MIMO Channel Blind Identification in the Presence of Spatially Correlated Noise

Walid Hachem, *Member, IEEE*, François Desbouvries, *Member, IEEE*, and Philippe Loubaton, *Member, IEEE*

***Abstract*—We address the problem of the second-order blind identification of a multiple-input multiple-output (MIMO) transfer function in the presence of additive noise. The additive noise is assumed to be (temporally) white, i.e., uncorrelated in time, but we do not make any assumption on its spatial correlation. This problem is thus equivalent to the second-order blind identification of a MIMO transfer function in the noiseless case but from a partial auto-covariance function $\{R_n\}_{n\neq 0}$. Our approach consists of computing the missing central covariance coefficient R_0 from this partial auto-covariance sequence. It can be described simply within the algebraic framework of rational subspaces. We propose an identifiability result that requires very mild assumptions on the transfer function to be estimated. Practical subspace-based identification algorithms are deduced and tested via simulations.**

***Index Terms*—Colored noise, MIMO blind identification, rational spaces, stochastic realization.**

I. Introduction

In DIGITAL communications, channel equalization/identification techniques are called "blind" when they do not require the use of a training sequence, i.e., a sequence of symbols known by the receiver and used to update the equalizer. This feature may be of interest, even in a cooperative transmission framework: When the channel impulse response varies with time, as in microwave data transmissions, a training sequence is usually sent periodically. Suppressing this sequence increases the effective data rate.

A number of papers in the early 1990s have given a large impetus for the research of blind identification algorithms based on second-order statistics; see, for instance, the recent survey papers [1] and [2] and books [3] and [4], as well as the references therein. When the received signal is sampled at symbol rate, the resulting channel model, in general, is nonminimum phase. It thus came as a breakthrough that blind identification *could* indeed be achieved by second-order techniques [5]. In the early algorithms of [5]–[7], the ambiguity about the position of the zeroes of the channel with respect to the unit circle is removed by introducing specific redundancy into the data. More precisely, the received signal is first oversampled, as in a fractionally spaced equalizer. The resulting cyclostationary signal, when vectorized appropriately, becomes multidimensional stationary. Under appropriate conditions, which indeed mean that a sufficient amount of diversity has been introduced, the (vector) channel becomes minimum phase, and the problem can be solved by second-order techniques.

The signal model used in the framework of second-order blind identification is thus, by essence, a multidimensional model. Later on, these ideas were generalized to the multi-input case (see, e.g., [8]–[16]). The noiseless transmission model is then, using the so-called operator notation

$$y_n^u = [H(z)]s_n \tag{1}$$

where s_n is a zero-mean p-dimensional vector process such that $E[s_n s_m^T] = \delta_{n,m} I_p$, which represents the sequences of symbols emitted by p sources sharing the same period; $H(z) = \sum_{k\geq 0} H_k z^{-k}$ is a $q \times p$ transfer function representing the channel; and y_n^u is the received $q \times 1$ signal. The parameter q is always supposed strictly greater than p. As stated above, this model accounts for oversampling the received signal at a multiple of the symbol period, or receiving the signal on an array of sensors (with more sensors than sources), or a combination of both.

Usually, the channel transfer function is modeled as FIR. We denote its degree by M: $H(z) = \sum_{k=0}^{M} H_k z^{-k}$. $H(z)$ is assumed to be irreducible, which means that its rank is equal to p for all $z \neq 0$ and for $z = \infty$ (see, e.g., [9], [13], and [17]). Irreducibility is a "generic" property, as soon as $q > p$. An immediate consequence is that $H(z)$ can theoretically be recovered up to a constant $p \times p$ orthogonal matrix from the sole knowledge of the covariance function of y_n^u by using a linear prediction algorithm (see, e.g., [11], [13], and [14]). In this paper, we will use a linear prediction algorithm as part of our identification method.

Let us now consider the more realistic case of a noisy model.

$$y_n = [H(z)]s_n + w_n. \tag{2}$$

We suppose here that w_n is an additive q-dimensional white noise [i.e., $E(w_n w_m^T) = 0$ if $n \neq m$] uncorrelated with s_n. Let $\Sigma = E(w_n w_n^T)$. Most existing identification algorithms based on second-order statistics consider the case where $\Sigma = \sigma^2 I_q$, where σ^2 is an unknown scalar parameter (see [9]–[11], [13], [14], and [16]). One can identify σ^2 from the covariance matrix of the vector $Y_N(n) = [y_n^T, \ldots, y_{n-N}^T]^T$, provided the parameter N is chosen "large enough." To show this, let $T_N(H)$

Manuscript received July 14, 1999; revised November 9, 2001. The associate editor coordinating the review of this paper and approving it for publication was Prof. Dimitrios Hatzinakos.

W. Hachem is with Supelec, Service Radioélectricité, Gif-sur-Yvette, France (e-mail: Walid.Hachem@supelec.fr).

F. Desbouvries is with the Institut National des Télécommunications, Département CITI, Évry, France (e-mail: Francois.Desbouvries@int-evry.fr).

P. Loubaton is with Equipe Signal pour les Communications/UMLV and Laboratoire Communication et Traitement de l'Information/ENST-CNRS, Marne-la-Vallée, France (e-mail: loubaton@univ-mlv.fr).

Publisher Item Identifier S 1053-587X(02)01349-1.

denote the $(N+1)q \times (M+N+1)p$ generalized Sylvester matrix associated with $H(z)$

$$T_N(H) = \begin{bmatrix} H_0 & \cdots & H_M & & 0 \\ & \ddots & & \ddots & \\ 0 & & H_0 & \cdots & H_M \end{bmatrix}. \quad (3)$$

Then, it is clear that $Y_N(n) = T_N(H)S_{M+N}(n) + W_N(n)$, in which $S_{M+N}(n) = [s_n^T, \ldots, s_{n-(M+N)}^T]^T$ and $W_N(n) = [w_n^T, \ldots, w_{n-N}^T]^T$. The covariance matrix $\mathcal{R}_N^y$ of $Y_N(n)$ is thus equal to

$$\mathcal{R}_N^y = T_N(H)T_N^T(H) + \sigma^2 I_{(N+1)q}. \quad (4)$$

Since $p < q$, one can choose N such that $(N+1)q > (M+N+1)p$. If this holds, then $T_N(H)T_N^T(H)$ is a singular matrix so that σ^2 is identified as the smallest eigenvalue of $\mathcal{R}_N$, which essentially leads us back to the noiseless case.

Let us now discuss the hypotheses on the additive noise. The fact that w_n is a white noise is a reasonable assumption, but the assumption that $\Sigma = \sigma^2 I_q$ may be restrictive in certain contexts. Very often, y_n represents the signal sampled behind a sensor array. If w is a thermal noise that stems from the acquisition devices of sensors with hardware nonidealities, or if the prevailing noise is ambient noise and the sensor array is sparse, then the noise components are likely to be decorrelated but their variances do not necessarily coincide [18]. As another example, the noise in underwater acoustics is the sum of ambient sea noise, flow noise, and traffic noise and may thus be spatially correlated [19], [20]. This is also the case in radio frequency communications when ambient noise is the dominant noise source, such as communications in the HF or VHF frequency bands [21].

Contributions [18], [20], and [21] address the problem of blind identification of an unknown channel in the presence of additive colored noise. In these approaches, an instantaneous mixture model [i.e., $H(z) = H_0$] is assumed, and some model for the noise covariance matrix is required since these methods actually rely on a pertinent exploitation of the underlying structure.

In this paper, we are going to show that *in the case where no a priori information on* Σ *is available*, it is possible, under certain conditions, to identify a convolutive mixture [i.e., $H(z) = \sum_{k=0}^{M} H_k z^{-k}$ with $M \neq 0$] (up to a $p \times p$ orthogonal matrix) from the second-order statistics of y_n.

This problem was first introduced in [22] in the case $p = 1$. There, it was shown that the unknown $q \times 1$ transfer function $H(z)$ is not necessarily identifiable. In case of identifiability, an identification procedure based on the classical stochastic realization theory was proposed. However, it is based on a difficult nonconvex optimization problem for which no satisfying solution was proposed. Later on, it was shown in [23] (still in the case $p = 1$) that the SIMO FIR subspace identification method, which was introduced in [6] for $\Sigma = \sigma^2 I$, could be generalized to the case where Σ is unknown if $q \geq 3$. The case $p > 1$ was first considered in [24] in the polynomial case, and some results are available in the rational case [25]. Finally, the algorithm of [26], based on state space models, gives a solution in the case $p = 1$.

The results of [24] hold under restrictive assumptions on the transfer function to be identified [in particular, the degrees of the columns of $H(z)$ must coincide]. In this paper, we propose a new approach that provides satisfying identifiability results without making any strong assumption on $H(z)$. The paper is organized as follows. Section II outlines the general methodology. Since we will use the notion of rational spaces, the necessary tools are also briefly recalled for convenience of the nonexpert reader. A general theoretical identifiability result is then given and formulated in terms of rational subspaces. In Section III, we first show that the results of [24] follow from our identifiability result. More importantly, we take benefit of it to introduce and motivate our new approach. Practical algorithms implementing the results of [24] ([24] contains no algorithm) as well as our new method are detailed in Section IV. Finally, some Monte Carlo simulations are given in Section V and followed by comments on the applicability of these methods.

II. Outline of the Method

A. Problem Statement and General Methodology

Let us get back to the model (2), where $H(z)$ is irreducible. We assume from now on that $E(w_n w_m^T) = \Sigma \delta_{n,m}$, where Σ is unknown. Therefore, (4) becomes

$$\mathcal{R}_N^y = T_N(H)T_N^T(H) + I_{N+1} \otimes \Sigma \quad (5)$$

where $\otimes$ denotes the Kronecker product. As a consequence, the eigenvalue considerations of Section I no longer hold.

