SIAM J. MATRIX ANAL. APPL.
Vol. 17, No. 2, pp. 355–382, April 1996
© 1996 Society for Industrial and Applied Mathematics
008

ON THE PROPERTIES OF A RELATIVE ENTROPY FUNCTIONAL*

J.-P. LE CADRE†

Abstract. The identification of the noise correlations (e.g., between the sensors of an array) is an important problem. It is also ill posed unless some additional conditions are verified. Here, these supplementary conditions are reduced to a low-rank hypothesis and to the knowledge (e.g., an upper bound of its length) of the general structure of the noise correlations. By introducing an original functional (named relative entropy functional), we develop a new approach for solving the above problem. In particular, it is shown that this functional inherits from its definition interesting and useful properties (such as location of the extrema, concavity, etc.). These properties are shown using elementary linear algebra.

Key words. entropy, noise, optimisation, eigenvalues

AMS subject classifications. 94A17, 94A12

1. Introduction and problem statement. Usually, the signal received on an array of sensors is composed of source and noise parts. The aim of array processing is to estimate source parameters from the array measurements [1]. However, in numerous practical situations, especially in the array processing context, the noise parameters are unknown.

Most of the practical array processing methods are based upon the properties of the covariance matrices (CM) of the various signals impinging on the array. This is particularly true for high-resolution methods for which source and noise parts play symmetric roles [2–7]. For readers unfamiliar with these methods, we note that they are rather similar, in spirit, to principal component analysis methods [8–10].

At a given frequency (after discrete Fourier transform, for example) the problem of separation in source and noise parts is reduced to the following matrix equation:

$$R = S + B$$

$$\text{with :}\quad \left|\begin{array}{lll} R & : & \text{sensor outputs } CM, \\ S & : & \text{source } CM, \\ B & : & \text{noise } CM, \\ R, S, B & & q \times q \text{ matrices.} \end{array}\right. \tag{1}$$

The matrix R is assumed to be known (it is actually estimated from the sensor outputs). The problem we deal with can be stated as follows:

How can we obtain an "estimate" of B from R?

The problem stated above is ill posed and is meaningless without the following additional hypotheses:

$$\left|\begin{array}{lll} H_1 & : & S \text{ is positive semidefinite and rank deficient,} \\ H_2 & : & \text{the general structure of } B \text{ is known,} \\ & & (B \text{ is positive definite}). \end{array}\right. \tag{2}$$

The above hypotheses are generally accepted in the array processing literature [1–7] even if H_2 is frequently replaced by a drastically simpler hypothesis, say, H_2'.

* Received by the editors July 6, 1993; accepted for publication (in revised form) by F. T. Luk May 30, 1995.

† IRISA/CNRS, Campus de Beaulieu, 35042 Rennes Cedex, France.

The matrix B is known except for a scalar multiplier λ (λ is the noise level). The hypothesis H_2 is thus far less restrictive than H_2'. The general structure of B may be simply a banded Toeplitz structure (with the positivity hypothesis) or given by the covariance structure of a moving average (MA) spatial process [11].

The hypothesis H_1 is also quite acceptable since the rank deficiency hypothesis amounts to assuming that the number of sources is strictly inferior to the number of sensors. This assumption is instrumental for high resolution methods.

After the noise matrix B is estimated, the source parameters (defining S) can be estimated [1, 2, 7]. We want to stress that, for this approach, the parameters defining B are estimated independently of the source ones, using only the observation (i.e., the matrix R). For that purpose, an original approach will be derived. It relies on the "separating" properties of a relative entropy functional (REF). Roughly speaking, this functional allows us to "extract" the smooth component (i.e., the noise) of the observations. Another approach consists of using an approximated likelihood functional. This functional involves only the eigenvalues of a whitened matrix. A complete description of this approach is given in [12]. This method presents some (hidden) similarities with the REF method since it too relies on a (hidden) barrier functional. However, it is much more classical in principle and does not present the same possibilities.

The optimal methods [13] (for the statistical meaning) consist in simultaneously estimating the source and noise parameters. These approaches are rather direct although they may involve rather intricate derivations. However, the main criticism comes from the absence of any convergence property for the iterative algorithms that optimize the related functionals. The practical interests of such methods may thus be greatly reduced despite their (theoretical) optimality.

The method that will be presented obeys the following general scheme: we define a barrier functional forbidding the description of sources by the noise model. We shall carefully study the estimation of the noise model (or equivalently of B) as well as iterative optimization of the functional (gradient-like procedure). We stress that this optimization is defined *only with respect to noise parameters*, which constitutes the major novelty of our approach. We recall that the present paper deals with the exact properties of the functional and, thus, that *statistical considerations are not in the paper's scope.*

Actually, the barrier property of the REF appears to be instrumental since it is a means to create a singularity at the boundary of the feasible region. This study can thus be included in the much more general context of barrier methods [14]. According to [14], barrier methods fell from favor during the 1970s partly because of inherent ill-conditioning in the Hessian matrix. We shall see that the proposed method does not suffer from this drawback and enjoys interesting properties.

We use the following notation throughout the text of this paper:

- matrices are represented by capital letters (e.g., R, S, B, U_i, Z_i); all the matrices are $q \times q$ except the matrices Z_i (5) and N,
- vectors are represented by capital bold letters (e.g., $\mathbf{X}, \mathbf{B}, \mathbf{V}, \mathbf{W}$),
- scalars are denoted by small letters (e.g., b_i, l), eigenvalues by small Greek letters (e.g., λ_i),
- R generally represents the observation (or a resume), S the source part, and B the noise part (noise parameters β_i or b_i),
- the symbols det and *tr* denote, respectively, the transpose and the trace (of a $q \times q$ matrix),

- diag denotes a diagonal matrix,
- A^t and A^* denote, respectively, the transpose and the hermitian adjoint of A,
- Id stands for the identity matrix,
- $R(k)$ denotes the spatial density (31) of the observation, and
- $\mathrm{Re}(z)$ denotes the real part of z, $\bar{z}$ the complex conjugate of z.

2. The relative entropy approach. This approach deals with a functional depending only on the matrices R and B. In what follows, this functional will be named the relative entropy functional (REF) and will play the central role in solving the problem (1) under the hypotheses (2). It is defined below as

$$H(B) = \log\det(R - B) + l.\log\det B, \tag{3}$$

where l is a scalar factor and $\det(A)$ denotes the determinant of the matrix A.

A brief statistical motivation of H is presented in Appendix A. The scalar factor l is considered (see Appendix A) as a redundancy factor since it is associated with the number of (statistically) independent noise vectors available along the array. Practically, the choice of the factor l is related to statistical considerations beyond the scope of the present paper.

Now a parameterization of the B matrix is necessary. For that purpose, a banded Toeplitz parameterization is quite convenient, i.e., [15, 16]:

$$B = \sum_{i=1}^{p} \beta_i U_i,$$

$$\text{where :} \left| \begin{array}{l} \beta_i \text{ are scalars (real),} \\ U_i \text{ is a } q \times q \text{ matrix defined as usual by:} \\ U_i(k,\ell) = \begin{cases} 1 & \text{if} \quad |k-\ell| = i-1, \\ 0 & \text{else.} \end{cases} \end{array} \right. \tag{4}$$

The number q represents the sensor number and is consequently the dimension of the square matrices R, S, B. The matrices $\{U_i\}_{i=1}^{q}$ constitute an orthogonal (for the euclidean product) basis of T_q (the vector space of q-dimensional symmetric Toeplitz matrices). For practical applications [7], p is small with respect to q.

Obviously, this parameterization does not ensure the positive definiteness of B, so to avoid such a problem B can be parameterized as the covariance matrix of a MA process:

$$B = \left(\sum_{i=0}^{p-1} b_i\ Z_i\right) \left(\sum_{i=0}^{p-1} b_i\ Z_i\right)^t.$$

(The symbol "t" denotes matrix transposition.)

Here the matrices Z_i are $(p \times p + q)$ rectangular matrices given by

$$Z_i = \begin{pmatrix} 0 & \cdots & 0 & 1 & 0 & 0 & \cdots & \cdots & 0 \\ 0 & \cdots & 0 & 0 & 1 & 0 & \cdots & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & 0 & 1 & \cdots & 0 \end{pmatrix}$$

or

$$Z_i(k,\ell) = \begin{cases} 1 \text{ if } |k-\ell| = i, \\ 0 \text{ else.} \end{cases} \tag{5}$$

(This model is the assumed minimum phase [11].)

This parameterization will be especially useful for the special case study of large arrays (§5), but for the moment our attention will be restricted to the banded Toeplitz parameterization. The REF will thus be defined as follows:

$$H\left(\beta_1, \ldots, \beta_p\right) = \log\det(R-B) + l.\log\det B$$

with $B = B\left(\beta_1, \ldots, \beta_p\right)$ (R, B are $q \times q$ matrices).

The general optimization problem takes the following form. Find

$$\max H\left(\beta_1, \ldots, \beta_p\right)$$

under the constraints

$$\mathcal{C} \left| \begin{array}{lcl} R-B & > & 0, \\ B & > & 0. \end{array} \right. \tag{6}$$

($A > B$ means, as usual, that $A - B$ is positive definite.)

Let $\mathbf{B}_*$ be the parameter vector maximizing H under $\mathcal{C}$, i.e.,

$$\mathbf{B}_* = \arg\max H\left(\beta_1, \ldots, \beta_p\right) \quad \text{under } \mathcal{C}.$$

As will be seen later, the functional H can be efficiently maximized by iterative methods. But let us first consider the properties of $\mathbf{B}_*$.

3. Properties of $\mathbf{B}_*$. Because the REF depends nonlinearly on the parameters $\{\beta_i\}$, it seems much simpler to formulate the problem in terms of the eigensystems.

The spatially white noise case is presented first. It is not relevant to our problem, but it allows us to obtain a simple result and enlightens the role of the factor l. Then the general case is considered.

3.1. Spatially white noise case. This case is very simple but the reasoning is rather similar to that used in the general case.

The noise is assumed to be uncorrelated (sensor to sensor), so

$$B = \lambda \,\mathrm{Id}$$

(Id : identity matrix, $\lambda > 0, \lambda = \beta_1$).

Consider now an eigendecomposition of R, i.e.,

$$\begin{aligned} R &= \sum_{i=1}^{q} \alpha_i \mathbf{V}_i \mathbf{V}_i^*, \quad \mathbf{V}_i \perp \mathbf{V}_j (i \neq j), \|\mathbf{V}_i\| = 1 \\ &\left(\alpha_1 \geq \alpha_2 \geq \cdots \geq \alpha_q > 0\right). \end{aligned} \tag{7}$$

($*$ denotes transposition and conjugation.)

Then by means of elementary algebra, we have

$$H(\lambda) = \sum_{i=1}^{q} \log\left(\alpha_i - \lambda\right) + ql.\log\lambda. \tag{8}$$

The problem now consists of determining the value of λ that maximizes $H(\lambda)$ under the constraints $\mathcal{C}$.

