Asymptotic Performance of Second-Order Algorithms

Jean-Pierre Delmas, *Member, IEEE*

***Abstract*—This paper re-examines the asymptotic performance analysis of second-order methods for parameter estimation in a general context. It provides a unifying framework to investigate the asymptotic performance of second-order methods under the stochastic model assumption in which both the waveforms and noise signals are possibly temporally correlated, possibly non-Gaussian, real, or complex (possibly noncircular) random processes. Thanks to a functional approach and a matrix-valued reformulated central limit theorem about the sample covariance matrix, the conditions under which the asymptotic covariance of a parameter estimator are dependent or independent of the distribution of the signal involved are specified. Finally, we demonstrate the application of our general results to direction of arrival (DOA) estimation, identification of finite impulse response models, sinusoidal frequency estimation for mixed spectra time series, and frequency estimation of sinusoidal signal with very lowpass envelope.**

***Index Terms*—Asymptotic covariance, asymptotic robustness, central limit theorem, direction-of-arrival estimation, finite-impulse response identification, sample covariance matrix, second-order algorithms, sinusoidal frequency estimator.**

I. Introduction

THE problem of estimating parameters of waveforms embedded in additive noise based on second-order algorithms (i.e., algorithms using second-order statistics from the data only) has been intensively studied in the signal processing community due to its wide applicability, mostly explicit physical interpretation, simplicity of implementation, and often good performance. Performance analyses of such algorithms derive from several signal models. The deterministic (or conditional) and the stochastic (or unconditional) model are the main models that have appeared in the literature (see, e.g., [1] and [2]). The noise is assumed to be a temporally uncorrelated Gaussian random process in these two models, but the waveforms are assumed to be fixed in all the realizations in the deterministic model and to be generally Gaussian temporally uncorrelated random processes in the stochastic model. Many authors have compared the asymptotic performance of parameter estimators with these two models and connected their performance to the deterministic or stochastic Cramer–Rao bound (see, e.g., [1] and [2] and the reference therein). Among the performance studies carried out for stochastic models, some authors have been interested in the invariance of the asymptotic distribution of parameter estimators to the distribution and to temporal correlation of the involved signals, which is often named *robustness to distribution and to temporal correlation*. Cardoso and Moulines [3] have shown that the asymptotic performance of most high resolution covariance-based DOA estimators is independent of the distribution of the source signals for independent snapshots. This robustness property was extended to the temporal correlation of the source signals and clarified in [4], where it is proved that Toeplitzation and the augmentation techniques are very sensitive to this correlation. Abed Meraim *et al.* [5] presented an asymptotic performance analysis of subspace methods for blind identification of single-input multiple-output FIR systems where it is shown that the higher than second-order statistics of the input signals do not affect the asymptotic covariance of the estimated impulse response. Besides these works, most asymptotic performance analyses rest on the assumption that the sample covariance matrix of the data has a Wishart distribution (see, e.g., [6]). This assumption, however, is valid only if the signals are Gaussian and i.i.d.

It is thus of importance to determine if the performance is affected by the joint distribution of the signals. The purpose of this contribution is to provide a unifying framework to investigate the asymptotic performance of second-order methods for parameter estimation in a general context under the stochastic model assumption in which both the waveforms and noise signals are possibly temporally correlated, possibly non-Gaussian, real, or complex (possibly noncircular) random processes. In this context, the performance is *a priori* expected to depend on the joint distribution of the signals involved. We adopt a functional approach in which the Gaussian asymptotic distribution of the covariance-based parameter estimates is derived from the Gaussian asymptotic distribution of the sample covariance matrix that is proved for this general model. This allows us to give closed-form expressions for the asymptotic covariance matrices of parameter estimates. We then examine under which conditions the asymptotic covariance of parameter estimators are dependent (or not) on the probability distribution and on the temporal correlation of the signals involved. In particular, for the DOA estimation, it is established that under mild assumptions and under the condition that the noise is temporally uncorrelated, the asymptotic covariance matrix of parameter estimates is independent of the distribution and of the temporal correlation of the waveforms. On the other hand, this asymptotic covariance is sensitive to the temporal correlation of the signals involved when the noise is temporally correlated, which is the case when the observed signal are oversampled or when the noise includes jammers. This result shows that the classic asymptotic robustness property (see, e.g., [2]) is only valid in the temporally

Manuscript received October 24, 2000; revised September 28, 2001. The associate editor coordinating the review of this paper and approving it for publication was Dr. Olivier Besson.

The author is with the Département CITI, Institut National des Télécommunications, Evry, France (e-mail: jean-pierre.delmas@int-evry.fr).

Publisher Item Identifier S 1053-587X(02)00403-8.

white noise case. Moreover, we prove that the noise whitening approach used classically when the spatial noise correlation is known is very sensitive to this correlation.

This paper is organized as follows. After the general data model and some assumptions are introduced, some examples are given in Section II. In Section III, some regularity conditions assumed for the algorithms under study are specified, and a general functional approach providing a unifying framework for asymptotic performance analysis is presented. In Section IV, the asymptotic normality of the sample covariance matrix is established for these general data models where a matrix valued reformulated central limit theorem is given. This methodology is then applied to direction-of-arrival (DOA), finite impulse response (FIR), and sinusoidal frequency estimators for mixed spectra times series and for sinusoidal signals with very lowpass envelopes. Finally, some remarks concerning the noise temporal correlation and about the whitening approach are given in Section V.

The following notations are used throughout the paper. T, H, and $*$ stand for transpose, conjugate transpose, and conjugate, respectively. $\mathbf{O}$ and $\mathbf{o}$ denote matrices and column vectors with zero entries, respectively. Vec$(.)$ is the "vectorization" operator that turns a matrix into a vector consisting of the columns of the matrix stacked one below another. Depending on whether the data are real or complex-valued, superscript $+$ stands for transpose or conjugate transpose, and the Kronecker product $\mathbf{A}\otimes\mathbf{B}$ is the block matrix, the (i,j) block element of which is $a_{i,j}\mathbf{B}$ or $b^*_{i,j}\mathbf{A}$.[1] The vec-permutation matrix $\mathbf{K}$ transforms Vec$(\mathbf{A})$ to Vec$(\mathbf{A}^T)$ for any square matrix $\mathbf{A}$. Diag$(a_1,\ldots,a_n)$ is a diagonal matrix with diagonal elements a_i. The symbol 1_A denotes the indicator function of the condition A, which assumes the value 1 if this condition is satisfied and 0 otherwise, and the symbol $\xrightarrow{\mathcal{L}}$ denotes the convergence in distribution.

II. Data Model

A. General Hypotheses

In many applications, it is of interest to estimate the parameter $\Theta\in\boldsymbol{\Theta}\subset R^q/C^q$ from the following p-variate real or complex (possibly noncircular) valued wide sense stationary time series

$$\mathbf{x}_t=\mathbf{E}_s(\Theta)\mathbf{s}_t+\mathbf{n}_t\quad t=1,\ldots,N. \tag{2.1}$$

$\mathbf{E}_s(\Theta)\mathbf{s}_t$ and $\mathbf{n}_t$ model the signals of interest and additive measurement noise, respectively. It is assumed that $\mathbf{E}_s(\Theta)$ is deterministic and known as a function of the unknown signal parameters Θ. Of course, the probability distribution of $(\mathbf{x}_t)_{t=1,\ldots,N}$ depends on extra parameters, which are also unknown, but we are only interested here in the estimation of parameters Θ. In this general model, $\mathbf{s}_t$ and $\mathbf{n}_t$ are multivariate independent, zero-mean, second-order stationary time series of covariance matrices $\mathbf{R}_s\stackrel{\text{def}}{=}\text{E}(\mathbf{s}_t\mathbf{s}_t^+)$ and $\mathbf{R}_n\stackrel{\text{def}}{=}\text{E}(\mathbf{n}_t\mathbf{n}_t^+)$. Thus, the covariance matrix of $\mathbf{x}_t$ is

$$\mathbf{R}_x\stackrel{\text{def}}{=}\text{E}\left(\mathbf{x}_t\mathbf{x}_t^+\right)=\mathbf{E}_s(\Theta)\mathbf{R}_s\mathbf{E}_s^+(\Theta)+\mathbf{R}_n. \tag{2.2}$$

[1]This unusual convention makes it easier to deal with complex matrices for which Vec$(\mathbf{ABC}^{\text{H}})=(\mathbf{A}\otimes\mathbf{C})Vec(\mathbf{B})$.

We suppose that the parameterization used is identifiable to the second-order for the signal parameter Θ only, i.e.,

$$\mathbf{E}_s(\Theta_1)\mathbf{R}_s^{(1)}\mathbf{E}_s^+(\Theta_1)+\mathbf{R}_n^{(1)}=\mathbf{E}_s(\Theta_2)\mathbf{R}_s^{(2)}\mathbf{E}_s^+(\Theta_2)+\mathbf{R}_n^{(2)}\Rightarrow\Theta_1=\Theta_2$$

whatever the expressions $\mathbf{R}_s^{(1)}$ and $\mathbf{R}_s^{(2)}$ [resp., $\mathbf{R}_n^{(1)}$ and $\mathbf{R}_n^{(2)}$] of $\mathbf{R}_s$ [resp., $\mathbf{R}_n$] compatible with their structure required by the algorithms. The studied model does not suppose that the nuisance parameter, i.e., the parameters that parametrize $\mathbf{R}_s$ and $\mathbf{R}_n$, are identifiable to the second order. Some examples of this data model are briefly described in Section II-B.

