

# Correspondence

## Wideband Maximum Likelihood Direction Finding and Signal Parameter Estimation by Using the Tree-Structured EM Algorithm

Nail Çadallı and Orhan Arıkan

**Abstract**—The expectation maximization (EM) algorithm is presented for the case of estimating direction of arrivals of unknown deterministic wideband signals. Alternative regularized least squares estimation techniques for the required signal estimation and a tree structure for the data mapping in the EM algorithm are proposed. Extensive simulation results are presented for comparison of the proposed algorithms with the conventional EM approach and the current high-resolution methods of wideband direction finding.

**Index Terms**—Array signal processing, EM algorithm, maximum likelihood estimation, signal parameter estimation, source localization, tree-structured EM algorithm, wideband direction finding.

### I. INTRODUCTION

The problem of direction-of-arrival estimation has been studied extensively both for the narrowband and wideband cases. Although the maximum likelihood (ML) estimates are the most preferable, due to its higher computational cost, ML approach has not found much use in practice. However, the superposition property of the data acquisition system can be exploited by using the expectation maximization (EM) algorithm in order to reduce greatly the complexity of the ML estimation [1]. The derivation of the EM algorithm for the direction finding problem in the narrowband case is available [2], and it is also applied to wideband signals [3].

In EM formalism, the observation, *incomplete data* is obtained via a many-to-one mapping from the *complete data* space that includes signals that we would obtain as the sensor outputs if we were able to observe the effect of each source separately. The EM algorithm iterates between estimating the likelihood of the complete data using the incomplete data and the current parameter estimates (E-step) and maximizing the estimated log-likelihood function to obtain the updated parameter estimates (M-step). Under mild regularity conditions, the iterations of the EM algorithm converge to a stationary point of the observed log-likelihood function, where at each iteration, the likelihood of the estimated parameters is increased [4], [5].

In the present study, for the estimation of unknown signals arriving from different directions to a passive array, alternative regularized estimation schemes to the common least squares solution are investigated. It has been demonstrated that when regularized methods are used in the estimation of the received signals, the EM algorithm has better convergence behavior. In addition, motivated with the ideas of two former extensions of the EM algorithm, a tree-structured hierarchy is used for the description of the relation between the

complete data space and the observations [6], [7]. It is shown that even for a moderate number of signals, the performance increases with the use of the tree-structured EM algorithm.

### II. THE DATA MODEL AND ML ESTIMATION VIA THE EM ALGORITHM

For the case of  $M$  wideband signal wavefronts from the directions  $\theta_1, \theta_2, \dots, \theta_M$  incident onto an array of  $P$  sensors, the output of the  $i$ 'th sensor, which is sampled at  $N$  sampling points with a sampling interval of  $T_s$ , can be written as

$$y_i(t) = \sum_{l=1}^M a_i(t, \theta_l) * s_l(t - \tau_i(\theta_l)) + u_i(t) \quad 1 \leq i \leq P, \quad t = 0, T_s, \dots, (N-1)T_s \quad (1)$$

where  $a_i(t, \theta_l)$  represents the frequency dependent directional sensor gain known *a priori*. The signal emitted from the direction  $\theta_l$  is denoted as  $s_l(t)$ . The noise at the  $i$ th sensor is  $u_i(t)$ , which is zero mean, spatially, and temporally white circularly symmetric complex Gaussian noise, and  $\tau_i(\theta_l)$  is the time delay (with respect to the phase center of the array) of the signal with the direction of arrival equal to  $\theta_l$ . Source signals are assumed to be nonparametric and unknown but deterministic wideband signals. When a prior parametric model of the source signals is available, this additional information can readily be utilized within the proposed framework with more reliable direction-of-arrival estimates. The number of the arriving signals  $M$  is assumed to be known, that is, the detection phase is performed accurately with the use of standard algorithms, which are reviewed in [8]. The above signal model assumes the absence of near-field sources since the added complexity of the data model for arbitrarily located sources only makes the presentation of the ideas more difficult. In this case, the impinging individual wavefields of the far-field sources have the same direction of arrival at the sensors, and the curvature effect of nearfield sources is avoided. The corresponding sensor output in the frequency domain can be closely approximated by using DFT, yielding

$$Y_i(k) = \sum_{l=1}^M A_i(k, \theta_l) e^{-j \frac{2\pi k \tau_i(\theta_l)}{F T_s}} S_l(k) + U_i(k) \quad 1 \leq i \leq P, \quad 0 \leq k < F \quad (2)$$

where  $Y_i(k)$ ,  $A_i(k, \theta_l)$ ,  $S_l(k)$ , and  $U_i(k)$  are the  $F$ -point discrete Fourier transformations (DFT) of  $y_i(t)$ ,  $a_i(t, \theta_l)$ ,  $s_l(t)$ , and  $u_i(t)$ , respectively. Since DFT is a unitary transformation, the transformed noise is still a spatially and temporally white complex Gaussian noise sequence. The following definitions are used to simplify the representation:

