July 1987 LIDS-P-1686

OPTIMUM AND SUBOPTIMUM ARRAY PROCESSING FOR THE ESTIMATION OF SUPERIMPOSED SIGNALS

by

Anthony J. Weiss*
Alan S. Willsky**
Bernard C. Levy**

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

ABSTRACT

This report addresses the problem of estimating the parameters of superimposed signals observed by an array of sensors. Some of the proposed techniques are equally useful for estimating the frequencies of sinusoids in noise. The methods used are direct iterative maximum likelihood, the EM algorithm, the eigenstructure approach and the polynomial approach. In addition to the traditional estimation of the source locations we also address the estimation of parameters related to the radiation patterns of the sources.

---

*The work of the first author was supported by a Rothschild fellowship.

**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 DAAG29-84-K-0005 and DAAL03-86-K-1071.

I. INTRODUCTION

This report addresses the problem of estimating the parameters of superimposed signals, occurring in a variety of fields ranging from radar, sonar and oceanography to seismology and radio-astronomy.

In this section we formulate the superimposed signals problem, review the relevant literature, and present a summary of the content and the contributions of this report.

1.1 Formulation of the problem

Our formulation of the superimposed signals problem is motivated by the specific problem of localizing N radiating sources using an array of M sensors. The signal at the output of the m-th sensor can be described by

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

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

where $\{s_n(t)\}_{n=1}^{N}$ are the radiated signals, $\{v_m(t)\}_{m=1}^{M}$ 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 patterns and relative locations of the n-th source and the m-th sensor. Note that α_{mn} is a function of the radiation pattern of the source in the direction of the sensor, of the radition pattern of the sensor in the direction of the source and of the distance between the source and the sensor. However, τ_{mn} is only a function of the distance between the source and the sensor. Hence the estimation of $\{\alpha_{mn}\}$ and $\{\tau_{mn}\}$ yields important information on the locations

and radiation patterns of the sources.

A convenient separation of the parameters of interest is obtained by using Fourier coefficients defined by

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

where $\omega_i = 2\pi(i_1+i)/T$, $i = 1,2,\ldots,I$; and i_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 $I<\infty$ coefficients.

Taking the Fourier coefficients of (1) we obtain:

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

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

$$\underline{X}(\omega_i) = A(\omega_i)\underline{S}(\omega_i) + \underline{V}(\omega_i); \quad i = 1,2,\ldots,I; \tag{3}$$

where

$$\underline{X}(\omega_i) = [X_1(\omega_i),\ X_2(\omega_i),\ldots,X_M(\omega_i)]^T$$

$$\underline{S}(\omega_i) = [S_1(\omega_i),\ S_2(\omega_i),\ldots,S_N(\omega_i)]^T$$

$$\underline{V}(\omega_i) = [V_1(\omega_i),\ V_2(\omega_i),\ldots,V_M(\omega_i)]^T$$

$$A(\omega_i) = [\underline{a}_i(\underline{\theta}_i),\ \underline{a}_i(\underline{\theta}_2),\ldots,\underline{a}_i(\underline{\theta}_N)]$$

$$\underline{a}_i(\underline{\theta}_n) = [\alpha_{1n}e^{-j\omega_i\tau_{1n}},\ \alpha_2 e^{-j\omega_i\tau_{2n}},\ldots,\alpha_{mn}e^{-j\omega_i\tau_{mn}}]^T .$$

We use $\underline{\theta}_n$ to represent all the parameters of interest associated with the n-th signal, namely $\{\alpha_{mn}\}_{n=1}^{M}$ and $\{\tau_{mn}\}_{m=1}^{M}$. Our main goal is to estimate the set $\{\underline{\theta}_n\}_{n=1}^{N}$. 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. I=1. In this case it is customary to use many short observation intervals or simply time samples, and the model becomes:

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

where the dependence on the single frequency ω_1 is suppressed, and j is 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 $A(\omega_i)$ is different in each of the I equations given by (3). In this report we concentrate on the <u>wideband</u> case whenever the proposed procedure can handle both the wideband case and the

narrowband case.

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

$$\{\hat{\underline{\theta}}_n\} = \arg\min_{\{\underline{\theta}_n\}\in\Theta} Q; \quad Q = \sum_{i=1}^{I} \left|\left|\underline{X}(\omega_i) - A(\omega_i)\underline{S}(\omega_i)\right|\right|^2 \tag{5}$$

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 $\{\underline{V}(\omega_i)\}$ are i.i.d. zero mean Gaussian with covariance $\sigma^2 I$.

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

$$\hat{\underline{S}}(\omega_i) = [A(\omega_i)^H A(\omega_i)]^{-1} A(\omega_i)^H \underline{X}(\omega_i) \tag{6}$$

as the estimates of $\underline{S}(\omega_i)$, where $(\ )^H$ denotes the Hermitian-transpose operation. Substituting (6) in (5) we obtain

$$\{\hat{\underline{\theta}}_n\} = \arg\max_{\{\underline{\theta}_n\}\in\Theta} \sum_{i=1}^{I} \underline{X}(\omega_i)^H A(\omega_i)[A(\omega_i)^H A(\omega_i)]^{-1} A(\omega_i)^H \underline{X}(\omega_i). \tag{7}$$

the maximization in (7) requires a multidimensional search over all the parameters α_{mn} and τ_{mn} and since this problem is difficult many papers and

books have proposed suboptimum estimation schemes.

I.2 Literature Survey

A comprehensive literature survey, including more than 120 references is included in [1]. Also see [2] for many other references not discussed in [1]. It is beyond the scope of this report to describe every algorithm and every set of assumptions in the hundreds of estimation schemes that have been proposed until now. Instead we confine our attention to the techniques that are currently the most promising.

The EM Algorithm [6]-[8]

The EM (Expectation-Maximization) algorithm was recently proposed by M. Feder and E. Weinstein for treating the problem described above. The EM procedure is essentially an iterative algorithm that is guaranteed to converge to a stationary point of the likelihood function. Hence, if it converges to the global maximum of the likelihood function the estimates are exactly the maximum likelihood estimates. The main disadvantages of this algorithm are:

1) If the likelihood function is not unimodal the algorithm may converge to a local maximum. Hence, it may be necessary to overcome this problem by appropriate measures.
2) The algorithm is often slow to converge and the amount of computation required for each iteration may be large.

In [6]-[8] Feder and Weinstein derived the EM algorithm for the general linear Gaussian case for known signals in noise and random signals in noise with known Gaussian statistics. Here we extend these results to the more realistic case of non-random unknown signals in noise. We also make use of

the EM algorithm in order to apply the polynomial approach to nonuniform arrays.

The Covariance Eigenstructure Approach

The covariance eigenstructure approach was first proposed in the time-domain by Schmidt [3], who called it the MUltiple Signal Identification and Classification (MUSIC) method, and by Bienvenu [10] who developed an equivalent frequency-domain procedure. Since its introduction, a large number of extensions and refinements of this method have been proposed, and this technique is therefore considered as one of the most practical for solving the superimposed signal identification and retrieval problem. Note however that since this method does not attempt directly or indirectly to maximize the likelihood function, it is suboptimal. Yet, it yields good results for sufficiently high signal to noise ratio (SNR). Its main advantages are:

1) It relies on an algorithm which is not iterative, and hence it eliminates the problem of converging to local stationary points.
2) The amount of computation is less than the amount required for the EM algorithm.

Its main drawbacks are:

1) The algorithm cannot be used in problems in which there are only a small number of observations.
2) There is no natural way to extend the algorithm to handle wideband signals. For different extensions for wideband signals, see [9]-[15].

In this report we show how the MUSIC algorithm can be used in the

rather interesting and practical case of non-omnidirectional sources in the near field of the array.

The Polynomial Approach [5], [16]-[18].

The polynomial approach in the context of maximum likelihood (ML) estimation was introduced only recently by Bresler and Macovski [5]. This approach is limited to the special, but important, case of linear, uniformly spaced, narrowband arrays.

Although the starting point of this technique is precisely the ML estimation problem described above, the algorithm proposed in [5] is not guaranteed to yield results that are confined to the a-priori known parameter space. However, in simulations, the algorithm converges within 5-10 iterations to the right results, for high enough SNR. In this report we extend the polynomial approach to nonuniform arrays.

Summary of Content and Contributions

In Section II we first briefly review the EM algorithm and then following the approach proposed by Feder and Weinstein, we derive an EM algorithm for the general case of superimposed, unknown deterministic signals in noise. This extension of the work in [6]-[8] is important for obvious reasons. However, the results are very general and require some refinement for practical computation. This is done in the last part of Section II in which we show how to obtain an estimation of the location and intensity parameters by a relatively efficient EM procedure. In Section III we describe a novel and efficient algorithm for computing the ML estimates of superimposed signals. The algorithm is equally applicable to wideband sources and narrowband sources and does not require a knowledge of the

statistical properties of the signals. Typically, it requires less iterations than the EM algorithm. In Section IV we briefly review the eigenstructure approach and then we employ some of the ideas of Section II to extend this approach to the estimation of the radiation patterns of the sources as well as the location of the sources. Some users may prefer this suboptimal approach, since it is very fast and yields good results at high enough SNR. Section V is devoted to linear, narrowband arrays with nonuniform sensor spacings. We make use of both the polynomial approach and the EM algorithm.

