Closed-Form Correlative Coding (CFC_2) Blind Identification of MIMO Channels: Isometry Fitting to Second Order Statistics

João Xavier, *Student Member, IEEE*, Victor A. N. Barroso, *Senior Member, IEEE*, and José M. F. Moura, *Fellow, IEEE*

***Abstract*—We present a blind *closed-form* consistent channel estimator for multiple-input multiple-output (MIMO) systems that uses only *second-order statistics*. We spectrally modulate the output of each source by correlative coding it with a distinct filter. The correlative filters are designed to meet the following desirable characteristics: No additional power or bandwidth is required; no synchronization between the sources is needed; the original data rate is maintained. We first prove an identifiability theorem: Under a simple spectral condition on the transmitted random processes, the MIMO channel is uniquely determined, up to a phase offset per user, from the second-order statistics of the received data. We then develop the closed-form algorithm that attains this identifiability bound. We show that minimum-phase finite impulse response filters with arbitrary memory satisfy our sufficient spectral identifiability condition. This results in a computationally attractive scheme for retrieving the data information sequences after the MIMO channel has been identified. We assess the performance of the proposed algorithms by computer simulations. In particular, the results show that our technique outperforms the recently introduced transmitter-induced conjugate cyclostationarity approach when there are carrier frequency misadjustments.**

I. Introduction

BLIND channel identification is an active topic of research. It is a bandwidth-attractive scheme for automatic link re-establishment that requires no training sessions and finds widespread application in the expanding field of mobile wireless communications. In this paper, we focus on closed-form blind channel identification based only on the second-order statistics (SOS) of the observations.

A. Monouser Context

For single-input multiple-output (SIMO) systems, the work by Tong *et al.* [1]–[3] was a major achievement. They exploit the information conveyed by certain correlation matrices of the channel outputs to derive an analytical, i.e., closed-form, solution for the unknown channel coefficients by the method of moments. See also [4] for a distinct alternative closed-form subspace-based method.

Manuscript received August 9, 1999; revised January 8, 2001. This work was supported in part by NATO Collaborative Research Grant (CRG.971184) 1202/97/JARC-501. The associate editor coordinating the review of this paper and approving it for publication was Prof. Michail K. Tsatsanis.

J. Xavier and V. A. N. Barroso are with the Instituto Superior Técnico—Instituto de Sistemas e Robótica, Lisbon, Portugal (e-mail: jxavier@isr.ist.utl.pt; vab@isr.ist.utl.pt).

J. M. F. Moura is with the Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213 USA (e-mail: moura@ece.cmu.edu).

Publisher Item Identifier S 1053-587X(01)03337-2.

B. Multiuser Context

Gorokhov and Loubaton [5] (see also [6]) generalize the subspace-based method for the multiple user situation. These references show that by exploiting only SOS, the multiple-input multiple-output (MIMO) channel can be recovered up to a block diagonal constant matrix. In equivalent terms, the convolutive mixture is converted into several instantaneous, or static, mixtures of the input signals, two sources being in the same group if and only if they share the same channel degree. In particular, the convolutive mixture is completely resolved if all users are exposed to distinct system orders, i.e., they exhibit memory diversity. The blind SOS-based whitening approach in [7]–[9] also converts a convolutive mixture into a static one, with a substantial weakening on the channel assumptions: Infinite-impulse response (IIR) channels can be accommodated, as well as minimum-phase common zeros among the subchannels; further, the usual column-reduced condition can be dropped. Still, these approaches do not resolve the residual static mixtures.

To resolve these residual static mixtures, one may employ one of several blind source separation (BSS) techniques, depending mainly on the characteristics of the sources, but also depending on the number of available samples, and the signal-to-noise ratio (SNR). Examples of these BSS techniques include

i) high-order statistics (HOS) approaches, e.g., the joint-diagonalization closed-form procedure in [10], which are feasible for non-Gaussian sources (although estimates of cumulants converge slower than SOS [3]);
ii) the analytical constant modulus algorithm (ACMA) [11], which provides a closed-form separation solution for constant modulus (CM) sources;
iii) separation of finite-alphabet (FA) sources, which may be tackled by locally-convergent iterative algorithms, see [12]–[18].

Chevreuil and Loubaton [20], [21] recently introduced the transmitter induced conjugate cyclostationarity (TICC) method. They provide a complete closed-form SOS-based solution for the MIMO channel with no extra BSS algorithmic step required. A distinct conjugate cyclic frequency per user is induced at the transmitter. This data preprocessing is then exploited at the receiver to reduce the problem to several SIMO channel estimation problems, which can then be solved by the subspace method. The main drawback of TICC is its high sensitivity to carrier frequency misadjustments. In multiuser scenarios, this distortion may be expected to appear in the baseband demodulated signals, as the receiver has to synchronize its local oscillator simultaneously with multiple independently generated carrier frequencies.

C. Contribution

In this paper, we propose a closed-form method for the blind identification of MIMO channels based only on SOS. As in the TICC approach [20], we require no additional BSS step. When compared with the TICC methodology, our channel assumptions are more restrictive since we assume, as in most approaches [5], [6], [18], that the transfer function is irreducible and column-reduced, but we gain robustness to baseband carrier phase drifts.

Our method uses spectral diversity at the sources, attained by correlative coding. We utilize correlative filters at the transmitters to assign distinct spectral patterns to the random messages emitted by the sources. We prove an identifiability theorem that establishes sufficient conditions on the correlative filters to ensure uniqueness of the MIMO channel matrix from the SOS of the channel outputs. We derive from this theoretical result a closed-form algorithm that blindly identifies the channel matrix by matching the theoretical and observed SOS.

Belouchrani *et al.* [19] derive an SOS-based algorithm for separation of static or instantaneous mixtures of sources with distinct spectra. Concatenating the techniques in [5] and [19] can then solve the convolutive mixture problem. In contrast, our work ensures the uniqueness of the channel matrix for general convolutive mixtures and not just instantaneous ones like in [19]; in addition, we recover the conditions in [19] as a special case from our theorem. Further, we provide an integrated analytical solution to resolve the convolutive mixture problem. Because the spectra of the sources are known, we can parallelize our solution, which is a fact that can be exploited in practical wireless systems.

D. Relation with Previous Work

We introduced the correlative framework in [22]. In that paper, the closed-form solution relies on a certain *quasi*-diagonal structure of the sources's correlation matrices. We obtain these by restricting the correlative filters to those that satisfy a *minimal memory length* condition: loosely, the channel order of each correlative filter is assumed to exceed the degree of intersymbol interference (ISI) experienced by each user. This condition imposes a significant lower bound on the computational complexity of the Viterbi decoding algorithm as the number of states in the trellis diagram increases with the order of the correlative filters. In this paper, we drop the *quasi*-diagonal property, which makes feasible correlative filters with arbitrary nonzero degree. Thus, the computational complexity of the Viterbi decoding step is significantly reduced. In fact, we prove that minimum-phase finite impulse response (FIR) filters of arbitrary nonzero degree can fulfill the requirements of the identifiability theorem. This allows for the direct inversion of the filters and leads to a simpler scheme to decode the original data sequences that may have phase-tracking capability. Since the sources's autocorrelation matrices do not have the *quasi*-diagonal structure, the method in [22] no longer applies. We develop here a new consistent closed form estimator for the MIMO channel.

E. Paper Organization

Section II establishes the data model and states our main assumptions. Section III presents the correlative framework for multiuser blind channel identification. We introduce the identifiability theorem and show that the requirements of the theorem are satisfied by minimum-phase FIR filters of arbitrary nonzero degree. Section IV describes the closed-form algorithm that estimates the MIMO channel up to a phase offset. Section V discusses a simple iterative technique for decoding the original symbol sequences once the MIMO channel is identified. This technique jointly tracks residual phase drifts in the baseband signals and reduces the mean-square error (MSE) of the closed-form channel estimate. Section VI evaluates the performance of our closed-form correlative coding (CFC_2) approach. We compare it with the transmitter induced conjugate cyclostationarity (TICC) approach in [20]. Our simulation results show that CFC_2 yields symbol estimates with lower probability of error than TICC in the presence of carrier frequency asynchronisms. Section VII concludes the paper.

F. Notation

$\mathbb{N}$, $\mathbb{Z}$, $\mathbb{R}$, and $\mathbb{C}$ denote the set of natural, integer, real, and complex numbers, respectively. Matrices (uppercase) and (column/row) vectors are in boldface type. $\mathbb{C}^{n\times m}$ and $\mathbb{C}^n$ denote the set of $n \times m$ matrices and the set of n-dimensional column vectors with complex entries, respectively. The notations $(\cdot)^{\mathrm{T}}$, $(\cdot)^*$, $(\cdot)^\dagger$, and tr $(\cdot)$ stand for the transpose, the Hermitean, the Moore–Penrose pseudo-inverse, and the trace operator, respectively; $\|\boldsymbol{A}\| = \sqrt{\mathrm{tr}\,(\boldsymbol{A}^*\boldsymbol{A})}$ denotes the Frobenius norm. The symbols $\boldsymbol{I}_n$, $\boldsymbol{0}_{n\times m}$, and $\boldsymbol{J}_n$ stand for the $n \times n$ identity, the $n \times m$ all-zero, and the $n \times n$ forward-shift (ones in the first lower diagonal) matrices, respectively. When the dimensions are clear from the context, the subscripts are dropped. For $k \in \mathbb{Z}$, we set $\boldsymbol{K}_n(k) = \boldsymbol{J}_n^k$, if $k \geq 0$, and $\boldsymbol{K}_n(k) = (\boldsymbol{J}_n^{\mathrm{T}})^{-k}$, if $k < 0$. The direct sum or diagonal concatenation of matrices is represented by diag $(\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_m)$; for $\boldsymbol{A} \in \mathbb{C}^{n\times m}$, vec $(\boldsymbol{A}) \in \mathbb{C}^{nm}$ consists of the columns of $\boldsymbol{A}$ stacked from left to right, and $\otimes$ represents the Kronecker product. For $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, $\sigma(\boldsymbol{A}) = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ denotes its spectrum, i.e., the set of its eigenvalues (including multiplicities). The set of polynomials with coefficients in $\mathbb{C}$ and indeterminate z^{-1} is denoted by $\mathbb{C}[z]$. The polynomial $f(z) = \sum_{k=0}^{d} f(k)z^{-k} \in \mathbb{C}[z]$ is said to have degree d, written deg $f(z) = d$, if $f(d) \neq 0$. The degree of the zero polynomial is not defined. The subsets of all polynomials with degree d and degree at most d are denoted by $\mathbb{C}^n_{(d)}[z]$ and $\mathbb{C}^n_d[z]$, respectively. Similar definitions hold for $\mathbb{C}^n[z]$ and $\mathbb{C}^{n\times m}[z]$, which are the set of $n \times 1$ polynomial vectors and $n \times m$ polynomial matrices, respectively. We will identify $\mathbb{C}^n_d[z]$ and $\mathbb{C}^{n(d+1)}$ by tacitly associating with $\boldsymbol{f}(z) = \sum_{k=0}^{d} \boldsymbol{f}(k)z^{-k}$ the vector $\boldsymbol{f} = (\boldsymbol{f}(0)^{\mathrm{T}}, \boldsymbol{f}(1)^{\mathrm{T}}, \ldots, \boldsymbol{f}(d)^{\mathrm{T}})^{\mathrm{T}}$. For $\boldsymbol{f}(z) =$

$\sum_{k=0}^{d} \boldsymbol{f}(k) z^{-k} \in \mathbb{C}_{(d)}^{n}[z]$ and $N \in \mathbb{N}$, we define the $n(N+1) \times (d+1+N)$ block-Sylvester matrix

$$\mathcal{T}_N(\boldsymbol{f}) = \begin{bmatrix} \boldsymbol{f}(0) & \cdots & \boldsymbol{f}(d) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{f}(0) & \cdots & \boldsymbol{f}(d) & \ddots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \boldsymbol{f}(0) & \cdots & \boldsymbol{f}(d) \end{bmatrix}.$$

Polynomial matrices $\boldsymbol{F}(z) \in \mathbb{C}_d^{n \times m}[z]$ operate on (input) signals $\{\boldsymbol{x}(t) \in \mathbb{C}^m : t \in \mathbb{Z}\}$, yielding (output) signals $\{\boldsymbol{y}(t) \in \mathbb{C}^n : t \in \mathbb{Z}\}$

$$\boldsymbol{y}(t) = [\boldsymbol{F}(z)]\boldsymbol{x}(t) \;\Leftrightarrow\; \boldsymbol{y}(t) = \sum_{k=0}^{d} \boldsymbol{F}(k)\boldsymbol{x}(t-k).$$

For $N \in \mathbb{N}$, the Nth-*stacking* operator $\mathcal{S}_N[\cdot]$ is defined as

$$\begin{aligned} \boldsymbol{y}(t) =& \mathcal{S}_N[\boldsymbol{x}(t)] \Leftrightarrow \;\; \boldsymbol{y}(t) = \big(\boldsymbol{x}(t)^{\mathrm{T}}, \boldsymbol{x}(t-1)^{\mathrm{T}}, \ldots \\ & \boldsymbol{x}(t-N)^{\mathrm{T}}\big)^{\mathrm{T}} \end{aligned}$$

where $\boldsymbol{x}(t)$ denotes a vector (or scalar) signal. Additional notation is introduced as needed.