- $\Theta = [\theta_1 \ \theta_2 \ \dots \ \theta_M]^T$ ;
- $\mathbf{b}(k, \theta) = [A_1(k, \theta) e^{-j \frac{2\pi k \tau_1(\theta)}{F T_s}} \ \dots \ A_P(k, \theta) e^{-j \frac{2\pi k \tau_P(\theta)}{F T_s}}]^T$ ;
- $\mathbf{B}(k, \Theta) = [\mathbf{b}(k, \theta_1) \ \dots \ \mathbf{b}(k, \theta_M)]$ ;
- $\mathbf{S}(k) = [S_1(k) \ \dots \ S_M(k)]^T$ ;
- $\mathbf{Y}(k) = [Y_1(k) \ \dots \ Y_P(k)]^T$ ;
- $\mathbf{U}(k) = [U_1(k) \ \dots \ U_P(k)]^T$ ;
- $\mathcal{B}(\Theta) = \text{diag}\{\mathbf{B}(0, \Theta) \ \dots \ \mathbf{B}(F-1, \Theta)\}$ ;
- $\mathcal{S} = [\mathbf{S}^T(0) \ \dots \ \mathbf{S}^T(F-1)]^T$ ;
- $\mathcal{Y} = [\mathbf{Y}^T(0) \ \dots \ \mathbf{Y}^T(F-1)]^T$ ;
- $\mathcal{U} = [\mathbf{U}^T(0) \ \dots \ \mathbf{U}^T(F-1)]^T$ .

Manuscript received August 11, 1996; revised June 9, 1998. The associate editor coordinating the review of this paper and approving it for publication was Dr. Petar M. Djuric.

N. Çadallı is with the Coordinated Science Laboratory and Department of Electrical and Computer Engineering, University of Illinois, Urbana, IL 61801 USA (e-mail: cadalli@csl.uiuc.edu).

O. Arıkan is with the Department of Electrical and Electronics Engineering, Bilkent University, Ankara, Turkey (e-mail: oarikan@ee.bilkent.edu.tr).

Publisher Item Identifier S 1053-587X(99)00169-5.

Using these definitions, (2) can be written as  $\mathcal{Y} = \mathcal{B}(\Theta)\mathcal{S} + \mathcal{U}$  or, equivalently, as

$$\mathbf{Y}(k) = \mathbf{B}(k, \Theta)\mathbf{S}(k) + \mathbf{U}(k) \quad 0 \leq k < F. \quad (3)$$

This final compact form of the measurement relation, which is the same as the signal model of the Cramér–Rao lower bound formula in [9], is used in our derivations.

The ML estimates for  $\Theta$  and  $\mathbf{S}$  are obtained by a joint maximization of the likelihood function of the observations. Direct solution of this maximization problem by using numerical search methods is not only computationally demanding, but in addition, due to the usually complicated local maxima structure of the likelihood function, it is not guaranteed to converge to the global maxima. The EM method of obtaining the ML estimates has been proposed to overcome the difficulty by an either parallel or sequential iterative search in much lower dimensional parameter spaces [1]. In our application, the most commonly used complete data specification is  $\mathbf{X}_l(k) = [X_{1l}(k) \cdots X_{Pl}(k)]^T$ , which is the spectrum of the signal that would be observed at the sensors if we were able to observe the effect of  $l$ th source separately. The mean of the complete data  $\mathbf{X}_l(k)$  is  $\mathbf{b}(k, \theta_l)S_l(k)$ , and its statistics are determined by the additive noise. Then, the many-to-one mapping for all sources from the complete data space to the incomplete data space can be written as  $\mathbf{Y}(k) = \sum_{l=1}^M \mathbf{X}_l(k)$  for  $0 \leq k < F$ . By means of such a mapping, the observed signal is decomposed into  $M$  constituents; hence, in the estimation of  $\theta_l$  and  $S_l(k)$ , only  $\mathbf{X}_l(k)$  is used along with the observations. At the  $n$ th iteration of the EM algorithm, the iterative update formulas are [10]

$$\text{E-step: } \mathbf{X}_l^n(k) = \mathbf{b}(k, \theta_l^n)S_l^n(k) + \frac{1}{M}[\mathbf{Y}(k) - \mathbf{B}(k, \Theta^n)\mathbf{S}^n(k)] \quad (4)$$

$$\text{M-step: } \theta_l^{n+1} = \arg \max_{\theta} \sum_{k=0}^{F-1} \frac{\mathbf{b}^\dagger(k, \theta) \mathbf{X}_l^n(k) \mathbf{X}_l^{n\dagger}(k) \mathbf{b}(k, \theta)}{\|\mathbf{b}(k, \theta)\|^2} \quad (5)$$

$$S_l^{n+1}(k) = \frac{\mathbf{b}^\dagger(k, \theta_l^{n+1}) \mathbf{X}_l^n(k)}{\|\mathbf{b}(k, \theta_l^{n+1})\|^2}. \quad (6)$$

In the above algorithm, direction-of-arrival estimation part of the maximization step is (5), whereas (6) is the signal estimation stage that uses the least squares solution, which is to be called LS-EM from now on. In the maximization step, the direction-of-arrival and signal estimation phases can be performed either one after the other for each  $l$  separately, as in the conventional EM algorithm, or the signal estimation stage can be performed after the direction-of-arrival estimation is completed for all  $l$ . Actually, for the latter case,  $\Theta^{n+1}$  is available after (5) and can be inserted into (3). Then, we can solve for  $\mathbf{S}^{n+1}(k)$  by using a number of alternatives such as the least squares (LS) solution, which can be performed to estimate all  $M$  signal waveforms at once by using all currently updated direction estimates, giving

$$\hat{\mathbf{S}}(k) = [\mathbf{B}^\dagger(k, \Theta)\mathbf{B}(k, \Theta)]^{-1}\mathbf{B}^\dagger(k, \Theta)\mathbf{Y}(k). \quad (7)$$