B. Examples of Application

1) Narrowband DOA Estimation: $\mathbf{x}_t$ represents the p-vector of the observed complex envelope at the sensor output. $\mathbf{s}_t\stackrel{\text{def}}{=}(s_{t,1},\ldots,s_{t,K})^T$, where $s_{t,k}$ is the complex envelope of the emitted signal by the source k at time t. $\mathbf{E}_s(\Theta)$ is the $p\times K$ "steering" matrix, and Θ is the spatial parameters of the K sources, which are referred to as the DOAs of the sources. $\mathbf{R}_s$ is the spatial covariance matrix (which is assumed positive definite and diagonal for the algorithms that require the sources spatially uncorrelated, e.g., in the Toeplization and the augmentation techniques; see, e.g., [4]). $\mathbf{n}_t$ and $\mathbf{R}_n=\sum_{l=1}^L a_l\mathbf{Q}_l$ are, respectively, the complex envelope and the spatial covariance matrix of the sensor output additive noise where a_l are unknown parameters and $\mathbf{Q}_l$ are known Hermitian weighting matrices (see, e.g., [7]). By choosing $L=1$, $a_1=\sigma^2$, and $\mathbf{Q}_1=\mathbf{I}_p$, the special case where the noise is spatially white is obtained.

2) Blind Identification of FIR Channels: $\mathbf{x}_t$ represents the p-vector of observed complex envelope of the channel output. $\mathbf{x}_t\stackrel{\text{def}}{=}(x_{t,1},\ldots,x_{t,K_o},x_{t-1,1},\ldots,x_{t-1,K_o},\ldots,x_{t-N,K_o})^T$, where $x_{t,k}$ is the kth output signal at time t. $\mathbf{E}_s(\Theta)$ is the $(N+1)K_o\times(M+N+1)K_i$ convolution matrix $\mathcal{T}(\mathbf{h})$ for SIMO channels [resp., $\mathcal{T}(\mathbf{H})$] for the MIMO channel. N and $M+1$ are the smoothing factor and the channel impulse response length, respectively, $p=(N+1)K_i$ (see, e.g., [8]). The components of Θ are the FIR coefficients of the channel. $\mathbf{s}_t\stackrel{\text{def}}{=}(s_{t,1},\ldots,s_{t,K_i},s_{t-1,1},\ldots,s_{t-M-N,K_i})^T$, where $s_{t,k}$ is the kth input signal at time t. $\mathbf{R}_s$ is the positive definite covariance matrix of the inputs (which is assumed diagonal for the linear prediction-based algorithms). $\mathbf{n}_t$ and $\mathbf{R}_n$ are, respectively, the complex envelope and the covariance matrix of the channel output additive noise. $\mathbf{R}_n=\sigma^2\mathbf{I}$ if the noise is spatially and temporally uncorrelated. This later assumption does not include jammers and supposes that the temporal correlation of noise due to the oversampling is not taken into account.

3) Sinusoidal Frequency Estimation for Mixed Spectra Times Series: $\mathbf{x}_t=(x_t,\ldots,x_{t-p+1})^T$ where in the complex case, x_t is a sum of sinusoid signals and a linear stationary process $n_t=\sum_{l=0}^\infty b_l u_{t-l}$ with $\sum_{l=0}^\infty|b_l|<\infty$. $\mathbf{E}_s(\Theta)=(\mathbf{e}_1,\ldots,\mathbf{e}_K)$ with $\mathbf{e}_k=\left(1,e^{i2\pi f_k},\ldots,e^{i2\pi(p-1)f_k}\right)^{\text{H}}$. Θ represents the K distinct frequencies in $]-1/2,+1/2[$. $\mathbf{s}_t=\left(a_1e^{i\phi_1}e^{i2\pi f_1t},\ldots,a_Ke^{i\phi_K}e^{i2\pi f_Kt}\right)^T$, where $(a_k)_{k=1,\ldots,K}$ are fixed positive real numbers, and $(\phi_k)_{k=1,\ldots,K}$ are random variables uniformly distributed on $[0,2\pi]$.

$\mathbf{n}_t = (n_t, \ldots, n_{t-p+1})^T$, $\mathbf{R}_n = \sigma^2 \mathbf{B}\mathbf{B}^{\mathrm{H}}$, where $\mathbf{B}$ is the $p \times \infty$ Toeplitz filtering matrix with first row $\mathbf{b}^T = (b_0, b_1, \ldots)$, and σ^2 is the power of the noise innovation. $\mathbf{B}$ is generally unknown, except in the whitening approach.

4) Frequency Estimation of Sinusoidal Signals With Very Lowpass Envelopes: The envelopes $a_{t,k}$ of sinusoids are stationary time series but slowly varying (w.r.t., f_k) and, therefore, are considered constant during the window of size p. The signal model is identical to the previous one with $\mathbf{s}_t = \left(a_{t,1}e^{i2\pi f_1 t}, \ldots, a_{t,K}e^{i2\pi f_K t}\right)^T$. Contrary to [9], which analyzes the degradation of performance induced by the aforementioned mismodeling, our paper is only devoted to the asymptotic covariance of the estimates.

III. Algorithms Under Study

A. Functional Approach

To consider the asymptotic performance of a second-order algorithm, we adopt a functional analysis that consists of recognizing that the whole process of constructing an estimate $\Theta(N)$ of Θ is equivalent to defining a functional relation linking this estimate $\Theta(N)$ to the statistics $\mathbf{R}_x(N) = (1/N)\sum_{t=1}^N \mathbf{x}_t\mathbf{x}_t^+$ from which it is inferred. This functional dependence is denoted $\Theta(N) = \mathrm{alg}(\mathbf{R}_x(N))$. By assumption, $\Theta = \mathrm{alg}(\mathbf{R}_x)$; therefore, the different algorithms alg(.) constitute distinct extensions of the mapping $\mathbf{R}_x \rightarrow \Theta$ generated by any unstructured real symmetric or Hermitian matrix $\mathbf{R}_x(N)$. In the following, we consider "regular" algorithms. We assume the regularity conditions stated below.

B. Regular Algorithms

1) The function alg(.) is differentiable in a neighborhood of $\mathbf{R}_x$, i.e., if $\mathbf{D}^{\mathrm{alg}}_{\Theta,\mathbf{R}_x}$[2] denotes the $q \times p^2$ matrix of this differential evaluated at point $\mathbf{R}_x$

$$\mathrm{alg}(\mathbf{R}_x + \delta\mathbf{R}) = \Theta + \mathbf{D}^{\mathrm{alg}}_{\Theta,\mathbf{R}_x}\mathrm{Vec}(\delta\mathbf{R}) + o(\delta\mathbf{R}). \tag{3.1}$$

2) For any $\Theta \in \boldsymbol{\Theta}$ and any covariance matrices $\mathbf{R}_s$ and $\mathbf{R}_n$ (structured[3] if the algorithm relies on this structure)

$$\mathrm{alg}\left(\mathbf{E}_s(\Theta)\mathbf{R}_s\mathbf{E}_s^+(\Theta) + \mathbf{R}_n\right) = \Theta. \tag{3.2}$$

These two requirements are met by most second-order algorithms, including the covariance matching estimation techniques [7]. We note that to fulfill requirement 1, the extension to $\mathbf{R}_x(N)$ of the mapping $\mathbf{R}_x \rightarrow \Theta$ sometimes needs regularization techniques (see, e.g., [10] for the linear prediction method in blind identification of FIR). Requirement 2 means that most second-order algorithms do not require the knowledge of covariance matrices $\mathbf{R}_s$ and $\mathbf{R}_n$. However, specified structures are sometimes needed. Some examples are given in Section II-B.

[2]Expressions of $\mathbf{D}^{\mathrm{alg}}_{\Theta,\mathbf{R}_x}$ are ordinarily deduced from perturbation calculus.

[3]Of course, the algorithm does not need to know $\mathbf{R}_s$ and $\mathbf{R}_n$.

C. Constraints Upon the Differential of the Algorithm

To specify the conditions under which the asymptotic distribution of the estimated parameter Θ is invariant with respect to the distribution and the temporal correlation of $\mathbf{s}_t$ and $\mathbf{n}_t$, we need to prove the following lemma.

Lemma 1: Under conditions 1 and 2, one has the following constraints on $\mathbf{D}^{\mathrm{alg}}_{\Theta,\mathbf{R}_x}$, according to the structure of the covariance matrices $\mathbf{R}_s$ and $\mathbf{R}_n$:

$$\mathbf{D}^{\mathrm{alg}}_{\Theta,\mathbf{R}_x}\left(\mathbf{E}_s(\Theta) \otimes \mathbf{E}_s(\Theta)\right) = \mathbf{O}, \text{ for } \mathbf{R}_s \text{ unstructured} \tag{3.3}$$

$$\mathbf{D}^{\mathrm{alg}}_{\Theta,\mathbf{R}_x}\left(\mathbf{e}_{s,k}(\Theta) \otimes \mathbf{e}_{s,k}(\Theta)\right) = \mathbf{0}$$
$$k = 1, \ldots, K, \text{ for } \mathbf{R}_s \text{ structured diagonal} \tag{3.4}$$

$$\mathbf{D}^{\mathrm{alg}}_{\Theta,\mathbf{R}_x}\mathrm{Vec}\left(\mathbf{E}_s(\Theta)\mathbf{E}_s^+(\Theta)\right) = \mathbf{0}$$
$$\text{for } \mathbf{R}_s \text{ structured proportional to the identity matrix} \tag{3.5}$$

$$\mathbf{D}^{\mathrm{alg}}_{\Theta,\mathbf{R}_x}\mathrm{Vec}(\mathbf{Q}_l) = \mathbf{0}, \quad l = 1, \ldots, L$$
$$\text{for } \mathbf{R}_n \text{ structured as a linear combination of } (\mathbf{Q}_l)_{l=1,\ldots,L} \tag{3.6}$$

with $\mathbf{E}_s(\Theta) = (\mathbf{e}_{s,1}(\Theta), \ldots, \mathbf{e}_{s,K}(\Theta))$.