II. Data Model and Problem Statement

Consider a noisy linear MIMO channel with P inputs, $s_p(t)$, $p = 1, 2, \ldots, P$, and M outputs collected in the output vector $\boldsymbol{y}(t)$:

$$\begin{aligned} \boldsymbol{y}(t) =& (y_1(t), y_2(t), \ldots, y_M(t))^{\mathrm{T}} && (1) \\ =& \sum_{p=1}^{P} [\boldsymbol{h}_p(z)]\, s_p(t) + \boldsymbol{w}(t) \\ =& \sum_{p=1}^{P} \left\{ \sum_{k=0}^{D_p} \boldsymbol{h}_p(k) s_p(t-k) \right\} + \boldsymbol{w}(t). && (2) \end{aligned}$$

Here, $\boldsymbol{h}_p(z) \in \mathbb{C}_{(D_p)}^{M}[z]$ denotes the $\mathcal{Z}$-transform of the multichannel filter corresponding to the pth user, and $\boldsymbol{w}(t) \in \mathbb{C}^M$ represents additive noise. For example, in spatial division multiple access (SDMA) networks for wireless mobile communications, several users impinge on an antenna array through a multipath space-time channel. In this scenario, (2) models the measured baseband array snapshots, and $\boldsymbol{h}_p(k)$ denotes the user multipath spatial signatures.

Problem Statement (Blind MIMO Channel Identification): Given the SOS of the observed output data samples $\boldsymbol{y}(t)$, i.e., the set of correlation matrices

$$\mathcal{R}_{\boldsymbol{y}} = \{\boldsymbol{R}_{\boldsymbol{y}}(\tau) : \tau \in \mathbb{Z}\}, \quad \boldsymbol{R}_{\boldsymbol{y}}(\tau) = \mathrm{E}\{\boldsymbol{y}(t)\boldsymbol{y}(t-\tau)^*\}$$

find the $M \times P$ MIMO polynomial matrix $\boldsymbol{\mathcal{H}}(z) = [\boldsymbol{h}_1(z)\boldsymbol{h}_2(z) \cdots \boldsymbol{h}_P(z)]$. □

Finding $\boldsymbol{\mathcal{H}}(z)$ amounts to finding

$$\begin{aligned} \boldsymbol{\mathcal{H}} =& [\boldsymbol{\mathcal{H}}_1 \boldsymbol{\mathcal{H}}_2 \ldots \boldsymbol{\mathcal{H}}_P], \boldsymbol{\mathcal{H}}_p \\ \equiv& \mathcal{T}_0(\boldsymbol{h}_p) = [\boldsymbol{h}_p(0)\boldsymbol{h}_p(1) \ldots \boldsymbol{h}_p(D_p)]. \end{aligned}$$

In the sequel, we will work with *stacked* (or *smoothed*) observations $\boldsymbol{y}_N(t)$

$$\begin{aligned} \boldsymbol{y}_N(t) =& \mathcal{S}_N[\boldsymbol{y}(t)] \\ =& \sum_{p=1}^{P} \boldsymbol{H}_p \boldsymbol{s}_p(t) + \boldsymbol{w}_N(t) = \boldsymbol{H}\boldsymbol{s}(t) + \boldsymbol{w}_N(t) \quad (3) \end{aligned}$$

where

$$\begin{aligned} \boldsymbol{H}_p =& \boldsymbol{\mathcal{T}}_N(\boldsymbol{h}_p) \\ \boldsymbol{H} =& [\boldsymbol{H}_1 \boldsymbol{H}_2 \ldots \boldsymbol{H}_P] \\ \boldsymbol{s}_p(t) =& \mathcal{S}_{D_{p+N}}[s_p(t)] \\ s(t) =& \big(s_1(t)^{\mathrm{T}}, \ldots, s_P(t)^{\mathrm{T}}\big)^{\mathrm{T}} \\ \boldsymbol{w}_N(t) =& \mathcal{S}_N[\boldsymbol{w}(t)]. \end{aligned}$$

Clearly, all the desired information, i.e., the entries of $\boldsymbol{\mathcal{H}}$, are contained in $\boldsymbol{H}$.

The following conditions are assumed:

A1) **Number of sources P and number of outputs M:** The number P of sources is known, and there are more outputs than sources $M > P$.

A2) **Irreducibility and column reducedness:** The MIMO polynomial matrix $\boldsymbol{\mathcal{H}}(z)$ satisfies three properties:

i) irreducible, i.e., rank $\boldsymbol{\mathcal{H}}(z) = P$, for all $z \in \mathbb{C} \setminus \{0\} \cup \{\infty\}$;

ii) column-reduced, i.e., rank $[\boldsymbol{h}_1(D_1)\boldsymbol{h}_2(D_2) \cdots \boldsymbol{h}_P(D_P)] = P$;

iii) the column degrees of $\boldsymbol{\mathcal{H}}(z)$ are uniformly upper bounded by some known integer $D_{\max}$, i.e., $0 \leq D_p \leq D_{\max}$, for $p = 1, 2, \ldots, P$.

As a consequence of **i)** and **ii)**, the MIMO channel matrix $\boldsymbol{H}$ in (3) is full column rank if N, which is the number of stacked observations, is greater than a certain integer. For example, the integer $D_1 + \cdots + D_P - 1$ is a lower bound, see [5] for a proof and a more detailed discussion on this topic. We assume in the sequel that $\boldsymbol{H}$ is full column rank and that $N \leq N_{\max}$ for some known $N_{\max}$.

A3) **Stationarity:** The sources $s_p(t)$, $p = 1, 2, \ldots, P$, and the noise $\boldsymbol{w}(t)$ are zero-mean wide-sense stationary processes, uncorrelated with each other. The noise correlation matrices $\{\boldsymbol{R}_{\boldsymbol{w}}(\tau) : \tau \in \mathbb{Z}\}$ are known, and without loss of generality, the sources have unit power,

$$r_{s_p}(0) = \mathrm{E}\{|s_p(t)|^2\} = 1, \quad p = 1, 2, \ldots, P.$$

III. Correlative Filtering

In digital communication systems, it is commonly assumed that each user transmits a spectrally white data stream, i.e.,

$$r_{s_p}(\tau) = \mathrm{E}\{s_p(t)s_p(t-\tau)^*\} = \delta(\tau),$$

where $\delta(\tau)$ denotes the Kronecker delta. As an example, take the usual scenario where $s_p(t)$ denotes an independent identically distributed (i.i.d.) sequence of information symbols drawn from a finite alphabet set $\mathcal{A} \subset \mathbb{C}$ like the binary alphabet $\mathcal{A}_{\mathrm{BSK}} = \{\pm 1\}$. The emitted signals $s_p(t)$ are indistinguishable from a statistical viewpoint. Their power spectral densities share the same flat pattern. Due to this spectral symmetry, the MIMO polynomial matrix $\boldsymbol{\mathcal{H}}(z)$ cannot, in general, be unambiguously determined from the SOS $\mathcal{R}_{\boldsymbol{y}}$. For a simple illustrative example, consider the two-users case $P=2$ and that $D_1 = D_2 = D$ for some $D \in \mathbb{N}$. In addition, for further reference, let $\mathcal{R}_{\boldsymbol{y}}(\boldsymbol{\mathcal{A}}(z))$ denote the SOS of $\boldsymbol{y}(t)$ given that $\boldsymbol{\mathcal{H}}(z) = \boldsymbol{\mathcal{A}}(z)$

$$\mathcal{R}_{\boldsymbol{y}}(\boldsymbol{\mathcal{A}}(z)) = \{\boldsymbol{R_y}(\tau) : \tau \in \mathbb{Z} \mid \boldsymbol{\mathcal{H}}(z) = \boldsymbol{\mathcal{A}}(z)\}.$$

It is readily seen that the two MIMO polynomial matrices $\boldsymbol{\mathcal{A}}(z) = [\boldsymbol{a}_1(z)\boldsymbol{a}_2(z)]$, where $\boldsymbol{a}_p(z) \in \mathbb{C}^M_{(D)}[z]$, $p = 1, 2$, and

$$\boldsymbol{\mathcal{B}}(z) = \boldsymbol{\mathcal{A}}(z)\boldsymbol{Q}, \quad \boldsymbol{Q} = \frac{1}{\sqrt{2}}\begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}$$

induce the same output correlation matrices, i.e., $\mathcal{R}_{\boldsymbol{y}}(\boldsymbol{\mathcal{A}}(z)) = \mathcal{R}_{\boldsymbol{y}}(\boldsymbol{\mathcal{B}}(z))$. Note also that

$$\boldsymbol{\mathcal{B}}(z) = \left[\frac{1}{\sqrt{2}}(\boldsymbol{a}_1(z) + \boldsymbol{a}_2(z)) \;\; \frac{1}{\sqrt{2}}(-\boldsymbol{a}_1(z) + \boldsymbol{a}_2(z))\right]$$

is useless for source separation purposes, as the two users are mixed together.

In conclusion, blind identification of $\boldsymbol{\mathcal{H}}(z)$ from $\mathcal{R}_{\boldsymbol{y}}$ is not a well-posed problem when all the input signals $s_p(t)$ are spectrally white. This motivates us to consider correlative coding at the source. Within this framework, the pth user, rather than transmitting a white data stream, say $a_p(t)$, emits the output of a correlative filter

$$\begin{aligned} s_p(t) =& [c_p(z)]a_p(t) \\ c_p(z) =& c(0) + c(1)z^{-1} + \cdots + c(C_p)z^{-C_p} \in \mathbb{C}_{(C_p)}[z]. \end{aligned} \quad (4)$$

To avoid increasing the transmitted power, the correlative filter is normalized to unit-norm

$$||\boldsymbol{c}_p|| = 1, \quad \boldsymbol{c}_p = (c(0), c(1), \ldots, c(C_p))^{\mathrm{T}}.$$

We will show in Theorem 1 below that the uniqueness of $\boldsymbol{\mathcal{H}}(z)$ up to a phase offset per user from $\mathcal{R}_{\boldsymbol{y}}$ is guaranteed by a certain spectral diversity condition on the correlation matrices of the colored signals $s_p(t)$.

First, we need some additional notation. For $M_p \in \mathbb{N}$, set $\boldsymbol{s}_p(t; M_p) = \mathcal{S}_{M_p}[s_p(t)]$. Since, in (4), $a_p(t)$ denotes a white sequence, i.e., $r_{a_p}(\tau) = \delta(\tau)$, the correlation matrices of $\boldsymbol{s}_p(t; M_p)$, hereafter $\boldsymbol{R}_{s_p}(\tau; M_p)$, are given by

$$\boldsymbol{R}_{s_p}(\tau; M_p) = \boldsymbol{\mathcal{T}}_{M_p}(c_p)\boldsymbol{K}_{M_p+C_p+1}(\tau)\boldsymbol{\mathcal{T}}_{M_p}(c_p)^* \quad (5)$$

where the shift matrix $\boldsymbol{K}_n(k)$ was defined previously for generic $n \in \mathbb{N}$ and $k \in \mathbb{Z}$. We formally state the spectral diversity condition.

A4) **Spectral Diversity:** The P users correlate their white information sequences $a_p(t)$ so that for each source $p = 1, 2, \ldots, P$, there is a correlation lag $\tau = \tau(p)$ such that

$$\sigma(\boldsymbol{A}_p(\tau; M_p)) \cap \sigma(\boldsymbol{A}_q(\tau; M_q)) = \emptyset$$

holds for every $q \in \{1, 2, \ldots, P\} \setminus \{p\}$ and every $0 \le M_p, M_q \le D_{\max} + N_{\max} + C_{\max} + 1$, $C_{\max} \equiv \max\{C_1, C_2, \ldots, C_P\}$; here

$$\boldsymbol{A}_p(\tau; M_p) = \boldsymbol{R}_{\boldsymbol{s_p}}(0; M_p)^{-1/2}\boldsymbol{R}_{\boldsymbol{s_p}}(\tau; M_p)\boldsymbol{R}_{\boldsymbol{s_p}}(0; M_p)^{-1/2} \quad (6)$$

denotes the normalized correlation matrix of $\boldsymbol{s}_p(t; M_p)$ at lag τ. In addition, the correlative filters $c_p(z)$, $p = 1, 2, \ldots, P$ are known by the receiver. Assumption A4) means that for each source $p \in \{1, 2, \ldots, P\}$, there must exist one correlation lag $\tau(p)$ that separates the spectra of $\boldsymbol{A}_p(\tau(p); M_p)$ from $\boldsymbol{A}_q(\tau(p); M_q)$, $q \neq p$, i.e., from the remaining sources, irrespective of the stacking parameters M_p, M_q taken in $\{0, 1, \ldots, D_{\max} + N_{\max} + C_{\max} + 1\}$. We now state the identifiability theorem.

Theorem 1 [Identifiability] : Consider the signal model in (2), and assume that conditions A1)-A4) are fulfilled. Let $\boldsymbol{\mathcal{G}}(z) = [\boldsymbol{g}_1(z)\boldsymbol{g}_2(z)\cdots\boldsymbol{g}_p(z)]$ be a polynomial matrix satisfying the same conditions as $\boldsymbol{\mathcal{H}}(z)$, i.e., $\boldsymbol{\mathcal{G}}(z)$ is $M \times P$, irreducible, column-reduced, and $deg\, \boldsymbol{g}_p(z) \le D_{\max}$. If $\boldsymbol{\mathcal{G}}(z)$ induces the same MIMO channel output statistics, i.e., $\mathcal{R}_{\boldsymbol{y}}(\boldsymbol{\mathcal{G}}(z)) = \mathcal{R}_{\boldsymbol{y}}(\boldsymbol{\mathcal{H}}(z))$, then $\boldsymbol{\mathcal{G}}(z) = \boldsymbol{\mathcal{H}}(z)\,\Theta$, where $\Theta = \mathrm{diag}(e^{i\theta_1}, e^{i\theta_2}, \ldots, e^{i\theta_P})$. □

Proof: See Appendix A. Theorem 1 ensures uniqueness of the MIMO channel matrix from the SOS of the system outputs, as [22, th. 1]. However, here, the minimum memory length restriction on the correlative filters of [22] has been dropped, and uniqueness of the MIMO channel is now ensured by a more general spectral condition.

