Blind Identification of FIR MIMO Channels by Decorrelating Subchannels

Yingbo Hua, *Fellow, IEEE*, Senjian An, and Yong Xiang

***Abstract*—We study blind identification and equalization of finite impulse response (FIR) and multi-input and multi-output (MIMO) channels driven by colored signals. We first show a sufficient condition for an FIR MIMO channel to be identifiable up to a scaling and permutation using the second-order statistics of the channel output. This condition is that the channel matrix is irreducible (but not necessarily column-reduced), and the input signals are mutually uncorrelated and of distinct power spectra. We also show that this condition is necessary in the sense that no single part of the condition can be further weakened without another part being strengthened. While the above condition is a strong result that sets a fundamental limit of blind identification, there does not yet exist a working algorithm under that condition. In the second part of this paper, we show that a method called blind identification via decorrelating subchannels (BIDS) can uniquely identify an FIR MIMO channel if a) the channel matrix is nonsingular (almost everywhere) and column-wise coprime and (b) the input signals are mutually uncorrelated and of sufficiently diverse power spectra. The BIDS method requires a weaker condition on the channel matrix than that required by most existing methods for the same problem.**

***Index Terms*—Adaptive signal processing, blind system identification, blind channel deconvolution, colored sources, decorrelation, MIMO channels, sensor array processing.**

I. Introduction

IDENTIFICATION and equalization of finite impulse response (FIR) and multi-input and multi-output (MIMO) channels channels driven by unknown colored signals are a fundamental problem encountered in a wide range of applications, which include speech enhancement using microphone arrays, wireless and mobile communications, video surveillance, and brain signal analysis. This paper focues on the important case where very little about the channel input is known. The FIR single-input multi-output (SIMO) channels studied in [1]–[3], [18], and [19] and the instantaneous MIMO channels studied in [4], [5], [18], and [19] are some extreme cases of the FIR MIMO channels. The FIR MIMO channels driven by colored signals are also in contrast to the FIR/IIR MIMO channels driven by white signals as studied in [7], [18], and [19]. In the latter case, the higher order statistics (HOS) have to be exploited to resolve the ambiguity of a unitary matrix that cannot otherwise be resolved using the second-order statistics. The previously developed methods for the same problem that we consider here are primarily the subspace algorithm [8] and the matrix pencil (MP) algorithm [9]. These two algorithms require the channel matrix to be irreducible and column reduced. This assumption is a barrier to many other methods. A recent work in [27] also requires the same condition and in addition a known input power spectra. Another recent work in [28] requires the channel matrix to have unit diagonal elements at all frequencies. The algorithms shown in [13] and [14] also assume the channel matrix to have unit diagonal elements. In this paper, we present a method called blind identification via decorrelating subchannels (BIDS). The BIDS method (see Fig. 1) is based on a set of decorrelators that decorrelate the output signals of subchannels. The BIDS method requires a channel condition weaker than "irreducible and column-reduced." Some prior developments of the BIDS method are available in [10]–[12].

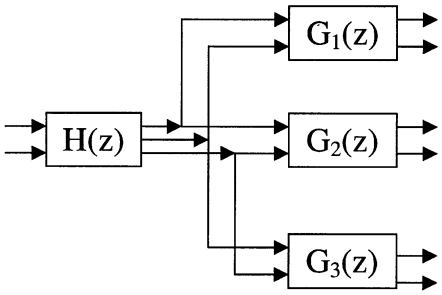


Fig. 1. Illustration of BIDS. $\mathbf{H}(z)$ denotes the channel matrix (i.e., the MIMO FIR channel transfer function matrix). The output signals of each $\mathbf{G}_i(z)$ are made mutually decorrelated. Each $\mathbf{G}_i(z)$ may have more input than output, depending on the dimension of $\mathbf{H}(z)$. BIDS-1 extracts out the input signals of $\mathbf{H}(z)$ from the output signals of $\mathbf{G}_i(z)$. BIDS-2 reconstructs $\mathbf{H}(z)$ from $\mathbf{G}_i(z)$ and then estimates the input signals of $\mathbf{H}(z)$.

