A MAXIMUM LIKELIHOOD SOLUTION TO DOA ESTIMATION FOR DISCRETE SOURCES

M. Lavielle E. Moulines J.F. Cardoso
Télécom Paris, CNRS URA 829, GDR TdSI
46 rue Barrault, 75634 Paris Cedex 13
email: cardoso, moulines@sig.enst.fr

ABSTRACT

In this contribution, we propose a maximum likelihood solution to the direction-of-arrival estimation for discrete sources (a problem which arises in digital communication context). The likelihood expression being in general very involved, direct solutions or approximations of the likelihood equations are likely to be rather messy. To alleviate this problem, we resort to the standard complete/incomplete data model, where the observations play the role of the incomplete data while the source signals are the missing data. We then maximize the incomplete likelihood (the likelihood of the observations) by iteratively maximizing the complete likelihood function using (i) the deterministic ECM algorithm and (ii) a stochastic version of it, the SEM, which is efficiently implemented by resorting to a Gibbs sampler. Extensive numerical simulations show that this method outperforms the standard higher-order statistics based techniques. Numerical investigation of the Cramer-Rao lower bound is also undertaken.

1. INTRODUCTION

Conventional array signal processing methods for multiple source localization are based on 2nd-order statistics of the received signals. These methods require the explicit knowledge of the noise characteristics. Most approaches assume that the noise covariance is either diagonal or known up to a scaling factor. Unfortunately, this is an unrealistic assumption in many practical applications and often yields poor estimates. Consequently, techniques that account for background noise spatial coloration have attracted great attention (see, among many other contributions, [1] and the references therein).

In the situations where the sources are non-Gaussian (for example, in digital communication context), a solution is to resort to higher-order statistics of the received signals; it is shown, for example, in [2] that fourth-order cumulant techniques can provide direction-of-arrival estimates which are essentially insensitive to the (Gaussian) noise structure (but obviously not to the noise variance).

The basic idea behind higher-order techniques is to exploit the statistical distribution of the sources (assumed to be non-Gaussian). Of course, higher-order methods do not retain all the pertinent statistical information carried by the distribution of the sources. This point is particularly striking when dealing with discrete sources, a situation which occurs in digital communication context.

In this contribution, we derive two algorithms for maximum likelihood DOA estimation for discrete sources in Gaussian noise with unknown covariance (some possible extensions to Non-Gaussian continuous sources are presented). The proposed algorithms share the same common framework of EM-type algorithms, where the observations play the role of incomplete data, whereas the source signals are the missing data.

2. PROBLEM FORMULATION

Assume that we have m narrow-band signals impinging on an array with p sensors (the condition p > m is not required, i.e. we may have more sources than sensors). Under classical assumptions, the array output vector at sampling time-instant t is modelled as

\[ Y(t) = \mathbf{A}(\theta) Z(t) + B(t) \]  

where \( \mathbf{A} = [a(\theta_1), \ldots, a(\theta_m)] \) is the p x m array response matrix, \( Z(t) \) is the m x 1 source signal vector and \( B(t) \) is the p x 1 noise vector. \( \mathbf{A}(\theta) \) is a p x 1 vector representing the response of the array to a signal arriving from direction \( \theta \) (assumed to be scalar and referred to as the direction of arrival). The array manifold \( \mathbf{A} = \{a(\theta), \theta \in \Theta\} \) is assumed to be known (either directly or via some calibration procedure), and to satisfy classical regularities conditions.

The observation noise \( B(t) \) is assumed to be a zero-mean, p-dimensional white Gaussian noise process, and is independent from the source signal (i.e. the random vectors \( B(t) \) and \( Z(t') \) are independent for all \( t,t' \)). Contrary to classical assumptions, its covariance matrix \( \mathbf{EB}(t)B(t)^\mathsf{T} = \Gamma \) is assumed to be unknown.