A. Connection with [19]

Assumption A4) entails a significant simplification when the channel degrees D_p in (2) are known *a priori*. In this situation, from the proof of Theorem 1, it can be seen that it suffices that for each pair of sources $p \neq q$, there exists a correlation lag $\tau = \tau(p, q)$ such that $\sigma(\boldsymbol{A}_p(\tau; N_p)) \cap \sigma(\boldsymbol{A}_q(\tau; N_q)) = \emptyset$, where $N_p = 1 + D_p + N$. Thus, for the special case of static mixtures ($D_p = 0$ and $N = 0$), we recover the spectral conditions of the identifiability Theorem 2 in [19].

Theorem 2 below shows that assumption A4) is not very restrictive and is generically satisfied by unit-norm minimum-phase filters of arbitrary nonzero degree. Before stating Theorem 2, we need a definition. For $d \in \mathbb{N}$, we let $\mathcal{M}_{(d)}[z] \subset \mathbb{C}_{(d)}[z]$ denote the subset of unit-norm minimum phase filters with degree d. Recall that we previouly identified $\mathbb{C}_d[z]$ with $\mathbb{C}^{d+1}$ by associating with the filter $c(z) = \sum_{k=0}^{d} c(k)z^{-k} \in \mathbb{C}_d[z]$ the vector $\boldsymbol{c} = (c(0), c(1), \ldots, c(d))^{\mathrm{T}} \in \mathbb{C}^{d+1}$. Thus, both $\mathbb{C}_{(d)}[z]$ and $\mathcal{M}_{(d)}[z]$ are subsets of $\mathbb{C}^{d+1}$ and take their metric structure from this identification. Now, for given P nonzero degrees in $\boldsymbol{d} = (d_1, d_2, \ldots, d_P)^{\mathrm{T}}$, let

$$\mathcal{M}_{\boldsymbol{d}}[z] = \mathcal{M}_{(d_1)}[z] \times \mathcal{M}_{(d_2)}[z] \times \cdots \times \mathcal{M}_{(d_P)}[z]$$

i.e., $\boldsymbol{f}(z) = (f_1(z), f_2(z), \ldots, f_P(z))^{\mathrm{T}} \in \mathcal{M}_{\boldsymbol{d}}[z]$ if and only if $f_p(z) \in \mathcal{M}_{(d_p)}[z]$ for $p = 1, 2, \ldots, P$.

Theorem 2: Consider the correlative filters in (4), where $a_p(t)$ denotes a white information signal, i.e., $r_{a_p}(\tau) = \delta(\tau)$, for $p = 1, 2, \ldots, P$. Let $\boldsymbol{c} = (C_1, C_2, \ldots, C_P)^{\mathrm{T}} \in \mathbb{N}^P$ be a P-tuple of nonzero correlative filter degrees, and let $\mathcal{F}[z] \subset \mathcal{M}_{\boldsymbol{c}}[z]$ denote the subset of correlative filters that satisfy assumption A4). Then, $\mathcal{F}[z]$ is dense in $\mathcal{M}_{\boldsymbol{c}}[z]$. ☐

Proof: See Appendix B.

IV. Algorithm

We exploit Theorem 1 to derive a closed-form algorithm that obtains the $M \times P$ MIMO polynomial matrix $\boldsymbol{\mathcal{H}}(z) = [\boldsymbol{h}_1(z)\, \boldsymbol{h}_2(z) \cdots \boldsymbol{h}_P(z)]$ up to a phase offset per user from the correlation matrices of the observed data samples $\boldsymbol{y}(t)$, i.e., from the set $\mathcal{R}_{\boldsymbol{y}} = \{\boldsymbol{R_y}(\tau) : \tau \in \mathbb{Z}\}$. The algorithm works in terms of the equivalent stacked data model defined in (3); it computes a MIMO channel matrix $\boldsymbol{G} = [\boldsymbol{G}_1 \boldsymbol{G}_2 \cdots \boldsymbol{G}_P]$ that reproduces the observed SOS of $\boldsymbol{y}_N(t)$, i.e., $\boldsymbol{G}$ satisfies

$$\boldsymbol{R}_{\boldsymbol{y}_N}(\tau) = \boldsymbol{G}\boldsymbol{R_s}(\tau)\boldsymbol{G}^* + \boldsymbol{R_{w_N}}(\tau), \quad \forall_{\tau \in \mathbb{Z}} \tag{7}$$

where $\boldsymbol{R_s}(\tau) = \mathrm{E}\{\boldsymbol{s}(t)\boldsymbol{s}(t-\tau)^*\}$ with $\boldsymbol{s}(t)$ as defined in (3). Theorem 1 asserts that if (7) holds, then $\boldsymbol{G}_p = \boldsymbol{H}_p e^{i\theta_p}$, or equivalently, since $\boldsymbol{G}_p = \boldsymbol{\mathcal{T}}_N(\boldsymbol{g}_p(z))$, $\boldsymbol{g}_p(z) = \boldsymbol{h}_p(z)e^{i\theta_p}$. The receiver knows

i) the correlation matrices of $\boldsymbol{y}_N(t)$ since we have (8), shown at the bottom of the page, and $\boldsymbol{R_y}(\tau)$ can be accurately estimated in practice, e.g., by a sample-mean operator;

ii) the sources's correlation matrices $\boldsymbol{R_s}(\tau)$ since the correlative filters are known by the receiver according to assumption A4);

iii) the noise correlation matrices $\boldsymbol{R_{w_N}}(\tau)$ since they are related to $\boldsymbol{R_w}(\tau)$; replace $\boldsymbol{y}(t)$ by $\boldsymbol{w}(t)$ in (8) and the latter are known by assumption A3).

Let $\boldsymbol{R}(\tau) = \boldsymbol{R_{y_N}}(\tau) - \boldsymbol{R_{w_N}}(\tau)$ denote the denoised correlation matrices of $\boldsymbol{y}_N(t)$. We compute $\boldsymbol{G}$ to satisfy (7) in four steps.

Step 1) **Determination of $\boldsymbol{G_0}$.** This step computes the matrix

$$\boldsymbol{G}_0 = \boldsymbol{H}\boldsymbol{R_s}(0)^{1/2}\boldsymbol{Q}^* \tag{9}$$

where $\boldsymbol{Q}$ is a residual (unknown) unitary mixing matrix. Given the signal model in (3) and assumptions A1)-A4), the denoised correlation matrix at lag 0 is given by the outer product

$$\boldsymbol{R}(0) = \boldsymbol{H}\boldsymbol{R_s}(0)\boldsymbol{H}^* = \left(\boldsymbol{H}\boldsymbol{R_s}(0)^{1/2}\right)\left(\boldsymbol{H}\boldsymbol{R_s}(0)^{1/2}\right)^*.$$

Thus, $L = \sum_{p=1}^{P}(D_p + N) + P$, which is the MIMO system order, i.e., the number of columns of $\boldsymbol{H}$ in (3), is obtained as $L = \mathrm{rank}\,\{\boldsymbol{R}(0)\}$. As it is well known, determination of rank in the presence of noise is a tricky problem. For robustness, it is more appropriate to evaluate the rank of $\boldsymbol{R}(0)$ through statistical methods, such as the minimum description length (MDL) test applied to $\boldsymbol{R_{y_N}}(0)$; see for example [23]. Once L is determined, we perform an L-truncated eigenvalue decomposition (EVD) of $\boldsymbol{R}(0)$ to obtain its square-root, say, $\boldsymbol{G}_0$, which automatically satisfies (9).

Step 2) **Joint Determination of $\boldsymbol{D_p (p = 1, 2, \ldots, P)}$ and $\boldsymbol{Q}$.** This step jointly estimates the channel degrees D_p ($p = 1, 2, \ldots, P$) in (2) and the unitary matrix $\boldsymbol{Q}$ in (9) up to a phase offset per user. Notice that estimating D_p is equivalent to estimating $L_p \equiv D_p + N$, as N, which is the smoothing factor, is chosen at the receiver, and so it is known. The indices $L_p, p = 1, 2, \ldots, P$ are obtained by minimizing a certain non-negative function $\varphi : \mathcal{L} \to \mathbb{R}_0^+$, where $\mathcal{L}$ denotes the finite set

$$\begin{aligned}\mathcal{L} = &\{(M_1, M_2, \ldots, M_P) \in \mathbb{N}^P : \\ &\quad M_1 + M_2 + \cdots + M_P = L - P\}.\end{aligned}$$

We will see that the point $\boldsymbol{l} = (L_1, L_2, \ldots, L_P)$ is the unique zero of φ. Moreover, the computation of $\varphi(\boldsymbol{l})$ automatically provides $\boldsymbol{Q}$ up to a phase offset per user. In order to define φ, we need some additional notation. Recall that the correlation matrices $\boldsymbol{R_{s_p}}(\tau; M_p)$ of $\boldsymbol{s}_p(t; M_p) = \mathcal{S}_{M_p}[s_p(t)]$ are given by (5) and its normalized correlation matrices $\boldsymbol{A}_p(\tau; M_p)$ by (6). For $\boldsymbol{m} = (M_1, M_2, \ldots, M_P) \in \mathcal{L}$, let

$$\boldsymbol{A}(\tau; \boldsymbol{m}) \equiv \mathrm{diag}(\boldsymbol{A}_1(\tau; M_1), \boldsymbol{A}_2(\tau; M_2), \ldots, \boldsymbol{A}_P(\tau; M_P))$$

and

$$\boldsymbol{B}(\tau) \equiv \boldsymbol{G}_0^{\dagger}\boldsymbol{R}(\tau)\boldsymbol{G}_0^{\dagger *} = \boldsymbol{Q}\boldsymbol{R_s}(0)^{-1/2}\boldsymbol{R_s}(\tau)\boldsymbol{R_s}(0)^{-1/2}\boldsymbol{Q}^*. \tag{10}$$

Note that both $\boldsymbol{A}(\tau; \boldsymbol{m})$ and $\boldsymbol{B}(\tau)$ are available to the receiver. The matrices $\boldsymbol{A}(\tau; \boldsymbol{m})$ can be prestored, and $\boldsymbol{B}(\tau)$ are computed from the received data. Associate with each P-tuple of integers $\boldsymbol{m} = (M_1, M_2, \ldots, M_P\} \in \mathcal{L}$, a unitary matrix $\boldsymbol{U}(\boldsymbol{m})$ as follows: Let $\boldsymbol{W}_p \in \mathbb{C}^{L \times (M_p+1)}$ be a global minimizer of

$$f(\boldsymbol{X}) = \sum_{\tau \in \mathbb{Z}} ||\boldsymbol{B}(\tau)\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}_p(\tau; M_p)||^2 \tag{11}$$

subject to $||\boldsymbol{X}|| = \sqrt{M_p + 1}$. Let $\boldsymbol{W} = [\boldsymbol{W}_1\, \boldsymbol{W}_2 \cdots \boldsymbol{W}_P] : L \times L$. We define $\boldsymbol{U}(\boldsymbol{m})$ as the nonlinear projection of $\boldsymbol{W}$ onto the group of $L \times L$ unitary matrices. Theorem 3 describes the function φ and states its main properties.

$$\boldsymbol{R_{y_N}}(\tau) = \begin{bmatrix} \boldsymbol{R_y}(\tau) & \boldsymbol{R_y}(\tau+1) & \cdots & \boldsymbol{R_y}(\tau+N) \\ \boldsymbol{R_y}(\tau-1) & \boldsymbol{R_y}(\tau) & \cdots & \boldsymbol{R_y}(\tau+N-1) \\ \vdots & \ddots & \ddots & \vdots \\ \boldsymbol{R_y}(\tau-N) & \boldsymbol{R_y}(\tau-N+1) & \cdots & \boldsymbol{R_y}(\tau) \end{bmatrix} \tag{8}$$

Theorem 3: Let $\varphi : \mathcal{L} \to \mathbb{R}_0^+$ be defined by

$$\begin{aligned}\boldsymbol{m} =& (M_1, M_2, \ldots, M_P) \in \mathcal{L}\\ \mapsto& \varphi(\boldsymbol{m}) = \sum_{\tau \in \mathbb{Z}} \|\boldsymbol{B}(\tau) - \boldsymbol{U}(m)\boldsymbol{A}(\tau; m)\boldsymbol{U}(m)^*\|^2. \quad (12)\end{aligned}$$

Then, $\boldsymbol{l} = (L_1, L_2, \ldots, L_P)$ is the unique zero of φ. Moreover, $\boldsymbol{U}(\boldsymbol{l}) = \boldsymbol{Q\Theta}$, for some diagonal matrix $\boldsymbol{\Theta} = \mathrm{diag}(e^{i\theta_1} \boldsymbol{I}_{L_1+1}, e^{i\theta_2}\boldsymbol{I}_{L_2+1}, \ldots, e^{i\theta_P}\boldsymbol{I}_{L_P+1})$. $\square$

Proof: See Appendix C.

The following remarks are in order.

Remark 1: Given (10), we have

$$\varphi(\boldsymbol{m}) = \sum_{\tau \in \mathbb{Z}} \|\boldsymbol{QA}(\tau; \boldsymbol{l})\boldsymbol{Q}^* - \boldsymbol{U}(\boldsymbol{m})\boldsymbol{A}(\tau; \boldsymbol{m})\boldsymbol{U}(\boldsymbol{m})^*\|^2. \quad (13)$$

Thus, as $\boldsymbol{m}$ runs over $\mathcal{L}$, the observed second-order moments $\boldsymbol{B}(\tau)$ tend to be matched by $\boldsymbol{U}(\boldsymbol{m})\boldsymbol{A}(\tau; \boldsymbol{m})\boldsymbol{U}(\boldsymbol{m})^*$.

Remark 2: The summations in (11) and (12) involve only a finite number of terms. They are carried over a finite subset $\mathcal{T} = \{\tau_1, \tau_2, \ldots, \tau_k\} \subset \mathbb{Z}$ since the correlative filters are FIR (finite memory span).

