in Problem 8.4.14.

Problem 8.6.11 (continuation, Problem 8.4.15)

Consider the 9-element linear array in Problem 8.4.15 and the same signal model. Find the CML estimate using the AP algorithm. Plot the normalized RMSE versus $ASNR$ and compare it to the CRB derived in Problem 8.4.15.

Problem 8.6.12 (continuation, Problem 8.4.16)

Consider the 10-element linear array in Problem 8.4.16 and the same signal model. Find the CML estimate using the AP algorithm. Plot the normalized RMSE versus $ASNR$ and compare it to the CRB derived in Problem 8.4.16.

P8.7 Polynomial Parameterization

Problem Note 8.7.1: The next several problems consider a SLA10 with a set of plane waves impinging on the array. In each problem, simulate IQML and IMODE. Plot the normalized RMSE ($10 \log(RMSE/BW_{NN})$) versus $ASNR$ for $K = 100$. Compare your results to the CRB and the previous ML implementations. Compare the number of iterations required.

Problem 8.7.1

There are two equal-power uncorrelated signals and $\Delta\psi = 0.05BW_{NN}$. (Test scenario 2.)

Problem 8.7.2

There are two equal-power coherent signals: $\Delta\psi = \Delta\psi_R$ and $\rho = 1 \exp(j\pi/2)$. (Test scenario 5.)

Problem 8.7.3

There are two uncorrelated signals: $\Delta\psi = \Delta\psi_R$ and $ASNR_2 = 100ASNR_1$. (Test scenario 3.)

Problem 8.7.4

There are two equal-power uncorrelated signals: $\Delta\psi = \Delta\psi_R$ and $K = 20$ instead of 100. (Test scenario 6.)

Problem 8.7.5

There are three equal-power uncorrelated signals: $\psi_1 = -\Delta\psi$, $\psi_2 = 0$, and $\psi_s = \Delta\psi$, where $\Delta\psi = \Delta\psi_R$. (Test scenario 7.)

Problem 8.7.6 (continuation, Problem 8.4.5)

Use the same signal model as in Problem 8.4.5. (Test scenario 8.)

Problem 8.7.7 (continuation, Problem 8.4.6)

Use the same signal model as in Problem 8.4.6. (Test scenario 10.)

Problem 8.7.8

In this problem, an alternative derivation of (8.509) is developed. Assume that D is odd and let

$$D = 2q + 1. \quad (8.707)$$

Define

$$\mathbf{b} = [\mathbf{b}_1^T \quad \mathbf{b}_2^T]^T, \quad (8.708)$$

where \mathbf{b} is given in (8.489) and \mathbf{b}_1 and \mathbf{b}_2 are $(q + 1) \times 1$ vectors where

$$\mathbf{b}_2 = \mathbf{J} \mathbf{b}_1^*. \quad (8.709)$$

The \mathbf{A}_k matrix in (8.498) is partitioned in two $(N - D) \times (q + 1)$ matrices,

$$\mathbf{A} = [\mathbf{A}_1 \mid \mathbf{A}_2], \quad (8.710)$$

where the k subscript is suppressed.

Now define

$$\tilde{\mathbf{A}} = [\mathbf{A}_1 + \mathbf{A}_2 \mathbf{J} \mid j(\mathbf{A}_1 - \mathbf{A}_2 \mathbf{J})], \quad (8.711)$$

and

$$\mathbf{c} = [[Re[\mathbf{b}_1]]^T \mid [Im[\mathbf{b}_1]]^T]. \quad (8.712)$$

Note that \mathbf{c} in (8.712) contains the same elements as the \mathbf{c} in (8.504), but they are arranged in a different order.

Show that

$$\mathbf{Ab} = \tilde{\mathbf{A}}\mathbf{c}. \quad (8.713)$$

Problem 8.7.9 (continuation)

In this problem, the IQML algorithm for even D is derived using the model in Problem 8.7.8. Let

$$D = 2q. \quad (8.714)$$

Define

$$\mathbf{b} = [\mathbf{b}_1^T \ b_3 \ \mathbf{b}_2^T]^T, \quad (8.715)$$

where \mathbf{b}_1 and \mathbf{b}_2 are $q \times 1$ vectors that satisfy (8.709) and b_3 is a real scalar. Partition \mathbf{A} as

$$\mathbf{A} = [\mathbf{A}_1 \ \mathbf{A}_3 \ \mathbf{A}_2], \quad (8.716)$$

where \mathbf{A}_1 and \mathbf{A}_2 are $(N - D) \times q$ matrices and \mathbf{A}_3 is an $(N - D) \times 1$ vector. Define

$$\tilde{\mathbf{A}} = [\mathbf{A}_1 + \mathbf{A}_2 \mathbf{J} \ \mathbf{A}_3 \ j(\mathbf{A}_1 - \mathbf{A}_2 \mathbf{J})]. \quad (8.717)$$

Define

$$\mathbf{c} = [[Re[\mathbf{b}_1]]^T \ b_3 \ [Im[\mathbf{b}_1]]^T]^T. \quad (8.718)$$

Show that

$$\mathbf{A} \mathbf{b} = \tilde{\mathbf{A}} \mathbf{c}. \quad (8.719)$$

Problem 8.7.10

In this problem, the IQML algorithm for even D is developed. In this case, the \mathbf{b} vector satisfies the constraints:

$$b_0 = b_{D-i}^*, \quad i = 0, \dots, D/2 - 1, \quad (8.720)$$

and

$$b_{\frac{D}{2}} = b_{\frac{D}{2}}^*, \quad (8.721)$$

so $b_{D/2}$ is real.

Define \mathbf{c} as

$$\mathbf{c} = [Re[b_0] \ Im[b_0] \ \dots \ Re[b_{\frac{D}{2}-1}] \ Im[b_{\frac{D}{2}-1}] \ b_{\frac{D}{2}}]. \quad (8.722)$$

Find \mathbf{T} such that

$$\mathbf{b} = \mathbf{T} \mathbf{c}. \quad (8.723)$$

Problem 8.7.11

FB-IMODE is the standard IMODE algorithm using FB averaging in the sample spectral matrix (e.g., [SJ97]). In this problem, we compare the performance of FB-IMODE and FO-IMODE for different parameter values. In each part, plot the normalized RMSE versus $ASNR$ for FB-IMODE and FO-IMODE. Also plot the normalized CRB. In the MODE algorithms, use $\epsilon = 0.01$ and a maximum of eight iterations. The array is a 10-element SLA and $K = 100$. There are two equal-power plane wave signals separated by $\Delta\psi$.

- (a) $\Delta\psi = 0.2165BW_{NN}$, $\rho = 0$;
- (b) $\Delta\psi = 0.05BW_{NN}$, $\rho = 0$;
- (c) $\Delta\psi = 0.4BW_{NN}$, $\rho = 0$;
- (d) $\Delta\psi = 0.2165BW_{NN}$, $\rho = |1|$, consider $\phi_\rho = 0, 0.9\pi, 0.8\pi, 0.85\pi, 0.7\pi, 0.6\pi$, and 0.5π ;

- (e) Repeat part (d) for $\Delta\psi = 0.05BW_{NN}$;
- (f) Repeat part (d) for $\Delta\psi = 0.4BW_{NN}$.

Discuss your results.

Problem 8.7.12

The vector \mathbf{b} is defined in (8.489). Consider the case of a single plane-wave signal impinging on the array from $\psi = 0$. Show that $Re[b_o] = 0$.

P8.8 Detection of Number of Signals

Problem Note 8.8.1: Our discussion and most of the discussions in the literature assume that the number of signals, D , is known. In practice, we usually have to estimate D . In the next nine problems we revisit Examples 8.7.1 and 8.7.2 and Problems 8.7.1–8.7.7.

