Correspondence

Maximum Likelihood Separation of Constant Modulus Signals

Amir Leshem

Abstract—In this correspondence, we derive a Newton scoring algorithm for the maximum likelihood separation and direction-of-arrival (DOA) estimation of constant modulus (CM) signals using a sensors array. We present a rapidly converging scheme for the joint estimation of the signals and their directions based on their CM property. We discuss initialization and show that the complexity is moderate.

Index Terms—Constant modulus, DOA estimation, maximum likelihood, source separation.

I. INTRODUCTION

Direction-of-arrival (DOA) estimation of multiple signals impinging on an antenna array is a fundamental problem in signal processing with applications to communication, radar, sonar, radio-astronomy, and many other fields. Super-resolution methods that exploit knowledge of the array manifold or its structure without using information on the signals have been thoroughly investigated. A good overview of these is given in [5]. Other methods exploit properties of the signals such as Gaussianity [2], non-Gaussianity [7], or cyclostationarity [13]. These methods are more robust to array manifold errors due to the extra information they use. In communication applications, the signals are typically structured. One very common property shared by frequency- and phase-modulated signals is the constant envelope or constant modulus (CM). Since the pioneering work of Treichler and Agee [11], it is known that the CM property is a strong property that is already sufficient for source separation. After separation of the signals, the DOA estimation problem is decoupled and can be done for each source individually. Such a scheme is proposed in [9], where the CM signals are sequentially separated using the so-called CM array. Weak points of this and related iterative CM algorithms are their initialization, the recovery of weak signals, and their unpredictable relatively slow convergence, which may require several hundred samples per signal.

Recent studies of the problem of DOA estimation based on the CM property [6] yielded good sub-optimal algorithms. However, a 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 an improvement can be achieved by using the maximum likelihood estimator (MLE). The main problem with using the MLE together with the CM property is the large dimension of the parameter space since unlike the case of arbitrary signals, one cannot perform the maximization of the likelihood under the CM constraint analytically. This makes the estimation very hard, even using iterative numerical methods.

In this correspondence, we derive an exact analytic expression for the inverse of the Fisher information matrix. We further show that the updating of a scoring type Newton algorithm (i.e., an algorithm that uses the expected value of the Hessian instead of the Hessian itself) can be done linearly in the number of samples, rather then cubic, as would be the case in direct numerical inversion. We then devise a scoring algorithm for maximum likelihood source separation based on initialization with a suboptimal method. Finally, we analyze the computational complexity of the algorithm and demonstrate its effectiveness by simulations. This correspondence continues the research presented in [6].

II. DATA MODEL

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

$$\boldsymbol{x}(t) = \boldsymbol{A}\boldsymbol{B}\boldsymbol{s}(t) + \boldsymbol{n}(t) \tag{1}$$

where we have used the following notation: $\boldsymbol{x}(t) = [x_1(t), \cdots, x_p(t)]^T$ is a $p \times 1$ vector of received signals at time t. $\boldsymbol{A} = \boldsymbol{A}(\boldsymbol{\theta}) = [\boldsymbol{a}(\theta_1), \cdots, \boldsymbol{a}(\theta_q)]$, $\boldsymbol{a}(\theta)$ 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. $\boldsymbol{B} = \text{diag}(\boldsymbol{\beta})$ 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 ith signal as received by the array. $\boldsymbol{s}(t) = [s_1(t), \cdots, s_q(t)]^T$ is a $q \times 1$ vector of source signals at time t. $\boldsymbol{n}(t)$ is the $p \times 1$ additive noise vector, which is assumed spatially and temporally white Gaussian distributed with covariance matrix $\nu \boldsymbol{I}$, where $\nu = \sigma^2$ is the noise variance.

The derivation can be extended to spatially colored Gaussian noise with known covariance matrix $\boldsymbol{Q}$ at the expense of greater notational complexity. In our problem, the array is assumed to be calibrated so that the array response vector $\boldsymbol{a}(\theta)$ is a known function. As usual, we require that the array manifold satisfies the uniqueness condition, i.e., every collection of p vectors on the manifold are linearly independent. 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 $\boldsymbol{B}$. 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 $\boldsymbol{X} = [\boldsymbol{x}(1), \cdots, \boldsymbol{x}(N)]$ are available. We assume that the signals are uncorrelated.

III. INFORMATION MATRIX AND THE NEWTON UPDATE FORMULA

In this section, we present the log-likelihood function and the Fisher information matrix. Then, we obtain an analytic expression for the inverse of the information matrix. This extends the derivation of the Cramér–Rao bound on the parameters, as given in [6].

Finally, we give a very simple update scheme for computing the Newton directions. This update formula is the heart of the MLE algorithm presented in Section IV.

