MAXIMUM LIKELIHOOD SEPARATION OF PHASE MODULATED SIGNALS

Amir Leshem

Delft University of Technology, Delft, The Netherlands
Email: leshem@cas.et.tudelft.nl

ABSTRACT

In this paper we present a Newton scoring algorithm for the maximum likelihood separation and direction of arrival estimation of constant modulus signals, using a calibrated array. The main technical step is the inversion of the Fisher information matrix, and an analytic formula for the update step in the Newton method. We present the algorithm based on the derived update and discuss potential initializations. We also present the computational complexity of the update. Finally we present simulation results comparing the method to the ESPRIT and the CM-DOA.

1. INTRODUCTION

Direction-of-arrival (DOA) estimation of multiple signals impinging on an antenna array is a well-studied problem in signal processing. "Traditional" methods exploit knowledge of the array manifold or its structure without using information on the signals. Example algorithms are MUSIC [1], MLE [2], and WSF [3]. In all these methods the estimation of the signals waveform is done by multiplying a weight matrix by the received data matrix. This does not enable introduction of constrains on the signals structure. Other, 'blind' methods exploit properties of the signals such as non-Gaussianity [4], or cyclostationarity [5]. These methods are more robust to array manifold errors due to the extra information they use.

Recent studies of the problem of DOA estimation based on the CM property [6], yielded good sub-optimal algorithms. However numerical study shows that these algorithms are indeed suboptimal in the sense that they do not achieve the CRB in some circumstances. This suggests that a possible improvement is possible by using the maximum likelihood estimator. The main problem with using the MLE together with the CM property is the large dimension of the parameter space. This makes the estimation even using iterative numerical methods very intractable.

In this paper we present exact analytic expression for the inverse of the Fisher information matrix. This implies that the updating of a scoring algorithm can be done linearly in the number of samples, rather then cubic as would be the case in direct numerical inversion of the information matrix. We then devise a scoring algorithm for maximum likelihood estimation based on initialization with a suboptimal method. Finally we present the computational complexity of the algorithm and demonstrate its effectiveness by simulations. Due to space limitations we omit all derivations. These will be presented in a more detailed version.

Amir Leshem was supported by the NOEMI project of the STW under contract no. DEL77-4476.

2. PROBLEM FORMULATION

Consider an array with p sensors receiving q narrow-band constant modulus signals. Under standard assumptions for the array manifold, we can describe the received signal as an instantaneous linear combination of the source signals, i.e.,

$$\mathbf{x}(t) = \mathbf{ABs}(t) + \mathbf{n}(t) \tag{1}$$

where

- $\mathbf{x}(t) = [x_1(t), \cdots, x_p(t)]^T$ is a $p \times 1$ vector of received signals at time t,
- $\mathbf{A} = \mathbf{A}(\boldsymbol{\theta}) = [\mathbf{a}(\theta_1), \cdots, \mathbf{a}(\theta_q)]$, where $\mathbf{a}(\theta)$ is the array response vector for a signal from direction θ, and $\boldsymbol{\theta} = [\theta_1, \cdots, \theta_q]$ is the DOA vector of the sources,
- $\mathbf{B} = \mathrm{diag}(\boldsymbol{\beta})$ is the channel gain matrix, with parameters $\boldsymbol{\beta} = [\beta_1, \cdots, \beta_q]^T$, where $\beta_i \in \mathbb{R}^+$ is the amplitude of the i-th signal as received by the array,
- $\mathbf{s}(t) = [s_1(t), \cdots, s_q(t)]^T$ is a $q \times 1$ vector of source signals at time t,
- $\mathbf{n}(t)$ is the $p \times 1$ additive noise vector, which is assumed spatially and temporally white Gaussian distributed with covariance matrix $\nu\mathbf{I}$, where $\nu = \sigma^2$ is the noise variance.

