

Parametric Methods for Spatial Signal Processing in the Presence of Unknown Colored Noise Fields

J. PIERRE LE CADRE

Abstract—This paper is devoted to the estimation of noise correlations along an array of sensors. The only hypothesis which is needed here is the adequacy of the ARMA model to noise description. Furthermore, the only available data are the sensor outputs. The aim of the methods which will be presented is thus to estimate a noise model from the sensor outputs in the presence of signal sources.

The eigenstructure methods developed thus far require that the additive noise be spatially white (uncorrelated between sensors) or that the noise correlation matrix be known to within a constant multiple. However, in most practical situations, this fundamental hypothesis is not verified, leading to important degradations of spatial processing in terms of bias, spurious peaks, angular resolution, and nondetection of small sources.

The additive noise received by the array is, in fact, the sum of ambient sea noise, flow noise, and traffic noise. It may be correlated along an important part of the array; thus, an ARMA modeling (involving only a small number of parameters) is well suited. Furthermore, the small number of AR model parameters (which are needed to describe the correlation sequence) is a great advantage for optimization of functionals.

In this paper, two types of methods for the estimation of noise parameters are presented. The first is related to the calculation of the likelihood of whitened observations (by means of ARMA modeling of noise) and the second is related to Pisarenko's method applied to whitened observations. Both methods are obtained by optimization of a criterion and are iterative. Obviously, noise estimates may be used for sensor outputs whitening and it is then a means to improve array processing performances. The two methods perform well, both on simulated and real data. However, the first method seems more attractive than the second (simpler and more robust).

I. INTRODUCTION

THIS paper deals with the estimation of noise correlations along an array of sensors. The following basic assumption is used in this paper: noise may be described by a (spatial) parametric model.

The aim of this paper is not to add a supplementary method for spectral (spatial) analysis; it is to present feasible methods for estimation of noise correlations in the presence of point sources (ships, etc.). Direct ARMA modeling of sensor outputs (sources plus noise) is a way to consider noise with arbitrary correlations, but it is not well suited to array processing, the main difficulty being due to high-order modeling (which is necessary for a great number of sources). The separation of the space of observations (sensor outputs) into two subspaces (i.e., noise

and sources subspaces) is a better way for noise correlation estimation. The main advantage of these approaches relies upon the low-order model of noise.

The adequacy of an AR(MA) modeling of noise is, obviously, a crucial point and will be considered later. It is also important to consider the improvements obtained by means of the noise correlation estimated.

The eigenstructure methods developed so far [1], [2] require that the additive noise be spatially white (uncorrelated between sensors) or that the noise correlation matrix be known to within a constant multiple. However, in most practical situations, the additive sensors noise is nonwhite and its correlations are unknown. If we persist to use standard methods, assuming that the sensor noise is white when in fact it is not so, then one notes important degradations in terms of bias, angular resolution, and especially spurious peaks, nondetection of weak sources. In fact, the main difference between point sources (isolated) and noise is their respective coherence. Sources are assumed to be perfectly coherent along the array, conversely to noise (even with important correlations). That means that source poles are on the unit circle, conversely to noise poles which are strictly inside the unit circle.

In this paper, we present principally two types of methods for the estimation of noise parameters (estimation of the ARMA coefficients). The first is related to the calculation of the likelihood of whitened observations (by means of ARMA noise modeling), the Gohberg formula [3], [4], and the perturbations of eigenvalues will be the keys of the method. This method does not use plane wave hypothesis unlike the second method. The second method relies upon the definition of a functional named whiteness functional which enjoys interesting properties. Both methods are obtained by optimization of a criterion and are iterative.

Obviously, the estimates of noise parameters may be used to whiten the observations (sensor outputs); efficient methods for that purpose are presented. It is then a means to improve array processing performances and to judge the proposed methods.

After a presentation of these two methods, some comparisons are presented for simulated data. We shall try to explain the results as also to justify the peremptory assertions of this Introduction. We shall consider especially the robustness of the proposed methods.

(*Notations:* Capital italic letters will denote matrices, while capital bold italic letters will denote vectors.)

Manuscript received November 17, 1986; revised June 2, 1988.
The author was with the Groupe d'Etudes et Recherche de Detection Sous-Marine, GERDSM, Le Brusq, 83140, Six-Fours-les-Plages, France. He is now with IRISA, Campus de Beaulieu, 35042, Rennes, France.
IEEE Log Number 8928129.

II. MAXIMUM LIKELIHOOD METHOD

A. Introduction and Formulation of the Problem

Let X be a stationary, n_s -dimensional random vector (n_s = number of sensors) constituted by the narrow-band outputs (e.g., after DFT) of an array sampling a homogeneous random field. Denote the covariance matrix of X by [5]

$$R = E(X \cdot X^*)$$

$$X^t = (x_1, \dots, x_{n_s}) \quad (1)$$

where the superscript * denotes transposition and conjugation, E meaning mathematical expectation.

Then

$$R = S + B \quad (2)$$

where S and B are the covariance matrices of sources and noise, respectively (sources and noise are assumed to be statistically independent).

The problem is to obtain an (accurate) estimation of the covariance matrix B (it is a cross-spectral matrix) from the only available statistics (i.e., \hat{R}).

The method relies upon three facts.

F1) The likelihood functional may be expressed as a functional of the eigenvalues of the whitened covariance matrix (of the outputs).

F2) The inverse B^{-1} of the noise covariance matrix admits an explicit formulation in terms of the AR(MA) coefficients of the noise model.

F3) The derivatives of the likelihood functional may be easily calculated using F1) and F2) and classical results for perturbations of eigenvalues.

B. Calculation of the Likelihood Functional

Consider, first, the white noise case (i.e., $B = \lambda \cdot Id$, $\lambda > 0$, Id meaning the identity matrix); assume furthermore that the source's number is q .

Let $\{X_1, \dots, X_N\}$ be a sequence of N independent complex Gaussian vectors (snapshot vectors) with covariance matrix R ; then the conditional density of this sequence is

$$f(X_1, \dots, X_N/\lambda, S)$$

$$= \frac{1}{\pi^{N \cdot n_s}} \cdot \frac{1}{(\det R)^N} \cdot \exp(-N \operatorname{tr}(\hat{R} \cdot R^{-1}))$$

where

$$\hat{R} = \frac{1}{N} \cdot \sum_{i=1}^N X_i \cdot X_i^*$$

$$R = S + \lambda Id$$

S being the unknown covariance matrix
of sources of rank $q < n_s$. (3)

We are now seeking the value of λ and the matrix S which maximize the likelihood. Forgetting the special structure of matrix S (plane wave hypothesis), we con-

sider that S is any rank q matrix and is described by its eigensystem [1]. A remarkable (and classical) fact is that the log likelihood may be simply expressed as a functional of the eigenvalues of \hat{R} , more precisely,

$$\sup f(X_1, \dots, X_N/(\lambda, S))$$

$$= N \cdot \left\{ -n_s \log \pi - n_s - \log \det \hat{R} + \sum_{i=q+1}^{n_s} \log \hat{\lambda}_i \right.$$

$$\left. - (n_s - q) \cdot \log \left(\frac{1}{n_s - q} \cdot \sum_{i=q+1}^{n_s} \hat{\lambda}_i \right) \right\} \quad (4)$$

Therefore, maximizing the likelihood amounts to maximizing the (negative) functional:

$$L_q(\lambda, S) = (n_s - q) \log ge(q)$$

$$- (n_s - q) \log ar(q) + cst \quad (5a)$$

where

$$ge(q) = \prod_{i=q+1}^{n_s} (\hat{\lambda}_i)^{1/n_s - q} \quad \text{geometric mean}$$

$$ar(q) = \frac{1}{n_s - q} \sum_{i=q+1}^{n_s} \hat{\lambda}_i \quad \text{arithmetic mean}$$

$$\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_{n_s} \quad \text{eigenvalues of } \hat{R}$$

$$cst = -\log \det \hat{R} - n_s \cdot (1 + \log \pi). \quad (5b)$$

Consider, now the case of an unknown correlated noise; then

$$L_q(B, S) = -\log \det R - \operatorname{tr}(R^{-1} \cdot \hat{R})$$

$$(R = S + \lambda \cdot B; \lambda \text{ scalar}). \quad (6)$$

Now B being a positive matrix (for the sequel, positive will mean positive definite [6]), it admits a Cholesky decomposition in triangular factors; let

$$B = L \cdot L^*. \quad (7)$$

Consider, furthermore, the whitened matrix of the exact covariances:

$$R_w = L^{-1} \cdot R \cdot L^{-1*} = L^{-1} \cdot S \cdot L^{-1*} + \lambda Id \quad (8a)$$

also as the whitened source's matrix:

$$S_w = L^{-1} \cdot S \cdot L^{-1*}. \quad (8b)$$

Then

$$\operatorname{tr}(R^{-1} \cdot \hat{R}) = \operatorname{tr}[(L^{-1} \cdot R \cdot L^{-1*})^{-1}$$

$$\cdot (L^{-1} \cdot \hat{R} \cdot L^{-1*})]$$

$$(\operatorname{tr}(AB) = \operatorname{tr}(B \cdot A))$$

$$= \operatorname{tr}(R_w^{-1} \cdot \hat{R}_w) \quad (9a)$$

and

$$\log \det R = \log \det [L \cdot (R_w) \cdot L^*]$$

$$= \log \det B + \log \det R_w. \quad (9b)$$

Then, A_1 and A_3 may be written in matrix form as

$$\begin{cases} A_1 = \sum_{i=0}^p a_i \cdot Z^i & (\text{by convention } a_0 = 1) \\ A_3 = \sum_{i=1}^p a_i \cdot Z^{n_s-i} \end{cases} \quad (16)$$

B^{-1} may be written as a quadratic form relative to the coefficients $\{a_i\}$ by (14) and (16). Note that B^{-1} defined by (14) is centrosymmetric [7] and its inverse (i.e., B) is a Toeplitz matrix [7].

The matrix A_3 can be considered as a corrective term. This formula is extended to the complex case in Appendix A.

To our knowledge, there is no explicit formulation of B^{-1} for the MA modeling; however, a simple expression of B is given by [8]

$$B = \left(\sum_{i=0}^p b_i Y_i \right) \left(\sum_{i=0}^p b_i Y_i \right)^t \quad (17)$$

when $\{b_i\}_{i=0}^p$ are the MA coefficients and $\{Y_i\}$ are rectangular ($n_s \times 2n_s$) matrices defined by

$$Y_i(j, k) = \begin{cases} 1 & \text{if } k - j = i \\ 0 & \text{else.} \end{cases}$$

The parametrization of B for ARMA modeling is obtained by combination of (14) and (17), i.e.,

$$B = \left(\sum_{i=0}^p b_i Y_i \right) \cdot R_{AR} \cdot \left(\sum_{i=0}^p b_i Y_i \right)^t \quad (18a)$$

where R_{AR} is the extended ($2n_s \times 2n_s$) matrix of the AR process, i.e.,

$$R_{AR} = \sigma^2 \cdot (A_1 \cdot A_1^t - A_3 \cdot A_3^t)^{-1}$$

with

$$A_1 = \sum_{i=0}^p a_i \cdot Z^i \quad (18b)$$

(Z^i defined as in (15) but in dimension $2n_s$).

The previous models assume that the receiver-generated component of the noise is of equal power in each sensor. If this assumption does not hold, a better model for the noise received on sensor m would be

$$X_m + a_1 \cdot X_{m-1} + a_p \cdot X_{m-p} = \beta_m \cdot w_m$$

w_m being a sequence of independent white noise.

The parameter β_m is an additional parameter which we would have to optimize with respect to.

Numerous parametrizations of noise covariance matrices may be used, but those which are considered have the greatest advantage to define a parametrization of B or even B^{-1} with a small number of parameters.

We shall now consider the maximization of the likelihood functional (13) in the case of a real AR model of noise.

D. Maximization of the Likelihood Functional (Real) AR Case

In the general case, iterative methods are needed for maximization of L_q ; the convergence of these methods will be studied in Appendix B.

A general form of a gradient algorithm is given by

$$A_{k+1} = A_k - \rho_k \cdot G_k$$

$$A_k^t = (\sigma_k^2, a_1^k, \dots, a_p^k)$$

being the estimated parameters at the k th iteration. (19)

(σ_k^2 has no theoretical necessity since the likelihood functional is invariant under scale change, but it may be preferable to include it, for practical applications, in the gradient algorithm.)