The log-likelihood function is given (after omitting some constants) by

$$
\mathcal{L}\left(\boldsymbol{X}|\boldsymbol{\phi}_1,\cdots,\boldsymbol{\phi}_N,\boldsymbol{\theta},\boldsymbol{\beta},\nu\right)
$$
$$
= -pN\log\nu - \frac{1}{\nu}\sum_{k=1}^{N}\left(\boldsymbol{x}(k)-\boldsymbol{ABs}(k)\right)^{*}\left(\boldsymbol{x}(k)-\boldsymbol{ABs}(k)\right). \tag{2}
$$

Following [10], the estimation of the noise variance is decoupled from all other parameters and can be estimated by the MLE of ν given all the other parameters

$$
\hat{\nu} = \frac{1}{pN}\sum_{k=1}^{N}\left(\boldsymbol{x}(k)-\boldsymbol{ABs}(k)\right)^{*}\left(\boldsymbol{x}(k)-\boldsymbol{ABs}(k)\right). \tag{3}
$$

Substituting this into $\mathcal{L}(\boldsymbol{X}|\boldsymbol{\phi}_1,\cdots,\boldsymbol{\phi}_N,\boldsymbol{\theta},\boldsymbol{\beta},\nu)$ and eliminating constant terms, we obtain

$$
\mathcal{L}(\boldsymbol{X}|\boldsymbol{\rho}) = -pN\log
$$
$$
\cdot\left(\sum_{k=1}^{N}\left(\boldsymbol{x}(k)-\boldsymbol{ABs}(k)\right)^{*}\left(\boldsymbol{x}(k)-\boldsymbol{ABs}(k)\right)\right)
$$

where the remaining parameters are collected in the vector $\boldsymbol{\rho} = [\boldsymbol{\phi}(1)^{T},\cdots,\boldsymbol{\phi}(N)^{T},\boldsymbol{\theta}^{T},\boldsymbol{\beta}^{T}]^{T}$. Thus, the maximization of the likelihood can be reduced to minimization of the least squares $\hat{\boldsymbol{\rho}} = \arg\min_{\boldsymbol{\rho}}\sum_{k=1}^{N}\|e(k)\|^{2}$, where

$$
e(k) = \boldsymbol{x}(k)-\boldsymbol{ABs}(k). \tag{4}
$$

Define

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

The Fisher information matrix associated to the estimation of the parameter vector can be derived as (see [6])

$$
\boldsymbol{F}_N = \frac{2}{\nu}\left[\begin{array}{ccc|cc} \boldsymbol{H}_1 & & 0 & \boldsymbol{\Delta}_1^T & \boldsymbol{E}_1^T \\ & \ddots & & \vdots & \vdots \\ 0 & & \boldsymbol{H}_N & \boldsymbol{\Delta}_N^T & \boldsymbol{E}_N^T \\ \hline \boldsymbol{\Delta}_1 & ,\cdots, & \boldsymbol{\Delta}_N & \boldsymbol{\Gamma} & \boldsymbol{\Lambda}^T \\ \boldsymbol{E}_1 & ,\cdots, & \boldsymbol{E}_N & \boldsymbol{\Lambda} & \boldsymbol{\Upsilon} \end{array}\right] \tag{5}
$$

where

$$
\begin{aligned}
\boldsymbol{H}_k &= \mathrm{Re}(\boldsymbol{S}_k^*\boldsymbol{B}^*\boldsymbol{A}^*\boldsymbol{ABS}_k); \\
\boldsymbol{\Delta}_k &= -\mathrm{Im}(\boldsymbol{S}_k^*\boldsymbol{B}^*\boldsymbol{D}^*\boldsymbol{ABS}_k); \\
\boldsymbol{E}_k &= -\mathrm{Im}(\boldsymbol{S}_k^*\boldsymbol{A}^*\boldsymbol{ABS}_k); \\
\boldsymbol{\Gamma} &= \sum_{k=1}^{N}\mathrm{Re}(\boldsymbol{S}_k^*\boldsymbol{B}^*\boldsymbol{D}^*\boldsymbol{DBS}_k); \\
\boldsymbol{\Lambda} &= \sum_{k=1}^{N}\mathrm{Re}(\boldsymbol{S}_k^*\boldsymbol{A}^*\boldsymbol{DBS}_k); \\
\boldsymbol{\Upsilon} &= \sum_{k=1}^{N}\mathrm{Re}(\boldsymbol{S}_k^*\boldsymbol{A}^*\boldsymbol{AS}_k).
\end{aligned}
$$

