June, 1987 LIDS-P-1670

Maximum Likelihood Array Processing for the Estimation of Superimposed Signals

Anthony J. Weiss[1], Alan S. Willsky[2], Bernard C. Levy[2]

Laboratory for Information and Decision Systems
Massachusetts Institute of Technology
Cambridge, Massachusetts 02139

Abstract

A novel and efficient algorithm for computing the maximum likelihood estimates of multiple signals observed by an array of sensors is presented. The algorithm provides estimates of parameters related to the directional patterns of the sources as well as estimates of the location parameters of the sources. Furthermore, the algorithm is equally applicable to wideband sources and narrowband sources and does not require a knowledge of the statistical properties of the signals.

---

[1]The work of the first author was supported by a Rothschild Fellowship.

[2]The work of these authors was supported in part by the National Science Foundation under Grant ECS-8312921 and in part by the Army Research Office under Grants DAAG-84-K-0005 and DAAL 03-86-K-0171.

I - Introduction

The localization of radiating sources by a passive array of sensors is a problem of considerable importance, occuring in a variety of fields ranging from radar, sonar, oceanography and seismology to radio-astronomy. Therefore, this problem has received considerable attention in the literature, resulting in a variety of estimation schemes, most of which are suboptimal or limited to special cases. A comprehensive literature survey, including more than 120 references is included in [1]; see also [2] for many other references not discussed in [1]. Here we concentrate on maximum likelihood estimation inspired by the recent work of Ziskind and Wax [3], and Feder and Weinstein [4]. Unlike [3], our approach is not limited to narrowband signals radiated by omnidirectional sources, and in contrast with [4], we do not assume known signals or random signals with known statistics. However, our approach can be considered as a modification of a special case of the EM (Expectation-Maximization) algorithm for unknown deterministic signals [5]. This modification enables the algorithm to converge considerably faster than the EM algorithm.

II - Problem Formulation

Consider N radiating sources with arbitrary radiation patterns observed by an array of M sensors. The signal at the output of the m-th sensor can be described by

$$x_m(t) = \sum_{n=1}^{N} \alpha_{mn} s_n(t-\tau_{mn}) + v_m(t) \quad ; \quad m = 1,2,\cdots,M \quad . \qquad (1)$$

$$-T/2 \leq t \leq T/2$$

where $\left\{s_n(t)\right\}_{n=1}^{N}$ are the radiated signals, $\left\{v_m(t)\right\}_{m=1}^{M}$ are additive noise processes, and T is the observation interval. The intensities α_{mn} and the delays τ_{mn} are parameters related to the directional pattern and relative location of the n-th source and the m-th sensor.

A convenient separation of the parameters to be estimated is obtained by using Fourier coefficients defined by

$$X_m(\omega_\ell) = \frac{1}{\sqrt{T}} \int_{-T/2}^{T/2} x_m(t) e^{-j\omega_\ell t}\, dt \quad ,$$

where $\omega_\ell = \frac{2\pi}{T}(\ell_1+\ell)$, $\ell = 1,2,\cdots,L$, and ℓ_1 is a constant. In principle the number of required coefficients tends to infinity. However, since we consider only finite bandwidth signals, we can use only $L < \infty$ coefficients. Taking the Fourier coefficients of (1) we obtain:

$$X_m(\omega_\ell) = \sum_{n=1}^{N} \alpha_{mn} e^{-j\omega_\ell \tau_{mn}} S_n(\omega_\ell) + V_m(\omega_\ell) \quad , \qquad (2)$$

where $S_n(\omega_\ell)$ and $V_m(\omega_\ell)$ are the Fourier coefficients of $s_n(t)$ and $v_m(t)$ respectively. Equation (2) may be expressed using vector notation as follows:

$$\underline{X}(\omega_\ell) = A(\omega_\ell)\underline{S}(\omega_\ell) + \underline{V}(\omega_\ell) \quad ; \quad \ell = 1,2,\cdots,L \quad ; \qquad (3)$$

where

$$\underline{X}(\omega_\ell) = \left[X_1(\omega_\ell),\ X_2(\omega_\ell),\cdots,X_M(\omega_\ell)\right]^T$$

$$\underline{S}(\omega_\ell) = \left[S_1(\omega_\ell),\ S_2(\omega_\ell),\cdots,S_N(\omega_\ell)\right]^T$$

$$\underline{V}(\omega_\ell) = \left[V_1(\omega_\ell),\ V_2(\omega_\ell),\cdots,V_M(\omega_\ell)\right]^T$$

$$A(\omega_\ell) = \left[\underline{a}_\ell(\theta_1),\ \underline{a}_\ell(\theta_2),\cdots,\underline{a}_\ell(\theta_N)\right]$$

$$\underline{a}_\ell(\underline{\theta}_n) = \left[\alpha_{1n} e^{-j\omega_\ell \tau_{1n}},\ \alpha_{2n} e^{-j\omega_\ell \tau_{2n}},\cdots,\alpha_{Mn} e^{-j\omega_\ell \tau_{Mn}}\right]^T \quad .$$