G_k is the gradient vector at the k th iteration, while ρ_k is the step size of the algorithm. A_k being given, $(B_k)^{-1}$ is obtained by (14), (15), i.e.,

$$B_k^{-1} = (\sigma_k^2)^{-1} \cdot [A_1^k \cdot (A_1^k)^t - A_3^k \cdot (A_3^k)^t]$$

with

$$A_1^k = \sum_{i=0}^p a_i^k \cdot Z^i. \quad (20)$$

The problems consist now in the calculation of the gradient vector G_k whose i th component is

$$G_k(i) = \frac{\partial}{\partial a_i} L(q, k) \quad (21)$$

($L(q, k)$ is the log-likelihood functional conditional to the q sources hypothesis and to B_k).

Calculation of G_k amounts to calculation of partial derivatives $(\partial/\partial a_i) \lambda_j^k$ (λ_j^k are eigenvalues of $\hat{R}_{w,k}$). However, some intermediate steps are needed; we shall consider, for the first time, the case of simple eigenvalues.

Lemma 1: The eigenvalues of $\hat{R}_{w,k}$ and $B_k^{-1} \cdot \hat{R}$ are identical.

Indeed,

$$\begin{aligned} \det(B_k^{-1} \cdot \hat{R} - \lambda Id) &= \det(L_k \cdot L_k^* \cdot \hat{R} - \lambda Id) \quad (L_k \cdot L_k^* = B_k^{-1}) \\ &= \det(L_k^* \cdot \hat{R} \cdot L_k - \lambda Id) \\ &= \det(\hat{R}_{w,k} - \lambda Id). \end{aligned} \quad (22)$$

\hat{R} being a positive matrix (covariance matrix), it can be decomposed in triangular factors; let $\hat{R} = T \cdot T^*$, leading to the following result.

Lemma 2: The eigenvalues of $R_{w,k}$ and $T^* \cdot B_k^{-1} \cdot T$ are identical.

Indeed,

$$\begin{aligned} \det(B_k^{-1} \cdot R - \lambda Id) &= \det(B_k^{-1} \cdot T \cdot T^* - \lambda Id) \\ &= \det(T^* \cdot B_k^{-1} \cdot T - \lambda Id). \end{aligned} \quad (23)$$

It can be deduced from Lemmas 1 and 2 that calculation of partial derivatives $(\partial/\partial a_i)\lambda_j$ amounts to calculation of partial derivatives $(\partial/\partial a_i)\mu_j$, $\{\mu_j\}$ being the eigenvalues of the Hermitian matrix $T^* \cdot B_k^{-1} \cdot T$.

This result has a great practical importance since B_k^{-1} is an explicit function of the AR coefficients.

Recall, furthermore, a classical result for the calculation of partial derivatives of a *simple* eigenvalue of a Hermitian matrix A depending (differentiably) upon a set of parameters [9]:

$$R_1: \frac{\partial}{\partial a_i} \lambda_j = \mathbf{V}_j^* \cdot \frac{\partial}{\partial a_i} A \cdot \mathbf{V}_j$$

$$\left(\begin{array}{l} \mathbf{V}_j \text{ eigenvector associated to } \lambda_j; \\ \left(\frac{\partial}{\partial a_i} A \right)(k, l) = \frac{\partial}{\partial a_i} (A(k, l)) \end{array} \right). \quad (24)$$

Calculation of vector \mathbf{G}_k is deduced from the preceding results and its steps are presented below.

Step 1: Calculation of partial derivatives Δ_i^k ($0 \leq i \leq p$) of B_k^{-1} relative to parameters $\{a_i\}$, i.e.,

$$\Delta_i^k = \frac{\partial}{\partial a_i} B_k^{-1} = \frac{1}{\sigma_k^2} [Z^i \cdot (A_1^k)^t + A_1^k \cdot (Z^i)^t - Z^{n_s-i} \cdot (A_3^k)^t - A_3^k \cdot (Z^{n_s-i})^t]$$

for $1 \leq i \leq p$

$$\Delta_0^k = -2(\sigma_k)^{-1} \cdot B_k^{-1}$$

(Δ_i^k defined by (25) is symmetric). (25)

Step 2: Calculation of derivative matrices $\Delta_i^{k'}$:

$$\Delta_i^{k'} = \frac{\partial}{\partial a_i} (T^* \cdot B_k^{-1} \cdot T) = T^* \cdot \Delta_i^k \cdot T$$

$$0 \leq i \leq p. \quad (26)$$

Step 3: Calculation of partial derivatives of the (simple) eigenvalues of $\hat{R}_{w,k}$:

$$\frac{\partial}{\partial a_i} \lambda_j^k = (\mathbf{U}_j^k)^* \cdot (\Delta_i^{k'}) \cdot (\mathbf{U}_j^k)^*, \quad (27)$$

\mathbf{U}_j^k being an eigenvector associated to the eigenvalue λ_j^k of $T^* \cdot B_k^{-1} \cdot T$.

Step 4: Calculation of the gradient vector \mathbf{G}_k , defined by its components $G_k(i)$:

$$G_k(i) = \sum_{j=q+1}^{n_s} (\lambda_j^k)^{-1} \cdot \frac{\partial}{\partial a_i} \lambda_j^k - (ar^k(q))^{-1} \cdot \sum_{j=q+1}^{n_s} \frac{\partial}{\partial a_i} \lambda_j^k$$

where

$$ar^k(q) = \frac{1}{n_s - q} \cdot \sum_{j=q+1}^{n_s} \lambda_j^k. \quad (28)$$

Calculation of the gradient vector \mathbf{G}_k is defined by formulas (25), (26), (27), (28) in the (real) AR case. Calculation of \mathbf{G}_k requires knowledge of the eigensystem $\{\lambda_j^k, \mathbf{U}_j^k\}$ of $T^* \cdot B_k^{-1} \cdot T$. These eigenvalues and eigenvectors may be exactly computed (by standard algorithms) or estimated by use of formulas (27) and (30) for the sake of computation time.

It is important, however, to compute a satisfying step size (denoted ρ) for the gradient algorithm in order to ensure convergence. This will be achieved by use of a first (or higher)-order approximation to the change in eigenvalues of the whitened matrix.

More precisely, let Λ_j^k be the vector of partial derivatives of λ_j^k , i.e.,

$$(\Lambda_j^k)^t = \left(\frac{\partial}{\partial a_1} \lambda_j^k, \dots, \frac{\partial}{\partial a_p} \lambda_j^k \right). \quad (29)$$

Then a first-order approximation of $\lambda_j^k(\rho)$, j th eigenvalue of the whitened matrix $\hat{R}_{w,k}$ or $[B_k(\rho)]^{-1} \cdot \hat{R}$ with

$$B_k^{-1}(\rho) = (\sigma^2)^{-1} \cdot [A_1^k(\rho) \cdot (A_1^k(\rho))^t - A_3^k(\rho) \cdot (A_3^k(\rho))^t]$$

and

$$A_1^k(\rho) = \sum_{i=0}^p (a_i - \rho \cdot G_k(i)) \cdot Z^i$$

is given by

$$\lambda_j^k(\rho) \approx \lambda_j^k - \rho \mathbf{G}_k^t \cdot \Lambda_j^k. \quad (30)$$

Then the approximation of $\hat{\lambda}_j^k(\rho)$ defined by (27), (28), and (30) is substituted in the likelihood functional L_q (13), and by means of a *unidimensional* method, an approximated ρ is determined.

A second-order approximation can be easily calculated [10]. However, in all practical situations, a first-order approximation seems to be sufficient in order to ensure convergence of the algorithm. Furthermore, a higher order approximation is more expensive in computation time.

E. Some Extensions

1) Parametrization by Reflection Coefficients: The parametrization of the noise matrix B may be advantageously expressed in terms of reflection coefficients. Using it this way, the stability of the AR model is easily ensured.

More precisely, using the formalism of [11], the Levinson recursion may be written in matrix form:

$$\mathbf{A}_{p+1} = (\mathbf{I}_{p+2} - k_p \cdot J) \Delta \mathbf{A}_p \quad (31a)$$

where Δ is the $(p+2) \times (p+1)$ matrix defined by

$$\Delta \equiv \begin{pmatrix} 1 & & & \mathbf{0} \\ \mathbf{0} & \dots & & 1 \\ & & & \mathbf{0} \end{pmatrix}. \quad (31b)$$

(J is the reflection matrix [11], $J^2 = \mathbf{I}_d$.)

A_p is now the vector of AR coefficients of the noise model (14), i.e.,

$$A_p^t = (1, a_1, \dots, a_p). \quad (32)$$

Using (32), it is then possible to express A_{p+1} as an explicit function of reflection coefficients $(1, k_1, \dots, k_p)$, leading to

$$A_{p+1} = (Id - k_p \cdot J) \cdot \Delta \cdot (Id - k_{p-1} \cdot J) \cdot \Delta \cdot \dots \cdot (Id - k_1 \cdot J) \cdot \Delta \cdot 1. \quad (33)$$

It is now possible to express the partial derivatives of the log-likelihood functional with respect to reflection coefficients; more precisely,

$$\frac{\partial}{\partial k_i} L_q = \sum_{j=1}^p \frac{\partial}{\partial a_j} L_q \cdot \frac{\partial}{\partial k_i} a_j. \quad (34)$$

Now $\partial L_q / \partial a_j$ have already been calculated (Section II-D) and $\partial a_j / \partial k_i$ is straightforwardly deduced from (33); therefore, the partial derivatives $(\partial / \partial k_i) L_q$ are calculated by use of (32), (33), (34), and Section II-D. The test of stability [for the polynomial $A(z)$] now reduces to the conditions $(0 \leq |k_i| < 1)$.

2) *AR(MA) Case*: The MA model leads us to replace the matrix Δ_i^k defined in (25) by the matrix δ_i^k defined by

$$\delta_i^k = -(B_k)^{-1} \cdot \left[Y_i \cdot \left(\sum_{l=0}^p b_l^k \cdot Y_l \right)^t + \left(\sum_{l=0}^p b_l^k \cdot Y_l \right) \cdot Y_i^t \right] \cdot (B_k)^{-1}$$

while the ARMA case is obtained by derivation of (18a). (The MA model is assumed to be minimum phase [12].)

3) *Multiple Eigenvalues*: In the case of multiple eigenvalues, the formula (27) of the eigenvalues' partial derivatives no longer holds. In this case, the eigenvalues of the perturbed matrix $R + \delta R$ are, up to the first order, the eigenvalues of the matrix [13]

$$\Delta = U^* \cdot (R + \delta R) \cdot U$$

(U is the matrix of an eigenvector basis associated to the multiple eigenvalue), leading us to replace (27) by the computation of eigenvalues of the matrix

$$U^* \cdot \Delta_i^k \cdot U.$$

The other steps of the algorithm are identical.

4) *Complex Case*: The nonsymmetry of the noise field with respect to the array broadside results in complex AR(MA) coefficients for the noise model. The procedure in the complex case is quite similar to the previous one (see Appendix A).

F. Convergence Analysis

We are now dealing with the convergence of the gradient algorithm defined in Section II-D.

Consider, for instance, a given AR model (second order); its covariance matrix may be calculated by (14); it is then possible to compute the functional $L_{a_1, a_2}(q)$ de-

finied by (13) where the eigenvalues λ_i are the eigenvalues of the matrix $B_{a_1, a_2}^{-1} \cdot R$.

$L_{a_1, a_2}(q)$ is the exact log-likelihood functional conditional to $\{a_1, a_2\}$.

Fig. 1 presents the shape of the surface described by $-L_{a_1, a_2}$ on the stability domain (corresponding to AR parameters). There are three sources. The maximum of the functional is reached for the exact values of AR parameters (i.e., a_1^0 and a_2^0), but the functional is not concave on the whole domain.

The functional is not concave (for AR modeling), but it enjoys the following properties (*regardless of estimation problems of \hat{R} and assuming an appropriate noise model*).

Property 1: The log-likelihood functional defined by (13) is negative; it is null if and only if the noise model is perfectly estimated.

This is a simple consequence of the properties of arithmetic and geometric means [14]; the nullity of the functional is equivalent to the equality of the $(n_s - q)$ lowest eigenvalues of \hat{R}_w (or $B^{-1} \cdot \hat{R}$). Assume now $(n_s - q)$ superior to p (AR order); then it can be easily proved (by consideration of the Jacobian of the function relating $\{a_i\}$ to eigenvalues of B^{-1}) that $B = B_0$.

Property 2: The gradient vector of the log-likelihood functional is null if and only if A is equal to A_0 (exact values of the coefficients).

The rather technical proof of Property 2 is skipped in Appendix B; from Properties 1 and 2, it is then possible to establish convergence of the gradient algorithm in the *asymptotic case*. Obviously, the estimation of \hat{R} leads to statistical results.