In each problem, simulate the AIC-FB and MDL-FB algorithm to estimate D . Plot

$$\begin{aligned} P_D &\triangleq Pr[\hat{D} = D] \\ P_M &\triangleq Pr[\hat{D} < D] \\ P_{FA} &\triangleq Pr[\hat{D} > D] \end{aligned}$$

versus $ASNR$ for $K = 100$. Use the output \hat{D} in both the IQML or IMODE estimation algorithm. Plot the normalized RMSE versus $ASNR$. In order to calculate the RMSE when $\hat{D} > D$, assume that the D estimates are paired with the closest true DOA and the excess estimates are ignored. Calculate the number of trials where $\hat{D} < D$, but do not assign an RMSE to the missed signal. Compare your results to those in the corresponding example or problem from Section 8.7.

Problem 8.8.1 (continuation, Example 8.7.1)

Use the signal model in Example 8.7.1.

Problem 8.8.2 (continuation, Example 8.7.2)

Use the signal model in Example 8.7.2.

Problem 8.8.3 (continuation, Problem 8.7.1)

Use the signal model in Problem 8.7.1.

Problem 8.8.4 (continuation, Problem 8.7.2)

Use the signal model in Problem 8.7.2.

Problem 8.8.5 (continuation, Problem 8.7.3)

Use the signal model in Problem 8.7.3.

Problem 8.8.6 (continuation, Problem 8.7.4)

Use the signal model in Problem 8.7.4.

Problem 8.8.7 (continuation, Problem 8.7.5)

Use the signal model in Problem 8.7.5.

Problem 8.8.8 (continuation, Problem 8.7.6)

Use the signal model in Problem 8.7.6.

Problem 8.8.9 (continuation, Problem 8.7.7)

Use the signal model in Problem 8.7.7.

Problem 8.8.10

Read Wax and Ziskind [WZ89]. Simulate their algorithm for the model in Problem 8.8.1. Compare your results to those in Problem 8.8.1. Discuss the relative computational complexity.

Problem 8.8.11

Read Wax [Wax91]. Simulate his algorithm for the model in Problem 8.8.4. Compare your results to those in Problem 8.8.4. Discuss the relative computational complexity.

Problem 8.8.12

Read Cho and Djuric [CD94]. Simulate their algorithm for the model in Problem 8.8.1. Compare your results to those in Problem 8.8.1. Discuss the relative computational complexity.

P8.9 Spatially Spread Signals**Problem 8.9.1**

Consider Example 8.9.3.

- (a) Discuss the relationship between σ_u and the CRB in terms of beam pattern of the array.
- (b) Show that the cross-matrices (8.564) are zero for the single signal case.

Problem 8.9.2 (continuation, Example 8.9.4)

Consider the 4-parameter model. Assume $\psi_{C_1} = -\Delta\psi/2$ and $\psi_{C_2} = \Delta\psi/2$. Assume $\sigma_{\psi_2} = \sigma_{\psi_1} = 0.3\Delta\psi/2$. Assume $K = 100$. Plot the CRB ($10 \log \text{CRB}^{1/2}$) versus $ASNR$ for $\Delta\psi = 0.2, 0.4$, and 0.6 . Note that the bound is not normalized.

Problem 8.9.3

Consider a SLA10. There is a single signal spatially spread signal impinging on the array. The spatial spectrum can be modeled as a complex AR(1) process with known power. Assume that the white noise is negligible.

Plot the CRB for various parameter values.

P8.10 Beamspace Algorithms**Problem Note 8.10.1:**

The first set of problems focuses on beamspace Cramér-Rao bounds. We consider many of same arrays and signals as in Section 8.4. In each problem, we consider three types of beamspace matrices:

- (a) DFT beamspace (8.581), (8.582).
- (b) Taylor series beamspace (8.588), (8.589).
- (c) DPSS beamspace (8.598), (8.599).

The dimension of the beamspace is a design parameter. In element space, the CRB in the ψ - or u -space only depends on the signal separation. In beamspace, the CRB also

depends on the actual location because the center of the beamspace may be mismatched. In each problem we consider three values for the location of the center of the signal set, $\psi_c = -\pi/N$, 0, and π/N .

In each problem, plot the normalized CRB versus $ASNR$ for $K = 100$ and compare them to the element-space CRBs in Section 8.4.

The first six problems consider a SLA10.

Problem 8.10.1 (continuation, Example 8.4.4)

Consider two equal-power uncorrelated plane-wave signals: $\psi_1 = \psi_c - \Delta\psi/2$, $\psi_2 = \psi_c + \Delta\psi/2$, where $\Delta\psi = \Delta\psi_R$. (Test scenario 1, modified.)

Problem 8.10.2 (continuation, Problem 8.4.1)

Consider two uncorrelated plane-wave signals: $\psi_1 = \psi_c - \Delta\psi/2$, $\psi_2 = \psi_c + \Delta\psi/2$, where $\Delta\psi = \Delta\psi_R$. $ASNR_2 = 100ASNR_1$. (Test scenario 3, modified.)

Problem 8.10.3 (continuation, Problem 8.4.2)

Consider the same signal directions as in Problem 8.10.1. The signals are correlated using the values in Problem 8.4.2. (Test scenario 4, modified.)

Problem 8.10.4 (continuation, Problem 8.4.3)

Consider three equal-power uncorrelated plane-wave signals: $\psi_1 = \psi_c - \Delta\psi$, $\psi_2 = \psi_c$, $\psi_3 = \psi_c + \Delta\psi$, where $\Delta\psi = \Delta\psi_R$. (Test scenario 7, modified.)

Problem 8.10.5 (continuation, Problem 8.4.4)

Consider the same signal model as in Problem 8.10.4, except the signals at ψ_1 and ψ_3 have an SNR that is 10 dB higher than the signal at ψ_2 . (Test scenario 9, modified.)

Problem 8.10.6 (continuation, Problem 8.4.6)

Consider five equal-power uncorrelated plane-wave signals: $\psi_1 = \psi_c - 2\Delta\psi$, $\psi_2 = \psi_c - \Delta\psi$, $\psi_3 = \psi_c$, $\psi_4 = \psi_c + \Delta\psi$, $\psi_5 = \psi_c + 2\Delta\psi$. (Test scenario 10, modified.)

Problem Note 8.10.2: The next six problems consider a SLA32.

Problem 8.10.7 (continuation, Problem 8.4.7)

Consider two equal-power uncorrelated signals at $\psi_1 = \psi_c - \pi/32$ and $\psi_1 = \psi_c + \pi/32$.

Problem 8.10.8 (continuation, Problem 8.4.8)

Consider two equal-power uncorrelated signals at $\psi_1 = \psi_c - \pi/128$ and $\psi_1 = \psi_c + \pi/128$.

Problem 8.10.9

Consider the same model as in Problem 8.10.7 except the signals are correlated. Let $\rho = 0.95$, $0.95 \exp(j\pi/4)$, and $0.95 \exp(j\pi/2)$.

Problem 8.10.10

- (a) Consider three equal-power uncorrelated signals located at $\psi_1 = \psi_c - \pi/32$, $\psi_2 = 0$, and $\psi_3 = \psi_c + \pi/32$.
- (b) Repeat part (a) with $ASNR_1 = ASNR_3 = 100ASNR_2$.

Problem 8.10.11

- (a) Consider five equal-power uncorrelated signals located at

$$\psi_i = \psi_c + \frac{\pi(i-3)}{32}, \quad i = 1, \dots, 5. \quad (8.724)$$

- (b) Consider seven equal-power uncorrelated signals located at

$$\psi_i = \psi_c + \frac{\pi(i-4)}{32}, \quad i = 1, \dots, 7. \quad (8.725)$$

Problem 8.10.12

Consider the signal model in Problem 8.10.7 and assume $\psi_c = \pi/128$. We use a DFT beamspace matrix with $N_{bs} = 5$. Find $\hat{\psi}_{cml,bs}$ (8.630) by using the AP algorithm.

Discuss your results.

Problem 8.10.13

Consider the signal model in Problem 8.10.10 and assume $\psi_c = 0$. We use a DFT beamspace matrix with $N_{bs} = 5$. Find $\hat{\psi}_{cml,bs}$ (8.630) by using the AP algorithm.

Discuss your results.