Remark 3: A closed-form minimizer of (11) is obtained as

$$\begin{aligned}f(\boldsymbol{X}) =& \sum_{\tau \in \mathcal{T}} \|\boldsymbol{B}(\tau)\boldsymbol{X} - \boldsymbol{XA}_p(\tau; M_p)\|^2\\ =& \sum_{i=1}^{k} \left\|(\boldsymbol{I}_{M_p} \otimes \boldsymbol{B}(\tau_i) - \boldsymbol{A}_p(\tau_i; M_p)^{\mathrm{T}} \otimes \boldsymbol{I}_L)\boldsymbol{x}\right\|^2\\ =& \|\boldsymbol{Tx}\|^2\\ =& \boldsymbol{x}^*\boldsymbol{T}^*\boldsymbol{Tx} \end{aligned} \quad (14)$$

where $\boldsymbol{x} = \mathrm{vec}(\boldsymbol{X})$, and

$$\begin{aligned}\boldsymbol{T} =& [\boldsymbol{T}_1^{\mathrm{T}}\boldsymbol{T}_2^{\mathrm{T}} \cdots \boldsymbol{T}_k^{\mathrm{T}}]^{\mathrm{T}}\\ \boldsymbol{T}_i =& \boldsymbol{I}_{M_p} \otimes \boldsymbol{B}(\tau_i) - \boldsymbol{A}_p(\tau_i; M_p)^{\mathrm{T}} \otimes \boldsymbol{I}_L.\end{aligned}$$

A global minimizer of $f(\boldsymbol{X})$ subject to $\|\boldsymbol{X}\| = \sqrt{M_p+1}$ can be obtained by scaling and reshaping into matrix format the eigenvector associated with the minimum eigenvalue of the semidefinite positive Hermitean matrix $\boldsymbol{T}^*\boldsymbol{T}$, say $\boldsymbol{u}$, i.e., $\boldsymbol{X}_p = \mathrm{vec}^{-1}(\sqrt{M_p+1}\boldsymbol{u})$. In addition, notice that $\boldsymbol{W} = [\boldsymbol{W}_1\boldsymbol{W}_2 \cdots \boldsymbol{W}_P]$ can be computed in P parallel threads, the pth thread leading to $\boldsymbol{W}_p$.

Remark 4: The nonlinear projection of $\boldsymbol{W} : L \times L$ onto $\mathcal{U}_L$, which is the group of $L \times L$ unitary matrices, is a special case of the classical orthogonal Procrustes problem [25]. It can be computed as follows. Let $\boldsymbol{W} = \boldsymbol{U\Sigma V}^*$ denote a singular-value decomposition (SVD) of $\boldsymbol{W}$. Then, $\boldsymbol{UV}^*$ (unitary) globally minimizes the Euclidean distance between $\boldsymbol{W}$ and $\mathcal{U}_L$.

Step 3) **Determination of $\boldsymbol{G}$:**. The previous step provides the unitary matrix $\boldsymbol{U} = \boldsymbol{U}(\boldsymbol{l})$. To obtain $\boldsymbol{G}$, which is a local copy of the channel matrix $\boldsymbol{H}$ up to a phase offset per user, set $\boldsymbol{G} = \boldsymbol{G}_0\boldsymbol{UR_s}(0)^{-1/2}$. In fact, using the expression of $\boldsymbol{G}_0$ in (9), we get

$$\boldsymbol{G} = \boldsymbol{HR_s}(0)^{1/2}\Theta\boldsymbol{R_s}(0)^{-1/2} = \boldsymbol{H}\Theta.$$

In the last equality, we have used the fact that $\boldsymbol{R_s}(0)^{1/2}$ and $\boldsymbol{\Theta}$ commute. They share the same block-diagonal structure, the pth block of $\boldsymbol{\Theta}$ being the matrix $e^{i\theta_p}\boldsymbol{I}_{L_p+1}$.

Step 4) **Determination of $\boldsymbol{\mathcal{H}}(z)$:** Once $\boldsymbol{G} = [\boldsymbol{G}_1\boldsymbol{G}_2 \cdots \boldsymbol{G}_P]$, $\boldsymbol{G}_p : L \times (L_p+1)$, is computed, the $M \times P$ MIMO polynomial matrix $\boldsymbol{\mathcal{H}}(z)$ up to a phase offset per user is easily retrieved. Notice the equation at bottom of the page. Thus, the coefficients of the pth filter

$$\boldsymbol{h}_p(z) = \sum_{k=0}^{D_p} \boldsymbol{h}_p(k)z^{-k}$$

i.e.,

$$\boldsymbol{\mathcal{H}}_p = [\boldsymbol{h}_p(0)\boldsymbol{h}_p(1) \cdots \boldsymbol{h}_p(D_p)]$$

may be obtained by averaging the $N+1$ copies available in $\boldsymbol{G}_p$.

V. Separation of Sources

After the MIMO polynomial matrix $\boldsymbol{\mathcal{H}}(z)$ is identified, we face the problem of detecting the unfiltered information sequences $a_p(t)$ in (4) from the observations $\boldsymbol{y}(t)$ in (2). In the sequel, we assume that the pth data sequence $a_p(t)$ consists of i.i.d. symbols drawn from a finite alphabet $\mathcal{A}_p \subset \mathbb{C}$. In addition, for simplicity, we assume that $\boldsymbol{w}(t)$ denotes spatio-temporal white Gaussian noise. Thus, the optimal maximum likelihood (ML) criterion leads to a generalized maximum likelihood sequence estimation (MLSE) Viterbi algorithm [24]. However, the computational cost of this approach is usually very high. We pursue a simpler, suboptimal technique to detect the data symbols. The proposed technique exploits the fact that the correlative filters are minimum-phase and permits to handle carrier frequency asynchronisms, e.g., Doppler effects. This distortion induces a baseband rotation in the received signals.

We have a data model similar to (2), except for the inclusion of the residual phase drifts

$$\boldsymbol{y}(t) = \sum_{p=1}^{P} [\boldsymbol{h}_p(z)]\tilde{s}_p(t) + \boldsymbol{w}(t) \quad (15)$$

$$\boldsymbol{G}_p = \boldsymbol{H}_p e^{i\theta_p} = \begin{bmatrix} \boldsymbol{h}_p(0) & \cdots & \boldsymbol{h}_p(D_p) & \boldsymbol{0} & \cdots & \boldsymbol{0}\\ \boldsymbol{0} & \boldsymbol{h}_p(0) & \cdots & \boldsymbol{h}_p(D_p) & \ddots & \boldsymbol{0}\\ \vdots & \ddots & \ddots & \ddots & \ddots & \vdots\\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{h}_p(0) & \cdots & \boldsymbol{h}_p(D_p)\end{bmatrix} e^{i\theta_p}$$

where $\tilde{s}_p(t) = s_p(t)e^{i\omega_p t}$, and ω_p denotes the baseband rotation frequency corresponding to the pth user. Although the filters $\boldsymbol{h}_p(z)$ in (15) are not exactly the same as in (2) (some phase offset corrections are needed), we maintain the notation for the sake of clarity.

Each iteration of the proposed iterative procedure consists of two steps: **Step A)** The data symbols $a_p(t)$ are detected, given the current estimate of $\boldsymbol{\mathcal{H}}(z)$, and **Step B)** the filtering matrix $\boldsymbol{\mathcal{H}}(z)$ is re-evaluated on the basis of the newly estimated data symbols. This resembles, in spirit, the methodology of the ILSP and ILSE approaches [12]. The added difficulty here is that the data symbols are prefiltered and distorted by baseband rotations. We now discuss these steps in more detail.

A. Step A

We are at the $(n+1)$th iteration cycle. Let

$$\boldsymbol{\mathcal{H}}^{(n)}(z) = \left[\boldsymbol{h}_1^{(n)}(z)\boldsymbol{h}_2^{(n)}(z)\cdots\boldsymbol{h}_P^{(n)}(z)\right]$$

denote the estimate of the MIMO channel matrix $\boldsymbol{\mathcal{H}}(z)$ obtained from the previous iteration cycle. The algorithm is initialized with $\boldsymbol{\mathcal{H}}^{(0)}(z)$, which is the closed-form solution of Section IV, and $n=0$. We reason as if $\boldsymbol{\mathcal{H}}^{(n)}(z) = \boldsymbol{\mathcal{H}}(z)$. Focus on the pth user. First, we extract the baseband rotated sequence $\tilde{s}_p(t) = e^{i\omega_p t}s_p(t)$ from the observations $\boldsymbol{y}(t)$ in (15). We employ a filter $\boldsymbol{f}_p(z) = \sum_{k=0}^{F_p}\boldsymbol{f}_p(k)z^{-k} \in \mathbb{C}_{F_p}^M[z]$ satisfying

$$\boldsymbol{f}_p(z)^{\mathrm{T}}\boldsymbol{\mathcal{H}}^{(n)}(z) = (0,\ldots,0,z^{-d_p},0,\ldots,0)$$

for some non-negative delay d_p in the pth entry, i.e., $\boldsymbol{f}_p(z)$ exactly nulls the intersymbol and co-channel interferences affecting the user p. Notice that for sufficiently high degree F_p, the existence of such a filter is guaranteed by the irreducibility of the channel matrix in assumption A2). The coefficients of the filter

$$\boldsymbol{f}_p^{\mathrm{T}} = \left(\boldsymbol{f}_p(0)^{\mathrm{T}}, \boldsymbol{f}_p(1)^{\mathrm{T}}, \ldots, \boldsymbol{f}_p(F_p)^{\mathrm{T}}\right)$$

can be obtained as the $((p-1)(1+F_p)+D_1+\cdots+D_{p-1}+d_p+1)$th row of the pseudoinverse of

$$\boldsymbol{H}^{(n)} = \left[\boldsymbol{\mathcal{T}}_{F_p}\left(\boldsymbol{h}_1^{(n)}\right)\boldsymbol{\mathcal{T}}_{F_p}\left(\boldsymbol{h}_2^{(n)}\right)\cdots\boldsymbol{\mathcal{T}}_{F_p}\left(\boldsymbol{h}_P^{(n)}\right)\right].$$

Let

$$\begin{aligned}\alpha_p(t) =& [\boldsymbol{f}_p(z)^{\mathrm{T}}]\boldsymbol{y}(t+d_p)\\ =& e^{i\omega_p t}\left(\sum_{k=0}^{C_p} c_p(k)a_p(t-k)\right) + n_p(t)\end{aligned} \tag{16}$$

where $n_p(t) = [\boldsymbol{f}_p(z)^{\mathrm{T}}]\boldsymbol{w}(t+d_p)$ denote the application of the separating filter to the observations' noise. We have to detect the data symbols $a_p(t) \in \mathcal{A}_p$ from (16).

i) First, we get rid of the correlative filter. Rewrite (16) as

$$\alpha_p(t) = \sum_{k=0}^{C_p}\tilde{c}_p(k)\tilde{a}_p(t-k) + n_p(t)$$

where $\tilde{c}_p(k) = c_p(k)e^{i\omega_p k}$, and $\tilde{a}_p(t) = a_p(t)e^{i\omega_p t}$. We have $\tilde{c}_p(k) \simeq c_p(k)$ since $e^{i\omega_p k} \simeq 1$ for small integers k and typical values of ω_p, as for example, $\omega_p = 2\pi/1000$ and $C_p = 5$. Thus, within this approximation

$$\alpha_p(t) = [c_p(z)]\tilde{a}_p(t) + n_p(t).$$

Since the correlative filter $c_p(z)$ is minimum-phase, we use its stable inverse, say $g_p(z)$, to recover the signal $\tilde{a}_p(t)$

$$\begin{aligned}\beta_p(t) =& [g_p(z)]\alpha_p(t) = \underbrace{e^{i\omega_p t}a_p(t)}_{\tilde{a}_p(t)} + u_p(t)\\ u_p(t) =& [g_p(z)]n_p(t).\end{aligned} \tag{17}$$

ii) Now, we handle the baseband rotation. Divide the available samples $\beta_p(t)$ in B blocks of equal size T_{blk}. Making T_{blk} small enough, we have, within each block $b = 1,2,\ldots,B$, the approximation

$$\beta_p(t) = e^{i\theta_p(b)}a_p(t) + u_p(t) \tag{18}$$

for some $\theta_p(b) \in \mathbb{R}$.

We process the data block by block. Within each block, we jointly estimate the symbols $a_p(t)$ and the phase offset $\theta_p(b)$. Assume we are processing the bth block. The algorithm starts with $\theta_p(0) = 0$ and $b = 1$. We use the estimate of the phase offset in the previous block and the fact that the phase varies smoothly between adjacent blocks to approximate $\theta_p(b-1) \approx \theta_p(b)$, from which we obtain almost phase-corrected symbols

$$\eta_p(t) = e^{-i\theta_p(b-1)}\beta_p(t) \approx a_p(t) + v_p(t)$$

where $v_p(t) = e^{-i\theta_p(b-1)}u_p(t)$ denotes complex Gaussian noise. The data symbol $a_p(t)$ is estimated by projecting $\eta_p(t)$ onto the alphabet $\mathcal{A}_p$. By the least-squares (LS) criterion

$$\widehat{a_p(t)} = \arg\min_{a\in\mathcal{A}_p} \|\eta_p(t) - a\|^2.$$

Now, we turn to the problem of estimating $\theta_p(b)$ in (18), given the symbols $a_p(t)$. Again, we follow an LS strategy

$$\widehat{\theta_p(b)} = \arg\min_{\theta\in\mathbb{R}} \sum_{t\in\mathcal{B}(b)} |\beta_p(t) - e^{i\theta}a_p(t)|^2. \tag{19}$$

The answer to this minimization is $\widehat{\theta_p(b)} = -\phi$, where

$$\sum_{t\in\mathcal{B}(b)} \beta_p(t)^* a_p(t) = \rho e^{i\phi}, \quad \rho \geq 0.$$

Here, $\mathcal{B}(b)$, with cardinality T_{blk}, denotes the set of indices t belonging to the bth block.

It should be noticed that the estimation of ω_p in (17) may also be efficiently solved by exploiting the fact that ω_p is a conjugate cyclic frequency of $\beta_p(t)$. The main advantage of the proposed methodology is that it permits us to handle more generic phase drifts, i.e., phase distortions of the form $e^{i\theta_p(t)}$, where the time-varying phase signal $\theta_p(t)$ is not necessarily given by $\theta_p(t) = \omega_p t$.