As usual, the Cramér–Rao bound on the variance of any unbiased estimator is then given by the diagonal elements of the inverse of $\boldsymbol{F}_N$. The bounds on the individual parameters were already obtained in [6]. We assume that the $\boldsymbol{H}_k$ are invertible (an assumption that follows from the independence condition on the array manifold and the independence of the sources). Let

$$
\left[\begin{array}{cc} \boldsymbol{\Xi}_{11} & \boldsymbol{\Xi}_{12} \\ \boldsymbol{\Xi}_{21} & \boldsymbol{\Xi}_{22} \end{array}\right] = \left[\begin{array}{cc} \sum_{k=1}^{N}\boldsymbol{\Delta}_k\boldsymbol{H}_k^{-1}\boldsymbol{\Delta}_k^T & \sum_{k=1}^{N}\boldsymbol{\Delta}_k\boldsymbol{H}_k^{-1}\boldsymbol{E}_k^T \\ \sum_{k=1}^{N}\boldsymbol{E}_k\boldsymbol{H}_k^{-1}\boldsymbol{\Delta}_k^T & \sum_{k=1}^{N}\boldsymbol{E}_k\boldsymbol{H}_k^{-1}\boldsymbol{E}_k^T \end{array}\right] \tag{6}
$$

and define the $q \times q$ matrix

$$
\boldsymbol{\Psi} = \left[\begin{array}{cc} \boldsymbol{\Gamma} & \boldsymbol{\Lambda}^T \\ \boldsymbol{\Lambda} & \boldsymbol{\Upsilon} \end{array}\right] - \left[\begin{array}{cc} \boldsymbol{\Xi}_{11} & \boldsymbol{\Xi}_{12} \\ \boldsymbol{\Xi}_{21} & \boldsymbol{\Xi}_{22} \end{array}\right]. \tag{7}
$$

The CRB for DOA's amplitudes and signal phases was presented in [6] without derivation. We extend the derivation of the CRB and completely invert the information matrix and not only its diagonal elements. This also provides the derivation omitted in [6]. For inverting the matrix $\boldsymbol{F}_N$, we first divide it into four different parts (to simplify notation, we disregard the constant $2/\nu$, remembering that its inverse has to multiply the final result).

$$
\boldsymbol{F}_{11} = \left[\begin{array}{ccc} \boldsymbol{H}_1 & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{H}_N \end{array}\right] \quad \boldsymbol{F}_{12} = \left[\begin{array}{cc} \boldsymbol{\Delta}_1^T & \boldsymbol{E}_1^T \\ \vdots & \vdots \\ \boldsymbol{\Delta}_N^T & \boldsymbol{E}_N^T \end{array}\right]
$$
$$
\boldsymbol{F}_{22} = \left[\begin{array}{cc} \boldsymbol{\Gamma} & \boldsymbol{\Lambda}^T \\ \boldsymbol{\Lambda} & \boldsymbol{\Upsilon} \end{array}\right] \tag{8}
$$

and $\boldsymbol{F}_{21} = \boldsymbol{F}_{12}^T$.

Using Schur's complement formula and (7), we obtain that the $2q \times 2q$ lower right part of $\boldsymbol{F}_N^{-1}$ is now given by

$$
(\boldsymbol{F}_N^{-1})_{22} = \boldsymbol{\Psi}^{-1} = \left(\left[\begin{array}{cc} \boldsymbol{\Gamma} & \boldsymbol{\Lambda}^T \\ \boldsymbol{\Lambda} & \boldsymbol{\Upsilon} \end{array}\right] - \left[\begin{array}{cc} \boldsymbol{\Xi}_{11} & \boldsymbol{\Xi}_{12} \\ \boldsymbol{\Xi}_{21} & \boldsymbol{\Xi}_{22} \end{array}\right]\right)^{-1}. \tag{9}
$$

The other parts of $\boldsymbol{F}^{-1}$ are given by

$$
(\boldsymbol{F}_N^{-1})_{11} = \left[\begin{array}{ccc} \boldsymbol{H}_1^{-1} & & 0 \\ & \ddots & \\ 0 & & \boldsymbol{H}_N^{-1} \end{array}\right] + \left[\begin{array}{cc} \boldsymbol{H}_1^{-1}\boldsymbol{\Delta}_1^T & \boldsymbol{H}_1^{-1}\boldsymbol{E}_1^T \\ \vdots & \vdots \\ \boldsymbol{H}_N^{-1}\boldsymbol{\Delta}_N^T & \boldsymbol{H}_N^{-1}\boldsymbol{E}_N^T \end{array}\right]
$$
$$
\cdot (\boldsymbol{F}_N^{-1})_{22} \left[\begin{array}{ccc} \boldsymbol{\Delta}_1\boldsymbol{H}_1^{-1} & ,\cdots, & \boldsymbol{\Delta}_N\boldsymbol{H}_N^{-1} \\ \boldsymbol{E}_1\boldsymbol{H}_1^{-1} & ,\cdots, & \boldsymbol{E}_N\boldsymbol{H}_N^{-1} \end{array}\right] \tag{10}
$$
$$
(\boldsymbol{F}_N^{-1})_{21} = -(\boldsymbol{F}_N^{-1})_{22}\left[\begin{array}{ccc} \boldsymbol{\Delta}_1\boldsymbol{H}_1^{-1} & ,\cdots, & \boldsymbol{\Delta}_N\boldsymbol{H}_N^{-1} \\ \boldsymbol{E}_1\boldsymbol{H}_1^{-1} & ,\cdots, & \boldsymbol{E}_N\boldsymbol{H}_N^{-1} \end{array}\right] \tag{11}
$$

and again, $(\boldsymbol{F}_N^{-1})_{21} = (\boldsymbol{F}_N^{-1})_{12}^T$. Now that we have explicit expressions for $\boldsymbol{F}_N^{-1}$, we are able to compute explicit expressions for the Newton update direction used in Section IV. To that end, let $\nabla_{\boldsymbol{\phi}(k)}\mathcal{L}$, $\nabla_{\boldsymbol{\theta}}\mathcal{L}$, $\nabla_{\boldsymbol{\beta}}\mathcal{L}$ be the derivatives of $\mathcal{L}$ with respect to the signal phases in the kth sample the signals DOA's and the signals power, respectively. Exact expressions for these are given in [6].