We use $\underline{\theta}_n$ to represent all the parameters of interest associated with the n-th signal, namely $\left\{\alpha_{mn}\right\}_{m=1}^{M}$ and $\left\{\tau_{mn}\right\}_{m=1}^{M}$. Our main goal is to estimate the set $\left\{\underline{\theta}_n\right\}_{n=1}^{N}$. Note that if the spectrum of the signals is concentrated around ω_1, with a bandwidth that is small compared to $2\pi/T$, then (3) reduces to a single relation between the observation vector $\underline{X}(\omega_1)$ and the parameters, i.e. L = 1. In this case, it is customary to use many short observation intervals or simply time samples, and the model becomes:

$$\underline{X}(j) = A\underline{S}(j) + \underline{V}(j) \quad ; \quad j = 1,2,\cdots,J \quad , \tag{4}$$

where the dependence on the single frequency ω_1 is suppressed, and j denotes the index of the different samples. Note that the main difference between the narrowband case and the wideband case is that A is the same in all the J equations specified by (4) while $A(\omega_\ell)$ is different in each of the L equations given by (3). However, the estimation procedure discussed here is equally applicable in both cases. In this communication we concentrate on the narrowband case. The modification for the wideband case is straightforward and is described in [5].

Under the assumption that the number of sources is known, the least squares estimates of $\left\{\underline{\theta}_n\right\}$ is given by:

$$\left\{\hat{\underline{\theta}}_n\right\}_{n=1}^{N} = \arg\min_{\left\{\theta_n\right\} \in \Theta} Q \quad ; \quad Q \overset{\Delta}{=} \sum_{j=1}^{J} \left|\left|\underline{X}(j) - A\underline{S}(j)\right|\right|^2 \quad ; \tag{5}$$

where $||\cdot||$ denotes the Euclidean norm and Θ is the given parameter space. Equation (5) also represents the maximum likelihood estimates under the assumption that the noise vectors $\left\{\underline{V}(j)\right\}$ are i.i.d. zero-mean Gaussian with covariance $\sigma^2 I$.

III - The Estimation Procedure

The minimization required in (5) is not trivial since the vector $\underline{S}(j)$ and the matrix A are not known to the observer. However, whenever A is known Q is minimized by choosing

$$\hat{\underline{S}}(j) = (A^H A)^{-1} A^H \underline{X}(j) \qquad (6)$$

as the estimate of $\underline{S}(j)$ for $j = 1,2,\cdots,J$, where $(\cdot)^H$ denotes the Hermitian-transpose operation. Relation (6) enables us to update the estimates $\hat{\underline{S}}(j)$ whenever we have a new estimate for A. The main principle of the algorithm is to perform successive minimization operations on each column of A, holding all the rest of the columns and the associated components of $\underline{S}(j)$ fixed. For example, suppose that we want to perform a minimization with respect to the k-th column vector, then Q can be rewritten as

$$Q = \sum_{j=1}^{J} || \underline{Y}^k(j) - \underline{a}(\underline{\theta}_k) S_k(j) ||^2 \qquad (7)$$

where $\underline{a}(\underline{\theta}_k)$ is the k-th column of A, $S_k(j)$ is the k-th component of $\underline{S}(j)$ and $\underline{Y}^k(j)$ is given by

$$\underline{Y}^k(j) = \underline{X}(j) - A\underline{S}^k(j) \quad , \qquad (8)$$

where $\underline{S}^k(j)$ is simply $\underline{S}(j)$ with the k-th component replaced by zero.

The minimization of (7) with respect to $\underline{a}(\underline{\theta}_k)$, using (6) with A replaced by $\underline{a}(\underline{\theta}_k)$, is given by

$$\hat{\underline{a}}(\underline{\theta}_k) = \arg \min_{\underline{\theta}_k \in \Theta} \sum_{j=1}^{J} || \underline{Y}^k(j) - \underline{a}(\underline{\theta}_k) \left[\underline{a}^H(\underline{\theta}_k) \underline{a}(\underline{\theta}_k) \right]^{-1} \underline{a}^H(\underline{\theta}_k) \underline{Y}^k(j) ||^2$$

which is equivalent to

$$\hat{\underline{a}}(\underline{\theta}_k) = \arg \min_{\underline{\theta}_k \in \Theta} || \underline{a}(\underline{\theta}_k) ||^{-2} \sum_{j=1}^{J} |(\underline{Y}^k(j))^H \underline{a}(\underline{\theta}_k)|^2 \quad . \qquad (9)$$

