An Error Bound for a Noise Canceller

PIERRE COMON, MEMBER, IEEE, AND DINH TUAN PHAM, MEMBER, IEEE

Abstract—Nonparametric estimation of a random signal from a set of observations is often carried out in two steps (for instance, especially in the MAP and Bayesian approaches). First, an optimal solution is sought by assuming that the deterministic parameters entering the probability distribution of signal or observation, such as the covariance matrices, are known. Then, in order to apply the latter solution, the covariances must actually be replaced by estimates, which confers the filter its adaptive character, and at the same time compromises its optimality, according to the criterion assumed initially. Taking advantage of the statistical properties of the complex Wishart matrix and its inverse, we investigate the decrease in performance stemming from this substitution, in the case of Gaussian signal and observations. Our performance criterion is based on the computation of the final output error, which includes covariance deviations. This is carried out for a certain class of linear observation models, involving an underlying problem of noise cancelling using Noise Alone References (NAR), namely, array processing techniques using a known array manifold.

I. Introduction

LINEAR Least Mean Square Estimation (LMSE) requires knowledge of the covariance structure of some system. Assuming stationarity, it consists of correlation functions in the time or space domain, and of power spectral densities in the frequency or wavenumber domain. These matrices are often unknown, and, in an adaptive context, they are replaced by estimates. It is of interest to evaluate the degradation in performance of the LMSE involved when replacing these matrices by estimates.

Several authors have already discussed the importance of this problem, and methods have been proposed that are robust with respect to covariance uncertainties [8], [15], [9]. In these techniques, uncertainty modeling is used, such as the set of all possible covariance matrices. In contrast, our approach is based on the computation of the final output noise variance, which includes estimation errors. This principle was first introduced in a paper of Akaike [1]. Our approach takes advantage of the Gaussian character of both signal and noise, and signal and noise are assumed independent throughout the paper. This paper extends the previous results stated in [4].

Consider the following linear statistical model:

$$R(k) = A(k)s(k) + e(k) \tag{1}$$

Manuscript received April 25, 1986; revised December 13, 1988.
P. Comon is with Thomson Sintra, BP 53, Cagnes sur Mer, France 06801.
D. T. Pham is with TIM3, Imag BP 68, St Martin d'Heres, France 38402.
IEEE Log Number 8930004.

where $R(k)$ denotes the $p \times 1$ observation vector, and k denotes either a time, space, frequency, or wavenumber known discrete variable; we do not specify this practical aspect for the moment, but this will be made clear by some examples. For each k, $s(k)$ and $e(k)$ are assumed to be independent complex normal vectors as defined in [6], of dimension r and p, respectively ($r < p$), with zero mean and covariance matrices:

$$G_s(k) = E\left\{s(k)s(k)^\dagger\right\}, \qquad G_e(k) = E\left\{e(k)e(k)^\dagger\right\} \tag{2}$$

where $^{(\dagger)}$ denotes transpose and complex conjugated. Recall that a random vector is said to be complex normal if its real and imaginary parts are jointly normal, and if additionally their covariance matrices are equal and the cross covariance matrix is skew-symmetric. This yields among other things that a complex normal random vector R satisfies $E\{RR'\} = 0$. Finally, $A(k)$ is a known deterministic p by r matrix of full rank r for each k, and each of its columns defines a 1-dimensional manifold in a p-dimensional space. Vectors $s(k)$ and $e(k)$ correspond to signal and noise, respectively, and our purpose is to estimate the signal, and its power, from the observations of $R(k)$. The variable k will be omitted when this can be done without ambiguity. In this paper, the sequences will be assumed complex-valued for more generality, but similar results hold in the real-valued case.

