Direction of Arrival Estimation

1 Introduction

We have seen that there is a one-to-one relationship between the direction of a signal and the associated received steering vector. It should therefore be possible to invert the relationship and estimate the direction of a signal from the received signals. An antenna array therefore should be able to provide for *direction of arrival estimation*. We have also seen that there is a Fourier relationship between the beam pattern and the excitation at the array. This allows the direction of arrival (DOA) estimation problem to be treated as equivalent to *spectral estimation*.
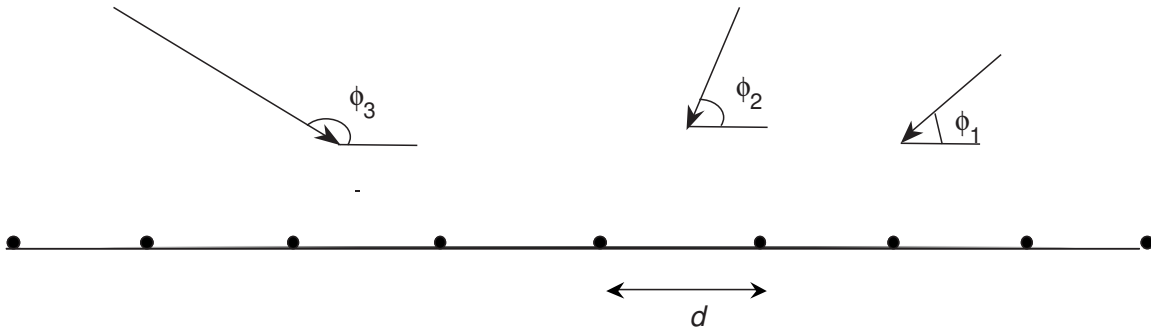

Figure 1: The DOA estimation problem.

The problem set up is shown in Fig. 1. Several (M) signals impinge on a linear, equispaced, array with N elements, each with direction ϕ_i. The goal of DOA estimation is to use the data received at the array to estimate ϕ_i, $i = 1, \ldots M$. It is generally assumed that $M < N$, though there exist approaches (such as maximum likelihood estimation) that do not place this constraint.

In practice, the estimation is made difficult by the fact that there are usually an unknown number of signals impinging on the array simultaneously, each from unknown directions and with unknown amplitudes. Also, the received signals are always corrupted by noise. Nevertheless, there are several methods to estimate the number of signals and their directions. Figure 2 shows some of these several spectral estimation [1] techniques[1]. Note that this is *not* an exhaustive list.

This chapter is organized as follows. We begin by determining the Cramer-Rao bound, the theoretical limit on how well the directions of arrival can be estimated. We then look at methods to estimates the directions assuming we know the number of incoming signals. We will only describe 5 techniques: correlation, Maximum Likelihood, MUSIC, ESPRIT and Matrix Pencil. Finally we

---

[1]I would like to acknowledge the contributions of Prof. Alex Gershman, Dept. of Elec. and Comp. Engg., McMaster University, for this figure [1]
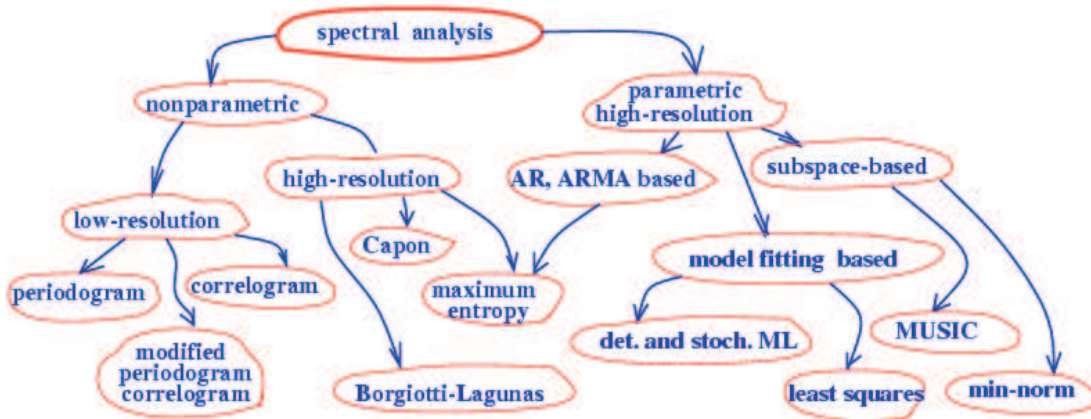
Figure 2: *Some* of the several approaches to spectral estimation

look at two methods to estimate the number of signals.

2 The Cramer-Rao Bound

We begin by realizing that the DOA is a parameter estimated from the received data. The minimum variance in this estimate is given by the Cramer-Rao bound (CRB).

The CRB theorem: Given a length-N vector of received signals $\mathbf{x}$ dependent on a set of P parameters $\boldsymbol{\theta} = [\theta_1, \theta_2, \ldots, \theta_P]^T$, corrupted by additive noise,

$$\mathbf{x} = \mathbf{v}(\boldsymbol{\theta}) + \mathbf{n}, \tag{1}$$

where $\mathbf{v}(\boldsymbol{\theta})$ is a *known* function of the parameters, the variance of an unbiased estimate of the p-th parameter, θ_p, is greater than the Cramer Rao bound

$$\mathrm{var}(\theta_p) \geq \mathbf{J}_{pp}^{-1}, \tag{2}$$

where $\mathbf{J}_{pp}^{-1}$ is the p-th diagonal entry of the *inverse* of the Fisher information matrix $\mathbf{J}$ whose $(i,j)^{\text{th}}$ is given by

$$\mathbf{J}_{ij} = -\mathrm{E}\left\{ \frac{\partial^2}{\partial\theta_i \partial\theta_j} \left[\ln f_{\mathbf{X}}(\mathbf{x}/\boldsymbol{\theta}) \right] \right\}, \tag{3}$$

where, $f_{\mathbf{X}}(\mathbf{x}/\boldsymbol{\theta})$ is the pdf of the received vector given the parameters $\boldsymbol{\theta}$ and $\mathrm{E}\{\cdot\}$ represents statistical expectation.

The CRB tells us that estimating parameters from noisy data will necessarily result in noisy estimates. Furthermore, the CRB is the best we can possibly do in minimizing the residual noise in unbiased estimates. Also, due to the fact that the minimum variance is dependent on the inverse

of the Fisher information matrix, we cannot ignore parameters that we are not interested in. The vector $\boldsymbol{\theta}$ must include *all* parameters in the model for $\mathbf{v}$.

2.1 CRB for DOA Estimation

As shown in Fig. 1, the model under consideration is a number of signals impinging at the array, corrupted by white noise. We will derive the CRB for a single signal corrupted by noise ($M = 1$). The data model is therefore

$$\mathbf{x} = \alpha \mathbf{s}(\phi) + \mathbf{n}, \tag{4}$$

where $\mathbf{s}(\phi)$ represents the steering vector of the signal whose direction (ϕ) we are attempting to estimate. The noise vector $\mathbf{n}$ is zero-mean Gaussian with covariance $\sigma^2 \mathbf{I}$. Note that, even though we are not interested in the amplitude, there are *two* unknown parameters, α and ϕ. Here α and ϕ are modelled as an unknown, but deterministic, constants, i.e., $\mathrm{E}\{\mathbf{x}\} = \alpha \mathbf{s}(\phi)$. In CRB literature, α would be considered a nuisance parameter, which must be accounted for because it is unknown. Finally, α itself represents two unknowns, its real and imaginary parts, or equivalently its magnitude and phase. Let $\alpha = ae^{jb}$. Therefore, $\boldsymbol{\theta} = [\,a,\,b,\,\phi\,]^T$. In our case,

$$\mathbf{v}(\boldsymbol{\theta}) \;=\; \alpha\,\mathbf{s}(\phi), \tag{5}$$

$$f_{\mathbf{X}}(\mathbf{x}/\boldsymbol{\theta}) \;=\; Ce^{-(\mathbf{x}-\mathbf{v})^H \mathbf{R}^{-1}(\mathbf{x}-\mathbf{v})}, \tag{6}$$

where $\mathbf{R} = \sigma^2 \mathbf{I}$ and C is a normalization constant.

$$\Rightarrow \ln f_{\mathbf{X}}(\mathbf{x}/\boldsymbol{\theta}) \;=\; \ln C - \frac{(\mathbf{x} - \mathbf{v})^H (\mathbf{x} - \mathbf{v})}{\sigma^2} \tag{7}$$

$$=\; \ln C + \frac{-\mathbf{x}^H \mathbf{x} + \mathbf{v}^H \mathbf{x} + \mathbf{x}^H \mathbf{v} - \mathbf{v}^H \mathbf{v}}{\sigma^2}, \tag{8}$$

$$=\; \ln C + \frac{-\mathbf{x}^H \mathbf{x} + \alpha^* \mathbf{s}^H(\phi)\mathbf{x} + \alpha \mathbf{x}^H \mathbf{s}(\phi) - |\alpha|^2 \mathbf{s}^H(\phi)\,\mathbf{s}(\phi)}{\sigma^2}. \tag{9}$$

Note that the first two terms in this final equation are constant with respect to the parameters θ_i. Since we are interested in taking derivatives of this expression, we can ignore these terms. Focusing on the important terms and writing the result in terms of the parameters $\boldsymbol{\theta} = [\,a,\,b,\,\phi\,]^T$,

$$g(\boldsymbol{\theta}) = \frac{1}{\sigma^2} \left[ae^{-jb}\mathbf{s}^H(\phi)\mathbf{x} + ae^{jb}\mathbf{x}^H \mathbf{s}(\phi) - a^2 \mathbf{s}^H(\phi)\mathbf{s}(\phi) \right]. \tag{10}$$

Also,

$$\mathbf{J} = \mathrm{E} \left\{ \begin{bmatrix} \dfrac{\partial^2 g}{\partial a^2} & \dfrac{\partial^2 g}{\partial a \partial b} & \dfrac{\partial^2 g}{\partial a \partial \phi} \\[2mm] \dfrac{\partial^2 g}{\partial b \partial a} & \dfrac{\partial^2 g}{\partial b^2} & \dfrac{\partial^2 g}{\partial b \partial \phi} \\[2mm] \dfrac{\partial^2 g}{\partial \phi \partial a} & \dfrac{\partial^2 g}{\partial \phi \partial b} & \dfrac{\partial^2 g}{\partial \phi^2} \end{bmatrix} \right\}. \tag{11}$$