The Newton update direction $\boldsymbol{v}$ is given now by $\boldsymbol{v} = (\nu/2)\boldsymbol{F}_N^{-1}\nabla\mathcal{L}$. This is a $q(N+2) \times 1$ vector function of the parameters. The components of $\boldsymbol{v}$ are given by

$$
(\boldsymbol{v})_{\boldsymbol{\phi}(k)} = \frac{\nu}{2}\boldsymbol{H}_k^{-1}\nabla_{\boldsymbol{\phi}(k)}\mathcal{L} + \frac{\nu}{2}\left[\boldsymbol{H}_k^{-1}\boldsymbol{\Delta}_k^T\ \ \boldsymbol{H}_k^{-1}\boldsymbol{E}_k^T\right]
$$
$$
\cdot (\boldsymbol{F}_N^{-1})_{22}\left[\begin{array}{c} \sum_{i=1}^{N}\boldsymbol{\Delta}_i\boldsymbol{H}_i^{-1}\nabla_{\boldsymbol{\phi}(i)}\mathcal{L} - \nabla_{\boldsymbol{\theta}}\mathcal{L} \\ \sum_{i=1}^{N}\boldsymbol{E}_i\boldsymbol{H}_i^{-1}\nabla_{\boldsymbol{\phi}(i)}\mathcal{L} - \nabla_{\boldsymbol{\beta}}\mathcal{L} \end{array}\right]
$$
$$
(\boldsymbol{v})_{\boldsymbol{\theta},\boldsymbol{\beta}} = \frac{\nu}{2}(\boldsymbol{F}_N^{-1})_{22}\left[\begin{array}{c} \nabla_{\boldsymbol{\theta}}\mathcal{L} - \sum_{i=1}^{N}\boldsymbol{\Delta}_i\boldsymbol{H}_i^{-1}\nabla_{\boldsymbol{\phi}(i)}\mathcal{L} \\ \nabla_{\boldsymbol{\beta}}\mathcal{L} - \sum_{i=1}^{N}\boldsymbol{E}_i\boldsymbol{H}_i^{-1}\nabla_{\boldsymbol{\phi}(i)}\mathcal{L} \end{array}\right] \tag{12}
$$

where $(\boldsymbol{v})_{\boldsymbol{\phi}(k)}$ are the components related to the phase parameters, and $(\boldsymbol{v})_{\boldsymbol{\theta},\boldsymbol{\beta}}$ are the components related to the DOA's and signal power parameters. Note that although the expressions above seems dependent on ν, this is not the case since $\nabla\mathcal{L}$ contains a factor $2/\nu$, which cancels with the leading $\nu/2$. This is very satisfactory as compared with the many gradient-based CMA algorithms, in which an arbitrary learning

constant appears. Note that although we invert the matrices $\boldsymbol{H}_k$, these matrices are small ($q \times q$) and typically have a reasonable condition number, which to a large extent is determined by the power ratio $\boldsymbol{B}$ and the angular separation between the sources $\boldsymbol{A}$.

IV. ALGORITHM

In this section, we describe a method for obtaining maximum likelihood estimates (MLE) of the parameters. The MLE is given by $\hat{\boldsymbol{\rho}} = \arg \max_{\boldsymbol{\rho}} L(\boldsymbol{X}|\boldsymbol{\rho})$, i.e., we would like to choose the parameters that maximize the likelihood of obtaining the observed data. Note that maximizing the likelihood function is equivalent to minimizing $-\mathcal{L}(\boldsymbol{X}|\boldsymbol{\rho})$ defined in Section III. To that end, we devise a Newton-type local optimization algorithm, using the formulas derived in the previous section. Since local optimization methods depends on good initializations, we describe in detail two possible initialization methods.