Examples: For $r = 1$, $A(k)$ could be a steering vector (known for each k). This includes the model of power selective analysis, where parameter k denotes either an angle of arrival (spatial analysis, detection of arrival) or the frequency (spectral analysis). Let us mention for instance the so-called Maximum Likelihood Method (MLM) introduced by Capon, and then used under different forms by Lacoss [11] and others [7]. Also, the MUSIC approach introduced by Schmidt [16] and Bienvenu-Kopp [2] is based on the same model, but takes advantage of a representation of the covariance matrix in terms of its eigenpairs. In these cases, vector $R(k)$ does not depend on k.

On the other hand, in the problems of noise cancellation, some observations contain noise alone [13], and can be represented by setting $A = [0, I_r]^T$ in our model (1). In that case, A does not depend on k, but R does.

In the next section, based on the model (1) where we see now the polyvalence, we stress the underlying role of Noise Alone References in multisensor processing as soon as the dimension of the signal subspace r is smaller than the number of sensors p.

0096-3518/89/1000-1513$01.00

II. Statement of the Problem

Since covariance matrices are unknown, a set of M independent realizations $\Omega(k) = \{R^1(k), \cdots, R^\mu(k), \cdots, R^M(k)\}$, each satisfying the model (1), will be necessary. Hence, we have $R^\mu(k) = A(k)s^\mu(k) + e^\mu(k)$ for any μ between 1 and M. It is well known that the best LMSE of s^μ given $\Omega(k)$, denoted by a superscript (°) standing for "optimal," is then given by

$$s^{o\mu}(k) = E(s^\mu(k)/\Omega(k)) = G_s A^\dagger G_R^{-1} R^\mu \quad (3)$$

where

$$G_R = E\{RR^\dagger\} = AG_sA^\dagger + G_e. \quad (4)$$

At this point, it is worthwhile to make a transformation of the observation, which provides an interesting interpretation of (3). Denote l_p^2, the Hilbert space of deterministic p-vectors, with the inner product $X^\dagger Y$, and h_p^2, the Hilbert space of random p-vectors of second order, with the inner product $E\{X^\dagger Y\}$. Now, since A is a p by r matrix of full rank, it is always possible to define a semi-unitary $p \times p - r$ matrix B so that the matrix $[A, B]$ is regular. Define the linear operators $P_A = (A^\dagger A)^{-1}A^\dagger$ and $P_B = B^\dagger$. Thus, AP_A and BP_B represent the operators which project onto the linear space $l_p^2(A)$ spanned by the columns of A, and onto its orthogonal complement, respectively. The following linear transform is bijective, and can be applied to any observation R^μ:

$$R^\mu \in h_p^2 \to P(R^\mu) = \begin{pmatrix} u^\mu \\ v^\mu \end{pmatrix}; \quad (5)$$

$$u^\mu = P_A R^\mu \in h_p^2, \qquad v^\mu = P_B R^\mu \in h_{p-r}^2.$$

This transform has the advantage of providing the Noise Alone Reference (NAR) vector v [14]. Indeed, we have from (1)

$$u(k) = s(k) + x(k); \qquad x(k) = P_A(k)e(k) \quad (6)$$

$$v(k) = y(k); \qquad y(k) = P_B(k)e(k). \quad (7)$$

Let us express (3) in terms of these new variables u and v. Denote the four blocks contained in the covariance matrix of $P(R)$ by G_u, C_{uv}, G_{vu}, and G_v. Then, using the inversion lemma for block matrices, it can be shown that [14]

$$s^{o\mu} = G_s(G_u - G_{uv}G_v^{-1}G_{vu})^{-1}(u^\mu - G_{uv}G_v^{-1}v^\mu). \quad (8)$$

This can be rewritten, with obvious notations, $s^{o\mu} = F_s(u^\mu - Fv^\mu)$, which shows that the optimal filter (3) can be split into two filters F and F_s as sketched in Fig. 1. In fact, $u^\circ = (u - Fv)$ provides a first estimate, removing the noise linked to the NAR and present in the signal space $l_r^2(A)$. This estimate is next improved by the second filter F_s, whose output s° is precisely the best Bayesian estimate $E(s/u^\circ)$, viz., s° is the projection of s onto the space $h_r^2(u^\circ)$.