In all the exposures, some parameters have been fixed; there are q (number of sources) and p (AR order). In practical situations, they are not known; we shall examine now the consequences of a misadjustment of these parameters.

1) *Number of Sources (q)*: By use of Property 1, it is possible to overdetermine the source's number as long as p (AR model) remains inferior to $(n_s - q)$ without degradation in the asymptotic case. If \hat{R} is estimated (practical case!), then the *source number overdetermination will lead to slightly inferior performances of the method (random eigenvalues), but the degradation is quite acceptable* (see Fig. 2). We now have to face a fundamental problem: what is the good "strategy" for the choice of q ? The answer is not evident.

The "classical" information criteria (Akaike, Rissanen, etc. [15]–[17]) do not provide a satisfying source number estimate because they use eigenvalues of \hat{R} and do not separate sources and (highly) correlated noise. For instance, the estimated source number is equal to the number of sensors for most of the simulations presented in Section IV. In order to remedy this problem, some solutions are presented below.

a) *Source Number Overdetermination*: It is possible to try consecutive values of q and to choose the value of q maximizing the log-likelihood functional.

b) *Modifications of Information Criteria*: The infor-

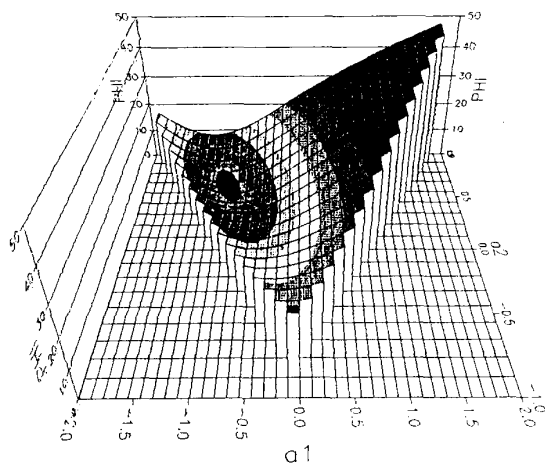


Fig. 1. Shape of the log-likelihood functional for three sources.

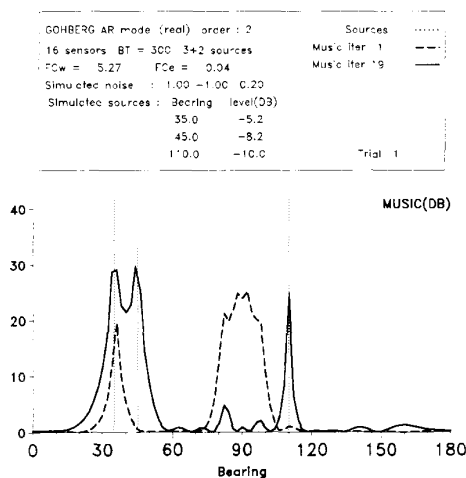


Fig. 2. Maximum likelihood method. Comparison between MUSIC method before whitening and after whitening.

mation theoretic criteria are based upon log-likelihood calculation (which is achieved by means of the presented gradient algorithm) and to its statistical behavior [8]. Its statistical behavior is characterized by the number of "free" parameters [12]; the inclusion of an AR noise model leads to add supplementary free parameters, but the spirit of the method is similar [19]; these approaches give satisfying results on simulated data.

More precisely, if the criterion of Rissanen [15], [16] is used, the following function is considered:

$$MDL(p, q) = -L(p, q) + \frac{\log N}{2N} \cdot [p \cdot (2n_s - p) + q + 2]$$

where

- N is the number of independent samples
- p is the source number
- q is the noise model order.

$L(p, q)$ is the likelihood functional conditional to the hypotheses: p sources and noise model order q . Obviously, $L(p, q)$ has no explicit formulation in the general case and must be calculated by use of the method presented in Section II-D.

The results of a simulation are now presented; the parameters of the simulation are the following:

$n_s = 32$, source number = 3, noise model order = 2.

Source bearing and corresponding powers:

Bearings: 40 deg 50 deg 150 deg

Powers: 0.3 0.4 0.2

Noise model:

$$a_0 = 1.00 \quad a_1 = -0.9.$$

The data are simulated by the method of Section IV-A, after which a covariance matrix is obtained. Table I presents the value of the MDL function for one trial (300 snapshots).

The minimum of the function $MDL(p, q)$ is attained at the exact values of p and q ; however, this minimum is not very sharp (in order). The above simulations have been repeated; then for ten trials (300 snapshots), 9 successes are obtained (estimation of the exact values of p and q). The results of this method for estimation of the parameters p and q are rather satisfying; however, its computation cost can be rather important. Therefore, a separated estimation of p and q can be an interesting way as in c) and d).

c) *Statistical Properties of the Eigenvalues of a "Corner" of Matrix R (MA Case)*: This original approach has been recently presented by Fuchs [13] and gives an accurate estimate of the sources number in the presence of correlated noise.

d) *Use of a State-Space Approach for Sensor Outputs Modeling (ARMA Case)*: By use of an information criterion like the predictive efficiency criterion presented by Arun and Kung [20], it is possible to obtain a satisfying estimate of the source's number without *a priori* information about the noise model. This application of Arun and Kung's method is presented and detailed in [21].

2) *Noise Order Model*: In the asymptotic case, noise model overdetermination leads to very slight degradations of the results (in terms of noise spatial density) as shown by Fig. 9 and by the proof of Property 2.

For all simulated data, it seems possible to overdetermine the noise model order without dramatic effects (see Fig. 9). The robustness of the method to misadjustments of these parameters is a crucial point; otherwise, the practical interest of the method should be very doubtful.

Obviously, the parameters p and q may be estimated simultaneously as shown previously, but a separate estimation of the order p seems rather difficult.

Another point to consider is the stability (with respect to the unit circle) of the AR model obtained by maximization of the functional. It is possible to compute the roots of $A_k(z)$ (polynomial associated to A_k) and to adjust con-

TABLE I

Noise Model Order	Source Number				
	1	2	3	4	5
1	21.688	13.851	3.786	4.764	5.782
2	25.378	16.644	3.686	4.745	5.776
3	25.591	16.953	3.710	4.761	5.793
4	25.697	16.891	3.713	4.768	5.799

sequently the step size of the gradient method, but the parametrization by reflection coefficients [Section II-E-1]) is the best suited approach. In practical situations, however, if the noise poles approach the unit circle, then the noise covariance matrix tends towards singularity, and therefore, the functional tends towards $-\infty$. When the step size is not too important, this fact avoids stability problems.

G. Estimation of Source Parameters

After runs of the preceding algorithm, an estimation of the AR coefficients is obtained. Let $A^l = (1, \hat{a}_1, \dots, \hat{a}_p)$ be this estimation.

The problem is now to translate this estimation in terms of spatial processing, and more precisely, to obtain source bearing estimates also as their powers at a given frequency. This last point has a great practical importance because power spectral density estimation is also a good means to make a distinction between true sources (positive powers) and false sources (negative powers). Furthermore, the derivation of source powers estimation will be a fundamental tool for the minimization of the whiteness functional (see Section III).

The whitened matrix \hat{R}_w given by

$$\hat{R}_w = L \cdot R \cdot L^*$$

with

$$B^{-1} = L \cdot L^* \quad [B \text{ given by (14)}] \quad (35)$$

is considered for source bearing estimation. Note that \hat{R}_w defined by (35) is not generally a Toeplitz matrix (even if \hat{R} is Toeplitz); it is possible to remedy this problem by use of the whitening filter (48).

Using formula (35), the steering vectors D_θ (corresponding to bearing θ) become whitened vectors, i.e., $L \cdot D_\theta$ by use of classical results [22].

The MUSIC method for bearing estimation consists of calculation of the sine of the steering vector $L \cdot D_\theta$ on the source subspace (i.e., the subspace spanned by the "greatest" eigenvectors of \hat{R}_w), i.e., [23],

$$\begin{aligned} \frac{1}{\sin^2(\theta)} &= \frac{\|L \cdot D_\theta\|^2}{\|(Id - \Pi) \cdot L \cdot D_\theta\|^2} \\ &= \frac{\|L \cdot D_\theta\|^2}{\|(\Pi^\perp) \cdot L \cdot D_\theta\|^2} \end{aligned} \quad (36a)$$

where

$\| \cdot \|$ is the Euclidian norm [6]

Π is the projector matrix, $\Pi = V \cdot V^*$

$V(n_s \times q)$ is the matrix constituted by the greatest

eigenvectors of \hat{R}_w . (36b)

We are now facing the problem of estimation of source powers. The MUSIC method [23] is also a means to estimate source powers (by eigendecomposition of the matrix R_w and identification), but that method seems very sensitive to noise estimation and has given unsatisfying results on our simulations. Another method based upon Pisarenko's method [24] has been considered.

More precisely, let $\{\hat{\theta}_1, \dots, \hat{\theta}_q\}$ be the estimated source bearings. Consider now the dyadic $(n_s \times n_s)$ associated matrices, i.e.,

$$\{D_{\hat{\theta}_i} \cdot D_{\hat{\theta}_i}^*\}_{i=1}^q$$

Then for each of these source matrices, pick the submatrices

$$S_{\hat{\theta}_i, l} = \begin{pmatrix} a_{l,1}^i & a_{l,2}^i & \dots & a_{l,p+1}^i \\ \vdots & \vdots & & \vdots \\ a_{l+p,1}^i & a_{l+p,2}^i & \dots & a_{l+p,p+1}^i \end{pmatrix} \quad (37a)$$

where

$a_{l,k}^i$ is the l th row, k th column element of the matrix

$$D_{\hat{\theta}_i} \cdot D_{\hat{\theta}_i}^* \quad (37b)$$

for $i = 1, 2, \dots, s$ and $l = 1, \dots, n_s - p$.

All the submatrices $S_{\hat{\theta}_i, l}$ are Toeplitz, but non-Hermitian. Then the whitened covariances of sources are given by (41)

$$s_w(\hat{\theta}_i, l) = A^* \cdot S_{\hat{\theta}_i, l} \cdot A \equiv s_w(\hat{\theta}_i, l). \quad (38)$$

Note that the covariances of sources $s_w(\hat{\theta}_i, l)$ correspond to source covariances after whitening by the inverse filter A as detailed in formulas (41a) and (41b).

Covariances of whitened array outputs $\hat{r}_w(l)$ are defined by the same method (38) using a Toeplitz estimate of the covariance matrix of the outputs. Such an estimate can be obtained by averaging along the diagonal [12] and amounts to an orthogonal projection on the Toeplitz subspace [19]. It is now possible to use the Pisarenko's method for source power estimation. Assuming spatial whiteness of the additive noise (after whitening), the estimated powers $\{\hat{q}_1, \dots, \hat{q}_q\}$ of sources are the solution of the following overdetermined linear system.

Covariances of whitened outputs $\hat{r}_w(l)$ are defined by the same method (38); then Pisarenko's method can be applied. The estimated powers $\{\hat{q}_1, \dots, \hat{q}_q\}$ of sources are the solution of the following overdetermined linear

system:

$$\begin{aligned} & \begin{bmatrix} s_w(\hat{\theta}_1, 1) & \dots & s_w(\hat{\theta}_q, 1) \\ \vdots & & \vdots \\ s_w(\hat{\theta}_1, n_s - p) & \dots & s_w(\hat{\theta}_q, n_s - p) \end{bmatrix} \begin{pmatrix} \hat{q}_1 \\ \vdots \\ \hat{q}_q \end{pmatrix} \\ &= \begin{bmatrix} \hat{r}_w(1) \\ \vdots \\ \hat{r}_w(n_s - p) \end{bmatrix}. \end{aligned} \quad (39)$$

Note that the above procedure corresponds to uncorrelated sources (the source matrix is $S = \sum_{i=1}^q q_i \cdot \mathbf{D}_{\theta_i} \cdot \mathbf{D}_{\theta_i}^*$) and is (theoretically) independent of the noise level since it uses covariances $\hat{r}_w(l)$ with $l \geq 1$.

The corresponding results are presented in Section IV. Let us now examine another approach.

III. PLANE WAVE HYPOTHESIS, WHITENESS FUNCTIONAL

A. Introduction

Converse to the preceding approach, special structures of source vectors (plane wave hypothesis) will be considered; thus, the problem will be restricted to the case of a linear array with equispaced sensors. The likelihood approach does not use the plane wave hypothesis or the spatial stationarity; it is maximized only relatively to noise parameters (for the sake of simplicity, it is, however, an advantage!).

