On the Application of the Global Matched Filter to DOA Estimation with Uniform Circular Arrays

Jean-Jacques Fuchs, *Member, IEEE*

***Abstract*—The problem of estimating the direction of arrivals (DOA) of narrowband sources impinging on a uniform circular array is considered. We present a method that uses as input the values of a small number of uniformly spaced beams and apply a model-fitting approach taking into account the statistical properties of the beams. The approach, which is called the "global matched filter," fits simultaneously to the observations all the elements needed to explain them. It chooses, among all the representations satisfying a constraint with a sensible physical interpretation, the one with minimal energy. The method drastically improves upon the conventional beamformer and has performance comparable with the best high-resolution (HR) techniques. It further applies when the number of sources exceeds the number of sensors: a situation that cannot be handled by standard HR techniques.**

***Index Terms*—Direction-of arrival estimation, matched filters, redundant parametrizations, regularized estimation, sparse signal representation.**

I. Introduction

THE MAJOR advantages of uniform circular arrays (UCAs) over other planar arrays are their 360° azimuthal coverage and their almost invariant directional pattern. This is in strong contrast with the widely studied uniform linear array that only covers 180° and whose beampattern significantly broadens when steered away from broadside. Indeed, even for a small number of sensors uniformly distributed around a circle, the fluctuation of the array pattern is extremely small and the azimuthal resolution almost constant, as the effective aperture analogy predicts. These interesting features of UCAs led to the development of experimental systems [1] employing the beamforming principle to obtain DOA estimates. It is known that in its basic form, this approach cannot resolve closely spaced sources (with spacing less than the main-lobe width of the array pattern). High-resolution (HR) techniques such as MUSIC and similar techniques can, of course, be applied to UCAs in a straightforward way and have better performance. Efforts on the applications of uniform linear array (ULA) techniques with UCAs appear in [2] and [3]. Modifications of HR techniques that allow the extension of these procedures to allow for 2-D angle estimations with UCAs have been proposed in [4].

It has been shown recently [5] that applying these HR techniques to a *rectified* estimated covariance matrix of the snapshots further increases the resolving power. *Rectifying* the estimate of the spectral density matrix means projecting it on the manifold in which its true value lies. While, for ULAs, this amounts to project it on the set of positive-definite Toeplitz matrices, for UCAs, this projection is slightly more intricate.

We show in Appendix A that forming a beam using the optimaly rectified matrix or the initial matrix yields the same value. This indicates that the conventional beamformer (CBF) somehow achieves this projection and already benefits from it. This also means that working with the outputs of the CBF is an excellent starting point for a localization algorithm, as has been recognized in [4], where a beam-space MUSIC algorithm is developed. We propose the application of a model-fitting approach to a small number of uniformly spaced beams. The method—*the global matched filter approach*—drastically improves upon the CBF and has performances comparable with the best HR techniques at a comparable cost.

II. Modeling and Problem Statement

We consider a circular array having N omnidirectional sensors in position $\{\rho\cos k\Delta,\ \rho\sin k\Delta\}$ with $k\in\{0,\cdots,N-1\}$ and $\Delta=2\pi/N$; see Fig. 1. To fix ideas, the radius of the array $\rho=\lambda/4\sin(\pi/N)$ is taken such that the distance between two neighboring sensors is $\lambda/2$, where λ is the wavelength for which the array is designed. We will assume that the array and the sources are coplanar. The sensor outputs are low-passed, sampled, and Fourier-transformed, and we will only be concerned with the narrowband signal corresponding to a single frequency bin of this discrete Fourier transform (DFT) of the sensor data. We denote by Z_k the complex order-N vector (snapshot) containing the corresponding components over the kth time interval. An estimate of the covariance matrix $\hat{R}$ of X_k is then given by

$$\hat{R}=\frac{1}{T}\sum_{k=1}^{T}Z_kZ_k^*. \tag{1}$$

This is also an estimate of the spectral density matrix of the sensor outputs at the considered temporal frequency. Assuming the source signals to be uncorrelated, its exact value, which is denoted R, can be decomposed into the sum of R_s, which is the contribution of the P sources, and R_n, which is the contribution of the noise

$$\begin{aligned} R&=R_s+R_n\\ R_s&=\sum_{p=1}^{P}\alpha_p d(\theta_p)d(\theta_p)^*\\ R_n&=vI \end{aligned} \tag{2}$$

Manuscript received December 28, 1999; revised December 13, 2001. The associate editor coordinating the review of this paper and approving it for publication was Prof. S. M. Jesus.

The author is with IRISA, Université de Rennes I, Rennes, France (e-mail: fuchs@irisa.fr).

Publisher Item Identifier S 1053-587X(01)02238-3.