III. Performance Limits of the Noise Canceller

It is quite clear that the performance of the estimation process depends on the correlation (or the coherency) between the NAR, v, and the noise disturbing the signal x. In the extreme case where the noises are uncorrelated, there is no gain in running any processing: one should take u^μ as the estimate of the signal s^μ. Moreover, as is shown below, due to errors in estimating the covariance matrices, it is still preferable to perform no processing when the coherency is not strong enough. This defines an error bound to the filter deviations.

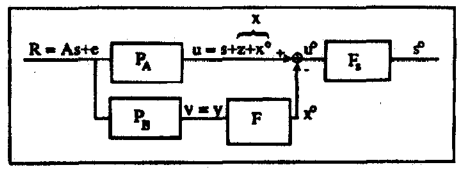

Fig. 1. Theoretical MS optimal filtering.

A. Qualitative Measure of Performance

From now on, and up to the end of the paper, we focus on the limitations of the first operator F defined in the previous section, namely, the noise canceller. The estimate $\hat{s}$ of the signal s is computed by first calculating an estimate $\hat{x}$ of the corrupting noise x, and then by a subtraction so that $\hat{s} = u - \hat{F}v$, where $\hat{F}$ denotes the estimate of F (see Fig. 2). Two kinds of errors must then be considered. First, the minimal error $z = x - x^\circ$, which cannot be outperformed. That is, if there were no estimation errors in the covariance matrices, the best estimate of x would be $x^\circ = x - z = Fv$. The noise z is therefore independent of x°. Second, there is an extraneous error, ϵ, made in estimating x° from the observation v, and due only to estimation errors, viz. $\epsilon = x^\circ - \hat{x} = Fv - \hat{x}$. This may be efficiently summarized by the relevant equations

$$u = s + z + x^\circ = u^\circ + x^\circ = s + x: \quad \text{observation}$$
$$\hat{s} = s + z + \epsilon = u^\circ + \epsilon = u - \hat{x}: \quad \text{estimation.} \quad (9)$$

We are now in a position to study the problem of sensitivity of the estimation $\hat{s}$, that is, the output of the noise canceller, to covariance uncertainties. We shall say that the estimated filter performs badly if it is "worse" than an arbitrarily chosen trivial filter, the quality criterion being the Output Signal-to-Noise Ratio (OSNR) for each component of the output:

$$\text{OSNR(estimated filter)}(k) \le \text{OSNR(trivial filter)}(k) \quad (10)$$

where the OSNR are defined for each k by the ratio of the quadratic norms in h_1^2.

$$\text{OSNR}(k) = \|\text{signal}(k)\|^2 / \|\text{noise}(k)\|^2.$$

Of course, this criterion is fundamentally linked to the choice of the trivial filter. Since we deal with a noise reduction problem, this criterion turns out to be very natural if we choose subtraction by zero from the observation as

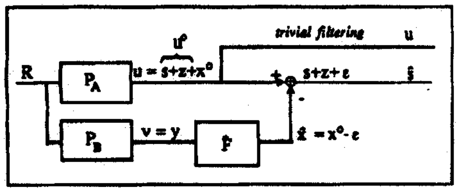

Fig. 2. Actual filtering with estimated covariances.