II. APPLICATION OF THE EM ALGORITHM TO THE ESTIMATION OF SUPERIMPOSED UNKNOWN SIGNALS IN NOISE

In this section we briefly review the EM method and then apply it to the ML problem described above.

II.1 The EM Method

Let $\underline{X}$ denote the observation vector and $\underline{\theta}$ represent the parameter vector. If $f_{\underline{x}}(\underline{X}|\underline{\theta})$ is the conditional probability density function of $\underline{x}$ given $\underline{\theta}$, then the ML estimate of $\underline{\theta}$ is:

$$\hat{\underline{\theta}} = \arg\max_{\underline{\theta}\in\Theta} f_{\underline{x}}(\underline{X}|\underline{\theta}) = \arg\max_{\underline{\theta}\in\Theta} \ln\{f_{\underline{x}}(\underline{X}|\underline{\theta})\}, \tag{8}$$

where Θ is the parameter space.

In many cases of interest one would like to observe $\underline{Y}$, the "complete data", instead of $\underline{X}$, the "incomplete data", where the relation between $\underline{X}$ and $\underline{Y}$ is given by some non-invertible mapping:

$$H(\underline{Y}) = \underline{X}.$$

From Bayes' rule we have

$$\ln\{f_{\underline{x}}(X|\underline{\theta})\} = \ln\{f_{\underline{y}}(\underline{Y}|\underline{\theta})\} - \ln\{f_{\underline{y}|\underline{x}}(\underline{Y}|\underline{X},\underline{\theta})\} \tag{9}$$

Taking the expectation of (9) over $\underline{y}$ given $\underline{X}$ and under the assumption that the parameter is equal to $\underline{\theta}'$, we obtain

$$L(\underline{\theta}) \triangleq \ln\{f_{\underline{X}}(\underline{X}|\underline{\theta})\} = Q(\underline{\theta}|\underline{\theta}') - H(\underline{\theta}|\underline{\theta}'), \tag{10}$$

where

$$Q(\underline{\theta}|\underline{\theta}') \triangleq E\{\ln\{f_{\underline{Y}}(\underline{Y}|\underline{\theta})\}|\underline{X},\underline{\theta}'\}$$

$$H(\underline{\theta}|\underline{\theta}') \triangleq E\{\ln\{f_{\underline{Y}|\underline{X}}(\underline{Y}|\underline{X},\underline{\theta})\}|\underline{X},\underline{\theta}'\}.$$

Using Jensen's inequality it is easy to verify that

$$H(\underline{\theta}|\underline{\theta}') \leq H(\underline{\theta}'|\underline{\theta}'). \tag{11}$$

The EM procedure may be described by the following sequence [20]:

(a) Initialization: set p=0, and $\underline{\theta}^{(p)} = \theta_0$.

(b) E-step: Determine $Q(\underline{\theta}|\underline{\theta}^{(p)})$.

(c) M-step: Choose $\underline{\theta}^{(p+1)}$ to be the value of $\underline{\theta}\varepsilon\Theta$ that maximizes $Q(\underline{\theta}|\underline{\theta}^{(p)})$.

(d) Check the convergence of $\underline{\theta}$. No - p=p+1, go to (b).
Yes - stop.

In every cycle of the algorithm the likelihood function $L(\underline{\theta})$ is increased, since:

$$L(\underline{\theta}^{(p+1)}) = Q(\underline{\theta}^{(p+1)}|\underline{\theta}^{(p)}) - H(\underline{\theta}^{(p+1)}|\underline{\theta}^{(p)})$$

$$\geq Q(\underline{\theta}^{(p)}|\underline{\theta}^{(p)}) - H(\underline{\theta}^{(p)}|\underline{\theta}^{(p)}) = L(\underline{\theta}^{(p)}).$$

The inequality holds due to (11) and due to the M-step. For proof of convergence of this procedure, see [21].

II.2 Application to Array Processing

In this section we apply the EM method to the case of superimposed unknown nonrandom signals. We concentrate on the wideband case described by equations (5). The modification for the narrowband case is straightforward.

Following [6]-[8] we choose the "complete data" as the observation of each of the signals separately. Hence

$$\underline{Y}_n(\omega_i) = \underline{a}_i(\underline{\theta}_n)S_n(\omega_i) + \underline{V}_n(\omega_i), \quad n = 1,2,\ldots,N; \quad i = 1,2,\ldots,I; \tag{12}$$

where the fictitious noises $\{\underline{V}_n(\omega_i)\}$ are chosen to be mutually uncorrelated, zero mean Gaussian vectors with covariance $\sigma_1^2 I$ satisfying

$$N\sigma_1^2 = \sigma^2, \text{ and } \sum_{n=1}^{N} V_n(\omega_i) = \underline{V}(\omega_i), \quad i = 1,2,\ldots,I. \tag{13}$$

The complete data vector is

$$\underline{Y} = [\underline{Y}^T(\omega_1), \underline{Y}^T(\omega_2),\ldots,\underline{Y}^T(\omega_I)]^T,$$

where

$$\underline{Y}(\omega_i) = [\underline{Y}_1^T(\omega_i),\ \underline{Y}_2^T(\omega_i),\ldots,\underline{Y}_N^T(\omega_i)]^T.$$

The "incomplete data" (i.e. the observed data) $\underline{X}$ is defined by

$$\underline{X} = [\underline{X}^T(\omega_1),\ \underline{X}^T(\omega_2),\ldots,\underline{X}^T(\omega_I)]^T,$$

and is obtained from $\underline{Y}$ by the linear transformation

$$\underline{X} = G\underline{Y},$$

where G is a block diagonal matrix with I blocks:

$$G = \begin{bmatrix} H & & 0 \\ & H & \\ & & \ddots \\ 0 & & H \end{bmatrix},$$

and the matrix H is constructed of N identity matrices of size MxM,

$$H = [I_M I_M \ \ldots\ I_M] .$$

We now turn to evaluate the functional $Q(\underline{\theta}|\underline{\theta}')$. We recall that $\underline{Y}$ is Gaussian with known covariance $\sigma_1^2 I$ and unknown mean $\underline{\mu}(\underline{\theta})$ (where the parameter vector $\underline{\theta}$ includes not only the parameters $\{\alpha_{mn}\}$ and $\{\tau_{mn}\}$ but also the signal parameters $S_n(\omega_i)$ with $1 \leq i \leq I$, $1 \leq n \leq N$), hence

$$\ln\{f_{\underline{Y}}(\underline{Y}|\underline{X},\underline{\theta})\} = -\ln \det(\pi G_1^2 I) - \frac{1}{\sigma_1^2}||\underline{Y} - \underline{\mu}(\underline{\theta})||^2, \tag{14}$$

and

$$Q(\underline{\theta}|\underline{\theta}') = K - \frac{1}{\sigma_1^2}||\hat{\underline{Y}} - \underline{\mu}(\underline{\theta})||^2 \tag{15}$$

where K represents terms independent of $\underline{\theta}$ and

$$\hat{\underline{Y}} = E\{\underline{Y}|\underline{X}, \underline{\theta}'\} = \underline{\mu}(\underline{\theta}') + G^H(GG^H)^{-1}(\underline{X} - G\underline{\mu}(\underline{\theta}')). \tag{16}$$

Equation (15) may be rewritten as

$$Q(\underline{\theta}|\underline{\theta}') = K - \frac{1}{\sigma_1^2}\sum_{n=1}^{N}\sum_{i=1}^{I}||\hat{\underline{Y}}_n(\omega_i) - \underline{a}_i(\underline{\theta}_n)S_n(\omega_i)||^2, \tag{17}$$

and using the block diagonal structure of G, equation (16) becomes:

$$\hat{\underline{Y}}_n(\omega_i) = \underline{a}_i(\underline{\theta}'_n)S'_n(\omega_i) + \frac{1}{N}[\underline{X}(\omega_i) - A'(\omega_i)\underline{S}'(\omega_i)] \tag{18}$$

The proposed EM algorithm may be summarized as follows:

(a) Guess initial values for the parameters $\{\alpha_{mn}\}$ and $\{\tau_{mn}\}$, and construct the matrices $A(\omega_i)$, i=1,2,...,I. Compute initial estimates for $\underline{S}(\omega_i)$ using (6).

(b) E-step: Substitute in (18) the current estimates of the parameters

and compute $\underline{\hat{Y}}_n^{(p)}(\omega_i)$ for $1 \leq n \leq N$, $1 \leq i \leq I$.

(c) M-step: Find the maximum of (17) for each $\underline{\theta}_n$. This is simply:

$$\underline{\theta}_n^{(p+1)} = \arg \max_{\underline{\theta}_n} \sum_{i=1}^{I} |\underline{a}_i^H(\underline{\theta}_n)\underline{\hat{Y}}_n^{(p)}(\omega_i)|^2 / ||\underline{a}_i(\underline{\theta}_n)||^2$$

$$S_n^{(p+1)}(\omega_i) = \underline{a}_i^H(\underline{\theta}_n^{(p+1)})\underline{\hat{Y}}_n(\omega_i) / ||\underline{a}_i(\theta_n^{(p+1)})||^2$$

(d) Check convergence of $\{\underline{\theta}_n\}$. If not: go to step (b).

If yes: done.