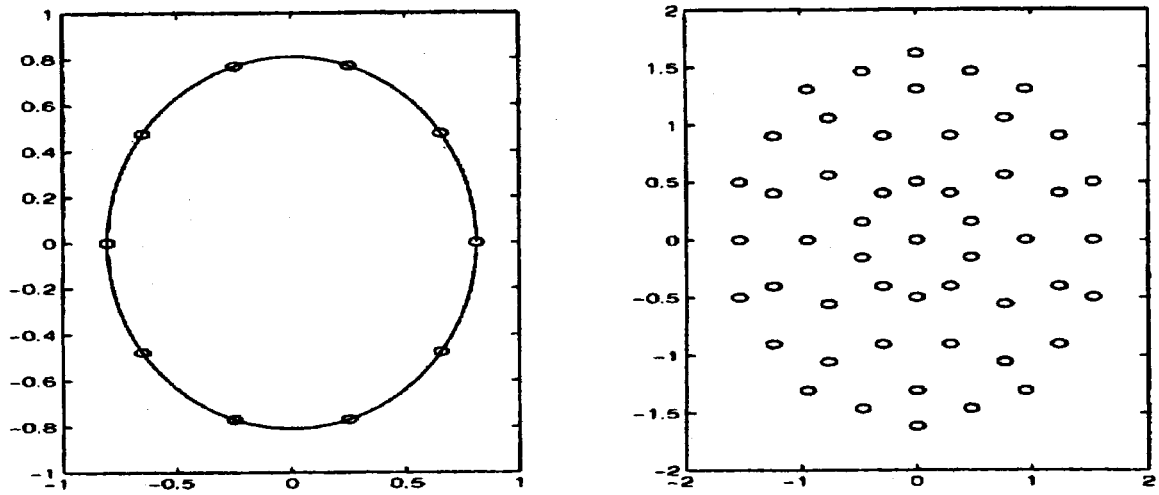

Fig. 1. Array and coarray.

where v is the power of the ambiant noise assumed spatially white, and $d(\theta_p)$ is the steering vector associated with the pth source whose power is denoted α_p. We define the signal-to-noise ratio (SNR) of the pth source to be α_p/v.

The steering-vector associated with bearing-angle $\phi \in [0, 2\pi[$ (measured counter-clockwise from the x-axis) is then, taking into account the position of the sensors

$$d(\phi) = \Big[e^{2i\pi(\rho/\lambda)\cos\phi}\ e^{2i\pi(\rho/\lambda)\cos(\phi-\Delta)}\ \cdots\ e^{2i\pi(\rho/\lambda)\cos(\phi-(N-1)\Delta)}\Big]^T.$$

Since we will later work with the beamformer outputs, let us introduce their definitions and properties. The beamformer output power (or beamformer output, for short) for bearing ϕ is

$$y(\phi) = \frac{1}{N^2} d(\phi)^* R d(\phi).$$

For R as in (2), it can be decomposed as

$$y(\phi) = \sum_{p=1}^{P} \alpha_p F(\phi, \theta_p) + \frac{1}{N} v \tag{3}$$

where

$$F(\phi, \theta_o) = \frac{1}{N^2} |d(\phi)^* d(\theta_o)|^2$$

is the output of the beamformer at bearing ϕ for a source with unit power located at bearing θ_o. This is also known as the array pattern when the sensor weights are all equal to one. For N as small as, say, ten, the array pattern is already almost invariant i.e.,

$$F(\phi, \theta_o) \simeq F(0, \phi - \theta_o) = F_o(\phi - \theta_o) = \frac{1}{N^2} |d(0)^* d(\phi - \theta_o)|^2.$$

This interesting property of UCAs guarantees that its performance is the same for all directions. Although we will use this property in the sequel to simplify the exposition, it is by no means necessary for our approach to apply.

As input to our algorithm, we evaluate L equispaced beams at bearings $\phi_k = 2\pi(k/L)$ for $k \in \{0, \cdots, L-1\}$ and denote $\hat{Y}$ the L-dimensional vector with components $\hat{y}_k = (1/N^2) d(\phi_k)^* \hat{R} d(\phi_k)$.

From (3), it follows that Y, which is the exact counterpart of $\hat{Y}$, can be decomposed into the sum of the contribution of each source and the contribution of the noise

$$Y = \sum_{p=1}^{P} \alpha_p f(\theta_p) + \left(\frac{v}{N}\right) \mathbf{e}_L \tag{4}$$

where $\mathbf{e}_L$ is an L-dimensional vector of ones modeling the noise contribution, and $f(\theta)$ represents the contribution to Y of a source at bearing θ and unit power.

For $\hat{Y}$, such a decomposition only holds approximatively because of the estimation errors due to the finite number of snapshots. In the next section, we indicate how such an approximate decomposition can be obtained.

III. Global Matched Filter Approach

A. Underdetermined System of Linear Equations

Let us introduce the set of M L-dimensional vectors $\{f_m, m \in \{0, \cdots, M-1\}\}$ with $f_m = f(2\pi m/M)$ (4). We will denote with F the $L \times (M+1)$-dimensional matrix whose columns are the f_m-vectors and the $\mathbf{e}_{M+1}$-vector that models the noise contribution, all normalized to one in ℓ_2-norm.

Our objective is to obtain a sparse representation of the observed vector $\hat{Y}$ in terms of the columns in F. Such a representation is of the form FX, where X is a vector having just a few nonzero components. Since the bearings θ_p in (4) do not, in general, belong to the discretization grid, more than P columns of F will be needed, and for a sparse representation to exist, one has to allow for reconstruction errors. These errors will not only take care of the mismatch due to the discretization of the bearings but will also serve to erase the estimation errors between $\hat{Y}$ and Y. The larger the allowed errors, the sparser the potential approximate representations.

What we seek is a sparse approximate solution to an underdetermined system of linear equations. In the present notations, a typical problem of this class is as follows: Find the (or a) vector X with the fewest number of nonzero components satisfying the constraint

$$||\hat{Y} - FX||_2^2 \leq \epsilon_2 \tag{5}$$

where $\epsilon_2 > 0$ is given, and $||X||_2^2 = \sum x_i^2$ is the Euclidean or ℓ_2 norm. This is a fundamental problem in matrix computation [6] that is know to be NP-hard. Other norms or different constraints lead to other problems in the class. The difficulty is to get close to the sparsest solution with a reasonable computational load. In the signal processing context, several different formulations have been considered; see, for instance, [7]–[11] and heuristic solutions proposed.