Taking into account the plane wave hypothesis leads us to define a functional depending upon both source and noise parameters. Converse to the preceding approach, the source and noise parameters will play symmetric roles.

This method is based on application of Pisarenko's method for harmonic retrieval [24] and, more precisely, on the analysis of its perturbations with respect to noise parameters. That leads us to define a functional involving differences between extradiagonal terms of a whitened matrix and extradiagonal terms of the covariance matrix of sources. These functionals will be minimized by means of an iterative (gradient) algorithm relative to both *noise and source parameters*. The method is mainly devoted to the case of AR noise modeling (simplicity of the whitening filter!). Notations and definitions are identical to those used in Section II.

After a general presentation of the whiteness functional, its properties and a gradient method (for its minimization) will be presented. This last part needs many intermediate steps and is rather technical.

B. The Whiteness Functional

Assume that R is a Toeplitz covariance matrix and that the source parameters (p.s.d. and bearings) are perfectly known and define the functional J by [25]

$$J = \sum_{l=1}^L |r(l) - r_s(l)|^2 \quad (L > 1). \quad (40)$$

Then, in the white noise case, J is null. If the noise is correlated, this conclusion is false, but is it possible to use whitened data?

Consider now the case of AR noise modeling. Let \mathbf{A} be the vector of coefficients of this model [i.e., $\mathbf{A}^t = (1, a_1, \dots, a_p)$] and $[A(z^{-1})]^{-1}$ the associated filter. The whitening filter is thus $A(z^{-1})$ and the covariances of whitened data (by use of this filter) can be written as

$$r_w(l) = E \left\{ \left(\sum_{i=0}^p a_i x_{t-i} \right) \cdot \left(\sum_{i=0}^p \bar{a}_i \bar{x}_{t-l-i} \right) \right\} \quad (a_0 = 1).$$

Therefore,

$$r_w(l) = \mathbf{A}^* \cdot R_l \cdot \mathbf{A} \quad (41a)$$

where R_l is the $(p+1) \times (p+1)$ matrix given by

$$R_l = \begin{bmatrix} r(l) & r(l-1) & \dots & r(l-p) \\ r(l+1) & & & \\ \vdots & & & \\ r(l+p) & \dots & r(l+1) & r(l) \end{bmatrix} \quad (41b)$$

(R_l is Toeplitz but non-Hermitian).

Define now the functional (conditionally to \mathbf{A})

$$J_w(\mathbf{A}) = \sum_{l=1}^L |r_w(l) - r_s(l)|^2 \quad (42)$$

[$r_w(l)$ defined by (41), $r_s(l)$ exact covariance of sources after the mapping \mathbf{A} (41)].

Then, when \mathbf{A} is equal to \mathbf{A}_0 (\mathbf{A}_0 : exact model), one has

$$r_w(l) = r_s(l) + E \left(\frac{A_0(z^{-1})}{A_0(z^{-1})} \cdot e(t) \cdot \frac{A_0(z^{-1})}{A_0(z^{-1})} \cdot e(t+l) \right) = r_s(l) \quad (43)$$

($e(t)$ is the white noise input of the AR model).

Therefore,

$$J_w(\mathbf{A}_0) = 0. \quad (44)$$

For practical applications, one uses the *whiteness functional* defined by

$$J_w(\hat{\mathbf{A}}) = \sum_{l=1}^L |r_w(l) - \hat{r}_s(l)|^2. \quad (45)$$

$\hat{r}_s(l)$ being defined by (41) (using \hat{R}_l available data instead of R_l), $\hat{r}_s(l)$ is estimated by means of Pisarenko's method (or another high-resolution (HR) method [23]) applied to whitened data, i.e.,

$$\hat{r}_s(l) = \sum_{i=1}^n \hat{q}_i \cdot \exp(-j\pi l f_i) \quad j^2 = -1 \quad (46)$$

with

n = number of sources (estimated by HR method)
 \hat{q}_i and \hat{f}_i are, respectively, power and spatial frequency of the i th sources obtained by HR method (see Section II-G) applied to the whitened matrix (for a given vector \mathbf{A}) \hat{R}_w .
 L is usually chosen equal to $n_s - p - 1$.

We are now coping with the minimization of the functional defined by (45) and (46). The minimization is defined in terms of AR coefficients vector \mathbf{A} only, but is relative to the terms $\hat{r}_w(l)$ and $\hat{r}_s(l)$ (\mathbf{A}); this fact implies the estimation of source parameters at each iterate of the gradient algorithm. The general procedure is summarized below.

It 0 (Starting Parameters):

$$\mathbf{A}'_0 = (1, 0, \dots, 0)$$

\hat{q}_i^0 and \hat{f}_i^0 obtained by spatial analysis (HR) of \hat{R}

$$\hat{r}_w(l) = \hat{r}(l)$$

(\hat{R} : Toeplitz matrix, orthogonal projection on the Toeplitz subspace [26]).

It k:

$$\mathbf{A}'_k = (1, a_1^k, \dots, a_p^k)$$

\hat{q}_i^k and \hat{f}_i^k obtained by spatial analysis of \hat{R}_w^k

$$\hat{r}_w^k(l) = \mathbf{A}'_k^* \cdot \hat{R}_l \cdot \mathbf{A}_k$$

(The sources number may be corrected at each iteration.)

It k \rightarrow It k + 1: Calculation of the gradient vector $\mathbf{G}_J(\mathbf{A}_k)$. Then

$$\mathbf{A}_{k+1} = \mathbf{A}_k - \rho_k \cdot \mathbf{G}_J(\mathbf{A}_k)$$

(ρ_k : gradient's step size). (47)

Section III-D will describe calculation of the gradient vector $\mathbf{G}_J(\mathbf{A}_k)$. In fact, the coherence of the above definitions requires invariance (by whitening) properties which will now be presented.

C. Whitening Invariance Properties

Formula (41) defining $r_w(l)$ may be written in matrix form, i.e.,

$$\mathbf{R}_w = \mathbf{A} \cdot \mathbf{R} \cdot \mathbf{A}^*$$

where $\mathbf{R}(n_s \times n_s)$ is the Toeplitz covariance matrix of the outputs, and \mathbf{A} is a rectangular ($L \times n_s$) matrix ($L \leq n_s - p$) defined by

$$\mathbf{A} = \begin{bmatrix} 1 & a_1 & \dots & a_p & 0 & \dots & 0 \\ 0 & 1 & \dots & a_1 & \dots & a_p & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \vdots \\ 0 & \dots & 0 & 1 & a_1 & \dots & a_p \\ 0 & \dots & 0 & 0 & 1 & \dots & a_1 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 0 & 0 & \dots & 1 \end{bmatrix}$$

(48)

The matrix \mathbf{R}_w defined by (48) is a Toeplitz positive matrix (\mathbf{R} positive) and $\mathbf{R}_w(1, l) = r_w(l)$.

This matricial translation of (41) allows us to study the effect of whitening on the covariance matrix of a source.

We shall consider now the effect of whitening [formula (48)] on a plane wave whose associated steering vector [1] is denoted \mathbf{D}_θ :

$$\mathbf{D}_\theta^l = (1, \exp(ja), \dots, \exp(j(n_s - 1)a),$$

$$j^2 = -1.$$

Then by use of (48), the covariance matrix of the source after whitening is given by

$$\mathbf{R}_w(\theta) = \mathbf{A} \cdot \mathbf{D}_\theta \cdot \mathbf{D}_\theta^* \cdot \mathbf{A}^* \quad (49)$$

$\mathbf{R}_w(\theta)$ is therefore a rank 1 Toeplitz matrix, and furthermore,

$$\mathbf{A} \cdot \mathbf{D}_\theta = \begin{bmatrix} 1 + a_1 e^{j\alpha} + \dots + a_p e^{j(p-1)\alpha} \\ e^{j\alpha} + a_1 e^{2j\alpha} + \dots + a_p e^{jp\alpha} \\ \vdots \\ e^{j(L-1)\alpha} + \dots + a_p e^{j(L+p-1)\alpha} \end{bmatrix}$$

or

$$\mathbf{A} \cdot \mathbf{D}_\theta = (1 + a_1 e^{j\alpha} + \dots + a_p e^{j(p-1)\alpha}) \cdot \mathbf{D}'_\theta, \quad (50)$$

\mathbf{D}'_θ being the ($L \times 1$) vector obtained by selection of the L first components of the vector \mathbf{D}_θ itself. Finally, the covariance matrix of a source (after whitening) may be written as

$$\mathbf{R}_w(\theta) = q(\theta) \cdot \mathbf{D}'_\theta \cdot \mathbf{D}'_\theta^* \quad (51)$$

with

$$q(\theta) = |1 + a_1 e^{j\alpha(\theta)} + \dots + a_p e^{j(p-1)\alpha(\theta)}|^2.$$

These results lead to Property 3.

Property 3: The whitening procedure defined by (48) enjoys the following properties.

a) Let \mathbf{R} be a Toeplitz covariance matrix; then $\mathbf{R}_w = \mathbf{A} \cdot \mathbf{R} \cdot \mathbf{A}^*$ is a Toeplitz matrix.

b) The procedure transforms a steering vector into another steering vector of (dimension L) corresponding to the same bearing.

These properties are, in fact, fundamental because they justify the whiteness functional definition; furthermore, they also constitute a justification of the method of calculation of the gradient defined in Section III-D.

In practical situations, exact source bearings are unknown; Property 3 states only that these bearings are invariant under the whitening procedure. In order to estimate these bearings, a high-resolution method is used. Obviously, in the general case, this method does not give the exact source bearings. But, using the whiteness functional, we hope to define an iterative method minimizing the differences between extradiagonal terms of the whitened matrix and the corresponding terms of the source matrix.

D. Calculation of the Gradient Vector (of Whiteness Functional)

The index k (of iterate number) are omitted in that subsection. The gradient vector is deduced from (45) and (48), i.e.,

$$G_J(\mathbf{A}) = 2 \operatorname{Re} \left[\sum_{l=1}^L (\mathbf{R}_l \mathbf{A} - \mathbf{G}_{r_s}(l)) \cdot (\mathbf{A}^* \cdot \mathbf{R}_l^* \cdot \mathbf{A} - \bar{r}_s(l)) \right]. \quad (52)$$

$\mathbf{G}_{r_s}(l)$ is the gradient vector of $r_s(l, \mathbf{A})$ relative to \mathbf{A} . It is the delicate part of the algorithm. It is defined by

$$\mathbf{G}_{r_s}(l) = \sum_{i=1}^n \mathbf{G} \hat{q}_i(\mathbf{A}) \cdot \exp(-j\pi l f_i) - j\pi l \cdot \sum_{i=1}^n \hat{q}_i \exp(-j\pi l \hat{f}_i) \cdot \mathbf{G} \hat{f}_i(\mathbf{A}). \quad (53)$$

The calculation of $\mathbf{G}_{r_s}(l)$ amounts to calculating the gradient vectors $\mathbf{G} \hat{f}_i(\mathbf{A})$ and $\mathbf{G} \hat{q}_i(\mathbf{A})$; this will be achieved in some steps. The first step consists of the calculation of partial derivatives of eigenvectors of \mathbf{R}_w . Derivatives of source parameters are then calculated in a second step by means of analysis of perturbations of a high-resolution method.

Step 1—Partial Derivatives of Eigenvectors of \mathbf{R}_w : Denote by $\hat{\mathbf{R}}_A$ the $(L+1) \times (L+1)$ Toeplitz matrix defined by

$$\hat{R}_A(i, j) = \hat{r}_A(i - j)$$

with

$$\hat{r}_A(l) = \mathbf{A}^* \cdot \mathbf{R}_l \cdot \mathbf{A}$$

(the matrix $\hat{\mathbf{R}}_A$ should be denoted $\hat{\mathbf{R}}_{w, A}$) and

$$\{ \mathbf{U}_1, \dots, \mathbf{U}_{L+1}; \lambda_1 \geq \dots \geq \lambda_{L+1} \} \quad (54)$$

an eigensystem of $\hat{\mathbf{R}}_A$.

Then a classical result of linear algebra [9] gives the partial derivatives of eigenvectors \mathbf{U}_k , i.e.,

$$\frac{\partial}{\partial a_i} \mathbf{U}_k = \sum_{\substack{j=1 \\ j \neq k}}^{L+1} \left\{ \frac{\mathbf{U}_j^* \cdot (\partial \hat{\mathbf{R}}_A / \partial a_i) \cdot \mathbf{U}_k}{\lambda_k - \lambda_j} \right\} \cdot \mathbf{U}_j \quad (55)$$

$(n+1 \leq k \leq L+1)$.