Observe that the EM algorithm presented here solves iteratively the original maximization problem over the parameters of N signals. At each iteration we have to solve N reduced maximization problems, one for each signal. However, even the reduced maximization problem specified by the M-step is not trivial, since 2(M-1) parameters $\{\alpha_{mn}\}$, $\{\tau_{mn}\}$ are not known. Therefore, in the following sections we will attempt to reduce further the computational requirements of each iteration.

II.3 Further Simplification of the EM Procedure

In this section we show how the parameters of each source can be estimated with minimal effort. We first define two vectors:

$$\underline{\alpha}_n = (\alpha_{1n}, \alpha_{2n}, \ldots, \alpha_{mn})^T, \quad \underline{\tau}_n = (\tau_{1n}, \tau_{2n}, \ldots, \tau_{mn})^T,$$

which we call the intensity vector and the delay vector of the n-th source.

Note that one can always choose the first component of $\underline{\tau}_n$ to be 0, and $||\underline{\alpha}_n|| = 1$ without loss of generality. This is true since we have extra degrees of freedom due to the estimation of both $\underline{\alpha}_n$ and $\{S_n(\omega_1)\}$. The n-th column of $A(\omega_i)$ is a function of both $\underline{\alpha}_n$ and $\underline{\tau}_n$, however, these parameters separate as follows:

$$\underline{a}_i(\theta_n) = \Gamma_i(\underline{\tau}_n)\underline{\alpha}_n \tag{19}$$

where $\Gamma_i(\underline{\tau}_n)$ is a diagonal matrix defined by

$$\Gamma_i(\underline{\tau}_n) = \text{diag}(1, e^{-j\omega_i\tau_{2n}}, \ldots, e^{-j\omega_i\tau_{mn}}),$$

Now, the maximization problem in the M-step becomes:

$$\underline{\theta}_n^{(p+1)} = \underline{\alpha}_n^{(p+1)}, \ \underline{\tau}_n^{(p+1)} = \arg\max_{\underline{\theta}_n} \sum_{i=1}^{I} \underline{a}_i^H(\underline{\theta}_n)\underline{Y}_n^{(p)}(\omega_i)(\underline{Y}_n^{(p)}(\omega_i))^H\underline{a}_i(\underline{\theta}_n)$$

$$= \arg\max_{\underline{\tau}_n, \underline{\alpha}_n} \underline{\alpha}_n^T[\sum_{i=1}^{I} \Gamma_i^H(\underline{\tau}_n)\underline{Y}_n^{(p)}(\omega_i)(\underline{Y}_n^{(p)}(\omega_i))^H\Gamma_i(\underline{\tau}_n)]\underline{\alpha}_n.$$

The solution of the above maximization problem is given by:

$$\underline{\tau}_n^{(p+1)} = \arg \max_{\underline{\tau}_n} \lambda^{max}\{C(\underline{\tau}_n)\} \tag{20a}$$

$$\underline{\alpha}_n^{(p+1)} = \underline{U}^{max} \tag{20b}$$

where $\lambda^{max}\{C(\underline{\tau}_n)\}$ is the largest eigenvalue of the matrix $C(\underline{\tau}_n)$ defined by

$$C(\underline{\tau}_n) = Re\{\sum_{i=1}^{I} \underline{\Gamma}_i^H(\underline{\tau}_n)\underline{Y}_n^{(p)}(\omega_i)(\underline{Y}_n^{(p)}(\omega_i))^H\underline{\Gamma}_i(\underline{\tau}_n)\}$$

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

Equation (20) requires a search over (M-1) parameters (the components of $\underline{\tau}_n$). However, even this problem can be reduced by recalling that the delay parameters are not independent. One can express each of the delays in $\underline{\tau}_n$ as a function of only two or three source location parameters. This is true since the delays are only a function of the distance between the source and the sensor (we assume that the speed of propagation in the medium is known and that the sensor location is known). Now, the search is limited to a three dimensional search over all possible individual source locations. If one is interested in the planar case or azimuth only system, the search is confined to only two or one dimension, respectively.

Finally, we note that the above method provides a very useful tool for estimating the vectors $\{\underline{\alpha}_n\}$ which in turn provide valuable information regarding the directional properties of the sources and/or the sensors, and also might be used to evaluate the attenuation of the medium in various

directions.

III. Direct Maximum Likelihood Approach

In this section we present a novel and efficient algorithm for computing the maximum likelihood estimates of multiple signals observed by an array of sensors. The algorithm provides estimates of parameters related to the directional patterns of the sources $\{\underline{\alpha}_n\}$ as well as estimates of the location parameters of the sources $\{\underline{\tau}_n\}$. 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. In this section we concentrate on the wideband case. The modification for the narrowband case is straightforward, and can be found in LIDS-P-1670.

We basically want to find a solution for equation (5). Relation (6) enables us to update the estimates $\underline{\hat{S}}(\omega_i)$ whenever we have new estimates for $A(\omega_i)$. The main principle of the algorithm is to perform successive minimization operations on the parameters of each signal, holding all the rest of the parameters fixed. For example, suppose that we want to perform a minimization with respect to the k-th signal parameters, then Q can be rewritten as

$$Q = \sum_{i=1}^{I} \left|\left| \underline{Y}^k(\omega_i) - \underline{a}_i(\underline{\theta}_k) S_k(\omega_i) \right|\right|^2 \tag{21}$$

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

$$\underline{Y}^{k}(\omega_i) = \underline{X}(\omega_i) - A(\omega_i)\underline{S}^{k}(\omega_i) \tag{22}$$

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

The minimization of (21) with respect to $\underline{\theta}k$, using (6) with $A(\omega_i)$ replaced by $\underline{a}_i(\underline{\theta}_k)$, is given by

$$\hat{\underline{\theta}}_k = \arg\min_{\underline{\theta}_k \varepsilon \Theta} \sum_{i=1}^{I} \left|\left|\underline{Y}^{k}(\omega_i) - \underline{a}_i(\underline{\theta}_k)[\underline{a}_i^{H}(\underline{\theta}_k)\underline{a}_i(\theta_k)]^{-1}\underline{a}_i^{H}(\underline{\theta}_k)\underline{Y}^{k}(\omega_i)\right|\right|^2$$

$$= \arg\max_{\underline{\theta}_k \varepsilon \Theta} \sum_{i=1}^{I} |(\underline{Y}^{k}(\omega_i))^{H}\underline{a}_i(\underline{\theta}_k)|^2 / ||\underline{a}_i(\underline{\theta}_k)||^2 \tag{23}$$

We now apply the assumption $||\underline{a}_i(\underline{\theta}_k)||^2 = 1$ and the decomposition (19) to obtain:

$$\hat{\underline{\tau}}_k = \arg\max \lambda^{\max}\{R_k\} \tag{24a}$$

$$\hat{\underline{\alpha}}_k = \underline{V}^{\max}$$

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

$$R_k = \mathrm{Re}\{\sum_{i=1}^{I} \Gamma_i^{H}(\underline{\tau}_k)\underline{Y}^{k}(\omega_i)(\underline{Y}^{k}(\omega_i))^{H}\Gamma_i(\underline{\tau}_k)\} \tag{25}$$

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

The maximization described by (24a) 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 $\{\underline{\tau}_n\}$, $\{\underline{\alpha}_n\}$. Set k=1.

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

(c) Compute R_k according to (22) and (25).

(d) Find $\hat{\underline{\tau}}_k$, $\hat{\underline{\alpha}}_k$ according to (24).

(e) Update the k-th column of $A(\omega_i)$ for $1 \leq i \leq I$ with the new $\hat{\underline{\tau}}_k$, $\hat{\underline{\alpha}}_k$; set k=k+1, if k>N then k=1.

(f) Check the convergence of $\{\underline{\alpha}_n\}_{n=1}^N$, $\{\underline{\tau}_n\}_{n=1}^N$. 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 estimates of $\underline{\alpha}_n$, $\underline{\tau}_n$ and on the structure of Q, the local minimum may or may not coincide with the global minimum.

This algorithm may be viewed as a modification of a special case of the EM algorithm. According to the theory of the EM method, 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 $\underline{S}(\omega_i)$ using (6) in step (b), $\underline{S}(\omega_i)$ is updated by replacing only the k-th component by the estimates, $\underline{a}_i^H(\underline{\theta}_k)\underline{Y}^k(\omega_i)$, which can be computed in step (d), following the computation of $\hat{\underline{\tau}}_k$ and $\hat{\underline{\alpha}}_k$. Note that $\underline{a}_i^H(\underline{\theta}_k)\underline{Y}^k(\omega_i)$ is simply the value of $S_k(\omega_i)$ that minimizes (21) whenever $\underline{a}_i(\underline{\theta}_k)$ is known. It is clear that the

last procedure typically will require more iterations than the proposed procedure since the updating of $\underline{S}(\omega_i)$ is done without using all the currently available information.

IV. EIGENSTRUCTURE APPROACH FOR ARRAY PROCESSING WITH UNKNOWN INTENSITY COEFFICIENTS

The eigenstructure approach for array processing is examined for the general case in which it is required to estimate parameters related to the directional patterns of the sources $\{\underline{\alpha}_n\}$ as well as parameters related to the location of the sources $\{\underline{\tau}_n\}$. In recent years there has been a growing interest in eigenstructure based methods, perhaps due to their applicability to general array configurations and due to their simplicity and relative efficiency. A comprehensive discussion of the method may be found in [3], while [1] contains a literature survey of most of the recently published results.