The source signal vectors \( Z(t) = [Z_1(t), \ldots, Z_m(t)]^\mathsf{T} \) are assumed to be composed of mutually independent complex discrete random variables; signals emitted at different times (by the same source or by different sources) are assumed to be statistically independent (in other words, the random variables \( Z_1(t) \) and \( Z_j(t') \) are independent whenever \( t \neq j \) or \( t \neq t' \)). Each random variable \( Z_i(t) \) takes its value in a finite set of values (or symbols) \( Z_i(a_i) = \{a_i^{z_1}, \ldots, a_i^{z_{N_i}}\} \), where \( z_i^{j} \) are some known complex values and \( a_i \) is an unknown complex amplitude (this data model is related to a linear modulation of a digital source over a Rayleigh channel).

Here, the a priori distribution of the discrete symbols is assumed to be known: it is specified by the vector \( \pi_i = [\pi_{i,1}, \ldots, \pi_{i,N_i}]^\mathsf{T} \), where \( \pi_i = P[Z_i(t) = z_i^{j}] \). Nevertheless, we must notice that the method can be extended for unknown a priori distributions. Based on the observations of T successive samples of the array output vector \( y(t) = [y_1(t), \ldots, y_T(t)] \), the problem consists in estimating (i) the direction-of-arrival \( \theta = [\theta_1, \ldots, \theta_m] \) (ii) the source complex amplitude \( \alpha = [\alpha_1, \ldots, \alpha_m] \) and (iii) the noise covariance matrix \( \Gamma \) (iv) optionally the symbols emitted by each source. For ease of notations, the unknown parameters are collectively referred to as \( \phi = (\theta, \alpha, \Gamma) \).

3. A MAXIMUM-LIKELIHOOD SOLUTION

Due to the intricate nature of the problem and the number of unknown parameters, the direct maximization (un-
where 'tr' stands for trace and C does not depend upon the parameter \( \phi \); the matrices \( R_T, P_T \) and \( Q_T \) are defined according to

\[
R_T = T^{-1} \sum_{t=1}^{T} y(t)y(t)'(5)
\]

\[
P_T = T^{-1} \sum_{t=1}^{T} \varepsilon y(t)'\varepsilon 1(z(t) = \varepsilon)
\]

\[
Q_T = T^{-1} \sum_{t=1}^{T} \varepsilon \varepsilon'\varepsilon 1(z(t) = \varepsilon)
\]

3.2. A deterministic EM algorithm: the ECM approach

**Algorithm**: As outlined above, the ECM algorithm maximizes the incomplete data likelihood by iteratively maximizing the complete likelihood. Each iteration of EM has two steps: an E-step and a CM-step. The (s+1)st E-step finds the conditional expectation of the complete data log-likelihood with respect to the conditional distribution of the missing data given the observations \( y([1:T]) \) and the current estimated parameter \( \phi^{(s)} \),

\[
\phi \leftarrow Q(\phi, \phi^{(s)}) = E[I(X[1:T]; \phi)/Y = y([1:T]); \phi^{(s)}(6)
\]

as a function of the unknown parameters \( \phi \) given the values of the observations \( y([1:T]) \) and the parameters \( \phi^{(s)} \). According to the particular decomposition of the complete log-likelihood function (4), the reestimation function \( Q(\phi, \phi^{(s)}) \) may be expressed as

\[
Q(\phi, \phi^{(s)}) = C - T \log |\Gamma| - T\text{tr}(\Gamma^{-1}R_T)
\]

\[
- T\text{tr}(\Gamma^{-1}2R(\phi^{(s)})d(\alpha)P_T^{(s)}))
\]

\[
+ T\text{tr}(\Gamma^{-1}A(\phi^{(s)})d(\alpha)Q_T^{(s)}A(\phi^{(s)})')
\]