As shown in the chapter on array theory, we can choose any convenient form of the steering vector, including the form where the phase reference is at the center of the array, i.e., for a array with an odd number of elements

$$\mathbf{s}(\phi) = \left[z^{-(N-1)/2}, \ z^{-(N-3)/2} \ \ldots, \ z^{-1}, \ 1, \ z, \ \ldots, \ z^{(N-3)/2}, \ z^{(N-1)/2} \right]^T, \tag{12}$$

where

$$z = e^{jkd\cos\phi}. \tag{13}$$

Consider the partial derivative of $\mathbf{s}(\phi)$ as a function of ϕ. Denoting this vector to be $\mathbf{s}_1(\phi)$,

$$\begin{aligned}
\mathbf{s}_1(\phi) &= \frac{\partial \mathbf{s}(\phi)}{\partial \phi}, \\
&= -jkd\sin\phi \left[\frac{-(N-1)}{2} z^{-(N-1)/2}, \ \frac{-(N-3)}{2} z^{-(N-3)/2}, \ \ldots \right. \\
&\qquad \left. -z^{-1}, \ 0, \ z \ \ldots \ \frac{(N-3)}{2} z^{(N-3)/2}, \ \frac{(N-1)}{2} z^{(N-1)/2} \right]^T,
\end{aligned}$$

$$\Rightarrow \mathbf{s}_1(\phi)_n = -jkdn\sin\phi \, z^n, \tag{14}$$

where $\mathbf{s}_1(\phi)_n$ is the n-th element in the vector $\mathbf{s}_1$. Similarly, denote as $\mathbf{s}_2(\phi)$ the second derivative of $\mathbf{s}(\phi)$ with respect to ϕ.

Using the definitions of $\mathbf{s}(\phi)$, $\mathbf{s}_1(\phi)$, and $z^* = z^{-1}$ we can derive some terms that will be useful later:

$$\mathrm{E}\left[\mathbf{v}\right] = \alpha\mathbf{s} = ae^{jb}\mathbf{s}, \tag{15}$$

$$\mathbf{s}^H(\phi)\mathbf{s}(\phi) = N, \tag{16}$$

$$\mathbf{s}_1^H(\phi)\mathbf{s}(\phi) = jkd\sin\phi \sum_{n=-(N-1)/2}^{(N-1)/2} n = 0, \tag{17}$$

$$\mathbf{s}_1^H(\phi)\mathbf{s}_1(\phi) = (kd)^2\sin^2\phi \sum_{n=-(N-1)/2}^{(N-1)/2} n^2 \triangleq B^2(kd)^2\sin^2\phi, \tag{18}$$

where B^2 represents the sum in Eqn. (18).

Having dealt with preliminaries, we obtain the CRB in the case of DOA estimation. Using Eqn. (5), we derive some of the entries in the Fisher information matrix given in Eqn. (11).

$$\frac{\partial g}{\partial a} = \frac{1}{\sigma^2}\left[e^{-jb}\mathbf{s}^H(\phi)\mathbf{x} + e^{jb}\mathbf{x}^H\mathbf{s}(\phi) - 2a\mathbf{s}^H(\phi)\mathbf{s}(\phi) \right], \tag{19}$$

$$\Rightarrow \frac{\partial^2 g}{\partial a^2} = \frac{1}{\sigma^2}\left[-2\mathbf{s}^H(\phi)\mathbf{s}(\phi) \right] \Rightarrow \mathrm{E}\left[\frac{\partial^2 g}{\partial a^2}\right] = -2\frac{\mathbf{s}^H(\phi)\mathbf{s}(\phi)}{\sigma^2}, \tag{20}$$

4

$$\frac{\partial^2 g}{\partial a \partial b} = \frac{1}{\sigma^2} \left[-je^{-jb}\mathbf{s}^H(\phi)\mathbf{x} + je^{jb}\mathbf{x}^H\mathbf{s}(\phi) \right] \tag{21}$$

$$\Rightarrow \mathrm{E}\left[\frac{\partial^2 g}{\partial a \partial b}\right] = \frac{1}{\sigma^2} \left[-je^{-jb}ae^{jb}\mathbf{s}^H(\phi)\mathbf{s} + je^{jb}ae^{-jb}\mathbf{s}^H(\phi)\mathbf{s}(\phi) \right] = 0 \tag{22}$$

$$\frac{\partial^2 g}{\partial a \partial \phi} = \frac{1}{\sigma^2} \left[e^{-jb}\mathbf{s}_1^H(\phi)\mathbf{x} + e^{jb}\mathbf{x}^H\mathbf{s}_1(\phi) - 2a\mathbf{s}_1^H(\phi)\mathbf{s}(\phi) - 2a\mathbf{s}^H(\phi)\,\mathbf{s}_1(\phi) \right] \tag{23}$$

$$\Rightarrow \mathrm{E}\left[\frac{\partial^2 g}{\partial a \partial \phi}\right] = \frac{1}{\sigma^2} \left[e^{-jb}ae^{jb}\mathbf{s}_1^H(\phi)\mathbf{s} + e^{jb}ae^{-jb}\mathbf{s}^H\mathbf{s}_1(\phi) - 2a\mathbf{s}_1^H(\phi)\mathbf{s} - 2a\mathbf{s}^H(\phi)\mathbf{s}_1 \right]$$

$$= 0, \tag{24}$$

where, in deriving Eqn. (24) we use the orthogonality between $\mathbf{s}_1$ and $\mathbf{s}$ derived in Eqn. (17). Similar to the derivation in Eqn. (22), we can show that $\partial^2 g / \partial b \partial \phi = 0$. Therefore, *all the non-diagonal terms in the Fisher information matrix $\mathbf{J}$ are zero*. The only term we have to worry about is $\partial^2 g / \partial \phi^2$.

$$\frac{\partial g}{\partial \phi} = \frac{1}{\sigma^2} \left[ae^{-jb}\mathbf{s}_1^H(\phi)\mathbf{x} + ae^{jb}\mathbf{x}^H\mathbf{s}_1(\phi) - a^2\mathbf{s}_1^H(\phi)\mathbf{s}(\phi) - a^2\mathbf{s}^H(\phi)\mathbf{s}_1(\phi) \right], \tag{25}$$

$$\Rightarrow \frac{\partial^2 g}{\partial \phi^2} = \frac{1}{\sigma^2} \Big[ae^{-jb}\mathbf{s}_2^H(\phi)\mathbf{x} + ae^{jb}\mathbf{x}^H\mathbf{s}_2(\phi) - a^2\mathbf{s}_2^H(\phi)\mathbf{s}(\phi) - a^2\mathbf{s}_1^H(\phi)\mathbf{s}_1(\phi) -$$

$$a^2\mathbf{s}_1^H(\phi)\mathbf{s}_1(\phi) - a^2\mathbf{s}^H(\phi)\mathbf{s}_2(\phi) \Big] \tag{26}$$

$$\Rightarrow \mathrm{E}\left[\frac{\partial^2 g}{\partial \phi^2}\right] = \frac{1}{\sigma^2} \Big[a^2 e^{-jb}e^{jb}\mathbf{s}_2^H(\phi)\mathbf{s} + a^2 e^{jb}e^{-jb}\mathbf{s}^H\mathbf{s}_2(\phi) - a^2\mathbf{s}_2^H(\phi)\mathbf{s}(\phi)$$

$$-a^2\mathbf{s}_1^H(\phi)\mathbf{s}_1(\phi) - a^2\mathbf{s}_1^H(\phi)\mathbf{s}_1(\phi) - a^2\mathbf{s}^H(\phi)\mathbf{s}_2(\phi) \Big] \tag{27}$$

$$\Rightarrow \mathrm{E}\left[\frac{\partial^2 g}{\partial \phi^2}\right] = -2a^2\frac{1}{\sigma^2}\mathbf{s}_1^H\mathbf{s}_1, \tag{28}$$

$$= -\frac{2a^2(kd\sin\phi)^2 B^2}{\sigma^2}, \tag{29}$$

where Eqn. (29) is obtained using Eqn. (18). Using the definition of B,

$$B^2 = \sum_{n=-\frac{N-1}{2}}^{\frac{N-1}{2}} n^2 = 2\sum_{n=1}^{\frac{N-1}{2}} n^2$$

$$= 2\left(\frac{N-1}{2}\right)\left(\frac{N+1}{2}\right)\frac{N}{6} = \frac{N(N^2-1)}{12} \tag{30}$$

Using $|\alpha| = a$, the CRB for the DOA estimation problem is therefore,

$$\mathrm{var}(\phi) \geq \left[\mathrm{E}\left(\frac{\partial^2 g}{\partial \phi^2}\right) \right]^{-1}$$

$$\geq \frac{6\sigma^2}{|\alpha|^2 N(N^2-1)(kd)^2\sin^2\phi}. \tag{31}$$

5

We now have the CRB of the estimate of the DOA. Note that there are clear physical interpretations to be made. The CRB sets the best possible estimation. As expected, as the SNR $(|\alpha|^2/\sigma^2)$ increases, the CRB is reduced. Further, the denominator is approximately proportional to $[(N-1)kd]^2$, which is proportional to the electrical length of the array, the length *in terms of wavelength*, i.e. as the array size increases, we can form a better estimate. The N term suggests that for a given overall electrical length the more samples (elements) we have the better the estimate we can obtain. Finally, the $\sin\phi$ term represents the fact that as we scan off broadside the beamwidth increases in ϕ terms, i.e., this represents the beam broadening factor making DOA estimates that much worse.

3 DOA Estimation using Correlation

Having determined how well we can do, we now turn to actual algorithms to determine the directions of arrival. The model is of M signals incident on the array, corrupted by noise, i.e.,

$$\mathbf{x} = \sum_{m=1}^{M} \alpha_m \mathbf{s}(\phi_m) + \mathbf{n}. \tag{32}$$

The goal therefore is to estimate ϕ_m, $m = 1, \ldots M$. The easiest way to estimate the angles is through correlation. We know that by the Cauchy-Schwarz inequality, as a function of ϕ, $\mathbf{s}^H(\phi)\mathbf{s}(\phi_m)$ has a maximum at $\phi = \phi_m$. Therefore, the correlation method plots $P_{\text{corr}}(\phi)$ versus ϕ where

$$P_{\text{corr}}(\phi) = \mathbf{s}^H(\phi)\mathbf{x}. \tag{33}$$

$P_{\text{corr}}(\phi)$ is a *non-adaptive* estimate of the spectrum of the incoming data. The M largest peaks of this plot are the estimated directions of arrival.