In Section II, we address some of the algebraic properties of the second-order statistics of the channel output to establish a sufficient condition under which an FIR MIMO channel is identifiable up to a scaling and permutation. In particular, we show that an FIR MIMO channel is identifiable if the channel matrix is irreducible and the input signals are mutually uncorrelated and of distinct power spectra. This condition is also shown to be necessary in that no single part of the condition can be weakened independently. This condition sets a theoretical limit of blind identification of FIR MIMO channels.

In Section III, we present the BIDS method. Assuming that there are more output signals than input signals, the BIDS

Manuscript received June 21, 2002; revised November 21, 2002. This work was supported in part by the Australian Research Council, Hong Kong University of Science and Technology, Microsoft Research, and the U. S. National Science Foundation. The associate editor coordinating the review of this paper and approving it for publication was Dr. Chong-Yung Chi.

Y. Hua is with the Department of Electrical Engineering, University of California, Riverside, CA 92521 USA (e-mail: yhua@ee.ucr.edu).

S. An is with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Victoria 3010, Australia.

Y. Xiang was with the Department of Electrical and Electronic Engineering, University of Melbourne, Parkville, Victoria 3010, Australia. He is now with Deakin University, Victoria, Australia.

Digital Object Identifier 10.1109/TSP.2003.810295

method first partition the original channel into a set of subchannels and then construct a set of decorrelators that decorrelate the output signals of each of the subchannels. The decorrelators are then exploited in two different fashions, leading to the BIDS-1 and BIDS-2 algorithms. To estimate the channel matrix, the BIDS-2 algorithm allows the channel matrix to have certain zeros (i.e., nonirreducible). This remarkable property makes the BIDS method more robust to noise than other methods like the subspace method and the matrix pencil method.

In Section IV, we compare the performance of the BIDS method with that of the above two methods.

II. Some Fundamental Limits

We consider the following FIR MIMO channel:

$$\mathbf{y}(n) = \mathbf{H}(n) * \mathbf{x}(n) + \mathbf{w}(n) = \sum_{l=0}^{L_H} \mathbf{H}(l)\mathbf{x}(n-l) + \mathbf{w}(n) \quad (1)$$

where $*$ denotes linear convolution, $\mathbf{x}(n)$ is the vector of I input signals, $\mathbf{y}(n)$ is the vector of J output signals, $\mathbf{H}(n)$ is the $J \times I$ matrix sequence of the channel impulse responses, L_H is the maximum length of the finite impulse responses, and $\mathbf{w}(n)$ is a noise vector uncorrelated with $\mathbf{x}(n)$. In this paper, all parameters in (1) are assumed to be real valued. A convenient form of (1) is

$$\mathbf{y}(n) = \mathbf{H}_z(z)\,[\mathbf{x}(n)] + \mathbf{w}(n) \quad (2)$$

where $\mathbf{H}_z(z) = \sum_{l=0}^{L_H} \mathbf{H}(l) z^{-l}$ is the channel matrix, and $\mathbf{H}_z(z)\,[\mathbf{x}(n)]$ denotes the operation of $\mathbf{H}_z(z)$ on $\mathbf{x}(n)$. Note that there is a one-to-one relationship between the polynomial matrix $\mathbf{H}_z(z)$ and the sequence $\mathbf{H}(n)$. The goal is to estimate $\mathbf{x}(n)$ and/or $\mathbf{H}_z(z)$ using $\mathbf{y}(n)$. In the sequal, we will use $\mathbf{H}(z)$ to represent $\mathbf{H}_z(z)$ for convenience.

Without any further constraint, the FIR MIMO system is ambiguous to any unimodular polynomial matrix[1] $\mathbf{T}(z)$ (the inverse of which is a polynomial matrix as well), i.e.,

$$\mathbf{y}(n) = \mathbf{H}(z)\,[\mathbf{x}(n)] + \mathbf{w}(n) = \mathbf{H}'(z)\,[\mathbf{x}'(n)] + \mathbf{w}(n)$$

where $\mathbf{H}'(z) = \mathbf{H}(z)\mathbf{T}(z)^{-1}$, and $\mathbf{x}'(n) = \mathbf{T}(z)[\mathbf{x}(n)]$. An additional constraint must be available to reduce the ambiguity.