B. Tentative Solutions

The matching pursuit algorithm [7] is one heuristic way to get close to a solution of (5). It is an iterative method that, at a given step, picks the element f_j that has maximal correlation with the present residual vector $E^k = \hat{Y} - FX^k$

$$j_k = \arg\max_j f_j^T E^k \tag{6}$$

and substracts the weighted contribution of this element to form the new residual observation. When the weight is chosen to minimize the energy in the residual vector, one gets $E^{k+1} = E^k - (f_{j_k}^T E^k / f_{j_k}^T f_{j_k}) f_{j_k}$. This is performed until a stopping criterion is satisfied. In an estimation context, where a true sparse model exists, only scenarios with well-separated sources can be handled with this approach. For two closely spaced sources, it will typically start by locating and substracting a source in between the two true ones, and further iterations are then needed to correct this initial mistake. The so-obtained representation is then of little help [8].

To avoid the drawbacks induced by this element-by-element search, a higher dimensional search is proposed in [12]. The ideal remains, however, to be able to select in one step a sparse set of elements that explains $\hat{Y}$.

In [10], the constraint (5) that defines the admissible solutions is replaced by

$$||\hat{Y} - FX||_\infty \leq \epsilon_\infty \tag{7}$$

where $||E||_\infty = \max_i e_i$ is the ℓ_∞ norm or *max norm*, and $\epsilon_\infty > 0$ is a fixed bound and an approximation to the sparsest solution obtained using a linear program. While (5) is a bound on the energy of the global error and is often used in signal processing, (7) is a bound on the individual or local errors since the inequality has to hold component-wise. Constraint (7) says that a representation FX is admissible if the reconstruction error nowhere exceeds ϵ_∞. Although working locally is important if one seeks spiky solutions, we do not retain (7) because it has poor invariance properties.

C. Global Matched Filter

We propose instead yet another constraint with a nice physical interpretation

$$F^T(\hat{Y} - FX) \leq h\mathbf{e}_{M+1} \tag{8}$$

where $h \in \mathbb{R}^+$, and $\mathbf{e}_{M+1}$ is an $(M+1)$-dimensional vector of ones. This constraint, which is a special case of $||F^T(\hat{Y} - FX)||_\infty \leq h$, lies in between (5) and (7) and preserves the advantages of both: It applies locally and is invariant with respect to unitary transformations applied to $\hat{Y}$.

From (6), it follows that $f_k^T(\hat{Y} - FX)$ is the output of the filter matched to a source with bearing $2k\pi/M$ when applied to the residual vector $E = \hat{Y} - FX$. The constraint (8) then implies that at an admissible representation X, the outputs of all the filters matched to the M potential sources are smaller than a chosen threshold h.

To find among all the representations satisfying (8), the sparsest one is, as for all the problems in this class, too time consuming, and we propose to get an approximation by selecting the representation with minimal energy. This leads to

$$\min_X \tfrac{1}{2}||FX||_2^2 \qquad \text{subject to } F^T(\hat{Y} - FX) \leq h\mathbf{e}_{M+1} \tag{9}$$

where $h \in \mathbb{R}^+$ is a parameter to be tuned. At the optimum, one has the representation of $\hat{Y}$ in terms of columns of F that has lowest energy and that yields a residual vector E whose correlation with any further potential source is below a threshold h. In contrast with the matching pursuit approach that is a relaxation-type algorithm, this technique selects in a *global* way (simultaneously) all the elements f_k needed to obtain the optimum.

The threshold h acts as a detection threshold, but, again, instead of detecting a source at a time, they are detected jointly. It should be taken both small to allow for the detection of weak sources and large to reject the estimation errors to prevent any attempt to model these errors.

At the optimum, the $(M+1)$-dimensional vector X is expected to have around $2P+1$ nonzero components only: one component in front of the column that models the noise contribution and a pair of neighboring components for each source. Indeed, since the true bearings θ_p will generically fall in between two discretization points, two columns of F will, in general, be needed to approximatively reconstruct $f(\theta_p)$.

The optimization of (9) may appear as a difficult task, but using duality, we show in Appendix B that it is equivalent to

$$\min_{X \geq 0} \tfrac{1}{2}||\hat{Y} - FX||_2^2 + h\mathbf{e}_{M+1}^T X \tag{10}$$

which is a quadratic program whose unique global minimum can be obtained using standard and robust routines for quadratic programs available in any scientific programs library.

IV. Application to the Localization Problem

A. Fixing the Discretization Steps

Besides h, we have to fix the two parameters L and M before applying criterion (9) or (10) to our problem. L is the number of beams to be evaluated to build the observation vector $\hat{Y}$, and

M fixes the discretization step $2\pi/M$ in bearing that separates two columns in F.

For an UCA having N sensors, we recommend the evaluation of $L \simeq 4N$ equispaced beams. To arrive at this figure, we note that the effective aperture of the array is $\simeq N\lambda/4\pi$, that the width of the mainlobe at halfpower (known as FWMH, [13]) is then roughly $300/N^\circ$, and in order to cover the 360° with about three samples per main lobe, one needs $4N$ beams. Although this is, in general, smaller than the number of real degrees of freedom (rdof) of the array, we believe that little information is lost. This further allows us to consider the beams to be almost decorrelated random variables.