Note that $|| \underline{a}(\underline{\theta}_k) ||^2 = \sum_{m=1}^{M} \alpha^2_{mk}$. Since there is an extra degree of freedom (due to the estimation of both $\{\alpha_{mn}\}$ and $\{\underline{S}(j)\}$), there is no loss of generality in assuming that $|| \underline{a}(\underline{\theta}_k) ||^2 = 1$. This simplifies (9) considerably. Now note that $\underline{a}(\underline{\theta}_k)$ may be decomposed as follows:

$$\underline{a}(\theta_k) = \Gamma(\underline{\tau}_k)\underline{\alpha}_k \tag{10}$$

where

$$\underline{\alpha}_k = (\alpha_{1k}, \alpha_{2k}, \cdots, \alpha_{Mk})^T$$

$$\Gamma(\underline{\tau}_k) = \text{diag}(e^{-j\omega_l \tau_{1k}}, e^{-j\omega_l \tau_{2k}}, \cdots, e^{-j\omega_l \tau_{Mk}})$$

$$\underline{\tau}_k = (\tau_{1k}, \tau_{2k}, \cdots, \tau_{Mk})^T \quad .$$

Using (10) in (9) we obtain

$$\hat{\underline{a}}(\underline{\theta}_k) = \arg \max_{||\underline{\alpha}_k|| = 1} \underline{\alpha}_k^T \left\{ \sum_{j=1}^{J} \Gamma^H(\underline{\tau}_k) \underline{Y}^k(j) (\underline{Y}^k(j))^H \Gamma(\underline{\tau}_k) \right\} \underline{\alpha}_k \tag{11}$$

Since $\underline{\alpha}_k$ is a real vector the solution of (11) is given by:

$$\hat{\underline{\tau}}_k = \arg \max_{\underline{\tau}_k} \lambda^{\max} \{R_k\} \tag{12.a}$$

$$\hat{\underline{\alpha}}_k = \underline{U}^{\max} \tag{12.b}$$

where $\lambda^{\max}\{R_k\}$ is the largest eigenvalue of the matrix R_k given by:

$$R_k = \text{Re}\left\{ \sum_{j=1}^{J} \Gamma^H(\underline{\tau}_k) \underline{Y}^k(j) (\underline{Y}^k(j))^H \Gamma(\underline{\tau}_k) \right\} \quad , \tag{13}$$

and $\underline{U}^{\max}$ is the associated normalized eigenvector.

The maximization described by (12.a) can be performed by a simple search over the space of $\underline{\tau}_k$, induced by all possible individual source locations, or by a simple gradient subalgorithm.

The algorithm is summarized as follows:

(a) Initialization: Select $A = A^{(0)}$. Set $k = 1$.

(b) Compute $\underline{S}(j)$ according to (6).

(c) Compute R_k according to (8), (13).

(d) Find $\hat{\underline{\tau}}_k$, $\hat{\underline{\alpha}}_k$ according to (12).

(e) Update A with the new $\hat{\underline{a}}(\underline{\theta}_k)$; set $k = k+1$; if $k > N$ then $k = 1$.

(f) Check the convergence of A; if yes: done; if no: go to (b).

Observe that at each updating step (i.e. steps (b) and (e)), we decrease the cost function Q defined in (5). Since $Q \geq 0$ the algorithm will converge at least to a local minimum of Q. Depending on the initial estimate of A and on the structure of Q, the local minimum may or may not coincide with the global minimum.

IV - An Example

To illustrate the behavior of the algorithm, let us consider an example. Specifically, consider a uniform linear array of 5 sensors separated by half a wavelength of the actual narrowband source signals. The sources are two narrowband emitters located in the far-field of the array. In this case, if γ_n denotes the bearing of the n-th source, n = 1,2, relative to the perpendicular to the array baseline, the differential delay is given by $\tau_{mn} = (m-1)\pi \sin(\gamma_n)$. The first source at a bearing of 10 degrees was observed with the intensity vector $\underline{\alpha}_1^T = [1,.8,.6,.4,.2]$; the second source at a bearing of 30 degrees was observed with $\underline{\alpha}_2^T = [1, 1, 1, 1, 1]$. In this case the difference in intensity may be viewed as caused by the directional pattern of the sensors rather than the directional pattern of the sources. We