Problem 8.10.14

Consider the same signal model as in Problem 8.10.12. Repeat Problem 8.10.12 using the beamspace IQML algorithm. Compare your results, including computational issues.

Problem 8.10.15 (continuation, Problem 8.10.13)

Repeat Problem 8.10.14 for the signal model in Problem 8.10.13.

P8.11 Sensitivity, Robustness, and Calibration

Problem Note 8.11.1: The next five problems consider a 10-element linear array that is nominally a SLA along the x -axis. The sensor location perturbation model is given in (8.661) and (8.662). The $p_{x,i}$ and $p_{y,i}$ perturbations are statistically independent with $\sigma_p = 0.05\lambda$, which is 10% of the nominal value. We consider different signal models and evaluate the element-space hybrid CRB, the beamspace hybrid CRB, and the RMSE of various algorithms:

- (a) Element-space hybrid CRB.
- (b) CML and MODE AP in element space.
- (c) CML and MODE AP in beamspace (DFT BS matrix with $N_{bs} = 5$).
- (d) Element-space IQML.
- (e) Beamspace IQML.

In each part, plot the normalized RMSE versus $ASNR$ for $K = 100$. Include the $\sigma_p = 0$ case as a reference. Note that the sensor positions do not change during the snapshot sequence. Discuss your results.

Problem 8.11.1

Two equal-power uncorrelated plane-wave signals impinge on the array from $\pm\Delta\psi/2$, where $\Delta\psi = \Delta\psi_R$. (Test scenario 1.)

Problem 8.11.2

Two uncorrelated plane-wave signals impinge on the array from $\pm\Delta\psi/2$, where $\Delta\psi = \Delta\psi_R$, $ASNR_2 = 100ASNR_1$. (Test scenario 3.)

Problem 8.11.3

Two equal-power correlated plane-wave signals impinge on the array from $\pm\Delta\psi/2$, where $\Delta\psi = \Delta\psi_R$, $\rho = 0.95 \exp(j\pi/4)$. (Test scenario 3.)

Problem 8.11.4

Three equal-power uncorrelated plane-wave signals impinge on the array; $\psi_1 = -\Delta\psi$, $\psi_2 = 0$, $\psi_3 = \Delta\psi$, where $\Delta\psi = \Delta\psi_R$. (Test scenario 7.)

Problem 8.11.5 (continuation)

Consider the same model as in Problem 8.11.4 except $ASNR_1 = ASNR_3 = 100ASNR_2$. (Test scenario 9.)

Problem Note 8.11.2: The next set of problems considers sensor gain and phase perturbation. We use the model in Section 6.3.3 with no sensor position perturbations. We let $\sigma_g = 0.02$ and $\sigma_\gamma = .05$.

In Problem 8.11.6, we derive the hybrid CRB for gain and phase perturbations. In Problems 8.11.7–8.11.11 we simulate the performance for the same signal models as in Problems 8.11.1–8.11.5. In each problem, do the five parts in Problem Note 8.11.1 and plot the results.

Problem 8.11.6

Use the expression in (8.677) as a starting point. Derive the hybrid CRB for the case of gain and phase perturbations.

Problem 8.11.7 (continuation, Problem 8.11.1)

Repeat Problem 8.11.1 for the sensor gain and phase perturbation parameters in Problem Note 8.11.2. (Test scenario 1.)

Problem 8.11.8 (continuation, Problem 8.11.2)

Repeat Problem 8.11.2 for the sensor gain and phase perturbation parameters in Problem Note 8.11.2. (Test scenario 3.)

Problem 8.11.9 (continuation, Problem 8.11.3)

Repeat Problem 8.11.3 for the sensor gain and phase perturbation parameters in Problem Note 8.11.2. (Test scenario 4.)

Problem 8.11.10 (continuation, Problem 8.11.4)

Repeat Problem 8.11.4 for the sensor gain and phase perturbation parameters in Problem Note 8.11.2. (Test scenario 7.)

Problem 8.11.11 (continuation, Problem 8.11.5)

Repeat Problem 8.11.5 for the sensor gain and phase perturbation parameters in Problem Note 8.11.2. (Test scenario 9.)

Bibliography

- [Abe90] J. S. Abel. A bound on mean-square-estimate error. *Proc. ICASSP*, Albuquerque, New Mexico, vol.3, pp. 1345–1348, April 1990.
- [Abe93] J. S. Abel. A bound on mean-square-estimate error. *IEEE Trans. Inf. Theory*, vol. IT-39, pp. 1675–1680, September 1993.
- [AGGS98] Y. I. Abramovich, D.A. Gray, A.Y. Gorokhov and N.K. Spencer. Positive-definite Toeplitz completion in DOA estimation for nonuniform linear antenna arrays—Part I: Fully augmentable arrays. *IEEE Trans. Signal Process.*, vol.SP-46, pp. 2458–2471, September 1998.
- [Aka74] H. Akaike. A new look at the statistical model identification. *IEEE Trans. Autom. Control*, vol.AC-19, pp. 716–723, June 1974.
- [AN95] S. Anderson and A. Nehorai. Optimal dimension reduction for array processing-generalized. *IEEE Trans. Signal Process.*, vol.SP-43, pp. 2025–2027, August 1995.
- [And63] T. W. Anderson. Asymptotic theory for principal component analysis. *Ann. Math. Stat.*, vol.34, pp. 122–148, 1963.
- [And71] T. W. Anderson. *The Statistical Analysis of Time Series*. Wiley, New York, 1971.
- [And84] T. W. Anderson. *An Introduction to Multivariate Statistical Analysis*. Wiley, New York 1984.
- [And91] S. Anderson. Optimal dimension reduction for sensor array signal processing. *Proc. 25th Asilomar Conf. on Signals, Systems and Computers*, Monterey, California, pp. 918–922, November 1991.
- [And93] S. Anderson. On optimal dimension reduction for sensor array signal processing. *IEEE Trans. Signal Process.*, vol.SP-30, pp. 245–256, January 1993.
- [Bag69] A. B. Baggeroer. Barankin bound on the variance of estimates of Gaussian random processes. Technical Report, MIT Lincoln Laboratory, Lexington, Massachusetts, January 1969.
- [Ban71] W. J. Bangs. *Array Processing with Generalized Beamformers*. PhD thesis, Yale University, New Haven, Connecticut, 1971.
- [Bar49] E. W. Barankin. Locally best unbiased estimates. *Ann. Math. Statist.*, vol.20, pp. 477–501, 1949.
- [Bar54] M. S. Bartlett. A note on the multiplying factors for various χ^2 approximations. *J. R. Stat. Soc.*, vol.16, pp. 296–298, 1954.
- [BCD+84] A. J. Barabell, J. Capon, D. F. DeLong, J. R. Johnson, and K. D. Senne. Performance comparison of superresolution array processing algorithms. Technical Report Project Report. TST-72, MIT Lincoln Laboratory, Lexington, MA, May 1984.
- [Bec77] D. Becker. Development of Cramér-Rao bounds on parameter estimation accuracy. Technical Report, Orincon, San Diego, California, January 1977.
- [Bel95] K. L. Bell. *Performance Bounds in Parameter Estimation with Application to Bearing Estimation*. PhD Thesis, George Mason University, Fairfax, Virginia, 1995.

