

Wide-Band Maximum Likelihood Direction Finding by Using Tree-Structured EM Algorithm

Nail Çadallı and Orhan Arıkan

Department of Electrical and Electronics Engineering, Bilkent University,
TR-06533 Ankara, TURKEY. Phone: 90-312-2664307, Fax: 90-312-2664126,
e-mail: cadalli@ee.bilkent.edu.tr, oarikan@ee.bilkent.edu.tr

Abstract

A tree structured Expectation Maximization (EM) algorithm is proposed and applied to the wide-band angle of arrival estimation. It may be seen as a generalization on EM using the ideas of Cascade EM algorithm and Space Alternating Generalized EM algorithm. Also, for passive data acquisition, robust and efficient alternatives for the estimation of the source signals are investigated.

1. Introduction

In many data acquisition systems, reception data acquired by an array of sensors are processed to obtain information about the source locations. When the sources are located relatively far away from the sensors, only the direction of arrivals of the acquired source signals can be reliably obtained. Although the Maximum likelihood (ML) estimation provides more accurate estimates for the direction of arrivals, due to the higher computational cost of obtaining the ML estimates, it has not found much use in practice. However, by exploiting the superposition property of the data acquisition system, the complexity of the ML estimation can be greatly reduced by using the Expectation Maximization (EM) algorithm [1, 2, 5]. In EM formalism, the observation, *incomplete data* is obtained via a many-to-one mapping from the *complete data* space that includes signals which 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 log-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 condi-

tions, the iterations of the EM algorithm converges to a stationary point of the observed log-likelihood function, where at each iteration the likelihood of the estimated parameters is increased [11]. In this study, a tree structured hierarchy is used for the description of relation between the complete data space and the observations. Within this hierarchy it is possible to combine in one algorithm the ideas of the Cascade EM and Space Alternating Generalized EM algorithms [3, 7]. 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. For this purpose two different methods are used. The first one is an adaptive Tikhonov type regularized least-squares (RLS) estimation method, which is computationally intensive and the second one is an averaged least-squares estimation (LSSET) method over a set of angles in a neighborhood of the nominal angles. It has been demonstrated that when RLS or LSSET methods are used in the estimation of the received signals, the EM algorithm has better convergence behavior.

2. Signal Model

For the case of M sources with direction of arrivals θ_l , $1 \leq l \leq M$, the measured signal at the i 'th sensor of an array with P sensors is

$$y_i(t) = \sum_{l=1}^M a_l(t, \theta_l) * s_l(t - \tau_i(\theta_l)) + u_i(t) \quad (1)$$
$$1 \leq i \leq P, \quad t=0, T, 2T, \dots, (N-1)T$$

where $s_l(t)$ is the wide-band signal of the l 'th source, $u_i(t)$ is the 0 mean spatially and temporally white Gaussian noise at the i 'th sensor, $\tau_i(\theta)$ is the time delay of the source signal from the direction θ as it propagates to the i 'th sensor relative to the phase center of

the array, $a_i(t, \theta)$ is the time domain function for the gain of the i 'th sensor which is dependent on frequency and the direction of arrival, θ . The frequency domain representation of (1) is,

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

where F is the DFT size which is chosen sufficiently large and $Y_i(k)$, $A_i(k, \theta)$, $S_l(k)$ ve $U_i(k)$ are the transforms of $y_i(t)$, $a_i(t, \theta)$, $s_l(t)$ and $u_i(t)$ respectively. Let the following definitions be made ($'$ is the transpose operator)

$$\begin{aligned} \mathbf{b}(k, \theta) &= [A_1(k, \theta) e^{-j \frac{2\pi k}{F} \frac{r_1(\theta)}{T}} \dots \\ &\quad A_P(k, \theta) e^{-j \frac{2\pi k}{F} \frac{r_P(\theta)}{T}}]' \\ &= [b_1(k, \theta) \dots b_P(k, \theta)]' \\ \mathbf{B}(k, \Theta) &= [\mathbf{b}(k, \theta_1) \dots \mathbf{b}(k, \theta_M)] \\ \mathbf{S}(k) &= [S_1(k) \dots S_M(k)]' \\ \mathbf{Y}(k) &= [Y_1(k) \dots Y_P(k)]' \end{aligned}$$

Using these definitions (2) becomes

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

This final compact form of the measurement relation, which is the same as the signal model of the Cramer-Rao Lower Bound formula in [8], will be used in our derivations.