Let us denote by $\{R_n\}_{n\in\mathbb{Z}}$ and by $\{R_n^y\}_{n\in\mathbb{Z}}$ the auto-covariance lags of the useful signal $[H(z)]s_n$ and of y_n, respectively. As w_n is assumed to be white and $H(z)$ is a degree M polynomial, it is clear that $R_0^y = R_0 + \Sigma$, $R_n^y = R_n$ for $1 \leq |n| \leq M$, and $R_n^y = R_n = 0$ for $|n| > M$. In particular, R_0^y contains little information on R_0; therefore, this term will simply be ignored (simulations will show that this indeed is pertinent at low SNR).

We will thus try to identify $H(z)$ from the lags $\{R_k\}_{k>0}$ by relying on the convolutive nature of the channel only. We thus reformulate our problem as follows.

Let $H(z) = \sum_{k=0}^{M} H_k z^{-k}$ be a $q \times p$ irreducible FIR transfer function, and let $\{R_n\}$ be the auto-covariance sequence associated with the $q \times q$ "spectral density" $S(z) = H(z)H^T(z^{-1})$. Set reasonable hypotheses under which $H(z)$ is identifiable (up to a $p \times p$ orthogonal matrix) from the partial sequence $\{R_n\}_{1\leq n\leq M}$, and deduce constructive identification algorithms.

We proceed by solving a completion problem. The general idea is as follows. Even though we do not know R_0, we do not deal with just any wide-sense stationary vector process y_n but with the noiseless MA model (1). Since $\text{Rank}(S(z)) = p < q$ for all z, this information can be exploited efficiently via the structural equation $H(z)H^T(z^{-1}) = S(z) = \sum_{k=-M}^{M} R_k z^{-k}$. This equation provides an implicit relation among the covariance lags $\{R_k\}_{k=0}^{M}$, which, in turn, enables (under certain sufficient conditions) recovery of the central lag R_0 from $\{R_k\}_{k=1}^{M}$. Last, once R_0 is known, $H(z)$ can be retrieved from $H(z)H^T(z^{-1})$ via any stochastic realization algorithm and, thus, for instance, by using the popular so-called linear prediction algorithm, which can be used here because of the irreducibility assumption.

Let us now outline our method. Our results are based on the following observation. Since $q > p$, there exist $1 \times q$ polynomials $g(z) = \sum_{k=0}^{N} g_k z^{-k}$ of degree N satisfying

$$g(z)H(z) = 0 \quad \text{for all} \quad z \tag{6}$$

or, equivalently, satisfying $g(z)H(z)H^T(z^{-1}) = 0$ for all z. Let us assume that we can compute a set of r such polynomials $g_j(z)$, and let $G(z)$ be the $r \times q$ matrix

$$G(z) = [g_1^T(z), \ldots, g_r^T(z)]^T = \sum_{n=0}^{N} G_n z^{-n} \tag{7}$$

(the parameters r and N will be specified later). Put $S(z) = \sum_{n=-M}^{M} R_n z^{-n} = R_0 + T(z)$, where $T(z)$ is supposed to be known. Since $S(z) = H(z)H^T(z^{-1})$, it is clear that

$$G(z)R_0 = -G(z)T(z). \tag{8}$$

Identifying on both sides the coefficient of z^{-k} for $0 \leq k \leq N$ and setting

$$\mathcal{G} = [G_0^T, \ldots, G_N^T]^T \tag{9}$$

we see that we can compute the matrix $\mathcal{G}R_0$ from the data. If $\mathcal{G}$ has full column rank, R_0 can be retrieved from (8), and finally, $H(z)$ is identifiable from the full sequence $\{R_n\}_{n=0}^{M}$.

Our method is, thus, twofold. It relies on the actual possibility of computing from $\{R_n\}_{n=1}^{M}$ polynomials $g(z)$ satisfying (6), but on the other hand, these polynomials should be "in sufficient number" and their degrees "sufficiently high." This is because the critical point is that $\mathcal{G}$ must have full column rank; intuitively, this is all the more likely to happen that $\mathcal{G}$ contains a large number of rows; therefore, the parameters r and N (which have not been determined for the moment) should be "large." This indeed will be confirmed later on by the identification theorems.

The properties of these polynomials (and particularly their degrees) naturally fit into the framework of rational spaces, i.e., of spaces of rational vectors over the field of rational functions. Consequently, we first recall some general results on this subject that will prove of importance to our problem. For a detailed study, see [27] and [28].

B. Brief Review of Rational Subspaces

Let us first recall that the set $\mathbb{R}^q(z)$ of all $q \times 1$ rational transfer functions is a q-dimensional subspace over the field $\mathbb{R}(z)$ of all scalar rational transfer functions. Let $\mathcal{S}$ be a d-dimensional (with $d < q$) subspace of $\mathbb{R}^q(z)$. $\mathcal{S}$ admits bases $U(z) = \{u_1(z), \ldots, u_d(z)\}$ characterized by the fact that $\text{Rank}(U(z)) = d$ for almost all z. The rational matrix-valued function $U(z)$ is then said to have a normal rank equal to d.

$\mathcal{S}$ admits, in particular, polynomial bases. One way of building polynomial bases is as follows. Among all polynomials belonging to $\mathcal{S}$, choose a polynomial $u_1(z)$ of minimum degree. Then, among all the remaining polynomials in $\mathcal{S}$, choose a polynomial $u_2(z)$ that is linearly independent with $u_1(z)$ and of minimum degree. Doing this, we get a so-called minimal polynomial basis $\{u_i(z)\}_{i=1}^{d}$. Minimal polynomial bases are characterized by the well-known criterion [27], [28] in the following.

Proposition 1: The polynomial basis $\{u_i(z)\}_{i=1}^{d}$ is minimal if and only if the matrix polynomial $U(z) = [u_1(z), \ldots, u_d(z)]$ is irreducible and column reduced.

Of course, there are many different minimal polynomial bases that can be obtained by the above procedure. However, an interesting feature is that all minimal polynomial bases share the same degrees $\{M_i = \deg(u_i(z))\}_{i=1}^{d}$. Usually, these minimal degrees $\{M_i\}_{i=1}^{d}$ are called the Kronecker indices associated with $\mathcal{S}$.

The dual space $\mathcal{S}^\perp$ of $\mathcal{S}$ is the $(q-d)$-dimensional subspace of all $1 \times q$ rational transfer functions $g(z)$ satisfying $g(z)f(z) = 0$ for all $f \in \mathcal{S}$. The Kronecker indices $\{M_j^\perp\}_{j=1}^{q-d}$ of $\mathcal{S}^\perp$ are the so-called dual Kronecker indices of $\mathcal{S}$. They satisfy the important equality

$$\sum_{i=1}^{d} M_i = \sum_{j=1}^{q-d} M_j^\perp. \tag{10}$$

C. Identifiability Result Based on Rational Subspaces

We now turn back to our problem and make things more precise. In particular, we will derive a simple identification lemma that will be used in two different ways in Section III.

Let us first trivially remark that a polynomial $g(z)$ satisfying (6) belongs to the dual of some rational space enclosing (but not necessarily equal to) the rational space $\text{Span}(H(z))$ generated by the columns of $H(z)$. Therefore, let $\mathcal{S}$ be some fixed rational space such that $\text{Span}(H(z)) \subseteq \mathcal{S}$, and let $d = \dim(\mathcal{S})$. By assumption, $H(z)$ is irreducible (see the problem formulation in Section II-A), and thus, it has full (normal) rank p; therefore, $\dim(\text{Span}(H(z))) = p$. We thus have $p \leq d \leq q$. Let $\mathcal{S}^\perp$ denote the dual space of $\mathcal{S}$. In addition, let $0 \leq M_1 \leq \cdots \leq M_d$ (respectively, $0 \leq M_1^\perp \leq \cdots \leq M_{q-d}^\perp$) denote the Kronecker indices of $\mathcal{S}$ (respectively, the dual Kronecker indices of $\mathcal{S}$) arranged in increasing order.

Now, assume that we have at hand some algorithm for computing the set $\mathcal{S}_N^\perp$ made of all polynomials of $\mathcal{S}^\perp$ but of degree lower than or equal to N. Then, one should stack in the matrix $G(z)$ defined by (7) "as many" $1 \times q$ polynomials $g_j(z)$ of $\mathcal{S}^\perp$ as possible; however, only polynomials that are linearly independent over the field $\mathbb{R}(z)$ should be considered. To see this, let $\{g_j(z)\}_{j=1}^{r}$ be a set of r polynomials of $\mathcal{S}_N^\perp$ that are linearly independent over $\mathbb{R}(z)$, and let $G(z) = [g_1^T(z), \ldots, g_r^T(z)]^T$. Then, adding a $(r+1)$st row $g_{r+1}(z)$ to $G(z)$ is useless [as regards the rank of the matrix $\mathcal{G}$ defined by (9)] if this polynomial $g_{r+1}(z)$ belongs to the rational space spanned by $\{g_j(z)\}_{j=1}^{r}$. In this case, it is easily seen that the $N+1$ rows brought by $g_{r+1}(z)$ to the matrix $\mathcal{G}$ are linear combinations of the $r(N+1)$ previous ones.

The maximum *useful* value for r, which is considered to be a function $r(N)$ of N, is thus equal to the dimension of the rational subspace of $\mathcal{S}^\perp$ generated by the set $\mathcal{S}_N^\perp$, but this dimension depends on the position of N with respect to the dual Kronecker indices $\{M_j^\perp\}$ of $\mathcal{S}$. If $N < M_1^\perp = \min_j M_j^\perp$, then the equation $g(z)f(z) = 0$ for all $f(z) \in \mathcal{S}$ holds if and only if $g(z) = 0$ for all z. Therefore, $G(z)$ is the null matrix. If $M_s^\perp \leq N < M_{s+1}^\perp$, then $r(N)$ is easily seen to be equal to s. Last, if $N \geq M_{q-d}^\perp = \max_j M_j^\perp$, then one can

extract from $\mathcal{S}_N^{\perp}$ a set of $(q-d)$ polynomials forming a basis of $\mathcal{S}^{\perp}$; therefore, $r(N) = q-d$. This, of course, is the most favorable situation. Therefore, it is desirable to have an algorithm for computing, from the sequence $\{R_n\}_{n=1}^{M}$, the set $\mathcal{S}_N^{\perp}$ for $N \geq M_{q-d}^{\perp} = \max_j M_j^{\perp}$. If this holds, R_0 can be identified from (8) under a simple additional condition, as we now see.