This LS estimation has been proposed as a generalization of the EM algorithm to speed up convergence [2]. However, since the array manifold matrix  $\mathbf{B}(k, \Theta)$  is used instead of the steering vector  $\mathbf{b}(k, \theta)$ , the required inversion in (7) may cause numerical instability problems during the iterations of the algorithm, especially for the case of sources with small separation. One way to avoid this is the use of more robust regularized least squares (RGLS) estimate

$$\hat{\mathbf{S}}(k) = [\mathbf{B}^\dagger(k, \Theta)\mathbf{B}(k, \Theta) + \mu\mathbf{I}]^{-1}\mathbf{B}^\dagger(k, \Theta)\mathbf{Y}(k). \quad (8)$$

It is important to properly choose the regularization parameter  $\mu$ . Some commonly used methods for this purpose have been investigated in [11] and [12]. In our implementations, two adaptive strategies for the choice of  $\mu$  are employed. The first one, which is referred to as RGLS-1, works with the known noise variance, whereas the other one, known as RGLS-2, can be used when the noise variance is unknown [13], [14].

An alternative to the adaptively regularized least squares estimates for source signals is the following, which is referred to as LS-SET solution:

$$\begin{aligned} \hat{\mathbf{S}}(k) &= \arg \min_{\mathbf{S}(k)} \int_{\mathcal{Z}} \|\mathbf{Y}(k) - \mathbf{B}(k, \Theta)\mathbf{S}(k)\|^2 d\Theta \\ &= \left[ \int_{\mathcal{Z}} \mathbf{B}(k, \Theta)^\dagger \mathbf{B}(k, \Theta) d\Theta \right]^{-1} \left[ \int_{\mathcal{Z}} \mathbf{B}(k, \Theta)^\dagger d\Theta \right] \mathbf{Y}(k) \end{aligned} \quad (9)$$

where  $\mathcal{Z}$  is a set of directions in a neighborhood of  $\Theta$ . Since, in the cost function of LS-SET, we use an average penalty in the neighborhood of the estimated direction of arrivals, the LS-SET solution provides signal estimates that is robust to the inaccuracies in the direction-of-arrival estimates. For the case of discrete neighborhood directions, the integrals reduce to summations over an  $M$ -dimensional grid of  $L$  directions in each dimension. The number of required nested summations over an  $M$ -dimensional grid in the above expression increases with the dimension of the direction vector  $\Theta$ . Thus, the computational complexity increases exponentially with  $M$  being in the order of  $L^M$ . Therefore, practically, this alternative, if used as above, is not preferable for large grid sizes and large number of superposed signals. However, since (9) is solely an alternative for the least squares solution, it can also be used instead of the LS-EM solution, performing the estimation in each complete data space separately as

$$\begin{aligned} \hat{S}_l(k) &= \arg \min_{S_l(k)} \int_{\mathcal{Z}_l} \|\mathbf{X}_l(k) - \mathbf{b}(k, \theta_l)S_l(k)\|^2 d\theta_l \\ &= \left[ \int_{\mathcal{Z}_l} \mathbf{b}^\dagger(k, \theta_l) \mathbf{b}(k, \theta_l) d\theta_l \right]^{-1} \left[ \int_{\mathcal{Z}_l} \mathbf{b}^\dagger(k, \theta_l) d\theta_l \right] \mathbf{X}_l(k). \end{aligned} \quad (10)$$

This solution is named LS-RSET, as an abbreviation for LS on a reduced set, since the size of the set is much smaller. In that case, the summation terms are not nested inside each other; hence, the computational load increases only linearly with  $M$  being in the order of  $L \times M$ . The EM algorithm with the above alternatives for the signal estimation stage is shown in Table I, and the computational issues are discussed in [10].

### III. TREE-STRUCTURED EM ALGORITHM

In this section, we propose to use a multilevel tree structured mapping between incomplete and complete data spaces rather than the commonly used data set up for the EM algorithm. In this way, we aim to bring together the superior features of two former extensions of the EM algorithm, namely, the cascade EM (CEM) algorithm and the space alternating generalized EM (SAGE) algorithm [6], [7]. In CEM, an intermediate data specification between the complete and the incomplete data of the conventional EM method is used, and intermediate EM steps at some iterations are performed. As a result, faster convergence with fewer computations per iteration is achieved when compared with the conventional EM algorithm. In SAGE, the parameters are sequentially updated by alternating between several hidden data spaces, unlike the EM algorithm, where the parameters are updated simultaneously. Moreover, the complete data spaces are organized such that the maximization step of the

TABLE I  
STEPS OF THE EM ALGORITHM WITH ALTERNATIVE SIGNAL ESTIMATION METHODS

Initialization

1. : Set  $n = 0$ .
2. : Obtain a rough estimate for  $\Theta^0$ .
3. : Using  $\Theta^0$ , solve (3) by LS, RGLS or LS-SET for  $\mathbf{S}^0(k)$

E-Step

4. : Decompose the observation to complete data by using (4)

M-Step

5. : Update  $\Theta^n$  by using (5) for all  $l$ . If LS-EM or LS-RSET is to be used in Step 6,  $\theta_l^n$  can be updated only.
6. : Update  $\mathbf{S}^n(k)$  if LS, RGLS or LS-SET is used. Update  $\mathbf{S}_l^n(k)$  if LS-EM or LS-RSET is used. Do the latter for each  $l$  if  $\Theta^n$  is updated in Step 5, otherwise goto Step 5 for the update of  $\theta_{l+1}^n$ .