Proof: The proof follows because for any perturbations $\delta\mathbf{R}_s$ unstructured, $\delta\mathbf{R}_s = \mathrm{Diag}\left(\delta\sigma_1^2, \ldots, \delta\sigma_K^2\right)$, $\delta\mathbf{R}_s = \delta\sigma_s^2\mathbf{I}_p$, or $\delta\mathbf{R}_n = \sum_{l=1}^L(\delta a_l)\mathbf{Q}_l$, the following equalities hold:

$$\begin{aligned}
&\mathrm{alg}\left(\mathbf{E}_s(\Theta)\left(\mathbf{R}_s + \delta\mathbf{R}_s\right)\mathbf{E}_s^+(\Theta) + \sum_{l=1}^L (a_l + \delta a_l)\mathbf{Q}_l\right)\\
&= \Theta\\
&= \Theta + \mathbf{D}^{\mathrm{alg}}_{\Theta,\mathbf{R}_x}\left(\mathrm{Vec}\left(\mathbf{E}_s(\Theta)\delta\mathbf{R}_s\mathbf{E}_s^+(\Theta)\right)\right.\\
&\quad\left. + \mathrm{Vec}\left(\sum_{l=1}^L \delta a_l\mathbf{Q}_l\right)\right) + o(\delta\mathbf{R}_s) + o(\delta\mathbf{a})\\
&= \Theta + \mathbf{D}^{\mathrm{alg}}_{\Theta,\mathbf{R}_x}\left(\left(\mathbf{E}_s(\Theta) \otimes \mathbf{E}_s(\Theta)\right)\mathrm{Vec}(\delta\mathbf{R}_s)\right.\\
&\quad\left. + \sum_{L=1}^L \delta a_l\mathrm{Vec}(\mathbf{Q}_l)\right) + o(\delta\mathbf{R}_s) + o(\delta\mathbf{a})
\end{aligned}$$

with $\delta\mathbf{a} \stackrel{\mathrm{def}}{=} (\delta a_1, \ldots, \delta a_L)^T$. When $\delta\mathbf{R}_s$ is diagonal, $\delta\mathbf{R}_s = \mathrm{Diag}\left(\delta\sigma_1^2, \ldots, \delta\sigma_K^2\right)$; thus, $\mathrm{Vec}\left(\mathbf{E}_s(\Theta)\delta\mathbf{R}_s\mathbf{E}_s^+(\Theta)\right) = \sum_{k=1}^K \delta\sigma_k^2\mathrm{Vec}\left(\mathbf{e}_{s,k}(\Theta)\mathbf{e}_{s,k}^+(\Theta)\right) = \sum_{k=1}^K \delta\sigma_k^2(\mathbf{e}_{s,k}(\Theta) \otimes \mathbf{e}_{s,k}(\Theta))$. When $\delta\mathbf{R}_s = \delta\sigma_s^2\mathbf{I}_p$, $\mathrm{Vec}\left(\mathbf{E}_s(\Theta)\ \delta\mathbf{R}_s\ \mathbf{E}_s^+(\Theta)\right) = \delta\sigma_s^2\mathrm{Vec}\left(\mathbf{E}_s(\Theta)\mathbf{E}_s^+(\Theta)\right)$. ■

Interpretation: We note that $\mathrm{Vec}\left(\mathbf{E}_s(\Theta)\mathbf{E}_s(\Theta)^+\right)$ is a linear combination of the vectors $\mathrm{Vec}\left(\mathbf{e}_{s,k}(\Theta) \otimes \mathbf{e}_{s,k}(\Theta)\right)$, $k = 1, \ldots, K$, the latter vectors being in the column space of $\mathbf{E}_s(\Theta) \otimes \mathbf{E}_s(\Theta)$. The larger the *a priori* knowledge about $\mathbf{R}_s$ is needed, the less severe the constraints (3.3)–(3.5) on $\mathbf{D}^{\mathrm{alg}}_{\Theta,\mathbf{R}_x}$ become. To derive the asymptotic distribution of second-order estimators, we need to know the asymptotic distribution of the sample covariance matrix $\mathbf{R}_x(N)$.

IV. Robustness of Parameter Estimates

A. Asymptotic Distribution of the Sample Covariance Matrix

For convenience, the definition of the complex Gaussian distribution is recalled. A complex random $p \times 1$ vector $\mathbf{y}$ has a zero-mean complex Gaussian distribution if the $2p$-joint distribution of the real and imaginary part of $\mathbf{y}$ is $2p$-zero-mean real Gaussian, i.e., for any complex $p \times 1$ vector $\mathbf{w}$; the real scalar $\mathbf{w}^{\mathrm{H}}\mathbf{y}+(\mathbf{w}^{\mathrm{H}}\mathbf{y})^{\mathrm{H}}$ has a zero-mean real Gaussian distribution with variance

$$2\mathbf{w}^{\mathrm{H}}\mathbf{\Sigma}_1\mathbf{w}+\mathbf{w}^{\mathrm{H}}\mathbf{\Sigma}_2\mathbf{w}^*+\mathbf{w}^T\mathbf{\Sigma}_2^*\mathbf{w}$$

where $\mathrm{E}\left(\mathbf{y}\mathbf{y}^{\mathrm{H}}\right) = bf\Sigma_1$, and $\mathrm{E}\left(\mathbf{y}\mathbf{y}^{T}\right) = \mathbf{\Sigma}_2$. This distribution, which is denoted $\mathcal{N}(\mathbf{0};\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)$, is specified by $p\times p$ positive definite matrix $\mathbf{\Sigma}_1$ and $p \times p$ symmetric matrix $\mathbf{\Sigma}_2$ and denoted $\mathcal{N}(\mathbf{0};\mathbf{\Sigma}_1,\mathbf{\Sigma}_2)$. For stationary processes $\mathbf{s}_t$ and $\mathbf{n}_t$ with finite fourth-order moments, the following theorem is proved.

Theorem 1: $\sqrt{N}\,(\mathrm{Vec}\,(\mathbf{R}_x(N))-\mathrm{Vec}(\mathbf{R}_x))$ converges in distribution to the zero-mean real [resp., complex] Gaussian distribution of covariance $\mathbf{C}_{R_x}$ [resp., $\mathbf{C}_{R_x}$, $\mathbf{C}_{R_x}\mathbf{K}$] in the real case [resp. in the complex case].

$$\sqrt{N}\,(\mathrm{Vec}(\mathbf{R}_x(N))-\mathrm{Vec}(\mathbf{R}_x)) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0};\mathbf{C}_{R_x}) \quad [\text{resp.}, \mathcal{N}(\mathbf{0};\mathbf{C}_{R_x},\mathbf{C}_{R_x}\mathbf{K})]. \tag{4.1}$$

Furthermore

$$\lim_{N\to\infty} N\mathrm{Cov}\,(\mathrm{Vec}(\mathbf{R}_x(N))) = \mathbf{C}_{R_x} \tag{4.2}$$

where $\mathbf{C}_{R_x}$ reads

$$\begin{aligned}\mathbf{C}_{R_x} =& (\mathbf{E}_s(\Theta)\otimes\mathbf{E}_s(\Theta))\,\mathbf{C}_{R_s}\left(\mathbf{E}_s^+(\Theta)\otimes\mathbf{E}_s^+(\Theta)\right)+\mathbf{C}_{R_n}\\ &+(\mathbf{E}_s(\Theta)\otimes\mathbf{I}_p)\,\mathbf{C}_{R_{s,n}}\left(\mathbf{E}_s^+(\Theta)\otimes\mathbf{I}_p\right)\\ &+(\mathbf{I}_p\otimes\mathbf{E}_s(\Theta))\,\mathbf{C}_{R_{n,s}}\left(\mathbf{I}_p\otimes\mathbf{E}_s^+(\Theta)\right)\end{aligned} \tag{4.3}$$