Now

$$\frac{\partial H}{\partial \lambda} = \frac{ql}{\lambda} + \sum_{i=1}^{q} \frac{-1}{(\alpha_q - \lambda) - (\alpha_q - \alpha_i)},$$

and denoting by $\tilde{\lambda}$ the following particular value of λ,

$$\tilde{\lambda} = \alpha_q \left(\frac{l}{l+1} \right),$$

we obtain

$$\text{(9)} \qquad \left(\frac{\partial H}{\partial \lambda} \right)_{\lambda = \tilde{\lambda}} = (l+1)\frac{q}{\alpha_q} + \sum_{i=1}^{q} \left[-\frac{1}{(l+1)}\alpha_q + (\alpha_q - \alpha_i) \right]^{-1}.$$

Now $\alpha_i - \alpha_q \geq 0$ for $i = 1, 2, \ldots, q$ and therefore

$$\left(\frac{\partial H}{\partial \lambda} \right)_{\lambda = \tilde{\lambda}} \geq 0.$$

Furthermore, if λ tends towards α_q, then $\partial H / \partial \lambda$ tends towards $-\infty$. Since H is a differentiable and concave function of λ in the interval $]\,0, \alpha_q\,[$, it follows that the maximum of the REF under the constraints $\mathcal{C}$ is attained for a value λ_* of λ satisfying the following inequalities:

$$\text{(10)} \qquad \alpha_q \left(\frac{l}{l+1} \right) \leq \lambda_* \leq \alpha_q.$$

3.2. General case. Let B_0 be the exact noise CM and assume that B_0 may be described by the parameterization defining the B matrices. Then the following proposition is valid.

PROPOSITION 3.1. *Let $\{\lambda_i^*\}_{i=1}^{q}$ be the eigenvalues of the (whitened) matrix $B_0^{-1}B_*$, where B_* denotes the matrix maximizing the REF H under the two constraints (B and $R - B$ positive definite). Then these eigenvalues satisfy the following inequalities:*

$$\frac{l}{l+1} \leq \lambda_i^* \leq 1, \qquad i = 1, 2, \ldots, q.$$

Proof. The decomposition of R in source and noise parts (i.e., $R = S + B_0$) is assumed to be unique (see Proposition 4.3). The rank of S is denoted by s (the source number). Note that s is strictly inferior to q. Then the REF H takes the general following form:

$$H(B) = \log\det\left(S + (B_0 - B)\right) + l.\log\det B.$$

Let us now examine the various terms of the functional $H(B)$. For that purpose, let us note the following assumptions:

1. B_0 is positive definite,
2. $(B_0 - B)$ is positive definite.

The first assumption is quite classical in the signal processing context since B_0 is a covariance matrix [1, 3]. The second will be justified later (Comment 1, pp. 361–362).

Since the matrix B_0 is positive definite, it can be factorized in triangular factors (Choleski factorization [17]), i.e.,

$$B_0 = T_0 T_0^*.$$

Thus

$$\begin{aligned}\log\det\left[S+(B_0-B)\right] &= \log\det\left[S(B_0-B)^{-1}+\mathrm{Id}\right]\\ &\quad+\log\det(B_0-B)\\ &= \log\det\left[S T_0^{-1*}\left(\mathrm{Id}-T_0^{-1}BT_0^{-1*}\right)^{-1}T_0^{-1}+\mathrm{Id}\right]\\ &\quad+\log\det(B_0-B)\,.\end{aligned} \tag{11}$$

$$\left((B_0-B)^{-1}=T_0^{-1*}\left(\mathrm{Id}-T_0^{-1}BT_0^{-1*}\right)^{-1}T_0^{-1}\right).$$

Since the matrix $\left(\mathrm{Id}-T_0^{-1}BT_0^{-1*}\right)^{-1}$ is hermitian it is diagonalizable, i.e.,

$$\left(\mathrm{Id}-T_0^{-1}BT_0^{-1*}\right)^{-1}=W\Delta W^*$$

(W : unitary matrix, Δ : diagonal) with

$$\Delta(i,i)=(1-\lambda_i)^{-1}.$$

(λ_i : eigenvalue of $B_0^{-1}B$.)

In what follows, it is necessary to preserve the symmetry of the problem obtained by using elementary algebra as follows:

$$\begin{aligned}\log\det\left[ST_0^{-1*}\left(\mathrm{Id}-T_0BT_0^{-1*}\right)T_0^{-1}+\mathrm{Id}\right] &= \log\det\left[ST_0^{-1*}W\Delta W^*T_0^{-1}+\mathrm{Id}\right]\\ &= \log\det\left[\Delta^{1/2}W^*S'W\Delta^{1/2}+\mathrm{Id}\right]\\ &\text{with : } S'=T_0^{-1}ST_0^{-1*}.\end{aligned} \tag{12}$$

(This last equality results from intensive use of the classical formula $\det(AB) = \det(BA)$ [17].)

The matrix $\Delta^{1/2}$ in (12) is the diagonal matrix defined by $\Delta^{1/2}(i,i)=(\Delta(i,i))^{1/2}$. Its existence follows from the hypothesis that (B_0-B) is positive definite. Thus, the following equality holds trivially:

$$\begin{aligned}\log\det\left[\Delta^{1/2}W^*S'W\Delta^{1/2}+\mathrm{Id}\right] &= \log\det\left[\Delta^{1/2}W^*\left(S'+W\Delta^{-1}W^*\right)W\Delta^{1/2}\right]\\ &= \log\det\Delta+\log\det\left(S'+W\Delta^{-1}W^*\right).\end{aligned} \tag{13}$$

We are now able to calculate the partial derivatives of the REF H with respect to the parameters λ_i. More precisely, using a classical formula for the differentiation of the determinant of a matrix $A(\lambda)$ [18] (i.e., $\partial/\partial\lambda \log\det A(\lambda) = tr\left(A^{-1}(\lambda)\partial/\partial\lambda A(\lambda)\right)$), the partial derivatives $\partial H/\partial\lambda_i$ take the following form:

$$\frac{\partial H}{\partial\lambda_i}=l.\frac{1}{\lambda_i}-\frac{1}{1-\lambda_i}+\left\{\frac{1}{1-\lambda_i}+tr\left[\left(S'+W\Delta^{-1}W^*\right)^{-1}W\left(\frac{\partial}{\partial\lambda_i}\Delta^{-1}\right)W^*\right]\right\} \tag{14}$$

(tr denotes the trace).

Now the following equality comes from the orthogonality property of the eigenvectors [17]:

$$(15)\qquad tr\left[\left(S' + W\Delta^{-1}W^*\right)^{-1} W \frac{\partial \Delta^{-1}}{\partial \lambda_i} W^*\right] = -\mathbf{W}_i^* \left(S' + W\Delta^{-1}W^*\right)^{-1} \mathbf{W}_i^*.$$

($\mathbf{W}_i$ is the ith column of the matrix W.)

Furthermore, one has

$$S' + W\Delta^{-1}W^* \geq W\Delta^{-1}W^*;$$

hence,

$$\left(S' + W\Delta^{-1}W^*\right)^{-1} \leq \left(W\Delta^{-1}W^*\right)^{-1},$$

so that, finally,

$$(16)\qquad -\mathbf{W}_i^* \left(S' + W\Delta^{-1}W^*\right)^{-1} \mathbf{W}_i \geq -\mathbf{W}_i^* (W\Delta W^*)^{-1} \mathbf{W}_i = -\frac{1}{1-\lambda_i}.$$

In conclusion, we note that the term between braces in (14) (i.e., $1/(1-\lambda_i) + tr[(S' + W\Delta^{-1}W^*)^{-1} W(\frac{\partial}{\partial \lambda_i}\Delta^{-1})W^*]$) is positive when λ_i runs throughout the open interval $]0,1[$. Consequently, the partial derivatives $\partial H/\partial \lambda_i$ are positive ($i = 1, \ldots, q$) when λ_i runs throughout the open interval $]\,0, l/l+1\,[$.

Furthermore, the equality

$$B_0 - B = T_0 \left(\mathrm{Id} - T_0^{-1} B T_0^{-1*}\right) T_0^*$$

proves that (under the basic assumptions) the matrix $\left(\mathrm{Id} - T_0^{-1} B T_0^{-1*}\right)$ must be positive definite and that all the eigenvalues (i.e., $1-\lambda_i$) of the matrix $\mathrm{Id} - B_0^{-1}B$ must be positive. It is thus sufficient to restrict our attention to the parameter values β_i such that all the eigenvalues λ_i are smaller than 1.

Now the following fact is instrumental for the proof of Proposition 3.1: the REF H is a concave functional on the whole domain $\mathcal{C}$ of the constraints (6). This property will be shown in the next section *independently* of Proposition 3.1.

Denote by $\mathcal{C}'$ the following (restricted) constraint domain defined as $\mathcal{C}' = \{\, B$ such that (s.t.) B and $B_0 - B$ are positive definite $\}$. Then it is directly shown that $\mathcal{C}'$ is a convex subset of $\mathcal{C}$.

When λ_i tends towards 1_- then H tends towards $-\infty$. Since all the partial derivatives $\partial H/\partial \lambda_i$ are positive when λ_i runs through the interval $]0, l/l+1[$ and are continuous on $\mathcal{C}'$, there exists a matrix B_* of $\mathcal{C}'$ such that the maximum of H on $\mathcal{C}'$ is attained for B_*. Note that this maximum is unique on $\mathcal{C}'$ (concavity) and is attained for a matrix B_* such that all the eigenvalues λ_i^* (of $B_0^{-1}B_*$) belong to the interval $]l/l+1, 1[$. So there is a point $\left(\beta_1^*, \ldots, \beta_p^*\right)$ of $\mathcal{C}'$ such that all the partial derivatives $\partial H/\partial \lambda_i$ are altogether null.

Because the REF H is concave on the whole domain $\mathcal{C}$, its maximum on $\mathcal{C}$ is unique and is attained for a matrix B_* of $\mathcal{C}'$. This proves Proposition 3.1. □

Comments. The preceding calculations require some comments.

1. In the proof of Proposition 3.1, the positive definite hypothesis $(R - B)$ has been replaced by the positive definite hypothesis $(B_0 - B)$.

Actually, the two subsets $\mathcal{C}$ and $\mathcal{C}'$ ($\mathcal{C}' = \{B$ s.t. B and $B_0 - B$ are positive definite$\}$) are convex and $\mathcal{C}$ contains $\mathcal{C}'$. Because the functional H is concave on $\mathcal{C}$ (Proposition 4.1) and attains its maximum value on $\mathcal{C}'$, this maximum is unique and satisfies Proposition 3.1 on the whole subset $\mathcal{C}$ [19].

2. Consider (16). Then this inequality becomes an equality (for all the values of i) if and only if the source matrix S' is null. In this case, all the partial derivatives $\partial H/\partial \lambda_i$ are null for $\{b_i^*\}$ values s.t.

$$B_0^{-1} B_* = \left(\frac{l}{l+1}\right) \mathrm{Id}$$

or

$$B_* = \left(\frac{l}{l+1}\right) B_0.$$

The matrix B_0 is thus perfectly "estimated" up to a scalar factor. We want to stress that this scalar factor (i.e., $l/l+1$) has no practical importance.