Loop

7. : Increment  $n$  by 1 and continue from Step 4.

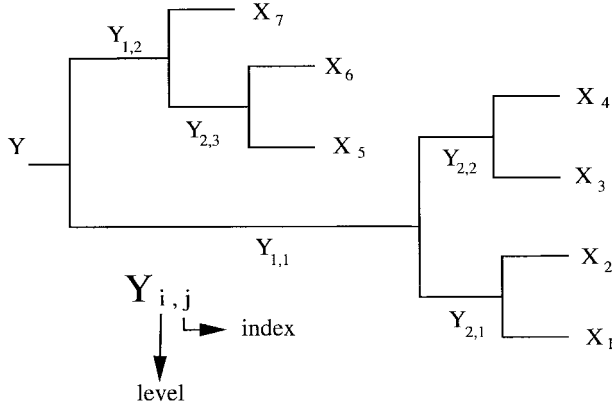


Fig. 1. Binary tree structure for the example case of seven sources.

EM algorithm is performed in less informative data spaces providing faster convergence. The sequential maximization of the expected likelihood function in each hidden data space has been reported to be the main factor of the superior performance of SAGE compared with the conventional EM algorithm.

We propose to use a multilevel data hierarchy as shown in Fig. 1 for the example case of seven sources. The leaves of the tree host the complete data  $\mathbf{X}_l(k)$ 's, and the root of the tree denotes the observation  $\mathbf{Y}(k)$ . The intermediate nodes correspond to the partial conditional incomplete data  $\mathbf{Y}_{i,j}(k)$ , which are also updated during the iterations. It should be noted that the intermediate data at a particular node is not simply obtained by summing the complete data of the leaves. The relevant branching indicates the relationship of the complete and incomplete data of the EM algorithm. The associated precise data assignment for the intermediate incomplete data is explained below.

In this setting, the EM algorithm can be run for two sources at a time using the incomplete data at the joint node of two leaves, which is obtained by using the intermediate data at the lower<sup>1</sup> branch node

<sup>1</sup> Just like in real trees, lower branch or lower level refers to a branch or level closer to the root of the tree.

and complete data, which is not to be updated by the current run. For instance, to run the EM algorithm for  $\mathbf{X}_1(k)$  and  $\mathbf{X}_2(k)$ , we form the required incomplete data as

$$\mathbf{Y}_{1,1}(k) = \mathbf{Y}(k) - \sum_{l=5}^7 \mathbf{b}(k, \theta_l) S_l(k)$$

and

$$\mathbf{Y}_{2,1}(k) = \mathbf{Y}_{1,1}(k) - \sum_{l=3}^4 \mathbf{b}(k, \theta_l) S_l(k) \quad (11)$$

where  $S_l(k)$  and  $\theta_l$  are the current values of the signal and direction of arrival, respectively. After a number of iterations has been performed on the branch of  $\mathbf{Y}_{2,1}(k)$ , the EM algorithm is run for the branch of  $\mathbf{Y}_{2,2}(k)$ , which can be found as  $\mathbf{Y}_{2,2}(k) = \mathbf{Y}_{1,1}(k) - \sum_{l=1}^2 \mathbf{b}(k, \theta_l) S_l(k)$ , where, in this case,  $S_l(k)$  and  $\theta_l$  for  $l = 1, 2$  are the updated values by the EM algorithm applied to the branch of  $\mathbf{Y}_{2,1}(k)$  before. This switching may be repeated a number of times or until a convergence criterion is satisfied. Having updated values of signal and direction of arrivals for  $l = 1 \cdots 4$ , the EM routine can be run for the branch of  $\mathbf{Y}_{1,2}(k)$  with the same strategy of data assignment. The switching between the branches of  $\mathbf{Y}_{1,1}(k)$  and  $\mathbf{Y}_{1,2}(k)$  can be repeated as well. The reason for forming  $\mathbf{Y}_{1,1}(k)$  as in (11) is now clear; although assignment of  $\mathbf{Y}_{2,1}(k)$  corresponds to

$$\mathbf{Y}_{2,1}(k) = \mathbf{Y}(k) - \sum_{l=3}^7 \mathbf{b}(k, \theta_l) S_l(k) \quad (12)$$

we assign a value to  $\mathbf{Y}_{2,1}(k)$  via  $\mathbf{Y}_{1,1}(k)$  because we want to switch between nodes of  $\mathbf{Y}_{2,1}(k)$  and  $\mathbf{Y}_{2,2}(k)$  and run the EM algorithm to enhance the estimates for  $\mathbf{X}_1(k) \cdots \mathbf{X}_4(k)$ . Then, we go for a lower level switching between  $\mathbf{Y}_{1,1}(k)$  and  $\mathbf{Y}_{1,2}(k)$ .

