OPTIMAL MULTIPLE SOURCE LOCATION ESTIMATION VIA THE *EM* ALGORITHM

Meir Feder and Ehud Weinstein

Department of Electrical Engineering and Computer Science
Research Laboratory of Electronics, Rm. 36-615
Massachusetts Institute of Technology, Cambridge, MA 02139
and
Department of Ocean Engineering
Woods Hole Oceanographic Institution, Woods Hole, MA 02543
and
Department of Electronic Systems, Faculty of Engineering,
Tel-Aviv University, Tel-Aviv, Israel

ABSTRACT

We developed an algorithm for multiple source localization based on the Estimate-Maximize (EM) method. The EM method is an iterative algorithm that converges to the Maximum Likelihood (ML) estimate of the unknown parameters by exploiting the stochastic syctem under consideration. In our case the algorithm will converge to the *exact* ML estimates of the various sources location parameters, where each iteration increases the likelihood of those parameters.

I. INTRODUCTION

The localization of a radiating acoustic source can be determined by observation of its signal at a spatially separated array of receivers. Localization of several spatially distributed signal sources is conceptually not very different, however, in actual practice, it is significantly more difficult to accomplish. In this study, we consider the problem of multiple source location estimation using the Maximum Likelihood (ML) criterion.

The ML estimation procedure requires the maximization of the likelihood function given the observed data with respect to all unknown location parameters. In the case of single source, the ML estimate is obtained by maximizing the array beamformer output with respect to one pair of location parameters, namely bearing and range. However, if there are several closely spaced signal sources, the likelihood functional assumes a significantly more complicated form, and its maximization with respect to all pairs of location parameters, using Newton-Raphson or any other brute force gradient-search iterative algorithm, is computationaly very intensive and time consuming.

In this study, we develop a new scheme for multiple source location estimation based on the Estimate Maximize (EM) algorithm [1]. The EM algorithm is basicly an iterative method to extract the ML estimate from "incomplete" data. However, unlike the brute force iterative algorithms, the EM algorithm can be used to carefully exploit the stochastic system under consideration.

II. MAXIMUM LIKELIHOOD ESTIMATION AND THE EM ALGORITHM

The ML criterion is widely regarded as the optimal criterion in parameters estimation problems. Given the observed data $\underline{y}$, the ML estimate of the vector parameters $\underline{\theta}$ is obtained by maximizing the likelihood function i.e.

$$\max_{\underline{\theta}} \log f_{\underline{y}}(\underline{y};\underline{\theta}) \longrightarrow \underline{\theta}_{ML}(\underline{y}) \tag{1}$$

where $f_{\underline{y}}(\underline{y};\underline{\theta})$ is the probability density function (p.d.f) of $\underline{y}$. If the vector $\underline{\theta}$ contains several unknowns, such as in the multiple source location problem, and since $\log f_{\underline{y}}(\underline{y};\underline{\theta})$ is generally a highly non-linear function of $\underline{\theta}$, the maximization required in (1) tends to be very complex.

One of the reasons why it is difficult to maximize the likelihood function is that the observed data $\underline{y}$ is incomplete, perhaps because of the limited available information in either the time or the spatial domain. Let $\underline{x}$ denote the complete data. For our purposes, the complete data is any vector $\underline{x}$ such that

$$\underline{y} = H(\underline{x})$$

where $H(\cdot)$ is a non-invertible transformation. This relation means that we observe the complete data only through a non-invertible transformation, where this transformation generally causes a reduction in the available data. In the multiple source location problem, the complete data $\underline{x}$ could be the observation of each source signal *separately* where the

CH2118-8/85/0000-1762 $1.00 © 1985 IEEE

incomplete (observed) data $\underline{y}$ is the *sum* of signal contributions from the various sources. Thus, in the transition from $\underline{x}$ to $\underline{y}$ there is a significant reduction in the data caused by the nature of the problem.