a trivial filter. Consequently, if the criterion is satisfied, with this choice of trivial filter, the adaptive filter will always improve the SNR by definition. Consider now the structure that we deal with presented in Fig. 2. Since u corresponds to the output (and input) of the trivial filter, the performance criterion is

$$\|z_i(k) + \epsilon_i(k)\|^2 \le \|z_i(k) + x_i^\circ(k)\|^2, \qquad 1 \le i \le r \tag{11}$$

for each component i of the output. We note that if the noise z is independent of the error ϵ, this criterion is equivalent to

$$\|\epsilon_i(k)\|^2 \le \|x_i^\circ(k)\|^2, \qquad 1 \le i \le r. \tag{12}$$

B. Computation of the Final Error for a Spectral Filter

As previously mentioned, we shall deal with observations $R(k)$ that are complex Gaussian random variables for each k, as introduced by Goodman [6]. They can be obtained, for example, by Discrete Fourier Transform of stationary time series. In the case of a spectral filter, the variable k denotes the frequency, and the $R(k)$'s are asymptotically independent. If we wanted to deal with a transversal filter acting in time domain instead, our derivation would still hold but would need the samples $R(k)$ to be independent, namely, the series $R(k)$ to be white. In this section, we present an exact calculation of the final error, which extends preceding results [4]. The basic idea lies in the use of the fact that the PSD estimates (Averaged Periodograms) are asymptotically complex Wishart matrices with M degrees of freedom up to a factor $1/M$ [3]. For instance,

$$MQ_v(k) \sim W_q^c(M, G_v(k)). \tag{13}$$

If the covariance matrices are estimated by Smoothed Averaged Periodograms (SAP), then

$$Q_{u_iv}(k) = \sum_{\mu=1}^{M} \sum_{n=0}^{N-1} w(n-k) u_i^\mu(n) v^{\mu\dagger}(n). \tag{14}$$

SAP estimates (14) are asymptotically complex Wishart variables with βM degrees of freedom, provided that the considered frequency k is sufficiently far from the zero and Nyquist frequencies [3], β being the bandwidth of the smoothing window $w(n)$ defined as

$$\beta = 1/\sum w^2(n), \qquad \text{with } \sum w(n) = 1. \tag{15}$$

Let us drop the variable k for convenience, and assume for the moment that we are dealing with Averaged Periodograms (no smoothing). The general case will be studied in a next subsection. Each component i of the error ϵ, $1 \le i \le r$, may be written as

$$\epsilon_i = x_i^\circ - \hat{F}_i v \tag{16}$$

$\hat{F}_i$ being the ith column of the estimated filter gain

$$\hat{F}_i = Q_{u_iv} Q_v^{-1}. \tag{17}$$

Remembering that the 1 by r row vector F_i is the true filter gain, this can be equivalently written

$$\epsilon_i = (F_i - \hat{F}_i)v. \tag{18}$$

Now assume the estimated filter $\hat{F}_i$ fitted to the series $\{u^\mu, v^\mu, 1 \le \mu \le M\}$ is applied to another realization v independent of $\{u^\mu, v^\mu\}$. This important assumption is very helpful and is always made in regression theory. We have then

$$\|\epsilon_i\|^2 = E\{\epsilon_i\epsilon_i^\dagger\} = E\{\text{trace}\,[G_v^{-1}(\hat{F}_i - F_i)^\dagger(F_i - F_i)]\}. \tag{19}$$

On the other hand, if we note that $u^\mu = s^\mu + z^\mu + Fv^\mu$, (13) yields

$$Q_{uv} = (1/M) \sum_{\mu=1}^{M} (s^\mu + z^\mu)v^{\mu\dagger} + FQ_v.$$

Thus, from (17)

$$(\hat{F}_i - F_i) = (1/M) \sum_{\mu=1}^{M} (s_i^\mu + z_i^\mu)v^{\mu\dagger} Q_v^{-1}. \tag{20}$$

Observe that s^μ as well as z^μ are independent of v^μ; hence, the conditional distribution of $(\hat{F}_i - F_i)^\dagger$ given $\{v^1, \cdots, v^\mu, \cdots, v^M\}$, that is, Q_v is given, is complex normal with zero mean and covariance matrix

$$\Theta_i = 1/M \operatorname{var}(s_i^\mu + z_i^\mu) Q_v^{-1}.$$

Thus, the conditional expectation of $|\epsilon_i|^2$ given $\{v^1, \cdots, v^\mu, \cdots, v^M\}$ is from (18), recalling that $u^\circ = s + z$

$$E\{|\epsilon_i|^2 / v^1, \cdots, v^\mu, \cdots, v^M\} = (1/M) G_{u^\circ_i} \text{trace}\,(G_v Q_v^{-1}). \tag{21}$$