We assume that the data sequence $\mathbf{y}(n)$ is long enough so that the second-order statistics (SOS) of $\mathbf{y}(n)$ can be exploited.[2] We write the autocorrelation of $\mathbf{R}_{\mathbf{yy}}(\tau)$ of $\mathbf{y}(n)$ as

$$\mathbf{R}_{\mathbf{yy}}(\tau) \hat{=} \lim_{N\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} \mathbf{y}(n)\mathbf{y}^T(n-\tau) \quad (3a)$$

and the autocorrelation $\mathbf{R}_{\mathbf{xx}}(\tau)$ of $\mathbf{x}(n)$ is defined in the same way. The power spectral matrix $\mathbf{S}_{\mathbf{yy}}(z)$ of $y(n)$ is defined as[3]

$$\mathbf{S}_{\mathbf{yy}}(z) \hat{=} \sum_{\tau=-\infty}^{\infty} \mathbf{R}_{\mathbf{yy}}(\tau) z^{-\tau} \quad (3b)$$

and the power spectral matrices $\mathbf{S}_{\mathbf{xx}}(z)$ and $\mathbf{S}_{\mathbf{ww}}(z)$ of $\mathbf{x}(n)$ and $\mathbf{w}(n)$, respectively, are similarly defined. It is natural to assume $\sum_{\tau=-\infty}^{\infty} |(\mathbf{R}_{\mathbf{xx}}(\tau))_{i,j}| < \infty$ such that all power spectra considered here exist. Then, it requires only a standard procedure [21] to show that

$$\mathbf{S}_{\mathbf{yy}}(z) = \mathbf{H}(z)\mathbf{S}_{\mathbf{xx}}(z)\mathbf{H}(z^{-1})^T + \mathbf{S}_{\mathbf{ww}}(z). \quad (3c)$$

Note that the power spectral matrix $\mathbf{S}_{\mathbf{yy}}(z)$ contains all the SOS information of the channel output. We say that an FIR MIMO channel is identifiable using the SOS of $\mathbf{y}(n)$ if $\mathbf{S}_{\mathbf{yy}}(z)$ implies $\mathbf{H}(z)$, $\mathbf{S}_{\mathbf{xx}}(z)$, and $\mathbf{S}_{\mathbf{ww}}(z)$ uniquely, up to some simple scaling and permutation.

Without a further constraint on $\mathbf{H}(z)$ and/or $\mathbf{x}(n)$, an FIR MIMO channel is still not identifiable, but for many applications, we can assume that the elements in the input vector $\mathbf{x}(n)$ are mutually uncorrelated, i.e., $\mathbf{S}_{\mathbf{xx}}(z)$ is a diagonal matrix, and the noise elements in $\mathbf{w}(n)$ are also mutually uncorrelated and of equal power spectra, i.e., $\mathbf{S}_{\mathbf{ww}}(z) = \rho(z)\rho(z^{-1})\mathbf{I}$, where $\rho(z)$ is a rational function. Furthermore, we assume that there are more output signals than input signals, i.e., $J > I$ (more sensors than sources). In this case, $\rho(z)\rho(z^{-1})$ is identifiable from $\mathbf{S}_{\mathbf{yy}}(z)$. Note that if z is on the unit circle, then $\rho(z)\rho(z^{-1})$ is real valued and is the least eigenvalue of $\mathbf{S}_{\mathbf{yy}}(z)$. If $\rho(z)\rho(z^{-1})$ is given for all z on the unit circle, then $\rho(z)\rho(z^{-1})$ at any z is given via analytic continuation. In other words, with $\rho(z)\rho(z^{-1})$ available on the unit circle, we can obtain the autocorrelation of the noise via an inverse Fourier transform and consequently obtain $\rho(z)\rho(z^{-1})$ by taking the Z-transform of the autocorrelation. A more general statement is given next.

Theorem 1: An FIR MIMO system is identifiable (i.e., $\mathbf{S}_{\mathbf{yy}}(z)$ implies $\mathbf{H}(z)$ uniquely up to a column-wise scaling and column-wise permutation) if we have the folowing.