The derivation can be easily extended to spatially colored Gaussian noise, with known covariance matrix $\mathbf{Q}$, at the expense of greater notational complexity. We prefer to let the interested reader do these changes, should he need them. In our problem, the array is assumed to be calibrated so that the array response vector $\mathbf{a}(\theta)$ is a known function. As usual, we require that the array manifold satisfies the uniqueness condition.

We further assume that all sources have constant modulus. This is represented by the assumption that for all t, $|s_i(t)| = 1$ $(i = 1, \cdots, q)$. Unequal source powers are absorbed in the gain matrix $\mathbf{B}$. Phase offsets of the sources after demodulation are part of the s_i. Thus we can write $s_i(t) = e^{j\phi_i(t)}$, where $\phi_i(t)$ is the unknown phase modulation for source i, and we define $\boldsymbol{\phi}(t) = [\phi_1(t), \cdots, \phi_q(t)]^T$ as the phase vector for all sources at time t.

Finally, we assume that N samples $[\mathbf{x}(1), \cdots, \mathbf{x}(N)]$ are available.

3. THE LIKELIHOOD, THE INFORMATION MATRIX AND THE NEWTON UPDATE FORMULA

In this section we present an analytic expression for the inverse of the information matrix, and the update formula for the scoring algorithm. The results below extends the results in [6], by computing analytically *all* the entries of the inverse information matrix, and not only the diagonal entries.

The likelihood function is given by

$$L(\mathbf{x}|\mathbf{s}, \boldsymbol{\theta}, \boldsymbol{\beta}, \nu) = c \exp\left\{-\tfrac{1}{\nu}\textstyle\sum_{k=1}^{N} \|\mathbf{x}(k) - \mathbf{ABs}(k)\|^2\right\}$$

where $c = \frac{1}{(2\pi)^{N(\frac{\nu}{2})^{pN}}}$.

Let $\mathcal{L}(\mathbf{x}|\mathbf{s}, \boldsymbol{\theta}, \nu) = \log L(\mathbf{x}|\mathbf{s}, \boldsymbol{\theta}, \boldsymbol{\beta}, \nu)$. After omitting constants we obtain

$$\mathcal{L}(\mathbf{x}|\mathbf{s}, \boldsymbol{\theta}, \boldsymbol{\beta}, \nu) = -pN \log \nu - \frac{1}{\nu}\sum_{k=1}^{N} \|\mathbf{x}(k) - \mathbf{ABs}(k)\|^2.$$

Following [7], the estimation of the noise variance is decoupled from all other parameters, and its bound can be computed separately as

$$\mathrm{CRB}_N(\nu) = \frac{\nu^2}{pN}.$$

The remaining parameters are collected in the vector

$$\boldsymbol{\rho} = [\boldsymbol{\phi}(1)^T, \cdots, \boldsymbol{\phi}(N)^T, \boldsymbol{\theta}^T, \boldsymbol{\beta}^T]^T.$$

Define

$$\mathbf{S}_k = \mathrm{diag}(\mathbf{s}(k)) \quad \text{and} \quad \mathbf{D} = \left[\frac{d\mathbf{a}}{d\theta}(\theta_1), \cdots, \frac{d\mathbf{a}}{d\theta}(\theta_q)\right].$$

The Fisher information matrix associated to the estimation of the parameter vector $\boldsymbol{\rho}$ can be derived as

$$\mathbf{F}_N = \frac{2}{\nu}\left[\begin{array}{ccc|cc} \mathbf{H}_1 & & 0 & \boldsymbol{\Delta}_1^T & \mathbf{E}_1^T \\ & \ddots & & \vdots & \vdots \\ 0 & & \mathbf{H}_N & \boldsymbol{\Delta}_N^T & \mathbf{E}_N^T \\ \hline \boldsymbol{\Delta}_1 & \cdots & \boldsymbol{\Delta}_N & \boldsymbol{\Gamma} & \boldsymbol{\Lambda}^T \\ \mathbf{E}_1 & \cdots & \mathbf{E}_N & \boldsymbol{\Lambda} & \boldsymbol{\Upsilon} \end{array}\right] \tag{2}$$