(Formula (55) is valid only for simple eigenvalues.)

The matrix $\partial \hat{\mathbf{R}}_A / \partial a_i$ is a Toeplitz matrix which may be defined by its first row as

$$\frac{\partial}{\partial a_i} \hat{R}_A(m, 1) = \mathbf{E}_i^* \cdot \mathbf{R}_{m-1} \cdot \mathbf{A} + \mathbf{A}^* \cdot \mathbf{R}_{m-1} \cdot \mathbf{E}_i$$

with

$$\mathbf{E}'_i(l) = \begin{cases} 1 & \text{if } l = i \\ 0 & \text{else.} \end{cases} \quad (56)$$

Step 2—Derivatives of Source Bearings (with Respect to the $\{a_i\}$): The i th source is characterized by its d.s.p.

\hat{q}_i and its estimated spatial frequency \hat{f}_i ; the problem consists of calculation of their partial derivatives with respect to the parameters $\{a_i\}$.

In order to solve it, recall that estimated source bearings (e.g., for the MUSIC method) minimize the projection of vector \mathbf{D}'_θ on the noise subspace.

Let Π be the projector on the noise subspace:

$$\Pi = \mathbf{U} \cdot \mathbf{U}^*$$

with

$$\mathbf{U} = (\mathbf{U}_{n+1}, \dots, \mathbf{U}_L) \quad (\text{“lowest” eigenvectors of } \mathbf{R}_w).$$

Then the projection of \mathbf{D}'_θ on the noise subspace is

$$\Pi(\theta) = \mathbf{D}'_\theta{}^* \cdot \Pi \cdot \mathbf{D}'_\theta. \quad (57)$$

If we consider only one eigenvector (the “lowest,” Pisarenko [24]), one obtains then, denoting

$$\hat{\mathbf{Z}}_i^l = (1, \hat{z}_i, \dots, \hat{z}_i^L)$$

$$\text{with } \hat{z}_i = \exp(-j\pi \hat{f}_i), j^2 = -1$$

and

$$\hat{\mathbf{Z}}_i^l = (0, z_i, \dots, L \cdot z_i^{L-1})$$

(the derivative vector of $\hat{\mathbf{Z}}_i$ relative to \hat{f}_i , $\hat{\mathbf{Z}}_i \equiv \mathbf{D}'_{\hat{\theta}_i}$) an expression of partial derivatives of the spatial frequencies $w \cdot r$ AR coefficients, i.e.,

$$\begin{aligned} & \frac{\partial}{\partial a_i} \hat{\mathbf{Z}}_i^* \cdot \Pi \cdot \hat{\mathbf{Z}}_i + \hat{\mathbf{Z}}_i^* \cdot \Pi \cdot \frac{\partial}{\partial a_i} \hat{\mathbf{Z}}_i + \hat{\mathbf{Z}}_i^* \\ & \cdot \frac{\partial}{\partial a_i} \Pi \cdot \hat{\mathbf{Z}}_i = 0 \end{aligned}$$

or, by factorization,

$$\frac{\partial}{\partial a_i} \hat{f}_i = -\frac{1}{\pi} \cdot \frac{\operatorname{Re} \left\{ \left[\left(\frac{\partial}{\partial a_i} \mathbf{U} \right)^* \cdot \hat{\mathbf{Z}}_i \right] \cdot [\hat{\mathbf{Z}}_i^* \cdot \mathbf{U}] \right\}}{\operatorname{Im} \left\{ (\mathbf{U}^* \cdot \hat{\mathbf{Z}}_i) \cdot (\hat{\mathbf{Z}}_i^* \cdot \mathbf{U}) \right\}}. \quad (58)$$

The partial derivatives of eigenvectors being calculated by (55), the spatial frequencies \hat{f}_i being estimated by an HR method, formula (58) allows us to calculate the partial derivatives $(\partial / \partial a_i) \hat{f}_i$. Formula (58) is obtained by differentiation of $\Pi(\theta)$ as given by (57). If we consider the projection on the whole noise subspace, then the derivative of the projector Π is given by

$$\begin{aligned} \frac{\partial}{\partial a_i} \Pi(\theta) &= \left(\frac{\partial}{\partial a_i} \mathbf{D}'_\theta \right)^* \cdot \Pi \cdot \mathbf{D}'_\theta + \mathbf{D}'_\theta{}^* \left(\frac{\partial}{\partial a_i} \Pi \right) \\ &\cdot \mathbf{D}'_\theta + \mathbf{D}'_\theta{}^* \cdot \Pi \cdot \left(\frac{\partial}{\partial a_i} \mathbf{D}'_\theta \right). \quad (59) \end{aligned}$$

The difficulty is now related to the calculation of partial derivatives of the projector Π (taking into account orthogonality of eigenvectors). Following [27], a second-

order expansion of Π is given by

$$\Pi_\delta = (U_0 + \delta U + \delta^2 U) \cdot (Id - \|\delta U\|^2) \cdot (U_0^* + \delta U^* + \delta^2 U^*)$$

with

$$\begin{aligned} U_i &\equiv (U_i)_0 + (\delta U_i) + (\delta^2 U_i) \\ (U_0 + \delta U + \delta^2 U)^* \cdot (U_0 + \delta U + \delta^2 U) \\ &= Id + \|\delta U\|^2 \end{aligned} \quad (60)$$

from which a first-order expansion of Π_δ is obtained:

$$\Pi_\delta = \Pi + U \cdot \delta U^* + \delta U \cdot U^*. \quad (61)$$

Hence,

$$\frac{\partial}{\partial a_l} \Pi = U \cdot \left(\frac{\partial}{\partial a_l} U \right)^* + \left(\frac{\partial}{\partial a_l} U \right) \cdot U^*$$

yielding $(\partial/\partial a_l) \hat{f}_i$; let

$$\begin{aligned} \frac{\partial}{\partial a_l} \hat{f}_i &= \frac{-1}{(L-n)\pi} \cdot \sum_{m=n+1}^l \\ &\cdot \frac{\operatorname{Re} \left\{ \left[\left(\frac{\partial}{\partial a_l} \cdot U_m \right)^* \cdot \hat{Z}_i \right] \cdot [\hat{Z}_i^* \cdot U_m] \right\}}{\operatorname{Im} \left\{ (U_m^* \cdot \hat{Z}_i) \cdot (\hat{Z}_i^* \cdot U_m) \right\}}. \end{aligned} \quad (62)$$

It is obviously possible to use a second-order expansion of Π_δ , but for our application, the first derivative gives satisfying results.

Step 3—Derivatives of Source Powers: Let F be the $(L \times q)$ matrix constituted by elementary theoretical covariances of estimated sources, i.e.,

$$F = \begin{pmatrix} s(\hat{\theta}_1, 1) & \dots & s(\hat{\theta}_q, 1) \\ \vdots & & \vdots \\ s(\hat{\theta}_1, L) & \dots & s(\hat{\theta}_q, L) \end{pmatrix}$$

with

$$s(\hat{\theta}_k, l) = -\sin(\pi l \hat{f}_k).$$

Furthermore, let \mathbf{J} be the $(L \times q)$ vector constituted by imaginary parts of whitened output covariances, i.e.,

$$\mathbf{J}' = \operatorname{Im}(\hat{r}_A(1), \dots, \hat{r}_A(L)) r_A(k) \text{ defined by (48)}$$

and

$$\mathbf{Q}(\mathbf{Q}' = (q_1, \dots, q_q) \text{ be the vector of source powers.} \quad (63)$$

Then this vector \mathbf{Q} would be the solution of the following linear system [24] (in the white noise case):

$$F \cdot \mathbf{Q} = \mathbf{J}. \quad (64)$$

The gradient vector $\mathbf{G}_l(\mathbf{G}_l' = ((\partial/\partial a_l)q_1, \dots, (\partial/\partial a_l)q_s))$ is deduced from (64) by differentiation, i.e.,

$$\mathbf{G}_l = - \left(F^{-1} \cdot \frac{\partial}{\partial a_l} F \cdot F^{-1} \right) \cdot \mathbf{J} + F^{-1} \cdot \left(\frac{\partial}{\partial a_l} \mathbf{J} \right)$$

where

$$\frac{\partial}{\partial a_l} F \text{ is deduced from (62), (55)}$$

$$\frac{\partial}{\partial a_l} \mathbf{J} \text{ is given by (56)}$$

$$F^{-1} = \text{pseudoinverse matrix of } F. \quad (65)$$

The vectors $\mathbf{G}\hat{q}_i(\mathbf{A})$ are deduced from vectors $\mathbf{G}_l(l = 1, 2, \dots, p)$.

Calculation of vectors $\mathbf{G}\hat{f}_i(\mathbf{A})$ and $\mathbf{G}\hat{q}_i(\mathbf{A})$ is now achieved; it remains to calculate a suited step size for the gradient method; this is achieved by means of a first- or second-order expansion of the functional.

Step 4 (Source Parameters Updating): After calculation of gradient vector $\mathbf{G}_j(\mathbf{A}_k)$ and estimation of a step size, it is necessary to use the new estimation of noise parameters in order to update the source parameters (by means of the HR method). It may be very convenient to eliminate spurious peaks (having a negative estimated p.s.d.).

E. Practical Utilization

The method starts with $\mathbf{A}'_0 = (1, 0, \dots, 0)$; L may be chosen equal to $n_s - 1 - p$. The parameters (p, q) are generally overdetermined. They can be adjusted by consideration of the whiteness functional (see Section II-F).

IV. SIMULATION RESULTS

These two methods have been tested on simulated signals. In the case of simulated signals, the degradations of high-resolution methods by noise correlations may be easily explained and performances easily quantified.

A. Simulation Method

The general aims of the simulation results that will be presented are the following:

- 1) accuracy of noise correlations estimates
- 2) improvements of high-resolution methods (using these estimates)
- 3) statistical behavior of these methods
- 4) explanation of results
- 5) robustness of the method.

For a given noise matrix \mathbf{B} and q uncorrelated sources (bearings θ_i , psd: σ_i^2), the random vectors \mathbf{X} are simulated as follows:

$$\mathbf{X} = \sum_{i=1}^q a_i \mathbf{D}_{\theta_i} + \mathbf{T} \cdot \mathbf{B}$$

where

$$\left\{ \begin{array}{l} \mathbf{D}_{\theta_i} \text{ is a steering vector [2] associated to bearing } \theta_i \\ a_i: \text{ complex Gaussian circular variable } N(0, \sigma_i^2) \\ \mathbf{B}: \text{ complex Gaussian vector } N(\mathbf{O}, Id) \\ \mathbf{B}_0 = \mathbf{T} \cdot \mathbf{T}^* \end{array} \right. \quad (66)$$

Then a covariance matrix $\hat{\mathbf{R}}$ is estimated by means of a periodogram method [31] (the parameter BT corresponds to the number of snapshots used for estimation of $\hat{\mathbf{R}}$). After runs of algorithms, an estimation of \mathbf{B} is obtained (named $\hat{\mathbf{B}}$) corresponding to estimates of noise parameters ($\hat{a}_1, \dots, \hat{a}_p$).

Several procedures are used in order to study the accuracy of noise parameter estimates.

1) *Computation of $\cos(\hat{\mathbf{B}}, \mathbf{B}_0)$:*

$$\cos(\hat{\mathbf{B}}, \mathbf{B}_0) = \frac{\text{tr}(\hat{\mathbf{B}} \cdot \mathbf{B}_0)}{\{\text{tr}(\mathbf{B}_0 \cdot \mathbf{B}_0) \text{tr}(\hat{\mathbf{B}} \cdot \hat{\mathbf{B}})\}^{1/2}} \quad (67)$$

A cosine near 1 means that \mathbf{B} and \mathbf{B}_0 are ‘‘almost’’ colinear, which is the aim of the method.

2) *Spatial Density of Noise:* We associate to an ARMA noise model its spatial density given by [12]

$$S(z) = \frac{\mathbf{B}(z) \cdot \mathbf{B}^*(z^{-1})}{\mathbf{A}(z) \cdot \mathbf{A}^*(z^{-1})}$$

with

$$\mathbf{B}(z) = b_0 + b_1 \cdot z + \dots + b_p \cdot z^p$$

$$\mathbf{A}(z) = a_0 + b_1 \cdot z + \dots + a_q \cdot z^q$$

and

$$z = \exp(-2j\pi d/\lambda \cos \theta), \quad j^2 = -1 \quad (68)$$

(λ : wavelength, θ : direction of arrival, d : intersensors distance usually for the simulations $d = \lambda/2$).