The proposed tree structure can be seen as a generalization of the CEM algorithm. Although the algorithm is presented for two levels of data hierarchy in [6], it is straightforward to formulate it for multiple levels. Note that there are three levels of data hierarchy in the binary tree example in Fig. 1, and a many-to-one mapping takes place from each level to a lower level of the tree. If we were to update the complete and intermediate data for all the nodes at a particular level

at once, that would be the case of CEM algorithm. However, after establishing that mapping structure in a binary tree, we take the route of the SAGE algorithm and restrict our attention to running EM on smaller sets of complete data, which can be associated with SAGE algorithm as being the hidden data spaces. Therefore, not all of the parameters are updated at a time, but only a subset of parameters are updated sequentially. Note that forming the intermediate data as given above is also related with the content of the information carried by intermediate data. By using that data assignment, we actually extract, from the observations at the lower branch node, the information carried by the parameters other than those to be updated at the current run of the EM algorithm. Hence, most of the noise in the intermediate data of the lower branch node remains on the intermediate data on the branch node where EM is to be applied. This makes the current branch a less informative space of complete data; hence, an increase in the convergence rate can be obtained [7]. Furthermore, in order to improve the signal estimation step of the EM algorithm, we use regularized estimators introduced in the previous section. Hence, the TSEM algorithm with the regularized signal estimation should provide better performance than both the CEM and SAGE algorithms.

Working in smaller dimensional spaces provides not only speed in convergence but also computational saving since in the tree structure, two sources are treated at a time with smaller dimensional matrices resulting in fewer computations per iteration. Therefore, it is suitable to use tree structure especially for a large number of sources and computationally expensive regularized signal estimation methods. The computation required for the overall tree-structured procedure depends on the number of switchings between lower branches and EM iterations performed at each branch and on the computational complexity of the EM algorithm itself, which is determined by the signal estimation methods used in the EM algorithm [10].

#### IV. SIMULATION RESULTS

In this section, we compare the proposed algorithms both with each other and with the conventional EM approach. In addition, we investigate the performance improvement with respect to one of the most improved subspace-based wideband direction-finding algorithms known as the two-sided correlation transformation (TCT) [15].

##### A. Simulation Set 1

This part of the simulations include the performance evaluation of the EM algorithm when, in the signal estimation stage of the maximization step, one of the proposed or the former signal estimation methods is used. The methods are termed as LS-EM, LS, RGLS, LS-SET, and LS-RSET, corresponding to (6)–(10) respectively. The scenario is as follows: Two wideband signals with true direction of arrivals  $\bar{\Theta} = [35^\circ \ -20^\circ]^T$  with respect to the normal of the array are incident onto a linear array of 19 sensors. The number of the sensors is chosen by using the result of another simulation, which is done to reveal the performance of the EM algorithm with changing sensor number [10]. Signals are taken as coherent linear FM waveforms with a bandwidth 116 MHz, a center frequency 106 MHz,  $N = 128$  time samples, and  $F = 256$  FFT points. For each of the signal estimation alternatives, the EM algorithm is run, within a convergence criterion and maximum number of iterations of 40, with an initial direction-of-arrival estimate  $\Theta^0 = [32^\circ \ -17^\circ]^T$  for 10 realizations and *signal-to-noise ratio* ranging from  $-10$  dB to  $10$  dB with  $5$ -dB increments. For initialization of the procedure, the initial signal estimation is performed by using RGLS-2, which is chosen by using the result of another simulation, where we have compared the performance of LS, RGLS and LS-SET in solving (3), given  $\Theta$  [10].