In the case of our linear, equispaced array, the steering vector $\mathbf{s}(\phi)$ is equivalent to Fourier coefficients, i.e., the correlation in Eqn. (33) is equivalent to a DFT of the data vector $\mathbf{x}$. We will see that this technique is optimal (in the maximum likelihood sense) in the single user situation.

4 Maximum Likelihood Estimator: Correlation of a different kind

One way of estimating the DOA of an incoming signal is to maximize the likelihood that that the signal came from that particular angle. The data model we use is the same as in Eqn. (4), i.e., we are focusing on estimating the DOA of a single user. However, here we generalize the vector $\mathbf{n}$ to be an *interference* vector, including the signals from other users. This interference vector is colored and, in general, $\text{E}\left[\mathbf{n}\mathbf{n}^H\right] = \mathbf{R}_n$. We are still attempting to estimate ϕ, though since we have two

unknown parameters, the magnitude and DOA, the maximum likelihood estimator (MLE) is given by

$$\hat{\phi}, \hat{\alpha} = \max_{\alpha, \phi} \left[f_{\mathbf{X}/\alpha, \phi}(\mathbf{x}) \right], \tag{34}$$

where $f_{\mathbf{X}/\alpha, \phi}(\mathbf{x})$ is the pdf of the data vector $\mathbf{x}$ *given* the parameters α, ϕ. Assuming that the interference vector is complex Gaussian,

$$f_{\mathbf{X}/\alpha, \phi}(\mathbf{x}) = \frac{1}{\pi^N \det(\mathbf{R}_n)} e^{-(\mathbf{x} - \alpha \mathbf{s})^H \mathbf{R}_n^{-1}(\mathbf{x} - \alpha \mathbf{s})}, \tag{35}$$

i.e., the maximization in Eqn. (34) is equivalent to

$$\hat{\phi}, \hat{\alpha} = \min_{\alpha, \phi} \left[(\mathbf{x} - \alpha \mathbf{s})^H \mathbf{R}_n^{-1} (\mathbf{x} - \alpha \mathbf{s}) \right],$$

$$= \min_{\alpha, \phi} \left[\mathbf{x}^H \mathbf{R}_n^{-1} \mathbf{x} - \alpha \mathbf{x}^H \mathbf{R}_n^{-1} \mathbf{s} - \alpha^* \mathbf{s}^H \mathbf{R}_n^{-1} \mathbf{x} + \alpha^* \alpha \mathbf{s}^H \mathbf{R}_n^{-1} \mathbf{s} \right]. \tag{36}$$

We must minimize this function over both α and ϕ. Starting first with α and remembering that we can differentiate with respect to α^* while treating α as an independent variable,

$$\frac{\partial}{\partial \alpha^*} = \mathbf{s}^H \mathbf{R}_n^{-1} (\mathbf{x} - \alpha \mathbf{s}),$$

$$\Rightarrow \hat{\alpha} = \frac{\mathbf{s}^H \mathbf{R}_n^{-1} \mathbf{x}}{\mathbf{s}^H \mathbf{R}_n^{-1} \mathbf{s}}. \tag{37}$$

Using this value of α, we get

$$\hat{\phi} = \max_{\phi} \left[P_{\text{MLE}}(\phi) \right] = \max_{\phi} \left[\frac{\left| \mathbf{s}^H \mathbf{R}_n^{-1} \mathbf{x} \right|^2}{\mathbf{s}^H \mathbf{R}_n^{-1} \mathbf{s}} \right]. \tag{38}$$

The function $P_{\text{MLE}}(\phi)$ is the *maximum likelihood estimate* of the spectrum of the incoming data. The DOA estimate is the point where this function takes its maximum.

An interesting aspect of this estimator is that if there is only one user and $\mathbf{R}_n = \sigma^2 \mathbf{I}$, the correlation matrix is diagonal and the MLE *reduces to the correlation technique of Section 3*. This is expected because the correlation technique there is equivalent to the matched filter, which is optimal in the single user case (more generally, in any case where the received data is a single data vector plus *white* Gaussian noise).

Note that if we define a new vector $\tilde{\mathbf{n}} = \mathbf{R}_n^{-1/2} \mathbf{n}$, $\mathrm{E}\left[\tilde{\mathbf{n}} \tilde{\mathbf{n}}^H \right] = \mathbf{R}_n^{-1/2} \mathrm{E}\left[\mathbf{n} \mathbf{n}^H \right] \mathbf{R}_n^{-1/2} = \mathbf{I}$, i.e. the new interference vector is *white*. The operation in Eqn. (38) is equivalent to taking the inner product of two *whitened* vectors, $\tilde{\mathbf{s}} = \mathbf{R}_n^{-1/2} \mathbf{s}$ and $\tilde{\mathbf{x}} = \mathbf{R}_n^{-1/2} \mathbf{x}$. Therefore, the MLE is equivalent to the correlation technique of Section 3, only in whitened data space.

The ML approach is optimal in the maximum likelihood sense. However, it is an impractical algorithm. The algorithm assumes knowledge of $\mathbf{R}_n$, the interference covariance matrix, something that is not available in practice. While it may be possible to estimate the covariance of $\mathbf{x}$, estimating

the covariance of the interference (by itself) is almost impossible. Also, the algorithm is highly computationally intensive. For each signal, a new interference covariance matrix is required. Even is this matrix were known, for each a matrix inverse and finally a search are required to find where $P_{\mathrm{MLE}}(\phi)$ reaches its maximum.

5 MUSIC: MUltiple SIgnal Classification

Of all techniques shown in Fig. 2, MUSIC is probably the most popular technique. MUSIC, as are many adaptive techniques, is dependent on the correlation matrix of the data. Using the data model in Eqn. (32),

$$\mathbf{x} = \mathbf{S}\boldsymbol{\alpha} + \mathbf{n}. \tag{39}$$

$$\mathbf{S} = [\mathbf{s}(\phi_1)\,\mathbf{s}(\phi_2)\,\ldots,\,\mathbf{s}(\phi_M)], \tag{40}$$

$$\boldsymbol{\alpha} = [\alpha_1,\,\alpha_2\,\ldots\,\alpha_M]^T. \tag{41}$$

The matrix $\mathbf{S}$ is a $N \times M$ matrix of the M steering vectors. Assuming that the different signals to be uncorrelated, the correlation matrix of $\mathbf{x}$ can be written as

$$\mathbf{R} = \mathrm{E}\left[\mathbf{x}\mathbf{x}^H\right], \tag{42}$$

$$= \mathrm{E}\left[\mathbf{S}\boldsymbol{\alpha}\boldsymbol{\alpha}^H\mathbf{S}^H\right] + \mathrm{E}\left[\mathbf{n}\mathbf{n}^H\right],$$

$$= \mathbf{S}\mathbf{A}\mathbf{S}^H + \sigma^2\mathbf{I}, \tag{43}$$

$$= \mathbf{R}_s + \sigma^2\mathbf{I}, \tag{44}$$

where

$$\mathbf{R}_s = \mathbf{S}\mathbf{A}\mathbf{S}^H \tag{45}$$

$$\mathbf{A} = \begin{bmatrix} \mathrm{E}\left[|\alpha_1|^2\right] & 0 & \cdots & 0 \\ 0 & \mathrm{E}\left[|\alpha_2|^2\right] & \cdots & 0 \\ 0 & 0 & \cdots & \mathrm{E}\left[|\alpha_M|^2\right] \end{bmatrix}. \tag{46}$$

The signal covariance matrix, $\mathbf{R}_s$, is clearly a $N \times N$ matrix with rank M. It therefore has $N - M$ eigenvectors corresponding to the zero eigenvalue. Let $\mathbf{q}_m$ be such an eigenvector. Therefore,

$$\mathbf{R}_s\mathbf{q}_m = \mathbf{S}\mathbf{A}\mathbf{S}^H\mathbf{q}_m = 0, \tag{47}$$

$$\Rightarrow \mathbf{q}_m^H\mathbf{S}\mathbf{A}\mathbf{S}^H\mathbf{q}_m = 0, \tag{48}$$

$$\Rightarrow \mathbf{S}^H\mathbf{q}_m = 0 \tag{49}$$

where this final equation is valid since the matrix $\mathbf{A}$ is clearly positive definite. Equation (49) implies that all $N - M$ eigenvectors ($\mathbf{q}_m$) of $\mathbf{R}_s$ corresponding to the zero eigenvalue are *orthogonal*

to all M signal steering vectors. This is the basis for MUSIC. Call $\mathbf{Q}_n$ the $N \times (N - M)$ matrix of these eigenvectors. MUSIC plots the *pseudo-spectrum*

$$P_{\text{MUSIC}}(\phi) = \frac{1}{\sum_{m=1}^{N-M} |\mathbf{s}^H(\phi)\mathbf{q}_m|^2} = \frac{1}{\mathbf{s}^H(\phi)\mathbf{Q}_n\mathbf{Q}_n^H\mathbf{s}(\phi)} = \frac{1}{||\mathbf{Q}_n^H\mathbf{s}(\phi)||^2} \tag{50}$$

Note that since the eigenvectors making up $\mathbf{Q}_n$ are orthogonal to the signal steering vectors, the denominator becomes zero when ϕ is a signal direction. Therefore, the estimated signal directions are the M largest peaks in the pseudo-spectrum. However, in any practical situation, the signal covariance matrix $\mathbf{R}_s$ would not be available. The most we can expect is to be able to estimate $\mathbf{R}$ the signal covariance matrix. The key is that the eigenvectors in $\mathbf{Q}_n$ can be estimated from the eigenvectors of $\mathbf{R}$.