B. Step B

Take the estimated symbols $a_p(t)$ and phase offsets $\{\theta_p(b) : b = 1, 2, \ldots, B\}$ in Step A, and let $\tilde{s}_p(t) = e^{i\theta_p(b)}[c_p(z)]a_p(t)$ whenever the time index t falls in the bth block. Rewrite (15) in matrix form as

$$\boldsymbol{Y} = [\boldsymbol{y}(1)\boldsymbol{y}(2)\cdots\boldsymbol{y}(T)] = \boldsymbol{\mathcal{H}}\tilde{\boldsymbol{S}} + \boldsymbol{W}.$$

Here

$$\begin{aligned}\tilde{\boldsymbol{\mathcal{S}}} =& [\tilde{\boldsymbol{s}}(1)\tilde{\boldsymbol{s}}(2)\cdots\tilde{\boldsymbol{s}}(T)]\\ \tilde{\boldsymbol{s}}(t) =& \left(\tilde{\boldsymbol{s}}_1(t)^{\mathrm{T}}, \tilde{\boldsymbol{s}}_2(t)^{\mathrm{T}}, \ldots \tilde{\boldsymbol{s}}\hat{P}(t)^{\mathrm{T}}\right)^{\mathrm{T}}\\ \tilde{\boldsymbol{s}}_p(t) =& \mathcal{S}_{D_p}[\tilde{s}_p(t)]\end{aligned}$$

and $\boldsymbol{W} = [\boldsymbol{w}(1)\boldsymbol{w}(2)\cdots\boldsymbol{w}(T)]$. The LS estimate of $\boldsymbol{\mathcal{H}}$ is given by

$$\tilde{\boldsymbol{\mathcal{H}}} = \underset{\boldsymbol{\mathcal{G}}\in\mathbb{C}^{M\times L}}{\arg\min}\, \|\boldsymbol{Y} - \boldsymbol{\mathcal{G}}\tilde{\boldsymbol{S}}\|^2 = \boldsymbol{Y}\tilde{\boldsymbol{S}}^{\dagger}.$$

We set $\boldsymbol{\mathcal{H}}^{(n+1)}(z) = \widehat{\boldsymbol{\mathcal{H}}}(z)$.

VI. Computer Simulations

We present two sets of simulations. In the first set, we consider $P = 2$ users , without carrier misadjustments. The performance of the proposed blind channel identification technique is evaluated in terms of the mean-square error (MSE) of the MIMO channel estimate. For separation of the sources and the equalization step, the performance criterion is the symbol error rate (SER) of the estimated data symbols. In the second set of simulations, we consider $P = 3$ users with residual phase drifts. We compare our technique with the TICC approach [20], both in terms of the MSE of the channel estimate and the SER of the resulting symbol detection scheme.

A. Scenario with Two Users

We consider $P = 2$ users with distinct digital modulation formats. User 1 employs the normalized (unit-power) quaternary amplitude modulation (QAM) digital format, and user 2 employs the binary phase keying modulation format (BPSK). Both users pass their i.i.d. information sequences through correlative filters prior to transmission. We used correlative filters with minimal memory, i.e., with just one zero $c_p(z) = \kappa_p(1 - z_p z^{-1})$. The zeros of the correlative filters for users 1 and 2 are $z_1 = 1/4e^{-j\pi/2}$ and $z_2 = 1/2e^{j\pi/4}$, respectively; the coefficients κ_p are normalizing constants to ensure unit-power outputs. For both users, the analog transmitter shaping filter $p(t)$ is a raised-cosine with $\alpha = 80\%$ excess bandwidth. Each communication channel is a realization of the continuous-time multipath model

$$g(t) = \gamma_0\delta(t) + \gamma_{\max}\delta(t - \Delta_{\max}) + \sum_{q=1}^{Q}\gamma_q\delta(t - \Delta_q). \quad (20)$$

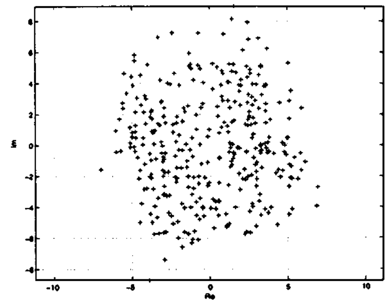

Fig. 1. Output of the unequalized channel.

Here, we have two fixed paths at $\Delta = 0$ and $\Delta_{\max} = 2.8T_0$, where T_0 is the symbol period, and a random number Q of extra paths Q is uniformly distributed in $\{5, 6, \ldots, 15\}$; for $q = 1, \ldots, Q$, the delays Δ_q are uniformly distributed in $[0, \Delta_{\max}]$, and γ_q, $q = 1, \ldots, Q$, as well as γ_0 and $\gamma_{\max}$ denote unit-power complex Gaussian random variables. Each composite continuous-time channel $h(t) = p(t) \odot g(t)$, where $\odot$ is the convolution operator, is then sampled at the baud rate and truncated at $4T_0$. This approximation deletes, at most, 4% of the channel power. Thus, in $\boldsymbol{\mathcal{H}}(z) = [\boldsymbol{h}_1(z)\boldsymbol{h}_2(z)]$, $D_p = \deg\boldsymbol{h}_p(z) = 3$, which is assumed known. The receiver has $M = 4$ antennas and processes packets of $T = 350$ data samples $\boldsymbol{y}(t)$ in (2), with smoothing factor $N = 3$; see (3).

For the identification technique, we fit six correlation matrices, i.e., $\mathcal{T} = \{\pm 1, \pm 2, \pm 3\}$, in (14). After channel identification, the sources are extracted from the observations $\boldsymbol{y}(t)$ by filters $\boldsymbol{f}_p(z)$ of degree $F_p = 2D_p = 6$ with delay $d_p = D_p + 1 = 4$. The observation noise $\boldsymbol{w}(t)$ in (2) is taken as spatio-temporal white Gaussian noise with power σ^2. The SNR is defined as

$$\mathrm{SNR} = \frac{\sum_{p=1}^{P}\mathrm{E}\{\|[\boldsymbol{h}_p(z)]s_p(t)\|^2\}}{\mathrm{E}\{\|\boldsymbol{w}(t)\|^2\}} = \frac{\|\boldsymbol{\mathcal{H}}\|^2}{M\sigma^2}.$$

We start by illustrating a typical run of our technique. Fig. 1 plots in the complex plane a typical received signal, i.e., an entry of the observed vector $\boldsymbol{y}(t)$ in (2). The joint effect of the intersymbol and co-channel interference is clearly noticeable. Fig. 2 (notice the difference in the vertical scale relative to Fig. 1) shows the output of the equalized channel, i.e., the signals $\beta_1(t)$ and $\beta_2(t)$ in (17). As seen, the algorithm recovers valid user signals from the observations. This example was generated with SNR $= 15$ dB. We evaluated more extensively the performance of our proposed technique. We varied the SNR between $\mathrm{SNR}_{\min} = 10$ dB and $\mathrm{SNR}_{\max} = 20$ dB in steps of $\mathrm{SNR}_{\mathrm{step}} = 2.5$ dB. For each SNR, $K = 500$ statistically independent trials were considered. For each trial, we generated $T = 350$ data samples and ran the proposed closed-form and iterative blind channel identification algorithms. We recorded the square-error (SE) of the channel estimate, i.e., SE $= \left\|\widehat{\boldsymbol{\mathcal{H}}} - \boldsymbol{\mathcal{H}}\right\|^2$. The symbol error rates for both sources were obtained by error counting. Fig. 3 displays the average results over the $K = 500$ trials for the mean-square error (MSE) of the channel estimate. This is monotonically decreasing, as expected. The dashed and solid curves refer to the closed-form and the iterative estimates, respectively. As seen, the iterative technique improves significantly over the closed-form estimate.

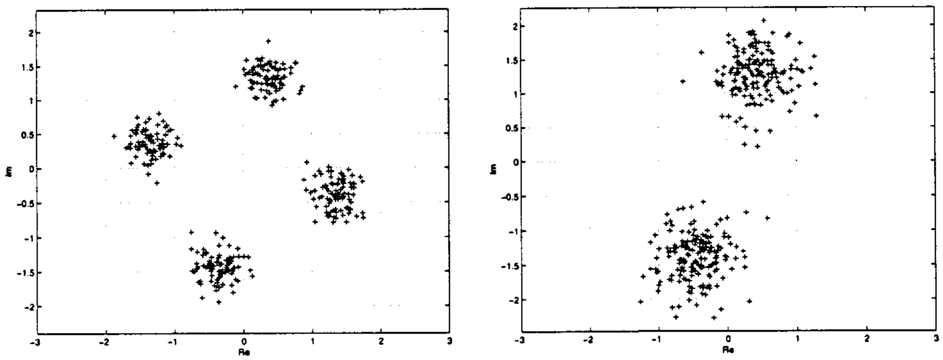

Fig. 2. Signal estimate for (left) user 1 and (right) user 2.

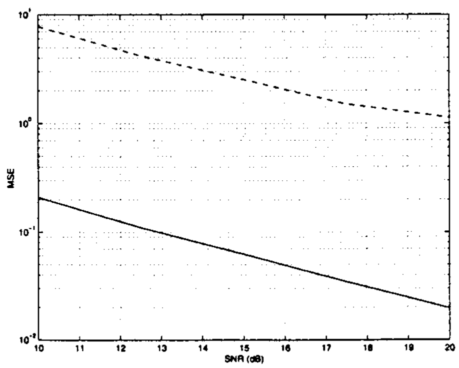

Fig. 3. MSE of the (dashed line) closed-form and the (solid line) channel estimate: SNR varies.

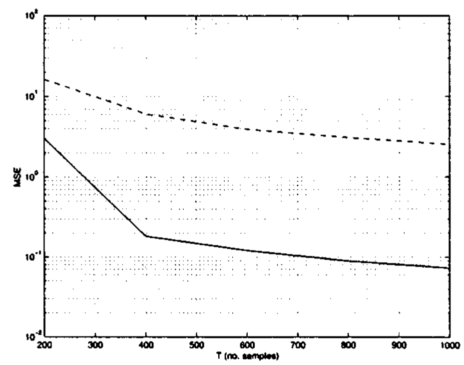

Fig. 4. MSE of the (dashed line) closed-form and the (solid line) channel estimate: T varies.

TABLE I
SYMBOL ERROR RATE (SER).
CLOSED-FORM ALGORITHM ($T = 350$, SNR VARIES)

SNR (dB)	User 1	User 2
10.0	0.0646	0.0276
12.5	0.0124	0.0032
15.0	0.0011	0.0001
17.5	0.0000	0.0000
20.0	0.0000	0.0000

TABLE II
SYMBOL ERROR RATE (SER). ITERATIVE ALGORITHM
($T = 350$, SNR VARIES)

SNR (dB)	User 1	User 2
10.0	0.0014	0.0040
12.5	0.0002	0.0002
15.0	0.0000	0.0000
17.5	0.0000	0.0000
20.0	0.0000	0.0000

TABLE III
SYMBOL ERROR RATE (SER). CLOSED-FORM ALGORITHM
(SNR= 10DB, T VARIES)

T	User 1	User 2
200	0.1527	0.0675
400	0.0461	0.0230
600	0.0218	0.0109
800	0.0168	0.0091
1000	0.0106	0.0075

TABLE IV
SYMBOL ERROR RATE (SER). ITERATIVE ALGORITHM
(SNR= 10DB, T VARIES)

T	User 1	User 2
200	0.0204	0.0156
400	0.0014	0.0041
600	0.0012	0.0041
800	0.0012	0.0041
1000	0.0012	0.0039

In Tables I and II, we show the symbol error rates (SER) associated to the two sources. These correspond to the symbol detectors implemented from the closed-form and iterative channel estimators, respectively. Notice that as user 1 employs the QAM format, the SNR per symbol is lower, and as a consequence, the SER is higher. Moreover, as expected, the better accuracy of the iterative MIMO channel estimate results in a lower probability of error. We also studied the performance of the proposed technique with respect to the packet size T. We fixed SNR $=$ 10 dB and varied T between $T_{\min} = 200$ and

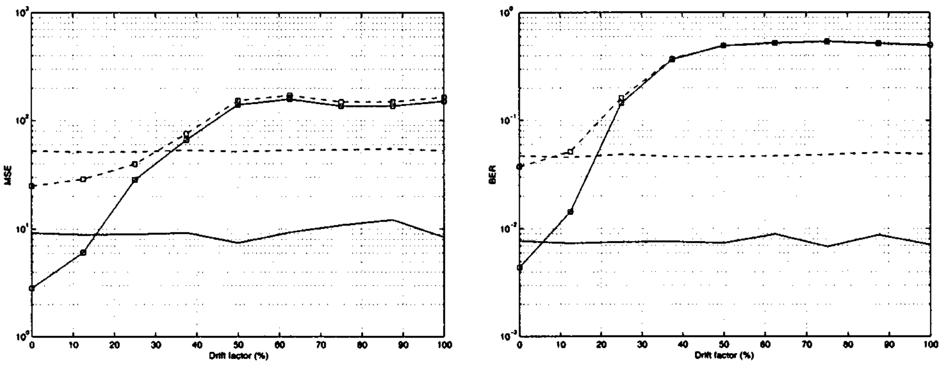

Fig. 5. (Left) MSE and (right) BER of user 1 for the proposed and TICC (with square marks) approaches. (Dashed) Closed-form and (solid) iterative algorithms (SNR $= 5$ dB).