The final prediction error is the expectation of the above right-hand side. From the lemma stated in the Appendix, we have

$$\alpha = E\{\text{trace}\,(G_v Q_v^{-1})\} = q/(M-q) \tag{22}$$

where q denotes the number of NAR, i.e., $q = p - r$. Furthermore, observe from Section II that $G_{u^\circ_i} = G_{u_i} - G_{x_i}$ (this is due to the fact that u° and x° are orthogonal in h_r^2) and also $G_{x^\circ_i} = G_{u_i} C_{u_iv}$, if $C_{u_iv}(k)$ denotes the multiple squared coherency between the component $u_i(k)$ and the vector $v(k)$

$$C_{U_iV}(k) = G_{U_iV}(k) G_V^{-1}(k) G_{VU_i}(k) / G_{U_i}(k).$$

Then the foregoing expressions are derived for each frequency k

$$E\{|\epsilon_i|^2\} = \alpha/M(G_{Ui} - G_{x^\circ_i}) = G_{x^\circ_i} + [\alpha/M - (1 + \alpha/M)C_{UiV}(k)]G_{Ui} \tag{23}$$

where $\alpha = q/(M - q)$ as defined in (22). This error expression includes the deviations in covariance matrices G_{UiV} and G_V, which are generally correlated, since they are estimated from the same set of observations Ω. It can easily be verified from the first equation in (23) that for large samples, the final error tends to zero. It may be seen from the right-hand side of (23) that the final error $\|\epsilon_i(k)\|^2$ cannot be evaluated. On the other hand, and this is the whole astuteness of the method, the difference $\|\epsilon_i(k)\|^2 - \|x_i^\circ(k)\|^2$ may be accessed, which is sufficient for our purpose.

C. Extension to Smoothed and Exponentially Averaged Periodograms

In adaptive processing, Smoothed Exponentially Averaged Periodograms (SEAP) are of interest because they allow an updating of covariance matrices on-the-fly. The result (23) is available for Averaged Periodograms, but may be extended to an exponential average instead. More precisely, a SEAP with an exponential averaging like

$$Q_{u_iv}(k, m) = (1 - \delta)Q_{UiV}(k, m - 1) + \delta \sum_{n=0}^{N-1} w(n - k)u_i^m(n)v^{m\dagger}(n) \tag{24}$$

is a block of a matrix which is asymptotically distributed as $W_p^c(L, Q_{u_iv}(k/L)$, where the number of degrees of freedom L is equal to

$$L = \beta(2 - \delta)/\delta, \quad 0 < \delta < 1. \tag{25}$$

Result (23) holds by replacing M by L, without forgetting that $\alpha = q/(M - q)$ has also to be replaced by $\alpha = q/(L - q)$. The same remark holds for usual SAP as in (14), where the number of degrees of freedom is $L = \beta M$.

D. A Robust Noise Canceller

The purpose of this section is now to use the performance criterion (12) and the final error expression (23) to build an adaptive noise canceller which is less sensitive to covariance deviations. A robust filter F^R can simply be designed by choosing the best filter from the estimated and the trivial, according to (12). Taking into account the result (23), we get the performance criterion

$$C_{u_iv}(k) \geq t; \quad t = 1/(1 + L/\alpha). \tag{26}$$

So, if the coherency is greater than the threshold t, the filter performs well, i.e., it reduces the noise, and increases the OSNR. On the other hand, if it is smaller, the harmful effect of the filter can be prohibited by setting it to zero. Hence, it coincides in this case with the trivial filter. Let $T_i(k)$ be a function whose values are one if the filter performs well, and zero otherwise. A robust filter can be defined as