For any eigenvector $\mathbf{q}_m \in \mathbf{Q}$,

$$\mathbf{R}_s\mathbf{q}_m = \lambda\mathbf{q}_m$$

$$\begin{aligned} \Rightarrow \mathbf{R}\mathbf{q}_m &= \mathbf{R}_s\mathbf{q}_m + \sigma^2\mathbf{I}\mathbf{q}_m, \\ &= (\lambda_m + \sigma^2)\mathbf{q}_m, \end{aligned} \tag{51}$$

i.e. any eigenvector of $\mathbf{R}_s$ is also an eigenvector of $\mathbf{R}$ with corresponding eigenvalue $\lambda + \sigma^2$. Let $\mathbf{R}_s = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$. Therefore,

$$\mathbf{R} = \mathbf{Q}\left[\mathbf{\Lambda} + \sigma^2\mathbf{I}\right]\mathbf{Q}^H$$

$$= \mathbf{Q}\begin{bmatrix} \lambda_1 + \sigma^2 & 0 & \cdots & 0 & 0 & \cdots & 0 \\ 0 & \lambda_2 + \sigma^2 & \cdots & 0 & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & \lambda_M^2 + \sigma^2 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & \sigma^2 & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & 0 & \cdots & \sigma^2 \end{bmatrix}\mathbf{Q}^H. \tag{52}$$

Based on this eigendecomposition, we can partition the eigenvector matrix $\mathbf{Q}$ into a signal matrix $\mathbf{Q}_s$ with M columns, corresponding to the M signal eigenvalues, and a matrix $\mathbf{Q}_n$, with $(N - M)$ columns, corresponding the noise eigenvalues (σ^2). Note that $\mathbf{Q}_n$, the $N \times (N - M)$ matrix of eigenvectors corresponding to the noise eigenvalue (σ^2), *is exactly the same as the matrix of eigenvectors* of $\mathbf{R}_s$ corresponding to the zero-eigenvalue. This is the matrix used in Eqn. (50). $\mathbf{Q}_s$ defines the signal subspace, while $\mathbf{Q}_n$ defines the noise subspace.

There are few important observations to be made:

- The smallest eigenvalues of $\mathbf{R}$ are the noise eigenvalues and are all equal to σ^2, i.e., one way of distinguishing between the signal and noise eigenvalues (equivalently the signal and noise subspaces) is to determine the number of small eigenvalues that are equal.

- By orthogonality of $\mathbf{Q}$, $\mathbf{Q}_s \perp \mathbf{Q}_n$

Using the final two observations, we see that *all noise eigenvectors are orthogonal to the signal steering vectors. This is the basis for MUSIC.* Consider the following function of ϕ:

$$P_{\text{MUSIC}}(\phi) = \frac{1}{\sum_{m=M+1}^{N} |\mathbf{q}_m^H \mathbf{s}(\phi)|^2} = \frac{1}{\mathbf{s}^H(\phi)\mathbf{Q}_n\mathbf{Q}_n^H\mathbf{s}(\phi)}, \tag{53}$$

where $\mathbf{q}_m$ is one of the $(N - M)$ noise eigenvectors. If ϕ is equal to DOA one of the signals, $\mathbf{s}(\phi) \perp \mathbf{q}_m$ and the denominator is identically zero. MUSIC, therefore, identifies as the directions of arrival, the peaks of the function $P_{\text{MUSIC}}(\phi)$.

5.1 MUSIC in Practice - or close to it

In practice, the correlation matrix $\mathbf{R}$ is unknown and must be estimated from the received data. This estimation requires averaging over several snapshots of data.

$$\mathbf{R} = \frac{1}{K} \sum_{k=1}^{K} \mathbf{x}_k \mathbf{x}_k^H, \tag{54}$$

where $\mathbf{x}_k$ is the k-th snapshot. If the received data is Gaussian, this estimate asymptotically converges to the true correlation matrix. As we will see later in this course, this matrix is used extensively in adaptive beamforming. In [2], the authors prove that one requires at least $K > 2N$ so that the signal-to-noise ratio is within 3dB of the optimum. While this result cannot be directly applied to the case of DOA estimation, this figure has been taken as a good rule of thumb.

One of the problems arising from using an estimate of the correlation matrix is that the noise eigenvalues are no longer the same. Figure 3 plots the eigenvalues in the ideal situation of a known correlation matrix and the more realistic situation of an estimated correlation matrix. The example uses an 11-element array with three incoming signals. As is clear, in the ideal case, the eight noise eigenvalues are all equal. The three signal eigenvalues are dependent on the signal power.

In the realistic case, where the correlation matrix is estimated, the eigenvalues are more of a continuum. There is no clear distinction between the signal and noise eigenvalues. Note that some noise eigenvalues are greater than their true value of 0dB. Several authors have suggested choosing the "knee" of this plot to estimate the number of signals. This is especially error prone in low SNR situations where the corresponding signal eigenvalue $(\lambda_M + \sigma^2)$ is not significantly different from the noise eigenvalue (σ^2).
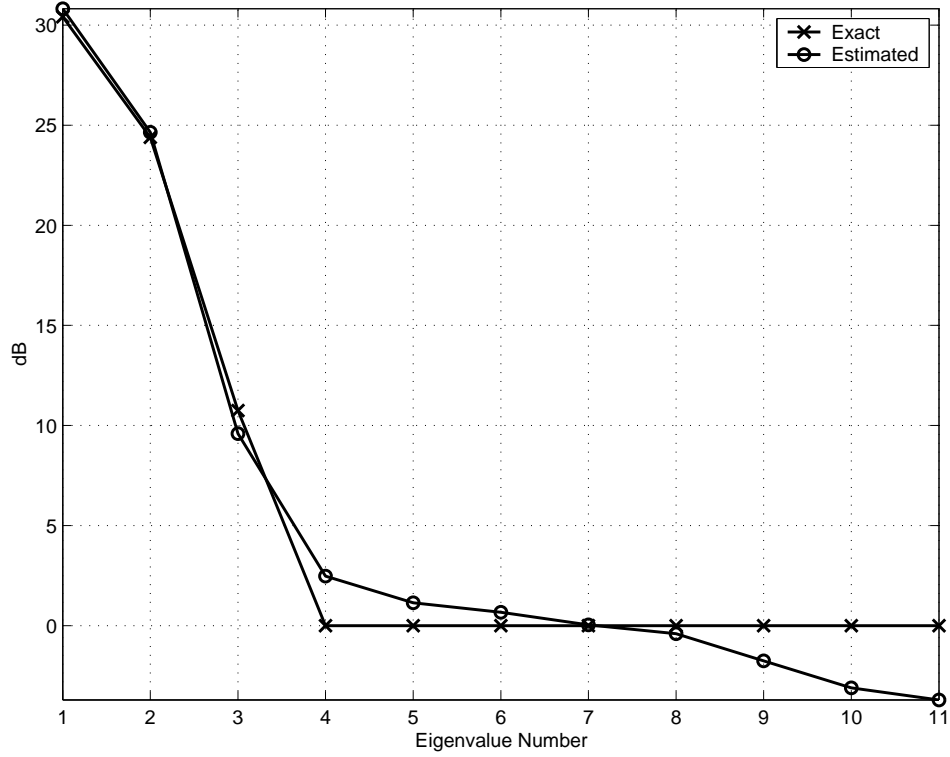
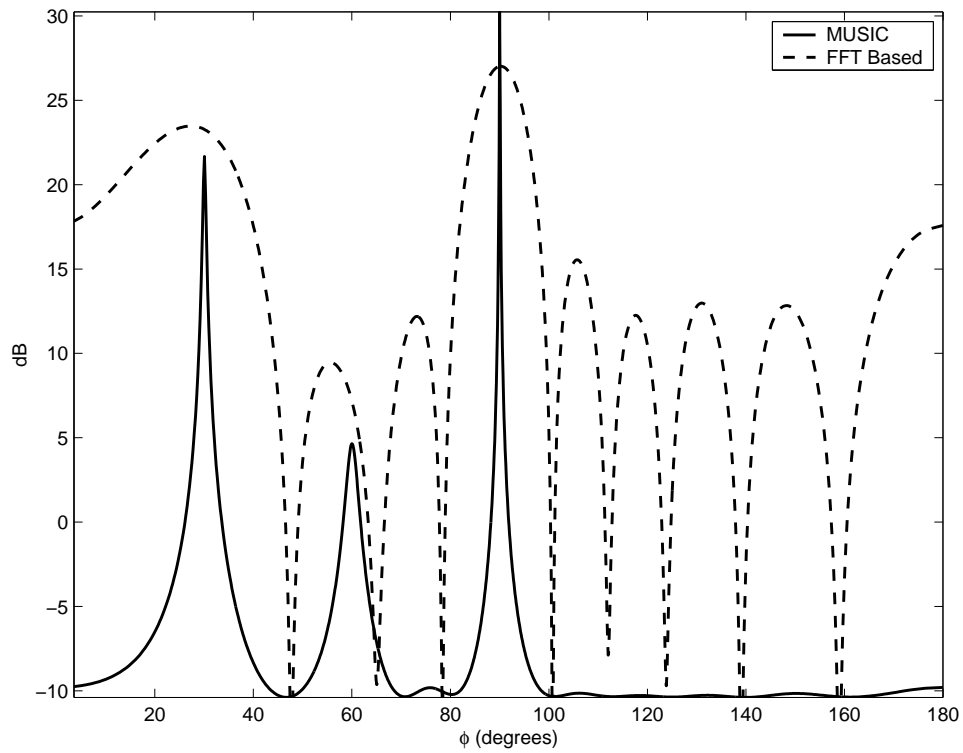Figure 3: The eigenvalues of the ideal and estimated correlation matrix

Figure 4: Comparision of the performance of the correlation (FFT) approach and the MUSIC algorithm

11

If the number of signals (M) is unknown, it is difficult then to decide which eigenvalues are equal, i.e., estimating the number of signals is error prone. The "equality" of the noise eigenvalues can be measured in a statistical sense. This is an issue we will address in Section 10. For now we will assume that M, the number of signals, is known.

In practice, therefore, the steps of MUSIC are:

1. Estimate the correlation matrix $\mathbf{R}$ using Eqn. (54). Find its eigendecomposition $\mathbf{R} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^H$.

2. Partition $\mathbf{Q}$ to obtain $\mathbf{Q}_n$, corresponding to the $(N - M)$ smallest eigenvalues of $\mathbf{Q}$, which spans the noise subspace.

3. Plot, as a function of ϕ, the MUSIC function $P_{\mathrm{MUSIC}}(\phi)$ in Eqn. (53).

4. The M signal directions are the M largest peaks of $P_{\mathrm{MUSIC}}(\phi)$.

Figure 4 plots the performance of the two algorithms (correlation and MUSIC) for the same example as in Fig. 3. The correlation plot is marked "FFT-based" since, for a linear array of equispaced isotropic sensors, the correlation approach is equivalent to taking a DTFT. Note the huge improvement of MUSIC over the non-adaptive correlation technique. The three peaks in MUSIC are clear and almost exactly on target. The signal arriving from angle 60° was the weakest, resulting in the broadest peak (worst accuracy). This is consistent with the CRB result of Section 2.