- A1) $\mathbf{H}(z)$ is irreducible.
- A2) $\mathbf{S}_{\mathbf{xx}}(z)$ is diagonal and of distinct diagonal (polynomial or rational) functions.
- A3) $\mathbf{S}_{\mathbf{ww}}(z) = \rho(z)\rho(z^{-1})\mathbf{I}$ and $J > I$.

Furthermore, if any of the above subconditions A1–A3 is weakened independently, the FIR MIMO system is not identifiable.

Remark: An irreducible $\mathbf{H}(z)$ is a polynomial matrix that has a full rank for every z except for $z = 0$. Given $J > I$, $\mathbf{H}(z)$ is a tall matrix. If all rows of $\mathbf{H}(z)$ are chosen randomly with a given maximum degree, $\mathbf{H}(z)$ is irreducible with probability one.

Proof: The sufficiency of Theorem 1 follows from the above discussion of $\rho(z)\rho(z^{-1})$ and a proof given in [15]. The necessity of Theorem 1 is shown next.

1) Necessity of A1: We show here two counter examples.[4]

Example 1: This example shows that there can be a maximum-phase channel matrix and a minimum-phase channel matrix such that both yield the same output power spectral

[1]A unimodular polynomial matrix is a polynomial matrix of constant determinant.

[2]Higher order statistics would require even a longer sequence to ensure the same level of accuracy as the second-order statistics.

[3]Strictly speaking, the power spectral matrix corresponds to the case where z is on the unit circle.

[4]The counter examples were discovered through a tedious and *ad hoc* procedure. Although the counter examples are sufficient and rigorous for the purpose here, a theory that governs the existence of such counter examples has yet to be developed.

matrix $\mathbf{S_{yy}}(z)$ *even if* the input signals have distinct power spectra. Let

$$\mathbf{S_{xx}}(z)=\begin{bmatrix}(z+2)\left(z^{-1}+2\right) & 0\\ 0 & 1\end{bmatrix}$$
$$\mathbf{H}(z)=\begin{bmatrix}-64 & 4z^{-1}+37\\ -32 & 4z^{-1}+23\\ -16 & 4z^{-1}+16\end{bmatrix}$$
$$\hat{\mathbf{H}}(z)=\begin{bmatrix}12 & 132z^{-1}+61\\ 4 & 68z^{-1}+31\\ 0 & 36z^{-1}+16\end{bmatrix}$$

where $\mathbf{H}(z)$ has a zero $-4/9$, and $\hat{\mathbf{H}}(z)$ has a zero $-9/4$. A zero of a polynomial matrix is a value of z at which the matrix does not have a full rank. $\mathbf{H}(z)$ is a minimum-phase matrix, and $\hat{\mathbf{H}}(z)$ is a maximum-phase matrix. One can verify that

$$\hat{\mathbf{H}}(z)\mathbf{S_{xx}}(z)\hat{\mathbf{H}}^T(z^{-1})=\mathbf{H}(z)\mathbf{S_{xx}}(z)\mathbf{H}^T(z^{-1}).$$

Example 2: This example[5] shows that there can be two different minimum-phase channel matrices such that both yield the same output power spectral matrix $\mathbf{S_{yy}}(z)$, *even if* the input signals have distinct power spectra. Let

$$\mathbf{H}(z)=\mathbf{H}_0(z)\begin{bmatrix}2z^{-1}+5 & 3z^{-1}+7\\ 1 & 1\end{bmatrix}$$
$$\mathbf{S}(z)=\begin{bmatrix}2(z+3)\left(z^{-1}+3\right) & 0\\ 0 & 2\end{bmatrix}$$
$$\hat{\mathbf{H}}(z)=\mathbf{H}_0(z)\begin{bmatrix}2z^{-2}+14z^{-1}+22 & 2z^{-1}+4\\ z^{-1}+4 & 1\end{bmatrix}$$
$$\hat{\mathbf{S}}(z)=\begin{bmatrix}1 & 0\\ 0 & (z+2)\left(z^{-1}+2\right)\end{bmatrix}$$