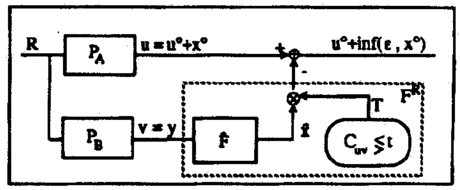

Fig. 3. Robust noise canceller.

$$F_i^R(k) = \hat{F}_i(k)T_i(k) \tag{27}$$

where

$$T_i(k) = 1 \quad \text{if } C_{u_iv}(k) \geq t$$

$$T_i(k) = 0 \quad \text{if } C_{u_iv}(k) < t.$$

This testing and correction procedure is sketched in Fig. 3, and may be applied independently to each component i of the input $u(k)$. Of course, coherency $C_{u_iv}(k)$ has to be itself estimated, and contains relative errors. So, the correction is not perfect, and some "ineffective ranges" might obviously still remain. As a result, the robust filter is not totally insensitive to covariance uncertainties. Taking into account the fact that α is a function of L and q yields

$$t = q/[q + (L - q)] = q/L. \tag{28}$$

Let us finally review the different possible forms, according to the assumed covariance estimates. For an Averaged Periodogram, we would have $t = q/M$, for an SAP defined in (14), we would have $t = q/\beta M$, and finally for an SEAP as defined in (24), we would have $t = q\delta/\beta(2 - \delta)$. Recall that q denotes the size of the NAR vector v, and that the expression for b is given in (15).

IV. Summary and Concluding Remarks

In Sections I and II, the Noise Alone References (NAR) have been shown to be relevant in a class of array processing models with a known array manifold matrix. In Section III, performance limits of the adaptive noise canceller have been investigated, specifically for the case of a spectral filter to illustrate the ideas. It has been shown that its performance depends on the quadratic coherency between the corrupted signal and the NAR. Indeed, it is preferable to run no processing when this coherency is not strong enough. A rule has been proposed in order to decide when to switch to the "no processing" scheme. Contrary to some previous results, the results given here are available without any approximations, provided the statistical properties are satisfied. In the case of a spectral noise canceller, these properties are asymptotically satisfied. On the other hand, in the case of a transversal noise canceller acting in time domain, these results are valid for white signal and noises.

Appendix

Lemma: Let Q be $W_q^c(M, \Sigma)$, then

$$E\{Q^{-1}\} = \Sigma^{-1}/(M - q).$$

Proof: Kshirsagar has shown that in the real case we have [10, lemma 10, p. 72]: $E[Q^{-1}] = 1/(M - q - 1)\Sigma^{-1}$. Our proof is very similar to that of this lemma. First, let $\Sigma = CC^{\dagger}$; by using a linear transform, it is easily seen that

$$E\{Q^{-1}\} = C^{-\dagger} \cdot E\{A^{-1}\} \cdot C^{-1},$$

where A is $W_q^c(M, I)$.

Next, if a^{ij} are the generic elements of A^{-1}, then $E\{a^{ij}\} = 0$ if $i \neq j$ and $E\{a^{ii}\} = E\{a^{qq}\}$, by the same arguments as used in [10]. So, it remains only to compute $E\{a^{qq}\}$.

The variable $1/a^{qq}$ is $W_1^c(M - q + 1)$; this means that $2/a^{qq}$ is $W_1(2(M - q + 1), 1)$, that is, a χ^2 variable with $2(M - q + 1)$ degrees of freedom. Thus, it is well known that

$$E\{a^{qq}/2\} = 1/[2(M - q + 1) - 2]$$

which yields finally $E\{a^{qq}\} = 1/(M - q)$. ◊

Acknowledgment

The authors wish to thank Prof. B. Picinbono for his especially detailed and helpful reviews.