with the equation shown at the bottom of the page, with $i,j=s,n$ or n,s, where $\mathbf{R}_s(N)\stackrel{\text{def}}{=}(1/N)\sum_{t=1}^N \mathbf{s}_t\mathbf{s}_t^+$, $\mathbf{R}_n(N) \stackrel{\text{def}}{=}(1/N)\sum_{t=1}^N \mathbf{n}_t\mathbf{n}_t^+$, $\mathbf{R}_s^{t-t'}\stackrel{\text{def}}{=}\mathrm{E}\left(\mathbf{s}_t\mathbf{s}_{t'}^+\right)$, $\mathbf{R}_n^{t-t'}\stackrel{\text{def}}{=}\mathrm{E}\left(\mathbf{n}_t\mathbf{n}_{t'}^+\right)$, $\mathbf{R}_s'^{\,t-t'}\stackrel{\text{def}}{=}\mathrm{E}\left(\mathbf{s}_t\mathbf{s}_{t'}^T\right)$ and $\mathbf{R}_n'^{\,t-t'}\stackrel{\text{def}}{=}\mathrm{E}\left(\mathbf{n}_t\mathbf{n}_{t'}^T\right)$. In the complex case, $\mathrm{Cov}\,(\mathrm{Vec}(\mathbf{R}_x(N)))$ denotes $\mathrm{E}(\mathrm{Vec}\left(\mathbf{R}_x(N)-\mathbf{R}_x\right)\mathrm{Vec}^{\mathrm{H}}(\mathbf{R}_x(N)-\mathbf{R}_x))$. We note that $\mathrm{Vec}^T(\mathbf{R}_x(N)-\mathbf{R}_x) = \mathrm{Vec}^T\left(\mathbf{R}_x^{\mathrm{H}}(N)-\mathbf{R}_x^{\mathrm{H}}\right) = \mathrm{Vec}^{\mathrm{H}}\left(\mathbf{R}_x^T(N)-\mathbf{R}_x^T\right) = \mathrm{Vec}^{\mathrm{H}}\left(\mathbf{R}_x(N)-\mathbf{R}_x\right)\mathbf{K}$, and therefore, $\mathrm{E}\left(\mathrm{Vec}(\mathbf{R}_x(N)-\mathbf{R}_x)\mathrm{Vec}^T(\mathbf{R}_x(N)-\mathbf{R}_x)\right) = \mathrm{E}\left(\mathrm{Vec}(\mathbf{R}_x(N)-\mathbf{R}_x)\mathrm{Vec}^{\mathrm{H}}(\mathbf{R}_x(N)-\mathbf{R}_x)\right)\mathbf{K}$. Therefore, the noncircular complex Gaussian asymptotic distribution of $\mathbf{R}_x(N)$ is characterized by $\mathbf{C}_{R_x}$ only.

Proof: In the example of application 3, where x_t is a sum of sinusoid signals and an MA process, this theorem is proved in [11]. The generalization to the data model (2.1) follows the same lines. First, (4.2) is straightforwardly proved after tedious but simple manipulations. Then, to prove (4.1), we adapt the steps of [12, sect. 7.3] to each model. ■

Remark 1: This theorem extends theorems following the classic stochastic model assumption (see, e.g., [1] and [2]) to accommodate non-Gaussian and temporally correlated noise. For Gaussian temporally uncorrelated noise, $\mathbf{C}_{R_n}$ and the cross-terms $\mathbf{C}_{R_{s,n}}$ and $\mathbf{C}_{R_{n,s}}$ reduce to $\mathbf{R}_n\otimes\mathbf{R}_n$ for circular complex case [resp. $(\mathbf{R}_n\otimes\mathbf{R}_n)(\mathbf{I}_p+\mathbf{K})$ for real case], $\mathbf{R}_s\otimes\mathbf{R}_n$ and $\mathbf{R}_n\otimes\mathbf{R}_s$, respectively. The non-Gaussian assumption simply adds a fourth-order term in the expression of $\mathbf{C}_{R_n}$, but the temporal correlation assumption completely modifies the expression of $\mathbf{C}_{R_n}$, $\mathbf{C}_{R_{s,n}}$ and $\mathbf{C}_{R_{n,s}}$. When the noise is possibly non-Gaussian and temporally correlated, $\mathbf{C}_{R_n}$ becomes in the circular complex ARMA case (see [4])

$$\mathbf{C}_{R_n} = \int_{-1/2}^{+1/2}\mathbf{S}_n(f)\otimes\mathbf{S}_n(f)df+\mathbf{Q}_n \tag{4.4}$$

where $\mathbf{S}_n(f)$ denotes the power cross-spectral density $p\times p$ matrix of $\mathbf{n}_t$. If the fourth-order polyspectrum of the components $(n_{t,k})_{k=1,\ldots,p}$ of $\mathbf{n}_t$ for $k_1,k_2,k_3,k_4=1,\ldots,p$ is defined as

$$\begin{aligned}&\rho_{k_1,k_2,k_3,k_4}(f,f',f'')\\ &\quad\stackrel{\text{def}}{=}\sum_{\tau,\tau',\tau''}\mathrm{Cum}\left(n_{0,k_1},n_{\tau,k_2}^*,n_{\tau',k_3},n_{\tau'',k_4}^*\right)\\ &\quad\times e^{i2\pi\left(f\tau+f'\tau'+f''\tau''\right)},\\ &[\mathbf{Q}_n]_{p(j-1)+i,p(l-1)+k}\\ &\quad=\int_{-1/2}^{+1/2}\int_{-1/2}^{+1/2}\rho_{i,j,l,k}(f,f',-f')dfdf'\end{aligned}$$

denotes the $p^2\times p^2$ fourth-order cumulant matrix. Simplified formulas of the expressions of the cross-terms $\mathbf{C}_{R_{s,n}}$ and $\mathbf{C}_{R_{s,n}}$ given in Theorem 1 can be obtained if the sequence $(\mathbf{R}_s^T\otimes\mathbf{R}_n^T)_{t=\cdots,-1,0,+1,\ldots}$ is assumed absolutely summable.[4] In this

[4]This condition is satisfied if the sequences $\mathbf{R}_s^t$ and $\mathbf{R}_n^t$ are absolutely summable or if one of the two sequences is absolutely summable and the other bounded.

$$\begin{aligned}\mathbf{C}_{R_s} &= \lim_{N\to\infty} N\mathrm{Cov}\,(\mathrm{Vec}(\mathbf{R}_s(N)))\\ \mathbf{C}_{R_n} &= \lim_{N\to\infty} N\mathrm{Cov}\,(\mathrm{Vec}(\mathbf{R}_n(N)))\\ \mathbf{C}_{R_{i,j}} &= \begin{cases}\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^N\sum_{t'=1}^N\mathbf{R}_i^{t-t'}\otimes\mathbf{R}_j^{t-t'}, & \text{in the circular complex case}\\ \lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^N\sum_{t'=1}^N\mathbf{R}_i^{t-t'}\otimes\mathbf{R}_j^{t-t'} & \\ \quad+\lim_{N\to\infty}\frac{1}{N}\sum_{t=1}^N\sum_{t'=1}^N\left(\mathbf{R}_i'^{\,t-t'}\otimes\mathbf{R}_j'^{\,t-t'}\right)\mathbf{K}, & \text{in the noncircular complex and real case}\end{cases}\end{aligned}$$

case, we get from [13, A10, p. 411], and from Parseval's theorem, we get (4.5), shown at the bottom of the page.

Remark 2: Detailed expressions of $\mathbf{C}_{R_s}$, $\mathbf{C}_{R_n}$, $\mathbf{C}_{R_{s,n}}$, and $\mathbf{C}_{R_{n,s}}$ depend on the application. For example

$$\begin{aligned}\mathbf{C}_{R_n} =& \int_{-1/2}^{+1/2} S_n^2(f)\left[\mathbf{e}_f\mathbf{e}_f^{\mathrm{H}} \otimes \mathbf{e}_f\mathbf{e}_f^{\mathrm{H}}\right] df \\ &+ \kappa_u \mathrm{Vec}\left(\mathbf{B}\mathbf{B}^{\mathrm{H}}\right)\mathrm{Vec}^{\mathrm{H}}\left(\mathbf{B}\mathbf{B}^{\mathrm{H}}\right) \\ \mathbf{C}_{R_{n,s}} =& \mathrm{Diag}\left(|a_1|^2\mathbf{e}_1\mathbf{e}_1^{\mathrm{H}}S_n(f_1),\ldots,|a_K|^2\mathbf{e}_K\mathbf{e}_K^{\mathrm{H}}S_n(f_K)\right)\end{aligned} \tag{4.6}$$

and $\mathbf{C}_{R_{s,n}}$ is the block matrix whose (i,j) block element is

$$\left[\mathbf{C}_{R_{s,n}}\right]_{i,j} = \mathrm{Diag}\Big(|a_1|^2\left[\mathbf{e}_1\mathbf{e}_1^{\mathrm{H}}\right]_{i,j} S_n(f_1),\ldots \\ |a_K|^2\left[\mathbf{e}_K\mathbf{e}_K^{\mathrm{H}}\right]_{i,j} S_n(f_K)\Big)$$

for sinusoidal frequency estimation with circular complex additive noise n_t of spectral density $S_n(f)$, $\kappa_u \stackrel{\mathrm{def}}{=} \mathrm{Cum}(u_t,u_t^*,u_t,u_t^*)$ and $\mathbf{e}_f \stackrel{\mathrm{def}}{=} \left(1,e^{i2\pi f},\ldots,e^{i2(p-1)\pi f}\right)^{\mathrm{H}}$. A similar expression is obtained for $\mathbf{C}_{R_s}$ for blind identification of SIMO FIR channels when the input s_t is circular complex temporally uncorrelated of power σ_s^2 with $\kappa_s \stackrel{\mathrm{def}}{=} \mathrm{Cum}\left(s_t,s_t^*,s_t,s_t^*\right)$ and $\mathbf{e}_f \stackrel{\mathrm{def}}{=} \left(1,e^{i2\pi f},\ldots,e^{i2L\pi f}\right)^{\mathrm{H}}$:

$$\mathbf{C}_{R_s} = \int_{-1/2}^{+1/2} \sigma_s^4\left[\mathbf{e}_f\mathbf{e}_f^{\mathrm{H}} \otimes \mathbf{e}_f\mathbf{e}_f^{\mathrm{H}}\right] df + \kappa_s \mathrm{Vec}(\mathbf{I}_p)\mathrm{Vec}^{\mathrm{H}}(\mathbf{I}_p). \tag{4.7}$$

We note that instead of the classic Bartlett formulation, which is concerned with the sample correlation coefficients sequence, Theorem 1 is devoted to the sample covariance matrix. This formulation is better adapted to deriving the asymptotic distribution of estimated parameters as is derived in Section IV-B.