It is then possible to compare the exact spatial density of noise to its estimate. Furthermore, it is a good means for explanation of degradations of HR methods (spurious peaks, nondetection of weak sources, etc.). The value of FC_w corresponds to the value of the initial likelihood functional (i.e., white noise assumption), whereas FC_e corresponds to the final value of this functional.

3) *HR Methods Behind Whitening:* The whitening is defined by (35) or (48). We consider also the results for the psd (of source) estimation (0° corresponds to the array axis, i.e., endfire direction).

B. Simulation Results

We shall now present the results obtained by maximization of the likelihood functional (Section II) for noise correlation estimation. The array is constituted by equally spaced sensors (on a line), the bearing 0° corresponds to the array axis (idem for real data), the parameter BT corresponds to the number of independent snapshots.

The first result (Fig. 2) corresponds to an AR modeling;

the simulation data are given in the caption of Fig. 2. Note that the dimension of sources subspace is assumed to be 5 (i.e., $q = 5$), whereas the results are quite satisfying.

The broken line corresponds to results of the MUSIC method without whitening ($\hat{\mathbf{R}}$ is a Toeplitz matrix), whereas the solid line corresponds to results after whitening. In this case, the two sources at bearings 35° and 45° are not separated, whereas the weak source is not detected by the MUSIC method (without whitening). Conversely, the whitening obtained by noise coefficient estimates yields the resolution of the two strong sources and the detection of the weak source. These good results may be easily explained by the value of $\cos(\hat{\mathbf{B}}, \mathbf{B}_0)$ ($=0.998$).

Fig. 3 represents the noise spatial density associated to the model. Obviously, its peak (at 90°) induces the spurious peaks (at the vicinity of 90°) and the nondetection of the weak source (110°) for the unwhitened data. The broken line (exact noise density) and the solid line (estimated noise density) are perfectly merged.

Fig. 4 presents the noise correlation shapes along the array; they decrease slowly as a function of the intersensor's distance. The simulated noise can be considered as highly correlated.

Fig. 5 illustrates the statistical stability of the method for ten trials of the same simulation. The mean of $\cos(\hat{\mathbf{B}}, \mathbf{B}_0)$ is just near one; the results are quite satisfying.

The second result (Fig. 6) corresponds to an MA noise model. The dimension of source subspace is assumed to be five. Then the whitening removes the spurious peaks and allows us to detect the weak source (bearing 110°).

Fig. 7 presents the exact spatial noise density (broken line); it explains the poor results of the MUSIC method without whitening (Fig. 6). The estimated spatial noise density (solid line) is quite acceptable, whereas Fig. 8 illustrates the statistical stability of the method.

Fig. 9 illustrates two important points. The exact AR noise model is an order two model (1, -0.4, 0.8); now we assume (in maximization of the likelihood functional) an order three AR model. The results are unaffected by this overdetermination of the noise model. Furthermore, the vertical lines represent the estimated source powers, the dashed lines correspond to results before whitening, whereas solid lines correspond to results after whitening [as defined by (38) and (39)]. One can see that estimated source psd's are greater for spurious peaks before whitening; on the other hand, after whitening, the estimated powers of sources correspond to simulation data. Note that the estimated powers (after whitening) corresponding to small local maxima are negative.

Fig. 10 illustrates the (direct) extension of the likelihood method to the complex AR noise model. In this case, the spatial noise density is nonsymmetric relative to the array's broadside. As previously, the broken line corresponds to exact noise density, whereas the solid line corresponds to its estimate. The results are quite satisfying.

For all the simulations, the number of iterations of the

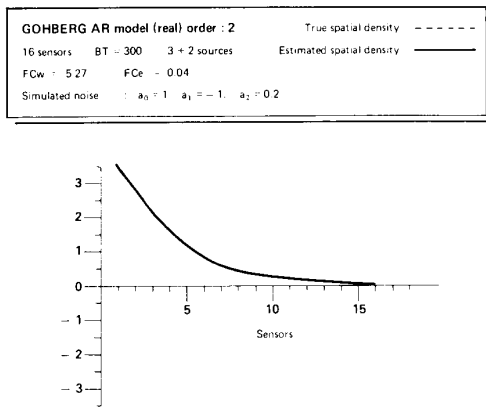


Fig. 3. Noise correlations of the noise model of Fig. 2.

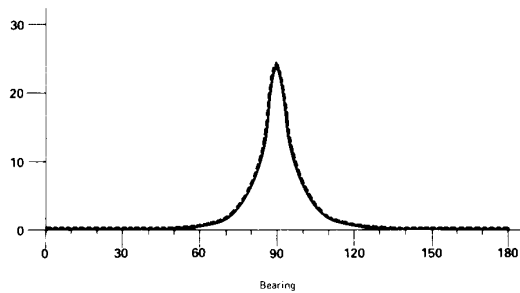


Fig. 4. Comparison between the exact spatial noise density and its estimate.

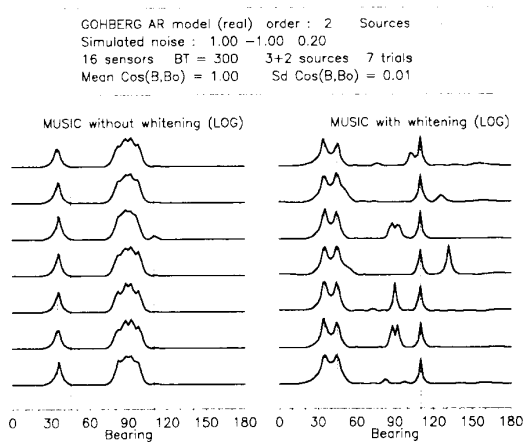


Fig. 5. Maximum likelihood method. Simulation parameters of Fig. 2. Statistical behavior of the method.

gradient method until convergence is between 10 and 20 (with a "first-order" determination of the step size).

The third type of result concerns the whiteness functional (Section III). Three sources are simulated; the simulated noise corresponds to a physical model (surface noise model [28]). The whitening obtained by means of estimated AR coefficients permits us to separate the two

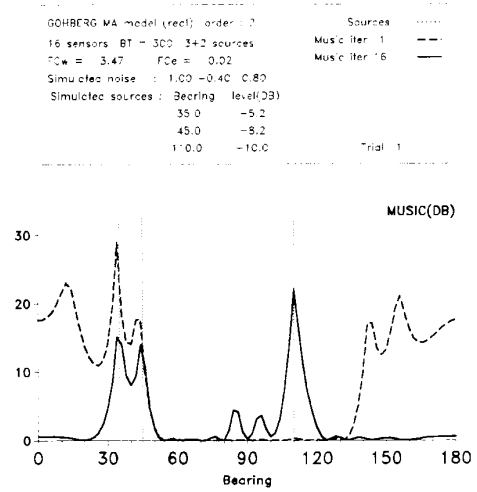


Fig. 6. Maximum likelihood method, MA noise model. Comparison between MUSIC method before whitening and after.

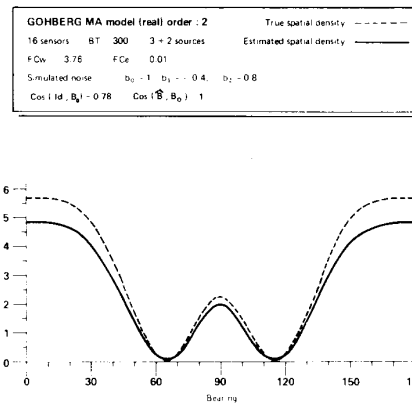


Fig. 7. Comparison between the exact noise density and its estimate.

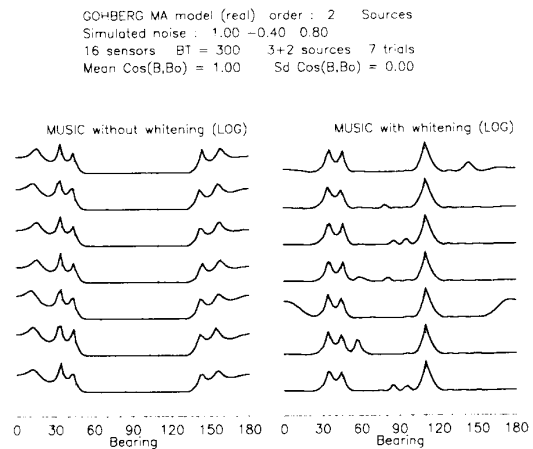


Fig. 8. Statistical behavior of the likelihood method.

```

GOHBERG AR model (real) order : 2
Simulated noise : 1.00 -0.40 0.80
16 sensors BT = 300 3+2 sources
FCw = 3.39 FCe = 0.03
    
```

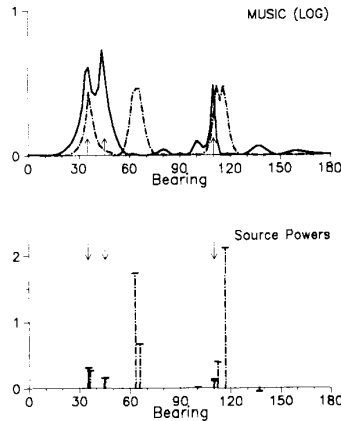


Fig. 9. Overdetermination of the noise model.

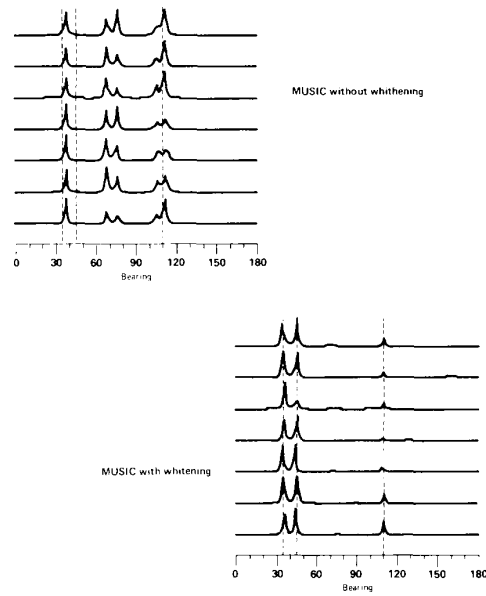


Fig. 11. Whiteness functional method. Statistical behavior of the method.

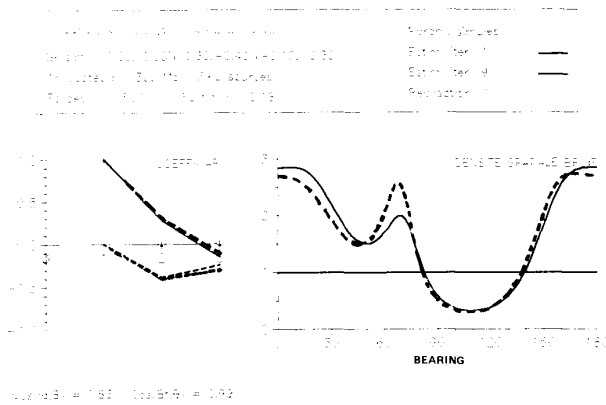


Fig. 10. Estimation of a complex AR noise model.

```

BARMAS
BT 300 3+2 sources
AR real coefficients: a0 = 1, a1 = -0.5, a2 = 0.3
7 trials
Mean Cos (B, B0) : 1 Sd Cos (B, B0) : 0.00
    
```

conclusion is valid for all the physical noise models [29], [30]. An AR(MA) noise model seems to be a good means for modeling any true noise [19].

D. Results for Real Data

We consider now the outputs of a linear array. This linear array is constituted by equispaced sensors (at half a wavelength for the upper frequency of the band). In order to obtain a noise model, we proceed as follows.

- 1) Compute the Fourier transforms of the outputs:

$$X_t \xrightarrow{FT} X(f).$$

- 2) The cross-spectral matrices of the outputs are estimated by means of the averaged periodogram method [31], i.e.,

$$\hat{R}(f) = \frac{1}{N} \cdot \sum_{i=1}^N X_i(f) \cdot X_i^*(f), \quad N = 300.$$

- 3) The source number is estimated (by use of \hat{R}). Generally, it is overdetermined.

- 4) Runs of the algorithm (likelihood functional), starting with

$$A_0^t = (1, 0, \dots, 0).$$

Estimation of the model order p .

A noise model $A^t = (\hat{a}_1, \dots, \hat{a}_p)$ is obtained for each matrix R (AR (3) for Fig. 12).

strong sources (at bearings 35° and 45°) and to eliminate the spurious peaks (Fig. 11). The statistical stability of the method is illustrated by Fig. 11.