References

[1] H. Akaike, "Fitting autoregressive models for prediction," *Ann. Inst. Stat. Math.*, vol. 21, no. 2, pp. 243–247, 1969.

[2] G. Bienvenu and L. Kopp, "Optimality of high-resolution array processing using the eigensystem approach," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-31, pp. 1235–1248, Oct. 1983.

[3] D. R. Brillinger, *Time Series, Data Analysis and Theory*. San Francisco, CA: Holden Day, 1981.

[4] P. Comon and J. L. Lacoume, "Noise reduction for an estimated Wiener filter using noise references," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 310–313, Mar. 1986.

[5] ——, "A robust adaptive filter for noise reduction problems," presented at the Int. Conf. ICASSP, Tokyo, Japan, Apr. 7–11, 1986.

[6] N. R. Goodman, "Statistical analysis based on a certain multivariate complex Gaussian distribution," *Ann. Math. Stat.*, vol. 34, pp. 152–177, 1963.

[7] S. Haykin, *Array Signal Processing*. Englewood Cliffs, NJ: Prentice-Hall, 1979.

[8] S. A. Kassam and T. L. Lim, "Robust Wiener filters," *J. Franklin Inst.*, vol. 304, pp. 172–185, Oct./Nov. 1977.

[9] S. A. Kassam and H. V. Poor, "Robust techniques for signal processing. A survey," *Proc. IEEE*, vol. 73, pp. 443–481, Mar. 1985.

[10] A. M. Kshirsagar, *Multivariate Analysis*. New York: Dekker, 1972.

[11] R. T. Lacoss, "Data adaptive spectral analysis methods," *Geophysics*, vol. 36, no. 4, pp. 661–675, 1971.

[12] M. A. Lagunas *et al.*, "An improved ML method for power spectral density estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-32, pp. 170–173, Feb. 1984.

[13] H. Mermoz, "Ecueils et diversite des traitements adaptatifs d'antenne," *Annales des Télécom.*, vol. 28, no. 5-6, pp. 244–254, 1973.

[14] B. Picinbono, "A geometrical interpretation of signal, detection and estimation," *IEEE Trans. Inform. Theory*, vol. IT-26, pp. 493–497, July 1980.

[15] M. V. Poor, "On robust Wiener filtering," *IEEE Trans. Automat. Contr.*, vol. AC-25, pp. 531–536, June 1980.

[16] R. O. Schmidt, "Multiple emmiter location and signal parameter estimation," in *Proc. RADC Spectrum Estimation Workshop*, Oct. 1979; reprinted in *IEEE Trans. Antennas Propagat.*, vol. AP-34, pp. 276–280, Mar. 1986.

[17] H. L. Van Trees, *Detection, Estimation and Modulation Theory*, vol. I. New York: Wiley, 1968.

Pierre Comon (M'87) was born in Paris, France, on January 28, 1959. He received the Ph.D. degree in electrical engineering in 1985 from the University of Grenoble, France.

From 1982 to 1985 he was with Crouzet Valence, France, and from 1981 to 1986 with the Center of Geophysics and Random Phenomena Studies (CEPHAG), France. In 1987 he was visiting the Information Systems Laboratory at Stanford. He has been with the Department of Submarine Acoustics, Thomson Sintra, France, since 1988. His current research interests include adaptive filtering, fast algorithms, sonar array processing, and neural nets.

Dinh Tuan Pham (M'88) was born in Hanoi, Vietnam, on February 15, 1945. He graduated from the Engineering School of Applied Mathematics and Computer Science (ENSIMAG) of the Polytechnic Institute of Grenoble in 1968. He received the Ph.D. degree in statistics in 1975 from the University of Grenoble.

He was a Postdoctoral Fellow at Berkeley (Department of Statistics) in 1977–1978, and a Visiting Professor at Indiana University (Department of Mathematics) at Bloomington in 1979–1980. His research interests include time series analysis, array processing, spectrum estimation and modeling, and biomedical signal processing.