The discretization step in bearing $2\pi/M$ has to be chosen so that it does not prevent the approach to attain the CR bounds. For standard values of T, the number of snapshots, and SNRs, this means that we expect to improve the standard Rayleigh resolution by a factor 2 or 3, i.e., to solve equipowered sources separated by $100/N$ degrees (FWMH/3). For this to be easily achievable, we take a step equal to about $20/N^\circ$. This allows us to have about three zero components in the X vector in between the components representing the two equipowered sources one expects to solve and leads to $M \simeq 18N$ if one wants to cover the 360°.

B. Using the Statistical Properties

To improve the performance of the method, let us now take into account the statistical properties of the beams [13]. As the number of beams we evaluate is quite low, it is realistic to assume that they are close to uncorrelated random variables. An estimate of the approximate *diagonal* covariance matrix of $\hat{Y}$ is then $\hat{\Sigma}_y = \text{diag}(\hat{Y})^2/T$, and we propose to whiten the observations before applying the criterion (9) or (10). This amounts to replacing $\hat{Y}$ with $\hat{Y}_w = \hat{\Sigma}_y^{-(1/2)}\hat{Y}$ and F with $F_w = \hat{\Sigma}_y^{-(1/2)}F$, where the columns in F_w are further normalized to one.

C. Fixing the Threshold

Finally, to set the value of h, we consider the constraint (8) in its whitened form

$$F_w^T(\hat{Y}_w - F_w X) \leq h\mathbf{e}_{M+1} \tag{11}$$

and assume that we are in the ideal situation where $X = X_o$, which is the exact model. The whitened residual vector $E_w = \hat{Y}_w - F_w X_o$ is then Gaussian with zero mean and covariance matrix identity. Because the columns in F_w are normalized to one, the $M+1$ components of $F_w^T(\hat{Y}_w - F_w X_o)$ are then scalar Gaussian random variables with zero mean and unit variance.

To make this ideal model admissible, we have to fix the threshold h in (11) at a value that guarantees that the probability that the maximum of these $M+1$ standard Gaussian random variables be larger than h is close to zero. Since these variables are correlated, a precise value is difficult to obtain. The algorithm is fortunately quite robust in this respect, and following [14], we suggest that we take $h \simeq \sqrt{\text{Log}2M}$.

D. Summary of the Algorithm

For an UCA with N sensors, we do the following.

- We evaluate $L = 4N$ uniformly spaced beams over the whole horizon and form the vector $\hat{Y}$.
- We built the matrix F with M L-dimensional vectors f_m *separated* by about $20/N^\circ$ and a vector $\mathbf{e}_L$ allowing for the noise contribution.
- We whiten both $\hat{Y}$ and F using the estimate $\hat{Y}$.
- We normalize the columns of F_w, and solve

$$\min_{X \geq 0} \tfrac{1}{2}\|\hat{Y}_w - F_w X\|_2^2 + h\mathbf{e}_{M+1}^T X \tag{12}$$

  with $h = \sqrt{\text{Log}2M}$ using a quadratic programming routine. For scenarios with P well-separated sources, the solution vector X will have about $2P+1$ nonzero components: one for the noise contribution and the others appearing as neighboring pairs.
- To each pair of nonzero components, one associates a unique source with power the sum of the two components and bearing obtained by linear interpolation.

For more difficult scenarios such as the one considered in Section V-B, one may have to resort to some thresholding [15] to get rid of small nonzero components in X. We further comment on this in Section V-B.

If one knows beforehand that sources are only present in a given sector of the horizon, one can use this information and only keep in F the columns belonging to this sector while keeping all the components in $\hat{Y}$. The results appear to be unsensitive with respect to this selection. In contrast with most deconvolution approaches, no performance gains should be expected. Such prior knowledge has nevertheless two important consequences: It leads to substantial savings in computation time and considerably reduces the number of spurious peaks.

V. Simulation Results

We consider a UCA with $N = 10$ sensors. To have an intersensor spacing of half a wavelength, its radius is then $\rho = 0.809\lambda$. A plot of the array and of its coarray are presented in Fig. 1 with λ taken as unit of length. Its effective aperture is 1.6, and its performance is thus comparable with those of an ULA having about four sensors separated by $\lambda/2$. The resolution of such an ULA, for sources around broadside, is roughly 100/4°, and indeed, the width of the mainlobe at halfpower is 27°; see Fig. 2. One can note that in the direction opposite to the main lobe, there is a quite important secondary lobe. While this leads to difficulties when using the conventional beamformer output, it is of no direct consequence in our approach.

Following the indications given in Section IV-D, we form $L = 40$ beams. Two neighboring beams are thus separated by 9°, which has to be compared to the 27° of the beamwidth. The column f_m we use to build F are separated by 2°. If we want to cover the whole horizon, there are then $M = 180 + 1$ columns in F, and we solve (12) using the NAG E04NCF routine with $h = \sqrt{\text{Log}2M}$.

As observed in [4] and exploited in [5], for this UCA with an even number of sensors, the geometrical symmetries induce obvious relations between specific components of the covariance matrix R that are no longer present in the estimated matrix $\hat{R}$.