C. Conclusions

The simulation results have shown the good performance of the two methods. Overall, the methods appear to be robust with respect to source number and noise model overdetermination. These two points have a fundamental practical importance.

The statistical behavior of these methods remains to be theoretically investigated in terms of spatial processing (bias in bearing estimation, angular resolution, etc.), but it appears to be a very difficult task in the general case.

The practical interest of these methods relies upon the ability of an AR(MA) noise model to describe physical noise by a high-order AR(MA) model; but a low-order model seems to be sufficient for most of the cases. This

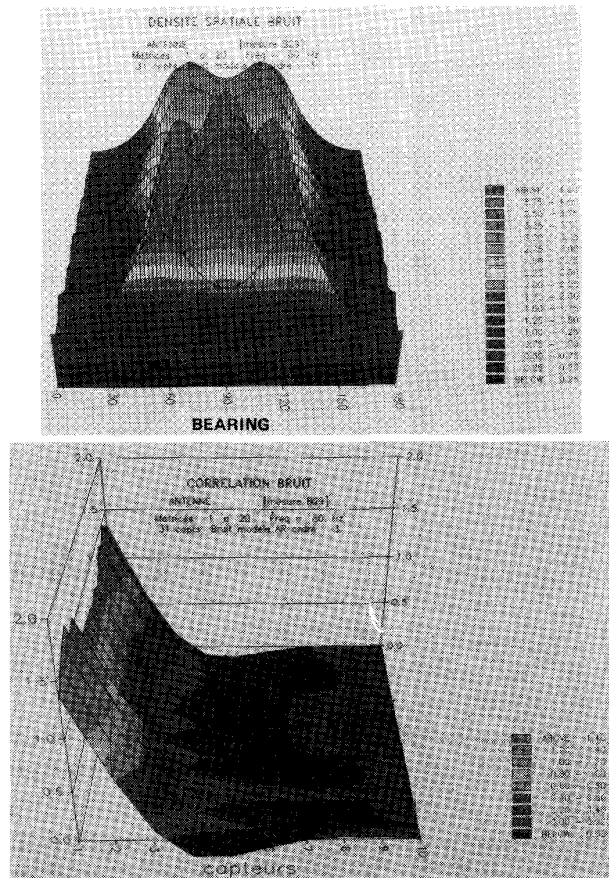


Fig. 12. Real data, linear towed array. Spatial noise density and associated noise correlations. Maximum likelihood methods.

5) The noise correlation as well as its spatial density are deduced from \mathbf{A} by (14).

The results are presented as follows.

1) *Spatial Noise Density*: Bearings are plotted on the x axis, spatial densities on the y axis, whereas the consecutive estimations are presented in depth (0° corresponds to the array axis).

2) *Spatial Noise Correlations*: The noise correlations are deduced from \mathbf{A} by (14) (they constitute the first row of \mathbf{B} which is Toeplitz). The values of correlations are plotted on the y axis (the number 1 corresponds to auto-correlation, 2 to correlation between sensors spaced by the elementary intersensor spacing). The consecutive results are presented in depth.

Fig. 12 presents the results obtained for the lowest frequency of the spectral band; the intersensors spacing correspond to $\lambda/4$ for that frequency. The coefficients of the AR noise model have been chosen real (the imaginary part seems to be negligible). The algorithm converges fast (at most ten iterations); the lowest eigenvalues of the whitened matrix R_w are well clustered.

The noise appears to be strongly spatially colored. However, the correlations decrease fast (with respect to sensor spacing). The statistical stability of these estimates is rather acceptable. Obviously, the noise model estima-

tion allows us to improve the results of spatial processing. Improvements may be significant if the noise is highly spatially colored. Generally, the noise correlations decrease when the frequency increases (for a given sensor spacing); these results are detailed in [19].

V. CONCLUSION

Two methods for estimation of noise correlations along an array of sensors have been presented. Both rely upon a parametric (ARMA) noise modeling. This modeling has the great advantage of describing the noise correlations by a small number of parameters and providing a satisfying modeling of a great variety of physical noises.

The likelihood method presents the great advantage of needing maximization with respect to noise parameters only. Results are quite satisfying, even with high and long correlations of the noise. An efficient numerical method for maximization has been presented; the convergence of the algorithm has been proven. Furthermore, the method appears to be robust with respect to the choice of parameters.

The whiteness functional method uses fundamentally the "plane wave" hypothesis. An efficient method for minimization of it has been developed; however, it is a little more complex than the previous algorithm.

Both methods are quite feasible and can be used in a great variety of situations. By providing a suitable algorithm, this paper has demonstrated the concrete possibility of separating the sensor outputs of a stationary random field in two parts (a noise part and a source part) by use of the classical hypothesis about the propagation of waves (plane wave or coherence properties).

APPENDIX A

We are now coping with the Gohberg formula. A great variety of presentations of this fundamental formula is available; among these are statistical methods [2], [32], a consequence of the Trench formula [33], and of the Darboux formula [3].

For our sake, we prefer to present a "statistical" proof. In fact, Gohberg's formula relies upon a remarkable matricial property. Using the notations of Johnson [2], an AR model is considered:

$$x_m + a_1 \cdot x_{m-1} + \dots + a_{M-1} \cdot x_{m-M+1} = w_m \quad (\text{A1})$$

(w_m : input white noise, variance 1).

Define $\mathbf{X}'_M = [x_1, \dots, x_M]$, \mathbf{R}_M , the covariance matrix of \mathbf{X}_M , and $p(\mathbf{X}_M)$, the density of the Gaussian vector \mathbf{X}_M ; then

$$p(\mathbf{X}_M) = p(\mathbf{X}_{M-1}, x_M).$$

Now

$$\begin{aligned} \mathbf{X}_M &= \begin{pmatrix} \mathbf{X}_{M-1} \\ x_M \end{pmatrix} \\ &= \left(\begin{array}{c|c} \text{Id}(M-1, M-1) & \mathbf{O} \\ \hline a_1 \cdots a_{M-1} & -1 \end{array} \right) \cdot \begin{pmatrix} \mathbf{X}_{M-1} \\ w_M \end{pmatrix} \end{aligned}$$

whence

$$R_M = \begin{pmatrix} Id & \mathbf{O} \\ \mathbf{A}' & -1 \end{pmatrix} \cdot \begin{pmatrix} R_{M-1} & \mathbf{O} \\ \mathbf{O}' & 1 \end{pmatrix} \cdot \begin{pmatrix} Id & \mathbf{A} \\ \mathbf{O}' & -1 \end{pmatrix}. \quad (\text{A2})$$

But the remarkable fact is the following (well known for automaticians):

$$\begin{pmatrix} Id & \mathbf{O} \\ \mathbf{A}' & -1 \end{pmatrix} \cdot \begin{pmatrix} Id & \mathbf{O} \\ \mathbf{A}' & -1 \end{pmatrix} = Id(M, M). \quad (\text{A3})$$

The Gohberg formula relies upon this property; indeed [by (A2) and (A3)], one has

$$R_M^{-1} = \begin{pmatrix} Id & \mathbf{A} \\ \mathbf{O}' & -1 \end{pmatrix} \cdot \begin{pmatrix} R_{M-1}^{-1} & \mathbf{O} \\ \mathbf{O}' & 1 \end{pmatrix} \cdot \begin{pmatrix} Id & \mathbf{O} \\ \mathbf{A}' & -1 \end{pmatrix}$$

whence

$$\begin{aligned} X_M^* \cdot R_M^{-1} \cdot X_M &= (X_{M-1}^* | X_M^* \cdot \mathbf{A}') \\ &\cdot \begin{pmatrix} R_{M-1} & \mathbf{O} \\ \mathbf{O}' & 1 \end{pmatrix} \cdot \begin{pmatrix} X_{M-1} \\ \mathbf{A}'^* \cdot X_M \end{pmatrix} \end{aligned}$$

with

$$\mathbf{A}' = (a_1, \dots, a_{M-1}, -1). \quad (\text{A4})$$

Now, the quadratic form $X_M^* \cdot R_M^{-1} \cdot X_M$ may be written as a sum of two quadratic forms, i.e.,

$$\begin{aligned} X_M^* \cdot R_M^{-1} \cdot X_M &= X_M^* \cdot \begin{pmatrix} R_{M-1}^{-1} & \mathbf{O} \\ \mathbf{O}' & 0 \end{pmatrix} \\ &\cdot X_M + X_M^* \cdot (\mathbf{A}' \cdot \mathbf{A}'^*) X_M \end{aligned}$$

and therefore

$$R_M^{-1} = \begin{pmatrix} R_{M-1}^{-1} & \mathbf{O} \\ \mathbf{O}' & +0 \end{pmatrix} + \mathbf{A}' \cdot \mathbf{A}'^*. \quad (\text{A5})$$

But R_M being Toeplitz, R_M^{-1} is centrosymmetric [3], and therefore

$$R_M^{-1}(i, j) = R_M^{-1}(M - j, M - i). \quad (\text{A6})$$

By use of (A5), the last column of R_M^{-1} is the last column of $\mathbf{A}' \cdot \mathbf{A}'^*$; that column is also [by (A6)] the first row of R_M^{-1} . By induction, it is now possible to deduce an explicit expression of R_M^{-1} (with respect to AR parameters) for which Gohberg's formula (14) is the matricial translation.

Using (A5), the Gohberg formula may be extended to complex AR model, yielding

$$B^{-1} = \frac{1}{\sigma^2} (A_1 \cdot A_1^* - A_3 \cdot A_3^*). \quad (\text{A7})$$

The rest of the algorithm is identical to the real case.

APPENDIX B

This Appendix deals with the convergence of gradient methods for maximization of the log-likelihood functional. This functional (13) depends upon eigenvalues of

\hat{R}_w , but it is not concave with respect to AR parameters (see Fig. 1). However, its gradient vector is null if and only if noise is perfectly estimated, as will be shown now. A direct analysis based on the analysis of eigenvalues seems very difficult or, at best, very cumbersome. Therefore, a direct approach is preferred.

Consider the log-likelihood functional (3):

$$L(q, \mathbf{A}) = -\log \det (R_{A,q}) - \text{tr} [(R_{A,q})^{-1} \cdot \hat{R}], \quad (\text{B1})$$

$R_{A,q}$ being the exact covariance matrix corresponding to q sources and an AR noise defined by $\mathbf{A}' = (a_1, \dots, a_p)$.

Consider the simplest case ($q = 0$); then

$$R_{A,q} = B \quad \text{and} \quad \frac{\partial L}{\partial a_i} = \text{tr} [(B - \hat{R}) \cdot \Delta_i] \quad (\text{B2})$$

(using the well-known theorem of linear algebra [6]; for a set of differentiable matrices $A(a)$, $(\partial/\partial a) \log \det A(a) = \text{tr} (A^{-1}(a) \cdot (\partial/\partial a) A)$).

Now

$$\begin{aligned} \text{tr} [(B - \hat{R}) \cdot \Delta_i] &= \text{tr} [B(\hat{R}^{-1} - B^{-1}) \cdot \hat{R} \cdot \Delta_i] \\ &= \text{tr} [(\hat{R}^{-1} - B^{-1}) \cdot \hat{R} \cdot \Delta_i B] \end{aligned} \quad (\text{B3})$$

[Δ_i being defined by (25)].

Thus, nullity of the gradient vector will induce the following equality:

$$\text{tr} \left[(\hat{R}^{-1} - B^{-1}) \cdot \hat{R} \left(\sum_{i=1}^p a_i \Delta_i \right) \cdot B \right] = 0$$

itself equivalent to

$$\begin{aligned} \text{tr} [(\hat{R}^{-1} - B^{-1}) \cdot \hat{R} B^{-1} \cdot B] \\ = \text{tr} (Id - B^{-1} \cdot \hat{R}) = 0 \end{aligned} \quad (\text{B4})$$

[$B^{-1} = \sum a_i \Delta_i$ from (25)].

Therefore, necessarily, at a point of nullity of gradient, the initial problem (max likelihood) amounts to the following problem:

$$P_2 \begin{cases} \max \log \det (B^{-1}) \\ \text{under the constraint} \\ \text{tr} (B^{-1} \cdot \hat{R}) = n_s. \end{cases} \quad (\text{B5})$$

Now, the inequality of arithmetic-geometric means [14] proves that all the eigenvalues of $B^{-1} \cdot \hat{R}$ are identical at a point solution of P_2 . Using Property 1, the nullity of gradient vector is equivalent to $B = \hat{R}$. Obviously, this conclusion is valid only in the case of a perfect adequacy of noise modeling.