- [Bell99] K. L. Bell. An uncorrelated maximum likelihood estimator. Internal report, Statistical Signal and Array Processing Group, George Mason University C3I Center, Fairfax, Virginia, April 1999.
- [Ben93] G. R. Benitz. Asymptotic results for maximum likelihood estimation with an array of sensors. *IEEE Trans. Inf. Theory*, vol. IT-39, pp. 1374–1386, July 1993.
- [BEVT95] K. L. Bell, Y. Ephraim, and H. L. Van Trees. Ziv-Zakai lower bounds in bearing estimation. *Proc. ICASSP*, Detroit, Michigan, vol.5, pp. 2852–2855, May 1995.
- [BEVT96] K. L. Bell, Y. Ephraim, and H. L. Van Trees. Explicit Ziv-Zakai lower bound for bearing estimation. *IEEE Trans. Signal Process.*, vol. SP-44, pp. 2810–2824, November 1996.
- [Bha46] A. Bhattacharyya. On some analogues of the amount of information and their use in statistical estimation. *Shankya*, vol.8, pp. 201–218, 1946.
- [BK84] G. Bienvenu and L. Kopp. Decreasing high resolution method sensitivity by conventional beamformer preprocessing. *Proc. ICASSP*, San Diego, California, vol.V pp. 3321–3324, 1984.
- [BK88] J. Bohme and D. Kraus. On least squares methods for DOA estimation in unknown noise fields. *Proc. ICASSP*, New York, New York, pp. 2833–2836, 1988.
- [BM86] Y. Bresler and A. Macovski. Exact maximum likelihood parameter estimation of superimposed exponential signals in noise. *IEEE Trans. Acoust., Speech, and Signal Process.*, vol. ASSP-34, pp. 1081–1089, October 1986.
- [Boh86] J. F. Böhme. Separated estimation of wave parameters and spectral parameters by maximum likelihood. *Proc. ICASSP*, Tokyo, Japan, pp. 2819–2822, 1986.
- [Boh89] J. F. Böhme. Location and spectrum estimation by approximate maximum likelihood. *Proc. SPIE*, Orlando, Florida, vol.1, pp. 326–337, 1989.
- [Bre88] Y. Bresler. Maximum likelihood estimation of a linearly structured covariance with application to antenna array processing. *Proc. 4th ASSP Workshop on Spectrum Estimation and Modeling*, Minneapolis, Minnesota, pp. 172–175, August 1988.
- [BT74] S. Bellini and G. Tartara. Bounds on error in signal parameter estimation. *IEEE Trans. Commun. Technol.*, vol. COM22, pp. 340–342, March 1974.
- [BX90] K. Buckley and X.L. Xu. Spatial-spectrum estimation in a location sector. *IEEE Trans. Acoust. Speech, Signal Process.*, vol. ASSP-38, pp. 1842–1852, November 1990.
- [CD94] C. M. Cho and P. M. Djuric. Detection and estimation of DOAs of signals via Bayesian predictive densities. *IEEE Trans. Signal Process.*, vol. SP-42, pp. 3051–3060, November 1994.
- [Cin75] E. Cinlar. *Introduction to Stochastic Processes*. Prentice-Hall, Englewood Cliffs, New Jersey, 1975.
- [CR51] D. G. Chapman and H. Robbins. Minimum variance estimation without regularity assumption. *Ann. Math. Stat.*, vol.21, pp. 581–586, 1951.

- [Cra46] H. Cramér. *Mathematical Methods of Statistics*. Princeton University Press, Princeton, New Jersey, 1946.
- [CS81] S. K. Chow and P. M. Schultheiss. Delay estimation using narrow-band processes. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-29, pp. 478–484, June 1981.
- [CS92] M. Clark and L. L. Scharf. On the complexity of IQML algorithms. *IEEE Trans. Signal Process.*, vol.SP-40, pp. 1811–1813, July 1992.
- [CTSD96] C. Chambers, T. C. Tozer, K. C. Sharman, and T. S. Durrani. Temporal and spatial sampling influence on the estimates of superimposed narrowband signals: when less can mean more. *IEEE Trans. Signal Process.*, vol.SP-44, pp. 3085–3098, December 1996.
- [CZZ75] D. Chazan, M. Zakai, and J. Ziv. Improved lower bounds on signal parameter estimation. *IEEE Transactions Inf. Theory*, vol.IT-21, pp. 90–93, January 1975.
- [Dju90] P. M. Djuric. *Selection of Signal and System Models by Bayesian Predictive Densities*. PhD Thesis, University of Rhode Island, Kingston, 1990.
- [DS83] J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs, New Jersey, 1983.
- [DS96] J. E. Dennis and R. B. Schnabel. *Numerical Methods for Unconstrained Optimization and Nonlinear Equations*. Prentice-Hall, Englewood Cliffs, New Jersey, 1996.
- [DW92] M. A. Doron and A. J. Weiss. On focusing matrices for wide-band array processing. *IEEE Trans. Signal Process.*, vol.SP-40, pp. 1295–1302, June 1992.
- [DWM93] M. A. Doron, A. J. Weiss, and H. Messer. Maximum-likelihood direction finding of wide-band sources. *IEEE Trans. Signal Process.*, vol.SP-41, pp. 411–414, January 1993.
- [EF73] A. G. Evans and R. Fischl. Optimal least squares time-domain synthesis of recursive digital filters. *IEEE Trans. Audio Electroacoust.*, vol.AE-21, pp. 61–65, February 1973.
- [FH94] J. A. Fessler and A. O. Hero. Space-alternating generalized expectation-maximization algorithm. *IEEE Trans. Signal Process.*, vol.SP-42, pp. 4664–4677, October 1994.
- [Fis22] R. A. Fisher. On the mathematical foundations of theoretical statistics. *Phil. Trans. R. Soc., London, Ser. A*, vol.222, pp. 309–368, 1922.
- [Fis25] R. A. Fisher. Theory of statistical estimation. *Proc. Cambridge Phil. Soc.*, vol.22, p. 700, 1925.
- [Fla00] B. P. Flanagan. *Self Calibration of Antenna Arrays with Large Perturbation Errors*. PhD Thesis, George Mason University, Fairfax, Virginia, 2000.
- [FP89] B. Friedlander and B. Porat. The exact Cramér-Rao Bound for Gaussian autoregressive processes. *IEEE Trans. Aerospace Electronic Syst.*, vol.AES-25, pp. 3–8, January 1989.
- [Fri84] B. Friedlander. On the computation of the Cramér-Rao Bound for ARMA parameter estimation. *IEEE Trans. Acoust., Speech Signal Process.*, vol.ASSP-32, pp. 721–727, 1984.

- [Fri90] B. Friedlander. A sensitivity analysis of the maximum likelihood direction finding algorithm. *IEEE Trans. Aerospace Electron. Systems*, vol.AES-25, pp. 953–968, November 1990.
- [Fuc88] J. J. Fuchs. Estimating the number of sinusoids in additive white noise. *IEEE Trans. Acoust., Speech, Signal Proc.*, vol.SSP-36, pp. 1846–1853, December 1988.
- [FV87] P. Forster and G. Vezzosi. Application of spheroidal sequences to array processing. *Proc. ICASSP*, Dallas, Texas, vol.4, pp. 2267–2271, 1987.
- [FW88] M. Feder and E. Weinstein. Parameter estimation of superimposed signals using the EM algorithm. *IEEE Trans. Acoust., Speech, and Sig. Proc.*, vol.ASSP-36, pp. 477–489, April 1988.
- [Gaa88] W.A Gardner. Simplification of MUSIC and ESPRIT by exploitation of cyclostationarity. *Proc. IEEE*, vol.76, pp. 845–847, July 1988.
- [GL86] R. Gooch and J. Lundell. The CM array: An adaptive beamformer for constant modulus signals. *Proc. ICASSP*, Tokyo, Japan, vol. 5, pp. 2523–2526, 1986.
- [GMW81] P. E. Gill, W. Murray, and M. H. Wright. *Practical Optimization*. Academic, London, 1981.
- [GP73] G. Golub and V. Pereyra. The differentiation of pseudoinverses and nonlinear least squares problems whose variables separable. *SIAM J. Numer. Anal.*, vol.10, pp. 413–432, 1973.
- [Gra82] D. A. Gray. Formulation of the maximum signal-to-noise ratio array processor in beam space. *J. Acoust. Soc. Am.*, vol.72, pp. 1195–1201, October 1982.
- [GS99] A. B. Gershman and P. Stoica. On unitary and forward-backward MODE. *Digital Signal Process*, vol.9, pp. 67–75, April 1999.
- [Gu00] H. Gu. Linearization method for finding Cramér-Rao bounds in signal processing. *IEEE Trans. Signal Process.*, vol.48, pp. 543–545, February 2000.
- [Gup65] R. P. Gupta. Asymptotic theory for principal component analysis in the complex case. *J. Indian Stat. Assoc.*, vol.3, pp. 97–106, 1965.
- [Ham50] J. M. Hammersley. On estimating restricted parameters. *J. R. Soc. Ser. B*, vol.12, pp. 192–240, 1950.
- [Hay91a] S. Haykin. *Adaptive Filter Theory*, 2nd ed. Prentice-Hall, Englewood Cliffs, New Jersey, 1991.
- [Hay91b] S. Haykin, editor. *Advances in Spectrum Analysis and Array Processing*, vol.2. Prentice-Hall, Englewood Cliffs, New Jersey, 1991.
- [HB91] Y. D. Huang and M. Barkat. Near-field multiple source localization by passive sensor array. *IEEE Trans. Antennas Propag.*, vol.AP-69, pp. 968–975, July 1991.
- [HS92] S. Haykin, and A. Steinhardt. *Adaptive Radar Detection and Estimation*. Wiley, New York, 1992.
- [HTK00] K. Harmanci, J. Tabrikian, and J. L. Krolik. Relationships between adaptive minimum variance beamforming and optimal source localization. *IEEE Trans. Signal Process.*, vol.SP-48, pp. 1–12, January 2000.