3. As has been seen in the proof of Proposition 3.1, the effect of sources is to move the maximum of H and to cancel the equality of all the λ_i^*. Thus, in the presence of sources, B_0 cannot be perfectly "estimated" by maximizing H. Of course, the "quality" of the estimate increases with the scalar l.
4. Proposition 3.1 is still valid when the noise model is overdetermined (i.e., $p_0 \leq p$); this fact follows clearly from the proof of Proposition 3.1. For practical applications, it is a fundamental point.
5. The following property seems valid (at least for sufficiently great values of q).

Conjecture 1. Consider two distinct values of the parameter l, say l_1 and l_2 ($l_1 > l_2$), and denote $\{\lambda_i^*\}$ (respectively, $\{\mu_i^*\}$) to be the eigenvalues of $B_0^{-1} B_{*l_1}$ (respectively, $B_0^{-1} B_{*l_2}$). Then the following property holds:

$$\lambda_i^* \geq \mu_i^*, \qquad i = 1, \ldots, q.$$

This property has been verified by numerous simulation results (see Figs. 3–6). A very rough "proof" is based on the following fact: two banded Toeplitz matrices commute (approximately).

Actually, the REF method can be considered as a way to tackle the following problem: how to determine the "more random" noise model (i.e., maximizing $\log\det B$) under the positive definiteness constraints (B and $R - B$ positive definite). Clearly, for this sense, the term $\log\det(R - B)$ appears as a barrier functional forbidding interaction between source and noise models. The factor l represents the compromise between the accuracy of estimated parameters (Proposition 3.1) and the statistical variability of the estimated data (i.e., $\hat{R}$). It can thus be considered as a redundancy factor (see Appendix A).

Obviously, this interpretation of the factor l relies upon statistical considerations that are not in the scope of this paper.

4. Maximization of the REF *H*. The numerical problem now consists of determining the values of the parameters $\{\beta_i\}_{i=1}^p$ that maximize the REF H under the positivity constraints. After a general study of the functional concavity, the problem of practical optimization will be considered.

Actually, the REF enjoys a useful property which has been instrumental in the proof of Proposition 3.1.

PROPOSITION 4.1. *On the constraints domain $\mathcal{C}$ (6) the REF is a concave functional with respect to the $\{\beta_i\}_{i=1}^p$ parameters.*

Proof. The proof of Proposition 4.1 relies upon classical results of linear algebra. More precisely, we use the following classical lemmas, valid for any differential family of isomorphisms [20]:

$$\left|\begin{array}{lcl} \dfrac{\partial}{\partial\beta}\log\det B(\beta) & = & tr\left(B^{-1}(\beta)\dfrac{\partial B}{\partial\beta}\right), \\ \dfrac{\partial}{\partial\beta}B^{-1}(\beta) & = & -B^{-1}(\beta)\dfrac{\partial B}{\partial\beta}B^{-1}(\beta). \end{array}\right. \tag{17}$$

Then the Hessian matrix (denoted H_2) of H with respect to the $\{\beta_i\}_{i=1}^p$ is easily obtained:

$$H_2(i,j) \triangleq \frac{\partial^2 H}{\partial\beta_i\partial\beta_j} = -tr\left[(R-B)^{-1}U_i(R-B)^{-1}U_j\right] - l\cdot tr\left(B^{-1}U_iB^{-1}U_j\right). \tag{18}$$

Let $\mathbf{X}$ be any vector of $\mathbb{R}^p$ ($\mathbf{X}^t = (x_1,\ldots,x_p)$); then

$$\mathbf{X}^t H_2\mathbf{X} = \sum_{i,j} x_i\frac{\partial^2 H}{\partial\beta_i\partial\beta_j}x_j,$$

and using (18) and the linearity of the trace we get

$$\begin{aligned}\mathbf{X}^t H_2\mathbf{X} = &-tr\left[(R-B)^{-1}\left(\sum_{i=1}^p x_iU_i\right)(R-B)^{-1}\left(\sum_{j=1}^p x_jU_j\right)\right]\\ &-l.tr\left[B^{-1}\left(\sum_{i=1}^p x_iU_i\right)B^{-1}\left(\sum_{j=1}^p x_jU_j\right)\right].\end{aligned} \tag{19}$$

The two terms $-tr(--)$ of (19) are of the form $-tr(AC\ AC)$ with

$$A = (R-B)^{-1} \text{ or } B^{-1} \text{ and } C = \sum_{i=1}^p x_iU_i.$$

Since the matrix A is assumed to be positive definite, it admits a Choleski factorization, say $A = TT^*$, so that

$$\begin{aligned}-tr\,[ACAC] &= -tr\,[TT^*CTT^*C]\\ &= -tr\left[(T^*CT)^2\right] = -\|T^*CT\|_F^2.\end{aligned} \tag{20}$$

(The symbol $\|\ \ \|_F$ denotes the Frobenius norm [17] of a matrix.)

Finally, the quadratic form $\mathbf{X}^tH_2\mathbf{X}$ is negative for any (nonnull) vector $\mathbf{X}$. The matrix H_2 is therefore negative definite and H is therefore a concave functional with respect to $\{\beta_i\}_{i=1}^p$ on $\mathcal{C}$. Consequently, gradient-like methods (with optimal stepsize) will converge on $\mathcal{C}$. □

Actually, this concavity property is very strong and quite dependent on the noise parameterization. Thus Proposition 4.1 holds for a linear parameterization but not for more restrictive (especially nonlinear) parameterizations. Consider, for instance, the MA parameterization of the noise (5); then the partial derivatives of the functional H (with respect to b_i) are directly calculated, yielding

$$\begin{aligned}\frac{\partial H}{\partial b_i} &= -tr\left[(R-B)^{-1}D_i^1\right] + l.tr\left[B^{-1}D_i^1\right]\\ \frac{\partial^2 H}{\partial b_i \partial b_j} &= -tr\left[(R-B)^{-1}D_i^1(R-B)^{-1}D_j^1\right] - tr\left[(R-B)^{-1}D_{i,j}^2\right]\\ &\quad -l.tr\left[B^{-1}D_i^1B^{-1}D_j^1\right] + l.tr\left[B^{-1}D_{i,j}^2\right]\end{aligned}$$

with

$$\begin{aligned}D_i^1 &= \frac{\partial B}{\partial b_i}\\ &= \left(\sum_{i=0}^{p-1} b_i Z_i\right) Z_j^t + Z_j\left(\sum_{i=0}^{p-1} b_i Z_i\right)^t\end{aligned}$$

and

$$\begin{aligned}D_{i,j}^2 &= \frac{\partial^2 B}{\partial b_i \partial b_j}\\ &= Z_i Z_j^t + Z_j Z_i^t.\end{aligned}$$

The sign of the quadratic form $\mathbf{X}H_2\mathbf{X}$ is thus not at all evident. So the reasoning previously used for proving Proposition 4.1 cannot be directly extended to this parameterization. The simplicity of the proof of Proposition 4.1 is essentially due to the linear parameterization of the noise matrix B.

The concavity property seems (generally) wrong for the MA parameterization. This is illustrated by Fig. 1, which represents the level curves of the functional $H(b_0, b_1)$. The cross corresponds to the exact values of b_0 and b_1. However, even if Proposition 4.1 is not (generally) valid for the MA parameterization, the following proposition holds.

PROPOSITION 4.2. *The coefficients $(b_0^*, \ldots, b_p^*)$ of the MA process maximizing H on the constraint domain satisfy the following inequalities:*

$$\left|\left(b_i^* - b_i^0\right)\frac{1}{b_i^0}\right| \leq 1 - \sqrt{\frac{l}{l+1}}.$$

Furthermore, the gradient of H is null for a unique point of the parameter set; this point verifies the above proposition. This is a direct consequence of Proposition 4.1 and the one-to-one mapping between the coefficients (say $\{b_i\}$) of a min-phase MA model and its covariance set (say $\{\beta_i\}$). Therefore, there is a unique maximum of H for the MA parameterization on the constraint domain $\mathcal{C}$. Locating this point is not at all obvious since the correspondence between the MA parameters and the eigenvalues of the matrix $B_o^{-1}B_*$ is highly nonlinear. So this property will be proved by analytic arguments (see the proof of Proposition 5.1). A direct algebraic proof of Proposition 4.2 seems unfeasible.

Let us now consider practical considerations and, more precisely, iterative methods for maximizing H.

We shall now briefly present the gradient method.

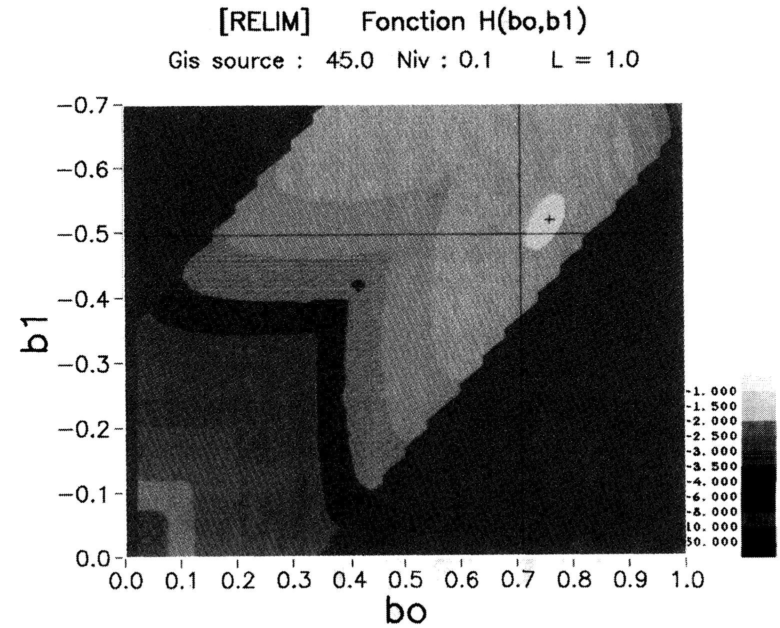

FIG. 1. *Values of the functional* $H(b_0, b_1)$, *one source* (*bearing* : 45 deg., -10dB).

The calculation of the gradient vector is straightforward. The ith component of the gradient vector G_k (at iteration k) is given by

$$G_k(i) = -tr\left[(R - B_k)^{-1} U_i\right] + l.tr\left(B_k^{-1} U_i\right). \tag{21}$$

(B_k is the noise matrix at iteration k.)

The gradient algorithm takes the following general form:

$$\mathbf{X}_{k+1} = \mathbf{X}_k - \rho_k \mathbf{G}_k \tag{22}$$

with

$$\mathbf{X}_k^t \triangleq \left(\beta_1^k, \dots, \beta_p^k\right)$$

and $\mathbf{G}$ defined by (21).

The scalar ρ_k is the stepsize of the algorithm. In order to ensure convergence (on $\mathcal{C}$) of the gradient algorithm, it is worth determining the optimal stepsize.

The matrix translation of (22) is

$$B_{k+1} = B_k - \rho_k D_k$$

with

$$\left\{\begin{array}{lcl} B_k & = & \sum_{i=1}^{p} \beta_i^k U_i, \\ D_k & = & \sum_{i=1}^{p} G_k(i) U_i. \end{array}\right. \tag{23}$$

The optimal stepsize ρ_k can be easily obtained by using eigendecompositions. The corresponding algorithm is presented below (and detailed in Appendix B).