5.2   Root-MUSIC: Model Based Parameter Estimation

There is a significant problem with MUSIC as described above. The accuracy is limited by the discretization at which the MUSIC function $P_{\mathrm{MUSIC}}(\phi)$ is evaluated. More importantly, it requires either human interaction to decide on the largest M peaks or a comprehensive search algorithm to determine these peaks. This is an extremely computationally intensive process. Therefore, MUSIC by itself is not very practical - we require a methodology that results directly in numeric values for the estimated directions. This is where Root-MUSIC comes in.

Note that MUSIC is a technique that estimates the *spectrum* of the incoming data stream, i.e., it is a spectral estimation technique. The end product is a function $P_{\mathrm{MUSIC}}(\phi)$ as a function of the DOA, ϕ. Root-MUSIC, on the other hand, is an example of a *model-based parameter estimation* (MBPE) technique. We use a *model* of the received signal as a function of the DOA - here, the model is the steering vector. The DOA, ϕ, is a *parameter* in this model. Based on this model and the received data, we will estimate this parameter.

A crucial aspect of MBPE is that the estimation technique is valid only as much as the model itself is valid. For example, our steering vector model is not valid when we take mutual coupling

into account or for a circular array. Without accounting for the change in model, the estimation results will be significantly off base [3]. However, we will not be addressing these concerns in this class. For now, define

$$z = e^{jkd\cos\phi}. \tag{55}$$

Then, *assuming the receiving antenna is a linear array of equispaced, isotropic, elements,*

$$\mathbf{s}(\phi) = \begin{bmatrix} 1, \ z, \ z^2, \ \ldots, \ z^{N-1} \end{bmatrix}^T, \tag{56}$$

$$\Rightarrow \mathbf{q}_m^H \mathbf{s} = \sum_{n=0}^{N-1} q_{mn}^* z^n = q_m(z), \tag{57}$$

i.e., the inner product of the eigenvector $\mathbf{q}_m$ and the steering vector $\mathbf{s}(\phi)$ is equivalent to a polynomial in z. Since we are looking for the directions (ϕ) where $\mathbf{q}_m \perp \mathbf{s}(\phi)$, $m = (M+1), \ldots, N$, we are looking for the *roots of a polynomial.*

To find the polynomial whose roots we wish to evaluate, we use

$$\begin{aligned} P_{\mathrm{MUSIC}}^{-1}(\phi) &= \mathbf{s}^H(\phi)\mathbf{Q}_n\mathbf{Q}_n^H\mathbf{s}(\phi) \\ &= \mathbf{s}^H(\phi)\mathbf{C}\mathbf{s}(\phi) \end{aligned} \tag{58}$$

where

$$\mathbf{C} = \mathbf{Q}_n\mathbf{Q}_n^H \tag{59}$$

$$\Rightarrow P_{\mathrm{MUSIC}}^{-1}(\phi) = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} z^n C_{mn} z^{-m} = \sum_{m=0}^{N-1}\sum_{n=0}^{N-1} z^{(n-m)} C_{mn} \tag{60}$$

The final double summation can be simplified by rewriting it as a single sum by setting $l = n - m$. The range on l is set by the limits on n and m, i.e. $-(N-1) \le l \le (N-1)$ and

$$\Rightarrow P_{\mathrm{MUSIC}}^{-1}(\phi) = \sum_{l=-(N-1)}^{(N+1)} C_l z^l, \tag{61}$$

$$C_l = \sum_{n-m=l} C_{mn}, \tag{62}$$

i.e., C_l is the sum of the elements of $\mathbf{C}$ on the n-th diagonal. Eqn. (61) defines a polynomial of degree $(2N-2)$ with $(2N-2)$ zeros. However, we can show that not all zeros are independent. If z is a zero of the above polynomial, and of $P_{\mathrm{MUSIC}}^{-1}(\phi))$, $1/z^*$ is also a zero of the polynomial. The zeros of $P_{\mathrm{MUSIC}}^{-1}(\phi)$ therefore come in pairs.

Since z and $1/z^*$ have the same phase and reciprocal magnitude, one zero is within the unit circle and the other outside. Note that we are using this root to estimate the signal angle. From

the definition of z, only the phase carries the desired information, i.e., both z and $1/z^*$ carry the same desirable information. Also, without noise, the roots would fall on the unit circle.

The steps of Root-MUSIC are:

1. Estimate the correlation matrix $\mathbf{R}$ using Eqn. (54). Find its eigendecomposition $\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$.

2. Partition $\mathbf{Q}$ to obtain $\mathbf{Q}_n$, corresponds to the $(N-M)$ smallest eigenvalues of $\mathbf{Q}$, which spans the noise subspace. Find $\mathbf{C} = \mathbf{Q}_n\mathbf{Q}_n^H$.

3. Obtain C_l by summing the l-th diagonal of $\mathbf{C}$.

4. Find the zeros of the resulting polynomial in terms of $(N-1)$ pairs.

5. Of the $(N-1)$ roots within the unit circle, choose the M closest to the unit circle (z_m, $m = 1, \ldots M$).

6. Obtain the directions of arrival using

$$\phi_m = \cos^{-1}\left[\frac{\Im\ln(z_m)}{kd}\right], \quad m = 1, \ldots M \tag{63}$$

As Root-MUSIC only worries about the phase of the roots, errors in the magnitude are irrelevant (in the ideal case the magnitude of the roots would be unity). In some cases, especially in low SNR situations, Root-MUSIC may provide better performance than MUSIC.

5.3 Smooth-MUSIC

There are several variants of the MUSIC algorithm, including Cyclic-MUSIC and Smooth-MUSIC. Smooth-MUSIC is interesting because it overcomes the MUSIC assumption that all incoming signals are uncorrelated (we had set the matrix $\mathbf{A}$ to be diagonal). In a communication situation, assuming flat fading, there may be multipath components from many directions. These components would be correlated with each other. Correlated components reduce the rank of the signal correlation matrix $\mathbf{R}_s$, resulting in more than $(N-M)$ noise eigenvalues.

In smooth-MUSIC, the N elements are subdivided into L overlapping subarrays, each with P elements. For example, subarray 0 would include elements 0 through $P-1$, subarray 1 elements 1 through P, etc. Therefore, $L = N - P + 1$. Using the data from each subarray, L correlation matrices are estimated, each of dimension $P \times P$. The MUSIC algorithm then continues using a *smoothed* correlation matrix correlation matrix

$$\mathbf{R}_L = \frac{1}{L}\sum_{l=0}^{L-1}\mathbf{R}_l \tag{64}$$

This formulation can detect the DOA of up to $L-1$ correlated signals. This is because the signal correlation matrix component of $\mathbf{R}_L$ becomes full rank again. See [4] for additional details.

6 ESPRIT: Estimation of Signal Parameters using Rotational Invariance Techniques

ESPRIT is another parameter estimation technique, based on the fact that in the steering vector, the signal at one element is a constant phase shift from the earlier element. As in Eqn. (55) in Section 5.2, let $z_m = e^{jkd\cos\phi_m}$. Using Eqn. (43), the correlation matrix is dependent on $\mathbf{S}$, the $N \times M$ matrix of steering vectors given by

$$\mathbf{S} = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_M \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{N-2} & z_2^{N-2} & \cdots & z_M^{N-2} \\ z_1^{N-1} & z_2^{N-1} & \cdots & z_M^{N-1} \end{bmatrix}. \tag{65}$$

Based on this matrix, define two $(N-1) \times M$ matrices, $\mathbf{S}_0$ and $\mathbf{S}_1$,

$$\mathbf{S}_0 = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_M \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{N-2} & z_2^{N-2} & \cdots & z_M^{N-2} \end{bmatrix} \quad \mathbf{S}_1 = \begin{bmatrix} z_1 & z_2 & \cdots & z_M \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{N-2} & z_2^{N-2} & \cdots & z_M^{N-2} \\ z_1^{N-1} & z_2^{N-1} & \cdots & z_M^{N-1} \end{bmatrix} \tag{66}$$

and note that $\mathbf{S}_1 = \mathbf{S}_0 \mathbf{\Phi}$ where $\mathbf{\Phi}$ is the $M \times M$ matrix

$$\mathbf{\Phi} = \begin{bmatrix} z_1 & 0 & \cdots & 0 \\ 0 & z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_M \end{bmatrix}, \tag{67}$$

i.e. $\mathbf{\Phi}$ is a diagonal matrix whose entries correspond to the phase shift from one element to the next due to each individual signal. We see that if we can estimate $\mathbf{\Phi}$, we can estimate the DOA of all signals using Eqn. (55).

If $\mathbf{S}_0$ and $\mathbf{S}_1$ were known, we could solve for $\mathbf{\Phi}$ easily. Of course, they are unknown matrices and we must use proxies to obtain the same result. The ESPRIT algorithm begins by recognizing that the steering vectors in matrix $\mathbf{S}$ span the same subspace the matrix $\mathbf{Q}_s$, the $N \times M$ matrix of *signal eigenvectors*. Since both these matrices span the same subspace, there exists an invertible matrix $\mathbf{C}$ such that

$$\mathbf{Q}_s = \mathbf{S}\mathbf{C} \tag{68}$$

Defining matrices $\mathbf{Q}_0$ and $\mathbf{Q}_1$ derived from $\mathbf{Q}_s$ just as $\mathbf{S}_0$ and $\mathbf{S}_1$ were derived from $\mathbf{S}$, i.e., $\mathbf{Q}_0$ comprises the first $(N-1)$ rows of $\mathbf{Q}_s$ and $\mathbf{Q}_1$ the last $(N-1)$ rows of $\mathbf{Q}_s$, and using Eqn. (68), we have

$$
\begin{aligned}
\mathbf{Q}_0 &= \mathbf{S}_0\mathbf{C}, \\
\mathbf{Q}_1 &= \mathbf{S}_1\mathbf{C} = \mathbf{S}_0\mathbf{\Phi}\mathbf{C}.
\end{aligned}
$$

Consider

$$
\mathbf{Q}_1\mathbf{C}^{-1}\mathbf{\Phi}^{-1}\mathbf{C} = \mathbf{S}_0\mathbf{\Phi}\mathbf{C}\mathbf{C}^{-1}\mathbf{\Phi}^{-1}\mathbf{C} = \mathbf{S}_0\mathbf{C} = \mathbf{Q}_0. \tag{69}
$$

Now, let

$$
\begin{aligned}
\mathbf{\Psi}^{-1} &= \mathbf{C}^{-1}\mathbf{\Phi}^{-1}\mathbf{C}, \\
\Rightarrow \mathbf{Q}_1\mathbf{\Psi}^{-1} &= \mathbf{Q}_0, \\
\Rightarrow \mathbf{Q}_1 &= \mathbf{Q}_0\mathbf{\Psi} \tag{70}
\end{aligned}
$$

where

$$
\mathbf{\Psi} = \mathbf{C}^{-1}\mathbf{\Phi}\mathbf{C} \tag{71}
$$

Equation (70) implies that the matrix $\mathbf{\Phi}$ is a diagonal matrix of the *eigenvalues* of $\mathbf{\Psi}$. Using Eqns. (70) and (71) we now have a complete algorithm.