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 to be "the standard against which other algorithms are measured" [3]. It solves the unconstrained minimization problem $\hat{\boldsymbol{\rho}} = \arg \min_{\boldsymbol{\rho}} f(\boldsymbol{\rho})$, and $f : \boldsymbol{R}^{(N+2)q} \rightarrow \boldsymbol{R}$ is usually referred to as the cost function. The basic algorithm is based on the following update scheme: Given an estimate $\boldsymbol{\rho}_n$ of the parameter vector $\boldsymbol{\rho}$, we improve the estimate by

$$\boldsymbol{\rho}_{n+1} = \boldsymbol{\rho}_n - \boldsymbol{H}^{-1}(\boldsymbol{\rho}_n)\nabla f(\boldsymbol{\rho}_n) \qquad (13)$$

where $\boldsymbol{H}(\boldsymbol{\rho}_n)$ is the Hessian matrix, and $\nabla f(\boldsymbol{\rho}_n)$ is the gradient of the cost function evaluated at $\boldsymbol{\rho}_n$. $\boldsymbol{\rho}_{n+1}$ is obtained by maximizing the quadratic approximation for f at $\boldsymbol{\rho}_n$.

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 [4, pp. 177–182]). Under this change, the algorithm is called a Newton-type scoring algorithm. In our case, the cost function is given by $-\mathcal{L}(\boldsymbol{X}|\boldsymbol{\rho})$, which we will denote from now on as $-\mathcal{L}(\boldsymbol{\rho})$. The expected value of the Hessian is just the information matrix, i.e., $E\boldsymbol{H} = \boldsymbol{F}_N$. Therefore, the update formula becomes

$$\boldsymbol{\rho}_{n+1} = \boldsymbol{\rho}_n - \boldsymbol{F}_N^{-1}(\boldsymbol{\rho}_n)\nabla \mathcal{L}(\boldsymbol{\rho}_n).$$

The Newton approach has another very appealing interpretation as a natural gradient method. Any parametric family of distributions can be considered to be a Riemannian manifold with the metric defined by the Fisher information matrix [1]. Under this metric, the scoring algorithm becomes the stochastic gradient method with respect to the above metric using the natural gradient on the manifold.

As is well known in the numerical analysis literature, although a unit step in the Newton direction

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

assures improvement in the quadratic approximation of 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 a suboptimal step. The update step becomes thus

$$\boldsymbol{\rho}_{n+1} = \boldsymbol{\rho}_n - \lambda \boldsymbol{F}_N^{-1}(\boldsymbol{\rho}_n)\nabla \mathcal{L}(\boldsymbol{\rho}_n) \qquad (14)$$

where λ is defined by [3]

$$\lambda = \arg \min_{\mu} -\mathcal{L}\left(\boldsymbol{\rho}_n - \mu \boldsymbol{F}_N^{-1}(\boldsymbol{\rho}_n)\nabla \mathcal{L}(\boldsymbol{\rho}_n)\right). \qquad (15)$$

The optimal λ can be computed very efficiently by standard one-dimensional (1-D) optimization methods using the good initialization

TABLE I
DESCRIPTION OF THE ALGORITHM

1. Generate a structured separating matrix $\hat{\mathbf{A}} = [\mathbf{a}(\hat{\theta}_1), \dots, \mathbf{a}(\hat{\theta}_q)]$ based on [6].

2. Compute an unstructured signal estimate by $\hat{\mathbf{s}} = \left(\hat{\mathbf{A}}^H \hat{\mathbf{A}}\right)^{-1} \hat{\mathbf{A}}^H \mathbf{x}(t)$.

3. Compute the initial phase estimates using the phases of the unstructured estimate, and estimate the power matrix $\mathbf{B}$ from the power of the unstructured signals.

4. Until termination condition is satisfied

   a. Estimate the Newton direction using (12)

   b. Compute λ using (15).

   c Update the parameters using (14).

5. End.

TABLE II
COMPLEXITY OF COMPUTATION OF THE SUB-MATRICES OF THE INFORMATION MATRIX

matrix	complexity	matrix	complexity	matrix	complexity	matrix	complexity
$\mathbf{A}^\star\mathbf{A}$	pq^2	$\mathbf{A}^\star\mathbf{D}$	pq^2	$\mathbf{D}^\star\mathbf{D}$	pq^2	$\mathbf{BS}_k$	q
$\mathbf{B}^\star\mathbf{A}^\star\mathbf{AB}$	q^2	$(\mathbf{B}^\star\mathbf{A}^\star\mathbf{AB})^{-1}$	q^3	$\mathbf{ABS}_k$	pq	$\mathbf{H}_k$	$2q^2$
$\mathbf{H}_k^{-1}$	q^3	$\boldsymbol{\Delta}_k$	q^2	$\mathbf{E}_k$	q^2	$\boldsymbol{\Gamma}$	$(q^2+1)N$
$\boldsymbol{\Lambda}$	$(q^2+1)N$	$\boldsymbol{\Upsilon}$	$(q^2+1)N$	$\mathbf{H}_k^{-1}\boldsymbol{\Delta}_k^T$	q^3	$\mathbf{H}_k^{-1}\mathbf{E}_k^T$	q^3
$\boldsymbol{\Delta}_k\mathbf{H}_k^{-1}\boldsymbol{\Delta}_k^T$	q^3	$\mathbf{E}_k\mathbf{H}_k^{-1}\mathbf{E}_k^T$	q^3	$\boldsymbol{\Delta}_k\mathbf{H}_k^{-1}\mathbf{E}_k^T$	q^3		

$\lambda_0 = 1$, which is optimal for the quadratic approximation of the likelihood. For more details about the 1-D optimization as well as the possible termination criteria of the Newton algorithm, see [3].