Step 1. Since B_k and $R - B_k$ are positive definite, decompose them in triangular factors:

$$B_k = T_k T_k^*, \quad R - B_k = S_k S_k^*.$$

Step 2. Compute the eigenvalues $\{\alpha_i^k\}$ and $\{\beta_i^k\}$ of the two hermitian matrices:

$$S_k^{-1} D_k S_k^{-1*} \text{ and } T_k^{-1} D_k T_k^{-1*}.$$

Step 3. Then the REF becomes an explicit function of the parameter (stepsize) ρ, given by

$$H(\rho) = \sum_{i=1}^{q} \log\left(1 + \rho\alpha_i^k\right) + l \sum_{i=1}^{q} \log\left(1 - \rho\beta_i^k\right) + cst. \tag{24}$$

Furthermore (it is perhaps the most important fact), the positivity constraints $\mathcal{C}$ are translated into explicit (relatively to ρ) constraints, i.e.,

$$\mathcal{C}\left\{\begin{array}{lcll} 1 + \rho\alpha_i^k & > & 0, & i = 1, \ldots, q, \\ 1 - \rho\beta_i^k & > & 0, & i = 1, \ldots, q. \end{array}\right. \tag{25}$$

Step 4. The optimal stepsize ρ is simply obtained by maximizing $H(\rho)$ given by (24) under the constraints (25). This task is easily achieved by means of a unidimensional Newton method initialized at 0.

Obviously, the gradient method may be replaced by more sophisticated iterative methods (Newton, BFGS, etc.). However, this does not appear drastically important since the number of parameters defining the noise model (i.e., p) is generally quite smaller than q and because of Propositions 4.1 and 4.2. A direct extension to the complex case (the noise parameters are complex) is provided in Appendix C.

We now consider the unicity of the decomposition in source and noise parts. This is an important problem of identifiability [21]. The source's CM matrix S is assumed to be Toeplitz. Physically, this corresponds to the plane wave hypothesis and a uniform linear array assumption [1–7]. Since S is rank deficient and semipositive definite, S may be written in the following form (thanks to the theorem of Caratheodory [22]):

$$\left|\begin{array}{l} S = \sum_{j=1}^{s} \sigma_j \mathbf{Z}_j \mathbf{Z}_j^* \\ \text{with } : \sigma_j > 0, \\ \mathbf{Z}_j = \left(1, z_j, \ldots, z_j^{q-1}\right)^t, \ |z_j| = 1, \\ \text{rank } (S) = s. \end{array}\right. \tag{26}$$

Then a sufficient (and very rough) condition ensuring unicity of the decomposition will be obtained as follows. Assume the existence of two such decompositions. Then

$$R = S_1 + B_1 = S_2 + B_2;$$

hence

$$S_1 - S_2 = B_2 - B_1$$

(B_1 and B_2 are p-banded Toeplitz matrices).

In order to annihilate the noise effect, we consider the $\left(\frac{q-p}{2} \times \frac{q-p}{2}\right)$ lower left submatrix L of $S_1 - S_2$ defined by

$$L(i,j) = (S_1 - S_2)\,(q + i - p, j)\,, \qquad 1 \leq i, j \leq \frac{q-p}{2}.$$

We assume, in the previous equation, that $q-p$ is even. Otherwise, it must be replaced by $q-p-1$. All the components of the matrix L must be null since B_1 and B_2 are two p-banded Toeplitz matrices. The following equality is then directly obtained from (26):

$$\sum_{j=1}^{s} \sigma_{j,1} \mathbf{Z}'_{j,1} (\mathbf{Z}''_{j,1})^* = \sum_{j=1}^{s} \sigma_{j,2} \mathbf{Z}'_{j,2} (\mathbf{Z}''_{j,2})^*$$

with

$$(27) \qquad \left| \begin{array}{l} \mathbf{Z}'_{j,1} \triangleq \left(z_{j,1}^{q-q'}, \ldots, z_{j,1}^{q-1} \right), \\ \mathbf{Z}''_{j,1} \triangleq \left(z_{j,1}^{0}, \ldots, z_{j,1}^{q'-1} \right), \\ q' \triangleq \dfrac{q-p}{2} \qquad (j = 1, \ldots, s). \end{array} \right.$$

Now there exist coefficients $\{a_0, a_1, \ldots, a_s\}$ such that the roots of the polynomial equation

$$A(z) = a_0 + a_1 z + a_2 z^2 + \cdots + a_s z^s = 0$$

are $\{z_{1,1}, \ldots, z_{s,1}\}$. Hence the $s \times (q-p)$ matrix N defined by

$$N \triangleq \begin{pmatrix} a_0 & \cdots & a_s & 0 & \cdots & 0 \\ 0 & a_0 & \cdots & a_s & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & a_0 & \cdots & a_s \end{pmatrix}$$

annihilates the columns of Z_1' (Z_1' is the rectangular matrix whose colums are the vectors $\mathbf{Z}'_i$; $i = 1, \ldots, s$) when $s \leq q'$. In other words,

$$N Z_1' = 0,$$

which implies that $N Z_2' = 0$.

This implication is easily shown by considering the columns of Z_1' and Z_2'. They span the same space since by (27)

$$Z'_1 \Delta_1 Z_1^{'',*} = Z'_2 \Delta_2 Z_2^{'',*}$$

with

$$\Delta_1 = \operatorname{diag}(\sigma_{1,1}, \ldots, \sigma_{s,1}), \qquad \Delta_2 = \operatorname{diag}(\sigma_{1,2}, \ldots, \sigma_{s,2}).$$

Consequently, $\{z_{1,2},\ldots,z_{s,2}\}$ are also the roots of $A(z)$. So there exists a permutation matrix P such that $Z_1 = Z_2P$.

Now assume that the dimension of $\mathbf{Z}'_j$ (i.e., q') is superior to s. Then using a basic result on Vandermonde determinants [23], the solution to (27) is either

$$\left| \begin{array}{ll} \sigma_{j,1} = 0, & j = 1,\ldots,s, \\ \sigma_{j,2} = 0, & j = 1,\ldots,s, \end{array} \right.$$

or, for each source index j (related to S_1), there exists a source index k (related to S_2) such that

$$\left| \begin{array}{l} \sigma_{j,1} = \sigma_{k,2}, \\ z_{j,1} = z_{k,2} \end{array} \right. \tag{28}$$

or, equivalently, there exists a $s \times s$ permutation matrix P such that

$$Z_1 = Z_2P. \tag{29}$$

(The matrices Z_1 and Z_2 are the rectangular matrices (27) whose columns are the source vectors.)

Thus, the following proposition holds.

PROPOSITION 4.3. *Assume that S and B are Toeplitz matrices and assume, furthermore, that $q - p$ is greater than $2s$. Then the decomposition in source and noise matrices is unique.*

Note that the plane wave assumption or, equivalently, the Toeplitz hypothesis, has been instrumental for proving Proposition 4.3, which may be only considered as a sufficient and rough identifiability condition. The identifiability problem is greatly complicated by the noise correlation. In this case, the noise subspace has no clear algebraic meaning, as opposed to the white noise case.

We would like to stress that the accuracy of noise parameter estimates (i.e., the $\{\beta_i\}$) is expressed only in terms of the eigenvalues of $B_0^{-1}B_*$ and not directly in terms of the β_i. However, these two subsets are strongly related even if these relations are nonexplicit and highly nonlinear in the general case.

Actually, there is a one to one correspondence between the noise parameter vector β and the vector of the eigenvalues of the matrix $B_0^{-1}B_*$.

Indeed, consider the Jacobian matrix of partial derivatives [17]:

$$J \triangleq \begin{pmatrix} \frac{\partial\lambda_1}{\partial\beta_1} & \cdots & \frac{\partial\lambda_q}{\partial\beta_1} \\ \vdots & & \vdots \\ \frac{\partial\lambda_1}{\partial\beta_p} & \cdots & \frac{\partial\lambda_q}{\partial\beta_p} \end{pmatrix}.$$

Using the classical lemma (simple eigenvalues) [17] we have

$$\frac{\partial\lambda_i}{\partial\beta_j} = \mathbf{V}_i^* \frac{\partial}{\partial\beta_\ell}\left(B_0^{-1}B_*\right)\mathbf{V}_i'$$

and thus

$$J = \begin{pmatrix} tr\left(B_0^{-1}U_1\mathbf{V}_1\mathbf{V}_1^*\right) & \cdots & tr\left(B_0^{-1}U_1\mathbf{V}_q\mathbf{V}_q^*\right) \\ \vdots & & \vdots \\ tr\left(B_0^{-1}U_p\mathbf{V}_1\mathbf{V}_1^*\right) & \cdots & tr\left(B_0^{-1}U_p\mathbf{V}_q\mathbf{V}_q^*\right). \end{pmatrix}. \tag{30}$$

Now the matrices $\{\mathbf{V}_i\mathbf{V}_i^*\}_{i=1}^q$ are linearly independent in $M_q(\mathbb{C})$ (the space of hermitian $q \times q$ matrices) and, consequently, the rank of the rectangular matrix J is generally equal to p ($q > p$).

Finally, when l becomes great the eigenvalues of $B_0^{-1}B_*$ approach 1 (Proposition 3.1) and the parameter vector $\mathbf{B}_*$ approaches $\mathbf{B}_0$. Using simple algebraic considerations, it seems difficult to go further, but as will be seen in §5, an analytic formulation of the REF will allow us to refine the results of Proposition 3.1.

5. Analytic properties of the REF. The REF properties, previously considered, rely upon algebraic considerations. We shall see now that the REF definition can be translated in terms of analytic functions, allowing us to make the REF properties precise.

Let $R(k)$ be the (spatial) density of the stationary process received by the array. For a uniform array, $R(k)$ is simply the Fourier transform of the covariance matrix R, i.e.,

$$R(k) = \sum_{j=1-q}^{q-1} r_j \exp\left(2i\pi kjd\right)$$

with

$$\begin{aligned} &R = \text{Toepl }\left(r_0, r_1, \dots, r_{q-1}\right), \\ &d \ : \ \text{intersensor distance}, \\ &k = (d/\lambda)\cdot \sin\theta, \qquad \lambda : \text{wavelength},\ \theta : \text{bearing}. \end{aligned} \tag{31}$$

Even if the scalar d has a physical meaning, this meaning may be forgotten for what follows. Using Szego's theorem [22] one obtains (for q large)

$$\lim_{q\to\infty} \frac{1}{q} \log\det R = \frac{1}{2w}\int_{-w}^{w} \log R(k)dk, \tag{32}$$

where w is the spatial bandwidth.