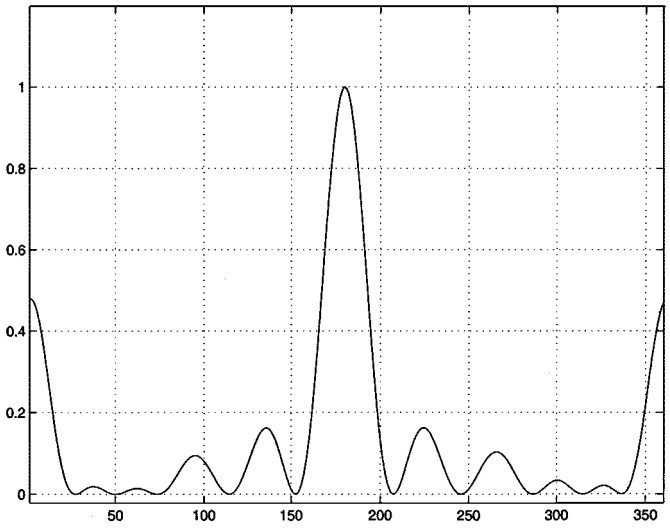

Fig. 2. Beampattern.

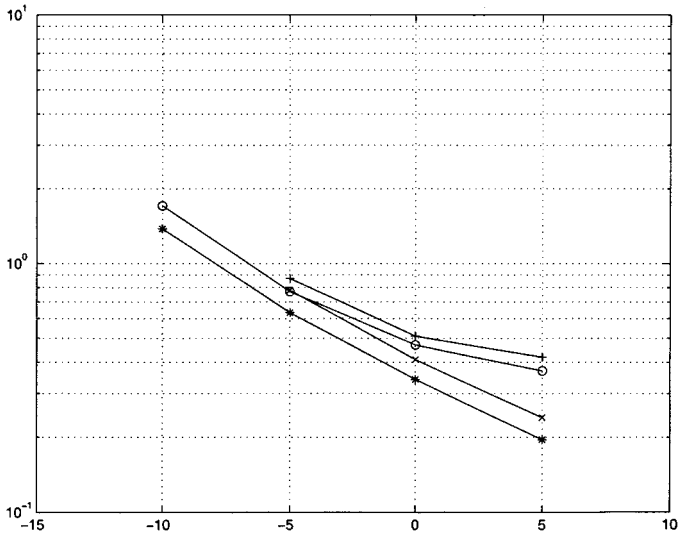

Fig. 3. Standard deviations with respect to SNR for different methods: Cramer–Rao bound "*," GMF "o," MUSIC "+," and R-MUSIC "x."

More specifically, for a ten-sensor array, the pairs of sensors 1-2 and 7-6 occupy similar positions, which means that the components (1, 2) and (7, 6) are identical in R. R is thus not just a Hermitian matrix with constant diagonal having 91 real degrees of freedom (rdof), but, since there are in fact 20 such identity relations that fix 40 rdof, R has only 51 rdof. The projection of the estimated covariance matrix $\hat{R}$ on its manifold (an operation called "rectification" in [5]) can simply be achieved by replacing each pair by its mean and averaging the diagonal. We denote $\hat{R}_R$ as the resulting matrix and R-MUSIC the MUSIC algorithm [16], [17] applied to $\hat{R}_R$. We will compare the results of the proposed approach to those of MUSIC and R-MUSIC in Section V-A.

A. Resolution of Two Equipowered Sources

Let us consider a scenario with two equipowered sources separated by 20°, $T = 400$ snapshots, and different common SNRs. Note that an SNR of 0 dB means $\alpha_1 = \alpha_2 = v = 1$ in (2). To fix ideas, we take $\theta_1 = 80$ and $\theta_2 = 100°$ and assume to know that the two sources are in between 60 and 120°. We plot in Fig. 3 the standard deviation in degrees for source 1 versus SNR for the global matched filter (GMF), the basic MUSIC algorithm, the MUSIC algorithm applied to the rectified covariance matrix $\hat{R}_R$ (R-MUSIC), and the Cramer–Rao bound (CRB). The results are obtained from 4000 independent trials. From Fig. 3, one sees that the methods have similar performance and that indeed, R-MUSIC performs better than MUSIC. For SNR $= -10$ dB, MUSIC failed to separate the two sources in about 15% of the simulations and R-MUSIC in 5%. The corresponding points are thus missing. GMF had no such difficulties; even at SNR $= -15$ dB; it only failed to separate the two sources in less than 1% of the simulations.

B. More Difficult Scenario

We now give a more detailed account upon the results of the global matched filter (GMF) for a more difficult scenario. The two equipowered sources are now separated by 10°, the number of snapshots is $T = 100$, and the SNR is 0 dB. To fix ideas, we take $\theta_1 = 80$ and $\theta_2 = 90$. While MUSIC-like algorithms separate the two sources in just about 30% of the realizations, the global matched filter approach encounters no difficulties as far as resolution is concerned. The results obtained by keeping the two strongest sources, in case more than two sources are detected, are very close to the CRBs.

TABLE I
RESULTS OVER 10 000 TRIALS WITH PRIOR KNOWLEDGE: FIRST TWO COLUMNS, WITHOUT PRIOR KNOWLEDGE: LAST TWO COLUMNS

source number	1	2	1	2
true bearing in degrees	80	90	80	90
estimated bearing	80.47	90.29	80.53	90.34
estim stdt dev. in deg.	1.45	1.53	1.40	1.41
CR stdt dev. in deg.	1.36	1.36	1.36	1.36
true amplitude	1	1	1	1
estimated amplitude	1.01	.90	1.02	.89
estim stdt dev.	.26	.25	.26	.27
CR stdt dev.	.26	.26	.26	.26

In the first two columns of Table I, we present the results obtained from 10 000 independent trials if it is known beforehand that the true sources lie between 60 and 120°. The matrix F has then 31 columns, and $h = 2.02$. The number of false alarms is quite low: Only 1.7% of the realizations detect one additional source with negligible amplitude.