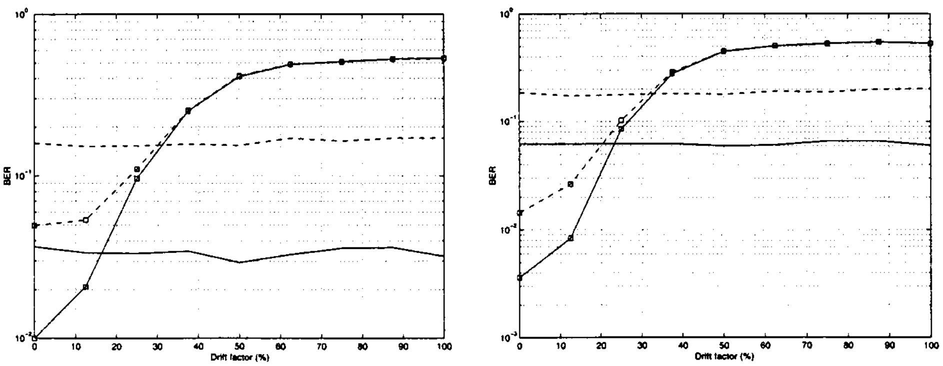

Fig. 6. (Left) BER of user 2 and (right) user 3 for the proposed and TICC (with square marks) approaches. (Dashed) Closed-form and (solid) iterative algorithms (SNR $= 5$ dB).

TABLE V
ZEROS OF THE CORRELATIVE FILTERS ($P = 3$ USERS)

User p	$z_{p,1}$	$z_{p,2}$
1	$\frac{1}{2}$	$-\frac{1}{2}$
2	$\frac{1}{3}e^{i\pi/3}$	$\frac{1}{4}e^{i\pi/2}$
3	$\frac{1}{2}e^{i3\pi/4}$	$-\frac{1}{4}e^{-i\pi/2}$

$T_{\max} = 1000$ in steps of $T_{\text{step}} = 200$. Fig. 4 shows the average results for the MSE, and Tables III and IV display the SER of both sources.

B. Scenario with Three Users

In this set of computer simulations, we consider $P = 3$ binary users, and compare our results with the TICC approach [20]. Each user employs a FIR correlative filter with two zeros, i.e., we have $c_p(z) = \kappa_p(1 - z_{p,1}z^{-1})(1 - z_{p,2}\,z^{-1})$. Table V discriminates the zeros of the correlative filters for each user. The multipath propagation model in (20) is maintained, but now, $\Delta_{\max} = 2.5T_0$, and the composite channel is truncated at $3T_0$. This suppresses, at most, 4% of the channel power. Thus, $D_p = 3$ for $p = 1, 2, 3$. An $M = 8$ antenna array sampled at the baud rate is assumed at the receiver. In addition, the data packet size $T = 750$, the smoothing factor $N = 2$, the correlation lags used in the closed-form identification are $\mathcal{T} = \{\pm 1, \pm 2, \pm 3\}$, and the degree and delay of the separating filters are $F_p = 2D_p = 6$ and $d_p = D_p + 1 = 4$. For the iterative algorithm, we considered $T_{blk} = 10$ samples. For the TICC approach, the three users employ cyclic frequencies given by $\alpha_1 = -0.35$, $\alpha_2 = 0$, and $\alpha_3 = 0.35$, respectively. In addition, the Wiener filters in [20] are implemented with parameters $\delta = 4$ and $L = 7$. The nominal baseband rotations for the three users in (15) are given by $\omega_1 = 2\pi/750$, $\omega_2 = -2\pi/1000$, and $\omega_3 = 2\pi/900$, respectively. The channel degrees are assumed known for both approaches.

We performed computer simulations to compare the performance of our proposed technique and the TICC approach. We considered residual baseband rotations given by $\lambda\omega_p (p = 1, 2, 3)$, where the drift factor λ varied between $\lambda_{\min} = 0$ and $\lambda_{\max} = 1$ in steps of $\lambda_{\text{step}} = 0.125$. For each λ, $K = 500$ statistically independent trials were performed. Each trial consisted of the generation of $T = 750$ data samples as well as subsequent channel estimation and symbol detection as in the previous scenario with two users. The left plot in Fig. 5 displays the average results over the $K = 500$ trials considered. The SNR was fixed at 5 dB. For both approaches, the dashed and solid curves correspond to the closed-form and iterative channel estimates, respectively. Additionally, the curves associated with the TICC approach are labeled with a square mark. As seen, the accuracy of the channel estimate by our technique is almost insensitive to the drift baseband rotation factor λ. In contrast, the performance of the TICC estimators degrades as the carrier's misadjustment gets worse. The right plot in Figs. 5 and 6 display the bit error rates (BERs) associated with the two approaches for the $P = 3$ users considered. As

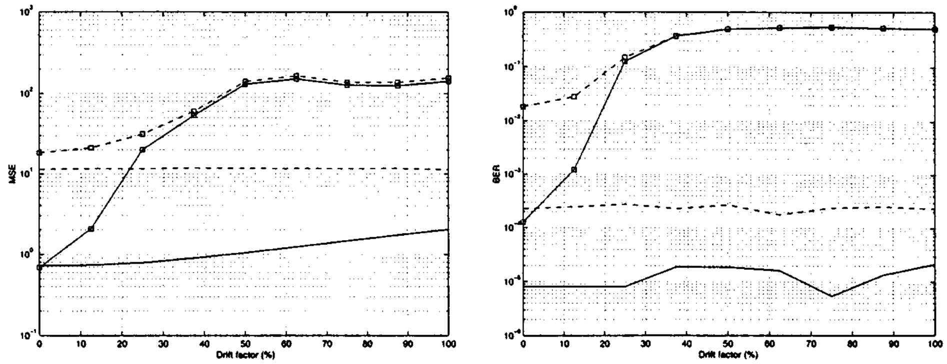

Fig. 7. (Left) MSE and (right) BER of user 1 for the proposed and TICC (with square marks) approaches. (Dashed) Closed-form and (solid) iterative algorithms (SNR $=$ 10 dB).

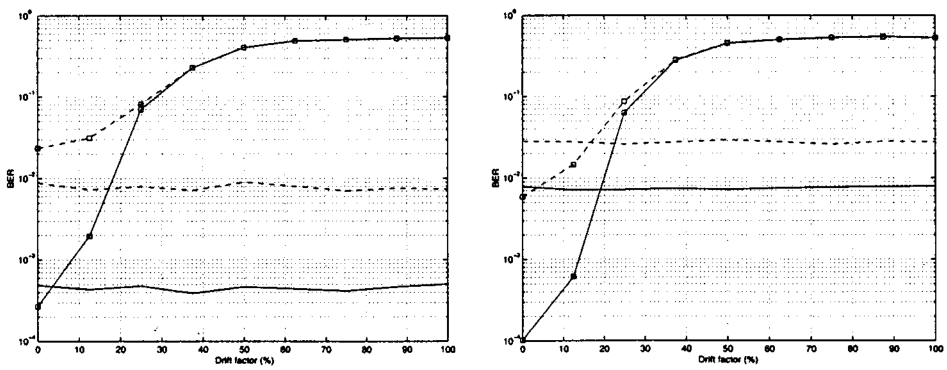

Fig. 8. (Left) BER of user 2 and (right) user 3 for the proposed and TICC (with square marks) approaches. (Dashed) Closed-form and (solid) iterative algorithms (SNR $=$ 10 dB).

seen, the proposed technique outperforms TICC. A similar set of simulations was performed under SNR $=$ 10 dB. The results are displayed in Figs. 7 and 8. We can infer the same conclusions as above.

VII. Conclusions

We described a novel blind *closed-form* estimator for the MIMO channel. This finds applications in multiuser environments like in wireless mobile communications. The proposed estimator is consistent and uses only *second-order statistics*. We *correlative code* the source outputs by filtering the data information sequences before transmission. This induces a spectral asymmetry between the sources and enables us to develop our closed-form solution. The correlative filters do not increase the power or the spectral bandwidth of the system, nor do they require any reduction in the original data rate. We identified the sufficient condition on the correlative filters that guarantees the desired spectral diversity. This condition is generic in the set of unit-power minimum-phase filters of arbitrary memory, in particular, filters with just one zero. Thus, the filters can be inverted directly, which results in a computationally attractive scheme to recover the original information sequences. We showed that in contrast to the TICC approach, our pre-processing is resilient to baseband phase drifts induced by carrier frequency misadjustments.

Appendix A
Proof of Theorem 1

First, we need some technical lemmas.

1) Lemma A: Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$ and $\boldsymbol{y} \neq \mathbf{o}$. If $\boldsymbol{x}^* \boldsymbol{K}_n(k) \boldsymbol{y} = 0$, for all $k \in \mathbb{Z}$, then $\boldsymbol{x} = \mathbf{0}$. $\square$

Proof: Write $\boldsymbol{y} = (0, \ldots, 0, y_l, \ldots, y_n)^{\mathrm{T}}$, where $y_l \neq 0$. Then, $\boldsymbol{\mathcal{Y}} = [\boldsymbol{y}(-l+1)\boldsymbol{y}(-l+2) \cdots \boldsymbol{y}(n-l)]$, where $\boldsymbol{y}(k) = \boldsymbol{K}_n(k)\boldsymbol{y}$ is a Toeplitz lower triangular matrix with y_l in the diagonal. Thus, $\boldsymbol{\mathcal{Y}}$ is nonsingular, and $\boldsymbol{x}^* \boldsymbol{\mathcal{Y}} = \mathbf{0}$ implies $\boldsymbol{x} = 0$.

Lemma B: Let $\boldsymbol{A} = \mathrm{diag}(\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n)$,where $\boldsymbol{A}_i \in \mathbb{C}^{n_i \times n_i}$ for $i = 1, 2, \ldots, n$. Assume that $\sigma(\boldsymbol{A}_i) \cap \sigma(\boldsymbol{A}_j) = \emptyset$ for $i \neq j$. If $\boldsymbol{B}$commutes with $\boldsymbol{A}$, then $\boldsymbol{B} = \mathrm{diag}(\boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_n)$ for some $\boldsymbol{B}_i \in \mathbb{C}^{n_i \times n_i}$, $i = 1, 2, \ldots, n$. $\square$

Proof: We use the fact that if $\boldsymbol{XY} - \boldsymbol{YZ} = \mathbf{o}$ and $\sigma(\boldsymbol{X}) \cap \sigma(\boldsymbol{Z}) = \emptyset$, then $\boldsymbol{Y} = \mathbf{0}$[25]. Write

$$\boldsymbol{B} = \begin{bmatrix} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} & \cdots & \boldsymbol{B}_{ln} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} & \cdots & \boldsymbol{B}_{2n} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{B}_{nl} & \boldsymbol{B}_{n2} & \cdots & \boldsymbol{B}_{nn} \end{bmatrix}$$

where $\boldsymbol{B}_{ij} \in \mathbb{C}^{n_i \times n_j}$. From $\boldsymbol{AB} = \boldsymbol{BA}$, $\boldsymbol{A}_i \boldsymbol{B}_{ij} = \boldsymbol{B}_{ij} \boldsymbol{A}_j$. Since $\sigma(\boldsymbol{A}_i) \cap \sigma(\boldsymbol{A}_j) = \emptyset$, for $i \neq j$, we have $\boldsymbol{B}_{ij} = \mathbf{0}$ for $i \neq j$.

Lemma C: Let $\boldsymbol{V} \in \mathbb{C}^{m \times n} (m \geq n)$ denote an isometry, i.e., $\boldsymbol{V}^* \boldsymbol{V} = \boldsymbol{I}_n$, and let $\boldsymbol{F}(k) = \boldsymbol{V}^* \boldsymbol{K}_m(k) \boldsymbol{V}$ for $k \in \mathbb{Z}$. If $\boldsymbol{X} \in \mathbb{C}^{n \times n}$commutes with $\boldsymbol{F}(k)$ for all $k \in \mathbb{Z}$, then $\boldsymbol{X} = \lambda \boldsymbol{I}_n$ for some $\lambda \in \mathbb{C}$. $\square$

Proof: First, consider that $\boldsymbol{X}$ is normal [25], i.e.,

$$\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^* \tag{21}$$

where $\boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1 \boldsymbol{I}_{n_1}, \lambda_2 \boldsymbol{I}_{n_2}, \ldots, \lambda_l \boldsymbol{I}_{n_l})(\lambda_i \neq \lambda_j$, for $i \neq j)$, and $\boldsymbol{Q}$ is unitary. Consider the hypothesis that $l \geq 2$. Using (21) in

$$\boldsymbol{F}(k)\boldsymbol{X} = \boldsymbol{X}\boldsymbol{F}(k) \tag{22}$$

we have $\boldsymbol{G}(k)\boldsymbol{\Lambda} = \boldsymbol{\Lambda}\boldsymbol{G}(k)$, where $\boldsymbol{G}(k) = \boldsymbol{W}^*\boldsymbol{K}_m(k)\boldsymbol{W}$, and $\boldsymbol{W} = \boldsymbol{V}\boldsymbol{Q}$ is an isometry. Since $\boldsymbol{G}(k)$ commutes with $\boldsymbol{\Lambda}$, it follows from Lemma B that $\boldsymbol{G}(k) = \mathrm{diag}(\boldsymbol{G}_1(k), \boldsymbol{G}_2(k), \ldots, \boldsymbol{G}_l(k))$, where $\boldsymbol{G}_i(k) \in \mathbb{C}^{n_i \times n_i}$. Thus, all matrices $\boldsymbol{G}(k)$ share at least one common zero entry (any entry off the block diagonal), i.e., there exists a pair of indices (i, j) such that $[\boldsymbol{G}(k)]_{i,j} = \boldsymbol{w}_i^* \boldsymbol{K}_m(k)\boldsymbol{w}_j = 0$ for all $k \in \mathbb{Z}$; here, $\boldsymbol{w}_i$ denotes the ith column of $\boldsymbol{W}$. Lemma A asserts that $\boldsymbol{w}_i = \boldsymbol{0}$ (which is a contradiction). Thus, we conclude that $l = 1$ and that $\boldsymbol{X} = \lambda \boldsymbol{I}_n$. We now turn to the general case. Given that $\boldsymbol{K}_n(k) = \boldsymbol{K}_n(-k)^*$, we have

$$\boldsymbol{F}(k)\boldsymbol{X}^* = \boldsymbol{X}^*\boldsymbol{F}(k). \tag{23}$$

Combining (22) and (23)

$$\begin{cases} \boldsymbol{F}(k)\boldsymbol{\mathcal{A}} = \boldsymbol{\mathcal{A}}\boldsymbol{F}(k) \\ \boldsymbol{F}(k)\boldsymbol{\mathcal{B}} = \boldsymbol{\mathcal{B}}\boldsymbol{F}(k) \end{cases}$$

for all $k \in \mathbb{Z}$ and where $\boldsymbol{X} = \boldsymbol{\mathcal{A}} + i\boldsymbol{\mathcal{B}}$ denotes the Cartesian decomposition of $\boldsymbol{X}$, i.e., $\boldsymbol{\mathcal{A}} = \boldsymbol{X} + \boldsymbol{X}^*/2$, $\boldsymbol{\mathcal{B}} = \boldsymbol{X} - \boldsymbol{X}^*/2i$. Since both $\boldsymbol{\mathcal{A}}$ and $\boldsymbol{\mathcal{B}}$ are normal (in fact, Hermitean) matrices, it follows from the first part of the proof that $\boldsymbol{\mathcal{A}} = \lambda_r \boldsymbol{I}_n$ and $\boldsymbol{\mathcal{B}} = \lambda_i \boldsymbol{I}_n$ for some $\lambda_r, \lambda_i \in \mathbb{R}$. Thus, $\boldsymbol{X} = (\lambda_r + i\lambda_i)\boldsymbol{I}_n$.