where the matrices \( P_T^{(s)} \) and \( Q_T^{(s)} \) are the conditional expectations of the ‘complete data’ statistics \( P_T \) and \( Q_T \) given the observations \( y([1:T]) \) and the current values of the parameters \( \phi^{(s)} \). Since \( E[I(1(Z(t) = \varepsilon)|Y(t) = y(t); \phi^{(s)}) = p(\varepsilon|Y(t) = y(t); \phi^{(s)}) \) is equal to the posterior distribution of the random variable \( \varepsilon \) given \( y(t) \) and \( \phi^{(s)} \), i.e. \( E[I(1(Z(t) = \varepsilon)|Y(t) = y(t); \phi^{(s)}) = p(\varepsilon|Y(t) = y(t); \phi^{(s)}) \), it is easily seen that

\[
P_T^{(s)} = T^{-1} \sum_{t=1}^{T} \sum_{\varepsilon \in \mathcal{Z}} y(t)'\varepsilon p(\varepsilon|Y(t) = y(t); \phi^{(s)})
\]

\[
Q_T^{(s)} = T^{-1} \sum_{t=1}^{T} \sum_{\varepsilon \in \mathcal{Z}} \varepsilon \varepsilon'\varepsilon p(\varepsilon|Y(t) = y(t); \phi^{(s)})
\]

The (s+1)st CM-step (conditional maximization) then finds \( \phi^{(s+1)} \) by conditionally maximizing \( \phi \rightarrow Q(\phi, \phi^{(s)}) \) over (i) the source complex amplitude (the DOAs and the noise covariance being fixed) (ii) the DOAs (the source complex amplitudes and the noise covariance being fixed) and (iii) the noise covariance (the source complex amplitudes and the noise covariance being fixed). The maximization steps involve the following equations:

\[
\alpha^{(s+1)} = [(A(\phi^{(s)})')^{-1} A(\phi^{(s)}) \circ (Q_T^{(s)})^{-1}]^{-1} \times (8)
\]

\[
\Gamma^{(s+1)} = R_T - 2R(A(\phi^{(s)})P_T^{(s)}) + A(\phi^{(s)})Q_T^{(s)}A(\phi^{(s)})'
\]
where the $A^*$ denotes the complex-conjugate of the matrix $A$ and $\otimes$ is the Schur (component-by-component) product. The parameters $\theta^{(s+1)}$ are obtained by explicit maximization over $\theta$, the other parameters being fixed. This can be achieved by using for example a Newton-Raphson technique, or relaxation-type optimization.

**convergence:** As all EM like algorithm, the proposed algorithm may be shown to converge, under appropriate conditions to a stationary point of the incomplete likelihood (the convergence rate is linear). In almost all cases, the limiting value will occur at a local, if not global maximum of $g(y_1^T), \phi$; provided appropriate initialization is undertaken, the algorithm converge, generally in very few iterations, say between 5 to 10; the proposed method is thus applicable in real situations. However, ill-convergence may appear, especially when the initial point is far from the true values of the parameters.

### 3.3. A Stochastic version of the EM algorithm: the SEM algorithm

In certain circumstances (large number of sources / large constellations), the expectation stage of the ECM algorithm is computationally expensive, impairing the usefulness of the algorithm for real-life applications (for example, for 3 16-QAM signals, there are $16^3 = 4096$ different configurations for the source symbols at each sampling time-instant $t$; one iteration of the ECM would require the computation of $4096 \times 7$ a posteriori probability $p(z|Y = y(t), \phi^{(s)}$ which is far beyond the performance of currently available hardware). In these cases (see also the next section for other applications), a computationally more efficient stochastic version of the EM algorithm can be implemented. The basic idea behind this technique consists in substituting the computationally involved expectation stage by a much faster stochastic simulation of the missing data.