Lemma 1: Let Span$(H(z))$ denote the rational subspace spanned by the columns of $H(z)$, and let $\mathcal{S}$ be some rational subspace such that Span$(H(z)) \subseteq \mathcal{S}$. Let d denote the dimension of $\mathcal{S}$, and let $0 \leq M_1 \leq \cdots \leq M_d$ and $0 \leq M_1^{\perp} \leq \cdots \leq M_{q-d}^{\perp}$ be its Kronecker and dual Kronecker indices, respectively. Let $\mathcal{S}_N^{\perp}$ be the set of all polynomials of $\mathcal{S}^{\perp}$ of degree lower than or equal to N. Assume that we have an algorithm for computing $\mathcal{S}_N^{\perp}$ for $N \geq M_{q-d}^{\perp} = \max_j M_j^{\perp}$. If $M_1 \geq 1$, then R_0 is identifiable from the sequence $\{R_n\}_{n=1}^{M}$.

Proof: As $N \geq M_{q-d}^{\perp}$, it is possible to extract from $\mathcal{S}_N^{\perp}$ a set of $(q-d)$ polynomials $(g_1(z), \ldots, g_{q-d}(z))$ forming a basis of $\mathcal{S}^{\perp}$. Let $G(z) = \sum_{n=0}^{N} G_n z^{-n} = (g_1(z)^T, \ldots, g_{q-d}^T(z))^T$ be the corresponding $(q-d) \times q$ polynomial matrix. Then, $G(z)$ satisfies (8). Let us now consider the matrix $\mathcal{G} = [G_0^T, \ldots, G_N^T]^T$ associated with $G(z)$. Let v be a q-dimensional constant vector satisfying $\mathcal{G}v = 0$ or, equivalently, $G(z)v = 0$ for all z. This condition holds if and only if the constant polynomial $v(z) = v$ belongs to the dual space of $\mathcal{S}^{\perp}$, i.e., to $\mathcal{S}$. On the other hand, the assumption $M_1 > 0$ means that $\mathcal{S}$ does not contain nonzero constant vectors, and thus, $v = 0$. Therefore, $\mathcal{G}$ is a full column rank matrix, and R_0 can be identified from (8), as expected. ∎

In the next section, we are going to propose two algorithms for computing (for N large enough) the sets $\mathcal{S}_N^{\perp}$ associated with two different rational subspaces $\mathcal{S}$ containing Span$(H(z))$. For the first algorithm, the space $\mathcal{S}$ coincides with Span$(H(z))$ itself, whereas for the second, it is no longer the case.

III. Subspace-Based Methods

In this section, we will state the principle of two algorithms for computing R_0 from $\{R_i\}_{i=1}^{M}$; both of them are practical applications of Lemma 1.

A. Making Use of the Block–Hankel Matrix Associated to the Sequence $\{R_n\}_{1 \leq n \leq M}$

In Section II, we considered a d-dimensional (with $d \geq p$) rational subspace $\mathcal{S}$ enclosing (but not necessarily equal to) the rational subspace Span$(H(z))$ generated by the columns of $H(z)$. Now, the most natural choice for $\mathcal{S}$ is, of course, Span$(H(z))$ itself. This choice was implicitly adopted in a method proposed in [24]. However, in order to compute the space $\mathcal{S}_N^{\perp}$ for N large enough, it is based on rather strong assumptions on $H(z)$. In order to better appreciate the content of Section III-B, we now briefly reformulate the method proposed in [24] in the light of the preceding results.

Let us assume the following.

- **C1)** The columns $h_i(z)$ of $H(z)$ all share the same degree $M > 0$.
- **C2)** $H(z)$ is irreducible, i.e., Rank$(H(z)) = p$ for all $z \neq 0$, including $z = \infty$.
- **C3)** $H(z)$ is column reduced, i.e., Rank$(H_M) = p$, because of **C1)**.

From **C2)** and **C3)**, the columns of $H(z)$ form a minimal polynomial basis of the subspace $\mathcal{S} = \text{Span}(H(z))$, due to Proposition 1. Thus, the Kronecker indices $\{M_i\}_{i=1}^{p}$ of $\mathcal{S}$ are all equal to M, and consequently, the identifiability condition of Lemma 1 (i.e., $M_1 > 0$) is, of course, satisfied.

Of course, according to that same lemma, we still have to extract the set $\mathcal{S}_N^{\perp}$ for $N \geq M_{q-p}^{\perp}$. However, the method proposed in [24] enables the computation of the set of all polynomials satisfying (6) *and* of degree lower or equal to $M-1$, i.e., $\mathcal{S}_{M-1}^{\perp}$; therefore, the identification scheme of [24] is successful provided $M - 1 \geq M_{q-p}^{\perp}$. As for size conditions, it is useful to notice that (10) reads $Mp = \sum_{j=1}^{q-p} M_j^{\perp} \leq (q-p) M_{q-p}^{\perp}$ so that the condition $M - 1 \geq M_{q-p}^{\perp}$ implies, in particular, that $q \geq 2p + (p/(M-1))$.

Let us explain how we compute the polynomials of $\mathcal{S}^{\perp}$ of degree $M-1$. Note that the assumption $(M-1) \geq M_{q-p}^{\perp}$ implies in particular that $(M-1) \geq M_1^{\perp}$; therefore, this subset is not trivially reduced to $\{0\}$ (see the discussion in Section II-C). The method is as follows. Let $\mathcal{H}$ be the $Mq \times Mq$ block Hankel matrix given by

$$\mathcal{H} = \begin{bmatrix} R_1 & \cdots & R_M \\ \vdots & \iddots & \\ R_M & & 0 \end{bmatrix}. \tag{11}$$

$\mathcal{H}$ can be factored as

$$\mathcal{H} = \begin{bmatrix} H_1 & \cdots & H_M \\ \vdots & \iddots & \\ H_M & & 0 \end{bmatrix} \begin{bmatrix} H_0^T & & 0 \\ \vdots & \ddots & \\ H_{M-1}^T & \cdots & H_0^T \end{bmatrix} = \mathcal{O}\mathcal{C}^T \tag{12}$$

where $\mathcal{O}$ and $\mathcal{C}$ are $Mq \times Mp$ matrices. Since H_M and H_0 have full column rank, it is clear that $\mathcal{O}$ and $\mathcal{C}$ also have full column rank Mp. Therefore, the rank of $\mathcal{H}$ is also equal to Mp. Let J be the q-block exchange matrix: $J = J_{M \times M} \otimes I_q$, where $\otimes$ denotes the Kronecker product, and $\{J_{M\times M}(i,j) = \delta_{i+j-(M+1)}\}_{i,j=1}^{M}$. It is easy to check that a Mq-dimensional row vector $g = [g_0, \ldots, g_{M-1}]$ (in which each g_k is q-dimensional) satisfies

$$gJ\mathcal{H} = 0 \quad \text{and} \quad \mathcal{H}g^T = 0 \tag{13}$$

if and only if $gT_{M-1}(H) = 0$. Therefore, $\text{Ker}^l(T_{M-1}(H)) = \text{Ker}^l(J\mathcal{H}) \cap \text{Ker}^l(\mathcal{H}^T) = \text{Ker}^l([J\mathcal{H}\,\mathcal{H}^T])$ (Ker^l stands for the left kernel). Let $g(z) = \sum_{k=0}^{M-1} g_k z^{-k}$ be the $(M-1)-$ degree polynomial associated with g. Then, it is easily seen that $gT_{M-1}(H) = 0$ if and only if $g(z)H(z) = 0$ for all z. Therefore, there is a one-to-one correspondence between the space $\text{Ker}^l(J\mathcal{H}) \cap \text{Ker}^l(\mathcal{H}^T)$ and the set $\mathcal{S}_{M-1}^{\perp}$ of all polynomials of degree less than or equal to $M-1$ belonging to $\mathcal{S}^{\perp}$.

Let us summarize the discussion thus far into the following theorem.

Theorem 1: Let $H(z) = \sum_{k=0}^{M} H_k z^{-k}$ be a $q \times p$ polynomial. Assume that **C1)** to **C3)** hold. Let $\mathcal{S} = \text{Span}(H(z))$ denote the p-dimensional rational subspace spanned by the columns of

$H(z)$, and let $\{M_j^\perp\}_{j=1}^{q-p}$, with $0 \le M_1^\perp \le \cdots \le M_{q-p}^\perp$, denote the dual Kronecker indices of $\mathcal{S}$. Let $\mathcal{S}_{M-1}^\perp = \{g(z)$, s.t. $\deg(g) \le M-1$ and $g(z)H(z) = 0\}$. Assume that $(M-1) \ge M_{q-p}^\perp$ [which, in particular, implies that $q \ge 2p + (p/(M-1))$].

Then, it possible to compute $\mathcal{S}_{M-1}^\perp$ and to extract from $\mathcal{S}_{M-1}^\perp$ $(q-p)$ linearly independent polynomials $\{g_j(z)\}_{j=1}^{q-p}$. Put $G(z) = (g_1^T(z), \ldots, g_{q-p}^T(z))^T = \sum_{k=0}^{M-1} G_k z^{-k}$. We have $G(z)H(z) = 0$, and the matrix $\mathcal{G} = [G_0^T, \ldots, G_{M-1}^T]^T$ has full column rank. Therefore, the matrix R_0 can be identified from (8).

The considerations exposed in this section are constructive since they form the material that will be used later on for building an algorithm for identifying R_0. However, the practical implementation is not straightforward, and more work is needed. For sake of clarity, the necessary discussion is postponed to Section IV.