An assumption common to all previously published contributions in this area is that any given source is observed by all the sensors with the same intensity. This assumption is reasonable only if the sources are in the far-field of the array and the sensors have identical radiation patterns. In this report we remove this rather restrictive assumption and thus extend the applicability of the eigenstructure approach to the case of near-field sources and/or sensors with unknown radiation patterns.

Since there are more than one extension of the eigenstructure approach to wideband signals we concentrate here on the narrowband case. The modification for each of the wideband extensions described in [9]-[15] is straightforward.

The following assumptions are made:

(a) The signals and noises are stationary over the observation interval.

(b) The number of sources in known and it is less than the number of sensors.

(c) The columns of A, in equation (4), are linearly independent.

(d) The signals are not completely correlated.

(e) The noise covariance matrix is known except for a multiplicative constant σ^2.

Recalling that equation (4) describes the narrowband case, the correlation matrices of the signal, noise and observation vectors are given by

$$R_s = E\{\underline{S}\ \underline{S}^H\}$$

$$\sigma^2\Sigma_0 = E\{\underline{N}\ \underline{N}^H\}$$

$$R_x = E\{\underline{X}\ \underline{X}^H\} = AR_sA^H + \sigma^2\Sigma_0 \tag{26}$$

where $(\ )^H$ represents the Hermitian transpose operation. The following theorem form the basis for the eigenstructure approach.

Theorem: Let λ_k and $\underline{u}_k$, k=1,2,...,M be the eigenvalues and corresponding eigenvectors of the matrix pencil (R_x, Σ_0), with λ_k in decending order. Then,

1) $\lambda_{N+1} = \lambda_{N+2} = \ldots = \lambda_M = \sigma^2$.

2) Each of the columns of A is orthogonal to the matrix $U = [\underline{u}_{N+1}, \underline{u}_{N+2}, \ldots, \underline{u}_M]$.

Proof: See [14].

This theorem suggests that reasonable estimates of the parameters

$\{\underline{\theta}_n\}_{n=1}^N$ may be obtained by first generating an estimate $\hat{\underline{U}}$ of U and then searching over all possible values of $\underline{\theta}_n$ for vectors $\underline{a}(\underline{\theta}_n)$ that are nearly orthogonal to $\hat{\underline{U}}$. This may be written as

$$\hat{\underline{\theta}}_n = \arg \min_{\underline{\theta}_n} \left|\left|\hat{\underline{U}}^H \underline{a}(\underline{\theta}_n)\right|\right|^2 \tag{27}$$

where $||\cdot||$ denotes the Euclidean norm. Since there is an extra degree of freedom, there is no loss of generality in assuming that $||\underline{a}(\underline{\theta}_n)|| = 1$. This also eliminates the trivial solution of (27). Note that (27) requires a multidimensional search over the parameters $\{\alpha_{mn}\}$ and $\{\tau_{mn}\}$. To overcome this difficulty we decompose $\underline{a}(\underline{\theta}_n)$ as follows:

$$\underline{a}(\underline{\theta}_n) = \Gamma(\underline{\tau}_n) \cdot \underline{\alpha}_n$$

where

$$\underline{\alpha}_n = (\alpha_{1n}, \alpha_{2n}, \ldots, \alpha_{mn})^T$$

$$\Gamma(\underline{\tau}_n) = \text{diag}(e^{-j\omega_1\tau_{1n}}, e^{-j\omega_1\tau_{2n}}, \ldots, e^{-j\omega_1\tau_{mn}})$$

$$\underline{\tau}_n = (\tau_{1n}, \tau_{2n}, \ldots, \tau_{mn})^T.$$

Using this notation, (27) becomes

$$\hat{\underline{\theta}}_n = \arg \min_{\underline{\alpha}_n, \underline{\tau}_n} \underline{\alpha}_n^T \Gamma^H(\underline{\tau}_n) \hat{U}\hat{U}^H \Gamma(\underline{\tau}_n) \underline{\alpha}_n, \tag{28}$$

and hence

$$\hat{\underline{\tau}}_n = \arg \min_{\underline{\tau}_n} \delta^{\min}\{C(\underline{\tau}_n)\}, \tag{29a}$$

$$\hat{\underline{\alpha}}_n = \underline{W}^{\min} \tag{29b}$$

where $\delta^{\min}\{C(\underline{\tau}_n)\}$ is the smallest eigenvalue of the matrix $C(\underline{\tau}_n)$ given by

$$C(\underline{\tau}_n) = \mathrm{Re}\{\Gamma^H(\underline{\tau}_n) \hat{U}\hat{U}^H \Gamma(\underline{\tau}_n)\}, \tag{30}$$

and $\underline{W}^{\min}$ is the associated normalized eigenvector. Equation (29) requires a simple search over the space of vectors $\underline{\tau}_n$, induced by all possible individual source locations.

The proposed algorithm may be summarized as follows:

(a) Estimate the observation covariance matrix:

$$\hat{R}_x = \frac{1}{J} \sum_{j=1}^{J} \underline{X}(j) \underline{X}(j)^H.$$

(b) Find the M-N eigenvectors, $\{\hat{\underline{u}}_k\}$, corresponding to the smallest M-N eigenvalues of the pencil $(\hat{R}_x, \Sigma_0)$, and construct the matrix:

$$\hat{\underline{U}} = [\hat{\underline{u}}_{N+1}, \hat{\underline{U}}_{N+2}, \ldots, \hat{\underline{u}}_{M}].$$

(c) Evaluate, for all possible source locations, the "spatial spectrum" given by:

$$P(\underline{\tau}) = \frac{1}{\delta^{\min}\{C(\underline{\tau})\}}$$

where $C(\underline{\tau})$ is defined by (30).

(d) Select the N highest peaks of $P(\underline{\tau})$. The corresponding values of $\underline{\tau}$ describe the source locations, and the corresponding eigenvectors describe the intensity vectors $\{\underline{\alpha}_n\}$.

Examples

To illustrate the behavior of the algorithm, let us consider two examples:

Example 1. Consider a uniform linear array of five sensors separated by half a wavelength of the actual narrowband source signals. The sources are two narrowband emitters located in the farfield 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 $\tau_{nm} = (m-1)\pi\sin(\gamma_n)$. The first source at a bearing of -9 degrees was observed with the intensity vector $\underline{\alpha}_1^T = [1,1,1,1,1]$, the second source at a bearing of 11 degrees was observed with $\underline{\alpha}_2^T = [1,.8,.6,.4,.2]$. 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 100 independent samples at a SNR of 20 dB. The spatial spectrum, $P(\gamma)$, is plotted versus the angle of arrival (bearing) in Figure 1. Two very sharp peaks are observed at -9 degrees and 11^{o} degrees. The associated estimates of the intensity vectors are $\hat{\underline{\alpha}}_1^T$ = (.99,1.0,.99,.99,.99) and $\hat{\underline{\alpha}}_2^T$ = (1.0,.79,.59,.40,.20). The spurious peak at 3 degree is associated with $\hat{\underline{\alpha}}^T$ = (1.0,0.7,0.3,-0.1,-0.3) and therefore can be easily eliminated, since under our assumptions the α_{mn}'s must be positive. For comparison, we plotted the result of the MUSIC algorithm [3] in Figure 2. Since only one source conforms with the assumptions of MUSIC, only one peak is observed.

Example 2.

Consider Example 1 except that here the SNR = 50 dB, γ_1=11 degrees γ_2=25 degrees and $\underline{\alpha}_1^T = \underline{\alpha}_2^T$ = (1,1,1,1,1). The spatial spectrum is plotted in Figure 3. We observe 2 peaks at 11 and 25 degrees and 3 more spurious peaks. The two peaks on the leftside are associated with nonphysical intensity vectors and therefore can be eliminated by post processing. The spurious peak at 18 degrees is associated with $\hat{\underline{\alpha}}^T$ = (.77,.94,1.0,.92,.73) and therefore is an ambiguous solution. Ambiguous solutions occur whenever the continuum $\underline{a}(\underline{\theta})$ ("array manifold") intersects the signal subspace (the space generated by the columns of A) in more than N points [3].

Conclusions

In this section the eigenstructure approach has been used to obtain estimates of source locations as well as estimates of the intensity vectors $\{\underline{\alpha}_n\}$, simultaneously. The estimates of $\{\underline{\alpha}_n\}$ may be useful in their own right, but their estimation is essential, even if one is only interested in the source locations, 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 calibration and a large memory).

We observed that in some cases post-processing is required to eliminate spurious solutions. The post processing decisions may rely on the sign of the intensity vectors $\{\hat{\underline{\alpha}}_n\}$ and on any prior knowledge concerning the source locations and expected intensity vectors.