generated 50 independent samples at a SNR of 30 dB. The initial guess was $\tau_1^{(0)} = 3°$, $\tau_2^{(0)} = 42°$, $\underline{\alpha}_1^{(0)} = \underline{\alpha}_2^{(0)} = [1, 1, 1, 1, 1]^T$. The algorithm converged to within one degree of the right result in 16 iterations as shown in Table 1. Note that the residual relative errors in $\underline{\alpha}_1$ and $\underline{\alpha}_2$ are respectively 4.8% and 12.3%. (The relative error is defined by $|| \hat{\underline{\alpha}}_i - \underline{\alpha}_i ||/|| \underline{\alpha}_i ||$.) This result is rather impressive if we consider that 110 independent parameters ($\left\{\underline{\alpha}_n\right\}$, $\left\{\tau_n\right\}$, $\underline{S}(j)$) have been estimated simultaneously.

V - Conclusion

We have presented a general algorithm for obtaining maximum likelihood estimates of superimposed signals. Perhaps one of the most distinctive features of the algorithm is its ability to obtain estimates of the intensity vectors $\left\{\underline{\alpha}_n\right\}$. These estimates may be useful in their own right, but also their estimation is essential even if one is only interested in the delays in cases where it is not appropriate to assume omnidirectionality. For example, whenever a source is in the near field of the array, its radiation pattern can rarely be assumed omnidirectional. This is also important in applications in which it is unrealistic to assume that the radiation pattern of each sensor is accurately known (this usually requires frequent calibrations and a large memory).

Other features of the algorithm, including its application to array processing of wideband signals, correlated signals (multipath), and cases when only few time samples are available, and its application to spectrum estimation of superimposed signals, are discussed in [5].

As mentioned in the introduction, the algorithm may be viewed as a modification of a special case of the EM algorithm proposed in [4]. According to the theory of the EM algorithm, the estimates generated in the M-step should be used in the E-step. This may be applied to the present algorithm as follows. Instead of updating $\underline{S}(j)$ using (6) in step (b), $\underline{S}(j)$ is updated by replacing only its k-th component by the estimate, $\underline{a}^H(\underline{\theta}_k)\underline{Y}^k(j)$, which can be computed in step (d), following the computation of $\hat{\underline{\tau}}_k$ and $\hat{\underline{\alpha}}_k$. Note that $\underline{a}^H(\underline{\theta}_k)\underline{Y}^k(j)$ is simply the value of $S_k(j)$ that minimizes (7) whenever $\underline{a}(\underline{\theta}_k)$ is known. It is clear that the last procedure typically will require more iterations than the proposed procedure since the updating of $\underline{S}(j)$ is done without using all the currently available information.

Finally, we would like to emphasize that the algorithm will converge to the right result only if the initial estimates are good enough. Fast initial estimates can be obtained by using simpler methods such as MLM, MEM, or MUSIC procedures (see, for example, [1]).

References

[1] M. Wax, "Detection and estimation of superimposed signals," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1985.

[2] U. Nickel, "Angular superresolution with phased array radar: A review of algorithms and operational constraints," IEE Proceedings, Vol. 134, No. 1, pp. 53-59, February 1987.

[3] I. Ziskind and M. Wax, "Maximum likelihood estimation via the alternating projection maximization algorithm," Proc. of the International Conference on Acoustics, Speech and Signal Processing (ICASSP), Dallas, TX, pp. 2280-2283, April 1987.

[4] M. Feder and E. Weinstein, "Parameter estimation of superimposed signals using the EM algorithm," preprint.

[5] A.J. Weiss, A.S. Willsky, and B.C. Levy, "Iterative maximum likelihood estimation of superimposed signals," in preparation.

Iteration No.	$\hat{\tau}_1$ degrees	$\hat{\underline{\alpha}}_1^T$	$\hat{\tau}_2$ degrees	$\hat{\underline{\alpha}}_2^T$
0	3.00	(1, 1, 1, 1, 1)	42.00	(1, 1, 1, 1, 1)
1	12.22		35.01	
2	11.41		33.91	
3	11.24		33.31	
4	11.16		32.87	
5	11.15	(1,.83,.63,.38,.15)	32.51	(1,.69,.62,.76,.85)
6	11.10		32.22	
7	11.09		31.98	
8	11.06		31.78	
9	11.04		31.61	
10	11.02	(1,.84,.63,.39,.15)	31.46	(1,.81,.75,.81,.87)
11	11.01		31.34	
12	10.99		31.23	
13	10.97		31.13	
14	10.96		31.09	
15	10.94		31.02	
16	10.93	(1,.84,.63,.39,.15)	30.98	(1,.86,.83,.86,.91)

Table 1.

Evolution of the algorithm for the example.