B. Alternate Rational Subspace Method

Conditions **C1)** and **C3)** require that H_M is a full column rank matrix. This assumption is often quite restrictive. Consider, for instance, a microwave multiple path transmission, and suppose that the symbol shaping filters are FIR. Element $h_{ij}(z)$ of $H(z)$ is the channel that conveys information from source i to sensor j. Its impulse response is often modeled as a finite sum of attenuated and time-delayed versions of the impulse response of the shaping filter. As a consequence, the degree of $h_{ij}(z)$ is related in a straightforward manner to the maximum delay of all paths between source i and sensor j. These delays might be distributed in such a way that two different columns of $H(z)$, which represent the vector transfer functions associated with two different sources, do not have the same degree. As a matter of fact, the transmission media between these sources and the sensor array might have different physical properties.

Furthermore, under **C1)**, the approach of Section III-A enables computation of polynomials belonging to the dual of the subspace $\mathcal{S} = \text{Span}(H(z))$ but of degrees bounded by $M-1$. On the other hand, the approach requires that $M-1 \ge M_{q-p}^\perp$. This additional condition is difficult to interpret and can also be restrictive. Consider, for example, the case where $p=2$, $q=5$, and $M=5$. Among the 14 triples $M_1^\perp$, $M_2^\perp$, $M_3^\perp$ satisfying $0 \le M_1^\perp \le M_2^\perp \le M_3^\perp$ and $\sum_{j=1}^3 M_j^\perp = 10$, only two satisfy the condition. If $M-1 < M_{q-p}^\perp$, one can compute $s < (q-p)$ linearly independent polynomials of $\mathcal{S}^\perp$ of degree $M-1$; however, we no longer have any reasonable condition guaranteeing that the corresponding matrices $\mathcal{G}$ have full column rank.

In this section, we propose a new alternative approach that overcomes these drawbacks. We will look for polynomials belonging to the dual space of a rational subspace $\mathcal{S}'$ strictly containing $\text{Span}(H(z))$; on the other hand, we will not encounter limitations on the degrees of these polynomials. On the whole, we will get new alternate identifiability conditions, which happen to be less severe than those of Theorem 1.

The idea is as follows. The unknown constant term R_0 in $S(z)$ can be eliminated by considering the difference of the function $S(z)$ at two different points. Our approach is based on the use of the function $Q_\alpha(z) = S(z) - S(ze^{i\alpha\pi})$, where α is a given real number. This function does not depend on R_0 and can thus be computed from the data. From $S(z) = H(z)H^T(z^{-1})$, we have

$$Q_\alpha(z) = \sum_{k=-M}^{M} Q_k^\alpha z^{-k} = [H(z) \quad H(ze^{i\alpha\pi})] \begin{bmatrix} H^T(z^{-1}) \\ -H^T(z^{-1}e^{-i\alpha\pi}) \end{bmatrix}. \quad (14)$$

In all this section, we will still assume that $H(z)$ is irreducible (see the problem formulation in Section II-A). We will also assume the following.

- **C4)** $q > 2p$.
- **C5)** Normal Rank$([H(z)\, H(ze^{i\alpha\pi})]) = 2p$.

Let us denote here by $\mathcal{S}'$ the rational subspace generated by the columns of $[H(z)\, H(ze^{i\alpha\pi})]$ and by $\mathcal{S}'^\perp$ its dual space. The space $\mathcal{S}'$, of course, contains $\text{Span}(H(z))$. Due to assumption **C5)**, $\mathcal{S}'$ (respectively, $\mathcal{S}'^\perp$) has dimension $2p$ (respectively, $q-2p$). The Kronecker indices of $\mathcal{S}'$ are denoted by $\{M_i'\}_{i=1}^{2p}$, with $0 \le M_1' \le \cdots \le M_{2p}'$, and those of $\mathcal{S}'^\perp$ by $\{M'^\perp_j\}_{j=1}^{q-2p}$, with $0 \le M'^\perp_1 \le \cdots \le M'^\perp_{q-2p}$.

Let us get into the derivation. Since $Q_\alpha(z)$ is known, one can extract for each N the set of all degree N $1 \times q$ polynomials $g(z) = \sum_{k=0}^N g_k z^{-k}$ satisfying $g(z)Q_\alpha(z) = 0$ for all z; this is because $g(z)Q_\alpha(z) = 0$ if and only if the $(N+1)q$ row vector $g = [g_0, \ldots, g_N]$ belongs to the left kernel of the *known* $(N+1)q \times (2M+N+1)q$ generalized Sylvester matrix $T_N(Q_\alpha)$ associated with $Q_\alpha(z)$

$$T_N(Q_\alpha) = \begin{bmatrix} Q_{-M}^\alpha & \cdots & Q_M^\alpha & & 0 \\ & \ddots & & \ddots & \\ 0 & & Q_{-M}^\alpha & \cdots & Q_M^\alpha \end{bmatrix}. \quad (15)$$

Now, **C5)** and (14) ensure that the rational spaces spanned by the columns of $Q_\alpha(z)$ and of $[H(z)\, H(ze^{i\alpha\pi})]$, respectively, are equal and, thus, that $g(z)Q_\alpha(z) = 0$ if and only if $g(z)[H(z)\ H(ze^{i\alpha\pi})] = 0$. This, in turn, implies that $g(z)H(z) = 0$ for all z. Therefore, using the function $Q_\alpha(z)$ enables extraction of polynomials of $\mathcal{S}'^\perp$ of degree N for all N. The following theorem is a consequence of Lemma 1.

Theorem 2: Let $H(z) = \sum_{k=0}^M H_k z^{-k}$ be a $q \times p$ polynomial. Let $\mathcal{S}'$ denote the rational subspace spanned by the columns of $[H(z)H(ze^{i\alpha\pi})]$. Let us assume that **C4)** and **C5)** hold, and let $\{M_i'\}_{i=1}^{2p}$, with $0 \le M_1' \le \cdots \le M_{2p}'$, denote the Kronecker indices of $\mathcal{S}'$. Let us assume the following.

- **C6)** The rational space $\mathcal{S}'$ does not contain nonzero constant vectors, i.e., that its smallest Kronecker index M_1' is nonzero.

Let N be an integer such that $N \ge 2\sum_{i=1}^p \deg(h_i(z))$, where $h_i(z)$ is the ith column of $H(z)$. Then, it is possible to compute $(q-2p)$ linearly independent degree N polynomials $\{g_j(z)\}_{j=1}^{q-2p}$ of $\mathcal{S}'^\perp$. Put $G(z) = [g_1^T(z), \ldots, g_{q-2p}^T(z)]^T = \sum_{k=0}^N G_k z^{-k}$. Then, the $(N+1)(q-2p) \times q$ matrix $\mathcal{G} = [G_0^T, \ldots, G_N^T]^T$ has full column rank; therefore, R_0 can be identified from (8).

Proof: The inequalities $M'^\perp_{q-2p} \le \sum_{j=1}^{q-2p} M'^\perp_j = \sum_{i=1}^{2p} M_i' < \sum_{i=1}^p (\deg(h_i(z)) + \deg(h_i(ze^{i\alpha\pi}))) = 2\sum_{i=1}^p \deg(h_i(z))$ hold (the second inequality is strict because

$[H(z)H(ze^{i\alpha\pi})]$ is not column reduced). Therefore, the condition $N \geq 2\sum_{i=1}^{p}\deg(h_i(z))$ implies that $N \geq M'^{\perp}_{q-2p}$. Hence, it is possible to compute $(q-2p)$ linearly independent degree N elements $(g_1(z), \ldots, g_{q-2p}(z))$ of $\mathcal{S}'^{\perp}$. Since we assume that $M'_1 > 0$, the matrix $\mathcal{G} = [G_0^T, \ldots, G_N^T]^T$ has full column rank (see Lemma 1). ■

As we now see, this new approach needs less restrictive assumptions on $H(z)$ than that exploiting the block Hankel matrix $\mathcal{H}$.

Let us first notice that the conditions $\deg(h_i(z)) = M$ for all i, $\text{Rank}(H_M) = p$, and $M - 1 \geq M^{\perp}_{q-p}$ of Theorem 1 are now relaxed.

On the other hand, Theorem 2 shows that R_0 can be identified from any basis of $\mathcal{S}'^{\perp}$, provided that **C6)** $M'_1 \geq 1$; it is thus interesting to evaluate whether this new identifiability condition is restrictive or not. If $\deg(H(z)) = 1$, one can easily verify that $\mathcal{S}'$ contains at least p independent constant vectors. Consequently, this rational subspace admits at least p zero Kronecker indices, and **C6)** is never satisfied. If $\deg(H(z)) = 2$, it can be established that **C6)** is not satisfied if $\text{rank}([H_1\ H_2]) \neq 2p$, which is rather restrictive. However, for degrees higher than two, it can be shown that **C6)** is not satisfied only for a few peculiar transfer functions $H(z)$ and, therefore, becomes a mild condition.

These results meet our expectations. Since the approach relies on the identification of R_0 from $\{R_n\}_{1\leq n\leq M}$, the fact that identifiability becomes difficult to satisfy when M is small is somewhat intuitive. In other words, the dispersive character of the channel plays a major role here.

Finally, it is interesting to notice that the identifiability condition $M'_1 \geq 1$ remains valid in the case where $H(z)$ is rational. The identification algorithm can be adapted to this case without major difficulty. We have chosen to restrict ourselves to the polynomial case for sake of clarity.

IV. Practical Considerations

In the preceding section, two constructive identification methods, based on Lemma 1, were proposed. However, some work is still needed before we effectively get practical algorithms. These implementation problems are considered in this section. In the remainder of this paper, the algorithm deduced from the method described in Sections III-A and B will be referred to as "algorithm A" and "algorithm B," respectively. We first begin with the implementation of algorithm B. The implementation of algorithm A will be treated at the end of this section.