3. Wide-Band EM Algorithm

Since, the measurement noise is modeled as normally distributed additive noise, the probability density of the observations are Gaussian. Hence, the log-likelihood function of the observations has the following familiar form (\dagger is the conjugate transpose operator),

$$\mathcal{L}(\Theta, \mathbf{S}; \mathbf{Y}) = - \sum_{k=0}^{F-1} [\mathbf{Y}(k) - \mathbf{B}(k, \Theta) \mathbf{S}(k)]^\dagger [\mathbf{Y}(k) - \mathbf{B}(k, \Theta) \mathbf{S}(k)] \quad (4)$$

In order to find the ML estimate, likelihood function of the observations should be maximized with respect to Θ and $\mathbf{S}(k)$. However, the direct maximization of this function is not only computationally demanding but also due to the local maxima structure of the likelihood function it is not guaranteed to converge to the global maxima. The Expectation Maximization (EM) method of obtaining the ML estimate overcomes this

difficulty by an iterative search in much lower dimensional parameter spaces [1]. The EM method requires the identification of so called complete data space. In our application the commonly used complete data is $\mathbf{X}_i(k) = [X_{1l}(k) \dots X_{Pl}(k)]'$ which is the signal that would be observed at the sensors if we were able to see the effect of l 'th source only. Then the many-to-one mapping for all sources from the complete data space to the incomplete data space can be written as

$$\mathbf{Y}(k) = \sum_{l=1}^M \mathbf{X}_l(k) \quad 0 \leq k < F \quad (5)$$

The mean of the complete data $\mathbf{X}_l(k)$ is $\mathbf{b}(k, \theta_l) S_l(k)$ and it is normally distributed. The log-likelihood function of the complete data is

$$\mathcal{L}_c(\Theta, \mathbf{S}; \mathbf{X}) = - \sum_{k=0}^{F-1} \sum_{l=1}^M \|\mathbf{X}_l(k) - \mathbf{b}(k, \theta) S_l(k)\|^2 \quad (6)$$

Here, the observed signal is decomposed to M constituents. Therefore to estimate θ_l and $S_l(k)$, only $\mathbf{X}_l(k)$ is used besides the observation. At the n 'th iteration of the EM algorithm expectation step conditionally estimates the likelihood of the complete data $\mathcal{L}_c(\Theta, \mathbf{S} \mid \Theta^n, \mathbf{S}^n)$. Maximization step then finds the maximizer of the estimated likelihood and assigns to θ_l^{n+1} . To find $\mathbf{b}(k, \theta_l) S_l(k)$ it is sufficient to know $\mathbf{X}_l(k)$, therefore in expectation step $\mathbf{X}_l(k)$ is estimated. It can be shown that, ([6], p. 164),

$$\begin{aligned} \mathbf{X}_l^n(k) &= \mathcal{E}\{\mathbf{X}_l(k) \mid \theta_l^n, S_l^n(k), \mathbf{Y}(k)\} \\ &= \mathbf{b}(k, \theta_l^n) S_l^n(k) + \frac{1}{M} [\mathbf{Y}(k) - \mathbf{B}(k, \Theta^n) \mathbf{S}^n(k)] \quad 0 \leq k < F \quad (7) \end{aligned}$$

In maximization step complete data likelihood which is formed by using $\mathbf{X}_l^n(k)$ is maximized with respect to θ_l and $S_l(k)$. The θ_l update is found as

$$\theta_l^{n+1} = \arg \max_{\theta} \left\{ \max_{S_l} \left\{ - \sum_{k=0}^{F-1} \sum_{l=1}^M \|\mathbf{X}_l^n(k) - \mathbf{b}(k, \theta) S_l(k)\|^2 \right\} \right\} \quad (8)$$

where there is two maximization problems inside one another. If $S_l(k)$ is unknown they must be simultaneously solved. For a given θ value, the solution of the inner maximization is

$$\begin{aligned} S_l(k) &= [\mathbf{b}(k, \theta) \mathbf{b}^\dagger(k, \theta)]^{-1} \mathbf{b}^\dagger(k, \theta) \mathbf{X}_l(k) \\ &= \frac{\mathbf{b}^\dagger(k, \theta) \mathbf{X}_l(k)}{\|\mathbf{b}(k, \theta)\|^2} \quad (9) \end{aligned}$$

Inserting this expression into (8) and solving for the outer maximization θ_l^{n+1} is found. For that maximization linear search may be used. Finally, at the n 'th iteration of the EM algorithm the update formulas are as follows,

$$\theta_l^{n+1} = \arg \max_{\theta} \sum_{k=0}^{F-1} \frac{\mathbf{b}^\dagger(k, \theta) \mathbf{X}_l^n(k) \mathbf{X}_l^{\dagger n}(k) \mathbf{b}(k, \theta)}{\|\mathbf{b}(k, \theta)\|^2} \quad (10)$$

$$S_l^{n+1}(k) = \frac{\mathbf{b}^\dagger(k, \theta_l^{n+1}) \mathbf{X}_l^n(k)}{\|\mathbf{b}(k, \theta_l^{n+1})\|^2} \quad (11)$$

If $S_l(k)$ is known, as in active array applications, (8) is simply reduced to one maximization problem and there remains no need for (11). If (10) and (11) are run together, i.e., in the case of unknown source signals, θ_l^{n+1} should be close to true direction values for S_l^{n+1} to converge to true signal waveforms.