The steps of ESPRIT are:

1. Estimate the correlation matrix $\mathbf{R}$ using Eqn. (54). Find its eigendecomposition $\mathbf{R} = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H$.

2. Partition $\mathbf{Q}$ to obtain $\mathbf{Q}_s$, corresponds to the M largest eigenvalues of $\mathbf{Q}$, which spans the signal subspace.

3. Using least squares, solve Eqn. (71) to obtain an estimate of the $M \times M$ matrix $\mathbf{\Psi}$.

4. Find the eigenvalues of $\mathbf{\Psi}$. Its diagonal elements are the estimates of z_m that we are looking for.

5. Obtain the DOA using Eqn.(63).

In practice, one would obtain the estimate of $\mathbf{\Psi}$ not using least squares, but Total Least Squares (TLS). This is an improved least squares technique detailed in Section 9.

Note that ESPRIT represents a significantly greater computation load than MUSIC. This is because we need two eigendecompositions, of the correlation matrix $\mathbf{R}$ and the estimated $\mathbf{\Psi}$. Furthermore, we need to solve a least squares problem to estimate $\mathbf{\Psi}$.

7 Matrix Pencil

So far, the "adaptive" algorithms we developed, the MLE, MUSIC and ESPRIT, are all dependent on an estimate of the correlation matrix $\mathbf{R}$. Estimating this matrix is a significant computation load as we need at least K samples of the data $\mathbf{x}$ (K snapshots) where $K > 2N$. The inherent assumption is that all K samples follow the same statistics, i.e., the data is homogeneous. In an environment in which the fading characteristics are rapidly changing, this may not be valid. More importantly, estimating the correlation matrix is computationally intensive.

This motivates the development of a "non-statistical" or "direct data domain" (D^3) technique known as Matrix Pencil [5]. Matrix Pencil was originally developed for the estimation of the poles of a system. However, it can be applied as well to DOA estimation [6]. In the original Matrix Pencil the received data at time index n is given by

$$x_n = \sum_{m=1}^{M} A_m z_m^n + n_n, \tag{72}$$

where $z_m = e^{jkd \cos \phi_m \Delta t}$ represent the *poles of the system*, n_n represents the AWGN. The goal is to estimate z_m given x_n, $n = 0, \ldots N-1$.

In our case, the data is received at the terminals of the N antenna elements, otherwise the formulation is exactly the same. Hence, the original Matrix Pencil algorithm is applicable to DOA estimation. It is interesting to note that Matrix Pencil has many similarities to the ESPRIT technique, however without estimating a correlation matrix. We begin by defining two $(N-L) \times L$ matrices $\mathbf{X}_0$ and $\mathbf{X}_1$ as

$$\mathbf{X}_0 = \begin{bmatrix} x_0 & x_1 & \cdots & x_{L-1} \\ x_1 & x_2 & \cdots & x_L \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-L-1} & x_{N-L} & \cdots & x_{N-2} \end{bmatrix}, \quad \mathbf{X}_1 = \begin{bmatrix} x_1 & x_2 & \cdots & x_L \\ x_2 & x_3 & \cdots & x_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-L} & x_{N-L+1} & \cdots & x_{N-1} \end{bmatrix}, \tag{73}$$

where L is a *pencil parameter* that must satisfy

$$M \leq L \leq N - L \qquad \text{N even,}$$
$$M \leq L \leq N - L + 1 \qquad \text{N odd.} \tag{74}$$

The basis of Matrix Pencil is that, based on the data model, we can write these matrices as

$$\mathbf{X}_0 = \mathbf{Z}_1 \mathbf{A} \mathbf{Z}_2, \tag{75}$$
$$\mathbf{X}_1 = \mathbf{Z}_1 \mathbf{A} \mathbf{\Phi} \mathbf{Z}_2, \tag{76}$$

where $\boldsymbol{\Phi}$ is the same as in ESPRIT, the diagonal matrix that we want to estimate. The four matrices are given by

$$
\mathbf{Z}_1 = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ z_1 & z_2 & \cdots & z_M \\ \vdots & \vdots & \ddots & \vdots \\ z_1^{(N-L-1)} & z_2^{(N-L-1)} & \cdots & z_M^{(N_-L-1)} \end{bmatrix}_{(N-L)\times M} \tag{77}
$$

$$
\mathbf{Z}_2 = \begin{bmatrix} 1 & z_1 & \cdots & z_1^{L-1} \\ 1 & z_2 & \cdots & z_2^{L-1} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & z_M & \cdots & z_M^{L-1} \end{bmatrix}_{M\times L} \tag{78}
$$

$$
\boldsymbol{\Phi} = \begin{bmatrix} z_1 & 0 & \cdots & 0 \\ 0 & z_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & z_M \end{bmatrix}_{M\times M} \tag{79}
$$

$$
\mathbf{A} = \begin{bmatrix} \alpha_1 & 0 & \cdots & 0 \\ 0 & \alpha_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \alpha_M \end{bmatrix}_{M\times M} \tag{80}
$$

Without noise, for the choice of pencil parameter L that satisfies the constraints in Eqn. (74), the matrices $\mathbf{X}_0$ and $\mathbf{X}_1$ have rank M. Consider the matrix pencil $\mathbf{X}_1 - \lambda \mathbf{X}_0 = \mathbf{Z}_1 \mathbf{A} \left[\boldsymbol{\Phi} - \lambda \mathbf{I} \right] \mathbf{Z}_2$. For arbitrary λ, this matrix difference also has rank M. However, if λ is one of the z_m, i.e. $\lambda = z_m$, for some $m \in [1, M]$, the rank of the matrix difference reduces by one to $M - 1$. This implies that we can find the poles (z_m) as the *generalized eigenvalues* of the matrix pair $[\mathbf{X}_0, \mathbf{X}_1]$, i.e.

$$
\mathbf{X}_1 \mathbf{q} = \lambda \mathbf{X}_0 \mathbf{q}. \tag{81}
$$

Note that $\mathbf{q}$, the generalized eigenvector, has *no relationship* to the eigenvectors of the correlation matrix. The M generalized eigenvalues of this matrix pair form the estimates of the z_m and the DOA may be obtained using Eqn. (63).

The steps of Matrix Pencil are therefore

1. Given N and M, choose L to satisfy Eqn. (74).

2. Form matrices $\mathbf{X}_0$ and $\mathbf{X}_1$.

3. Find z_m as the generalized eigenvalues of the matrix pair $[\mathbf{X}_0, \mathbf{X}_1]$.

4. Find the DOA using Eqn. (63).

Note that finding the generalized eigenvalues of the matrix pair $[\mathbf{X}_0, \mathbf{X}_1]$ is equivalent to finding the eigenvalues of $\left[\mathbf{X}_0^H \mathbf{X}_0\right]^{-1} \mathbf{X}_0^H \mathbf{X}_1$.

The similarities of Matrix Pencil and ESPRIT are clear. Both algorithms estimate a diagonal matrix whose entries are the poles of the system (what we call z_m). The major difference is that ESPRIT works with the signal subspace as defined by the correlation matrix, while Matrix Pencil works with the data directly. This represents a significant savings in terms of computation load.

As with ESPRIT, in practice, one would implement a TLS version of Matrix Pencil as described in Section 9.2.

8 Comparison Of Methods

Number of Resolvable Signals : In MUSIC we assumed that the number of elements, N, was greater than the number of signals, M. This is required because MUSIC depends on the existence of a noise subspace. Therefore with N elements, MUSIC can resolve a maximum of $(N-1)$ signals. In ESPRIT, a similar argument holds.

On the other hand, in Matrix Pencil, due to Eqn. (74), the maximum value of the pencil parameter L (and hence M) is $N/2$ for even N and $(N+1)/2$ for odd N. This is the penalty we must pay for not estimating a covariance matrix.

Accuracy : In terms of accuracy, all the adaptive techniques have similar accuracy. Figure 5 plots an example case of applying Root-MUSIC to a CDMA situation, while Fig. 6 plots the performance of Matrix Pencil for the same situation. This example uses a $N=7$ element array receiving two incoming signals with 3 multipath components each. The CDMA receiver is matched to the first path of the first signal. Due to the spreading gain (and the fact that we are using 4 samples per chip), the other 5 "interfering" signals are suppressed leaving only one signal whose DOA is to be estimated. This signal has nominal SNR = -20dB, which with the +27dB gain $(10\log_{10}(128 \times 4))$ is effectively an SNR of 7dB.

As can be seen, for both Root-MUSIC and Matrix Pencil, the DOA estimation is very accurate. Both techniques yield a root mean square error of about 0.3 degrees. This is despite the fact that only 5 snapshots are used to estimate a 7×7 correlation matrix. This is because there is effectively only one signal to locate. The other signals have been suppressed due to the spreading gain. 5 snapshots are adequate to estimate the signal subspace (which is orthogonal to the noise subspace).