where $\mathbf{H}_0(z)$ is any irreducible matrix. It is easy to verify that $\mathbf{H}(z)$ has a zero $-1/2$, $\hat{\mathbf{H}}(z)$ has a zero $-1/3$, and furthermore

$$\hat{\mathbf{H}}(z)\hat{\mathbf{S}}_{\mathbf{xx}}(z)\hat{\mathbf{H}}^T\left(z^{-1}\right)=\mathbf{H}(z)\mathbf{S_{xx}}(z)\mathbf{H}^T\left(z^{-1}\right)$$

where both $\mathbf{H}(z)$ and $\hat{\mathbf{H}}(z)$ are minimum-phase.

Necessity of A2: If $\mathbf{S_{xx}}(z)$ has some identical diagonal functions (disregarding scaling), then the corresponding columns of $\mathbf{H}(z)$ are ambiguous by a unitary rotation $\mathbf{Q}$ (in addition to column-wise scaling and permutation), i.e., there is a unitary $\mathbf{Q}$ such that $\mathbf{H}(z)\mathbf{S_{xx}}(z)\mathbf{H}(z^{-1})^T = \hat{\mathbf{H}}(z)\mathbf{Q}\hat{\mathbf{S}}_{\mathbf{xx}}(z)\mathbf{Q}^T\hat{\mathbf{H}}(z^{-1})^T$, where $\hat{\mathbf{H}}(z)$ is a column-wise scaled version of $\mathbf{H}(z)$, and $\hat{\mathbf{S}}_{\mathbf{xx}}(z)$ is a corresponding diagonal-wise scaled version of $\mathbf{S_{xx}}(z)$.

Necessity of A3: If $\mathbf{S_{ww}}(z)$ is not necessarily diagonal and is completely unknown, it is clearly not distinguishable from the other unknown component $\mathbf{H}(z)\mathbf{S_{xx}}(z)\mathbf{H}^T(z^{-1})$.

The sufficiency and "necessity"[6] of the conditions A1–A3 of Theorem 1 are now established. It sets a fundamental limit on the blind identifiability of FIR MIMO channels. However, thus far there exists no working algorithm under the condition. In the next section, we show the method of BIDS, which exploits the spectral diversity of the input signals and consequently requires a weaker condition on the channel matrix.

[5]This example unfortunately counters a previously reported result shown in [10] and [23].

[6]The necessity implied here should not be confused with other possible meanings of the same term.

III. Blind Identification via Decorrelating Subchannels

A. Decorrelating Subchannels

In the sequel, we will ignore the noise term for convenience. The effect of white noise can be mitigated in a conventional fashion as mentioned later. We now write

$$\mathbf{y}(n)=\mathbf{H}(z)[\mathbf{x}(n)]. \tag{4}$$

Let $\mathbf{S}_i$ be a $I\times J$ selection matrix. The BIDS algorithms first form subchannel output vectors as follows:

$$\mathbf{y}_i(n)\hat{=}\mathbf{S}_i\mathbf{y}(n) \tag{5}$$

where $i=1,2,\ldots,M$, and $M\hat{=}J!/((J-I)!I!)$, which is the total number of subchannels. For each i, the BIDS method searches for a decorrelator $\mathbf{G}_i(z)$ such that the power spectral matrix $\mathbf{S}_{\mathbf{u}_i\mathbf{u}_i}(z)$ of

$$\mathbf{u}_i(n)\hat{=}\mathbf{G}_i(z)\mathbf{y}_i(n) \tag{6}$$

is diagonal. Let $\mathbf{G}_i(z)=\sum_{l=0}^{L_G}\mathbf{G}_i(l)z^{-1}$. Then, we can write

$$\mathbf{u}_i(n)=\begin{bmatrix}\mathbf{G}_i(0) & \mathbf{G}_i(1) & \cdots & \mathbf{G}_i(L_G)\end{bmatrix}\begin{bmatrix}\mathbf{y}_i(n)\\ \mathbf{y}_i(n-1)\\ \vdots\\ \mathbf{y}_i(n-L_G)\end{bmatrix}$$

or simply $\mathbf{u}_i(n)=\bar{\mathbf{G}}_i\bar{\mathbf{y}}_i(n)$. The autocorrelation matrix of $\mathbf{u}_i(n)$ can be computed (estimated) as

$$\hat{\mathbf{R}}_{\mathbf{u}_i\mathbf{u}_i}(\tau)=\frac{1}{N}\sum_{n=0}^{N-1}\mathbf{u}_i(n)\mathbf{u}_i^T(n-\tau).$$