As the ECM algorithm, the SEM is an iterative procedure that goes as follows; at iteration $s$:

1. **i)** Draw a single realization of the missing data $\hat{z}^{(s)} = \{z_1^{(s)}, \ldots, z_3^{(s)}\}$ (the source symbols) under the a posteriori distribution $p(\hat{z}|Y = y(1, T), \phi^{(s)}$.
2. **ii)** Compute the statistics $P_T(\hat{z}^{(s)}, y(1 : T))$ and $Q_T(\hat{z}^{(s)}, y(1 : T))$ for the current realization of the missing data according to (8):

$$
P_T(\hat{z}^{(s)}, y(1 : T)) = T^{-1} \sum_{t=1}^{T} z_1^{(s)} y(t)^H (9)$$

$$
Q_T(\hat{z}^{(s)}, y(1 : T)) = T^{-1} \sum_{t=1}^{T} z_3^{(s)} (\hat{z}^{(s)})^H (10)
$$

3. **iii)** Update the values of $P_T$ and $Q_T$ using the following stochastic averaging technique

$$
P_T^{(s+1)} = P_T^{(s)} + \gamma_s \left(P(\hat{z}^{(s)}, y) - P_T^{(s)}\right)$$

$$
Q_T^{(s+1)} = Q_T^{(s)} + \gamma_s \left(Q(\hat{z}^{(s)}, y) - Q_T^{(s)}\right)
$$

where $\gamma$ is a decreasing sequence of positive step sizes such that $\sum \gamma_s = \infty$ and $\sum \gamma_s^2 < \infty$.

4. **iv)** Update the value of the unknown parameters $\phi$ by conditional-maximization of the complete likelihood $r(R_T, P_T^{(s+1)}, Q_T^{(s+1)}, \phi)$ over $\phi$, (4).

Note that the last step (maximization over the unknown parameters) is equivalent to the maximization step in the ECM algorithm. The simulation of the joint distribution $p(z|Y = y(1 : T), \phi^{(s)})$ can easily be implemented by resorting to the computationally efficient Gibbs sampling technique (see, for example [5]).

**convergence:** The convergence of the stochastic EM can be studied using results drawn from the stochastic approximation theory; in particular, it can be shown that:

- The stationary points of the ordinary differential equation (ODE) associated with the stochastic approximation procedure (9) (10) are the stationary points of the (incomplete) likelihood function $g$.
- The asymptotically stable points of the ODE are the (possibly local) maxima of the incomplete likelihood function $g(y(1 : T), \phi)$

As for the ECM, convergence to the global maximum of the incomplete likelihood function $\phi$ occurs provided the initial values stand ‘not too far’ from the true values of the parameters. Practical evidence demonstrates that the dependance of the SEM solution on the initial guess is weaker than for the ECM. In fact, the stochastic behavior of the algorithm avoids ill-convergence on saddle points or a spurious local maxima of the likelihood.

### 3.4. Some possible extensions

The use of the ECM algorithm is restricted to discrete sources: the expectation stage (6) cannot be carried out for continuous sources (by continuous sources, we mean that the probability distribution of the sources $z_s(t)$ have a density with respect to the Lebesgue measure on the complex plane); it requires the evaluation of multi-dimensional integrals, a rather involved numerical task. On the contrary, the SEM algorithm can be applied to continuous sources: the expectation stage is, in some sense, replaced by a Monte-Carlo integration, which is implicitly performed in the stochastic iterations.

These interesting features of the SEM algorithm will be developed in a forthcoming publication.

### 4. SIMULATIONS

This section contains some numerical results illustrating the performance of the algorithms discussed in the previous sections.

We use the following parameters. Number of sources: $m = 2$; number of sensors: $p = 4$; array configuration: linear; element spacing: $\lambda/2$ (where $\lambda$ is the wavelength); DOA’s: $\theta_1 = 10$ deg and $\theta_2 = 15$ deg; signal type: quaternary QAM; signal amplitudes: both have unit amplitude; noise model: Gaussian, with correlation between sensors $k$ and $l$ equal to $\rho^{k-l}$ (spatially autoregressive noise); noise amplitude $\sigma$ at each sensor; number of measurements: $T = 500$ (see [2] for details).