The results in the case where no prior knowledge upon the sources location is assumed are presented in the last two columns of Table I. The matrix F then has $180 + 1$ columns and $h = 2.42$. There is no real difference in performance. The situation is, however, less favorable with respect to the number of spurious sources that are present. The average number of nonzero components in X was 8.52 and once the useful components (between 3 and 5) that model the two strongest sources and the noise have been removed, this figure drops to

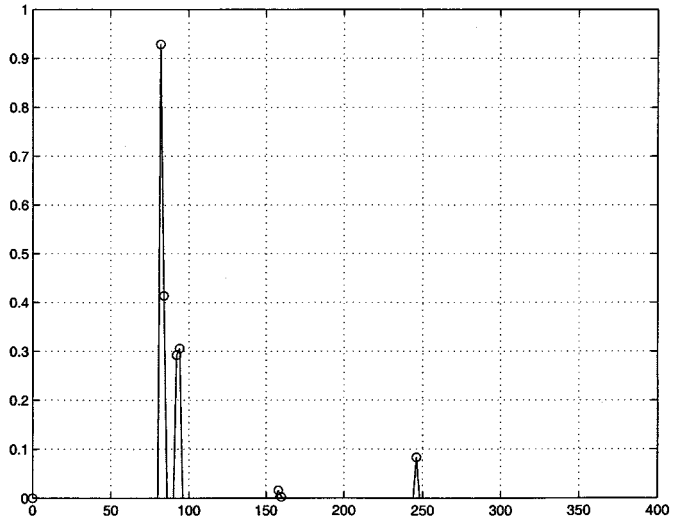

Fig. 4. Typical bad output. Among the 180 weights covering 360°, just eight, which are marked by o's, are nonzero: two of them associated with each true source and four with spurious sources.

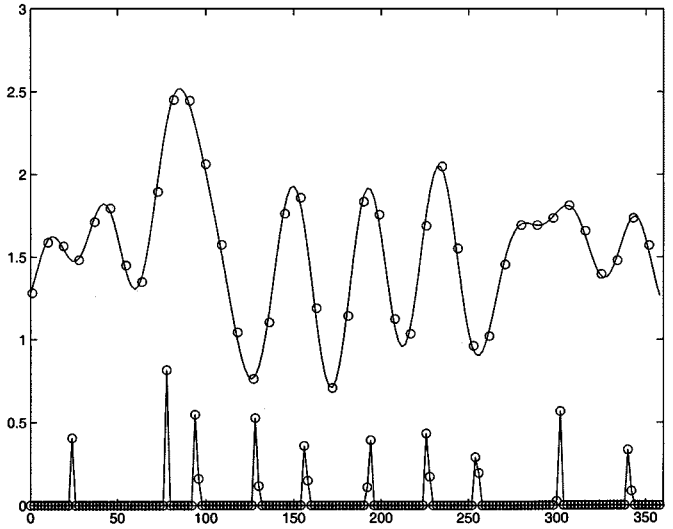

Fig. 5. Top: Output of the standard beamformer with the 40 values used by the algorithm marked by o's. Bottom: Output of the algorithm. Among the 180 components of X, only 18 are nonzero.

4.25. These components correspond to false alarms or spurious sources that have to be discarded by a thesholding operation. A closer look at the results shows that 2% of the realizations suffer from three false alarms, 43% have two false alarms, 44% have one false alarm, and 11% have no false alarm. Fortunately, looking at the realization among the 10 000 that yielded the strongest spurious source, its amplitude was 21% of the amplitude of the weakest of the two true sources. This means that in the worst realization, there was a ratio of about five between the amplitude of the second and third strongest source.

The output of a typically bad realization with three false alarms (four additional nonzero weights) is presented in Fig. 4. The weakest true source at bearing 93.0 has two neighboring nonzero weights that add to 0.59 and the strongest spurious source at bearing 246.0 has amplitude 0.08. There are three other, scarcely visible, nonzero weights marked by "o"s.

As explained in Section III, the larger h is, the smaller the number of spurious sources will be, and indeed, multiplying h by 2 in our simulations drastically reduces the number of false alarms. In the latter case, from 11%, the percentage of false-alarm-free realizations rises to 76%. Just 22% of the realizations have one false alarm left and 2% two false alarms. This is feasible in our scenario because the SNR is large (0 dB), but if weak sources are to be detected, one should stay with the initial value. Remember that this is a difficult scenario as far as resolution is concerned and that MUSIC-like algorithms in most cases simply fail to separate the two sources. On the other hand, in most HR techniques, one has either to do a similar thesholding operation or know beforehand the number of sources, and this is also the information we used since we sytematically kept the two strongest sources.

C. Locating Numerous Sources

As an interesting property of the proposed approach that it shares with nonparametric methods such as the standard beamformer or with certain nonuniformly spaced linear arrays [18], we present just one result for a ten-source scenario. The characteristics are as follows:

- $N = 10$;
- $P = 2$:
- $T = 100$;
- $\alpha_i = v = 1$;
- $\theta_i = 20, 80, 90, 120, 150, 190, 220, 260, 300, 340°$.