where

$$\begin{aligned}
\mathbf{H}_k &:= \tfrac{\nu}{2}E\tfrac{\partial \mathcal{L}}{\partial \boldsymbol{\phi}(k)}(\tfrac{\partial \mathcal{L}}{\partial \boldsymbol{\phi}(k)})^T = \mathrm{Re}(\mathbf{S}_k^*\mathbf{B}^*\mathbf{A}^*\mathbf{ABS}_k) \\
\boldsymbol{\Delta}_k &:= \tfrac{\nu}{2}E\tfrac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}}(\tfrac{\partial \mathcal{L}}{\partial \boldsymbol{\phi}(k)})^T = -\mathrm{Im}(\mathbf{S}_k^*\mathbf{B}^*\mathbf{D}^*\mathbf{ABS}_k) \\
\mathbf{E}_k &:= \tfrac{\nu}{2}E\tfrac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}}(\tfrac{\partial \mathcal{L}}{\partial \boldsymbol{\phi}(k)})^T = -\mathrm{Im}(\mathbf{S}_k^*\mathbf{A}^*\mathbf{ABS}_k) \\
\boldsymbol{\Gamma} &:= \tfrac{\nu}{2}E\tfrac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}}(\tfrac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}})^T = \textstyle\sum_{k=1}^{N} \mathrm{Re}(\mathbf{S}_k^*\mathbf{B}^*\mathbf{D}^*\mathbf{DBS}_k) \\
\boldsymbol{\Lambda} &:= \tfrac{\nu}{2}E\tfrac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}}(\tfrac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}})^T = \textstyle\sum_{k=1}^{N} \mathrm{Re}(\mathbf{S}_k^*\mathbf{A}^*\mathbf{DBS}_k) \\
\boldsymbol{\Upsilon} &:= \tfrac{\nu}{2}E\tfrac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}}(\tfrac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}})^T = \textstyle\sum_{k=1}^{N} \mathrm{Re}(\mathbf{S}_k^*\mathbf{A}^*\mathbf{AS}_k)
\end{aligned}$$

Let $\nabla_{\boldsymbol{\phi}(k)}\mathcal{L}, \nabla_{\boldsymbol{\theta}}\mathcal{L}, \nabla_{\boldsymbol{\beta}}\mathcal{L}$ be the partial gradients of $\mathcal{L}$ with respect to the signals' phases in the the k'th sample, the signals' DOA's and the signals' power respectively. They can be computed as

$$\begin{aligned}
\nabla_{\boldsymbol{\phi}(k)}\mathcal{L} &= \tfrac{\partial \mathcal{L}}{\partial \boldsymbol{\phi}(k)} = \frac{2}{\nu}\mathrm{Im}\left(\mathbf{S}(k)^*\mathbf{B}^*\mathbf{A}^*\mathbf{e}(k)\right). \\
\nabla_{\boldsymbol{\theta}}\mathcal{L} &= \tfrac{\partial \mathcal{L}}{\partial \boldsymbol{\theta}} = \frac{2}{\nu}\sum_{k=1}^{N}\mathrm{Re}\left(\mathbf{S}^*(k)\mathbf{B}^*\mathbf{D}^*\mathbf{e}(k)\right) \\
\nabla_{\boldsymbol{\beta}}\mathcal{L} &= \tfrac{\partial \mathcal{L}}{\partial \boldsymbol{\beta}} = \frac{2}{\nu}\sum_{k=1}^{N}\mathrm{Re}\left(\mathbf{S}^*(k)\mathbf{A}^*\mathbf{e}(k)\right).
\end{aligned} \tag{3}$$