We ran six cases of the SEM algorithm with 100 independent Monte-Carlo runs for each case. The six cases correspond to $\sigma = \sqrt{0.1}$ and $\sigma = 1$, $\rho = 0$ (i.e., spatially uncorrelated noise), $\rho = 0.5$ and $\rho = 0.99$. Table 1 shows the means and the standard deviations (in degree) of the estimates obtained in these situations. For reference, the values of these quantities for DOAs estimation using the standard (second order) MUSIC algorithm and the optimal 4-th order MUSIC (MUSIC-4) are given in table 2 (these values are taken from [2], table 1).

The inspection of the tables demonstrate that the SEM algorithm outperforms 2nd order and 4-th order identification techniques. The SEM DOAs estimators are virtually
unbiased; the standard deviation of the estimates even decreases when the spatial correlation coefficient is increased. In all cases, the SEM DOA estimates show significantly smaller mean-square errors (MSE) than MUSIC-4 DOA estimates. The comparison with 2nd-order MUSIC is somewhat unfair for $\rho \neq 0$, because the DOA estimator is severely biased in this context; however, one should notice that the SEM outperforms 2nd-order MUSIC, even for $\rho = 0$. The level of performance achieved by the SEM algorithm is even more striking at low SNR: by comparing Table 1 and Table 2, it appears that the SEM DOA’s MSE at $\sigma = 1.0$ are comparable to MUSIC-4 DOA MSE at $\sigma = \sqrt{0.1}$. The SEM algorithm achieves an impressive 10 dB improvement over the optimal 4th-order method.

5. CONCLUSION
In this contribution, two maximum likelihood solutions to DOA estimation for discrete sources have been presented. These algorithms belong to the general class of EM algorithms, where the array output (measures) play the role of the incomplete data and the signal emitted by the sources (unobserved) are the missing data. Numerical simulations evidence the performance of these approaches in circumstances (low SNR, unknown noise spatial correlation) where traditional methods (MUSIC, Gaussian-ML) fail.

REFERENCES

Figure 1. An EM trajectory: 2 quaternary QAM signals. 500 samples. Noise amplitude: 1 at each sensor

Figure 2. A SEM trajectory: 2 quaternary QAM signals. 500 samples. Noise amplitude: 1 at each sensor

<table>
<thead>
<tr>
<th>noise ampl. $\sqrt{0.1}$</th>
<th>noise ampl. 1</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho = 0$ Mean</td>
<td>$\theta_1$</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.036</td>
</tr>
<tr>
<td>MSE</td>
<td>0.036</td>
</tr>
<tr>
<td>$\rho = 0.5$ Mean</td>
<td>10.000</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.033</td>
</tr>
<tr>
<td>MSE</td>
<td>0.033</td>
</tr>
<tr>
<td>$\rho = 0.99$ Mean</td>
<td>10.001</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.009</td>
</tr>
<tr>
<td>MSE</td>
<td>0.009</td>
</tr>
</tbody>
</table>

Table 1. SEM algorithm: Mean and standard deviation of the DOA estimates. $(\theta_1, \theta_2) = (10\, \text{deg}, 15\, \text{deg})$. Spatial autoregressive noise with parameter $\rho$.

<table>
<thead>
<tr>
<th>2nd-order MUSIC</th>
<th>4th-order MUSIC</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\rho = 0$ Mean</td>
<td>$\theta_1$</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.35</td>
</tr>
<tr>
<td>MSE</td>
<td>0.41</td>
</tr>
<tr>
<td>$\rho = 0.5$ Mean</td>
<td>9.45</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.30</td>
</tr>
<tr>
<td>MSE</td>
<td>1.58</td>
</tr>
<tr>
<td>$\rho = 0.99$ Mean</td>
<td>7.98</td>
</tr>
<tr>
<td>Std. Dev.</td>
<td>0.14</td>
</tr>
<tr>
<td>MSE</td>
<td>2.02</td>
</tr>
</tbody>
</table>

Table 2. 2nd-order and 4th-order MUSIC: Mean and standard deviation of the DOA estimates. $(\theta_1, \theta_2) = (10\, \text{deg}, 15\, \text{deg})$. Spatial autoregressive noise with parameter $\rho$. Noise amplitude: $\sqrt{0.1}$ at each sensor.