Applying the same algorithm as above leads typically to an X vector with about 20 nonzero components plus one for the noise variance. In Fig. 5, we present the output of the algorithm i.e., the 180 components of X (we removed the *noise* component) for one realization. There are 18 nonzero components that appear either as singletons or by pairs. One further associates with each pair a unique source by summing the weights and interpolating the bearings. One then gets the ten estimated bearings that are close to the true ones.

D. Computational Complexity

As far as the computational cost of quadratic programming routines is concerned, one admits in general that the number of operations required to perform one iteration is of the order of the square of $\min(L, M)$ and that the number of iterations is linear in this number. The computational cost is thus cubic in $\min(L, M)$, i.e., cubic in N, which is the number of sensors. This is quite comparable with other HR techniques such as MUSIC. Indeed, to extract the signal or noise subspace from $\hat{R}$ takes on the order of N^3 operations to which one has to add the cost of the evaluation of the spectrum at a number of points that is far from negligible. Knowing the sector in which the sources are present reduces the computational load of both approaches.

VI. Conclusions

A global matched filter (GMF) approach has been proposed and applied to localizing sources. It simultaneously fits to the observations all the multiple elements needed in order to explain

them. It is straightforward to implement, and there are few parameters to be chosen. We have indicated precisely how to tune them for UCAs, but GMF applies to any array and more generally to any situation where one has to fit to noisy observations a model composed of a unknown number of superimposed signals.

For a ten-sensor UCA, the method allows us to monitor the whole horizon (360°) using just 40 beams, and its performance is that of an HR method. For the simulated two closely spaced equipowered sources scenarios, the method attains performances close to the CRB for as little as 100 snapshots and outperforms the MUSIC-like algorithm at low SNR for similar computational loads. One should also note that GMF gives accurate estimates of the amplitude of the sources: information that is not available with MUSIC.

The method applies to scenarios where the number of sources is equal or greater than the number of sensors. Very few approaches able to handle this situation exist.

Appendix A
Property of the Beamformer

The coarray of a uniform circular array (see Fig. 1) with $N = 10$ sensors has 51 elements (including symmetry), indicating that the array has 51 real degrees of freedom (d.o.f.). This means that while the estimated covariance matrix $\hat{R}$ in (1) has 100 real d.o.f., its exact counterpart R in (2) has only 51 real d.o.f. Projecting $\hat{R}$ onto this reduced-dimensional subspace before using it in any suboptimal localization technique will increase the performance.

In [5] , this operation is called *rectification*; it is shown how such a projection can be achieved in an approximate way, and the corresponding gain in performance for a MUSIC-type algorithm is highlighted. In this Appendix, we show that the beamformer actually performs such a projection, or more precisely, the outputs of the beamformer are the same when using $\hat{R}$ or $\hat{R}_p$ its projection on this subspace. Since the output of the beamformer $d(\phi)^*\hat{R}d(\phi)^*$ can be rewritten $\mathrm{tr}(\hat{R}d(\phi)d(\phi)^*)$, where $\mathrm{tr}(\cdot)$ is the trace operator, one has to establish that

$$\mathrm{tr}(\hat{R}d(\phi)d(\phi)^*) = \mathrm{tr}(\hat{R}_p d(\phi)d(\phi)^*).$$

The proof uses the fact that $\mathrm{tr}AB^*$ is a inner product on the set of square matrices that corresponds to the usual Euclidean inner product when matrices are considered as N^2-dimensional vectors. Denoting with $\mathbb{P}$ the operator that achieves the desired projection, one has not only $\mathbb{P}\hat{R} = \hat{R}_p$ by definition, but $\mathbb{P}R = R$ as well, and $\mathbb{P}d(\phi)d(\phi)^* = d(\phi)d(\phi)^*$ since R and $d(\phi)d(\phi)^*$ already belong to the subspace. To establish the result, we successively write

$$\begin{aligned}\mathrm{tr}(\hat{R}d(\phi)d(\phi)^*) &= \mathrm{tr}(\hat{R}\mathbb{P}d(\phi)d(\phi)^*) = \mathrm{tr}(d(\phi)d(\phi)^*\mathbb{P}\hat{R}) \\ &= \mathrm{tr}(\hat{R}_p d(\phi)d(\phi)^*)\end{aligned}$$

where the second equality follows from the Hermitian property of the inner product.

Appendix B
Note on Duality

We will establish that

$$\min_{X\geq 0} \tfrac{1}{2}\,||\hat{Y} - FX||_2^2 + h\mathbf{e}_{M+1}^T X \tag{13}$$

and

$$\min_{X} \tfrac{1}{2}\,||FX||_2^2 \qquad \text{subject to } F^T(\hat{Y} - FX) \leq h\mathbf{e}_{M+1} \tag{14}$$

where $\mathbf{e}_{M+1}$ is a column of dimension $M+1$, are indeed equivalent. The aim of duality is to provide different formulations of an optimization problem. In our case, (13) is more convenient computationally, whereas (14) has more physical significance. Writing $FX = U$ in (14) yields yet another formulation:

$$\min_{U} \tfrac{1}{2}\,||U||_2 \qquad \text{subject to } F^T(\hat{Y} - U) \leq h\mathbf{e}_{M+1}. \tag{15}$$

Somehow, surprisingly, this means that the optimum U^* to this last problem can be rewritten as FX^* with a generically sparse X^*. The recovery of X^* from U^* is immediate by observing that X^* is the Lagrange multiplier vector for the constraints in (15); thus, $x_j^* = 0$ if j corresponds to an inactive constraint at the optimum of (15).