Let

$$\left[\begin{array}{cc} \boldsymbol{\Xi}_{11} & \boldsymbol{\Xi}_{12} \\ \boldsymbol{\Xi}_{21} & \boldsymbol{\Xi}_{22} \end{array}\right] = \sum_{k=1}^{N}\left[\begin{array}{cc} \boldsymbol{\Delta}_k\mathbf{H}_k^{-1}\boldsymbol{\Delta}_k^T & \boldsymbol{\Delta}_k\mathbf{H}_k^{-1}\mathbf{E}_k^T \\ \mathbf{E}_k\mathbf{H}_k^{-1}\boldsymbol{\Delta}_k^T & \mathbf{E}_k\mathbf{H}_k^{-1}\mathbf{E}_k^T \end{array}\right] \tag{4}$$

and define the $2q \times 2q$ matrix

$$\boldsymbol{\Psi} = \left[\begin{array}{cc} \boldsymbol{\Gamma} & \boldsymbol{\Lambda}^T \\ \boldsymbol{\Lambda} & \boldsymbol{\Upsilon} \end{array}\right] - \left[\begin{array}{cc} \boldsymbol{\Xi}_{11} & \boldsymbol{\Xi}_{12} \\ \boldsymbol{\Xi}_{21} & \boldsymbol{\Xi}_{22} \end{array}\right]. \tag{5}$$

The bounds on the individual parameters can be explicitly computed. The CRB for DOAs amplitudes and signal phases is

$$\begin{aligned}
\mathrm{CRB}_N(\boldsymbol{\theta}) &= \mathrm{diag}\left(\left[\boldsymbol{\Psi}_{11} - \boldsymbol{\Psi}_{12}\boldsymbol{\Psi}_{22}^{-1}\boldsymbol{\Psi}_{21}\right]^{-1}\right) \\
\mathrm{CRB}_N(\boldsymbol{\beta}) &= \mathrm{diag}\left(\left[\boldsymbol{\Psi}_{22} - \boldsymbol{\Psi}_{21}\boldsymbol{\Psi}_{11}^{-1}\boldsymbol{\Psi}_{12}\right]^{-1}\right)
\end{aligned} \tag{6}$$

$$\mathrm{CRB}_N(\boldsymbol{\phi}(k)) = \mathrm{diag}\left\{\mathbf{H}_k^{-1}\left[\mathbf{I} + [\boldsymbol{\Delta}_k^T \quad \mathbf{E}_k^T]\boldsymbol{\Psi}^{-1}\left[\begin{array}{c} \boldsymbol{\Delta}_k \\ \mathbf{E}_k \end{array}\right]\mathbf{H}_k^{-1}\right]\right\}. \tag{7}$$

Note that the number of samples and the quality of DOA estimation affects the bound on the phase estimation only through the matrix $\boldsymbol{\Psi}^{-1}$.

Further use of the Schur formula enables us to completely invert the information matrix and not only its diagonal elements. For inverting the matrix we first define four different parts of the matrix. To simplify notation we disregard the constant $\frac{2}{\nu}$, remembering that its inverse has to multiply the final result.

$$\mathbf{F}_{11} = \left[\begin{array}{ccc} \mathbf{H}_1 & & 0 \\ & \ddots & \\ 0 & & \mathbf{H}_N \end{array}\right] \quad \mathbf{F}_{12} = \left[\begin{array}{cc} \boldsymbol{\Delta}_1^T & \mathbf{E}_1^T \\ \vdots & \vdots \\ \boldsymbol{\Delta}_N^T & \mathbf{E}_N^T \end{array}\right]$$

$$\mathbf{F}_{21} = \begin{bmatrix} \boldsymbol{\Delta}_1 & ,\cdots, & \boldsymbol{\Delta}_N \\ \mathbf{E}_1 & ,\cdots, & \mathbf{E}_N \end{bmatrix} \quad \mathbf{F}_{22} = \begin{bmatrix} \boldsymbol{\Gamma} & \boldsymbol{\Lambda}^T \\ \boldsymbol{\Lambda} & \boldsymbol{\Upsilon} \end{bmatrix}$$