Proof of Theorem 1: To settle notation

$$\begin{aligned} \boldsymbol{\mathcal{H}}(z) &= [\boldsymbol{h}_1(z) \cdots \boldsymbol{h}_p(z)] \\ \boldsymbol{h}_p(z) &= \sum_{k=0}^{D_p} \boldsymbol{h}_p(k) z^{-k} \in \mathbb{C}^M_{(D)_p}[z] \\ \boldsymbol{\mathcal{H}}_p &= [\boldsymbol{h}_p(0)\boldsymbol{h}_p(1) \cdots, \boldsymbol{h}_p(D_p)] \\ \boldsymbol{\mathcal{G}}(z) &= [\boldsymbol{g}_1(z) \cdots \boldsymbol{g}_p(z)] \\ \boldsymbol{g}_p(z) &= \sum_{k=0}^{E_p} \boldsymbol{g}_p(k) z^{-k} \in \mathbb{C}^M_{(E)_p}[z] \\ \boldsymbol{\mathcal{G}}_p &= [\boldsymbol{g}_p(0)\boldsymbol{g}_p(1) \cdots, \boldsymbol{g}_p(E_p)]. \end{aligned}$$

The proof $\mathcal{R}_{\boldsymbol{y}}(\boldsymbol{\mathcal{H}}(z)) = \mathcal{R}_{\boldsymbol{y}}(\boldsymbol{\mathcal{G}}(z)) \Rightarrow \boldsymbol{\mathcal{H}}(z) = \boldsymbol{\mathcal{G}}(z)$ (up to a phase offset per column) is carried out incrementally in three steps in which we establish

4) $\sum_{p=1}^P D_p = \sum_{p=1}^P E_p$;
5) $D_p = E_p$;
6) $\boldsymbol{\mathcal{G}}_p = \boldsymbol{\mathcal{H}}_p e^{i\theta_p}$.

In addition, for simplicity, we consider noiseless samples $\boldsymbol{y}(t)$ in (2), as our proof only relies on the SOS of $\boldsymbol{y}(t)$.

Step 1) Let $\mathcal{F}(z)$ denote the vector space of M-tuples of rational functions (with indeterminate z^{-1}) over the field of rational functions [6], [26]. Let $\mathcal{S}_{\boldsymbol{H}}(z) \subset \mathcal{F}(z)$ denote the P-dimensional subspace spanned by $\boldsymbol{\mathcal{H}}(z)$, and $\mathcal{S}^{\perp}_{\boldsymbol{H}}(z) \subset \mathcal{F}(z)$ its $(M-P)$-dimensional dual subspace [6], [26]. Similar definitions hold for $\boldsymbol{\mathcal{G}}(z)$. As shown in [6], if K (a stacking parameter) is high enough, then $\mathcal{S}^{\perp}_{\boldsymbol{H}}(z)$ is uniquely determined from the correlation matrix $\boldsymbol{R}_{\boldsymbol{y}_{\boldsymbol{K}}}(0)$, $\boldsymbol{y}_K(t) = \mathcal{S}_K[\boldsymbol{y}(t)]$. Thus, $\mathcal{S}^{\perp}_{\boldsymbol{H}}(z) = \mathcal{S}^{\perp}_{\boldsymbol{G}}(z)$ and $\mathcal{S}_{\boldsymbol{H}}(z) = \mathcal{S}_{\boldsymbol{G}}(z) = \mathcal{S}(z)$. Because both $\boldsymbol{\mathcal{H}}(z)$ and $\boldsymbol{\mathcal{G}}(z)$ are irreducible and column-reduced, they are minimal polynomial basis for $\mathcal{S}(z)$ [26]; consequently, they have the same order [26], i.e., $\sum_{p=1}^P D_p = \sum_{p=1}^P E_p$.

Step 2) Choose $K \leq N_{\max}$ [recall (A2)] such that $\boldsymbol{H} = [\boldsymbol{H}_1\boldsymbol{H}_2 \cdots \boldsymbol{H}_p]$ is full column rank, where $\boldsymbol{H}_p = \boldsymbol{\mathcal{T}}_K(\boldsymbol{h}_p)$. Let $\boldsymbol{R}_{\boldsymbol{y}_{\boldsymbol{K}}}(0) = \boldsymbol{V}\boldsymbol{\Sigma}^2\boldsymbol{V}^*$ denote a r-truncated EVD $r = \mathrm{rank}\{\boldsymbol{H}\}$, $\boldsymbol{V}^*\boldsymbol{V} = \boldsymbol{I}_r$, $\boldsymbol{\Sigma} = \mathrm{diag}(\sigma_1, \ldots, \sigma_r)(\sigma_i > 0)$. Set $\boldsymbol{P} = \boldsymbol{V}\boldsymbol{\Sigma}$. Then

$$\boldsymbol{P} = \boldsymbol{H}\boldsymbol{R}_{s_m}(0)^{1/2}\boldsymbol{Q}^* \tag{24}$$

where $\boldsymbol{m} = (M_1, \ldots, M_P)^{\mathrm{T}}$, $M_p = D_p + K$, $\boldsymbol{s}_{\boldsymbol{m}}(t) = (\boldsymbol{s}_1(t; M_1)^{\mathrm{T}}, \ldots, \boldsymbol{s}_P(t; M_P)^{\mathrm{T}})^{\mathrm{T}}$; $\boldsymbol{Q} : r \times r$ denotes some unitary matrix. Thus, $\boldsymbol{B}(\tau) \equiv \boldsymbol{P}^{\dagger}\boldsymbol{R}_{\boldsymbol{y}_{\boldsymbol{K}}}(\tau)\boldsymbol{P}^{\dagger *}$ is given by

$$\boldsymbol{B}(\tau) = \boldsymbol{Q}\boldsymbol{A}(\tau; m)\boldsymbol{Q}^* \tag{25}$$

$\boldsymbol{A}(\tau; \boldsymbol{m}) = \mathrm{diag}(\boldsymbol{A}_1(\tau; M_1), \ldots, \boldsymbol{A}_P(\tau; M_P))$. The same reasoning, in terms of $\boldsymbol{G} = [\boldsymbol{G}_1\boldsymbol{G}_2 \cdots \boldsymbol{G}_P]$, $\boldsymbol{G}_p = \boldsymbol{\mathcal{T}}_K(\boldsymbol{g}_p)$, leads to

$$\boldsymbol{P} = \boldsymbol{G}\boldsymbol{R}_{s_n}(0)^{1/2}\boldsymbol{W}^* \tag{26}$$

where $\boldsymbol{W} : r \times r$ (unitary), $\boldsymbol{n} = (N_1, \ldots, N_P)^{\mathrm{T}}$, $N_p = E_p + K$. Thus

$$\boldsymbol{B}(\tau) = \boldsymbol{W}\boldsymbol{A}(\tau; \boldsymbol{n})\boldsymbol{W}^* \tag{27}$$

We must prove $M_p = N_p$ for $p = 1, \ldots, P$. Assume the opposite. Since $\sum_p D_p = \sum_p E_p$ (Step 1), then $M_p > N_p$ for some p. Let $\tau_0 = \tau(p)$ be a correlation lag satisfying A4). In the sequel, the notation $x(t) \sim \boldsymbol{X}$ means that $x(t)$ is the characteristic polynomial of the $n \times n$ matrix $\boldsymbol{X}$, i.e., $x(t) = \det(t\boldsymbol{I}_n - \boldsymbol{X})$. In addition, let $\mathcal{R}\{x(t)\}$ denote the set of roots of $x(t)$ (including multiplicites); of course, $\sigma(\boldsymbol{X}) = \mathcal{R}\{x(t)\}$. Let $b(t) \sim \boldsymbol{B}(\tau_0)$. From (25), $b(t) = g_p(t)\prod_{q \neq p} g_q(t)$, where $g_l(t) \sim \boldsymbol{A}_l(\tau_0; M_l)$ for $l = 1, \ldots, P$. From (27), $b(t) = h_p(t)\prod_{q \neq p} h_q(t)$, where $h_l(t) \sim \boldsymbol{A}_l(\tau_0; N_l)$ for $l = 1, \ldots, P$. By A4), $\mathcal{R}\{g_p(t)\} \cap \mathcal{R}\{h_q(t)\} = \varnothing$ for $q \neq p$. Thus, necessarily, $\mathcal{R}\{g_p(t)\} \subset \mathcal{R}\{h_p(t)\}$; this is a contradiction since the cardinality of $\mathcal{R}\{g_p(t)\}$ (i.e., $M_p + 1$) is greater than the cardinality of $\mathcal{R}\{h_p(t)\}$ (i.e., $N_p + 1$). Thus, $D_p = E_p$ for $p = 1, \ldots, P$.

Step 3) From Step 2, $\boldsymbol{m} = \boldsymbol{n}$. In the sequel, $\boldsymbol{A}_p(\tau) \equiv \boldsymbol{A}_p(\tau; M_p) = \boldsymbol{A}_p(\tau; N_p)$. Thus, $\boldsymbol{A}(\tau) \equiv \mathrm{diag}(\boldsymbol{A}_1(\tau), \ldots, \boldsymbol{A}_P(\tau)) = \boldsymbol{A}(\tau; \boldsymbol{m}) = \boldsymbol{A}(\tau; \boldsymbol{n})$. Let $\boldsymbol{\Theta} \equiv \boldsymbol{Q}^*\boldsymbol{W}$ (unitary). From (25) and (27), $\boldsymbol{A}(\tau)\boldsymbol{\Theta} = \boldsymbol{\Theta}\boldsymbol{A}(\tau)$ for all $\tau \in \mathbb{Z}$. Letting

$$\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\Theta}_{11} & \boldsymbol{\Theta}_{12} & \cdots & \boldsymbol{\Theta}_{1P} \\ \boldsymbol{\Theta}_{21} & \boldsymbol{\Theta}_{22} & \cdots & \boldsymbol{\Theta}_{2P} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{\Theta}_{Pl} & \boldsymbol{\Theta}_{P2} & \cdots & \boldsymbol{\Theta}_{PP} \end{bmatrix}$$

$\boldsymbol{\Theta}_{pq} : (M_p+1) \times (M_q+1)$, we have $\boldsymbol{A}_p(\tau)\boldsymbol{\Theta}_{pq} = \boldsymbol{\Theta}_{pq}\boldsymbol{A}_q(\tau)$. Consider $p \neq q$. By A4), there is a correlation lag τ_0 such that $\sigma(\boldsymbol{A}_p(\tau_0)) \cap \sigma(\boldsymbol{A}_q(\tau_0)) = \emptyset$. Thus, $\boldsymbol{\Theta}_{pq} = \mathbf{0}$ for $p \neq q$, and $\boldsymbol{\Theta} = \mathrm{diag}(\boldsymbol{\Theta}_{11}, \ldots, \boldsymbol{\Theta}_{PP})$ with

$$\boldsymbol{A}(\tau)\boldsymbol{\Theta}_{pp} = \boldsymbol{\Theta}_{pp}\boldsymbol{A}(\tau), \quad \forall_{\tau \in Z}. \tag{28}$$

Let $\boldsymbol{\mathcal{T}}_{M_p}(\boldsymbol{c}_p) = \boldsymbol{U}_p\boldsymbol{\Sigma}_p\boldsymbol{V}_p^*$ denote a SVD, where $\boldsymbol{U}_p$ is unitary, $\boldsymbol{\Sigma}_p$ is nonsingular (because, without loss of generality, $c_p(0) \neq 0$, and $\boldsymbol{\mathcal{T}}_{M_p}(\boldsymbol{c}_p)$ has full row rank), and $\boldsymbol{V}_p$ is an isometry. Using this representation in (5) and (6), we have

$$\boldsymbol{A}_p(\tau) = \boldsymbol{U}_p\boldsymbol{V}_p^* = \boldsymbol{K}^{M_p+C_p+1}(\tau)\boldsymbol{V}_p\boldsymbol{U}_p^*. \tag{29}$$

Apply Lemma C to (28) and (29) to obtain $\boldsymbol{\Theta}_{pp} = e^{i\theta_p}\boldsymbol{I}_{M_p+1}$, for some $\theta_p \in \mathbb{R}$. From (24) and (26)

$$\begin{aligned} \boldsymbol{G} =& \boldsymbol{PWR}_{s_n}(0)^{-1/2} \\ =& \boldsymbol{HR}_{s_m}(0)^{-1/2}\boldsymbol{\Theta}\boldsymbol{R}_{s_n}(0)^{-1/2} \\ =& \boldsymbol{H}\boldsymbol{\Theta}. \end{aligned}$$

Thus, $\boldsymbol{G}_p = \boldsymbol{H}_p e^{i\theta_p}$, i.e., $\boldsymbol{\mathcal{G}}_p = \boldsymbol{\mathcal{H}}_p e^{i\theta_p}$.