- [Hua94] Y. Hua. The most efficient implementation of the IQML algorithm. *IEEE Trans. Signal Process.*, vol.SP-42, pp. 2203–2204, August 1994.
- [HUSF97] A.O. Hero, M. Usman, A.C. Sauve, and J.A. Fessler. Recursive algorithms for computing the Cramér-Rao bound. *IEEE Trans. Signal Process.*, vol.SP-45, pp. 803–807, March 1997.
- [IH81] I. A. Ibragimov and R. Z. Has'minski. *Statistical Estimation*. Springer-Verlag, New York, 1981.
- [Jaf85] A. G. Jaffer. Maximum likelihood angular resolution of multiple sources. *Proc. 19th Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, California, pp. 68–72, November 1985.
- [Jaf88] A. G. Jaffer. Maximum likelihood direction finding of stochastic sources: A separable solution. *Proc. ICASSP*, New York, vol.5, pp. 2893–2896, April 1988.
- [JGO99] M. Jansson, B. Goransson, and B. Ottersten. A subspace method for direction of arrival estimation of uncorrelated emitter signals. *IEEE Trans. Signal Process.*, vol.SP-47, pp. 945–956, April 1999.
- [JSO98] M. Jansson, A. L. Swindlehurst, and B. Ottersten. Weighted subspace fitting for general array error models. *IEEE Trans. Signal Process.*, vol.SP-46, pp. 2484–2498, September 1998.
- [Kau75] L. Kaufman. A variable projection method for solving separable nonlinear least squares problems. *BIT*, vol.15, pp. 49–57, 1975.
- [Kay84] S. M. Kay. Accurate frequency estimation at low signal-to-noise ratio. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-32, pp. 540–547, June 1984.
- [Kay88] S. M. Kay. *Modern Spectral Estimation*. Prentice-Hall, Englewood Cliffs, New Jersey, 1988.
- [KB93] D. Kraus and J.F. Böhme. Maximum likelihood location estimation of wideband sources using the EM algorithm. *Adaptive Systems in Control and Signal Processing*, 4th IFAC Symposium. pp. 487–491, Pergamon Press, Oxford, UK, 1993.
- [KDB93] D. Kraus, A. Dhaouadi and J.F. Böhme. Dual maximum likelihood estimation for wideband source location. *Proc. ICASSP*, New York, vol.I, pp. 257–260, 1993.
- [Kie52] J. Kiefer. On minimum variance estimators. *Ann. Math. Stat.*, vol.23, pp. 627–629, 1952.
- [KM81] S.M. Kay and S.L. Marple. Spectrum analysis: A modern perspective. *Proc. IEEE*, vol.69, pp. 1380–1419, 1981.
- [KMB92] D. Kraus, D. Maiwald and J.F. Böhme. Maximum likelihood source location estimation via EM algorithm. *Proc. 6th European Signal Process. Conf.* Amsterdam, Netherlands, vol.2, pp. 649–452, 1992.
- [KS61] M. G. Kendall and A. Stuart. *The Advanced Theory of Statistics*. Hafner, New York, 1961.
- [KS85] R. Kumaresan and A. K. Shaw. High resolution bearing estimation without eigendecomposition. *Proc. ICASSP*, Tampa Bay, Florida, vol.1 pp. 576–579, April 1985.

- [KSS86] R. Kumaresan, L. L. Scharf, and A. K. Shaw. An algorithm for pole-zero modelling and spectral analysis. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-34, pp. 637–640, June 1986.
- [KS88] R. Kumaresan, L.L. Scharf, and A.K. Shaw. Superresolution by structured matrix approximation. *IEEE Trans. Antennas Propag.*, vol.AP, pp. 36–44, 1988.
- [KT83] R. Kumaresan and D. W. Tufts. Estimating the angles of arrival of multiple plane waves. *IEEE Trans. Aerospace Electron. Syst.*, vol.AES-19, pp. 134–139, January 1983.
- [Kum85] R. Kumar. A fast algorithm for solving a Toeplitz system of equations. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-33, pp. 254–267, February 1985.
- [Law56] D. N. Lawley. Tests of significance of the latent roots of the covariance and correlation matrices. *Biometrika*, vol.43, pp. 128–136, 1956.
- [LC93] J. Li and R.T. Compton. Maximum likelihood angle estimation for signals with known waveforms. *IEEE Trans. Signal Process.*, vol.SP-41, pp. 2850–2862, September 1993.
- [LDD94] D.A. Linebarger, R.D. DeGroat, and E.M. Dowling. Efficient direction-finding methods employing forward/backward averaging. *IEEE Trans. Signal Process.*, vol.SP-42, pp. 2136–2145, August 1994.
- [LDR77] N. M. Laird, A. P. Dempster, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Ann. R. Stat. Soc.*, pp. 1–38, December 1977.
- [LeC89] J. Le Cadre Parametric methods for spatial signal processing in unknown colored noise fields. *IEEE Trans. Acoust. Speech, Signal Process.*, pp. 965–983, July 1989.
- [Lee92] H. B. Lee. The Cramér-Rao bound on frequency estimates of signals closely spaced in frequency. *IEEE Trans. Signal Process.*, vol.SP-40, pp. 1508–1517, June 1992.
- [Lee94] H. B. Lee. The Cramér-Rao bound on frequency estimates of signals closely spaced in frequency (unconditional case). *IEEE Trans. Signal Process.*, vol.SP-42, pp. 1569–1572, June 1994.
- [Leh83] E. L. Lehmann. *Theory of Point Estimation*. Wiley, New York, 1983.
- [LL94] H. Lee and F. Li. An eigenvector technique for detecting the number of emitters in a cluster. *IEEE Trans. Signal Process.*, vol.SP-42, pp. 2380–2388, September 1994.
- [LSL98] J. Li, P. Stoica and Z. Liu. Comparative study of IQML and MODE direction-of arrival estimators. *IEEE Trans. Signal Process.*, vol.SP-46, pp. 149-160, January 1998.
- [LW88] H.B. Lee and M. S. Wengrovitz. Improved high-resolution direction-finding through use of homogeneous constraints. *Proc. IEEE ASSP Workshop on Spectrum Estimation and Modeling*, Minneapolis, Minnesota, vol.1 pp. 152–157, August 1988.