Once again the physical meaning of w is not at all fundamental for what follows. Usually it is assumed to be $1/2$. Hence for a large value of q, the REF can be expressed as follows:

$$H = \int_{-w}^{w} \log\left(R(k) - B(k)\right)dk + l. \int_{-w}^{w} \log B(k)dk. \tag{33}$$

An MA noise modelling (5) seems to be quite convenient since it avoids the positivity problems while conserving the banded Toeplitz structure of the covariance matrix. For this model, $B(k)$ is given by

$$\left| \begin{aligned} &B(k) = \mid F(z) \mid^2 \\ &\text{with} \\ &F(z) = b_0 + b_1 z + \cdots + b_{p-1}z^{p-1}, \\ &z = \exp(2i\pi kd), \\ &i^2 = -1. \end{aligned} \right. \tag{34}$$

Then the following proposition of the REF holds.

PROPOSITION 5.1. *Assume that the noise may be exactly modelled by an MA process* $(b_0^0, b_1^0, \ldots, b_{p-1}^0)$. *Then for any MA modelling of the same order* (p), *the coefficients* $(b_0^*, b_1^*, \ldots, b_{p-1}^*)$ *of the MA process that maximizes* H (33) *under* C *satisfy the following set of inequalities:*

$$\left|(b_i^* - b_i^0)\frac{1}{b_i^0}\right| \leq 1 - \sqrt{\frac{l}{l+1}}, \qquad i = 0, 1, \ldots, p-1.^{1}$$

Proof. Previously, the proofs basically used the tools of linear algebra. From now on they will be replaced by complex analysis arguments. A direct approach will be considered. More precisely, the study of the sign of functionals involving partial derivatives (e.g., $\sum \lambda_i \partial H/\partial b_i$) will be instrumental.

The REF H is given by (33) and its partial derivatives (with respect to the $\{b_i\}$) are straightforwardly obtained:

$$\frac{\partial H}{\partial b_i} = \int_{-w}^{w} \frac{\mathrm{Re}\left(z^i \bar{F}(k)\right) \cdot [lR(k) - (l+1)B(k)]}{(R(k) - B(k))\, B(k)} dk. \tag{35}$$

(Re: real part of a complex number, $\bar{z}$: complex conjugate of z.)

We first consider the noise alone case. Then $R(k) = B_0(k)$ and the partial derivatives $\partial H/\partial b_i$ become

$$\frac{\partial H}{\partial b_i} = \int_{-w}^{w} \frac{\mathrm{Re}\left(z^i \bar{F}(k)\right) [lB_0(k) - (l+1)B(k)]}{(B_0(k) - B(k))\, B(k)} dk.$$

Because of the independence of the functions $\{\mathrm{Re}\left(z^i \bar{F}(k)\right)\}$ in the polynomial space, there exists a set of scalars $\{\lambda_i\}$ such that

$$\sum \lambda_i \mathrm{Re}\left(z^i \bar{F}(k)\right) = lB_0(k) - (l+1)B(k),$$

which implies

$$\sum \lambda_i \frac{\partial H}{\partial b_i} = \int_{-w}^{w} \frac{(lB_0(k) - (l+1)B(k))^2}{(B_0(k) - B(k))\, B(k)} dk. \tag{36}$$

Let us now examine the right member of (36). Clearly, the integrand is positive since the function $B_0(k) - B(k)$ must be positive on $[-w, +w]$ because of the definition of the REF. Therefore, $\sum \lambda_i \partial H/\partial b_i$ is positive and is null if and only if the numerator $(lB_0(k) - (l+1)B(k))$ is almost everywhere (a.e.) null on $[-w, w]$ or

$$B(k) = (l/l+1)B_0(k), \qquad k \in [-w, w],$$

which implies

$$b_i = \sqrt{\frac{l}{l+1}} \cdot b_i^0, \qquad i = 0, 1, \ldots, p-1.$$

Consequently, the gradient of H is null only when $b_i = \sqrt{l/l+1} \cdot b_i^0 (\, i = 0, 1, \ldots, p-1)$. Proposition 5.1 is thus proved for this (special) case.

[1] The coefficients b_i are assumed to be real.

We now assume that at least a source is present and we consider the following functional of the partial derivatives:

$$\begin{aligned}\sum_{i=0}^{p-1} b_i \frac{\partial H}{\partial b_i} &= \int_{-w}^{w} \frac{\operatorname{Re}\left(\left(\sum_{i=1}^{p-1} b_i z^i\right) \bar{F}(k)\right) [lR(k) - (l+1)B(k)]}{(R(k) - B(k))\, B(k)} dk \\ &= \int_{-w}^{w} \frac{lR(k) - (l+1)B(k)}{R(k) - B(k)}\, dk \\ &= I_1 + I_2\end{aligned}$$

with

$$I_1 = l \int_{-w}^{w} \frac{S(k)}{R(k) - B(k)}\, dk,$$

$$I_2 = l \int_{-w}^{w} \frac{B_0(k) - (l + 1/l)\, B(k)}{R(k) - B(k)}\, dk, \tag{37}$$

and

$$R(k) = S(k) + B_0(k), \qquad B(k) \text{ given by (34).}$$

The scalar product $\sum b_i \partial H/\partial b_i$ is thus written as the sum of the terms I_1 and I_2. We shall now examine them.

First, note that $(R(k) - B(k))$ is positive no matter what k is in the interval $[-w, +w]$. This is due to the definition of the REF and involves the barrier functional $\log(R(k) - B(k))$. Furthermore, $S(k)$ is also positive (it is a power spectral density), so that the term I_1 is always positive.

Second, now assume that a single source is present. Then

$$S(k) = \frac{\sigma^2}{|z - z_0|^2}$$

with z_0 the pole of the source, $z_0 \in D(0,1)$.

Then the following inequality holds:

$$\int_{-w}^{w} \frac{S(k)}{R(k) - B(k)} dk \geq \frac{1}{\alpha} \int_{-w}^{w} S(k) dk = \frac{1}{\alpha} \frac{\sigma^2}{1 - |z_0|^2}. \tag{38}$$

(α : lower bound of $R(k) - B(k)$ on $[-w, +w]$, σ^2 source power.)

Let us consider the term I_2. For that purpose, it is worth partitioning the parameter domain $(\{b_i\}_i)$ into zones, as depicted below. For the sake of clarity, only the two-dimensional (2-D) case will be presented in Fig. 2.

Let us now prove that the maximum of H cannot be attained on $\mathcal{Z}_1$. More precisely, assume that the coefficients $\{b_i\}$ satisfy the following inequalities (defining $\mathcal{Z}_1$):

$$|b_i| < |b_i^0| \sqrt{\frac{l}{l+1}} \quad \text{for} \quad i = 0, 1, \ldots, p-1.$$

Now Parseval's equality asserts that

$$\int_{-w}^{w} B_0(k) = \sum_{i=0}^{p-1} (b_i^0)^2$$

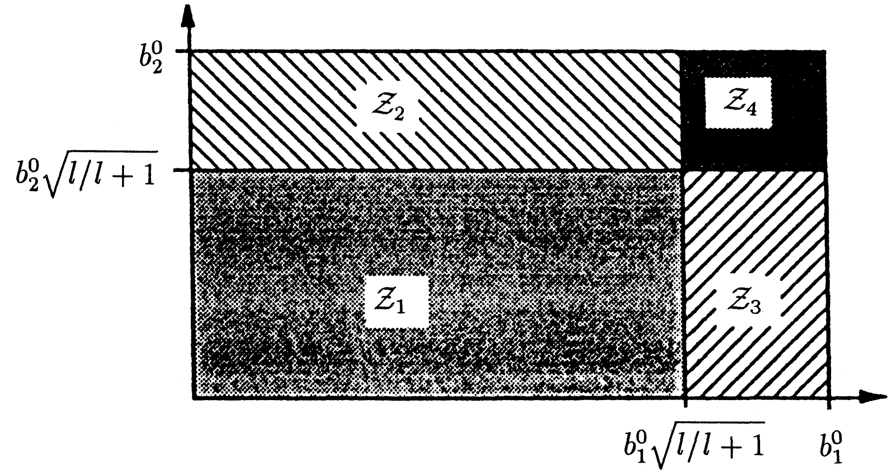

FIG. 2. *Decomposition of the positive orthant in* 4 *zones* $(\mathcal{Z}_1, \mathcal{Z}_2, \mathcal{Z}_3, \mathcal{Z}_4)$ b_1^0, b_2^0, $b_1^0\sqrt{l/l+1}$, $b_2^0\sqrt{l/l+1}$.

and, consequently,

$$\int_{-w}^{w} \left(B_0(k) - \left(\frac{l+1}{l}\right) B(k)\right) dk = \sum_{i=0}^{p-1} \left[\left(b_i^0\right)^2 - \left(\frac{l+1}{l}\right)\left(b_i^2\right)\right],$$

so that

$$(39) \qquad | I_2 | \le \left(\frac{1}{\alpha}\right) \cdot \left(\sum_{i=0}^{p-1} \left[\left(b_i^0\right)^2 - \left(\frac{l+1}{l}\right)\left(b_i^2\right)\right]\right).$$

Therefore, the term I_2 is bounded on $\mathcal{Z}_1$. So when the source pole z_0 approaches the unit circle, then I_1 tends towards $+\infty$. Note that this is equivalent to the rank deficiency hypothesis for source CM. Finally, the following result has been obtained:

$$(40) \qquad \sum_{i=0}^{p-1} b_i \frac{\partial H}{\partial b_i} > 0 \quad \text{on } \mathcal{Z}_1.$$

When the source contribution is null (i.e., $S = 0$), then the partial derivatives $\partial H/\partial b_i$ are null for $b_i^* = b_i^0\sqrt{l/l+1}, i = 0, \dots, p-1$. The effect of the source is thus to displace the location of the maximum of H.

Let us now consider the zones $\mathcal{Z}_2$ and $\mathcal{Z}_3$.

If the coefficients $\{b_i\}$ belong to $\mathcal{Z}_2$ or $\mathcal{Z}_3$, then the term I_2 of (37) is not necessarily positive, but it remains bounded. Therefore when $|z_0|$ tends towards 1 (plane wave hypothesis) then one has once again

$$(41) \qquad \sum_{i=1}^{p-1} b_i \frac{\partial H}{\partial b_i} > 0 \quad \text{on } \mathcal{Z}_2, \mathcal{Z}_3.$$

Finally, if the coefficients $\{b_i\}$ approach their exact values $\{b_i^0\}$, then H tends toward $-\infty$. According to (40), (41) the maximum of H is thus attained on $\mathcal{Z}_4$, achieving the proof.

Obviously, the reasoning is strictly similar for the multiple source case. □

Jensen's theorem [24] can be used to calculate the REF. Thus, since $B(k)$ is analytic in $D(0,1)$ we obtain ($w = 1/2$)

$$\int_{-1/2}^{1/2} \log(B(k))dk = 2\left(\log|F(0)| - \sum_{i-1}^{p} \log(|z_i|\right),$$

where $\{z_i\}$ are the zeros of $B(k)$ inside the unit circle.

Thus, the following equality holds:

$$\int_{-1/2}^{1/2} \log(B(k))dk = 2\log(b_0).$$

It is rather surprising that the limit ($q \to \infty$) of the term $(1/q)\log(\det B(b_0,\dots,b_p))$ is simply $2\log(b_0)$. The first part (noise alone case) of Proposition 5.1 can be proved in this way (Jensen's theorem). However, practically, this proof is restricted to first-order MA models.