Finally, we would like to discuss the choice of initial parameter estimate. A good initial point is very important to any local optimization scheme to prevent convergence to a local minimum that is not a global minimum. For that purpose, we propose two alternatives. The first is computationally simple but might fail in hard cases of closely spaced sources, whereas the other is almost optimal and assures convergence to the MLE at the price of further complexity. The exposition of the two initialization has been done in [6], where analysis and simulations of the initialization is done. We will shortly describe the methods.

The first initialization scheme we propose is by the ESPRIT algorithm [8] (or any other super-resolution DOA estimation method, which is based the sample covariance matrix without taking the signal structure into account). First, we estimate the DOA's. Then, the signals are estimated using the unconstrained ML estimator. Finally, the signals are projected to the family of CM signals.

The second initialization method is based on the suboptimal algorithm presented in [6]. Using the ACMA, the CM sources are blindly 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. Then, we can compute improved signal estimates using the DOA estimates.

The algorithm using the weighted ACMA initialization is described in Table I.

We now estimate the computational complexity of the algorithm. Since the initialization method can vary and, in each case, has a known complexity, we will concentrate on the complexity of the update step. The details are described in Tables II and III. The overall complexity is given by $(6q^3 + 5pq + 12q^2 + p + 8q + 7)N + q^2(3p + 9q + 13) + 5q$. Finally, we comment that the algorithm usually converges after a few iterations due to the quadratic convergence properties of Newton algorithms. This means that another small multiplicative constant has to be added to the expression above. In the simulation section, we will demonstrate the convergence. Note that the overall complexity is not prohibitive, i.e., not exponential in the problem dimension, and linear in the data length. This puts the algorithm in the class of moderate complexity algorithms.

Expression	defined by equation	complexity
$\mathbf{e}(k), 1 \leq k \leq N$	(4)	$p(q+1)N$
$\nabla \mathcal{L}$	see [6]	$3q(1 + p + pN)$
$\Xi_{11}, \ldots, \Xi_{22}$	(6)	$4N$
$\mathbf{F}_{22}$	(8)	$4q^2$
$(\mathbf{F}_{22})^{-1}$	(9)	$8q^3$
$\sum_{i=1}^{N} \boldsymbol{\Delta}_i \mathbf{H}_i^{-1} \nabla_{\phi(i)} \mathcal{L} - \nabla_\theta \mathcal{L}$		$(q+1)N+q$
$\sum_{i=1}^{N} \mathbf{E}_i \mathbf{H}_i^{-1} \nabla_{\phi(i)} \mathcal{L} - \nabla_\beta \mathcal{L}$		$(q+1)N+q$
The Newton direction for θ, β	(12)	$8q^2$
The Newton direction for each $\phi(k)$	(12)	$5q^2$
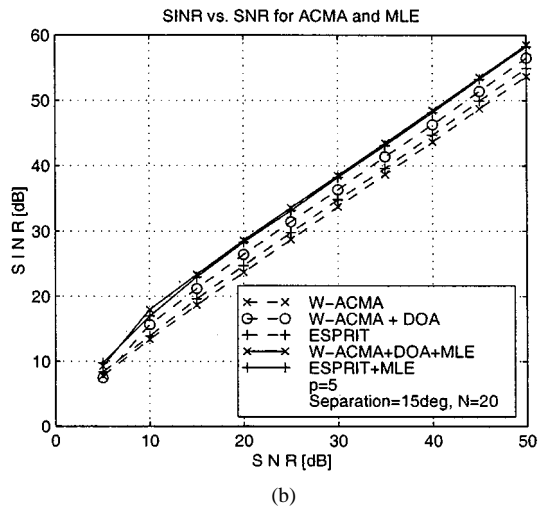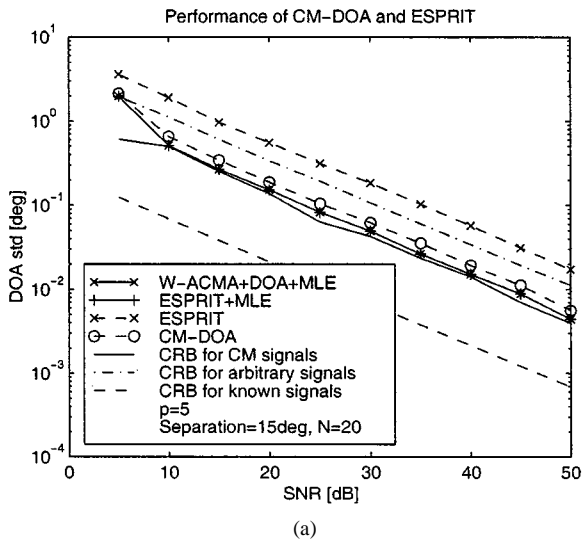
Fig. 1. DOA estimation accuracy and SINR versus SNR. (a) DOA estimation accuracy versus SNR. (b) SINR versus SNR.