After some algebraic manipulations involving the Schur decomposition and Woodbury's identity we obtain

$$(\mathbf{F}^{-1})_{22} = \boldsymbol{\Psi}^{-1}$$

$$(\mathbf{F}^{-1})_{11} = (\mathbf{F}_{11})^{-1} + \mathbf{G}\boldsymbol{\Psi}^{-1}\mathbf{G}^T$$

$$(\mathbf{F}^{-1})_{12} = \mathbf{G}\boldsymbol{\Psi}^{-1}$$

$$(\mathbf{F}^{-1})_{21} = -\boldsymbol{\Psi}^{-1}\mathbf{G}^T$$

where

$$\mathbf{G} = \begin{bmatrix} \mathbf{H}_1^{-1}\boldsymbol{\Delta}_1^T & \mathbf{H}_1^{-1}\mathbf{E}_1^T \\ \vdots & \vdots \\ \mathbf{H}_N^{-1}\boldsymbol{\Delta}_N^T & \mathbf{H}_N^{-1}\mathbf{E}_N^T \end{bmatrix}.$$

Finally to obtain the update direction define the $2q \times 1$ vector

$$\boldsymbol{\delta} = \begin{bmatrix} \sum_{i=1}^N \boldsymbol{\Delta}_i \mathbf{H}_i^{-1} \nabla_{\boldsymbol{\phi}(i)} \mathcal{L} - \nabla_{\boldsymbol{\theta}} \mathcal{L} \\ \sum_{i=1}^N \mathbf{E}_i \mathbf{H}_i^{-1} \nabla_{\boldsymbol{\phi}(i)} \mathcal{L} - \nabla_{\boldsymbol{\beta}} \mathcal{L} \end{bmatrix}$$

The Newton update direction in the scoring algorithm is given by:

$$\begin{aligned} \tfrac{\nu}{2}\left(\mathbf{F}^{-1}\nabla\mathcal{L}\right)_{\boldsymbol{\phi}(k)} &= \tfrac{\nu}{2}\mathbf{H}_k^{-1}\nabla_{\boldsymbol{\phi}(k)}\mathcal{L} + \\ & \quad \tfrac{\nu}{2}\begin{bmatrix} \mathbf{H}_k^{-1}\boldsymbol{\Delta}_k^T & \mathbf{H}_k^{-1}\mathbf{E}_k^T \end{bmatrix} \boldsymbol{\Psi}^{-1}\boldsymbol{\delta} \\ \tfrac{\nu}{2}\left(\mathbf{F}^{-1}\nabla\mathcal{L}\right)_{\boldsymbol{\theta},\boldsymbol{\beta}} &= -\tfrac{\nu}{2}\boldsymbol{\Psi}^{-1}\boldsymbol{\delta} \end{aligned} \tag{8}$$

Note that although the expressions above seems dependent on ν this is not the case since $\nabla\mathcal{L}$ contains a factor $\frac{2}{\nu}$, which cancels with the leading $\frac{\nu}{2}$, this is very satisfactory as compared to the many gradient based CMA algorithms, in which an arbitrary learning constant appears.

4. THE ALGORITHM

It is well known that optimization methods based on second order derivatives are superior to other methods, and that the Newton algorithm is usually considered as "the standard against which other algorithms are measured" [8]. The basic algorithm is based on the following update scheme: Given an estimate $\boldsymbol{\rho}_k$ of the parameter vector $\boldsymbol{\rho}$ we improve the estimate by

$$\boldsymbol{\rho}_{k+1} = \boldsymbol{\rho}_k - \mathbf{H}^{-1}(\boldsymbol{\rho}_k)\nabla f(\boldsymbol{\rho}_k) \tag{9}$$