B. Asymptotic Distribution of the Estimated Parameter

By the regularity condition (3.1), the asymptotic behaviors of $\Theta(N)$ and $\mathbf{R}_x(N)$ are directly related. The standard theorem on regular functions of asymptotically normal statistics (see e.g., [14, p. 122]) applies.

Theorem 2:

$$\sqrt{N}\left(\Theta(N)-\Theta\right) \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0};\mathbf{C}_\Theta) \text{ for } \Theta \in \mathcal{R}^q \\ [\text{resp., } \mathcal{N}(\mathbf{0};\mathbf{C}_\Theta,\mathbf{C}'_\Theta) \text{ for } \Theta \in \mathcal{C}^q] \tag{4.8}$$

with

$$\begin{aligned}\mathbf{C}_\Theta =& \lim_{N\to\infty} N\mathrm{E}\left((\Theta(N)-\Theta)(\Theta(N)-\Theta)^+\right) \\ =& \mathbf{D}_{\Theta,\mathbf{R}_x}^{\mathrm{alg}}\mathbf{C}_{R_x}\left(\mathbf{D}_{\Theta,\mathbf{R}_x}^{\mathrm{alg}}\right)^+\end{aligned} \tag{4.9}$$

$$\begin{aligned}\mathbf{C}'_\Theta =& \lim_{N\to\infty} N\mathrm{E}\left((\Theta(N)-\Theta)(\Theta(N)-\Theta)^T\right) \\ =& \mathbf{D}_{\Theta,\mathbf{R}_x}^{\mathrm{alg}}\mathbf{C}_{R_x}\mathbf{K}\left(\mathbf{D}_{\Theta,\mathbf{R}_x}^{\mathrm{alg}}\right)^T.\end{aligned} \tag{4.10}$$

Consequently, for second-order algorithms satisfying the regularity conditions of Section III-B, the expressions of the asymptotic covariances $\mathbf{C}_\Theta$ and $\mathbf{C}'_\Theta$ of estimators can be simplified thanks to the constraints on $\mathbf{D}_{\Theta,\mathbf{R}_x}^{\mathrm{alg}}$ of Lemma 1 and become $\mathbf{C}_\Theta = \mathbf{D}_{\Theta,\mathbf{R}_x}^{\mathrm{alg}}\mathcal{C}_{R_x}\left(\mathbf{D}_{\Theta,\mathbf{R}_x}^{\mathrm{alg}}\right)^+$, $\mathbf{C}'_\Theta = \mathbf{D}_{\Theta,\mathbf{R}_x}^{\mathrm{alg}}\mathcal{C}_{R_x}\mathbf{K}\left(\mathbf{D}_{\Theta,\mathbf{R}_x}^{\mathrm{alg}}\right)^T$, where $\mathcal{C}_{R_x}$ is deduced from the expression of $\mathbf{C}_{R_x}$ (4.3) by suppressing some terms. This result is specialized to the examples described in Section II-B and implies that these covariance matrices are invariant to the distribution and/or to the temporal correlation of $\mathbf{s}_t$ and $\mathbf{n}_t$, depending on the application. This admits the following interpretation: The larger the *a priori* knowledge about $\mathbf{R}_s$ and $\mathbf{R}_n$ is needed, the less severe the constraints on $\mathbf{D}_{\Theta,\mathbf{R}_x}^{\mathrm{alg}}$ are, and the less robust to the distribution of the signals the second-order estimators become.

C. Examples of Applications

1) Narrowband DOA Estimation: Depending on whether $\mathbf{n}_t$ is assumed temporally uncorrelated (an assumption admitted in all papers devoted to performance analysis) or correlated, the following results hold.

Result 1: If $\mathbf{n}_t$ is assumed temporally uncorrelated, the algorithms that do not suppose the sources to be spatially uncorrelated are robust to the distribution and to the temporal correlation of the sources $\mathbf{s}_t$.

Proof: Thanks to the first constraint (3.3), $\mathcal{C}_{R_x}$ is deduced from the expression of $\mathbf{C}_{R_x}$ by suppression of its first term. Furthermore, because the terms $\mathbf{C}_{R_{s,n}}$ and $\mathbf{C}_{R_{n,s}}$ of $\mathbf{C}_{R_x}$ reduce, respectively, to the spatial terms $\mathbf{R}_s \otimes \mathbf{R}_n$ and $\mathbf{R}_n \otimes \mathbf{R}_s$, $\mathcal{C}_{R_x}$ reduces to

$$\mathbf{C}_{R_n} + (\mathbf{E}_s(\Theta)\otimes\mathbf{I}_p)(\mathbf{R}_s\otimes\mathbf{R}_n)\left(\mathbf{E}_s^+(\Theta)\otimes\mathbf{I}_p\right) \\ + (\mathbf{I}_p\otimes\mathbf{E}_s(\Theta))(\mathbf{R}_n\otimes\mathbf{R}_s)\left(\mathbf{I}_p\otimes\mathbf{E}_s^+(\Theta)\right).$$

∎

This extends the results by Cardoso and Moulines [3] that have shown that the asymptotic performance of most high-resolution covariance-based DOA estimators is independent of the distribution of the source signals for independent snapshots. It is shown [4] that the Toeplization and augmentation techniques, which are based on the source spatial uncorrelation assumption, are very sensitive to the distribution and to the temporal correlation of the sources in the case of several sources because the

$$\mathbf{C}_{R_{s,n}} = \begin{cases} \sum_{t=-\infty}^{+\infty}\left(\mathbf{R}_s^t\otimes\mathbf{R}_n^t\right)\left(\mathbf{I}_{p^2}+\mathbf{K}\right) = \left(\int_{-1/2}^{+1/2}\mathbf{S}_n(f)\otimes\mathbf{S}_s(f)df\right)\left(\mathbf{I}_{p^2}+\mathbf{K}\right), & \text{in the real case} \\ \sum_{t=-\infty}^{+\infty}\mathbf{R}_s^t\otimes\mathbf{R}_n^t = \int_{-1/2}^{+1/2}\mathbf{S}_s(f)\otimes\mathbf{S}_n(f)df, & \text{in the complex case.}\end{cases} \tag{4.5}$$

constraint (3.3) is not satisfied. For only one source, the robustness is preserved thanks to constraint (3.4).

Result 2: If $\mathbf{n}_t$ is assumed temporally correlated, all the second-order algorithms are sensitive to the temporal correlation of the sources.

Proof: This is due to the contribution of terms $\mathbf{C}_{R_{s,n}}$ and $\mathbf{C}_{R_{n,s}}$ [see (4.5)] in $\mathbf{C}_{R_x}$. ■

A realistic example of this situation is given in Section V-A.

2) Blind Identification of FIR Channels: From the general methodological viewpoint, the second-order algorithms may be classified as methods that do not suppose that the inputs $s_{t,k}$ are temporally uncorrelated (e.g., the subspace methods which exploit low-rank space–time properties; see, e.g., [5] and references therein) and methods that explicitly suppose that the inputs $s_{t,k}$ are temporally uncorrelated (e.g., the linear prediction methods, using specific invertibility properties of FIR models; see, e.g., [10] and references therein).

Result 3: The blind SIMO identification methods that do not suppose that the inputs $s_{t,k}$ are temporally uncorrelated are robust to the distribution of the inputs but sensitive to the temporal correlation of the inputs.

Proof: Thanks to the first constraint (3.3), $\mathcal{C}_{R_x}$ is deduced from the expression of $\mathbf{C}_{R_x}$ by suppression of its first term, which is the only term that depends on the fourth-order properties of the inputs $s_{t,k}$ by way of $\mathbf{C}_{R_s}$. Therefore, $\mathcal{C}_{R_x}$ reduces to

$$\mathbf{C}_{R_n} + (\mathbf{E}_s(\Theta)\otimes\mathbf{I}_p)\,\mathbf{C}_{R_{s,n}}\left(\mathbf{E}_s^{+}(\Theta)\otimes\mathbf{I}_p\right) + (\mathbf{I}_p\otimes\mathbf{E}_s(\Theta))\,\mathbf{C}_{R_{n,s}}\left(\mathbf{I}_p\otimes\mathbf{E}_s^{+}(\Theta)\right).$$

These methods are sensitive to the temporal correlation of the inputs because the terms $\mathbf{C}_{R_{s,n}}$ and $\mathbf{C}_{R_{n,s}}$ depend on the temporal correlation of the inputs, including the case where the noise is temporally uncorrelated because in this case, $\mathbf{C}_{R_{s,n}} = \mathbf{R}_s\otimes\mathbf{R}_n$ and $\mathbf{C}_{R_{n,s}} = \mathbf{R}_n\otimes\mathbf{R}_s$ [see (4.5), where $\mathbf{R}_n^t = 1_{t=0}\mathbf{R}_n$], where $\mathbf{R}_s$ includes temporal correlation in this space–time application. ■

We note that this result does not extend to blind MIMO identification methods. In this case, time series $(s_{t,k})_{k=1,\dots,K}$ are assumed independent, and consequently, $\mathbf{R}_s$ is structured block diagonal, and the first constraint (3.3) no longer applies.

Result 4: The blind SIMO identification methods that explicitly suppose that the inputs $s_{t,k}$ are temporally uncorrelated are robust to the distribution of the inputs.