The first step of algorithm B consists of finding, for N large enough, a degree N $(q-2p)\times q$ polynomial matrix $G(z) = \sum_{k=0}^{N} G_k z^{-k}$, the rows of which belong to $\mathcal{S}'^{\perp}$ and are linearly independent. In other words, we look for a polynomial basis of the rational subspace $\mathcal{S}'^{\perp}$.

Let us first notice that $G(z)[H(z)H(ze^{i\alpha\pi})] = 0$ if and only if $[G_0 \cdots G_N]T_N(Q_\alpha) = 0$, where $T_N(Q_\alpha)$ is defined in (15). Therefore, it is obvious that the subspace where the rows of the $(q-2p)\times(N+1)q$ matrix $[G_0, \ldots, G_N]$ must be looked for is the left kernel of the $(N+1)q \times (2M+N+1)q$ generalized Sylvester matrix $T_N(Q_\alpha)$. In practice, this means that we are going to minimize a quadratic form built on the Grammian $T_N(\widehat{Q_\alpha})T_N(\widehat{Q_\alpha})^T$ of an empirical estimate $T_N(\widehat{Q_\alpha})$ of $T_N(Q_\alpha)$. To avoid trivial solutions, some constraint should also be introduced into this minimization problem.

Now, remember that the rows of $G(z)$ are linearly independent (over the field of rational functions) if and only if $G(z)$ has full normal rank $(q-2p)$. Therefore, our constraint should be chosen to enforce this theoretical requirement. Since $G(z)$ is a polynomial matrix, it has full normal row rank $(q-2p)$ if and only if it has full row rank $(q-2p)$ for at least one value $z_0 \in \{\mathbb{C}^\star \cup \infty\}$. On the other side, there exist bases of $\mathcal{S}'^{\perp}$ that have full rank at infinity. For instance, minimal polynomial bases are irreducible and thus verify this property. This means that we can choose without restriction $G(z)$ in such a way that $G_0 = G(z=\infty)$ has full row rank. Finally, we can assume without loss of generality that the rows of the $(q-2p)\times q$ matrix G_0 are orthogonal.

The problem thus amounts to looking for the solution of the constrained optimization problem

$$\begin{cases} \text{minimize} \quad \text{trace}\left(\left[\hat{G}_0, \ldots, \hat{G}_N\right]\left(T_N\left(\hat{Q}_\alpha\right)T_N^T\left(\hat{Q}_\alpha\right)\right)\right. \\ \qquad\qquad\qquad \left.\cdot\left[\hat{G}_0, \ldots, \hat{G}_N\right]^T\right) \\ \text{under the constraint} \qquad \hat{G}_0\hat{G}_0^T = I_{q-2p} \end{cases} \tag{16}$$

for which a solution is presented in the Appendix.

As for the second step of the algorithm, we estimate the missing covariance lag R_0 using (8). Stated as a matrix equality, this equation becomes

$$\begin{bmatrix} \hat{G}_0 \\ \vdots \\ \hat{G}_N \end{bmatrix} \hat{R}_0 = -\begin{bmatrix} 0 & & \hat{G}_0 & \cdots & \hat{G}_M \\ & \iddots & & & \hat{G}_{M+1} \\ \hat{G}_0 & & & & \vdots \\ \vdots & & & & \hat{G}_N \\ & & & \iddots & \\ \hat{G}_{N-M} & \cdots & \hat{G}_N & & 0 \end{bmatrix} \begin{bmatrix} \hat{R}_M \\ \vdots \\ 0 \\ \vdots \\ \hat{R}_{-M} \end{bmatrix} \tag{17}$$

in which the first matrix of the left-hand side (resp. of the right-hand side) is $(N+1)(q-2p)\times q$ [resp. $(N+1)(q-2p)\times(2M+1)q$].

Of course, this system must be solved in the mean-square sense. At this point, it is interesting to make use of the Hermitian character of the auto-covariance function (i.e., enforce $\hat{R}_{-M} = \hat{R}_M^T$), as well as to constrain the solution $\hat{R}_0$ to be a Hermitian matrix; besides, this Hermitian character of $\hat{R}_0$ will also be used in the identification algorithm of $H(z)$ when it appears later in (19). Hence, instead of solving (17), we solve the system deduced from (17) by vectorizing the $q(q+1)/2$ unknown entries of $\hat{R}_0$, and modify the left-hand and right-hand side matrices accordingly (this equation is easy to write and, thus, is omitted here for lack of space).

The purpose of the third step is to identify $H(z)$ up to an orthogonal matrix from $\hat{R}_0$ and from $(\hat{R}_n)_{n=1,\ldots,M}$. This can be

performed by any blind second-order MIMO identification technique that is available for the noiseless case. Now, for simulation purposes, we considered the version of the MIMO linear prediction technique presented in [11] and based on a least square estimate of $H(z)$.

A brief outline of this method is given here for convenience. We begin by assuming that $H(z)$ is irreducible and column reduced. Consequently, the degrees of the columns of $H(z)$ coincide with the Kronecker indices $\{M_i\}_{i=1}^{p}$ of the rational subspace spanned by its columns. The MIMO linear prediction technique relies on the following result. On the one hand, the process y_n^u described in (1) is a moving average process; on the other hand, it is known that for any $K \geq \sum_{i=1}^{p} M_i$, there exists a $q \times q$ polynomial matrix $A(z) = I_q + \sum_{i=1}^{K} A_i z^{-i}$ that satisfies

$$[A(z)]y_n^u = H_0 s_n. \tag{18}$$

This means that y_n^u is also finite-order autoregressive and that its innovation process is $H_0 s_n$. Now, let $D = H_0 H_0^T$ be the covariance matrix of the innovation, and let $\mathcal{R}_K = T_K(H)T_K^T(H)$ be the $q(K+1) \times q(K+1)$ covariance matrix associated with y_n^u. Then, the matrices $\mathcal{A} = [A_1, \ldots, A_K]$ and D are given by the multivariate Yule–Walker equation $[I_q, \mathcal{A}]\mathcal{R}_K = [D, 0]$. The so-called minimum norm solution of this equation is

$$\mathcal{A} = -\mathbf{r}\mathcal{R}_{K-1}^{\dagger}, \qquad D = R_0 + \mathcal{A}\mathbf{r}^T \tag{19}$$

where $()^{\dagger}$ stands for the Moore–Penrose pseudo-inverse, and $\mathbf{r} = [R_1, \ldots, R_K]$. Note that in practice, it is important in these kinds of methods to estimate the dimension of the "signal subspace" (i.e., the rank of $\mathcal{R}_{K-1}$) properly. We have chosen to estimate this dimension as $\arg\min(\hat{\lambda}_{i+1}/\hat{\lambda}_i)$, where $\{\hat{\lambda}_i\}$ is the set of eigenvalues of $\hat{\mathcal{R}}_{K-1}$ arranged in decreasing order.

We still need to compute the unknown transfer function $H(z)$ from $\mathcal{A}$ and D. This can be done only up to some $p \times p$ constant orthogonal matrix, and as is well known, this final indeterminacy cannot be removed by using second-order techniques only. Therefore, let $F_0 F_0^T$ be a factorization of the positive semi-definite matrix D. Then, F_0 is related to H_0 by $F_0 = H_0 U$ for some orthogonal $p \times p$ matrix U. From (1) and (18), we have $A(z)H(z) = H_0$, and thus, $A(z)F(z) = F_0$, where $F(z) = \sum_{k=0}^{M} F_k z^{-k} = H(z)U$. In other words, the coefficients $F_0, F_1, \ldots, F_M$ are the unique solutions of the equation

$$\begin{bmatrix} I_q & & 0 \\ A_1 & \ddots & \\ \vdots & \ddots & I_q \\ & & A_1 \\ A_K & & \vdots \\ & \ddots & \\ 0 & & A_K \end{bmatrix} \begin{bmatrix} F_0 \\ F_1 \\ \vdots \\ F_M \end{bmatrix} = \begin{bmatrix} F_0 \\ 0 \\ \vdots \\ \\ \\ 0 \end{bmatrix} \tag{20}$$

in which the first matrix of the left-hand side is $(K + M + 1)q \times (M+1)q$, and its solution in the least-squares sense gives a consistent estimate of $F(z)$ (see [11] for more details). Note that

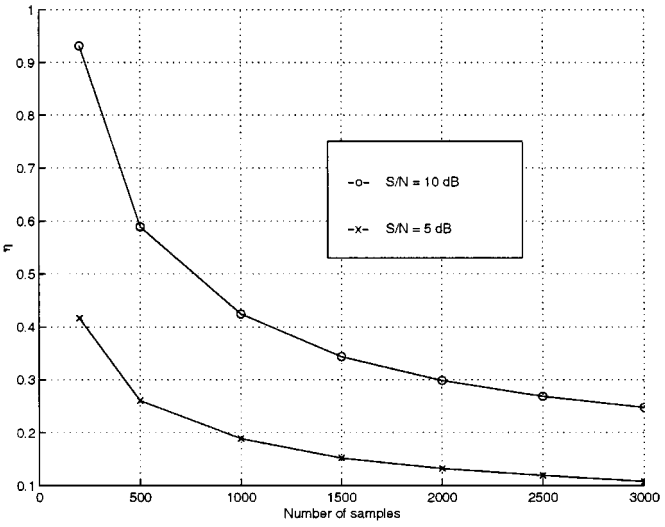

Fig. 1. Noise covariance matrix (Σ) estimation.