After (10), Θ^{n+1} is available. If it is inserted into (3), $\mathbf{S}^{n+1}(k)$ can be solved for by using a number of alternatives. For instance the least squares (LS) solution is as follows,

$$\begin{aligned} \hat{\mathbf{S}}(k) &= \arg \min_{\mathbf{S}(k)} \|\mathbf{Y}(k) - \mathbf{B}(k, \Theta) \mathbf{S}(k)\|^2 \\ &= [\mathbf{B}^\dagger(k, \Theta) \mathbf{B}(k, \Theta)]^{-1} \mathbf{B}^\dagger(k, \Theta) \mathbf{Y}(k) \end{aligned} \quad (12)$$

Regularization may be applied on the LS solution which is called regularized least squares, RLS,

$$\hat{\mathbf{S}}(k) = [\mathbf{B}^\dagger(k, \Theta) \mathbf{B}(k, \Theta) + \mu \mathbf{I}]^{-1} \mathbf{B}^\dagger(k, \Theta) \mathbf{Y}(k) \quad (13)$$

It is important to choose μ in the regularization and it can be chosen optimally [4, 9]. Another alternative in source signal estimation may be the following which will be referred to as LSSET solution, where \mathcal{K} is a set of angles in a neighborhood of Θ ,

$$\hat{\mathbf{S}}(k) = \arg \min_{\mathbf{S}(k)} \int_{\mathcal{K}} \|\mathbf{Y}(k) - \mathbf{B}(k, \Theta) \mathbf{S}(k)\|^2 d\Theta \quad (14)$$

EM algorithm starts with $n = 0$ at which time Θ^0 is available obtained by using a rough estimation. To find $\mathbf{X}_l^0(k)$ in (7), S_l^0 is needed and it is estimated by one of the methods mentioned above. EM shows monotonic increase of the likelihood and its convergence issues have been investigated [1, 11].

4. Tree-Structured EM

In this section we will use a different mapping from the complete data to the incomplete data which is

structured as a binary tree as shown in Figure 1. $\mathbf{Y}_{i,j}(k)$ is the intermediate incomplete data between the observation $\mathbf{Y}(k)$ and the complete data $\mathbf{X}_l(k)$'s. In this setting EM algorithm is run for two sources at a time using the intermediate data at the joint node of two leaves. This provides an update for the corresponding DOA and source signals. The value of the intermediate data is found by using, in (3), the original observation $\mathbf{Y}(k)$ and the current source signal estimates other than the ones which are to be updated by the current run. For instance, to run EM algorithm for $\mathbf{X}_1(k)$ and $\mathbf{X}_2(k)$ we form the required incomplete data as

$$\mathbf{Y}_{2,1}(k) = \mathbf{Y}_{1,1}(k) - \mathbf{B}(k, \begin{bmatrix} \theta_3 \\ \theta_4 \end{bmatrix}) \begin{bmatrix} \mathbf{S}_3(k) \\ \mathbf{S}_4(k) \end{bmatrix} \quad (15)$$

where $\mathbf{Y}_{1,1}(k)$ is found by using $\mathbf{Y}(k)$ and the current estimates for the last three source signals in (3). $\mathbf{Y}_{2,2}(k)$ may be found similarly and EM algorithm is run for that branch too. This may be repeated a number of times and then by using the updates obtained for the first 4 source signals and DOA's, branch of $\mathbf{Y}_{1,2}(k)$ may be processed. The idea of putting inter-

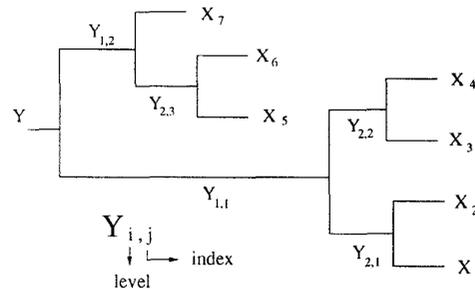


Figure 1. An example for the tree structure.

mediate data mappings between $\mathbf{Y}(k)$ and $\mathbf{X}_l(k)$'s can be associated with that of the Cascade EM, CEM, algorithm but here there is more than one intermediate data space. Due to the limited space, the generalization of CEM to multiple levels is not presented here. The tree structure may also be associated with Space Alternating Generalized EM algorithm in the sense that not all of the parameters are updated at a time. Also EM is run on a more noisy data reducing the information content of intermediate observations and this is reported to speed the convergence [3].