Proof: Thanks to the third constraint (3.5) applied to the expression (4.7) of $\mathbf{C}_{R_s}$ and thanks to the equality

$$\begin{aligned}\left[\mathbf{E}_s(\Theta)\otimes\mathbf{E}_s^{+}(\Theta)\right]\mathrm{Vec}(\mathbf{I}) &= \mathrm{Vec}\left[\mathbf{E}_s(\Theta)\mathbf{I}\mathbf{E}_s^{+}(\Theta)\right]\\ &= \mathrm{Vec}\left[\mathbf{E}_s(\Theta)\mathbf{E}_s^{+}(\Theta)\right]\end{aligned}$$

the contribution of the cumulant κ_s of the input signal is canceled in the expression of $\mathcal{C}_{R_x}$, which reduces to

$$\begin{aligned}&(\mathbf{E}_s(\Theta)\otimes\mathbf{E}_s(\Theta))\left(\int_{-1/2}^{+1/2}\sigma_s^4[\mathbf{e}_f\mathbf{e}_f^{\mathrm{H}}\otimes\mathbf{e}_f\mathbf{e}_f^{\mathrm{H}}]df\right)\\ &\times\left(\mathbf{E}_s^{+}(\Theta)\otimes\mathbf{E}_s^{+}(\Theta)\right)+\mathbf{C}_{R_n}\\ &+(\mathbf{E}_s(\Theta)\otimes\mathbf{I}_p)\,\mathbf{C}_{R_{s,n}}\left(\mathbf{E}_s^{+}(\Theta)\otimes\mathbf{I}_p\right)\\ &+(\mathbf{I}_p\otimes\mathbf{E}_s(\Theta))\,\mathbf{C}_{R_{n,s}}\left(\mathbf{I}_p\otimes\mathbf{E}_s^{+}(\Theta)\right).\end{aligned}$$ ■

We note that this result does not extend to blind MIMO identification methods because in this case, the fourth-order term of $\mathbf{C}_{R_s}$ is no longer structured in the form $\kappa_s\mathrm{Vec}(\mathbf{I}_p)\mathrm{Vec}^{\mathrm{H}}(\mathbf{I}_p)$.

This robustness property extends to any second-order algorithm, where the robustness result is proved by [5] for some subspace methods after calculating $\mathbf{D}_{\Theta,\mathbf{R}_x}^{\mathrm{MUSIC}}$. We note that proving this robustness property directly from the expression of $\mathbf{D}_{\Theta,\mathbf{R}_x}^{\mathrm{alg}}$ for each specific algorithm would be tedious and cumbersome; see, e.g., the intricate expression of $\mathbf{D}_{\Theta,\mathbf{R}_x}^{\mathrm{LP}}$ in [10].

Furthermore, as a byproduct of our results, we note that the asymptotic covariance $\mathbf{C}_{\Theta}$ has the same expression under the assumptions "$\mathbf{s}_t$ is Gaussian i.i.d." and "s_t are temporally uncorrelated with any distribution" for all second-order algorithms that suppose s_t temporally uncorrelated. Therefore, our result validates the asymptotic performance and limitation results of Zeng and Tong [6] (which were based on the i.i.d. Gaussian assumption of $\mathbf{s}_t$) for an input s_t temporally uncorrelated with any distribution.

3) Sinusoidal Frequency Estimation for Mixed Spectra Times Series:

Result 5: The second-order algorithms are robust to the distribution of the noise $\mathbf{n}_t$ but sensitive to its temporal correlation.

Proof: Thanks to the fourth constraint (3.6) (where $\mathbf{Q}_1 = \mathbf{B}\mathbf{B}^{\mathrm{H}}$) applied to the expression (4.6) of $\mathbf{C}_{R_n}$, the contribution of the cumulant κ_u of the noise innovation is canceled in the expression of $\mathcal{C}_{R_x}$, which reduces to

$$\begin{aligned}&(\mathbf{E}_s(\Theta)\otimes\mathbf{E}_s(\Theta))\,\mathbf{C}_{R_s}\left(\mathbf{E}_s^{+}(\Theta)\otimes\mathbf{E}_s^{+}(\Theta)\right)\\ &+\int_{-1/2}^{+1/2}S_n^2(f)\left[\mathbf{e}_f\mathbf{e}_f^{\mathrm{H}}\otimes\mathbf{e}_f\mathbf{e}_f^{\mathrm{H}}\right]df\\ &+(\mathbf{E}_s(\Theta)\otimes\mathbf{I}_p)\,\mathbf{C}_{R_{s,n}}\left(\mathbf{E}_s^{+}(\Theta)\otimes\mathbf{I}_p\right)\\ &+(\mathbf{I}_p\otimes\mathbf{E}_s(\Theta))\,\mathbf{C}_{R_{n,s}}\left(\mathbf{I}_p\otimes\mathbf{E}_s^{+}(\Theta)\right).\end{aligned}$$ ■

This result apparently contradicts a Monte Carlo simulation recently presented in [15] in which the frequency estimators degrade with an heavy-tailed probability distribution of the noise. In fact, this simulation is presented with a complex circular symmetric α-stable distribution of the noise with $\alpha = 1$, for which neither $\mathrm{E}(n_t)$ nor $\mathrm{E}|n_t|^2$ are defined, and our analysis is devoted to second-order processes only. Fig. 1 illustrates the asymptotic performance of the MUSIC algorithm for two equipowered sinusoids with two complex circular distributions of the noise [Gaussian (a) and with the heavy-tailed probability distribution of normalized p.d.f. $2/\left(\pi(1+x^2)^2\right)$ (b)]. Fig. 1 shows the similarity of the behavior of this algorithm with these two distributions. We notice good agreement between the theoretical and the estimated MSE with a domain of validity reducing with p increasing ($N \approx 100$, 500, and 3000 for, respectively, $p = 4$, 6, and 12). This remark extends to non-Gaussian noise, as observed in [16] (where, e.g., for $p = 8$ and $SNR = 20$ dB, a good agreement requires $N = 10\,000$).

4) Frequency Estimation of Sinusoidal Signals With Very Lowpass Envelopes:

Result 6: The second-order algorithms are robust to the distribution of the envelopes $a_{t,k}$ of the sinusoidal signals and of the noise $\mathbf{n}_t$ but sensitive to their temporal correlation.

Proof: Thanks to the first (3.4) and fourth (3.6) constraint (where $\mathbf{Q}_1 = \mathbf{B}\mathbf{B}^{\mathrm{H}}$) applied to the first term of $\mathbf{C}_{R_x}$ (4.3) and

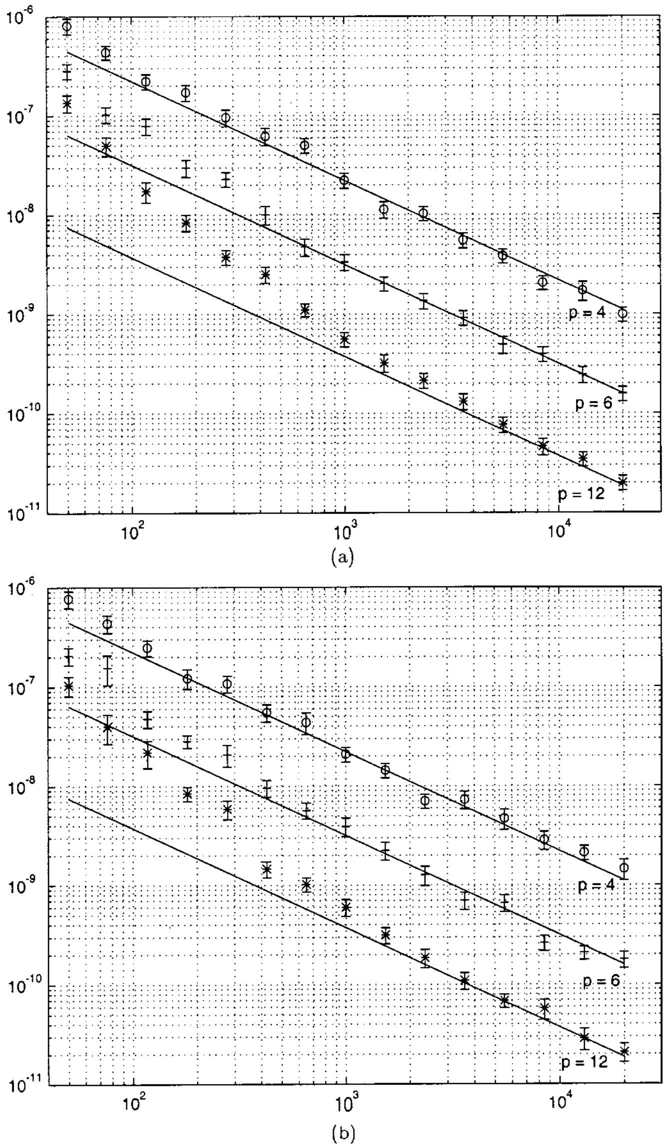

Fig. 1. Theoretical and estimated (100 runs) MSE of $f_1(N)$ with 95% confidence interval (with error bars) by the MUSIC algorithm for two equipowered sinusoids and white noise ($f_1 = 0.1$, $f_2 = 0.2$ and $SNR = 10$ dB) for $p = 4$, 6, and 12 versus N when (a) the noise n_t is circular Gaussian distributed and (b) circularly distributed with the normalized p.d.f. $2/(\pi(1+x^2)^2)$.