- [LW90] H.B. Lee and M. S. Wengrovitz. Resolution threshold of beamspace MUSIC for two closely spaced emitters. *IEEE Trans. Acoust. Speech, Signal Process.*, vol.ASSP-38, pp. 1545–1559, September 1990.
- [Mar87] S.L. Marple. *Digital Spectral Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey, 1987.
- [Mcl91] J.H. McClellan and D. Lee. Exact equivalence of the Steiglitz-McBride iteration and IQML. *IEEE Trans. Signal Process.*, vol.SP-39, pp. 509–512, February 1991.
- [MF90] M. I. Miller and D. R. Fuhrmann. Maximum-likelihood narrow-band direction finding and the EM algorithm. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-38, pp. 1560–1577, September 1990.
- [MH71] R. J. McAulay and E. M. Hofstetter. Barankin bounds on parameter estimation. *IEEE Trans. Inf. Theory*, vol.IT-17, pp. 669–676, November 1971.
- [MKB79] K. V. Mardia, J. T. Kent, and J. M. Bibby. *Multivariate Analysis*. Academic Press, New York, 1979.
- [Moo96] T. K. Moon. The expectation maximization algorithm. *IEEE Signal Process. Magazine*, pp. 47–60, November 1996.
- [MS69] R. J. McAulay and L. P. Seidman. A useful form of the Barankin lower bound and its application to PPM threshold analysis. *IEEE Trans. Inf. Theory*, vol.IT-15, pp. 273–279, March 1969.
- [MSR83] M. R. Matausek, S. S. Stankovic, and D. V. Radovic. Iterative inverse filtering approach to the estimation of frequencies of noisy sinusoids. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-31, pp. 1456–1463, December 1983.
- [MWW93] Y. Meng, K. M. Wong, and Q. Wu. Estimation of the direction-of-arrival of spread sources in sensor array processing. *Proc. Int. Conf. on Signal Processing.*, Beijing, China, vol.1 pp. 430–434, October 1993.
- [NK94] V. Nagesha and S. Kay. On frequency estimation with the IQML algorithm. *IEEE Trans. Signal Processing*, vol.SP-42, pp. 2509–2513, September 1994.
- [NK96] V. Nagesha and S. Kay. Maximum likelihood estimation for array processing in colored noise. *IEEE Trans. Signal Processing*, vol.SP-44, pp. 169–180, February 1996.
- [NS96] S. G. Nash and A. Sofer. *Linear and Nonlinear Programming*. McGraw-Hill, New York, 1996.
- [OL89] B. Ottersten and L. Ljung. Asymptotic results for sensor array processing. *Proc. ICASSP*, Glasgow, Scotland, vol.4 pp. 2266–2269, May 1989.
- [ORK89] B. Ottersten, R. Roy, and T. Kailath. Signal waveform estimation in sensor array processing. *Proc. 23rd Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, California, pp. 787–791, 1989.
- [Ott89] B. Ottersten. *Parametric Subspace Fitting Methods for Array Signal Processing*. PhD Thesis, Stanford University, Stanford, California, 1989.
- [OVK92] B. Ottersten, M. Viberg, and T. Kailath. Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data. *IEEE Trans. Signal Process.*, vol.SP-40, pp. 590–600, March 1992.

- [OVSN93] B. Ottersten, M. Viberg, P. Stoica, and A. Nehorai. Exact and large sample maximum likelihood techniques for parameter estimation and detection in array processing. *Radar Array Processing*, Springer-Verlag, Berlin, pp. 99–151, 1993.
- [Pie63] J. N. Pierce. Approximate error probabilities for optimal diversity combining. *IEEE Trans. Commun. Syst.*, vol.CS-11, pp. 352–354, September 1963.
- [PF86] B. Porat and B. Friedlander. Computation of the exact information of Gaussian time series with stationary random components. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-34, pp. 118–130, January 1986.
- [PF88] B. Porat and B. Friedlander. Analysis of the asymptotic relative efficiency of the MUSIC algorithm. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-36, pp. 532–544, 1988.
- [PF97] B. Porat and B. Friedlander. Accuracy requirements in off-line array calibration. *IEEE Trans. Aerospace Electron. Syst.*, vol.AES-33, pp. 545–556, April 1997.
- [PK86] A. Paulraj and T. Kailath. Eigenstructure methods for direction of arrival estimation in the presence of unknown noise fields. *IEEE Trans. Acoust. Speech, Signal Process.*, vol.ASSP-34, pp. 13–20, February 1986.
- [PWMS88] S. Prasad, R. T. Williams, A. K. Mahalanabis, and L. H. Sibul. A transform-based covariance differencing approach for some classes of parameter estimation problems. *IEEE Trans. Acoust. Speech, and Sig. Proc.*, vol.ASSP-36, pp. 631–641, May 1988.
- [Rao46] C. R. Rao. *Linear Statistical Inference and Its Applications*. Wiley, New York, 1946.
- [RKR94] R. Rajagopal, K. A. Kumar, and P. R. Rao. High-resolution beamforming in the presence of coherent interferences and unknown correlated noise. *Sixth IEEE Digital Signal Process. Workshop*, pp. 245–248, 1994.
- [Ris78] J. Rissanen. Modeling by shortest data description. *Automatica*, vol.14, pp. 465–471, 1978.
- [RM91] M.J.D Rendas and J.M.F. Moura. Cramér-Rao bound for location systems in multipath environments. *IEEE Trans. Signal Process.*, vol.SP-39, pp. 2593–2610, December 1991.
- [RS87a] Y. Rockah and P. M. Schultheiss. Array shape calibration using sources in unknown locations-part II: Near-field sources and estimator implementation. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-35, pp. 724–735, June 1987.
- [RS87b] Y. Rockah and P. M. Schultheiss. Array shape calibration using sources in unknown locations-part I: Far-field sources. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-35, pp. 286–299, March 1987.
- [RW80] A. Ruhe and P. A. Wedin. Algorithms for separable nonlinear least squares problems. *SIAM Rev.*, vol.22, pp. 318–327, July 1980.
- [RW92] J. P. Reilly and K. M. Wong. Estimation of directions of arrival of signals in unknown correlated noise, Part II: Asymptotic behavior and performance of the MAP approach. *IEEE Trans. Signal Process.*, vol.SP-40, pp. 2018–2028, August 1992.

- [SB87] U. Sandkühler and J. F. Böhme. Accuracy of maximum likelihood estimates for array processing. *Proc. ICASSP*, Dallas, Texas, vol.4 pp. 2015–2018, April 1987.
- [Sch68] F. Schweppe. Sensor array data processing for multiple signal sources. *IEEE Trans. Inf. Theory*, vol. IT-4, pp. 294–305, March 1968.
- [Sch78] G. Schwartz. Estimating the dimension of a model. *Ann. Stat.*, vol.6, pp. 461–464, 1978.
- [Sch79] R. O. Schmidt. Multiple emitter location and signal parameter estimation. *Proc. RADC Spectrum Estimation Workshop*, Griffiths AFB, Rome, New York, pp. 243–258, 1979.
- [Sch81] R. O. Schmidt. *Multiple Emitter Location and Signal Parameter Estimation*. Ph.D. Dissertation, Stanford University, Stanford, California, 1981.
- [Sch91] L. L. Scharf. *Statistical Signal Processing; Detection, Estimation, and Time Series Analysis*. Wiley, New York, 1991.
- [Sch94] S. V. Schell. Performance analysis of the cyclic MUSIC method of direction estimation for cyclostationary signals. *IEEE Trans. Signal Process.*, vol. SP-42, pp. 3043–3050, November 1994.
- [SG92] S. V. Schell. and W.A. Gardner. Cramér-Rao bound for directions of arrival of Gaussian cyclostationary signals. *IEEE Trans. Signal Process.*, vol. SP-38, pp. 1418–1422, July 1992.
- [Sha87] A.K. Shaw. *Structured Matrix Approximation Problems in Signal Processing*. PhD Dissertation, University of Rhode Island, Kingston, 1987.
- [Sha95] A.K. Shaw. Maximum likelihood estimation of multiple frequencies with constraints to guarantee unit circle roots. *IEEE Trans. Signal Process.*, vol. SP-43, pp. 796–799, March 1995.
- [Sim80] D. N. Simkins. *Multichannel Angle-of-Arrival Estimation*. PhD thesis, Stanford University, Stanford, California, 1980.
- [SJ97] P. Stoica and M. Jansson. On forward-backward MODE for array signal processing. *Digital Signal Process.*, vol. 7, pp. 239–252, October 1997.
- [SL01] P. Stoica and E. G. Larson. Comments on “Linearization method for finding Cramér-Rao bounds in signal processing.” *IEEE Trans. Signal Process.*, vol. SP-49, pp. 3168–3169, December 2001.
- [Sle78] D. Slepian. Prolate spheroidal wave functions, Fourier analysis, and uncertainty, V: The discrete case. *Bell Syst. Tech. J.*, vol. 57, pp. 1371–1430, May-June 1978.
- [SLG01] P. Stoica, E. G. Larson, and A. B. Gershman. The stochastic CRB for array processing: A textbook derivation. *IEEE Signal Process. Lett.*, vol. 8, pp. 148–150, May 2001.
- [SLS97] P. Stoica, J. Li, and T. Soderstrom. On the inconsistency of IQML. *Signal Process.*, vol. 56, pp. 185–190, 1997.
- [SM65] K. Steiglitz and L. E. McBride. A technique for the identification of linear systems. *IEEE Trans. Autom. Control*, vol. AC-10, pp. 461–465, October 1965.
- [SM97] P. Stoica and R. Moses. *Introduction to Spectral Analysis*. Prentice-Hall, Englewood Cliffs, New Jersey, 1997.