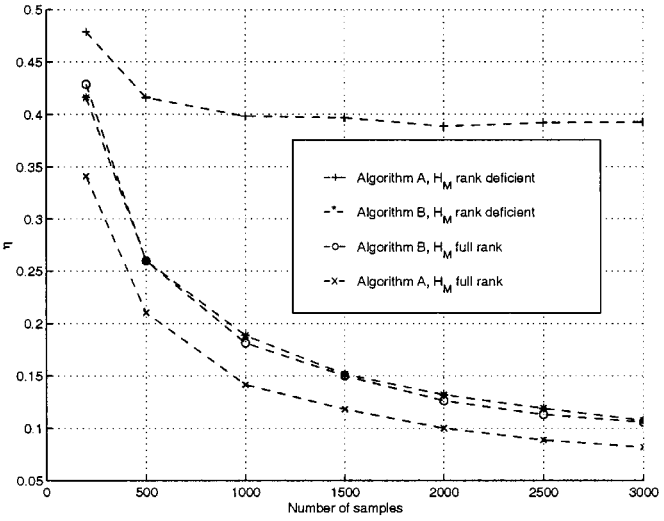

Fig. 2. Effect of Rank(H_M) on the estimation of Σ.

in [11], a weighted least square estimate was proposed. In this paper, we just consider the standard plain least square estimate.

Let us now summarize the discussion. The whole identification procedure consists of the following steps.

1) Compute $\{\hat{Q}_k^{\alpha} = (1 - e^{-ik\alpha\pi})\hat{R}_k\}_{k=1}^{M}$, where $\hat{R}_k$ is an empirical estimate of R_k.
   Compute $T_N(\hat{Q}_\alpha)$ (for N large enough) given in (15), and solve (16) to obtain $\hat{G}(z)$.
2) Compute $\hat{R}_0$ from $\hat{G}(z)$.
3) Build the estimated covariance matrix $\hat{\mathcal{R}}_K$, and estimate $\mathcal{A}$ and D as in (19).
4) Compute a $q \times p$ matrix $\hat{F}_0$ satisfying $\hat{F}_0 \hat{F}_0^T = \hat{D}$.
5) Finally, estimate $H(z)$ (up to a $p \times p$ orthogonal matrix) by solving (20) in the least-squares sense (of course, $\mathcal{A}$ and F_0 have to be replaced by their estimates).

Finally, let us sketch the implementation of algorithm A. From (13), we see that the rows of $\mathcal{G}$ now belong to the left kernel of the $Mq \times 2Mq$ matrix $\mathcal{Q} = [J\mathcal{H}\mathcal{H}^T]$, and all the discussion above can be adapted immediately. The only

difference is that the first step is now replaced by the solution of the minimization problem

$$\begin{cases} \text{minimize} & \text{trace}\left(\left[\hat{G}_0, \ldots, \hat{G}_{M-1}\right]\left(\hat{\mathcal{Q}}\hat{\mathcal{Q}}^T\right)\right. \\ & \left.\cdot\left[\hat{G}_0, \ldots, \hat{G}_{M-1}\right]^T\right) \\ \text{under the constraint} & \hat{G}_0\hat{G}_0^T = I_{q-p} \end{cases} \tag{21}$$

where $\hat{\mathcal{Q}}$ represents a consistent estimate of $\mathcal{Q}$.

V. Experimental Results

In this section, some numerical simulations are performed on algorithms A and B. Two channel models are used. The first model is used to illustrate the theoretical results of the preceding sections. The second model is more suited to the digital communications context, which is one possible application. In both models, the number of sources p is equal to 2, and that of the sensors q is equal to 7. In addition, the covariance matrix of the noise process is chosen as $\Sigma = [\sigma_{ij}]$, with $\sigma_{ij} = (-\mu)^{|i-j|}/(1-\mu^2)$ (it is thus a "stationary spatial AR process of order 1"). The parameter μ is chosen equal to 0.9.

A. Simulations

The first model is used in the first three experiments. We set $M = 4$, and the $qp(M+1) = 70$ unknown entries in $H(z) = \sum_{k=0}^{4} H_k z^{-k}$ are chosen randomly, under a zero-mean unit-variance Gaussian probability law. When needed (i.e., in the second and third experiments), we force H_M to be rank deficient by setting its last column to zero; otherwise, H_M generically has full column rank. The parameter α of Section III-B is chosen equal to 1. The reason why is that in the expression of $Q_1(z)$, all the covariance lags R_k with even indices k are canceled. Therefore, these terms need not be estimated.

Last, in each of the three experiments show later, our results for a given channel are averaged over 50 runs; then, these results are averaged over 20 channels.

In the first experiment, we compute $\hat{\Sigma}$ as $\hat{R}_0^y - \hat{R}_0$. Fig. 1 shows the relative estimation error $\eta = \|\Sigma - \hat{\Sigma}\|_{fro}/\|\Sigma\|_{fro}$ ($\|\cdot\|_{fro}$ denotes the Frobenius norm) versus the samples block size for signal-to-noise ratios equal to 5 and 10 dB. This is a good measure of the performance of the estimate $\hat{R}_0$ of R_0. Note that this relative error is lower for the noise having the largest power; this property is shown here for algorithm B only but, indeed, is shared by both algorithms. It might be justified intuitively by the following observation. Let $S^y(z) = H(z)H^T(z^{-1}) + \Sigma$. Then, from (6), the Σ-identification equation can also be written $G(z)S^y(z) = G(z)\Sigma$; consequently, estimation errors on $G(z)$ affect more sensitively the estimation of Σ when Σ is a matrix with small norm.

In the second experiment, we visualize the effect of the rank of H_M on the identification of Σ. Remember that algorithm A relies heavily on the assumption that H_M has full column rank but that algorithm B does not exploit this property. Fig. 2 also exhibits values of the relative estimation error η. As we can see, algorithm A performs slightly better than algorithm B if the full rank assumption on H_M is satisfied; if not, it does not converge.

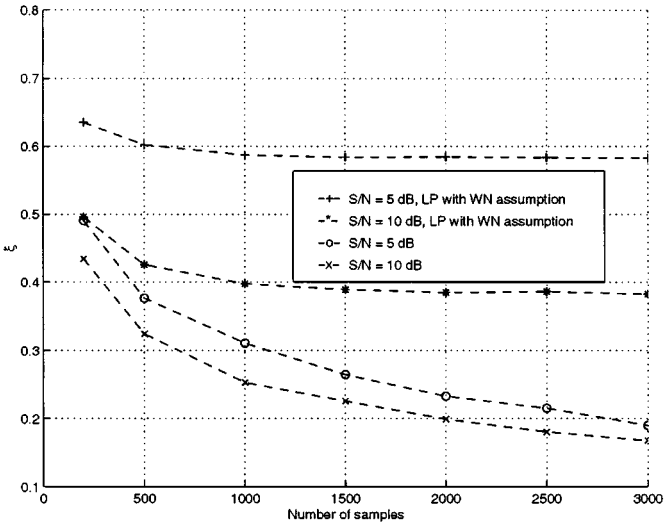

Fig. 3. Transfer function estimation (algorithm A).

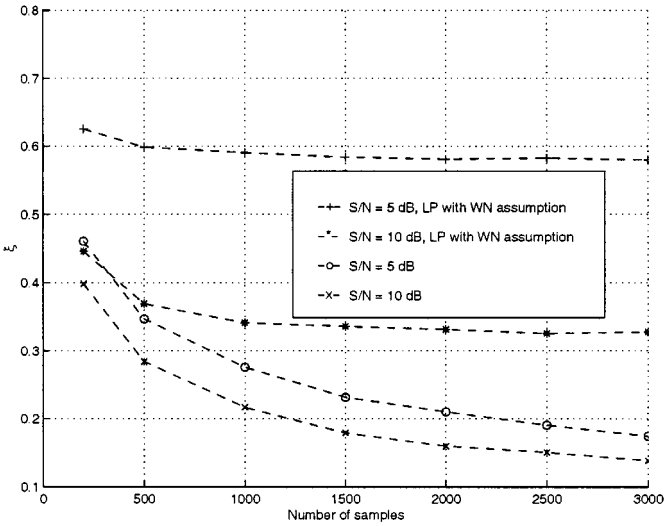

Fig. 4. Transfer function estimation (algorithm B).

On the other hand, simulations show that algorithm B is nearly insensitive to the rank of H_M.

The third experiment measures, for both algorithms, a relative error on $\hat{H}(z)$ after the whole identification process described in Section IV has been performed. This experiment thus combines both the Σ identification procedure, which is specific to this paper, and the effect of the linear prediction algorithm, followed by a least-squares identification. Since $H(z)$ can be identified only up to a $p \times p$ arbitrary orthogonal matrix, the error criterion we choose is

$$\xi = \sqrt{\frac{\int_0^{2\pi} \left\|S(e^{i\omega}) - \hat{S}^{(F)}(e^{i\omega})\right\|_{fro}^2 d\omega}{\int_0^{2\pi} \|S(e^{i\omega})\|_{fro}^2 d\omega}} = \sqrt{\frac{\sum_k \left\|R_k - \hat{R}_k^{(F)}\right\|_{fro}^2}{\sum_k \|R_k\|_{fro}^2}} \tag{22}$$

where $\hat{S}^{(F)}(z) = \hat{F}(z)\hat{F}^T(z^{-1}) = \sum_k \hat{R}_k^{(F)} z^{-k}$ is the covariance function deduced from the estimated transfer function $\hat{F}(z)$.

Fig. 3 (resp. Fig. 4) visualizes ξ as a function of the samples block size in the case of algorithm A with H_M having full column-rank (resp. algorithm B with H_M rank deficient). The error is compared with that obtained with the help of a "classical" linear prediction algorithm, i.e., an algorithm that assumes (wrongly) that $\Sigma = \sigma^2 I_q$. As has been said in the introduction, we begin in this classical algorithm by identifying σ^2 as the smallest eigenvalue of the covariance matrix $\mathcal{R}_N^y$ given in (4); it is obvious that the estimates obtained with such an algorithm do not converge to the exact solution. Indeed, we have observed that the benefit of our algorithms becomes clearer and clearer as Σ departs from $\sigma^2 I$.