Consider now the general case (q sources); then by use of classical results of linear algebra (B3), one obtains

$$\frac{\partial}{\partial a_i} \log \det R = \text{tr} \left(R^{-1} \cdot \frac{\partial}{\partial a_i} R \right).$$

Now

$$\frac{\partial}{\partial a_i} B^{-1} = -B^{-1} \cdot \frac{\partial}{\partial a_i} B \cdot B^{-1}$$

and

$$\frac{\partial}{\partial a_i} B = -B \cdot \Delta_i \cdot B. \quad (\text{B6})$$

Therefore,

$$\frac{\partial}{\partial a_i} \log \det R = -\text{tr} (R^{-1} \cdot B \cdot \Delta_i \cdot B). \quad (\text{B7})$$

Furthermore,

$$\begin{aligned} \frac{\partial}{\partial a_i} \text{tr} (R^{-1} \cdot \hat{R}) &= \text{tr} \left(\frac{\partial}{\partial a_i} R^{-1} \cdot \hat{R} \right) \\ &= \text{tr} -(R^{-1} \cdot B \cdot \Delta_i \cdot B \cdot R^{-1} \cdot \hat{R}) \end{aligned} \quad (\text{B8})$$

so that, finally,

$$\frac{\partial}{\partial a_i} L(q, \mathbf{A}) = \text{tr} [B \cdot R^{-1} \cdot (R - \hat{R}) \cdot R^{-1} \cdot B \cdot \Delta_i]. \quad (\text{B9})$$

(This formula is, in fact, valid for any source number.)

Therefore, nullity of the gradient vector implies

$$\begin{aligned} \sum_{i=1}^p a_i \cdot \frac{\partial}{\partial a_i} L(q, \mathbf{A}) \\ = \text{tr} [(Id - \hat{R} \cdot R^{-1}) \cdot B \cdot R^{-1}] = 0. \end{aligned} \quad (\text{B10})$$

[Note the difference between (B4) and (B10).]

Equation (B10) can also be written as

$$\begin{aligned} \sum_{i=1}^p a_i \cdot \frac{\partial}{\partial a_i} L(q, \mathbf{A}) \\ = \text{tr} [T^* \cdot R^{-1} \cdot (R - \hat{R}) \cdot R^{-1} \cdot T] = 0 \end{aligned} \quad (\text{B11})$$

(with $B = T \cdot T^*$).

Now it is always possible to choose starting values of AR parameters and step sizes in order to assume the positivity of the matrix $(\hat{R} - R)$, which amounts [by the definition of likelihood functional (13)] to positivity of the matrix $(B_0 - B)(R \equiv S + B_0)$.

Consequently, nullity of the gradient vector implies [by (B11)] equality of exact and estimated parameters of the noise. Stress that this conclusion is valid only in the case of perfect adequacy of noise modeling and without consideration of estimation errors of \hat{R} .

This reasoning can be extended, by induction, to any source number, leading to the same result (i.e., $\mathbf{A}_{\max} = \mathbf{A}_0$). It may also be extended to the MA model.

Furthermore, H denoting the Hessian matrix of the likelihood functional, the following result can be easily

proven [19]:

$$\begin{aligned} \mathbf{X}^* \cdot H \cdot \mathbf{X} &= -\text{tr} \left[R^{-1} \cdot (2\hat{R} - R) \cdot R^{-1} \right. \\ &\quad \cdot \left(\sum_i x_i^* U_i^* \right) \cdot R^{-1} \cdot \left(\sum_j x_j U_j \right) \Big] \\ &\quad - \text{tr} \left[R^{-1} \cdot (\hat{R} - R) \cdot R^{-1} \right. \\ &\quad \cdot \left. \left(\sum_{i,j} x_i^* x_j V_{ij} \right) \right] \end{aligned}$$

with

$$U_i = \frac{\partial}{\partial a_i} B; V_{ij} = \frac{\partial^2}{\partial a_i \partial a_j} B. \quad (\text{B12})$$

It can be deduced from (B12) that the functional is concave in the vicinity of the exact parameter values.

Examine now the convergence of the gradient method. By use of (B9), the following inequality is obtained:

$$\begin{aligned} \|\nabla L(q, \mathbf{A}) - \nabla L(q, \mathbf{A}')\| \\ \leq C \cdot \|\mathbf{R}_{\mathbf{A},q} - \mathbf{R}_{\mathbf{A}',q}\|_F \leq C' \|\mathbf{A} - \mathbf{A}'\|. \end{aligned} \quad (\text{B13})$$

(C' : positive constant, and ∇ denoting the gradient vector.)

Consider now the equivalent problem of minimization of the functional $-L(q, \mathbf{A})$ [denoted $L'(\mathbf{A})$]; then

$$\begin{aligned} L'(\mathbf{A}) - L'(\mathbf{A}') \\ = \int_0^1 \langle \nabla L'(\mathbf{A} + \tau(\mathbf{A} - \mathbf{A}')), \mathbf{A} - \mathbf{A}' \rangle \cdot d\tau \\ = \langle \nabla L'(\mathbf{A}), \mathbf{A} - \mathbf{A}' \rangle \\ + \int_0^1 \langle \nabla L'(\mathbf{A} + \tau(\mathbf{A} - \mathbf{A}')) \\ - \nabla L'(\mathbf{A}), \mathbf{A} - \mathbf{A}' \rangle \cdot d\tau \\ \geq \langle \nabla L'(\mathbf{A}), \mathbf{A} - \mathbf{A}' \rangle \\ - \int_0^1 \|\nabla L'(\mathbf{A} + \tau(\mathbf{A} - \mathbf{A}')) \\ - \nabla L'(\mathbf{A})\| \cdot \|\mathbf{A} - \mathbf{A}'\| d\tau \end{aligned} \quad (\text{B14})$$

by the use of the Cauchy-Schwarz inequality [6] ($\langle \nabla L', \mathbf{A} \rangle$ denoting the scalar product $\sum a_i \cdot (\partial/\partial a_i) L'$).

Therefore, using (B10) and (B11), the following inequality is obtained:

$$\begin{aligned} L'(\mathbf{A}) - L'(\mathbf{A}') \\ \geq \langle \nabla L'(\mathbf{A}), \mathbf{A} - \mathbf{A}' \rangle - C' \|\mathbf{A} - \mathbf{A}'\|^2. \end{aligned} \quad (\text{B15})$$

Consider now the (classical) gradient algorithm (19):

$$\mathbf{A}_{k+1} = \mathbf{A}_k - \rho_k \nabla L'(\mathbf{A}_k).$$

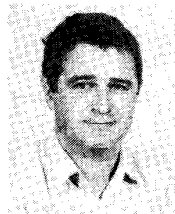
Then [by use of (B12)]

$$\begin{aligned} L'(A_k) - L'(A_{k+1}) \\ \geq (1 - C' \cdot \rho_k) \rho_k \cdot \|\bar{V}L'(A_k)\|^2. \quad (\text{B16}) \end{aligned}$$

Consequently, the sequence $\{L'(A_k)\}$ decreases (as k increases) if the step size ρ is sufficiently small (i.e., inferior to C'^{-1}). Now, by Property 1, this sequence is superior to zero; hence, it converges.

REFERENCES

- [1] 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. 1235-1247, Oct. 1983.
- [2] D. H. Johnson, "The application of spectral estimation methods to bearing estimation problems," *Proc. IEEE*, vol. 70, pp. 1018-1028, Sept. 1982.
- [3] T. Kailath, A. Vieira, and M. Morf, "Inverse of Toeplitz operators, innovations and orthogonal polynomials," *SIAM Rev.*, vol. 20, pp. 106-119, Jan. 1978.
- [4] A. Ben Artiz and T. Shalom, "On inversion of Toeplitz and close to Toeplitz matrices," *Linear Algebra Appl.*, vol. 75, pp. 173-192, 1986.
- [5] R. J. Muirhead, *Statistical Multivariate Analysis*. New York: Wiley, 1981.
- [6] P. Lancaster and M. Tismenestsky, *The Theory of Matrices with Applications*, 2nd ed. New York: Academic, 1985.
- [7] G. Heinig and K. Rost, *Algebraic Methods for Toeplitz-Like Matrices and Operators*. Basel: Birkhäuser Verlag, 1984.
- [8] B. Friedlander, "On the computation of the Cramer-Rao bound for ARMA estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 721-727, Aug. 1984.
- [9] G. H. Golub and C. F. Van Loan, *Matrix Computations*. Oxford: North Oxford Academic, 1983.
- [10] J. Vom Scheidt and W. Purket, *Random Eigenvalue Problems* (North Holland Ser. Prob. Appl. Math.). Amsterdam: North-Holland, 1983.
- [11] C. J. Demeure and L. L. Scharf, "Linear statistical models for stationary sequences and related algorithms for Toeplitz matrices," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 29-43, Jan. 1987.
- [12] S. Haykin, Ed., *Non-Linear Methods in Spectral Analysis*. New York: Springer-Verlag, 1983.
- [13] J. J. Fuchs, "Estimation du nombre de sinusoides dans du bruit coloré," *Actes du Onzième Colloque GRETSI*, vol. 1, pp. 197-200, June 5, 1987.
- [14] G. Polya and G. Szegő, *Problems and Theorems in Analysis I*. New York: Springer-Verlag, 1978, p. 64.
- [15] M. Wax and T. Kailath, "Detection of signals of information theoretic criteria," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 387-392, 1985.
- [16] J. Rissanen, "A universal prior for integers and estimation by minimum description length," *Ann. Statist.*, vol. 11, no. 2, pp. 416-431, 1983.
- [17] L. C. Zhao, P. R. Krishnaiah, and Z. D. Bai, "Remarks on certain criteria for detection of number of signals," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-35, pp. 129-133, Feb. 1987.
- [18] H. Akaike, "A new look at the statistical model identification," *IEEE Trans. Automat. Contr.*, vol. AC-19, pp. 716-723, Dec. 1974.
- [19] J. P. Le Cadre, "Contributions à l'utilisation de méthodes paramétriques en traitement d'antenne," thèse de Doctorat es Sciences, Univ. Grenoble, INPG, Oct. 1987.
- [20] K. S. Arun and S. Y. Kung, "Generalized principal components analysis and its applications in approximate stochastic realizations," in *Modeling and Application of Stochastic Process*, U.B. Desai, Ed. Boston: Klüwer Academic, 1986, pp. 75-105.
- [21] J. P. Le Cadre and P. Ravazzola, "Utilisation de modélisation d'état en traitement d'antenne," *Actes du Onzième Colloque GRETSI*, vol. 2, pp. 385-388, June 5, 1987.
- [22] G. Bienvenu and L. Kopp, "Principe de la goniométrie passive adaptative," *Actes du Septième Colloque GRETSI*, pp. 106-110, 1979.
- [23] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 276-280, Mar. 1986.
- [24] V. F. Pisarenko, "The retrieval of harmonics from a covariance function," *Geophys. J. Roy. Astron. Soc.*, no. 33, pp. 347-366, 1973.
- [25] A. Sano and K. Hashimoto, "Adaptive recursive scheme for spectral analysis of sinusoids in signals with unknown colored spectrum," in *Proc. Int. Conf. Acoust., Speech, Signal Processing*, Tampa, FL, 1985, pp. 109-112.
- [26] J. P. Burg, G. Luenberger, and D. L. Wenger, "Estimation of structured covariance matrices," *Proc. IEEE*, vol. 70, pp. 963-974, Sept. 1982.
- [27] G. Vezzosi, "Estimation of phase angles from the cross-spectral matrix," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-34, pp. 405-422, June 1986.
- [28] R. J. Urick, *Principles of Underwater Sound*. New York: McGraw-Hill, 1983.
- [29] W. S. Burdick, *Underwater Acoustic System Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1984.
- [30] E. M. Arase and T. Arase, "Correlations of ambient sea noise," *J. Acoust. Soc. Amer.*, vol. 40, no. 1, pp. 205-210, 1966.
- [31] D. R. Brillinger, *Time Series (Data Analysis and Theory)*, expanded ed. San Francisco, CA: Holden-Day, 1981.
- [32] S. Orfanidis, *Optimum Signal Processing*. New York: Wiley.
- [33] A. Botcher and B. Silbermann, *Invertibility and Asymptotics of Toeplitz Matrices*. Berlin: Akademie Verlag, 1983.



J. Pierre Le Cadre was born in Elven, France, in 1953. He received the M.S. degree in mathematics in 1976, the third cycle Doctorate in 1982, and the Docteur d'Etat degree in 1987, from INPG, Grenoble, France.

Since 1980 he has worked mainly on array signal processing at GERDSM. Since 1983 he has been with CNRS. He is now joining IRISA, Rennes.