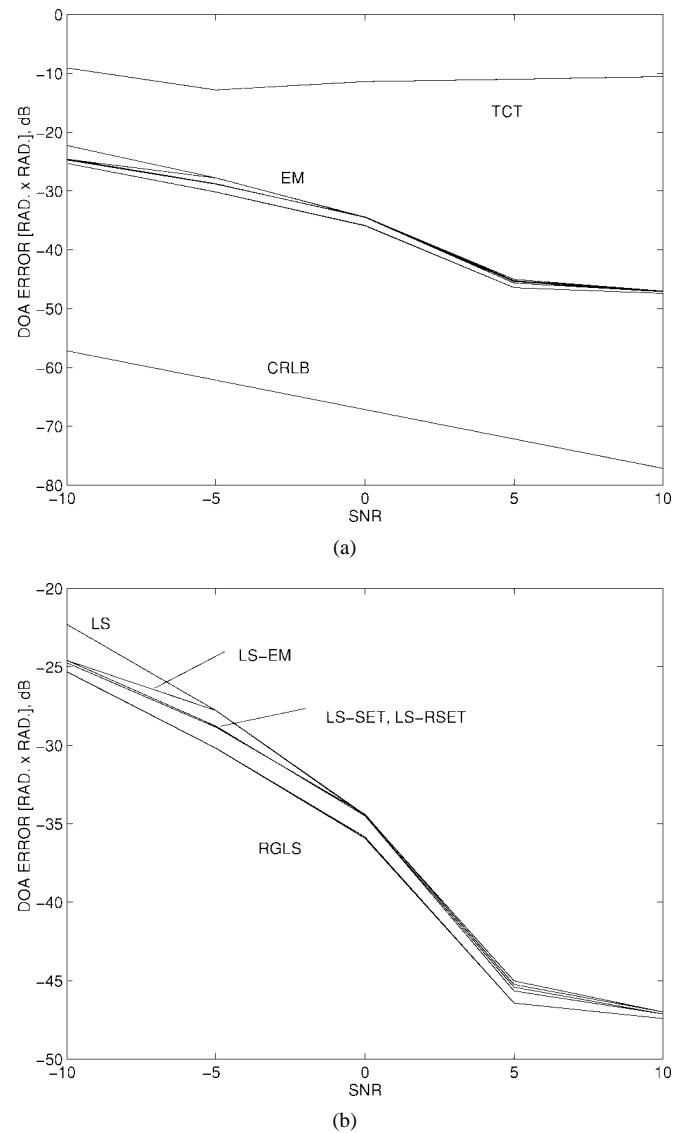


Fig. 2. (a) DOA estimation error of the EM algorithm together with CRLB and result of TCT. (b) Magnified view of EM result with different signal estimation methods.

According to that study, performances can be descendingly ordered as that of RGLS-2, RGLS-1, LS-SET, and LS, with LS-SET being the least sensitive of the methods to the error in initial direction estimates.

The performance of the EM algorithm with alternative signal estimation methods is evaluated in terms of both the direction of arrival and the signal estimation error, which are defined as

$$\epsilon_\Theta = \|\hat{\Theta} - \bar{\Theta}\|^2 \quad \epsilon_s = \frac{1}{MN} \sum_{l=1}^M \sum_{t=0}^{N-1} |\hat{s}_l(t) - \bar{s}_l(t)|^2 \quad (13)$$

where  $\bar{\Theta}$  and  $\bar{s}_l(t)$  are the true direction of arrival and the true signal waveform and  $\hat{\Theta}$  and  $\hat{s}_l(t)$  are their estimates.

The direction-of-arrival estimation error of the EM algorithm using different signal estimation methods is shown in Fig. 2(a) together with the Cramér–Rao lower bound on the variance of the estimates and the results produced by applying the TCT algorithm to our problem. As seen in Fig. 2, the performance of the RGLS method, which is almost the same for RGLS-1 and RGLS-2 algorithms, seems to be superior to the others at the simulated SNR values. This behavior is apparent, especially at low SNR values, where the performance of LS degrades due to the numerical instability

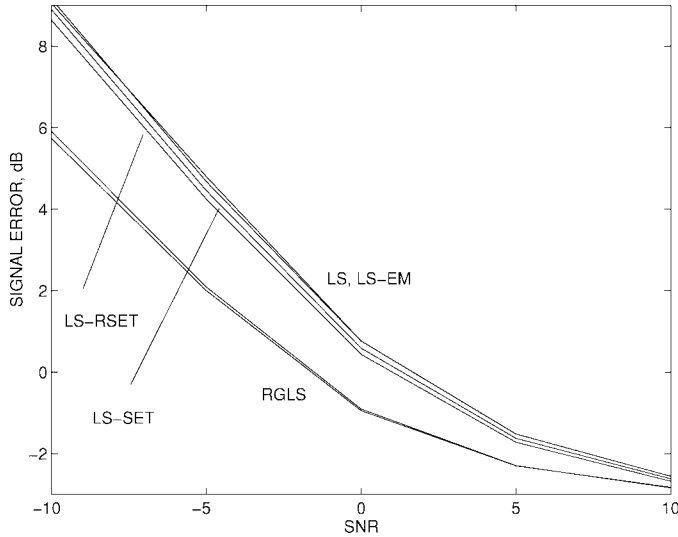


Fig. 3. Signal estimation performance of the EM algorithm with different signal estimation methods.

effect, which was mentioned earlier. Since, if it converges, the EM algorithm provides the same estimates as those of CEM and SAGE algorithms (where the signal estimation is performed by ordinary least squares techniques), the ordinary least squares performance curves in this figure also correspond to the performances of CEM and SAGE algorithms. LS-SET and LS-RSET, having almost identical behavior throughout the range of SNR values, also do better than LS or LS-EM for low SNR regions. Correspondingly, it is clear from Fig. 3 that the RGLS method is superior in producing signal estimates. The LS-SET performs slightly better than LS-RSET, both of which produce less error than the LS or LS-EM, which are almost on the same curve. As expected, these two figures reveal the fact that the signal and direction-of-arrival estimations are closely related. The TCT algorithm has not produced good results in the simulations. This may be due to the large bandwidth of the signals used. In addition, the large number of frequency bins used in our simulations may cause difficulty in focusing different signal subspaces associated with particular frequency bins to a common focusing subspace.

We have also investigated the convergence behavior of the EM algorithm, which employs the signal estimation alternatives. The configuration is the same as in the previous set of simulations. The simulation is done for  $\text{SNR} = 0$  dB and the direction-of-arrival error is traced as the EM algorithm iterates. The result shown in Fig. 4 is the average error of direction-of-arrival estimates over 10 realizations as a function of iteration number of the EM algorithm. In the EM algorithm, if the complete data were available, then the best estimate of the signals would be given by LS-EM. However, since only the incomplete data is available, the greatest increase in the complete data likelihood is obtained by jointly estimating all of the signal components given the currently updated direction values [2]. Notice in Fig. 4 that when joint signal estimation is performed as in LS, RGLS, or LS-SET, convergence improvement over LS-EM and LS-RSET is obtained. Furthermore, with the same reasoning, a regularized signal estimator that provides better signal estimates than an ordinary least squares estimate is expected to increase the estimated likelihood further. In accordance with our expectations, the convergence of RGLS methods clearly outperforms the others, providing a significant gain in the number of iterations that should be performed to reach a satisfactory convergence level. The argument that better signal estimates and joint signal estimation enhance the convergence is totally supported by the results in Figs. 3 and 4.