V. Nonuniform Array Processing Via the Polynomial Approach

Recently an effective technique for computing the maximum likelihood (ML) estimates of the signals was introduced by Bresler and Macovski [5] and Kumaresan, Scharf and Shaw [17], [18]. We refer to this technique as the "polynomial approach" since it is based on expressing the ML criterion in terms of the prediction polynomial of the noiseless signals. The polynomial approach relies on the assumption that the array of sensors is uniformly spaced. It is well known [22] that the optimal sensor configuration is not uniform under many reasonable criteria. For example, minimum bearing variance is obtained by placing half of the sensors (with a spacing of half of the design wavelength) at each end of the given aperture; minimum range variance is obtained by placing one fourth- of the element at each end and half in the middle; and optimal position estimation is obtained by placing one third of the sensors at each end and the middle. Furthermore, when operating long uniform arrays, often some of the sensors do not function and their outputs must be ignored, yielding in effect a sublattice array. In this section we present a method for extending the polynomial approach to sublattice arrays. We treat the sublattice array output as an incomplete data observation. Therefore the EM algorithm is directly applicable. This algorithm was only recently applied to array processing problems by Feder and Weinstein [6]. However, in [6] the EM algorithm is used to enable the estimation of one signal at a time, while here it is used to enable the use of the polynomial approach which estimates all the signals simultaneously. Since the polynomial approach is not widely known, the basic principles of this technique are briefly reviewed here for clarity. Note that although we

concentrate on the array problem, all the results are equally applicable to the corresponding time series problem discussed in [5], namely, the estimation of superimposed complex exponential signals in noise.

This section is organized as follows. The polynomial approach for processing data collected over a uniform array is described in V.1. In V.2 it is shown how the EM algorithm can be used to adapt the polynomial approach to the case of sublattice arrays. Several examples of our procedure are presented in Section V.3, and Section V.4 contains some conclusions.

V.1 Uniform Arrays and the Polynomial Approach

Consider N narrowband radiating sources observed by a linear uniform array composed of M sensors. The sources are assumed to be far enough from the array, compared to the array length so that the signal wavefronts are effectively planar over the array. The signal at the output of the m-th sensor can be expressed by

$$x_m(t) = \sum_{n=1}^{N} s_n(t-(m-1)\tau_n) + v_m(t);\ m = 1,2,\ldots,M, \qquad -T/2 \leq t \leq T/2, \tag{31}$$

where $\{s_n(t)\}_{n=1}^{N}$ are the radiated signals, $\{v_m(t)\}_{m=1}^{M}$ are additive noise processes, and T is the observation interval. The delay of the n-th wavefront at the m-th sensor, relative to the first sensor, is given by $(m-1)\tau_n$. The parameter, τ_n, can be expressed in terms of the sensor spacing, d, the propagation velocity, c, and the source bearing, γ_n,

relative to the array perpendicular as

$$\tau_n = (d/c)\sin(\gamma_n).$$

A convenient separation of the parameters $\{\tau_n\}_{n=1}^{N}$ to be estimated maybe obtained by using Fourier coefficients, defined by

$$X_m = \frac{1}{T} \int_{-T/2}^{T/2} x_m(t) e^{-j\omega_0 t} dt .$$

Since we assume that the spectrum of the signals is concentrated around ω_0, with a bandwidth that is small compared to $2\pi/T$, a single Fourier coefficient is enough to completely describe the signals. Taking the Fourier coefficients of (1) we obtain:

$$X_m = \sum_{n=1}^{N} e^{-j\omega_0(m-1)\tau_n} S_n + V_m, \quad m = 1,2,\ldots,M; \tag{32}$$

where S_n and V_m are the Fourier coeffients of $S_n(t)$ and $v_m(t)$ respectively. Equation (32) may be expressed using vector notation as

$$\underline{X} = A\underline{S} + \underline{V}, \tag{33}$$

where

$$\underline{X} = [X_1, X_2, \ldots, X_M]^T,$$

$$\underline{S} = [S_1, S_2, \ldots, S_N]^T,$$

$$\underline{V} = [V_1, V_2, \ldots, V_M]^T,$$

$$A = [\underline{a}_1, \underline{a}_2, \ldots, \underline{a}_N],$$

$$\underline{a}_n = [1, \lambda_n, \lambda_n^2, \ldots, \lambda_n^{M-1}]^T, \quad n=1,2,\ldots,N;$$

$$\lambda_n = e^{-j\omega_0\tau_n} .$$

In many cases estimation is based on more than one realization of equation (33), corresponding for example to several time samples or observation intervals. In that case we use the index j to denote the different realizations:

$$\underline{X}_j = A\underline{S}_j + \underline{V}_j \qquad j = 1,2,\ldots,J. \tag{34}$$

Instead of estimating $\{\tau_n\}$ directly we concentrate on estimating $\{\lambda_n\}_{n=1}^N$. Under the assumption that the vectors $\{\underline{V}_j\}_{j=1}^J$ are i.i.d. zero mean and Gaussian with covariance $\sigma^2 I$, the maximum likelihood estimates are given by

$$\{\hat{\lambda}_n\}_{n=1}^N = \arg\min_{\lambda \varepsilon UC} \{R\}, \quad R = \sum_{j=1}^{J} ||\underline{X}_j - A\underline{S}_j||^2 \tag{35}$$

where $||\cdot||$ denotes the Euclidean norm and UC stands for the unit circle which is the parameter space, in this case.

The minimization required in (35) is not trivial since the vectors $\{\underline{S}_j\}$ and the matrix A are not known to the observer. However, whenever A is known, R is minimized by choosing

$$\hat{\underline{S}}_j = (A^H A)^{-1} A^H \underline{X}_j \tag{36}$$

as the estimate of $\underline{S}_j$, for j=1,2,...,J, where $(\ )^H$ represents the Hermitian-transpose operation. Substituting (36) in (37) we obtain:

$$R = \sum_{j=1}^{J} \left\| \underline{X}_j - A(A^H A)^{-1} A^H \underline{X}_j \right\| = \sum_{j=1}^{J} \underline{X}_j^H P_B \underline{X}_j, \tag{37}$$

where

$$P_B = I - A(A^H A)^{-1} A^H.$$

The polynomial approach relies on the introduction of the polynomial $b(z) = b_0 z^N + b_1 z^{N-1} + \ldots + b_N$, whose zeros are the parameters of interest $\{\lambda_n\}_{n=1}^{N}$. Observe that by definition the Mx(M-N) Toeplitz matrix B defined by

$$B^H = \begin{bmatrix} b_N & b_{N-1} & \cdots & b_0 & & \\ & b_N & b_{N-1} & \cdots & b_0 & \\ & & \ddots & \ddots & & \ddots \\ & & & b_N & b_{N-1} & \cdots & b_0 \end{bmatrix}$$

is orthogonal to A, i.e. $B^H A = 0$, and hence $P_B = B(B^H B)^{-1} B^H$. Now the minimization in (35) can be expressed in terms of the coefficients $\{b_i\}_{i=0}^N$ as

$$\underline{b} = \arg \min_{\underline{b} \in \Theta_b} \sum_{j=1}^{J} \underline{x}_j^H B(B^H B)^{-1} B^H \underline{x}_j, \tag{38}$$

where $\underline{b} = [b_N, b_{N-1}, \ldots, b_0]^T$, and Θ_b is the space of all the vectors whose associated polynomials have zeros only on the unit circle. It can be shown that since b(z) has its roots on the unit circle, its coefficient vector is α-conjugate-symmetric; i.e. $\underline{b} = \alpha[b_0, b_1, \ldots, b_N]^H$ where α is a constant of unit modulus.

The algorithm for the minimization required in (38) is based on the relation

$$B^H \underline{X}_j = \tilde{X}_j \underline{b}, \tag{39}$$

where $\tilde{X}_j$ is the (M-N)x(N+1) matrix defined by:

$$\tilde{\underline{X}}_j = [\underline{X}_j(N+1:M),\ \underline{X}_j(N:M-1),\ldots,\underline{X}_j(1:M-N)],$$

and $\underline{X}_j(k:r)$ describes a subvector of $\underline{X}_j$ consisting of all of the components from the k-th component to the r-th component. Substituting (39) in (38) we obtain:

$$\underline{b} = \arg \min_{b\varepsilon\Theta_b} \underline{b}^H C\, \underline{b}, \quad C = \sum_{j=1}^{J} \tilde{\underline{X}}_j^H (B^H B)^{-1} \tilde{\underline{X}}_j. \tag{40}$$

This relation is used in the minimization algorithm [5], [16]-[18]. The algorithm starts with any initial estimate $\underline{b}^{(0)}$ of $\underline{b}$ and proceeds as follows:

a) Initialization k=0, $\underline{b}=b^{(0)}$

b) Compute $C^{(k)}$ according to (40) using $\underline{b}^{(k)}$ to construct the matrix $B^{(k)}$.

c) Find $\underline{b}^{(k+1)} = \min_{b\varepsilon\Theta_b} \underline{b}^H C^{(k)} \underline{b}$

d) Check convergence. NO - k = k+1, go to (b).

YES - Continued.

e) Find the roots of the polynomial $b^{(k+1)}(z)$ whose coefficients are given by $\underline{b}^{(k+1)}$.

In [1] the relation $\underline{b} = \alpha[b_0, b_1, \ldots, b_N]^H$ was incorporated in step (c) to yield a simple quadratic minimization problem. We now turn to the more practical situation of nonuniform arrays.