As for the second model, we simulate a linear narrowband array of sensors equispaced by a quarter of a wavelength. Both sources transmit a BPSK i.i.d. digital sequence shaped by a raised cosine filter with a roll-off factor of 0.2. In our simulations, the impulse response of this filter is truncated to its main lobe. Signals from each source impinge on the array along six paths. For each path, the angle of arrival is a uniformly distributed variable in the range $[0;\ \pi]$, the amplitude is zero-mean unit-variance Gaussian, and the delay is uniformly distributed between zero and four times the symbol period. Let us write $H(z)$ as $[h_1(z)\, h_2(z)]$. As the received signal is sampled at symbol rate, the maximum value of the transfer function degree is also $M = 4$ here. This model does not generically guarantee the full-column rank condition on H_0 and H_M needed by algorithm A. Suppose, for instance, that for $h_1(z)$, all path delays are less than three times the symbol period. Then, $\deg(h_1(z)) = 3$, and H_4 is obviously rank deficient. Simulations are performed for algorithm B only because algorithm A no longer provides consistent estimates. Here, $\alpha = 0.8$. Our results for a given channel are averaged over 40 runs; then, these results are averaged over ten channels.

The curve of Fig. 5 measures the quality of the estimation of Σ. It depicts η for $S/N = 5$ dB. Fig. 6 shows ξ for $S/N = 5$ dB as well. We have observed that under realistic conditions, the algorithm behaves well in rather adverse conditions (i.e., for low SNRs) but degrades if the SNR increases. This is because for low SNR, the central covariance lag R_0^y of the received signal is very different from that of the useful signal, whereas if the noise is weak, estimating Σ properly becomes difficult, and one should make the approximation $\Sigma = \sigma^2 I$.

B. Applicability of the Method

Let us now make some comments on these results. In this paper, we considered the general problem of second-order blind identification of a system corrupted by a noise, the covariance matrix Σ of which is totally unknown. Since there was no *a priori* assumption on Σ, we were led to ignore the central covariance lag R_0^y. Therefore, we tried to identify the channel from the lags $\{R_k\}_{k>0}$ by relying on the convolutive nature of the channel only.

Now, from a practical point of view, the price to pay for dropping R_0^y is, of course, a severe loss in performance. In fact, the proposed algorithms are pertinent when $R_0^y = R_0 + \Sigma$ cannot be modeled as $R_0^y = R_0 + \sigma^2 I$ without a great error. Therefore, in practice, the method is useful at low SNR and when Σ moves away from a multiple of the identity matrix. If the SNR is high, then estimating Σ consistently is not so important, and indeed, in this case, a classical subspace algorithm, i.e., one which assumes that the noise is spatially white, would be more convenient.

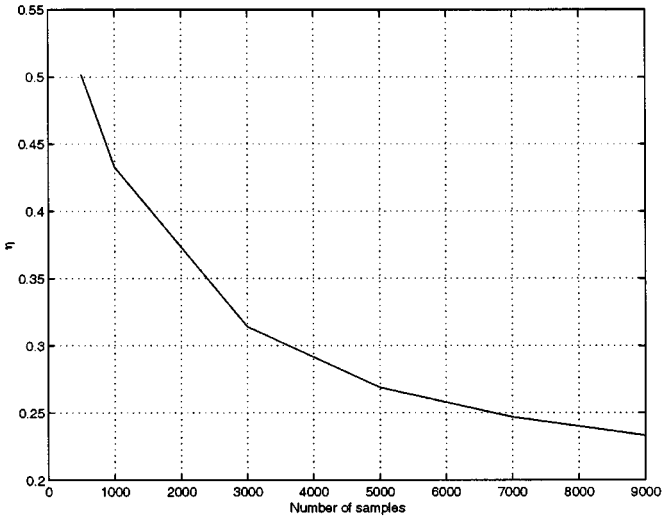

Fig. 5. Noise covariance matrix (Σ) estimation.

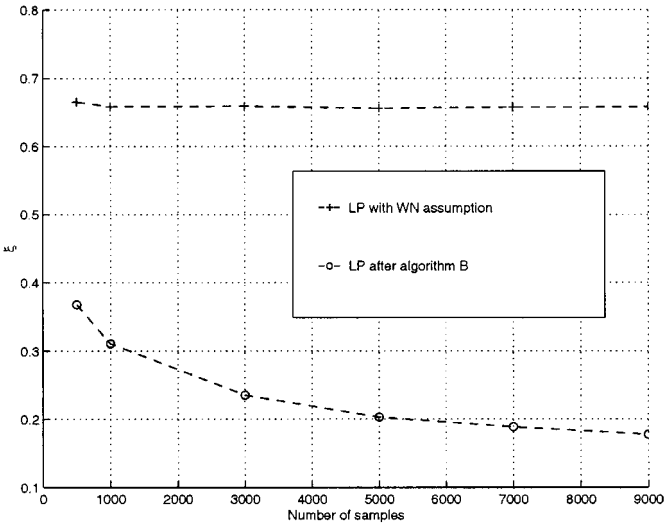

Fig. 6. Transfer function estimation.

This is the reason why a large number of samples is required to estimate the channel properly. Our algorithms need a larger number of samples than those that are usually considered in other related works [8]–[16]. In particular, in all experiments, reasonable results are obtained if the number of samples exceeds 1000 (even more in the second model). This shows that our results are not applicable in the context of highly mobile cooperative systems. However, they can be useful in military applications such as passive listening, in which it is necessary to extract the symbols transmitted by an unknown emitter. In this case, the receiver has, of course, no knowledge of any training sequence, and the identification of the channel has to be performed blindly.

VI. Conclusion

In this paper, we proposed a new subspace-based approach, based on second-order statistics, for identifying a discrete-time noisy MIMO channel up to a constant orthogonal matrix. Due to the specific assumptions on the noise, the problem has been formulated as the identification of the unknown filter $H(z)$ from the truncated auto-covariance sequence $(R_n)_{n\neq 0}$ associated with the spectral density $S(z) = H(z)H^T(z^{-1})$. The method consists of identifying R_0 from $(R_n)_{n\neq 0}$; once R_0 has been identified, the subsequent retrieval of $H(z)$ from the full auto-covariance sequence can be performed by a variety of classical methods and, for example, by the linear prediction approach. Our approach consists of identifying R_0 directly from certain $1 \times q$ polynomials $g(z)$ satisfying $g(z)H(z) = 0$ and, thus, naturally fits in the field of rational subspaces. A general identification lemma is provided in this framework. In practice, computing the filters $g(z)$ can be done either from the block Hankel matrix associated with the sequence $(R_n)_{n\geq 1}$ or from a function such as $S(z) - S(ze^{i\alpha\pi})$, which requires fewer restrictive assumptions on $H(z)$. Finally, practical implementation considerations are addressed, and both algorithms are tested via numerical simulations.

Appendix
Solution of the Constrained Minimization Problem

The minimization problems (16) and (21) are two particular cases of the following problem. Let A and B be two matrices sharing the same number of rows l. Find a matrix ϕ of dimensions $r \times l$, where $r \leq r_B = \text{rank}(B)$, which satisfies simultaneously

$$\chi \stackrel{\text{def}}{=} \text{trace}(\phi AA^T\phi^T) \text{ is minimum} \tag{23}$$

and

$$\phi BB^T\phi^T = I_r. \tag{24}$$

For instance, in (16), $A = T_N(\hat{Q})$, $B = [I_q, 0, \ldots, 0]^T$, $\phi = [\hat{G}_0, \ldots, \hat{G}_N]$, $l = q(N+1)$, and $r = q - 2p$.

We begin with parameterizing the constraint (24). Let

$$BB^T = [U\ V]\begin{bmatrix}\Lambda & \\ & 0\end{bmatrix}\begin{bmatrix}U^T \\ V^T\end{bmatrix}$$

be an orthogonal eigenvalue decomposition of the symmetric semi-definite positive matrix BB^T. In this decomposition, Λ is a diagonal matrix with real positive entries, the matrix $[U\ V]$ is orthogonal, and Λ, U, and V are of dimensions $r_B \times r_B$, $l \times r_B$, and $l \times (l - r_B)$, respectively. As $[U\ V]$ is a regular matrix, it is always possible to set $\phi = [\phi_0\ \phi_1][U\ V]^T$, where ϕ_0 and ϕ_1 are partitioned in conformity with U and V. Let $\Lambda^{1/2}$ be the diagonal square-root of Λ with positive diagonal entries, and let $\Lambda^{-1/2} = (\Lambda^{1/2})^{-1}$. Plugging this expression into (24), ϕ is a solution of (24) if and only if

$$\phi = K_0\Lambda^{-1/2}U^T + \phi_1 V^T \tag{25}$$

for some $r \times r_B$ matrix K_0 satisfying $K_0K_0^T = I_r$.

We still need to specify the two matrices K_0 and ϕ_1. Injecting (25) into (23), we see that the quantity to minimize is

$$\begin{aligned}\chi = \text{trace}\Bigg(&[K_0 \quad \phi_1] \\ &\cdot\begin{bmatrix}\underbrace{\Lambda^{-1/2}U^TAA^TU\Lambda^{-1/2}}_{D_{00}} & \underbrace{\Lambda^{-1/2}U^TAA^TV}_{D_{01}} \\ \underbrace{V^TAA^TU\Lambda^{-1/2}}_{D_{10}} & \underbrace{V^TAA^TV}_{D_{11}}\end{bmatrix} \\ &\cdot\begin{bmatrix}K_0^T \\ \phi_1^T\end{bmatrix}\Bigg) \\ = \text{trace}(&K_0D_{00}K_0^T + \phi_1D_{10}K_0^T \\ &+ K_0D_{01}\phi_1^T + \phi_1D_{11}\phi_1^T).\end{aligned} \tag{26}$$

We will minimize (26) with respect to ϕ_1 for a given K_0. The proper framework for performing this minimization is the inner-product space of $r \times (l - r_B)$ matrices, endowed with the inner product $\langle X, Y\rangle = \text{trace}(X^TY)$. The gradient of χ with respect to ϕ_1 is given by $\nabla_{\phi_1}\chi = 2\phi_1D_{11} + 2K_0D_{01}$. It is then clear that a minimizer of χ is given by $\nabla_{\phi_1}\chi = 0$. From the definitions of D_{10} and D_{11}, we see that

$$\text{Span}(D_{10}) = \text{Span}(D_{01}^T) \subseteq \text{Span}(D_{11}) = \text{Span}(V^TA) \tag{27}$$

where $\text{Span}(\cdot)$ stands for the range space; therefore, the equation $\nabla_{\phi_1}\chi = 0$ is consistent and, thus, always has solutions. These are given by

$$\phi_1 = -K_0D_{01}D_{11}^{\dagger} + W \tag{28}$$

where $D_{11}^{\dagger}$ denotes the Moore–Penrose pseudo inverse of D_{11}, and W is any matrix belonging to $\text{Ker}^l(D_{11})$ [29], [30].