Fig. 2. DOA estimation accuracy and SINR versus separation. (a) DOA estimation accuracy versus separation. (b) SINR versus separation.

V. 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 estimation performance.

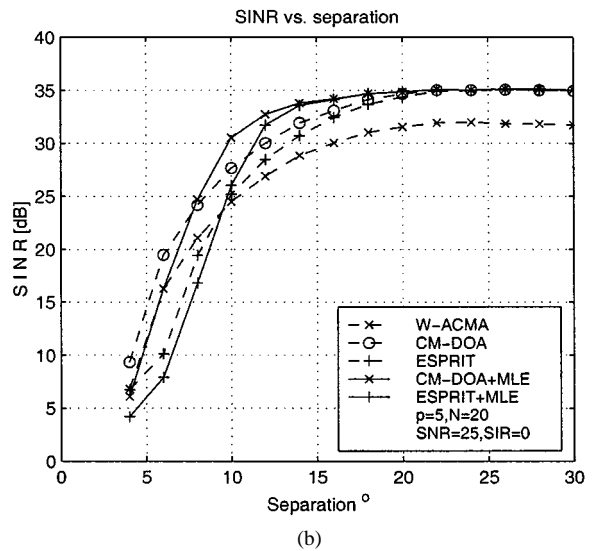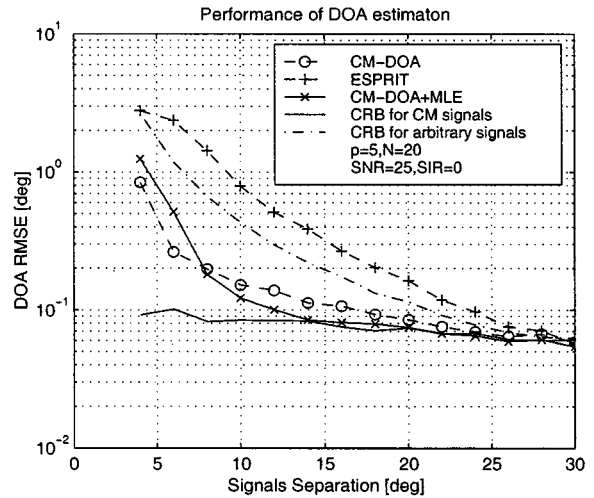In the first two experiments, we have tested the performance as a function of source separation and signal-to-noise ratio (SNR). We have used a five-element ULA and three sources. The number of samples was held fixed as $N = 20$. Each experiment included 100 Monte Carlo trials.

The first experiment tested the dependence on SNR. Three equipowered sources were located at $-15°$, $0°$, $15°$, and the SNR was varied from 5 to 50 dB. We can clearly see the convergence of the DOA estimation to the CRB, as expected from a maximum likelihood estimator.

The second experiment tested the performance as a function of the separation. We have used three equipowered signals. The central source was fixed at $0°$, whereas the two other sources were located at $-\Delta°$, $\Delta°$, and Δ was changed from $4°$ to $30°$ at steps of $2°$. As was demonstrated in [12], in large separation, the ESPRIT algorithm tends to have better performance than the ACMA in terms of SINR. 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.

It is interesting to notice that the DOA estimation achieved the same performance no matter what the initialization was. This is caused by the accuracy of the DOA's, which depends on $1/M$, in contrast to the SINR, which depends only on the array gain, and SNR. This effect is similar to the difference in phase and frequency estimation for sinusoid in noise (see Figs. 1 and 2).
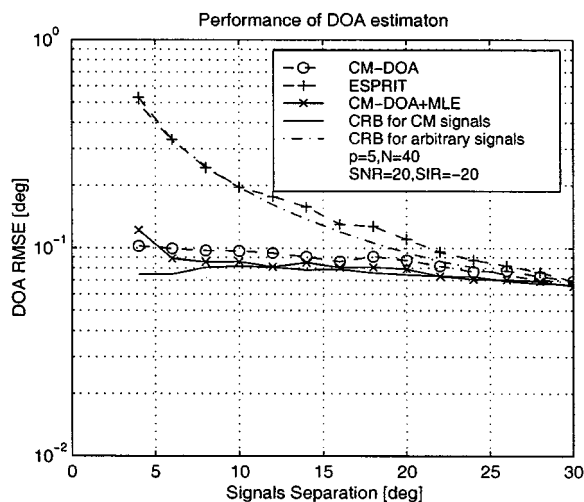
(a)
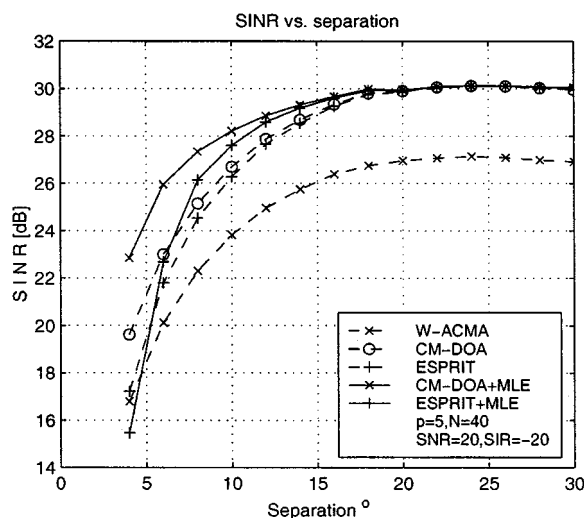
(b)