Indeed, (13)–(15) are all quadratic programs [19], and there is no way to distinguish a *primal* or *dual* problem. To establish the equivalence between (13) and (14), let $\mu \geq 0$ be the Lagrange multiplier associated with the inequality constraint on X in (13); the standard Lagrange duality applied to the convex problem (13) then says that it is equivalent to

$$\max_{\mu\geq 0} \min_{X} \tfrac{1}{2}\,||\hat{Y} - FX||_2^2 + h\mathbf{e}_{M+1}^T X - \mu^T X. \tag{16}$$

Since F^TF is only positive semidefinite, the minimum in X may not be finite for all $\mu \geq 0$, but these cases can be ignored since we will take the maximum in μ. The minimum in X is finite if and only if there is an X satisfying

$$-F^T(\hat{Y} - FX) + h\mathbf{e}_{M+1} - \mu = 0$$

and all Xs satisfying this relation yield the minimum. Substituting this relation premultiplied by X^T into (16), one gets the following equivalent form:

$$\max_{X,\mu\geq 0} \tfrac{1}{2}\,X^TF^TFX + \tfrac{1}{2}\,\hat{Y}^T\hat{Y}$$

subject to

$$F^T(\hat{Y} - FX) = h\mathbf{e}_{M+1} - \mu$$

where the maximum is taken over both X and $\mu \geq 0$. As $\hat{Y}^T\hat{Y}$ is a constant, it can be ignored in the criterion. Finally, eliminating $\mu \geq 0$ from the equality constraint transforms it into the inequality constraint in (14) and establishes the equivalence of (13) and (14).

References

[1] D. E. N. Davies, *The Handbook of Antenna Design*. London, U.K.: Peregrinus, 1983, vol. 2, ch. 12.

[2] A. H. Tewfik and W. Hong, "On the applications of uniform linear array bearing estimation techniques to uniform circular arrays," *IEEE Trans. Signal Processing*, vol. 40, pp. 1008–1011, Apr. 1992.
[3] B. Friedlander and A. Weiss, "Direction finding using spatial smoothing with interpolated arrays," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 28, pp. 574–587, Apr. 1992.
[4] C. P. Mathews and M. Zoltowski, "Eigenstructure techniques for angle estimation with UCA's," *IEEE Trans. Signal Processing*, vol. 42, pp. 2395–2407, Sept. 1994.
[5] P. Forster and T. Ast, "Rectification of cross spectral matrices for arrays of arbitrary geometry," in *Proc. IEEE ICASSP*, Phoenix, AZ, 1999, pp. 1311–1314.
[6] B. K. Natarjan, "Sparse approximate solutions to linear systems," *SIAM J. Comput.*, vol. 24, no. 2, pp. 227–234, Apr. 1995.
[7] S. G. Mallat and Z. Zhang, "Matching pursuit with time-frequency dictionaries," *IEEE Trans. Signal Processing*, vol. 41, pp. 3397–3415, Dec. 1993.
[8] A. H. Tewfik *et al.*, "Optimal subset selection for adaptive signal representation," in *Proc. IEEE ICASSP*, Atlanta, GA, 1996, pp. 2511–2514.
[9] G. Harikumar and Y. Bresler, "A new algorithm for computing sparse solutions to linear inverse problems," in *Proc. IEEE ICASSP*, Atlanta, GA, 1996, pp. 1331–1334.
[10] J. J. Fuchs, "Extension of the Pisarenko method to sparse linear arrays," *IEEE Trans. Signal Processing*, vol. 45, pp. 2413–2421, Oct. 1997.
[11] M. D. Sacchi, T. J. Ulrych, and C. J. Walker, "Interpolation nas extrapolation using a high-resolution discrete Fourier transform," *IEEE Trans. Signal Processing*, vol. 41, pp. 31–38, Jan. 1998.
[12] M. Goodwin, "Matching pursuit with damped sinusoids," in *Proc. IEEE ICASSP*, Munich, Germany, 1997, pp. 2037–2040.
[13] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1993.
[14] M. R. Leadbetter, G. Lindgren, and H. Rootzen, *Extremes and Related Properties of Random Processes and Sequences*. New York: Springer-Verlag, 1983.
[15] J. J. Fuchs, "Detection and estimation of superimposed signals," in *Proc. IEEE ICASSP*, vol. III, Seattle, WA, 1998, pp. 1649–1652.
[16] G. Bienvenu and L. Kopp, "Optimality of high resolution array processing using the eigensystem approach," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 1235–1248, Oct. 1983.
[17] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 276–280, Mar. 1986.
[18] Y. U. Abramovich *et al.*, "Positive-definite toeplitz completion in DOA estimation for nonuniform linear antenna arrays," *IEEE Trans. Signal Processing*, vol. 45, pp. 2413–2421, Oct. 1997.
[19] R. Fletcher, *Practical Methods of Optimization*. New York: Wiley, 1987.

Jean-Jacques Fuchs (M'81) was born in France in 1950. He graduated from the École Supérieure d'Électricité, Paris, France, in 1973 and received the M.S. degree in electrical engineering from the Massachusetts Institute of Technology, Cambridge, in 1974.

After a short period in industry with Thomson-CSF, he joined the Institut de Recherches en Informatique et Systèmes Aléatoires (IRISA), Rennes, France, in 1976. Since 1983, he has been Professor of Electrical Engineering at the Université de Rennes 1. His research interests shifted from adaptive control and identification to signal processing. He is now involved in array processing and radar.