APPENDIX B
PROOF OF THEOREM 2

Given $\boldsymbol{f}(z) = (f_1(z), \ldots, f_P(z))^{\mathrm{T}} \in \mathcal{M}_{\boldsymbol{C}}[z]$ and $\epsilon > 0$, we must provide $\boldsymbol{g}(z) = (g_1(z), \ldots, g_P(z))^{\mathrm{T}} \in \mathcal{F}[z]$ such that $\mathrm{dist}(\boldsymbol{f}(z), \boldsymbol{g}(z)) = \sqrt{\sum_{p=1}^{P} \left\| \boldsymbol{f}_p - \boldsymbol{g}_p \right\|^2} < \epsilon$. Before proceeding, we need some definitions. For $\theta \in \mathbb{R}$ and $p(z) = \sum_{k=0}^{d} p(k)z^{-k} \in \mathbb{C}[z]$, let $p(z;\theta) \equiv \sum_{k=0}^{d}(p(k)e^{ik\theta})z^{-k}$. For $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^{\mathrm{T}} \in \mathbb{R}^n$ and $\boldsymbol{p}(z) \in \mathbb{C}^n[z]$, set $\boldsymbol{p}(z;\boldsymbol{\theta}) \equiv (p_1(z;\theta_1), \ldots, p_n(z;\theta_n))^{\mathrm{T}}$. Finally, for $M_p \in \mathbb{N}$ and $c_p(z) \in \mathbb{C}_{(C_p)}[z]$, let $\boldsymbol{A}_p(\tau; M_p, c_p(z))$ denote the normalized correlation matrix of $\boldsymbol{s}_p(t) = \mathcal{S}_{M_p}[s_p(t)]$ at lag τ induced by $c_p(z)$; recall (5) and (6).

Notice that $\boldsymbol{f}(z;\boldsymbol{\theta}) \in \mathcal{M}_{\boldsymbol{C}}[z]$ is a continuous function of $\boldsymbol{\theta}$. In addition, some algebra reveals that irrespective of M_p, we have $\boldsymbol{A}_p(\tau; M_p, f_p(z;\theta_p)) \sim e^{i\tau\theta_p}\boldsymbol{A}_p(\tau; M_p, f_p(z))$, and $\boldsymbol{A}_p(C_p; M_p, f_p(z))$ is nonsingular. Thus, as θ_p varies, the spectrum of $\boldsymbol{A}_p(C_p; M_p, f_p(z;\theta_p))$ is rotated in the complex plane (origin excluded by the nonsingularity). Clearly, we can choose θ_p (as small as we want) such that $\sigma(\boldsymbol{A}_p(C_p; M_p, f_p(z;\theta_p))$ does not intersect a given finite set of points in $\mathbb{C}$. Apply this property for all possible M_p in order to satisfy A4), and let $\boldsymbol{g}(z) = \boldsymbol{f}(z;\boldsymbol{\theta})$.

APPENDIX C
PROOF OF THEOREM 3

Suppose $\varphi(\boldsymbol{l}) = 0$. Then, by the arguments in Step 2 of the proof of Theorem 1, we have $\boldsymbol{l} = \boldsymbol{m}$. In the sequel, $\boldsymbol{A}_p(\tau) = \boldsymbol{A}_p(\tau; M_p)$, and $\boldsymbol{A}(\tau) = \boldsymbol{A}(\tau; \boldsymbol{m})$. Equation (11) reads as

$$\begin{aligned} f(\boldsymbol{X}) &= \sum_{\tau \in Z} \|\boldsymbol{B}(\tau)\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}_p(\tau)\|^2 \\ &= \sum_{\tau \in Z} \|\boldsymbol{Q}\boldsymbol{A}(\tau)\boldsymbol{Q}^*\boldsymbol{X} - \boldsymbol{X}\boldsymbol{A}_p(\tau)\|^2. \end{aligned} \tag{30}$$

Write $\boldsymbol{Q} = [\boldsymbol{Q}_1\boldsymbol{Q}_2 \cdots \boldsymbol{Q}_P]$, where $\boldsymbol{Q}_p : L \times (M_p+1)$. Suppose that $f(\boldsymbol{X}) = 0$. Then

$$\boldsymbol{Q}\boldsymbol{A}(\tau)\boldsymbol{Q}^*\boldsymbol{X} = \boldsymbol{X}\boldsymbol{A}_p(\tau) \tag{31}$$

for all $\tau \in \mathbb{Z}$. Let $\boldsymbol{Y}_p = \boldsymbol{Q}_p^*\boldsymbol{X}$. Equation (31) is equivalent to

$$\begin{aligned} \boldsymbol{A}_p(\tau)\boldsymbol{Y}_p =& \boldsymbol{Y}_p\boldsymbol{A}_p(\tau) \\ \boldsymbol{A}_q(\tau)\boldsymbol{Y}_q =& \boldsymbol{Y}_q\boldsymbol{A}_p(\tau) \end{aligned} \tag{32}$$

for $q \neq p$ and $\tau \in \mathbb{Z}$. In the proof of Theorem 1, it was shown that (32) implies $\boldsymbol{Y}_p = \lambda_p\boldsymbol{I}_{M_p+1}$ for some $\lambda_p \in \mathbb{C}$. On the other hand, from A4), there is a correlation lag τ_0 that separates the spectra of $\boldsymbol{A}_p(\tau_0)$ and $\boldsymbol{A}_q(\tau_0)$. Thus, $\boldsymbol{Y}_q = \mathbf{0}$ for $q \neq p$. We conclude that $\boldsymbol{X}_p = \lambda_p\boldsymbol{Q}_p$. Normalizing $\boldsymbol{X}_p$ to get $\|\boldsymbol{X}_p\| = \sqrt{M_p+1}$ yields $\boldsymbol{W}_p = e^{i\theta_p}\boldsymbol{Q}_p$.

REFERENCES

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

[2] ——, "Fast blind equalization via antenna array," in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 3, 1993, pp. 272–275.

[3] ——, "Blind identification and equalization based on second-order statistics: a time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, pp. 340–349, Mar. 1994.

[4] E. Moulines, P. Duhamel, J. 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.

[5] A. Gorokhov and P. Loubaton, "Subspace-based techniques for blind separation of convolutive mixtures with temporally correlated sources," *IEEE Trans. Circuits Syst. I*, vol. 44, pp. 813–820, Sept. 1997.

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

[7] J. K. Tugnait, "On linear predictors for MIMO channels and related blind identification and equalization," *IEEE Signal Processing Lett.*, vol. 5, pp. 289–291, Nov. 1998.

[8] 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.*, vol. 5, 1999, pp. 2707–2710.

[9] ——, "Multi-step linear predictors-based blind equalization of multiple-input multiple-output channels," in *Proc. IEEE Int. Conf. Acoust., Speech. Signal Process.*, vol. 5, 1999, pp. 2949–2952.

[10] J. F. Cardoso and A. Souloumiac, "Blind beamforming for nongaussian signals," in *Proc. Inst. Elect. Eng. F, Radar Signal Process.*, vol. 140, Dec. 1993, pp. 362–370.

[11] A. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Trans. Signal Processing*, vol. 44, pp. 1136–1155, May 1996.

[12] ——, "Blind estimation of multiple co-channel digital signals using an antenna array," *IEEE Signal Processing Lett.*, vol. 1, pp. 29–31, Feb. 1994.

[13] ——, "Blind separation of synchronous co-channel digital signals using an antenna array-Part I: algorithms," *IEEE Trans. Signal Processing*, vol. 44, pp. 1184–1197, May 1996.

[14] S. Talwar and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array-part II: performance analysis," *IEEE Trans. Signal Processing*, vol. 45, pp. 706–718, Mar. 1997.

[15] B. Halder, B. Ng, A. Paulraj, and T. Kailath, "Unconditional naximum likelihood approach for blind estimation of digital signals," in *Proc. IEEE Int. Acoust., Speech Signal Process.*, vol. 2, 1996, pp. 1081–1084.

[16] V. Barroso, J. M. F. Moura, and J. Xavier, "Blind array channel division multiple access (AChDMA) for mobile communications," *IEEE Trans. Signal Processing*, vol. 46, pp. 737–752, Mar. 1998.

[17] J. Xavier and V. Barroso, "Blind source separation, ISI cancelation and carrier phase recovery in SDMA systems for mobile communications," *Wireless Personal Commun., Special Issue on Wireless Broadband Communications*, vol. 10, pp. 36–76, June 1999.

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

[19] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, “A blind source separation technique using second-order statistics,” *IEEE Trans. Signal Processing*, vol. 45, pp. 434–444, Feb. 1997.

[20] A. Chevreuil and P. Loubaton, “MIMO blind second-order equalization method and conjugate cyclostationarity,” *IEEE Trans. Signal Processing*, vol. 47, pp. 572–578, Feb. 1999.

[21] ——, “On the use of conjugate cyclostationarity: a blind second-order multi-user equalization method,” in *Proc. IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, 1996, pp. 2441–2445.

[22] J. Xavier, V. Barroso, and J. M. F. Moura, “Closed form blind channel identification and source separation in SDMA systems through correlative coding,” *IEEE J. Select. Areas Commun.*, vol. 16, pp. 1506–1517, Oct. 1998.

[23] M. Wax and T. Kailath, “Detection of signals by information theoretic criteria,” *IEEE Trans. Acoust., Speech, Signal Processing*, vol. ASSP-33, pp. 387–392, Apr. 1985.

[24] G. Forney Jr, “Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference,” *IEEE Trans. Inform. Theory*, vol. IT-8, pp. 363–378, May 1972.

[25] G. Golub and C. van Loan, *Matrix Computations*, 3rd ed. Baltimore, MD: John Hopkins Univ. Press, 1996.

[26] D. Forney, “Minimal bases of rational vector spaces, with applications to multivariable linear systems,” *SIAM J. Contr.*, vol. 13, no. 3, pp. 493–520, May 1975.

João Xavier (S’97) received the E.C.E degree in 1995 and the M.Sc. degree in electrical and computer engineering in 1997, both from the Instituto Superior Tecnico (IST), Lisbon, Portugal. Presently, he is pursuing the Ph.D. degree at IST.

In 1997, he joined the faculty of the Department of Electrical and Computer Engineering at IST, where he is with the Signal and Image Processing Group of the Institute of Systems and Robotics (ISR). His current research interests are in the area of statistical signal processing with applications to mobile wireless digital communications.

Victor A. N. Barroso (M’89–SM’00) was born in Lisbon, Portugal, on December 22, 1952. He received the E.E. degree in 1976 and the Ph.D. degree in electrical and computer engineering in 1990, both from Instituto Superior Técnico (IST), Lisbon, Portugal.

In 1976, he joined the faculty of the Department of Electrical and Computer Engineering at IST. He was promoted to Assistant Professor in 1990 and to Associate Professor in 1993. He has taught courses in systems and signal theory, control systems, signal and array processing, computer networks, and communications. Currently, he is the Head of the M.Sc. and the Ph.D. programs in Electrical and Computer Engineering at IST. He is also a Researcher at the Institute for Systems and Robotics (ISR), Lisbon, Portugal, where he is the coordinator of the Signal and Image Processing Group. His research interests are in statistical signal/image and array processing, wireless communications, and underwater acoustics. He has published over 70 journal and conference papers.

Dr. Barroso was an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1996 to 1999 and the representative of the IEEE Signal Processing Society at the IEEE Wireless Periodicals Working Group in 2000. He is affiliated with several IEEE societies.

José M. F. Moura (S’71–M’75–SM’90–F’94) received the engenheiro electrotécnico degree in 1969 from Instituto Superior Técnico (IST), Lisbon, Portugal, and the M.Sc., E.E., and D.Sc. degrees in electrical engineering and computer science from the Massachussetts Institute of Technology (MIT), Cambridge, in 1973 and 1975, respectively.

Since 1986, he has been a Professor of Electrical and Computer Engineering at Carnegie Mellon University (CMU), Pittsburgh, PA. From 1999 to 2000, he was a Visiting Professor of Electrical Engineering at MIT. Prior to this, he was on the faculty of IST as an Assistant Professor (1975), Professor Agregado (1978), and Professor Catedítico (1979), and held visiting appointments at MIT (Genrad Associate Professor of Electrical Engineering and Computer Science, from 1984 to 1986) and the University of Southern California, Los Angeles (Research Scholar, Department of Aerospace Engineering, Summers of 1978 to 1981). His research interests include statistical signal processing and telecommunications and image and video processing. He has published over 240 journal and conference papers, is the Co-Editor of two books, holds three U.S. patents in the areas of image and video processing and digital communications, and has given numerous invited seminars at U.S. and European Universities and Laboratories.

Dr. Moura serves as Vice President for Publications and as a member of the Board of Governors of the IEEE Signal Processing Society. He is also Vice President for publications for the IEEE Sensors Council. He has been on the Editorial Board of PROCEEDINGS OF THE IEEE since January 1999. He chairs the IEEE Signal Processing Society Publications Board. He was the Editor in Chief for the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1995 to 1999, and he is currently a member of the Multimedia Signal Processing Technical Committee, a founding member of the Sensor Array and Multichannel Technical Committee, and a past member of the Underwater Acoustics Technical Committee. He was a member of the IEEE Press Board from 1991 to 1995, an Associate Editor for several journals, and has been on the program committee for several IEEE international conferences and workshops. He is a corresponding member of the Academy of Sciences of Portugal. He is affiliated with several IEEE societies, Sigma Xi, AMS, and SIAM. In 2000, he received an IEEE Third Millennium Medal.