V.2 Sublattice Arrays and the EM Algorithm

In this paper we are primarily interested in the problem where the measurements are taken along a sublattice arrays of M' sensors. The sublattice array may be described by a binary vector, $\underline{1}$, of length M. The m-th component of $\underline{1}$ is 1 if the m-th sensor of the full array is part of the subarray, and it is zero if the sensor is missing. Equation (34) may be converted to describe a sublattice array through a left-multiplication by a transformation matrix G. The M'xM matrix G is constructed by eliminating all the zero rows in diag($\underline{1}$). For example an array of three elements in positions 1,2,5 is described by $\underline{1}^T = (1,1,0,0,1)$ and

$$G = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

Multiplying equation (34) by G we obtain, for a given sublattice array, the equation:

$$\underline{Y}_j = G\underline{X}_j = G(A\underline{S}_j + \underline{V}_j), \quad j = 1,2,\ldots,J \tag{41}$$

We refer to $\{\underline{X}_j\}$ as the (unavailable) "complete data" and to $\{\underline{Y}_j\}$ as the observed data. Let $\underline{Y} = [\underline{Y}_1^T, \underline{Y}_2^T, \ldots, \underline{Y}_J^T]^T$ denote the observation vector and $\underline{\theta} = [A\underline{S}_1, A\underline{S}_2, \ldots, A\underline{S}_J]$ represent the parameter vector. If $f_{\underline{Y}}(\underline{Y}|\underline{\theta})$ is the conditional probability density function of $\underline{y}$ given $\underline{\theta}$, then the maximum likelihood estimate of $\underline{\theta}$ is:

$$\hat{\underline{\theta}} = \arg\max_{\underline{\theta}\in\Theta} f_{\underline{Y}}(\underline{Y}|\underline{\theta}) = \arg\max_{\underline{\theta}\in\Theta} \ln\{f_{\underline{Y}}(\underline{Y}|\underline{\theta})\} \tag{42}$$

where Θ is the parameter space. In order to use the polynomial approach it would be useful to express $\underline{Y}$ in terms of the complete data vector $\underline{X} = [\underline{X}_1^T, \underline{X}_2^T, \ldots, \underline{X}_J^T]^T$. This relation is given by

$$\underline{Y} = F\underline{X} , \tag{43}$$

where F is the block diagonal matrix with J blocks:

$$F = \begin{bmatrix} G & & 0 \\ & G & \\ & & \ddots \\ 0 & & G \end{bmatrix}.$$

The application of the EM algorithm to the problem at hand requires only the determination of $Q(\underline{\theta}|\underline{\theta}')$. We recall that $\underline{X}$ is Gaussian with given covariance $\sigma^2 I$ and unknown mean $\underline{\theta}$, where $\underline{\theta} = (\underline{\theta}_1^T, \underline{\theta}_2^T, \ldots, \underline{\theta}_J^T)^T$ and

$$\underline{\theta}_j = A\underline{S}_j \qquad j = 1,2,\ldots,J.$$

Hence,

$$\ln\{f_{\underline{X}}(\underline{X}|\underline{\theta})\} = -M\cdot J\cdot \ln\{\pi\sigma^2\} - \frac{1}{\sigma^2}(||\underline{X}||^2 + ||\underline{\theta}||^2 - \underline{\theta}^H\underline{X} - \underline{X}^H\underline{\theta}).$$

Therefore,

$$Q(\underline{\theta}|\underline{\theta}') = K - \frac{1}{\sigma^2}(||\underline{\theta}||^2 - \underline{\theta}^H\hat{\underline{X}} - \hat{\underline{X}}^H\underline{\theta}) \tag{44}$$

where K represents all the terms independent of $\underline{\theta}$ and

$$\hat{\underline{X}} = E\{\underline{X}|\underline{Y},\underline{\theta}'\} = \underline{\theta}' + F^H(FF^H)^{-1}(\underline{Y}-F\underline{\theta}') \quad . \tag{45}$$

Note that the specific $\underline{\theta}$ that maximizes (17) is the same $\underline{\theta}$ that minimizes the functional

$$R_1 = ||\hat{\underline{X}}-\underline{\theta}||^2 = \sum_{j=1}^{J} ||\hat{\underline{X}}_j - \underline{\theta}_j||^2 = \sum_{j=1}^{J} ||\hat{\underline{X}}_j - A\underline{S}_j||^2. \tag{46}$$

Hence, the M-step of the EM algorithm maybe performed by using the polynomial approach for minimizing (46).

The following relations are useful for the actual proposed procedure. Using the block diagonal structure of F and the relations $GG^H = I$ and $G^HG = \text{diag}(\underline{1})$, equation (45) maybe rewritten as:

$$\hat{\underline{X}}_j = \text{diag}(\bar{\underline{1}})\underline{\theta}_j' + G^H\underline{Y}_j \tag{47}$$

where $\bar{\underline{1}}$ is the complement of $\underline{1}$ (zeros and ones are interchanged). The parameter vector $\underline{\theta}_j'$ is simply the estimate of $A\underline{S}_j$ obtained in the previous cycle and therefore (47) may be written also as:

$$\overset{\bullet}{\underline{X}}_j^{(p+1)} = \text{diag}(\underline{\bar{1}})\{A(A^HA)^{-1}A^H\overset{\bullet}{\underline{X}}_j\}^{(p)} + G^H\underline{Y}_j = \text{diag}(\underline{\bar{1}})\{(I-B(B^HB)^{-1}B^H)\overset{\bullet}{\underline{X}}_j\}^{(p)} +$$

$$+ G^H\underline{Y}_j,$$

using the notation of the polynomial approach. As one would expect, equation (47) states that the components of $\overset{\bullet}{\underline{X}}_j$ that correspond to existing sensors are always equal to the observed data, i.e. the corresponding components of $\underline{Y}_j$.

The proposed EM algorithm may be summarized as follows:

(a) Initialization: Select initial values for $\{\lambda_n\}_{n=1}^N$; find the corresponding $\underline{b}^{(0)}$.

Compute: $A_1=GA$; $\underline{S}_j=(A_1^HA_1)^{-1}A_1^H\underline{Y}_j$;

$$\overset{\bullet}{\underline{X}}_j^{(0)} = \text{diag}(\underline{\bar{1}})A\underline{S}_j + G^H\underline{Y}_j \qquad \text{(See (47))}$$

Set: p=0.

(b) Use the minimization algorithm for uniform arrays:

(b.1) Constuct $\tilde{X}_j = [\underset{\sim}{\hat{X}}_j^{(p)}(N+1:M),\ldots,\underset{\sim}{\hat{X}}_j^{(p)}(1:M-N)]$;

set k=0, $\underset{\sim}{b}_1^{(0)} = \underset{\sim}{b}^{(p)}$.

(b.2) Construct B using $\underset{\sim}{b}_1^{(k)}$.

Compute $C = \sum_{j=1}^{J} \tilde{x}_j^H (B^H B)^{-1} \tilde{X}_j$.

(b.3) Compute $\underset{\sim}{b}_1^{(k+1)} = \arg\min_{b_1 \varepsilon \Theta_b} \underset{\sim}{b}_1^H C \underset{\sim}{b}_1$.

(b.4) Check convergence of $\underset{\sim}{b}_1$. No - k=k+1; go to (b.2).

Yes - $\underset{\sim}{b}^{(p)} = \underset{\sim}{b}_1^{(k+1)}$; continue.

(c) Construct B using $\underset{\sim}{b}^{(p)}$

Compute:

$$\underset{\sim}{\hat{X}}_j^{(p+1)} = \text{diag}(\underline{\bar{1}})(I - B(B^H B)^{-1} B^H)\underset{\sim}{\hat{X}}_j^{(p)} + G^H \underset{\sim}{Y}_j.$$

(d) Check the convergence of $\hat{\tilde{X}}_j$. No - p=p+1, go to (b).

Yes - continue.

(e) Find the roots of the polynomial $\underset{\sim}{b}^{(p)}(z)$ whose coefficients are given by $\underset{\sim}{b}^{(p)}$.

V.3 Examples

To illustrate the behavior of the algorithm, let us consider two examples:

Example 1: Consider a uniform linear array of 6 sensors separated by half a wavelength of the actual narrowband source signals. Now, assume that the two middle sensors are missing (i.e. $\underline{l}^T = (1\ 1\ 0\ 0\ 1\ 1)$); this is the optimal configuration for bearing estimation when the given aperture is 2.5 wavelengths and the number of sensors is limited to 4.

The sources are two narrowband emitters located in the far field of the array. One source is located at a bearing of 10 degrees, while the second source is located at a bearing of 25 degrees. We generated only 10 independent samples with a SNR of 30 dB. The initial guess was $\gamma_1^{(0)} = 3^o$, $\gamma_2^{(0)} = 17^o$. The algorithm converged to within one degree of the right result in 8 iterations, as shown in Table 1.

Example 2: Consider example 1 where the array is reconfigured so that $\underline{l}^T = (1\ 0\ 1\ 0\ 0\ 1)$. Note that only 3 sensors are used and they are separated by one wavelength and 1.5 wavelengths. Nevertheless, the algorithm converged to within one degree of the right result in only 7 iterations, as shown in Table 2. The initial guess was $\gamma_1^{(0)} = 3^o$, $\gamma_2^{(0)} = 35^o$.

V.4 Summary

We have proposed a novel EM algorithm for the estimation of superimposed signals observed by nonuniform arrays. The algorithm is efficient and provides accurate results even when the number of samples is small and the sensors are separated by more than half a wavelength.