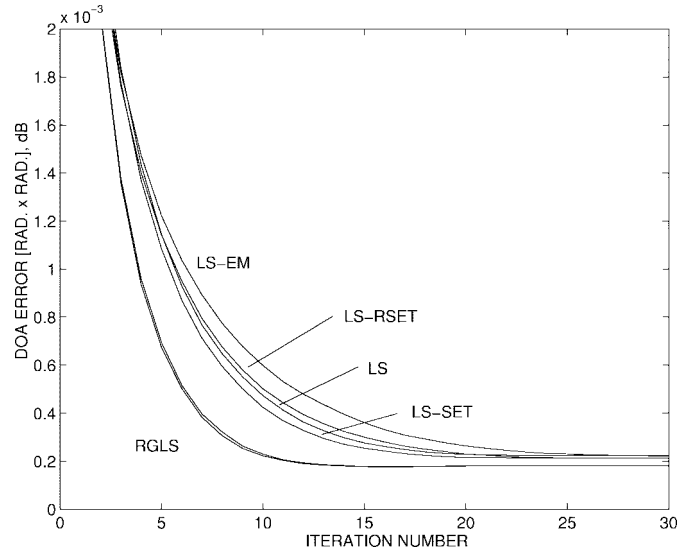


Fig. 4. Trace of direction of arrival error.

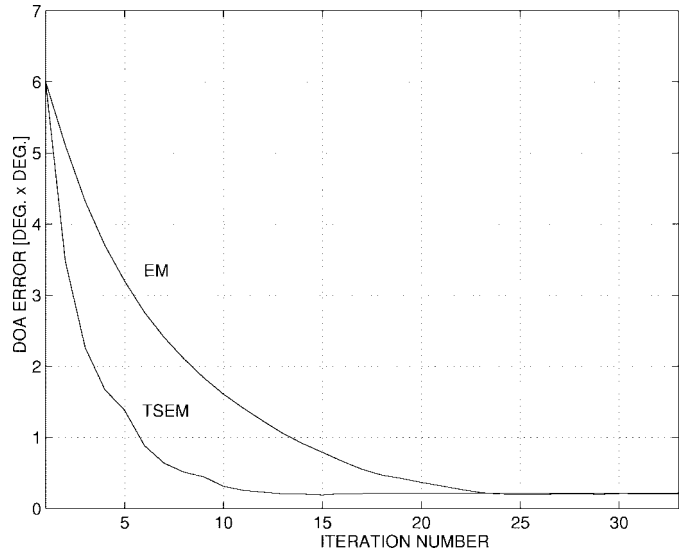


Fig. 5. TSEM versus EM.

## B. Simulation Set II

In this part of the simulations, the tree-structured EM (TSEM) algorithm is compared with the conventional EM algorithm for the case of four sources from directions  $\Theta = [35^\circ -20^\circ -50^\circ 60^\circ]^T$  at  $\text{SNR} = 0$  dB. Initial directions are given as  $\Theta_0 = [32^\circ -17^\circ -47^\circ 57^\circ]^T$ . In TSEM, the EM algorithm, which uses the LS-RSET in its signal estimation, is run for two sources at a time, with a maximum number of iterations of five at each branch. After five iterations at one branch, the algorithm switches to the other branch. This switching is also repeated eight times. Therefore, in total, 40 iterations of the EM algorithm are performed for each direction of arrival. The reason we use LS-RSET solution for the signal estimation stage is that this method is robust to inaccuracies in the initial direction estimates as LS-SET because of the smoothing it performs over a set of neighboring directions. Furthermore, it needs fewer computations than LS-SET.

For comparison purposes, the EM algorithm is also run for four sources with a maximum number of iterations of 40; hence, the number of EM iterations performed on each source is the same in both

cases. The direction-of-arrival estimation error is traced throughout the processes and displayed in Fig. 5, where it can be clearly seen that the tree-structured EM algorithm has a significantly higher speed of convergence. In addition to this important speedup, the tree structure provides savings in each iteration due to the smaller size of matrices and vectors handled in the algorithm. Notice also that Figs. 2 and 3 present comparison results between the TSEM and EM algorithms at their converged estimates. Moreover, since it is known that both CEM and SAGE algorithms provide the same estimates as the EM, only more efficiently, those figures also provide comparison between the converged estimates of TSEM and CEM and SAGE when the curves of EM are interpreted as those of CEM and SAGE.

## V. CONCLUSION

For the case of estimating direction of arrival of unknown deterministic wideband signals arriving from different directions to a passive array, a generalization on the expectation-maximization algorithm is proposed by using a tree-structured multiple-level data mapping. It is demonstrated by simulations that the proposed tree-structured EM (TSEM) algorithm converges faster than the conventional EM algorithm. In addition to speeding up the convergence, it provides considerable saving in computation, especially for a large number of superposed signals, since in the tree structure, two signals are processed at a time with lower dimensional data, unlike the conventional EM algorithm, where the size of data matrices is larger.