5. Simulations and Conclusions

Observation signals are obtained by simulation of a linear array of sensors. The number of signals are assumed to be known since there are studies in detection

[10]. The source signals are taken as coherent pulse modulated chirp signals with bandwidth comparable to the center frequency. Noise is assumed to be independent identical Gaussian distributed. First the

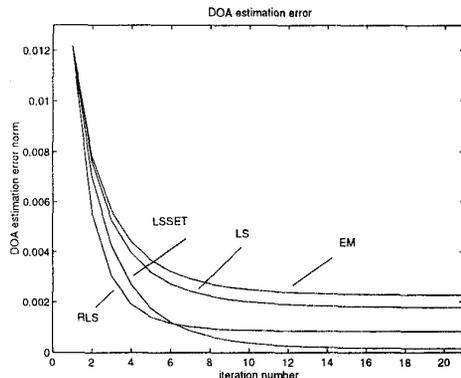


Figure 2. Averaged traces of DOA error norms.

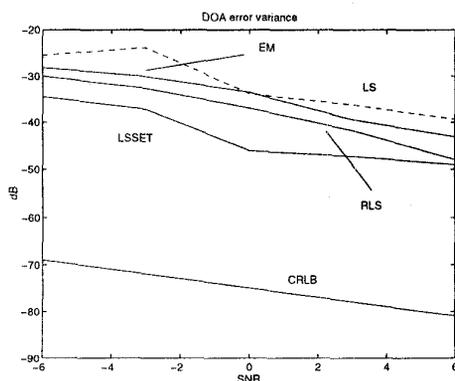


Figure 3. DOA error variances.

signal estimation alternatives are inserted in the EM algorithm and their relative performances are compared. EM algorithm is run for two sources impinging from 35° and -50° at an SNR level of 0dB. The averaged traces of error norm of DOA estimation, which describes the convergence behaviours, can be seen in Figure 2 where EM, LS, RLS, LSSET refers to (11), (12), (13), (14) respectively. The DOA error variance together with the CRLB for each alternative is plotted in Figure 3. For LSSET, \mathcal{K} consists of 5 angles in a 1° neighborhood of the current DOA. This figures out to be computationally less complex than RLS.

To compare the tree-structured EM algorithm with the original EM algorithm four sources from directions $\Theta = [35^\circ - 50^\circ - 20^\circ 50^\circ]'$ are used at SNR=10dB. Initial DOA's are given as $\Theta_0 = [33^\circ - 48^\circ - 18^\circ 48^\circ]'$.

The DOA error norms for iterations of original EM and tree-structured EM algorithms are shown in the next table. The original EM algorithm, could not converge to true DOA values. Furthermore, it diverges from the initial angle values. But within the same number of total iterations the tree-structured EM converges with much lower DOA error to $\Theta = [35.3 - 50.0 - 20.0 50.7]'$.

iteration no. \rightarrow	10	20	50	100
EM (10^{-3})	5.3	5.5	5.5	5.5
Tree-EM (10^{-4})	6.3	6.1	4.6	2.2

By this study, an improvement on EM algorithm is realized not only by using robust signal estimation schemes but also by changing the data mapping of the original algorithm.

References

- [1] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *J. Roy. Stat. Soc.*, B-39:1-37, 1977.
- [2] M. Feder and E. Weinstein. Parameter estimation of superimposed signals using the EM algorithm. *IEEE Trans. Acoust., Speech, Signal Processing*, 36(4):477-489, April 1988.
- [3] J. A. Fessler and A. O. Hero. Space alternating generalized expectation-maximization algorithm. *IEEE Trans. Signal Processing*, 42(10):2664-2677, Oct. 1994.
- [4] A. Hoerl and W. Kennard. Ridge regression: biased estimation for non-orthogonal problems. *Technometrics*, 12:55-68, 1970.
- [5] M. I. Miller and D. R. Fuhrmann. Maximum-likelihood narrow-band direction finding and the EM algorithm. *IEEE Trans. Acoust., Speech, Signal Processing*, 38(9):1560-1577, Sep. 1990.
- [6] A. Papoulis. *Probability, Random Variables and Stochastic Processes*. McGraw-Hill, 3rd edition, 1991.
- [7] M. Segal and E. Weinstein. The cascade em algorithm. *Proc. IEEE*, 76(10):1388-1390, Oct. 1988.
- [8] P. Stoica and A. Nehorai. MUSIC, Maximum Likelihood and Cramer-Rao Bound. *IEEE Trans. Acoust., Speech, Signal Processing*, 37(5):720-741, May 1989.
- [9] H. D. Vinod. Survey of ridge regression and related techniques for improvements over ordinary least squares. *Review of Economics and Statistics*, 60:121-131, 1978.
- [10] H. Wang and M. Kaveh. Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-33(4):823-831, Aug. 1985.
- [11] C. F. J. Wu. On the convergence properties of the EM algorithm. *The Annals of Statistics*, 11(1):95-103, 1983.