Note that convergence theorems exist for the EM method. However, convergence theorems for the polynomial approach are not yet available and therefore further investigation is required to prove the convergence of the proposed technique. Finally we would like to emphasize that the EM algorithm is guaranteed to converge to a local maximum of the likelihood function. Thus we would expect that the algorithm described here will converge to the globally optimum result only if the initial estimates are good enough. Fast initial estimates can be obtained by using simpler methods such as the MLM, MEM or the MUSIC techniques (see [1] for a review of these methods).

Iterations No.	$\hat{\gamma}_1$ degrees	$\hat{\gamma}_2$ degrees
0	3.00	17.0
1	6.15	19.38
2	7.29	20.49
3	8.16	21.47
4	8.78	22.30
5	9.23	22.95
6	9.55	23.46
7	9.77	23.83
8	9.93	24.12
9	10.04	24.32
10	10.12	24.47

Table 1: Evolution of the algorithm for $\underline{1}^T = (1\ 1\ 0\ \ 0\ 1\ 1)$.

Iterations No.	$\hat{\gamma}_1$ degrees	$\hat{\gamma}_2$ degrees
0	3.00	35.00
1	-0.01	18.13
2	3.46	18.27
3	7.18	20.16
4	8.74	21.90
5	9.39	23.01
6	9.68	23.69
7	9.84	24.10
8	9.92	24.35
9	9.96	24.51
10	9.99	24.61

Table 2: Evolution of the algorithm for $\underline{1}^T = (1\ 0\ 1\ 0\ 0\ 1)$.

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," IEEE Proceedings, Vol. 134, No. 1, pp. 53-59, February 1987.

[3] R.O. Schmidt, "A Signal Subspace Approach to Multiple Emitter Location and Spectral Estimation," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1981.

[4] D.H. Johnson and S.R. DeGraff, "Improving the Resolution of Bearing in Passive Sonar Arrays by Eigenvalue Analysis," IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. 30, No. 4, pp. 638-647, August 1982.

[5] Y. Bresler and A. Macovski, "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Signals in Noise," IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. 34, No. 5, pp. 1081-1089, October, 1986.

[6] M. Feder and E. Weinstein, "Parameter Estimation of Composite Signals using the EM Algorithm," preprint.

[7] M. Feder and E. Weinstein, "Optimal Multiple Source Location Estimation Via the EM Algorithm", Proceedings of ICASSP 1985, pp. 1762-1765.

[8] M. Feder and E. Weinstein, "Multipath and Multiple Source Array Processing Via the EM Algorithm," Proceedings of ICASSP 1986.

[9] M. Coker and E. Ferrara, "A New Method for Multiple Source Location," ISASSP 1982, pp. 411-415.

[10] G. Bienvenu, "Eigensystem Properties of the Sampled Space Correlation Matrix," ISASSP 83, pp. 332-335.

[11] G. Su and M. Morf, "The Signal Subspace Approach for Multiple Wideband Emitter Location," IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-31, No. 6, pp. 1502-1522, Dec. 1983.

[12] M. Wax, T. Shan, and T. Kailath, "Spatio-Temporal Spectral Analysis By Eigenstructure Methods," IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-32, No. 4, pp. 817-827, Aug. 1984.

[13] H. Wang and M. Kaveh, "Estimation of Angle of Arrival for Wideband

Sources," ICASSP 84, pp. 7.5.1-7.5.4.

[14] H. Wang and M. Kaveh, "Coherent Signal-Subspace Processing for the Detection and Estimation of Angles of Arrival of Multiple Wideband Sources," IEEE Trans. on Acoustics, Speech and Signal Processing, Vol. ASSP-33, No. 4, pp. 823-831, Aug. 1985.

[15] K.M. Buckley and L.J. Griffiths, "Eigenstructure Based Broadband Source Location Estimation", ICASSP 86, pp. 1869-1872.

[16] Y. Bresler and A. Macovski, "Exact Maximum Likelihood Parameter Estimation of Superimposed Exponential Siganls in Noise," ICASSP 85, pp. 1824-1827.

[17] R. Kumaresan and A.K. Shaw, "High Resolution Bearing Estimation without Eigen Decomposition," ICASSP 85, pp. 576-579.

[18] R. Kumaresan, L.L. Scharf, and A.K. Shaw, "An Algorithm for Pole-Zero Modeling and Spectral Analysis," IEEE Trans. on Acoustics, Speech, and Signal Processing, Vol. ASSP-34, No. 3, pp. 637-640, June 1986.

[19] C.R. Rao, Linear Statistical Inference and Its Applications, New York, Wiley 1965.

[20] A.P. Dempster, N.M. Laird and D.B. Rubin, "Maximum Likelihood From Incomplete Data Via the EM Algorithm", J. Roy. Statist. Soc. Ser. 39, pp. 1-38.

[21] C.F. Jeff Wu, "On the Convergence Properties of the EM Algorithm," The Annals of Statistics, Vol. 11, No. 1, pp. 95-103, 1983.

[22] G.C. Carter, "Coherence and Time Delay Estimation," Proceedings of the IEEE, Vol. 75, No. 2, pp. 236-255, Feb. 1987.

Figure Captions

Figure 1: Spatial spectrum of the proposed procedure for two far-field sources.

Figure 2: Spatial spectrum of the MUSIC procedure for the case of Figure 1.

Figure 3: Spatial spectrum of the proposed procedure for two far-field sources with equal intensity vectors.

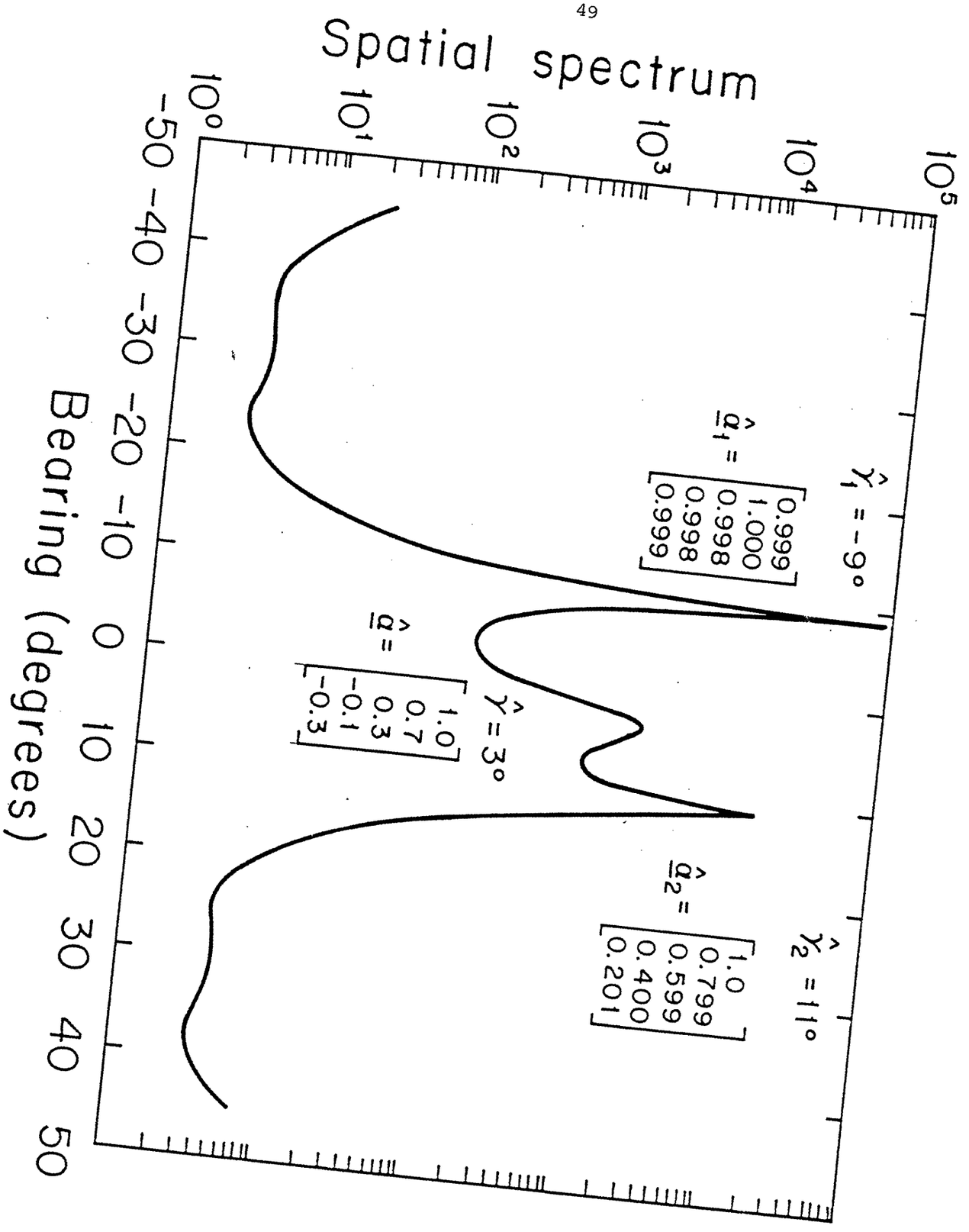
Spatial spectrum
10^0
10^1
10^2
10^3
10^4
10^5
-50 -40 -30 -20 -10 0 10 20 30 40 50
Bearing (degrees)
$\hat{\gamma}_1 = -9°$
$\hat{\underline{a}}_1 = \begin{bmatrix} 0.999 \\ 1.000 \\ 0.998 \\ 0.998 \\ 0.999 \end{bmatrix}$
$\hat{\gamma} = 3°$
$\hat{\underline{a}} = \begin{bmatrix} 1.0 \\ 0.7 \\ 0.3 \\ -0.1 \\ -0.3 \end{bmatrix}$
$\hat{\gamma}_2 = 11°$
$\hat{\underline{a}}_2 = \begin{bmatrix} 1.0 \\ 0.799 \\ 0.599 \\ 0.400 \\ 0.201 \end{bmatrix}$