Given a complete data specification, the EM method is characterized by the following iterative algorithm:

$$\max_{\underline{\theta}} E\{\log f_{\underline{x}}(\underline{x};\underline{\theta})/\underline{y},\underline{\theta}^{(n)}\} \longrightarrow \underline{\theta}^{(n+1)} \qquad (2)$$

where $\underline{\theta}$ is the current estimate after n iterations of the algorithm. $E\{\log f_{\underline{x}}(\underline{x};\underline{\theta})/\underline{y},\underline{\theta}^{(n)}\}$ is the conditional expectation of $\log f_{\underline{x}}(\underline{x};\underline{\theta})$ given the observed data $\underline{y}$ and the current estimate, and $\log f_{\underline{x}}(\underline{x};\underline{\theta})$ is the likelihood functional of the complete data. Thus each iteration cycle consists of a conditional expectation (estimate) step followed by a maximization step. This is why the algorithm is termed the Estimate-Maximize algorithm.

The heuristic idea here is that we would like to choose $\underline{\theta}$ that maximizes the likelihood of the complete data. However since $\log f_{\underline{x}}(\underline{x};\underline{\theta})$ is not available to us (since the complete data $\underline{x}$ is not available), we maximize instead its expectation given the observed data $\underline{y}$ and we use for that expectation the current estimate $\underline{\theta}^{(n)}$ rather then the actual value of the parameter which is unknown. The conditional expectation is thus not exact, but the algorithm iterates using each new parameter estimate to improve the conditional expectation on the next iteration cycle and thus to improve the next parameter estimate.

In [1] it is shown that under certain regularity conditions, the algorithm converges essentially exponentially to the desired result, i.e. $\underline{\theta}^{(n)} \longrightarrow \underline{\theta}_{ML}(\underline{y})$, where each iteration increases the likelihood of the estimated parameters.

Thus, the direct maximization in (1), which is commonly accomplished via one of the brute force gradient-search iterative algorithm, is substituted by the iterative algorithm given in (2). Hence, the EM algorithm is useful only if the latter procedure is computationally simpler to carry out.

The conditional expectation and the maximization required in each iteration of the EM algorithm critically depends on the complete data specification. Obviously, there are many possible complete data specifications that will generate the observed data, and unfortunate choice of the complete data may yield a completely useless algorithm. However, as previously mentioned, in the multiple source location problem, there is a nartural choice of complete data, leading to a surprisingly simple algorithm to extract the ML astimates of the various source location parameters.

III.MULTIPLE SOURCE LOCATION ESTIMATION VIA THE *EM* ALGORITHM

The basic system of interest consists of several spatially distributed sources, radiating noise-like signals towards the receiving array as illustrated in figure 1. Each receiver output is contaminated by an additive noise, so that the actual waveform observed at the output of the m^{th} receiver is given by

$$y_m(t) = \sum_{k=1}^{K} \alpha_{km} s_k(t-\tau_{km}) + n_m(t) \quad m = 1,2,\ldots M \qquad (3)$$

where M is the number of elements in the receiving array, K is the assumed number of signal sources, α_{km} are possible amplitude scales, and τ_{km} is the travel time of the signal from the k^{th} source to the m^{th} receiver. To simplify the exposition, we shall assume that the various signal wavefronts are essentially planar across the array baseline so that

$$\tau_{km} = \tau_m(\theta_k)$$

where θ_k is the bearing angle of the k^{th} source measured relative to the center of the array.

We shall further assume that tha various source signals $s_k(t)$ and the additive noises $n_m(t)$ are statistically uncorrelated, zero-mean and Gaussian.

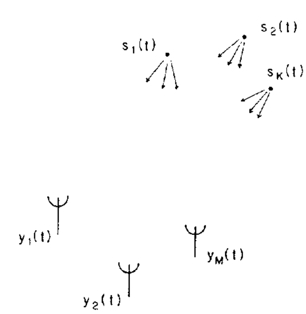
Finally, we shall assume that the observation time T is large compared with the correlation time (inverse bandwidth) of the signals and the noises (i.e. $WT \gg 1$), and with the propagation time of the signals across the array.