where $\mathbf{H}(\boldsymbol{\rho}_k)$ is the Hessian matrix and $\nabla f(\boldsymbol{\rho}_k)$ is the gradient of the target function evaluated at $\boldsymbol{\rho}_k$. In statistical inference when maximizing the likelihood function one usually prefers to replace the Hessian by its expected value, i.e. the Fisher information matrix. This change contributes to the numerical stability of the algorithm (see [9] pp. 177-182). Under this change the algorithm is called a Newton type scoring algorithm. As is well known in the numerical analysis literature although a unit step in the Newton direction

$$\mathbf{F}_N^{-1}(\boldsymbol{\rho}_k)\nabla\mathcal{L}(\boldsymbol{\rho}_k)$$

asures improvement in the quadratic approximation of the the likelihood function, it does not assure improvement in the likelihood function itself. To overcome this difficulty a line search along the Newton direction is devised, either for an optimal step or for suboptimal step. The update step becomes thus

$$\boldsymbol{\rho}_{k+1} = \boldsymbol{\rho}_k - \lambda\mathbf{F}_N^{-1}(\boldsymbol{\rho}_k)\nabla\mathcal{L}(\boldsymbol{\rho}_k) \tag{10}$$

where λ is defined by

$$\lambda = \arg\min_{\lambda} \mathcal{L}(\boldsymbol{\rho}_k - \lambda\mathbf{F}_N^{-1}(\boldsymbol{\rho}_k)\nabla\mathcal{L}(\boldsymbol{\rho}_k)) \tag{11}$$

The optimal λ can be computed very efficiently using the good initialization $\lambda_0 = 1$ which is optimal for the quadratic approximation of the likelihood. For more details about the one-dimensional optimization as well as the possible termination criteria of the Newton algorithm the reader is referred to [8].

Finally we would like to discuss the initialization of the search. For that purpose we propose two alternatives. The first is computationally simple but might fail in hard cases of closely spaced sources while the other is almost optimal and assures convergence to the MLE at the price of further complexity.

The first initialization scheme we propose is by the ESPRIT algorithm. After estimating the DOA's an estimate of the signal using the unconstrained ML estimator, and projecting to the family of CM signals.

$$\hat{\mathbf{s}}'(t) = \left(\mathbf{A}^*(\hat{\boldsymbol{\theta}}_i)\mathbf{A}(\hat{\boldsymbol{\theta}}_i)\right)^{-1}\mathbf{A}^*(\hat{\boldsymbol{\theta}}_i)\mathbf{x}(t) \tag{12}$$

$$\hat{\mathbf{s}}_i(t) = \frac{\hat{s}_i'(t)}{|\hat{s}_i'(t)|} \tag{13}$$

where $\hat{\theta}_i$ is the estimate of the DOA of the i'th signal and $\hat{s}_i'(t)$ is the i'th component of $\hat{\mathbf{s}}(t)$.

The second initialization method is based on the weighted ACMA method [10]. Using the ACMA the CM sources are separated based on their CM property. The DOA's are then estimated by fitting the weight vectors given by the ACMA to the array manifold [6].

To summarize we describe the algorithm using the weighted ACMA initialization:

1. Compute a separating matrix $\hat{\mathbf{W}}$ using the W-ACMA algorithm.
2. Estimate the initial DOA's by fitting the unstructured matrix $\hat{\mathbf{W}}$ to the array manifold. Generate a structured separating matrix $\hat{\mathbf{A}}$ based on the DOA's.
3. Estimate the initial signal estimates using (12)-(13)