Figure 1

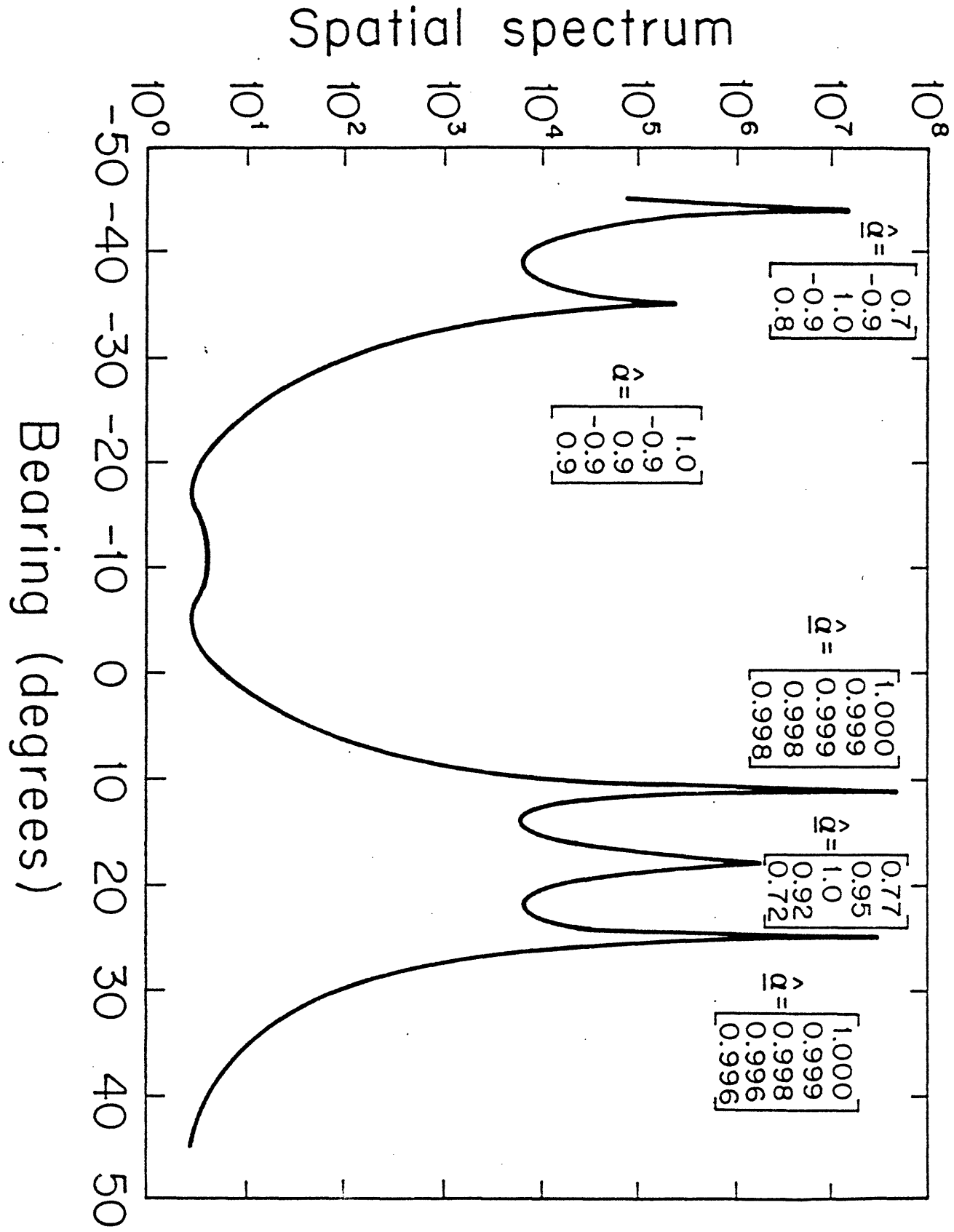
Spatial spectrum
10^0 10^1 10^2 10^3 10^4 10^5 10^6 10^7 10^8
-50 -40 -30 -20 -10 0 10 20 30 40 50
Bearing (degrees)
$\hat{\underline{\alpha}}$= [0.7 -0.9 1.0 -0.9 0.8]
$\hat{\underline{\alpha}}$= [1.0 -0.9 0.9 -0.9 0.9]
$\hat{\underline{\alpha}}$= [1.000 0.999 0.998 0.998]
$\hat{\underline{\alpha}}$= [0.77 0.95 1.0 0.92 0.72]
$\hat{\underline{\alpha}}$= [1.000 0.999 0.998 0.996]

Figure 2

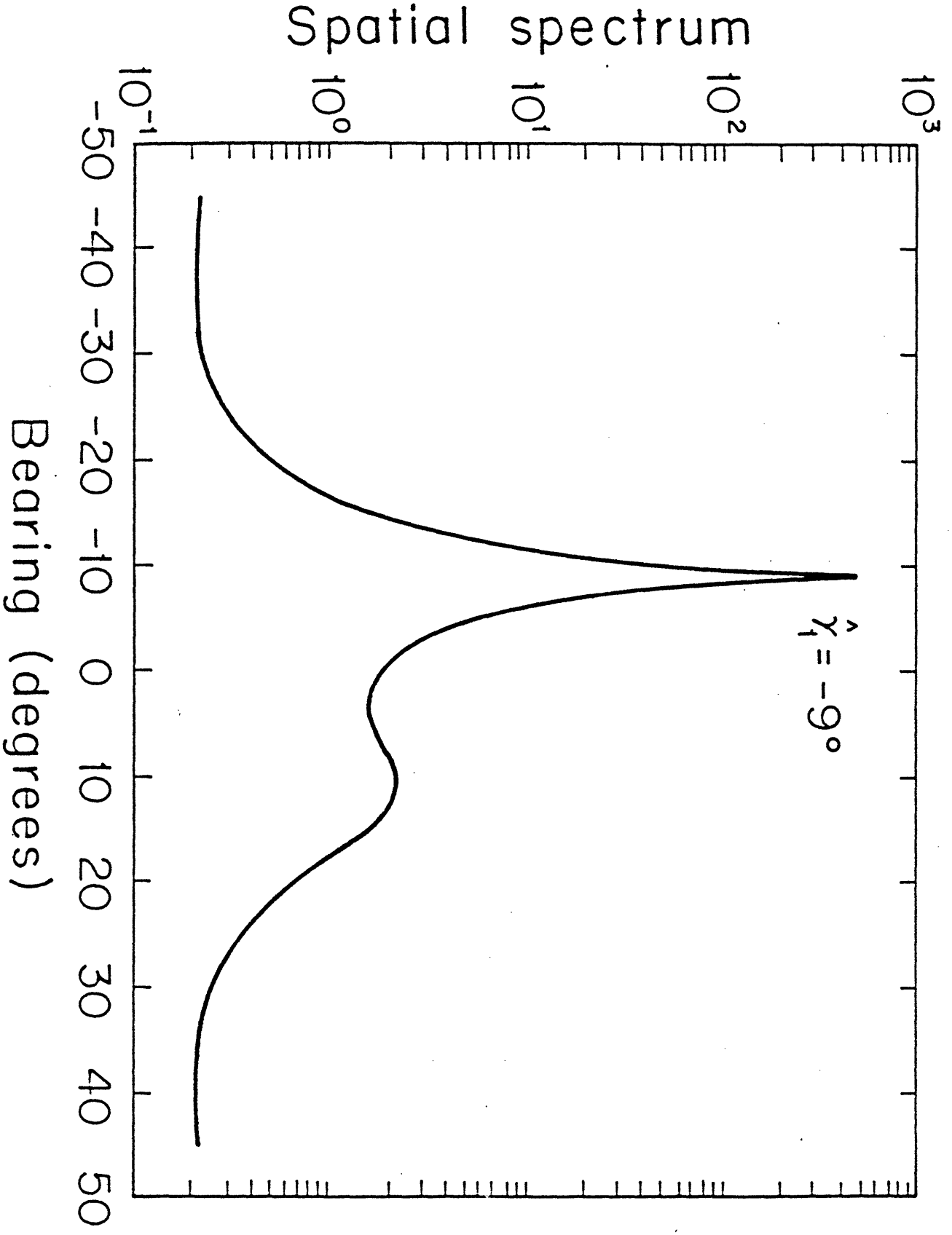
Spatial spectrum
10^{-1}
10^{0}
10^{1}
10^{2}
10^{3}
-50 -40 -30 -20 -10 0 10 20 30 40 50
Bearing (degrees)
$\hat{\gamma}_1 = -9°$

Figure 3