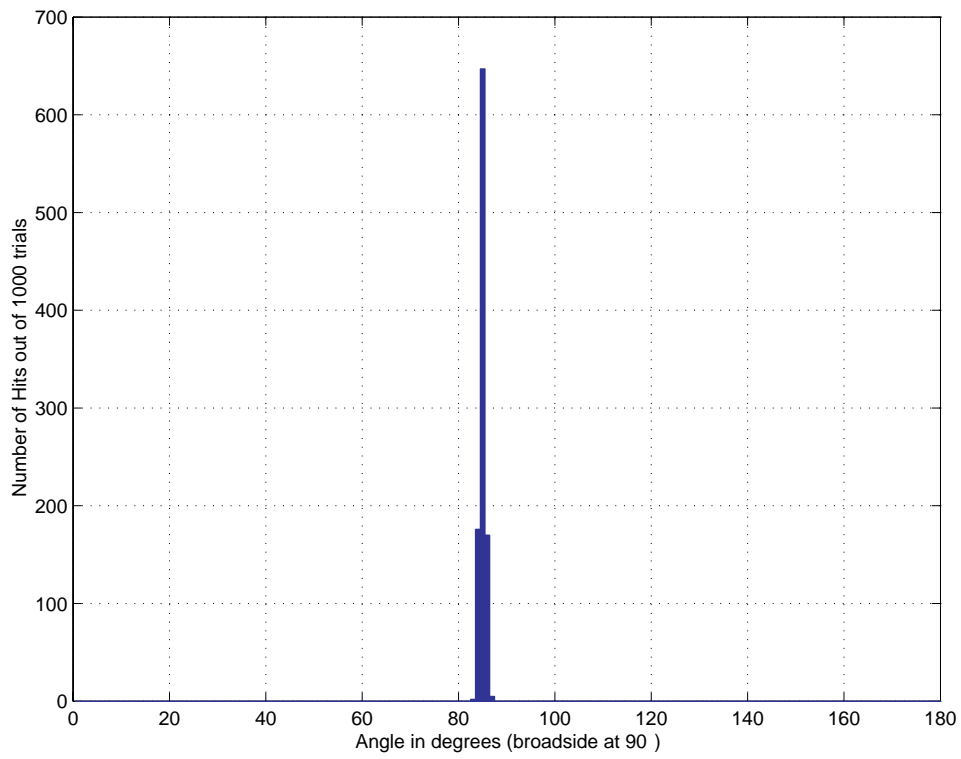One problem with applying Matrix Pencil with multiple snapshots is that one needs a coherent

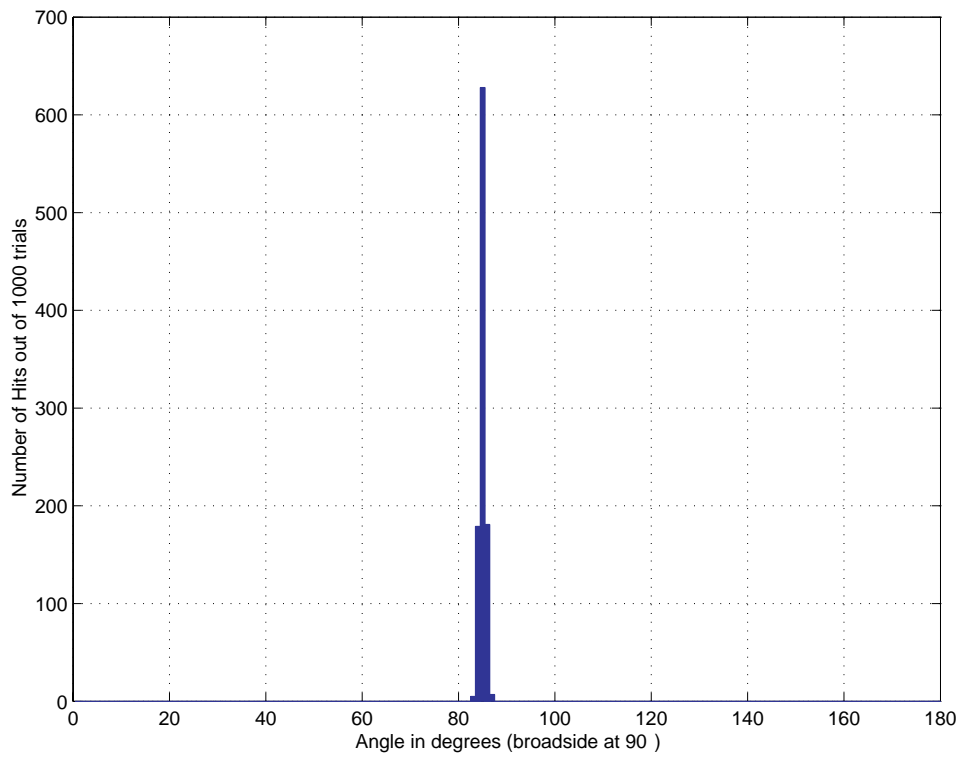Figure 5: Accuracy of Root MUSIC. 5 snapshots.

Figure 6: Accuracy of Matrix Pencil. 5 snapshots.

20

detector and knowledge of the data being transmitted (training) . The SNR is improved by averaging the received data. If the phase information is not removed, due to the random phase, the average would tend to zero! However, assuming we have a coherent detector, Matrix Pencil has the same accuracy as Root-MUSIC.

A significant advantage of Matrix Pencil is that it *does not* require multiple snapshots. Figure 7 plots the performance of Matrix Pencil using a single snapshot. As can be seen, while the accuracy is lower than in Fig. 6, it is still quite accurate. The RMSE is about 1.8degrees. In a similar situation, Root-MUSIC has a RMSE of greater than 50 degrees (not shown)! This is clearly due to the inaccurate estimate of the correlation matrix.

Figure 8 illustrates another significant advantage of Matrix Pencil. In comparing the computation loads Matrix Pencil is at least twice as fast at Root-MUSIC. This is because of the fact that Matrix Pencil does not require the estimation of a correlation matrix.

9   Total Least Squares

The least squares approach was developed to solve an over-determined system of equations

$$\mathbf{A}\mathbf{x} = \mathbf{b}, \tag{82}$$

where $\mathbf{A}$ is a $N \times M$ matrix, $\mathbf{x}$ is a length M vector of unknowns and $\mathbf{b}$ is a length N vector of observations. The LS approach finds $\mathbf{x}$ such that the residual error, $||\mathbf{A}\mathbf{x} - \mathbf{b}||^2$, is minimized. The LS solution is given by $\mathbf{x}_{LS} = \left[\mathbf{A}^H\mathbf{A}\right]^{-1}\mathbf{A}^H\mathbf{b}$. The least squares approach can also be looked finding the error matrix $\mathbf{E}$ with minimum Frobenius norm[2] such that

$$[\mathbf{A} + \mathbf{E}]\,\mathbf{x} = \mathbf{b}.$$

This approach is optimal if the vector $\mathbf{b}$ is noise-free and all the measurement errors are in $\mathbf{A}$. However, if both $\mathbf{A}$ and $\mathbf{b}$ are noisy, then the TLS approach is optimal. The TLS approach reorders the original equation to

$$[\mathbf{A} \mid -\mathbf{b}]\begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} = 0. \tag{83}$$

Now, any rectangular $N \times M$ matrix $\mathbf{A}$ can be written in its singular value decomposition (SVD) $\mathbf{A} = \mathbf{U}\mathbf{\Sigma}\mathbf{V}^H$. Assuming $N > M$, the $N \times N$ matrix of left singular vectors $\mathbf{U}$ are the N eigenvectors of $\mathbf{A}\mathbf{A}^H$, the $M \times M$ matrix of right singular vectors $\mathbf{V}$ the M eigenvectors of $\mathbf{A}^H\mathbf{A}$ and $\mathbf{\Sigma}$ is a $N \times M$ matrix whose diagonal elements are the square roots M of the non-zero eigenvalues of $\mathbf{A}^H\mathbf{A}$. Furthermore, $\mathbf{V}^H\mathbf{V} = \mathbf{I}_{M \times M}$ and $\mathbf{U}^H\mathbf{U} = \mathbf{I}_{N \times N}$. Note that the singular values are always real. The SVD is also defined for $N < M$. Please see [7] for additional details.

---

[2]The Frobenius norm of a $N \times M$ matrix $\mathbf{A}$, denoted as $||A||_F$, is defined by $||A||_F^2 = \sum_{n=1}^N \sum_{m=1}^M |A_{nm}|^2$
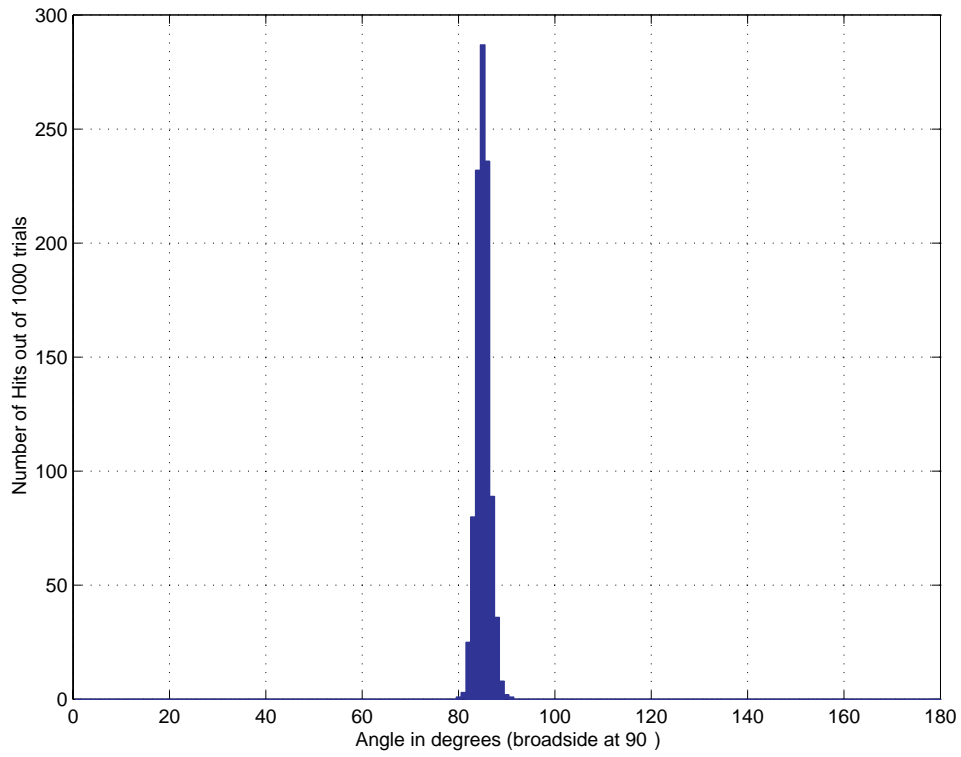
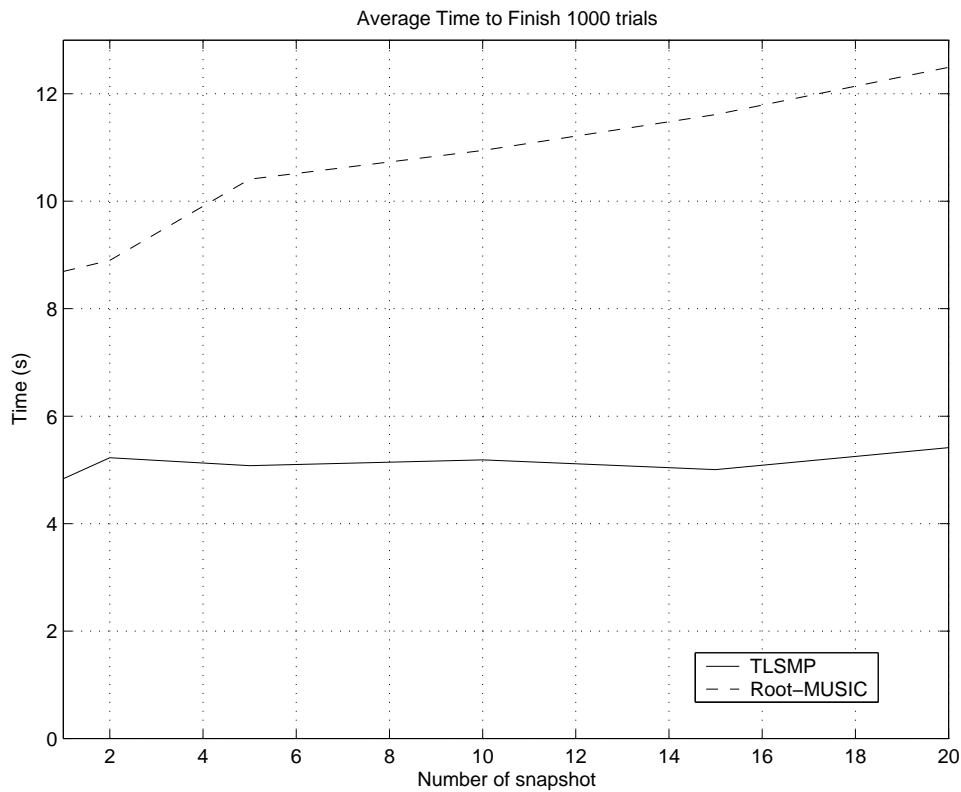Figure 7: Accuracy of Matrix Pencil. One snapshot.