Let us now minimize χ with respect to K_0, taking into account the constraint $K_0K_0^T = I_r$. Plugging $\nabla_{\phi_1}\chi = 0$ into (26), the expression to be minimized is now $\text{trace}(\{K_0(D_{00} - D_{01}D_{11}^{\dagger}D_{10}) + WD_{10}\}K_0^T)$. From (27), we have $\text{Ker}^l(D_{11}) \subset \text{Ker}^l(D_{10})$, and thus, $WD_{10} = 0$; since the choice of W is irrelevant, we can choose the "minimum norm" solution $W = 0$ in (28). Let us now consider the Schur complement matrix $D_{sc} = D_{00} - D_{01}D_{11}^{\dagger}D_{10}$. It can be shown easily that D_{sc} is symmetric semi-definite positive. Therefore, due to a classical result, the matrix K_0 that minimizes $\text{trace}(K_0D_{sc}K_0^T)$ is obtained by arranging the eigenvectors corresponding to the r smallest eigenvalues of D_{sc}. In summary, the procedure is as follows.

- Perform an orthogonal eigenvalue decomposition (EVD) of BB^T to deduce U, V and Λ.
- Compute the matrices D_{00}, D_{01}, and D_{11} defined in (26).
- Compute an orthogonal EVD of D_{sc}, and deduce K_0.
- Compute $\phi_1 = -K_0D_{01}D_{11}^{\dagger}$; the solution is then given by $\phi = K_0\Lambda^{-1/2}U^T + \phi_1V^T$.

REFERENCES

[1] H. Liu, G. Xu, L. Tong, and T. Kailath, "Recent developments in blind channel equalization: From cyclostationarity to subspaces," *Signal Process.*, vol. 50, pp. 83–99, 1996.

[2] L. Tong and S. Perreau, "Multichannel blind identification: From subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, pp. 1951–1968, Oct. 1998.

[3] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds., *Signal Processing Advances in Wireless and Mobile Communications, Vol. 1: Trends in Channel Estimation and Equalization*: Prentice-Hall, 2001.

[4] G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds., *Signal Processing Advances in Wireless and Mobile Communications, Vol. 2: Trends in Single- and Multi-User Systems*. Englewood Cliffs, NJ: Prentice-Hall, 2001.

[5] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Proc. 25th Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, 1991, pp. 856–860.

[6] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.

[7] D. T. M. Slock, "Blind fractionally-spaced equalization, perfect reconstruction filter banks and multichannel linear prediction," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, Adelaide, Australia, May 1994, pp. 585–588.

[8] K. Abed Meraim and Y. Hua, "Blind identification of multi-input multi-output system using minimum noise subspace," *IEEE Trans. Signal Processing*, vol. 45, pp. 254–258, Jan. 1997.

[9] K. Abed Meraim, Ph. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Trans. Inform. Theory*, vol. 43, pp. 499–511, Mar. 1997.

[10] L. Deneire and D. T. M. Slock, "A Schur method for multiuser multichannel blind identification," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Phoenix, AZ, 1999, pp. 2905–2908.

[11] A. Gorokhov and Ph. Loubaton, "Blind identification of MIMO-FIR systems: A generalized linear prediction approach," *Signal Process.*, vol. 73, pp. 105–124, 1999.

[12] H. Liu and G. Xu, "Closed form blind symbol estimation in digital communications," *IEEE Trans. Signal Processing*, vol. 43, pp. 2714–2723, Nov. 1995.

[13] D. T. M. Slock, "Blind joint equalization of multiple synchronous mobile users using oversampling and/or multiple antennas," in *Proc. Asilomar Conf. Signals, Syst., Comput.*, Pacific Grove, CA, Nov. 1994, pp. 1154–1158.

[14] J. K. Tugnait and B. Huang, "Blind channel estimation and equalization of multiple input multiple output channels," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Phoenix, AZ, Mar. 1999, pp. 2707–2710.

[15] A. J. Van der Veen, S. Talwar, and A. Paulraj, "A subspace approach to blind space time processing for wireless communication systems," *IEEE Trans. Signal Processing*, vol. 45, pp. 173–190, Jan. 1997.

[16] J. Xavier, V. Barroso, and J. M. F. Moura, "Closed-form blind identification of MIMO channels," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Seattle, WA, 1998, pp. 3165–3168.

[17] J. K. Tugnait, "Blind spatio-temporal equalization and impulse response estimation for MIMO channels using a godard cost function," *IEEE Trans. Signal Processing*, vol. 45, pp. 268–271, Jan. 1997.

[18] M. Pesavento and A. B. Gershman, "Maximum-likelihood direction-of-arrival estimation in the presence of unknown nonuniform noise," *IEEE Trans. Signal Processing*, vol. 49, pp. 1310–1324, July 2001.

[19] W. S. Burdic, *Underwater Acoustic System Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1984.

[20] J. P. Le Cadre, "Parametric methods for spatial signal processing in the presence of unknown colored noise fields," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 37, pp. 965–983, July 1989.

[21] B. Friedlander and A. J. Weiss, "Direction finding using noise covariance modeling," *Signal Process.*, vol. 43, pp. 1557–1567, 1995.

[22] I. Fijalkow and P. Loubaton, "Identification of rank one rational spectral densities from noisy observations: A stochastic realization approach," *Syst. Contr. Lett.*, vol. 24, pp. 201–205, 1995.

[23] K. Abed Meraim, Y. Hua, P. Loubaton, and E. Moulines, "Subspace method for blind identification of multichannel FIR systems in noise field with unknown spatial covariance," *IEEE Signal Processing Lett.*, vol. 4, pp. 135–137, May 1997.

[24] F. Desbouvries, I. Fijalkow, and P. Loubaton, "On the identification of noisy MA models," *IEEE Trans. Automat. Contr.*, vol. 41, pp. 1810–1814, Dec. 1996.

[25] F. Desbouvries and P. Loubaton, "On the identification of certain rational transfer functions from truncated autocovariance sequences," *IEEE Trans. Automat. Contr.*, vol. 44, pp. 1078–1081, May 1999.

[26] P. Vandaele and M. Moonen, "A stochastic subspace algorithm for blind channel identification in noise fields with unknown spatial color," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, Phoenix, AZ, Mar. 1999, pp. 2619–2622.

[27] G. D. Forney, "Minimal bases of rational vector spaces, with applications to multivariable linear systems," *SIAM J. Contr.*, vol. 13, pp. 493–520, 1975.

[28] T. Kailath, *Linear Systems*. Englewood Cliffs, NJ: Prentice-Hall, 1980.

[29] C. R. Rao and S. K. Mitra, *Generalized Inverses of Matrices and Its Applications*. London, U.K.: Wiley, 1971.

[30] A. Bjorck, *Numerical Methods for Least Squares Problems*. Philadelphia, PA: SIAM, 1996.

Walid Hachem (M'00) was born in Bhamdoun, Lebanon, in 1967. He received the engineering degree in telecommunications from St. Joseph University (ESIB), Beirut, Lebanon, in 1989 and the masters degree in signal processing from Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France, in 1990. He received the Ph.D. degree in signal processing from Université de Marne-la-Vallée, Champs sur Marne, France, in 2000.

He was with Philips T.R.T., working on telephone modems, and then joined FERMA, working on DSP processing for voice servers. He is currently an Associate Professor with the Ecole Supérieure d'Electricité, Paris. His research interests concern blind estimation, performance evaluation of multiple access techniques, and multiple antenna wireless communications.

François Desbouvries (M'01) received the Engineer degree in telecommunications in 1987 and the Ph.D. degree in signal processing in 1991, both from the Ecole Nationale Supérieure des Télécommunications (ENST), Paris, France. He received the Habilitation à Diriger des Recherches (HDR) degree from the Université de Marne-la-Vallée, Champs sur Marne, France, in 2001.

Since 1991, he has been an Associate Professor with the Institut National des Télécommunications, Evry, France. His research interests are in statistical signal processing and digital communications.

Philippe Loubaton (M'88) was born in 1958 in Villers Semeuse, France. He received the M.Sc. and the Ph.D. degrees from Ecole Nationale Supérieure des Télécommunications, Paris, France, in 1981 and 1988, respectively.

From 1982 to 1986, he was Member of Technical Staff with Thomson-CSF/RGS, where he worked in digital communications. From 1986 to 1988, he was with the Institut National des Télécommunications as an Assistant Professor of electrical engineering. In 1988, he joined the Ecole Nationale Supérieure des Télécommunications, Paris, in the Signal Processing Department. Since 1995, he has been Professor of electrical engineering at Marne-la-Vallée University, Champs sur Marne, France. From 1996 to 2000, he was director of the Laboratoire Système de Communication of Marne la Vallée University and is now a member of the Laboratoire Traitement et Communication de l'Information (CNRS/Ecole Nationale Supérieure des Télécommunications). His present research interests are in statistical signal processing and digital communications with a special emphasis on blind equalization, multiuser communication systems, and multicarrier modulations.

Dr. Loubaton is currently an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and IEEE COMMUNICATIONS LETTERS and is a member of the IEEE Signal Processing for Communications Technical Committee.