Practically, $R(k)$ must be replaced by an estimate $\hat{R}(k)$, generally obtained by Fourier transform of the spatial covariances:

$$\hat{R}(k) = \sum_{j=-q+1}^{j=q-1} \hat{r}(jd)w_j \exp(2i\pi kjd), \tag{42}$$

where $\hat{r}(jd)$ are estimates of the spatial covariances.

Estimates of $\hat{r}(jd)$ are themselves obtained by replacing the exact matrix R by an orthogonal projection of the periodogram matrix [11] on the Toeplitz subspace [15,16]. The scalars w_j represent the array weighting. They are necessary for sidelobe reduction and, overall, to ensure the positivity of $\hat{R}(k)$. For this purpose, the following weighting ensures the positivity constraint of $\hat{R}(k)$:

$$\mathbf{W}^t = (1, 1-1/q, 1-2/q, \dots, 1/q),$$

since it amounts to a consideration of $\hat{R}(k)$ defined by

$$\hat{R}(k) = \mathbf{D}_k^* \hat{R} \mathbf{D}_k.$$

The REF method can be easily extended to multifrequency analysis. Under the independence assumption, the following formulation of the REF is obtained:

$$H = \sum_{f=f\min}^{f\max} \left[\int \left[\log\left(\hat{R}(f_i,k) - B(f_i,k)\right) + l \cdot \log B(f_i,k)\right] dk\right]. \tag{43}$$

6. The whitening procedure. The more classical and direct whitening procedure consists of performing a Choleski factorization of the matrix B_* (i.e., $B_* = TT_*$, T triangular factor) and defining the whitened matrix R_w by

$$R_w = T^{-1}RT^{-1*}.$$

However, this approach suffers from some drawbacks, which may become important. Among them are the computation cost (for a large array) and the numerical conditioning if the matrix B_* is near to singularity. So the following procedure is generally preferable.

1. Determine an autoregressive (AR) model "equivalent" to the MA model. Let p' be the AR model order. The term "equivalent" means that the covariance sequence of the AR model is as close as possible to the MA ones. Usually, this is achieved by means of the Yule–Walker equation [11]. Standard procedures exist for this problem [25].

2. Consider the whitened matrix defined as follows:

$$R_w = A_w R A_w^*,$$

where A_w is a rectangular $q - p' \times q$ defined by

$$\begin{pmatrix} a_0 & \cdots & a_{p'} & 0 & \cdots & 0 \\ 0 & a_0 & \cdots & a_{p'} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & \cdots & 0 & a_0 & \cdots & a_{p'} \end{pmatrix}. \tag{44}$$

This whitening method enjoys the following properties.

1. R_w is a Toeplitz matrix.

2. The transform of a source CM matrix (i.e., $\mathbf{D}_\theta \mathbf{D}_\theta^*$) is a rank-one matrix associated with the same bearing and given by

$$A_w \left(\mathbf{D}_\theta \mathbf{D}_\theta^*\right) A_w^* = q(\theta)\, \mathbf{D}_\theta' \mathbf{D}_\theta'^*$$

with

$$q(\theta) = |A(z)|^2.$$

Both the Toeplitz and plane wave structures are preserved by using this whitening procedure. In the case of a very large array, the above formula suggests the following (approximated) whitening:

$$R_w(k) = \left(\frac{1}{q(\theta)}\right) \cdot R(k).$$

7. Computation results and further comments. In this section, the behavior of the REF will be illustrated by computation and simulation results. The covariance matrix of the sensor outputs is given by

$$R = \sum_{i=1}^{s} \sigma_i^2 D_{\theta_i} D_{\theta_i}^* + B_0 \tag{45}$$

(B_0 exact noise CM; D_{θ_i} steering vector [1] associated with a source coming from the bearing θ_i and with spectral density σ_i^2).

The aim of the following results is to illustrate the REF properties.

1. *Effects of signal to noise ratios and of the factor* l. The effects of the signal to noise ratios are illustrated by Figs. 3 and 4. The eigenvalues of $B_0^{-1} B_{l,*}$ are ranked in increasing order. The index of each eigenvalue is plotted in the x-axis and its corresponding value in the y-axis. For these two figures, both the source bearings and the noise parameters are similar. They differ only by the source powers ("level" indicates the source spectral density σ_i^2).

The REF (3) is maximized by using a standard gradient algorithm, initialized on $(\beta_1 = 0.1, \beta_2 = .0, \ldots, \beta_5 = .0)$ or, in other words $B_{init} = \lambda \cdot .\mathrm{Id}$ (λ is chosen "small").

Thanks to Propositions 3.1 and 4.1, the convergence of the iterative maximization algorithm is ensured no matter what initialization satisfies the constraints $\mathcal{C}$. The previous choice (for initialization) appears to the simpler. Once the gradient method has converged, a matrix $B_{l,*}$ is obtained for each value of l. For each value of l a horizontal dotted line ($y = l/l+1$) is plotted and the eigenvalues of $B_0^{-1}B_{l,*}$ are compared to this line.

Proposition 3.1 is verified no matter what the value of l and the signal to noise ratios are. The lowest eigenvalue of $B_0^{-1}B_{l,*}$ may be slightly inferior to the theoretical lower bound (i.e., $l/l+1$) because of the stopping rule of the iterative algorithm.

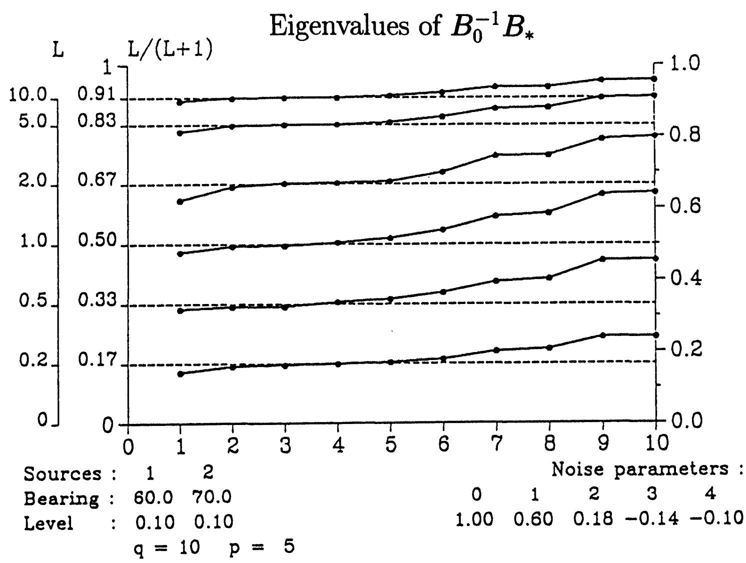

FIG. 3. *Verification of Proposition* 3.1. *Eigenvalues of the matrix* $B_0^{-1}B_{l,*}$ *for various values of* $l, p=5, q=10, s=2$. *Noise parameters :* $(\beta_1 = 1, \beta_2 = 0.6, \beta_3 = 0.18, \beta_4 = -0.14, \beta_5 = -0.10)$, *two sources* (*bearings :* 60 *and* 70 *deg., powers :* 0.1 *and* 0.1).

2. *Noise modelling overdetermination and Proposition* 4.3. The proof of Proposition 3.1 shows that it still holds when the noise model is overdetermined. This fact is illustrated by Fig. 5, where we assumed that the noise model was defined by 7 parameters when the true order was 5. The REF has been maximized with respect to $\beta_1, \ldots, \beta_7$. Note that the true parameters are those of Fig. 4. Proposition 3.1 is still verified in Fig. 4 even if an effect of overdetermination is an increase in the greater eigenvalues of $B_0^{-1}B_{l,*}$, thus enlarging the dispersion of the eigenvalues. Conversely, Proposition 3.1 is not verified if the noise model is underdetermined.

The effects of a "large" source number are presented in Fig. 6. Obviously, the dispersion of the eigenvalues is enlarged, but Proposition 3.1 still holds when the hypotheses of Proposition 4.3 are not satisfied. Proposition 4.3 thus appears to be very pessimistic. Note that Conjecture 1 is valid for all these simulations.

3. *Verification of Proposition* 5.1. Proposition 5.1 is illustrated by Table 1, for which the computation parameters are

$$q = 32,$$
$$B_0 \;\; MA(2) : b_0 = 1, \;\; b_1 = 0.3, \;\; b_2 = -0.3.$$

As can be seen in Table 1, Proposition 5.1 is verified no matter what the value of l is. The parameters $\{b_i^*\}$ have been computed by using a gradient algorithm for

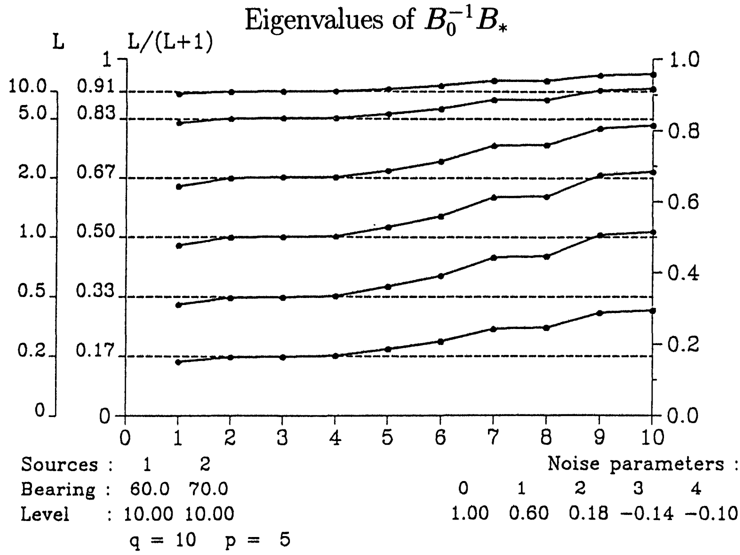

FIG. 4. *Verification of Proposition* 3.1. *Eigenvalues of the matrix* $B_0^{-1}B_{l,*}$ *for various values of* $l, p = 5, q = 10, s = 2$. *Noise parameters :* $(\beta_1 = 1, \beta_2 = 0.6, \beta_3 = 0.18, \beta_4 = -0.14, \beta_5 = -0.10)$, *two sources* (*bearings :* 60 *and* 70 *deg., powers :* 10 *and* 10).