Fig. 3.   DOA estimation accuracy and SINR versus separation for the weak source. (a) DOA estimation accuracy versus separation. (b) SINR versus separation.

We have performed a third experiment to demonstrate the robustness of the estimator to source power variability. We have located a first source at $0°$ and a second strong source at $\Delta°$, where Δ varied from $4°$ to $30°$. The weak source was 20 dB below the strong source, and the SNR for the weak source was 20 dB. This demonstrates possible near-far robustness of the method. At each separation, we have performed 400 Monte Carlo trials. Fig. 3 presents the RMSE of DOA and the SINR for the weak source. We can clearly see the improved performance, even at separation of $5°$. We see again the importance of exploiting both the array manifold and the CM property.

Finally, we demonstrate the statistics of number of iterations until convergence as function of SNR in the first experiment. The average number of iterations is shown in Fig. 4. We can see that as expected, the better ACMA initialization yielded faster convergence.

VI. CONCLUSIONS

In this correspondence, we have derived an algorithm for the maximum likelihood separation of constant modulus signals. We
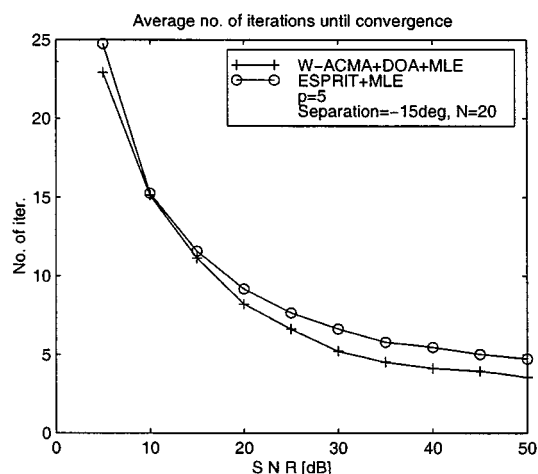

Fig. 4.   Average nunber of iterations versus SNR.

have demonstrated the good performance as well as the reasonable computational complexity of the algorithm. We have also tested various initialization methods. The algorithm yields good results for diverse signal to interference ratios with as little as 20 samples, which is a feature that is not offered by any adaptive CM array algorithm.

An important conclusion is that if one is only interested in the DOA's and not in the signals, then initialization by any method that yields good DOA estimation for the unstructured signals will suffice; however, if one would like to estimate the signals, then the good initialization given by the ACMA gives improved SINR and more robustness to local minima in the high-dimensional parameter space.

REFERENCES

[1] S. Amari, *Differential-Geometrical Methods in Statistics*.   New York: Springer-Verlag, 1985, vol. 28.

[2] J. F. Böhme, "Estimation of spectral parameters of correlated signals in wavefields," *Signal Process.*, pp. 329–337, Oct. 1986.

[3] P. Gill, W. Murray, and M. H. Wright, *Practical Optimization*.   New York: Academic, 1981.

[4] S. M. Kay, *Fundumentals of Statistical signal Processing: Estimation Theory*.   Englewood Cliffs, NJ: Prentice-Hall, 1993.

[5] H. Krim and M. Viberg, "Sensor array signal processing: The parameter estimation approach," *Signal Process.*, June 1995.

[6] A. Leshem and A. J. van der Veen, "Direction-of-arival estimation for constant modulus signals," *IEEE Trans. Signal Processing*, vol. 47, pp. 3125–3129, Nov. 1999.

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

[8] R. Roy, A. Paulraj, and T. Kailath, "ESPRIT—A subspace rotation approach to estimation of parameters of cisoids in noise," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 1340–1342, Oct. 1986.

[9] J. J. Shynk and R. P. Gooch, "The constant modulus array for cochannel signal copy and direction finding," *IEEE Trans. Signal Processing*, vol. 44, pp. 652–660, Mar. 1996.

[10] 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.

[11] J. R. Treichler and B. G. Agee, "A new approach to multipath correction of constant modulus signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 459–471, Apr. 1983.

[12] A. J. van der Veen, "Blind source separation based on combined direction finding and constant modulus properties," in *Proc. IEEE SP Workshop Stat. Signal Array Process.*, Portland, OR, Sept. 1998.

[13] 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.