In order to obtain reliable estimates, even at low signal-to-noise ratio, and to speed up the convergence of the EM algorithm, alternative regularized least squares estimation techniques are proposed with significant improvement over the standard least squares techniques. It is shown that with better signal estimation during the iterations, the EM algorithm converges faster to more reliable direction-of-arrival and signal estimates.

## REFERENCES

- [1] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the EM algorithm," *J. R. Stat. Soc. B*, vol. 39, pp. 1–37, 1977.
- [2] M. I. Miller and D. R. Fuhrmann, "Maximum-likelihood narrow-band direction finding and the EM algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1560–1577, Sept. 1990.
- [3] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 36, pp. 477–489, Apr. 1988.
- [4] R. A. Boyles, "On the convergence of the EM algorithm," *J. R. Stat. Soc. B*, vol. 45, no. 1, pp. 47–50, 1983.
- [5] C. F. J. Wu, "On the convergence properties of the EM algorithm," *Ann. Stat.*, vol. 11, no. 1, pp. 95–103, 1983.
- [6] M. Segal and E. Weinstein, "The cascade EM algorithm," *Proc. IEEE*, vol. 76, pp. 1388–1390, Oct. 1988.
- [7] J. A. Fessler and A. O. Hero, "Space alternating generalized expectation-maximization algorithm," *IEEE Trans. Signal Processing*, vol. 42, pp. 2664–2677, Oct. 1994.
- [8] D. B. Williams, "Counting the degrees of freedom when using AIC and MDL to detect signals," *IEEE Trans. Signal Processing*, vol. 42, pp. 3282–3284, Nov. 1994.
- [9] P. Stoica and A. Nehorai, "MUSIC, maximum likelihood and cramer-rao bound," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 720–741, May 1989.
- [10] N. Çadallı, "Wide-band maximum likelihood direction finding by using the tree-structured EM algorithm," Master's thesis, Bilkent Univ., Ankara, Turkey, July 1996.
- [11] A. Hoerl and W. Kennard, "Ridge regression: Biased estimation for nonorthogonal problems," *Technometrics*, vol. 12, pp. 55–68, 1970.
- [12] H. D. Vinod, "Survey of ridge regression and related techniques for improvements over ordinary least squares," *Rev. Econ. Stat.*, vol. 60, pp. 121–131, 1978.
- [13] P. A. V. B. Swamy, J. S. Mehta, and P. N. Rappoport, "Two methods of evaluating Hoerl and Kennard's ridge regression," *Commun. Stat.—Theory Methods*, vol. A7, no. 12, pp. 1133–1155, 1978.
- [14] F. Chebil, "Robust estimation of unknowns in a linear system of equations with modeling uncertainties," Master's thesis, Bilkent Univ., Ankara, Turkey, July 1997.
- [15] S. Valaee and P. Kabal, "Wideband array processing using a two-sided correlation transformation," *IEEE Trans. Signal Processing*, vol. 43, pp. 160–172, Jan. 1995.

## Time Series Analysis in the Frequency Domain

Rik Pintelon and J. Schoukens

**Abstract**—This correspondence presents a parametric frequency domain identification algorithm for autoregressive moving average (ARMA) processes that does not suffer from spectral leakage errors. It is based on an extended transfer function model that takes into account the begin and end effect of the finite data record. The relationship with the one-step-ahead prediction error method is established. The advantages of the proposed method are easy prefiltering and leakage-free spectral representation of the raw data.

**Index Terms**—Autoregressive moving average processes, frequency domain analysis, time series.

## I. INTRODUCTION

Time series analysis has a lot of applications such as forecasting in econometrics [1], spectral estimation in signal processing [2], and parametric noise modeling in time-domain system identification [3]. It consists of modeling a time series  $y(t)$  as an autoregressive moving average (ARMA) process

$$y(t) = H(q, \theta)e(t) \quad t = 0, 1, \dots, N-1 \quad (1)$$

where  $e(t)$  is a zero mean, white noise sequence with unknown variance  $\sigma^2$ .  $q$  stands for the backward shift operator,  $H(q, \theta)$  for the parametric noise model

$$H(z^{-1}, \theta) = \frac{C(z^{-1}, \theta)}{D(z^{-1}, \theta)} = \frac{1 + \sum_{m=1}^{n_c} c_m z^{-m}}{1 + \sum_{m=1}^{n_d} d_m z^{-m}} \quad (2)$$

and  $\theta = [c_1, c_2, \dots, c_{n_c}, d_1, d_2, \dots, d_{n_d}]^T$  for the noise model (ARMA) parameters. Note that in (2), the parameter constraints  $c_0 = d_0 = 1$  have been imposed in order to make the estimation problem identifiable [ $\sigma^2 = \text{var}(e(t))$  is unknown]. Since (1) depends

Manuscript received December 19, 1996; revised June 8, 1998. This work was sponsored by the National Fund for Scientific Research of Belgium, the Flemish Government (GOA-IMMI), and the Belgian Program on Interuniversity Poles of Attraction initiated by the Belgian State, Prime Minister's Office, Science Policy Programming (IUAP 4/2). The associate editor coordinating the review of this paper and approving it for publication was Prof. M. H. Er.

The authors are with the Vrije Universiteit Brussel, Brussels, Belgium (e-mail: rpintel@vnet3.vub.ac.be).

Publisher Item Identifier S 1053-587X(99)00149-X.