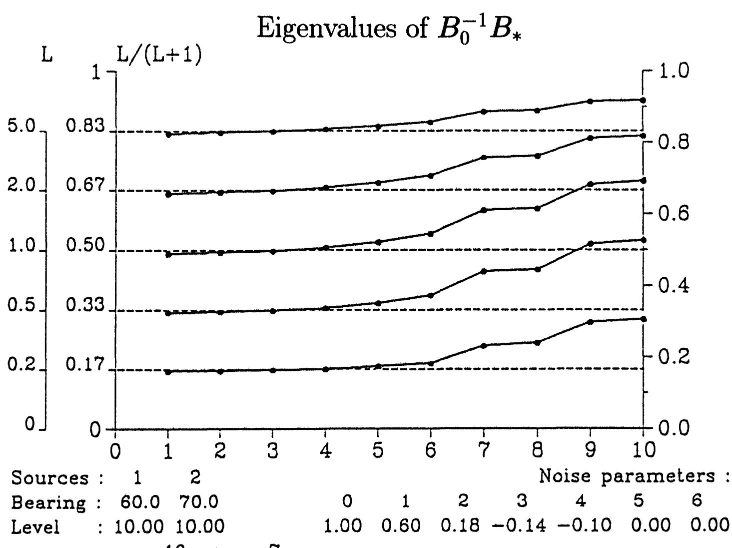

FIG. 5. *Effect of the noise model overdetermination. Eigenvalues of the matrix* $B_0^{-1}B_{l,*}$ *for various values of* $l, p = 7, q = 10, s = 2$. *Noise parameters :* $(\beta_1 = 1,\ \beta_2 = 0.6,\ \beta_3 = 0.18, \beta_4 = -0.14, \beta_5 = -0.10, \beta_6 = 0.0, \beta_7 = 0.0)$, *two sources* (*bearings :* 60 *and* 70 *deg., powers :* 10 *and* 10).

maximizing the REF H (30). Because it is quite direct, the calculation of the gradient is skipped. No convergence problem occurs.

4. *Simulation results.* Practically, the REF H (3) is replaced by the following functional:

$$H = \log\det\left(\hat{R} - B\right) + l \log\det B, \tag{46}$$

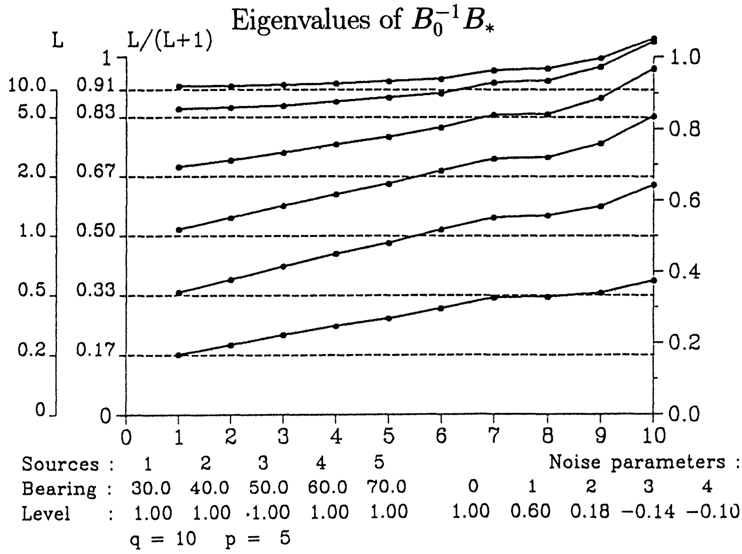

FIG. 6. *Verification of Proposition* 4.3. *Eigenvalues of the matrix* $B_0^{-1}B_{l,*}$ *for various values of* $l, p = 5, q = 10, s = 5$. *Noise parameters :* $(\beta_1 = 1, \beta_2 = 0.6, \beta_3 = 0.18, \beta_4 = -0.14, \beta_5 = -0.10)$, *five sources (bearings :* 30, 40, 50, 60, *and* 70 *deg., powers :* 1).

where $\hat{R}$ is an estimated CM of the sensor outputs.

The vectors $\mathbf{X}_i$ of array outputs have then been simulated. The general scheme of the simulation is presented below.

1. Let B_0 be the exact noise matrix, which performs a Choleski factorization of B_0, say,

$$B_0 = TT^*.$$

2. Let $\mathbf{Y}_i$ be a zero-mean gaussian complex vector of dimension q with covariance matrix Id; then a noise vector is $\mathbf{Y}_i' = T\mathbf{Y}_i$.

3. A source vector $\mathbf{S}_i$ is simulated by

$$\mathbf{S}_i = \sum_{j=1}^{s} \alpha_{i,j} \mathbf{D}_{\theta,j} \text{ with : } \alpha_{i,j} \ \mathcal{N}\left(0, \sigma_j^2\right).$$

The covariance matrix $\hat{R}$ is then estimated by the following.

1. $\hat{R}_1 = (1/N) \sum_{i=1}^{n} \mathbf{X}_i \mathbf{X}_i^*, \quad \mathbf{X}_i = \mathbf{S}_i + \mathbf{Y}_i'$.

2. $\hat{R} = \text{proj}(\hat{R}_1)$.

The projection is the orthogonal projection on the Toeplitz subspace, which is simply obtained by averaging along the diagonals [15–16]. The gradient algorithm is once again used for maximizing the REF (cf. §4). Since the initialization is not critical, we simply choose $B_{init} = \lambda$ Id. The value of λ must be inferior to the lowest eigenvalue of $\hat{R}$. After runs of the algorithm, noise estimates are obtained. The eigenvalues of the matrix $B_0^{-1}B_*$ are presented in Figs. 7 and 8 for 10 trials, each corresponding to $N = 300$. In other words, the snapshot number is 300.

Figures 7 and 8 correspond to the same simulated data; they differ only by the value of l. Proposition 3.1 advocates choosing a large l. This is not true for simulated data. If the value of l is 3, then Proposition 3.1 is "almost" valid. The statistical

TABLE 1
Values of $b_i^*(l)$ for various values of l.

Value of l		$\{b_i^*\}$ without source	$\{b_i^*\}$ with source	b_i^0	$b_i^0\sqrt{l/l+1}$
1	b_0	0.71	0.73	1.0	0.71
	b_1	0.21	0.22	0.3	0.21
	b_2	−0.21	−0.23	−0.3	−0.21
2	b_0	0.82	0.84	1.0	0.82
	b_1	0.24	0.25	0.3	0.24
	b_2	−0.24	−0.27	−0.3	−0.24
3	b_0	0.87	0.89	1.0	0.87
	b_1	0.26	0.26	0.3	0.26
	b_2	−0.26	−0.28	−0.3	−0.26
4	b_0	0.89	0.91	1.0	0.89
	b_1	0.27	0.27	0.3	0.27
	b_2	−0.27	−0.28	−0.3	−0.27
5	b_0	0.91	0.93	1.0	0.91
	b_1	0.27	0.28	0.3	0.27
	b_2	−0.27	−0.29	−0.3	−0.27
10	b_0	0.95	0.97	1.0	0.95
	b_1	0.29	0.29	0.3	0.29
	b_2	−0.29	−0.29	−0.3	−0.29

dispersion of the results of the various trials is rather reduced. This is not the case when the value of l is 10. It thus seems that there is an optimal value of l.

The choice of the optimal value of the parameter l results from statistical considerations relating the values of the parameters p, q, l with the statistical properties of the b_i''s estimates. Actually, the quantities defining the statistical behavior (standard deviation bias) of the b_i' estimates can be calculated by using an expansion of the $\{\hat{b}_i^*\}$ around their asymptotic values b_i^*.

This kind of calculation presents no major difficulty, but it is omitted here since it is beyond the scope of this paper. Roughly, there is a compromise between the accuracy of the b_i''s estimates (large values of l) and their variance. Thus l appears as an uncertainty factor describing the redundancy of information relative to the noise structure.

8. Conclusion. The properties of an original functional have been studied. They appear to be quite interesting, proving furthermore that the maximization of the REF is easy and reliable. The REF thus appears to be a promising method for solving an ill-posed problem.

Appendix A: A definition of the REF. We shall consider, for the REF definition, that the *physical* array is constituted of n_s equispaced sensors impinged by s sources, and we shall assume that the correlation length of the noise is null beyond p sensors.

Now consider the vector $\mathcal{B}$ defined by

$$\mathcal{B}^t = \left(\mathbf{B}_1^t, \mathbf{B}_2^t, \dots, \mathbf{B}_l^t\right),$$

where the vectors $\mathbf{B}_i$ are (statistically) independent sample vectors of the noise im-

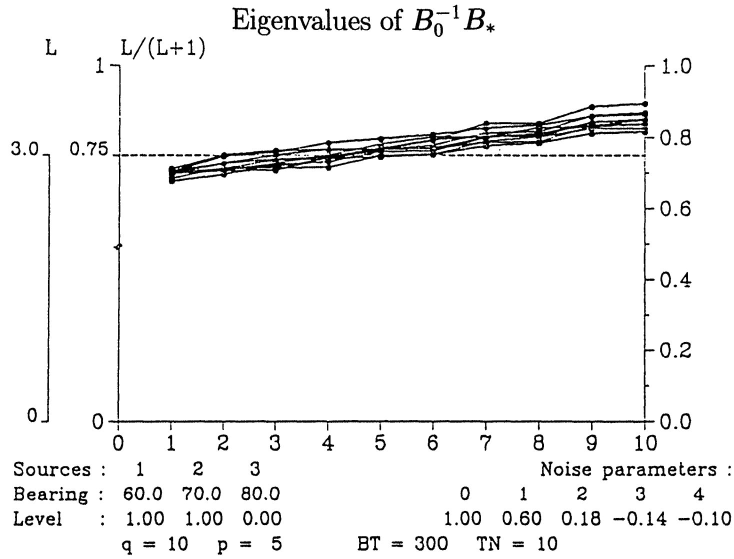

FIG. 7. *Simulated data. Eigenvalues of the matrix* $B_0^{-1}B_{l,*}$ *for* $l = 3$. *Ten trials, number of snapshots* $N = 300$, *two sources (bearings : 60 and 70 deg., powers:1 and 1), noise parameters* :($\beta_1 = 1, \beta_2 = 0.6, \beta_3 = 0.18, \beta_4 = -0.14, \beta_5 = -0.10$).