- [SN89a] P. Stoica and A. Nehorai. MODE, maximum likelihood, and Cramér-Rao bound: Conditional and unconditional results. Report No. 8901, Center for Systems Science, Yale University, New Haven, Connecticut, 1989.
- [SN89b] P. Stoica and A. Nehorai. MUSIC, maximum likelihood, and Cramér-Rao bound. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-37, pp. 720–741, May 1989.
- [SN90a] P. Stoica and A. Nehorai. MUSIC, maximum likelihood, and Cramér-Rao bound: Further results and comparisons. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-38, pp. 2140–2150, December 1990.
- [SN90b] P. Stoica and A. Nehorai. Performance study of conditional and unconditional direction-of-arrival estimation. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-38, pp. 1783–1795, October 1990.
- [SN92] D. Starer and A. Nehorai. Newton algorithm for conditional and unconditional maximum likelihood estimation of the parameters of exponential signals in noise. *IEEE Trans. Signal Process.*, vol.SP-40, pp. 1528–1533, June 1992.
- [SOV92] P. Stoica, B. Ottersen, and M. Viberg. An instrumental variables approach to array processing in spatially correlated noise fields. *Proc. ICASSP*, San Francisco, California, vol.II pp. 421–424, 1992.
- [SOVM96] P. Stoica, B. Ottersen, M. Viberg and R. L. Moses. Maximum likelihood array processing for stochastic coherent sources. *IEEE Trans. Signal Process.*, vol.SP-44, pp. 96–105, January 1996.
- [SS90a] P. Stoica and K. Sharman. Maximum likelihood methods for direction-of-arrival estimation. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-38, pp. 1132–1143, July 1990.
- [SS90b] P. Stoica and K. Sharman. A novel eigenanalysis method for direction estimation. *Proc. IEE*, pp. 19–26, 1990.
- [SVO94] P. Stoica, M. Viberg, and B. Ottersten. Instrumental variable approach to array processing in spatially correlated noise fields. *IEEE Trans. Signal Process.*, vol.SP-42, pp. 121–133, 1994.
- [SVWW96] P. Stoica, M. Viberg, K. M. Wong, and Q. Wu. Maximum-likelihood bearing estimation with partly calibrated arrays in spatially correlated noise fields. *IEEE Trans. Signal Processing*, vol.SP-44, pp. 888–899, April 1996.
- [SVWW98] P. Stoica, M. Viberg, M. Wong and Q. Wu. A unified instrumental variable approach to direction finding in colored noise fields. *Digital Signal Processing Handbook*, V. Madisetti and D. Williams, Eds, pp. 64.1–64.18, CRC Press and IEEE Press, 1998.
- [Swe59] P. Swerling. A proposed stagewise differential correction procedure for satellite tracking and prediction. *J. Astronaut. Sci.*, vol.6, 1959.
- [Swi93] D. N. Swigler. Simple approximations to the Cramér-Rao lower bound on directions of arrival for closely spaced sources. *IEEE Trans. Signal Process.*, vol.SP-41, pp. 1668–1672, April 1993.
- [Tia01] Z. Tian. Beamspace weighted subspace fitting. Submitted to *Signal Process. Lett.*.

- [TK82] D. W. Tufts and R. Kumaresan. Singular value decomposition and improved frequency extimation using linear prediction. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-30, pp. 671–675, August 1982.
- [TVT00] Z. Tian and H. L. Van Trees. Beamspace IQML. *Proc. IEEE Sensor Array and Multichannel Signal Processing Workshop*, Lexington, Massachusetts, March 2000.
- [Vib89] M. Viberg. *Subspace Fitting Concepts in Sensor Array Processing*. PhD Thesis, Linköping University, Linköping, Sweden, 1989.
- [VO91] M. Viberg and B. Ottersten. Sensor array processing based on subspace fitting. *IEEE Trans. Signal Process.*, vol.SP-39, pp. 1110–1121, May 1991.
- [VOK91] M. Viberg, B. Ottersten, and T. Kailath. Detection and estimation in sensor arrays using weighted subspace fitting. *IEEE Trans. Signal Process.*, vol.SP-39, pp. 2436–2449, November 1991.
- [VS94] M. Viberg and A. Swindlehurst. Analysis of the combined effects of finite samples and model errors on array processing performance. *IEEE Trans. Signal Process.*, vol.SP-42, pp. 3073–3083, November 1994.
- [VT68] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I*. Wiley, New York, 1968.
- [VT71] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part III*. Wiley, New York, 1971.
- [VT01a] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part I*. Wiley Interscience, New York, 2001.
- [VT01b] H. L. Van Trees. *Detection, Estimation, and Modulation Theory, Part III*. Wiley Interscience, New York, 2001.
- [VWV88a] B. Van Veen and B. Williams. Structured covariance matrices and dimensionality reduction in array processing. *Proc. 4th Workshop Spectr. Est. and Modeling*, pp. 168–171, August 1988.
- [VWV88b] B. Van Veen and B. Williams. Dimensionality reduction in high resolution direction of arrival estimation. *Proc. Asilomar Conf. on Signals, Systems, and Computers*, Pacific Grove, California, vol.2, pp. 588–592, October 1988.
- [Wax85] M. Wax. *Detection and Estimation of Superimposed Signals*. PhD thesis, Stanford University, Stanford, California, 1985.
- [Wax91] M. Wax. Detection and localization of multiple sources via the stochastic signals model. *IEEE Trans. Signal Process.*, vol.SP-39, pp. 2450–2456, November 1991.
- [Wax92] M. Wax. Detection and localization of multiple sources in noise with unknown covariance. *IEEE Trans. Signal Process.*, vol.SP-40, pp. 245–249, January 1992.
- [Wei87] A. J. Weiss. Bounds on time-delay estimation for monochromatic signals. *IEEE Trans. Aerospace Electron. Syst.*, vol.AES-23, pp. 798–808, November 1987.
- [WF89] A. J. Weiss and B. Friedlander. Array shape calibration using sources in unknown locations: A maximum likelihood approach. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-37, pp. 1958–1966, December 1989.