to the expression (4.6) of $\mathbf{C}_{R_n}$, respectively, the term $\mathbf{C}_{R_s}$ and the contribution of the cumulant κ_u of the noise innovation is canceled in the expression of $\mathcal{C}_{R_x}$, which reduces to

$$\int_{-1/2}^{+1/2} S_n^2(f)\left[\mathbf{e}_f\mathbf{e}_f^{\mathrm{H}} \otimes \mathbf{e}_f\mathbf{e}_f^{\mathrm{H}}\right] df + (\mathbf{E}_s(\Theta)\otimes\mathbf{I}_p)\,\mathbf{C}_{R_{s,n}} \times \left(\mathbf{E}_s^{+}(\Theta)\otimes\mathbf{I}_p\right) + (\mathbf{I}_p\otimes\mathbf{E}_s(\Theta))\,\mathbf{C}_{R_{n,s}}\left(\mathbf{I}_p\otimes\mathbf{E}_s^{+}(\Theta)\right).$$

■

V. Further Illustrations

A. Temporally Correlated Noise

One would expect that the temporal correlation of the noise would modify the asymptotic performance of the covariance-based frequency estimation of sinusoidal signals because the asymptotic Cramér–Rao bounds of the estimated frequencies are inversely proportional to the local signal-to-noise ratio $a_k^2/S_n(f_k)$ [17], but in the case of the DOA and FIR parameters, the asymptotic robustness property (see, e.g., [2]) proved in the temporally white noise case is questioned. To show the influence of this noise temporal correlation, we concentrate on the circular complex narrowband DOA estimation example. In this case, thanks to result 1 and (4.4) and (4.5), the asymptotic covariance of the parameter estimates reduces to

$$\begin{aligned}\mathbf{C}_\Theta = &\mathbf{D}_{\Theta,\mathbf{R}_x}^{\mathrm{alg}}\left(\int_{-1/2}^{+1/2}\mathbf{S}_n(f)\otimes\mathbf{S}_n(f)df + \mathbf{Q}_n\right.\\ &+ (\mathbf{E}_s(\Theta)\otimes\mathbf{I}_p)\left(\int_{-1/2}^{+1/2}\mathbf{S}_s(f)\otimes\mathbf{S}_n(f)df\right)\\ &\times\left(\mathbf{E}_s^{\mathrm{H}}(\Theta)\otimes\mathbf{I}_p\right) + (\mathbf{I}_p\otimes\mathbf{E}_s(\Theta))\\ &\times\left(\int_{-1/2}^{+1/2}\mathbf{S}_n(f)\otimes\mathbf{S}_s(f)df\right)\\ &\left.\times\left(\mathbf{I}_p\otimes\mathbf{E}_s^{\mathrm{H}}(\Theta)\right)\right)\left(\mathbf{D}_{\Theta,\mathbf{R}_x}^{\mathrm{alg}}\right)^{\mathrm{H}}.\end{aligned}\tag{5.1}$$

Usually, performance analyses are evaluated as a function of the number of observed snapshots without taking the sampling rate into account. In fact, depending on the value of this sampling rate, the collected samples are more or less temporally correlated, and performance is affected. Thus, the interesting question arises as to how the asymptotic covariance of the parameter estimators varies with this sampling rate $1/T_s$ for a fixed observation interval T. This will be investigated by considering the preprocessing operation. The received signals are bandpass filtered (with bandwidth B) around the center frequency of interest. After frequency down-shifting the sensor signals to baseband, the complex envelope is generated. If the background noise is white, the continuous-time noise envelope $\mathbf{n}_t$ is white in the bandwidth $[-B/2, +B/2]$ with a power spectral density N_0. $\mathbf{n}_t$ is circular complex and assumed Gaussian and spatially uncorrelated. The cross-power spectral density matrix of the continuous-time source envelope is denoted $\mathbf{S}_s^c(f)$ and lies in $[-B/2, +B/2]$. Under these conditions, after sampling the complex envelope signals at the rate $1/T_s$, the power spectra of the discrete-time signals in the bandwidth $[-B/2, +B/2]$ become

$$\begin{aligned}\mathbf{S}_s(f) = &\frac{1}{T_s}\sum_{k=-\infty}^{+\infty}\mathbf{S}_s^c\left(f-\frac{k}{T_s}\right)\\ \mathbf{S}_n(f) = &\frac{1}{T_s}\sum_{k=-\infty}^{+\infty}\mathrm{Diag}\Big(N_0 1_{[-\frac{B}{2}-\frac{k}{T_s},\frac{B}{2}-\frac{k}{T_s}]}(f),\ldots\\ &N_0 1_{[-\frac{B}{2}-\frac{k}{T_s},\frac{B}{2}-\frac{k}{T_s}]}(f)\Big).\end{aligned}$$

Therefore, according to whether the signals are oversampled or subsampled with respect to the Nyquist frequency, the integrals of (5.11) become the following.

- If $1/T_s > B$

$$\begin{aligned}\mathbf{C}_{R_n} =& T_s \int_{-1/2T_s}^{+1/2T_s} \mathbf{S}_n^c(f) \otimes \mathbf{S}_n^c(f) df \\ =& \text{Diag}\left(\frac{N_0^2 B}{T_s}, \dots, \frac{N_0^2 B}{T_s}\right) = \frac{\sigma_n^4}{BT_s}\mathbf{I}_{p^2} \\ \mathbf{C}_{R_{s,n}} =& T_s \int_{-1/2T_s}^{+1/2T_s} \mathbf{S}_s^c(f) \otimes \mathbf{S}_n^c(f) df \\ =& \frac{1}{T_s}\int_{-B/2}^{+B/2} \mathbf{S}_s^c(f) df \otimes \text{Diag}(N_0, \dots, N_0) \\ =& \frac{1}{BT_s}\mathbf{R}_s \otimes \sigma_n^2 \mathbf{I}_p\end{aligned}$$

  where $\sigma_n^2 = N_0 B$ denotes the noise power.
- If $1/T_s < B$, the separate terms in the previous integrals overlap and, according to the value of T_s, $\mathbf{C}_{R_n}$, and $\mathbf{C}_{R_{s,n}}$, fluctuate around their limit value when $T_s \to \infty$, viz.

$$\begin{aligned}\mathbf{C}_{R_n} =& \mathbf{R}_n \otimes \mathbf{R}_n = \sigma_n^4 \mathbf{I}_{p^2} \\ \mathbf{C}_{R_{s,n}} =& \mathbf{R}_s \otimes \mathbf{R}_n = \mathbf{R}_s \otimes \sigma_n^2 \mathbf{I}_p.\end{aligned}$$

  These values are obtained when the successive snapshots are assumed independent.

The asymptotic error covariance matrix of the parameter Θ is now considered as a function of the observation interval $T = NT_s$. The previous values of $\mathbf{C}_{R_n}$ and $\mathbf{C}_{R_{s,n}}$ show that if the signals are oversampled

$$\text{E}\left((\Theta(T) - \Theta)(\Theta(T) - \Theta)^T\right) \sim \frac{1}{BT}\mathbf{C}_\Theta > \frac{1}{N}\mathbf{C}_\Theta \quad \text{for } N \gg 1$$

irrespective of the sample rate $1/T_s$, and if the signals are subsampled

$$\begin{aligned}\text{E}\left((\Theta(T) - \Theta)(\Theta(T) - \Theta)^T\right) \sim& \frac{T_s}{T}\mathbf{C}_\Theta = \frac{1}{N}\mathbf{C}_\Theta \\ &> \frac{1}{BT}\mathbf{C}_\Theta \\ &\text{for } N \gg 1 \text{ and } BT_s \gg 1\end{aligned}$$

where $\mathbf{C}_\Theta$ denotes the asymptotic covariance matrix of estimated DOA parameters under the snapshot independence assumption. Therefore, *the array must be oversampled, and the parameter of interest that characterizes performance is not the number of snapshots N but the observation interval T.*

B. Whitening Approach

In the special case where the covariance matrix $\mathbf{R}_n$ is known up to a multiplicative constant, the whitening of the noise used classically in direction of arrival estimation (DOA) (see, e.g., [18]) can be used to advantage because many second-order algorithms require $\mathbf{R}_n$ to be proportional to the identity matrix. In this approach, after $\mathbf{n}_t$ is whitened by a linear transformation applied to $\mathbf{x}_t$, many covariance-based methods based on the white noise assumption can be used. In these circumstances, it makes sense to study the influence of the correlations between the components of $\mathbf{n}_t$ and the selected linear transformation on the performance of this covariance-based estimator. Considering our functional analysis, Theorem 2 answers this question. The process $\mathbf{n}_t$ is whitened using the Cholesky decomposition $\mathbf{L}^+\mathbf{L}$ of $\mathbf{R}_n^{-1}$ and any unitary matrix $\mathbf{Q}$:

$$\mathbf{R}_n^{-1} = \mathbf{L}'^+\mathbf{L}' \text{ with } \mathbf{L}' \overset{\text{def}}{=} \mathbf{Q}\mathbf{L}$$

and the covariance matrix of $\mathbf{x}_t$ becomes

$$\mathbf{R}_x' = \mathbf{L}'\mathbf{E}_s(\Theta)\mathbf{R}_s\left(\mathbf{L}'\mathbf{E}_s(\Theta)\right)^+ + \sigma_n^2\mathbf{I}.$$

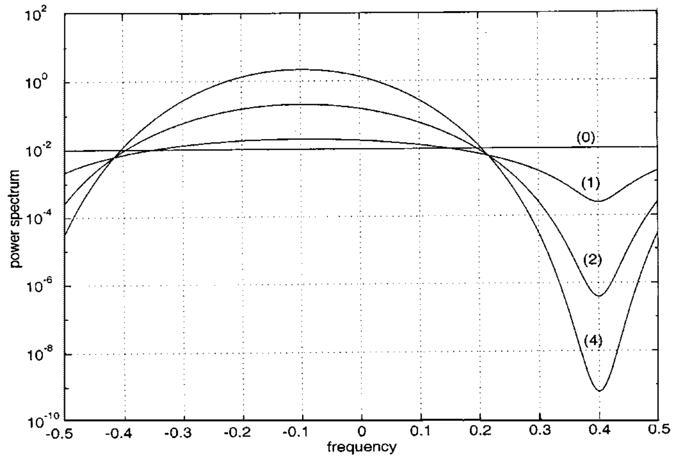

Fig. 2. Power density spectra for white (0) and MA noise of order (1), (2), (4) of zero $be^{i2\pi f_0}$ for $b = 0.8$ and $f_0 = 0.4$.