With those simplifying assumptions, it can be shown that the likelihood function to be maximized is given by:

$$\log f_{\underline{y}}(\underline{y};\underline{\theta}) = \frac{T}{2\pi}\int_{\omega}[\log\det\Lambda(\omega;\underline{\theta}) + \underline{Y}^*(\omega)\Lambda^{-1}(\omega;\underline{\theta})\underline{Y}(\omega)]d\omega \qquad (4)$$

where the integration in (4) is carried out over all ω in the signal frequency band. $\underline{\theta}$ is the $K \times 1$ vector of unknown bearing parameters. $Y^*(\omega)$ denotes the conjugate-trasspose of $Y(\omega)$, where $Y(\omega)$ is the $M \times 1$ vector whose m^{th} element is given by:

$$Y_m(\omega) = \frac{1}{\sqrt{T}}\int_{-T/2}^{T/2} y_m(t)e^{-j\omega t}dt$$



ARRAY - SOURCE GEOMETRY
Figure 1

$\Lambda(\omega;\underline{\theta})$ is the $M\times M$ data covariance matrix at the frequency ω which is given by:

$$\Lambda(\omega;\underline{\theta}) = U(\omega;\underline{\theta})P_s(\omega)U^*(\omega;\underline{\theta}) + P_n(\omega)$$

where $U(\omega;\underline{\theta})$ is the $M\times K$ matrix whose m,k element is given by:

$$U_{mk}(\omega;\underline{\theta}) = \alpha_{km}e^{-j\omega\tau_m(\theta_k)}$$

$P_s(\omega)$ and $P_n(\omega)$ are the diagonal matrices

$$P_s(\omega) = diag(S_1(\omega), S_2(\omega), \ldots S_K(\omega))$$

$$P_n(\omega) = diag(N_1(\omega), N_2(\omega), \ldots N_M(\omega))$$

where $S_k(\omega)$ is the spectral density of the k^{th} signal source, and $N_m(\omega)$ is the spectral density of the additive noise observed at the output of the m^{th} receiver.

Thus, just the calculation of the likelihood function (4) requires the inversion and determinant calculation of $\Lambda(\omega;\underline{\theta})$ for each frequency ω in the signal band. It is clear that brute force maximization of (4) with respect to the parameters vector $\underline{\theta}$ that might require also the calculation of the gradient of (4) is computationally very complex.

An alternative approach to this problem is suggested by the EM algorithm. The EM algorithm is directed at finding a value of $\underline{\theta}$ that maximizes the likelihood function given the observed (incomplete) data, however, it does so by making an essential use of the way in which the complete data is specified. A natural choice of the complete data will be

$$\underline{x}(t) = (\underline{x}_1^T(t), \underline{x}_2^T, \ldots \underline{x}_K^T(t))^T \tag{5}$$

where

$$\underline{x}_k(t) = (x_{k1}(t), x_{k2}(t), \ldots x_{kM}(t))^T$$

and

$$x_{km}(t) = \alpha_{km}s(t-\tau_m(\theta_k)) + n_{km}(t)$$

Thus, the components of $\underline{x}_k(t)$ are the signals observed from the k^{th} source in the presence of additive noise. The noises $n_{km}(t)$ are zero-mean Gaussian processes satisfying the constraint

$$\sum_{k=1}^{K} n_{km}(t) = n_m(t) \quad m = 1, 2, \ldots M$$

It immediately follows that

$$y_m(t) = \sum_{k=1}^{K} x_{km}(t) \quad m = 1, 2, \ldots M \tag{6}$$

Concatenting the various M equations into a vector form, the relation between the complete and incomplete (observed) data is given by:

$$\underline{y}(t) = \sum_{k=1}^{K} \underline{x}_k(t) = H\underline{x}(t) \tag{7}$$

where

$$H = [\underbrace{I\ I\ \cdots\ I}_{K\ terms}]$$

and I is the $M\times M$ identity matrix

Ignoring constant terms that are independent of the unknown parameters, the likelihood function of the complete data is given by:

$$\log f_{\underline{x}}(\underline{x};\underline{\theta}) = \sum_{k=1}^{K} \int_{\omega} |F_k(\omega)\underline{V}^*(\omega)\underline{X}_k(\omega)|^2 d\omega \tag{8}$$

where $F_k(\omega)$ is a shaping filter that depends on the signal to noise ratio, the amplitude scales α_{km}, and the number of receivers M.