- [WF90] A. J. Weiss and B. Friedlander. On the Cramér-Rao bound for direction finding of correlated signals. Technical report, Signal Processing Technology, July 1990.
- [WF91a] A. J. Weiss and B. Friedlander. On the Cramér-Rao bound for direction finding of correlated signals. Technical report, Signal Processing Technology, October 1991.
- [WF91b] A. J. Weiss and B. Friedlander. Array shape calibration via eigenstructure methods. *J. Signal Process.*, vol.122, pp. 251–258, 1991.
- [WF93] A. J. Weiss and B. Friedlander. On the Cramér-Rao bound for direction finding of correlated signals. *IEEE Trans. Signal Process.*, vol.SP-41, pp. 495–499, January 1993.
- [WF94] A. J. Weiss and B. Friedlander. Preprocessing for direction finding with minimal variance degradation. *IEEE Trans. Signal Process.*, vol.SP-42, pp. 1478–1485, June 1994.
- [WK85] M. Wax and T. Kailath. Detection of signals by information theoretic criteria. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-33, pp. 387–392, April 1985.
- [WJ65] J. M. Wozencraft and I. M. Jacobs. *Principles of Communication Engineering*. Wiley, New York, 1965.
- [WOV91] B. Wahlberg, B. Ottersten, and M. Viberg. Robust signal parameter estimation in the presence of array perturbations. *Proc. ICASSP*, Toronto, Canada, vol.5 pp. 3277–3280, May 1991.
- [WRWQ92] K. M. Wong, J. P. Reilly, Q. Wu, and S. Qiao. Estimation of directions of arrival of signals in unknown correlated noise: Parts I and II. *IEEE Trans. Signal Process.*, vol.SP-40, pp. 2007–2028, August 1992.
- [WRWQ92] K. M. Wong, J. P. Reilly, Q. Wu, and S. Qiao. Estimation of directions of arrival of signals in unknown correlated noise, Part I: The MAP approach and its implementation. *IEEE Trans. Signal Process.*, vol.SP-40, pp. 2007–2017, August 1992.
- [WSK82] M. Wax, T. J. Shan, and T. Kailath. Location and spectral density estimation of multiple sources. *Proc. 16th Asilomar Conf. on Signals, Systems, and Computers*, pp. 322–326, 1982.
- [Wu83] N.-L. Wu. An explicit solution and data extension in the maximum entropy method. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-31, pp. 486–491, April 1983.
- [WW82] A. J. Weiss and E. Weinstein. Composite bound on the attainable mean-square error in passive time delay estimation from ambiguity prone signals. *IEEE Trans. Inf. Theory*, vol.IT-28, pp. 977–979, November 1982.
- [WW83] A. J. Weiss and E. Weinstein. Fundamental limitations in passive time delay estimation, Part I: Narrow-band systems. *IEEE Trans. Acoust., Speech Signal Process.*, vol.ASSP-31, pp. 472–486, April 1983.
- [WW84] E. Weinstein and A. J. Weiss. Fundamental limitations in passive time delay estimation, Part II: Wide-band systems. *IEEE Trans. Acoust., Speech Signal Process.*, vol.ASSP-32, pp. 1064–1078, October 1984.

- [WW88] E. Weinstein and A. J. Weiss. A general class of lower bounds in parameter estimation. *IEEE Trans. Inf. Theory*, vol. IT-34, pp. 338–342, March 1988.
- [WW94] Q. Wu and K. M. Wong. UN-MUSIC and UN-CLE: An application of generalized correlation to the estimation of directivities of signals in unknown correlated noise. *IEEE Trans. Signal Process.*, vol. SP-42, pp. 2331–2343, September 1994.
- [WW95] Q. Wu and K. M. Wong. Estimation of DOA in unknown noise: Performance analysis of UN-MUSIC and UN-CLE, and the optimality of CCD. *IEEE Trans. Signal Process.*, vol. SP-43, pp. 454–467, February 1995.
- [WWL88] A. J. Weiss, A. S. Willsky, and B. C. Levy. Eigenstructure approach for array processing with unknown intensity coefficients. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-36, pp. 1613–1617, October 1988.
- [WZ89] M. Wax and I. Ziskind. Detection of the number of coherent signals by the MDL principle. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-37, pp. 1190–1196, August 1989.
- [WZRY90] K. M. Wong, Q. T. Zhang, J. P. Reilly, and P. C. Yip. On information theoretic criteria for determining the number of signals in high resolution array processing. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-38, pp. 1959–1971, November 1990.
- [XB88] X. L. Xu and K. Buckley. Reduced-dimension beamspace broadband source localization: Preprocessor design and evaluation. *Proc. 4th ASSP Workshop Spectrum Estimation and Modeling*, Minneapolis, Minnesota, pp. 22–27, August 1988.
- [XB89] X. L. Xu and K. M. Buckley. Statistical performance comparison of MUSIC in element-space and beam-space. *Proc. ICASSP*, Glasgow, Scotland, vol. 4 pp. 2124–2127, May 1989.
- [XK92] G. Xu and T. Kailath. Direction of arrival estimation via exploitation of cyclostationarity—A combination of temporal and spatial processing. *IEEE Trans. Signal Process.*, vol. SP-40, pp. 1775–1786, July 1992.
- [XK95] W. Xu and M. Kaveh. Analysis of the performance and sensitivity of eigendecomposition-based detectors. *IEEE Trans. Signal Process.*, vol. SP-43, pp. 1413–1426, June 1995.
- [XRK94] G. Xu, R. H. Roy, and T. Kailath. Detection of number of sources via exploitation of centro-symmetry property. *IEEE Trans. Signal Process.*, vol. SP-42, pp. 102–112, January 1994.
- [YB92] S. F. Yau and Y. Bresler. Worst case Cramér-Rao bounds for parametric estimation of superimposed signal with applications. *IEEE Trans. Signal Process.*, vol. SP-40, pp. 2973–2986, December 1992.
- [YDG95] H. Ye and R. D. DeGroat. Maximum likelihood DOA estimation and asymptotic Cramér-Rao bounds for additive unknown colored noise. *IEEE Trans. Signal Process.*, vol. SP-43, pp. 938–949, April 1995.
- [YK87] Y. Q. Yin and P. R. Krishnaiah. On some nonparametric methods for detection of the number of signals. *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, pp. 1533–1538, November 1987.

- [YS95] J. Yang and A. L. Swindlehurst. The effects of array calibration errors on DF-based signal copy performance. *IEEE Trans. Signal Processing*, vol.SP-43, pp. 2724–2732, November 1995.
- [Zha94] Q. T. Zhang. Asymptotic performance analysis of information-theoretic criteria for determining the number of signals in spatially correlated noise. *IEEE Trans. Signal Process.*, vol.SP-42, pp. 1537–1539, June 1994.
- [ZKB86] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai. On detection of the number of signals in the presence of white noise. *J. Multi-Variate Anal.*, vol.20, pp. 1–25, 1986.
- [ZKB87] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai. Remarks on certain criteria for detection of number of signals. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-35, pp. 129–132, February 1987.
- [ZKS93] M. D. Zoltowski, G.M. Kautz, and S.D. Silverstein. Beamspace root-MUSIC. *IEEE Trans. Signal Process.*, vol.SP-41, pp. 344–364, February 1993.
- [ZL91] M. D. Zoltowski, T.S. Lee. Maximum likelihood based sensor array signal processing in the beamspace domain for low angle radar tracking. *IEEE Trans Signal Process.*, vol.SP-39, pp. 656–671, March 1991.
- [Zol88] M. D. Zoltowski. High resolution sensor array signal processing in the beamspace domain: Novel techniques based on the poor resolution of Fourier beam-forming. *Proc. ASSP*, pp. 350–355, 1988.
- [ZW88a] J. X. Zhu and H. Wang. Effects of sensor position and pattern perturbations on CRLB for direction finding of multiple narrow-band sources. *Proc. 4th ASSP Workshop on Spectrum Estimation and Modeling*, Minneapolis, Minnesota, pp. 98–102, August 1988.
- [ZW88b] I. Ziskind and M. Wax. Maximum likelihood localization of multiple sources by alternating projection. *IEEE Trans. Acoust., Speech, Signal Process.*, vol.ASSP-36, pp. 1553–1560, October 1988.
- [ZZ69] J. Ziv and M. Zakai. Some lower bounds on signal parameter estimation. *IEEE Trans. Inf. Theory*, vol.IT-15, pp. 386–391, May 1969.

Copyright of Optimum Array Processing is the property of John Wiley & Sons, Inc. 2002 and its content may not be copied or emailed to multiple sites or posted to a listserv without the copyright holder's express written permission. However, users may print, download, or email articles for individual use.