If alg(.) denotes a second-order algorithm based on the new data model $\mathbf{x}_t' = (\mathbf{L}'\mathbf{E}_s(\Theta))\mathbf{s}_t + \mathbf{n}_t'$ and white noise assumption, the parameters are estimated with the following scheme:

$$\begin{aligned}&\mathbf{R}_x(N) \mapsto \mathbf{R}_x'(N) \overset{\text{def}}{=} \mathbf{L}'\mathbf{R}_x(N)\mathbf{L}'^+ \overset{\text{alg}}{\mapsto} \Theta(N) \\ &\qquad \Rightarrow \mathbf{R}_x(N) \overset{\text{alg}'}{\mapsto} \Theta(N).\end{aligned}$$

Applying the chain differential rule, Theorem 2 applies in this situation by replacing in (4.9) and (4.10), $\mathbf{D}_{\Theta,\mathbf{R}_x}^{\text{alg}}$ by $\mathbf{D}_{\Theta,\mathbf{R}_x}^{\text{alg}'} = \mathbf{D}_{\Theta,\mathbf{R}_x'}^{\text{alg}}(\mathbf{L}' \otimes \mathbf{L}')$ because $\text{Vec}(\mathbf{R}_x'(N)) = (\mathbf{L}' \otimes \mathbf{L}')$ $\text{Vec}(\mathbf{R}_x(N))$.

To illustrate the sensitivity of performance to the coloring of the noise, a numerical study is presented. First, we note that taking into account the expression of $\mathbf{D}_{\Theta,\mathbf{R}_x}^{\text{MUSIC}}$, it is easily proved that the MUSIC algorithm associated with the prewhitening of the data is insensitive to the choice of the unitary matrix $\mathbf{Q}$. Consider a complex sinusoid corrupted additively by an MA process of transfer function $\left(1 - be^{i2\pi(f_0 - f)}\right)^r$ of order $r = 1$, 2, or 4. The power density spectrum of the noise is shown in Fig. 2. The sinusoid frequency is estimated from the sample covariance matrix of order $p = 3$ by the standard MUSIC algorithm after noise whitening. Fig. 3 plots the theoretical MSE of sinusoid frequency $f_1(N)$, where $SNR = 0$ dB, and $N = 200$ as a function of f_1 for white and

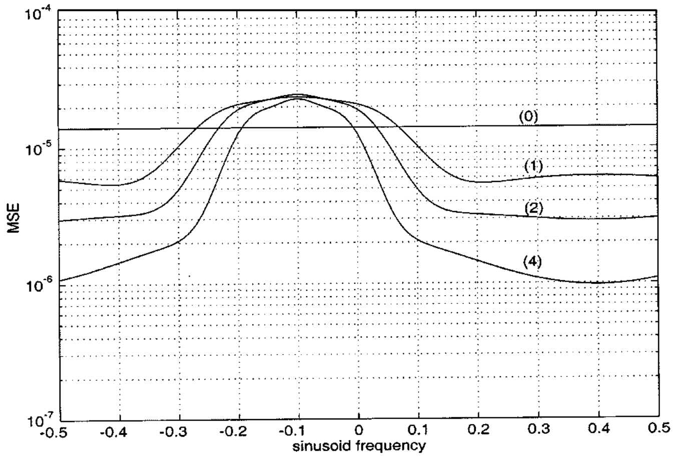

Fig. 3. Theoretical MSE of $f_1(N)$ versus the frequency f_1 of the complex sinusoid for (0) white and MA noise of order (1), (2), (4) with $p = 3, N = 200$, and $SNR = 0$ dB.

MA noise of order 1, 2, or 4. This figure shows a degradation of the performance when the frequency f_1 is in the vicinity of the maximum of the power density spectra of the noise. This result is similar to the performance of the nonlinear least square estimator [17], where the asymptotic variances are proportional to $S_n(f_1)/a_1^2$, although the MUSIC algorithm uses knowledge of the noise color contrary to the nonlinear least square estimator.

VI. Conclusion

In this paper, we have provided a unifying framework to investigate the asymptotic performance of second-order methods for parameter estimation under the stochastic model assumption. Thanks to a functional approach and a matrix-valued reformulated central limit theorem about the sample covariance matrix, we have specified conditions under which the second-order algorithms are robust to the temporal correlation and to the distribution of the signals involved. Our results have been illustrated in the context of DOA, FIR, and frequency estimators, and particular attention has been given to the temporal correlation of the noise and to the whitening approach.

For complex noncircular signals, our analysis did not take into account the second covariance matrix of the data. An asymptotic analysis of second-order based-algorithms dedicated to these specific signals is underway.

References

[1] P. Stoica and A. Nehorai, "Performance study of conditional and unconditional direction of arrival estimation," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 1783–1795, Oct. 1990.

[2] B. Ottersten, M. Viberg, and T. Kailath, "Analysis of subspace fitting and ML techniques for parameter estimation from sensor array data," *IEEE Trans. Signal Processing*, vol. 40, pp. 590–599, Mar. 1992.

[3] J. F. Cardoso and E. Moulines, "A robustness property of DOA estimators based on covariance," *IEEE Trans. Signal Processing*, vol. 42, pp. 3285–3287, Nov. 1994.

[4] J. P. Delmas and Y. Meurisse, "Asymptotic performance analysis of DOA algorithms with temporally correlated narrow-band signals," *IEEE Trans. Signal Processing*, vol. 48, pp. 2669–2674, Sept. 2000.

[5] K. A. Meraim, J. F. Cardoso, A. Y. Gorokhov, and P. Loubaton, "On subspace methods for blind identification of single-input multiple output FIR systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 42–55, Jan. 1997.

[6] H. H. Zeng and L. Tong, "Blind channel estimation using the second-order statistics: Asymptotic performance and limitations," *IEEE Trans. Signal Processing*, vol. 45, pp. 2060–2071, Aug. 1997.

[7] B. Ottersten, P. Stoica, and R. Roy, "Covariance matching estimation techniques for array signal processing," *Digital Signal Process.*, vol. 8, pp. 185–210, July 1998.

[8] P. Loubaton, E. Moulines, and P. Regalia, "Subspace methods for blind identification and deconvolution," in *Signal Processing Advances in Wireless and Mobile Comunications*. Englewood Cliffs, NJ: Prentice-Hall, 2001, vol. 1.

[9] O. Besson and P. Stoica, "Analysis of MUSIC and ESPRIT frequency estimates for sinusoidal signals with lowpass envelopes," *IEEE Trans. Signal Processing*, vol. 44, pp. 2359–2364, Sept. 1996.

[10] K. A. Meraim, E. Moulines, and P. Loubaton, "Prediction error method for second-order blind identification," *IEEE Trans. Signal Processing*, vol. 45, pp. 694–705, Mar. 1997.

[11] J. P. Delmas, "Asymptotic normality of sample covariance matrix for mixed spectra time series: Application to sinusoidal frequencies estimation," *IEEE Trans. Inform. Theory*, vol. 47, pp. 1681–1687, May 2001.

[12] P. J. Brockwell and R. A. Davis, *Time Series, Theory and Methods*, Second ed. New York: Springer-Verlag, 1991.

[13] B. Porat, *Digital Processing of Random Signals, Theory and Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1993.

[14] R. J. Serfling, *Approximation Theorems of Mathematical Statistics*. New York: Wiley, 1980.

[15] S. Visuri, H. Oja, and V. Koivunen, "Nonparametric statistics for subspace based frequency estimation," in *Proc. Conf. EUSIPCO*, Tampere, Finland, Sept. 2000.

[16] P. Stoica and T. Söderström, "Statistical analysis of MUSIC and subspace rotation estimates of sinusoidal frequencies," *IEEE Trans. Signal Processing*, vol. 39, pp. 1836–1847, Aug. 1991.

[17] P. Stoica, A. Jakobsson, and J. Li, "Cissoid parameter estimation in the colored noise case: Asymptotic Cramer–Rao bound, maximum likelihood, and nonlinear least-squares," *IEEE Trans. Signal Processing*, vol. 45, pp. 2048–2059, Aug. 1997.

[18] R. O. Schmidt, "A signal subspace approach to multiple emitter location and spectral estimation," Ph.D. dissertation, Stanford Univ., Stanford, CA, 1981.

Jean-Pierre Delmas (M'00) was born in France in 1950. He received the engineering degree from Ecole Centrale de Lyon, Lyon, France, in 1973 the Certificat d'études supérieures from the Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1982, and the Habilitation à diriger des recherches (HDR) degree from the University of paris, Orsay, France, in 2001.

Since 1980, he has been with the Institut National des Télécommunications, Evry, France, where he is currently a Professor. His teaching and research interests focus on statistical signal processing.