$\underline{V}(\omega;\theta_k)$ is the steering vector of the array in direction θ_k i.e.

$$\underline{V}(\omega;\theta_k) = (e^{-j\omega\tau_1(\theta_k)}, e^{-j\omega\tau_2(\theta_k)}, \ldots e^{-j\omega\tau_M(\theta_k)})^T$$

and $\underline{X}_k(\omega)$ is the $M\times 1$ vector whose m^{th} element is given by:

$$X_{km}(\omega) = \frac{1}{\sqrt{T}}\int_{-T/2}^{T/2} x_{km}(t)e^{-j\omega t}dt$$

We recognize each term in (8) as the output of a beamformer steered to direction θ_k. Thus, the likelihood function of the complete data consists of K terms, each term is generated by a beamforming operation.

Now we must substitute (8) into (2) and carry out the indicated operation. Since the k^{th} term in the sum in (8) depends only on θ_k, the maximize step in the algorithm decouples into the K separate single-parameter maximizations

$$\max_{\theta_k}\int_{\omega} |F_k(\omega)|^2\underline{V}^*(\omega;\theta_k)Q_k(\omega;\underline{\theta}^{(n)})\underline{V}(\omega;\theta_k)d\omega \longrightarrow \theta_k^{(n+1)} \tag{9}$$

where

$$Q_k(\omega;\underline{\theta}^{(n)}) = E\{\underline{X}_k(\omega)\underline{X}_k^*(\omega)/\underline{Y}(\omega), \underline{\theta}^{(n)}\}$$

Thus, the EM algorithm requires the maximization of K beamformers in parallel, where we substitute the sufficient statistic $\underline{X}_k(\omega)\underline{X}_k^*(\omega)$ by its current expectation. Since $\underline{X}_k(\omega)$ and $\underline{Y}(\omega)$ are related by a linear transformation, they are jointly Gaussian. In that case the conditional expectation required in order to get Q_k is readily available in the literature. Specifically we get

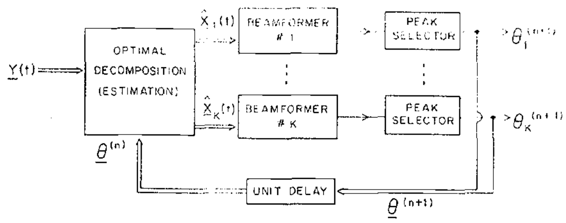
$$Q_k(\omega;\underline{\theta}^{(n)}) = B_k +$$

$$B_k\Lambda^{-1}(\omega;\underline{\theta}^{(n)}[\underline{Y}(\omega)\underline{Y}^*(\omega) - \Lambda(\omega;\underline{\theta}^{(n)})]\Lambda^{-1}(\omega;\underline{\theta}^{(n)})B_k \tag{10}$$

where B_k is the cross-spectrum matrix of the array due to $\underline{x}(t)$ using the value $\underline{\theta}^{(n)}$ for the unknown parameters i.e.

$$B_k = S_k(\omega)\underline{V}(\omega;\theta_k^{(n)})\underline{V}^*(\omega;\theta_k^{(n)}) + N_{km}(\omega) \tag{11}$$

(9) together with (10) and (11) completely specify the algorithm. The algorithm is illustrared schematically in figure 2.



MULTIPLE - SOURCE LOCALIZATION via the EM ALGORITHM

Figure 2

Perhaps the most attractive feature of the algorithm is that it does not require the multiple parameter maximization assosiated with the brute force gradient-search iterative scheme. The algorithm naturally decouples the multiple parameter maximization to K seperate single parameter maximizations using the computationally covenient beamformer. The extention to bearing and range estimation is straightforward. The basic scheme is still characterized by the above equations (illustrated by figure 2), where each beamformer output is now maximized with respect to a pair of bearing and range parameters. Thus, the complexity of the algorithm is unaffected by the assumed number of sources. As the number of sources increases, we have tu use more beamformers in parallel, however, each beamformer output is maximized seperately.

Since the algorithm is based on the Estimate-Maximize method, it will converge to the optimal (ML) solution, where each iteration increases the likelihood of the estimated parameters. Hence, the computationally efficient algorithm presented here can be used to *simultaneously* extract the *exact* ML estimates of all source location parameters without any loss in performance accuracy.

REFERENCES

[1] A. P. Dempster, N. M. Laird, and D. B. Rubin, *"Maximum Likelihood from Incomplete Data via the EM Algorithm,"* Ann. of the Royal Stat. Soc., pp 1-38, Jan. 1977.