Figure 8: Comparing the computation load of Matrix Pencil and MUSIC

Total Least Squares finds the error matrix $[\mathbf{E} \mid \mathbf{e}]$ with minimum Frobenius norm such that

$$[\mathbf{A} + \mathbf{E} \mid -\mathbf{b} + \mathbf{e}] \begin{bmatrix} \mathbf{x} \\ 1 \end{bmatrix} = 0. \tag{84}$$

The solution can be shown to be the singular vector in $\mathbf{V}$ that corresponds to the minimum singular value. This singular value is normalized such that its final entry is 1 to satisfy Eqn. (83).

9.1 TLS ESPRIT

In ESPRIT, the estimate of the matrix $\boldsymbol{\Phi}$ is obtained using the matrix equation in Eqn. (70), $\mathbf{Q}_1 = \mathbf{Q}_0 \boldsymbol{\Psi}$. Note that both matrices arise from the eigenvalue decomposition of a noise (estimated) correlation matrix. There are, therefore, errors in both matrices and a simple LS algorithm is not completely valid. The TLS version of ESPRIT is based on the TLS method described above [4]. We start by writing the K data snapshots as a $K \times N$ data matrix (each snapshot is a length N vector of data received at the N elements of the array)

$$\mathbf{X} = \begin{bmatrix} x_{01} & x_{11} & \cdots & x_{(N-1)1} \\ x_{02} & x_{12} & \cdots & x_{(N-1)2} \\ \vdots & \vdots & \ddots & \vdots \\ x_{0K} & x_{1K} & \cdots & x_{(N-1)K} \end{bmatrix} \tag{85}$$

A SVD of this matrix $\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{Q}^H$. Note that $\mathbf{Q}$ is the matrix of eigenvectors of $\mathbf{X}^H \mathbf{X}$, which is proportional to the correlation matrix $\mathbf{R}$, i.e., the matrix of singular vectors $\mathbf{Q}$ is same as the eigenvector matrix $\mathbf{Q}$ defined in Section 6. Define matrices $\mathbf{Q}_0$ and $\mathbf{Q}_1$ as in Section 6. Form the $(N-1) \times 2M$ matrix $[\mathbf{Q}_0 \, \mathbf{Q}_1]$ and take the SVD of this matrix, $[\mathbf{Q}_0 \, \mathbf{Q}_1] = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H$. Note that this new $\mathbf{U}$ and $\boldsymbol{\Sigma}$ has nothing to do with the same matrices in the SVD of the matrix $\mathbf{X}$. The matrix we are interested in is the $2M \times 2M$ matrix $\mathbf{V}$. Partition this matrix into four $M \times M$ matrices

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix} \tag{86}$$

Now, the poles z_m are the eigenvalues of $-\mathbf{V}_{12}\mathbf{V}_{22}^{-1}$.

9.2 TLS Matrix Pencil

A similar TLS approach can be applied to Matrix Pencil. However, since Matrix Pencil uses only one snapshot, the data matrix is defined as the $(N-L) \times (L+1)$ matrix

$$\mathbf{X} = \begin{bmatrix} x_0 & x_1 & \cdots & x_L \\ x_1 & x_2 & \cdots & x_{L+1} \\ \vdots & \vdots & \ddots & \vdots \\ x_{N-L-1} & x_{N-L} & \cdots & x_N \end{bmatrix} \tag{87}$$

A SVD of this matrix yields $\mathbf{X} = \mathbf{U\Sigma V}^H$. Only M of the singular values of this matrix correspond to signals, while the rest correspond to noise. We generate a new, *filtered* version of the data matrix $\tilde{\mathbf{X}} = \mathbf{U'\Sigma'V'}$, where $\mathbf{U'}$ is the first M left singular vectors, $\mathbf{V'}$ the first M right singular vectors and $\mathbf{\Sigma'}$ are the first M signal singular values.

Matrix Pencil continues as before using the *filtered* data with $\mathbf{X}_0'$ is the first L columns of $\tilde{\mathbf{X}}$ and $\mathbf{X}_1'$ is the last L columns of the same matrix.

10 Estimating Number of Signals

So far we have assumed that we knew M, the number of signals that are incident on the array. This number is always important, setting the partition of the correlation matrix $\mathbf{R}$ in MUSIC and ESPRIT and determining the number of vectors to use in the TLS Matrix Pencil case. In some cases, this assumption is valid - if there is a base station controlling the number of users entering the system. However, even in this case, there may be external interference sources *that also act as incident signals*. The number M includes *all* incident signals. Estimating the number of signals is therefore a crucial function.

We present here two techniques to estimate the number of signals based on the work in [8]. The algorithms start with realizing that the number of signals M is the number of elements N minus the number of noise eigenvalues. These eigenvalues are all equal in the ideal case and easy to identify. In practice, due to the estimation of the correlation matrix, the noise eigenvalues are not equal, but are close to each other. The algorithms therefore use an estimate of the *closeness of the eigenvalues*. If K snapshots are used to estimate the correlation matrix and assuming there are d signals, a measure of closeness of the noise eigenvalues would be the ratio of their geometric mean to their arithmetic mean.

$$L(d) = -K(N-1)\log\left\{\frac{\left[\prod_{n=d+1}^{N}\lambda_n\right]^{1/(N-d)}}{\frac{1}{N-d}\sum_{n=d+1}^{N}\lambda_n}\right\} \tag{88}$$

Based on this measure of closeness, Wax and Kailath define two information theoretic criteria. The first is the Akaiake Information Criterion (AIC) and the second uses the Minimum Description Length (MDL)

$$
\begin{aligned}
\mathrm{AIC}(d) &= L(d) + d(2N - d) \\
&= -K(N-d)\log\left\{\frac{\left[\prod_{n=d+1}^{N}\lambda_n\right]^{1/(N-d)}}{\frac{1}{N-d}\sum_{n=d+1}^{N}\lambda_n}\right\} + d(2N - d),
\end{aligned}
\tag{89}
$$

$$\begin{aligned} \text{MDL}(d) &= L(d) + \frac{1}{2}d(2N-d)\log K \\ &= -K(N-d)\log\left\{\frac{\left[\prod_{n=d+1}^{N}\lambda_n\right]^{1/(N-d)}}{\frac{1}{N-d}\sum_{n=d+1}^{N}\lambda_n}\right\} + \frac{1}{2}d(2N-d)\log K. \end{aligned} \qquad (90)$$

The number of signals is the point at which these measures achieve their minimum. Unfortunately, these two techniques do not always give the same number of signals. The authors show that the MDL approach results in unbiased estimates, while the AIC approach yields biased estimates. In general, therefore, it is better to use the MDL than the AIC approach. Also, recently, Chen *et.al.* demonstrated an approach based on ranking and selection theory that shows promise [9].

References

[1] A. Gershman. Class slides. Advanced Topics in DSP. McMaster University. Personal Communication.

[2] I. S. Reed, J. Mallett, and L. Brennan, "Rapid convergence rate in adaptive arrays," *IEEE Transactions on Aerospace and Electronic Systems*, vol. 10, No. 6, pp. 853–863, Nov. 1974.

[3] C. K. E. Lau, R. S. Adve, and T. K. Sarkar, "Mutual coupling compensation based on the minimum norm with applications in direction of arrival estimation," *IEEE Trans. on Antennas and Propagation*, vol. 52, pp. 2034 – 2041, August 2004.

[4] J. C. Liberti and T. S. Rappaport, *Smart Antennas for Wireless Communications: IS-95 and Third Generation CDMA Applications*. Upper Saddle River, New Jersey: Prentice-Hall, Inc., 1997.

[5] Y. Hua and T. Sarkar, "Matrix pencil method for estimating parameters of exponentially damped/undamped sinusoids in noise," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 38, pp. 814–824, May 1990.

[6] R. Adve, *Elimination of the Effects of Mutual Coupling in Adaptive Thin Wire Antennas*. PhD thesis, Syracuse University, December 1996.

[7] G. H. Golub and C. F. V. Loan, *Matrix Computations*. Johns Hopkins University Press, Baltimore, 1983.

[8] M. Wax and T. Kailath, "Detection of signals by information theoretic criteria," *IEEE Transactions on Acoustics, Speech and Signal Processing*, vol. 33, pp. 387–392, April 1985.

[9] P. Chen, M. Wicks, and R. Adve, "Development of a procedure for detecting the number of signal in a radar measurement," *IEE Proceedings on Radar Sonar and Navigation*, vol. 148, pp. 219–226, August 2001.