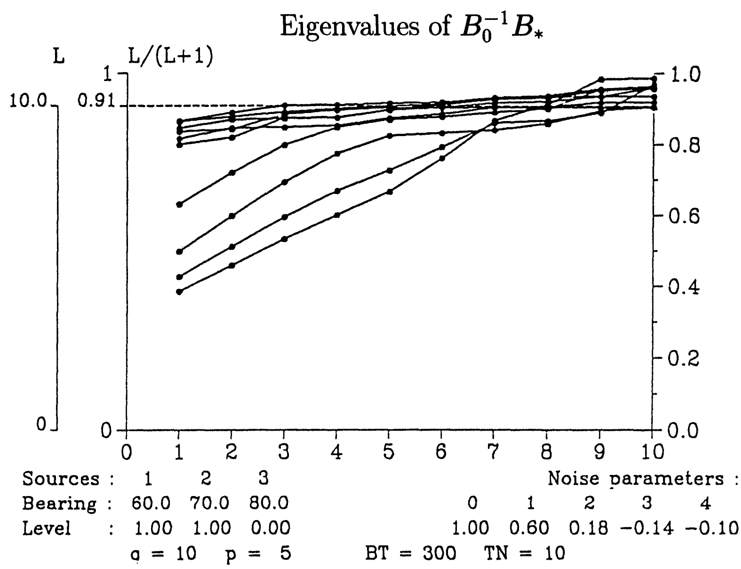

FIG. 8. *Simulated data. Eigenvalues of the matrix* $B_0^{-1}B_{l,*}$ *for* $l = 10$. *Ten trials, number of snapshots* $N = 300$, *two sources (bearings : 60 and 70 deg., powers:1 and 1), noise parameters* :($\beta_1 = 1, \beta_2 = 0.6, \beta_3 = 0.18, \beta_4 = -0.14, \beta_5 = -0.10$).

pinging on the array and with

$$\mathbf{B}_i \quad q - \text{dimensional vector.} \tag{47}$$

Furthermore, let $\mathbf{X}$ be an observation vector (sensor outputs) of the same dimension q. Then denote $\mathbf{B}_i|\mathbf{X}$ to be the linear minimum variance estimate of the zero-mean random vector $\mathbf{X}$ (i.e., the orthogonal projection of the random vector $\mathbf{B}_i$

on the Hilbert space spanned by $\mathbf{X}$). The following results:

$$\mathbf{B}_i|\mathbf{X} = \mathbb{E}\left(\mathbf{B}_i\mathbf{X}^*\right)\left[\mathbb{E}\left(\mathbf{X}\mathbf{X}^*\right)\right]^{-1}\mathbf{X},$$

$\mathbb{E}$ denotes expectation, and by a slight abuse of notation (concatenation of the projections), we have

$$\mathcal{B}|\mathbf{X} = \mathbb{E}\left(\mathcal{B}\mathbf{X}^*\right)\left[\mathbb{E}\left(\mathbf{X}\mathbf{X}^*\right)\right]^{-1}\mathbf{X}. \tag{48}$$

Then a "measure" of the "uncertainty" upon $\mathcal{B}$ which is not "explained" by $\mathbf{X}$ is deduced from the conditional variance of $\mathcal{B}$ and is equal to

$$I(\mathcal{B},\mathbf{X}) = \log\det\left[\operatorname{covar}\left(\mathcal{B} - \mathcal{B}|\mathbf{X}\right)\right]. \tag{49}$$

Actually, the observation vector $\mathbf{X}$ is the sum of a source part $(\mathbf{X})$ and a noise part $(\mathbf{B}_1)$, say,

$$\mathbf{X} = \mathbf{S} + \mathbf{B}_1. \tag{50}$$

Using (48) and (49), the following expression of $I(\mathcal{B},\mathbf{X})$ is easily derived, yielding

$$\begin{aligned} I(\mathcal{B},\mathbf{X}) &= \log\det\left\{\begin{pmatrix} B & & & \\ \ddots & \ddots & & 0 \\ 0 & \ddots & & \\ & & & B \end{pmatrix} - \begin{pmatrix} B \\ 0 \\ \vdots \\ 0 \end{pmatrix} R_q^{-1}(B\ \ 0\ \ 0)\right\} \\ &= \log\det\left(R_q - B\right) + l\cdot\log\det B - \log\det R_q. \end{aligned} \tag{51}$$

Appendix B: Optimization of the stepsize. This appendix is devoted to the calculation of the optimal stepsize ρ of the gradient's algorithm on $\mathcal{C}$.

The major aim of this appendix is to obtain an explicit formulation of the REF $H\left(R, B_{k+1}\right)$. For that purpose, consider the following factorization:

$$B_k = T_k T_k^* \quad\text{and}\quad R - B_k = S_k S_k^*$$

(by assumption B_k and $R - B_k$ are positive definite) so that

$$\begin{aligned} \log\det\left(R - B_k + \rho D_k\right) &= \log\det\left(S_k S_k^* + \rho D_k\right) \\ &= \log\det\left[S_k\left(\mathrm{Id} + \rho S_k^{-1} D_k S_k^{-1*}\right)S_k^*\right] \\ &= \log\det\left(R - B_k\right) + \log\det\left(\mathrm{Id} + \rho S_k^{-1} D_k S_k^{-1*}\right). \end{aligned} \tag{52}$$

Similarly, one obtains

$$\log\det\left(B_k - \rho D_k\right) = \log\det B_k + \log\det\left(\mathrm{Id} - \rho T_k^{-1} D_k T_k^{-1*}\right). \tag{53}$$

Therefore, using (52) and (53), the following results:

$$H(\rho) = \log\det\left(\mathrm{Id} + \rho S_k^{-1} D_k S_k^{-1*}\right) + l.\log\det\left(\mathrm{Id} - \rho T_k^{-1} D_k T_k^{-1*}\right) + cst. \tag{54}$$

The two matrices $S_k^{-1} D_k S_k^{-1*}$ and $T_k^{-1} D_k T_k^{-1*}$ are hermitian and therefore diagonalizable. Let $\left\{\alpha_i^k\right\}$ and $\left\{\beta_i^k\right\}$ be their respective eigenvalues. The following explicit form of $H(\rho)$ is thus

$$H(\rho) = \sum_{i=1}^{q}\log\left(1 + \rho\alpha_i^k\right) + l.\sum_{i=1}^{q}\log\left(1 - \rho\beta_i^k\right) + cst. \tag{55}$$

The two constraints $\mathcal{C}$ (6) are translated into explicit constraints (with respect to ρ), i.e.,

$$\left| \begin{array}{lll} 1+\rho\alpha_i^k > 0, & i=1,2,\ldots,q, & R-B_{k+1} \text{ positive definite}, \\ 1-\rho\beta_i^k > 0, & i=1,2,\ldots,q, & B_{k+1} \text{ positive definite}. \end{array} \right. \tag{56}$$

The optimal stepsize ρ_k is obtained by maximizing $H(\rho)$ (55) under the constraints (56). Practically, ρ_k is obtained by means of a unidimensional Newton method initialized at $\rho = 0$. The convergence of Newton's method on $\mathcal{C}$ is ensured since $H(\rho)$ is concave on this domain.

Appendix C: The complex case. The gradient algorithm for REF maximization will now be extended to the complex case. For that purpose, let V_i be the $q \times q$ matrix defined by

$$V_i = \left| \begin{array}{lll} 1 & \text{if} & l-k=i-1, \\ 0 & \text{else}. & \end{array} \right.$$

The noise matrix B then takes the following form:

$$B = \beta_1 U_1 + \beta_2 V_2 + \bar{\beta}_2 V_2^t + \cdots + \beta_p V_p + \bar{\beta} p V_p^t.$$

A real gradient vector $\mathbf{G}_k$ is then defined by

$$\mathbf{G}_k = \left| \begin{array}{l} g_k^1 = -tr\left(\Delta_k U_1\right), \\ g_k^2 = tr\left(\Delta_k (V_2 + V_2^t)\right), \\ g_k^{2'} = itr\left(\Delta_k (V_2 - V_2^t)\right), \\ \vdots \\ g_k^p = tr\left(\Delta_k (V_p + V_p^t)\right), \\ g_k^{p'} = itr\left(\Delta_k (V_p - V_p^t)\right), \end{array} \right.$$

with

$$\Delta_k = l \cdot \left(B_k{}^{-1}\right) - \left(R - B_k\right)^{-1}.$$

The gradient iteration takes then the following form:

$$B_{k+1} = B_k - \rho_k D_k$$

with

$$D_k = g_k^1 U_1 + g_k^2 (V_2 + V_2^t) + i g_k^{2'} (V_2 - V_2^t) + \cdots + g_k^p (V_p + V_p^t) + i g_k^{p'} (V_p - V_p^t).$$

The rest of the algorithm is strictly similar to the real case.

Acknowledgments. A major part of this research was done while the author worked with GERDSM, a laboratory of the "Direction des Constructions Navales" (DCN). The author wishes to thank DCN for its constant support, his former colleagues of GERDSM for their friendship, and the referees for their comments.

REFERENCES

[1] W. S. Burdic, *Underwater Acoustic System Analysis*, 2nd ed., Signal Proc. series, Prentice–Hall, Englewood Cliffs, NJ, 1991.

[2] G. Bienvenu and L. Kopp, *Optimality of high resolution array processing using the eigensystem method*, IEEE Trans. on ASSP, ASSP–31 (1983), pp. 1235–1247.

[3] D. H. Johnson, *The application of spectral estimation methods to bearing estimation problems*, Proc. of the IEEE, 70 (1982), pp. 1018–1028.

[4] W. S. Liggett, *Passive sonar: Fitting models to multiple time series*, in Signal Processing, J.W.R. Griffiths, P.L. Stocklin, and C. Van Schooneveld, eds., Academic Press, New York, 1979, pp. 327–345.

[5] S. U. Pillai, *Array Signal Processing*, Springer–Verlag, New York, Berlin, 1989.

[6] S. Haykin, ed., *Advances in Spectrum Analysis and Array Processing*, Vols. II and III, Prentice–Hall, Englewood Cliffs, NJ, 1991–1992.

[7] M. Bouvet and G. Bienvenu, eds., *High-resolution methods in underwater acoustic*, Lecture Notes in Control and Inform. Sci., 155, Springer–Verlag, New York, Berlin, 1991.

[8] M. S. Srivastava and G. G. Khatri, *An Introduction to Multivariate Statistics*, North–Holland, Amsterdam, 1979.

[9] R. J. Muirhead, *Aspects of Multivariate Statistical Theory*, John Wiley, New York, 1982.

[10] C. R. Rao, *The Use and Interpretation of Principal Component Analysis in Applied Research*, Sankhya Series A, Vol. 26, 1964, pp. 329–358.

[11] S. Kay, *Modern Spectral Analysis*, Prentice–Hall, Englewood Cliffs, NJ, 1988.

[12] J. P. Le Cadre, *Parametric methods for spatial signal processing in the presence of unknown colored noise fields*, IEEE Trans. on ASSP, 37, pp. 965–983.

[13] M. Wax, *Detection and localization of multiple sources in noise with unknown covariance*, IEEE Trans. on Signal Processing, 40 (1992), pp. 245–249.

[14] W. Murray and M. H. Wright, *Line search procedures for the logarithmic barrier function*, SIAM J. Optim., 4 (1994), pp. 229–246.

[15] P. A. Roebuck and S. Barnet, *A survey of Toeplitz and related matrices*, Internat. J. Systems Sci., 9 (1978), pp. 921–934.

[16] B. N. Mukherjee and S. S. Maiti, *On some properties of positive definite Toeplitz matrices and their possible applications*, Linear Algebra Appl., 102 (1988), pp. 211–240.

[17] G. H. Golub and C. F. Van Loan, *Matrix computations*, The Johns Hopkins University Press, Baltimore, MD, 1983.

[18] R. A. Horn and C. R. Johnson, *Matrix Analysis*, Cambridge University Press, Cambridge, UK, 1985.

[19] P. Stoica, personal communication.

[20] G. W. Stewart and J. G. Sun, *Matrix perturbation theory*, in Computer Science and Scientific Computing. Academic Press, New York, 1990.

[21] L. Ljung and T. Glad, *On global identifiability for arbitrary model parametrizations*, Automatica, 30 (1994), pp. 265–276.

[22] U. Grenander and G. Szego, *Toeplitz Forms and their Applications*, University of California Press, Berkeley, CA, 1958.

[23] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, 2nd ed., Academic Press, New York, 1985.

[24] R. P. Boas, *Invitation to Complex Analysis*, The Random House Birkhauser Mathematic Series, Random House, New York, 1987.

[25] A. Tikhonov and V. Arsenine, *Méthodes de résolution de problémes mal posés*. Editions MIR, Moscow, 1974.