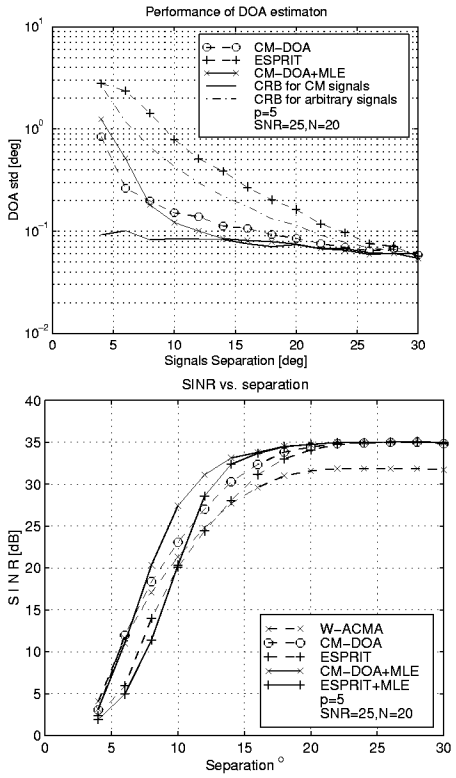

Figure 1: DOA estimation accuracy and SINR vs. separation (a) DOA estimation accuracy vs. separation, (b) SINR vs. separation.

4. Until termination condition is satisfied

   a. Estimate the Newton direction using (8)

   b. Compute λ using (11).

   c. Update the parameters using (10).

End.

A detailed computation of the complexity of the update step yields $(6q^3 + 5pq + 7q^2 + p + 8q + 7)N + q^2(3p + 9q + 13) + 5q$ many operations. The dominant number of operations is $O(6q^3N)$ *real* operations.

5. SIMULATION RESULTS

In this section we describe some simulations demonstrating the efficiency of the proposed MLE method. We present both signal to interference plus noise (SINR) improvement, and DOA performance. The interesting conclusion is that while both initialization methods are good for obtaining the DOA estimation, and yield identical results, the ESPRIT initialization performs worse for separation of the signals, and the phase estimates tends to converge to local minima.

In the experiment we have used a 5 elements ULA, and three sources. The number of samples was held fixed as $N = 20$. We have performed 400 Monte-Carlo trials. The central source was fixed at 0^o while the two other sources were located at $-\Delta^o, \Delta^o$, and Δ was changed from 4^o to 30^o at steps of 2^o. We show that the MLE outperforms both the ESPRIT and the CM-DOA algorithms over the complete range of separations. However at the very small separations the MLE iterations increased the DOA estimation RMSE. This is caused by convergence failure of the iterations in some cases.

Notice that since the DOA estimation achieved the same performance no matter what the initialization was. we present only the initialization with the CM-DOA.

6. REFERENCES

[1] R. Schmidt, *A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation*. PhD thesis, Stanford university, 1981.

[2] I. Ziskind and M. Wax, "Maximum likelihood localization of multiple sources by alternating projections," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 36, pp. 1553–1560, Oct. 1988.

[3] M. Viberg and B. Ottersten, "Sensor array processing based on subspace fitting," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 39, pp. 1110–1121, May 1991.

[4] B. Porat and B. Friedlander, "Direction finding algorithms based on higher order statistics," *IEEE Trans. Signal Processing*, vol. 37, pp. 2016–2024, Sept. 1991.

[5] G. Xu and T. Kailath, "DOA estimation via exploitation of cyclostationarity-A combination of spatial and temporal processing," *IEEE Trans. Signal Processing*, vol. 40, pp. 1775–1785, July 1992.

[6] A. Leshem and A. van der Veen, "Bounds and algorithm for direction finding of phase modulated signals," in *Proc. IEEE SP Workshop on Stat. Signal and Array Proc.*, Sept. 1998.

[7] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood, and Cramer-Rao bound," *IEEE Trans. Acoust. Speech Signal Processing*, vol. 37, pp. 720–743, May 1989.

[8] P. Gill, W. Murray, and M. Wright, *Practical optimization*. Academic Press, 1981.

[9] S. Kay, *Fundumentals of statistical signal processing: Estimation theory*. PTR, Prentice Hall, 1993.

[10] A. van der Veen, "Weighted ACMA." Preprint July 1998. To be submitted to ICASSP 1999.