Equivalently, $\hat{\mathbf{R}}_{\mathbf{u}_i\mathbf{u}_i}(\tau)=\bar{\mathbf{G}}_i\hat{\mathbf{R}}_{\bar{\mathbf{y}}_i\bar{\mathbf{y}}_i}(\tau)\bar{\mathbf{G}}_i^T$, where $\hat{\mathbf{R}}_{\bar{\mathbf{y}}_i\bar{\mathbf{y}}_i}(\tau)$ is the autocorrelation matrix of $\bar{\mathbf{y}}_i(n)$, i.e., we have the equation at the bottom of the next page, with

$$\hat{\mathbf{R}}_{\mathbf{yy}}(\tau)=\frac{1}{N}\sum_{n=0}^{N-1}\mathbf{y}(n)\mathbf{y}^T(n-\tau).$$

If the noise is spatially and temporally white, then $\mathbf{R_{yy}}(\tau)=\mathbf{R_{ss}}(\tau)+\delta(\tau)\sigma^2\mathbf{I}$, where $\mathbf{s}(n)=\mathbf{H}(z)[\mathbf{x}(n)]$, and σ^2 is the noise variance. Given that $\mathbf{H}(z)$ is a tall matrix, σ^2 is the smallest eigenvalue of $\mathbf{R_{yy}}(0)$. Hence, the noise contribution can be asymptotically removed from $\mathbf{R_{yy}}(\tau)$. Several variations for estimating σ^2 from a set of the smallest eigenvalues of $\mathbf{R_{yy}}(0)$ are possible. However, we will not address this issue further here. The least eigenvalue will be used to estimate σ^2 in our simulation.

The cost function for constructing the decorrelator can be defined as the mean squared values of the off diagonal elements of $\hat{\mathbf{R}}_{\mathbf{u}_i\mathbf{u}_i}(\tau)$ over a sufficient range of τ, i.e.,

$$E_i=\frac{1}{L_yI(I-1)}\sum_{\tau=0}^{L_y}\sum_{j\neq k}\left(\mathbf{g}_{i,k}^T\hat{\mathbf{R}}_{\bar{\mathbf{y}}_i\bar{\mathbf{y}}_i}(\tau)\mathbf{g}_{i,j}\right)^2$$

where $g_{i,j}$ is defined by

$$\bar{\mathbf{G}}_i^T=\begin{bmatrix}\mathbf{g}_{i,1} & \mathbf{g}_{i,2} & \cdots & \mathbf{g}_{i,I}\end{bmatrix}.$$

The cost function E_i is a nonquadratic function of $\bar{\mathbf{G}}_i$ in the strict sense, but it is quadratic with respect to each (individual) row of $\bar{\mathbf{G}}_i$. This leads to a simple algorithm to minimize E_i, which consists of a sequence of the following sweeps until convergence. During each sweep, the rows of $\bar{\mathbf{G}}_i$ are updated sequentially. When updated, each row of $\bar{\mathbf{G}}_i$ minimizes E_i with all other rows fixed. To avoid trivial solutions, each row of $\bar{\mathbf{G}}_i$ is constrained to yield a constant diagonal of $\hat{\mathbf{R}}_{\mathbf{u}_i\mathbf{u}_i}(0)$, i.e.,

$$\mathbf{g}_{i,k}^T \hat{\mathbf{R}}_{\bar{\mathbf{y}}_i\bar{\mathbf{y}}_i}(0)\mathbf{g}_{i,k} = 1.$$

This minimization with respect to each row of $\bar{\mathbf{G}}_i$ is a standard linear-quadratic problem, but the overall minimization with respect to the whole matrix $\bar{\mathbf{G}}_i$ may converge to a local minimum. Therefore, in practice, several initial values for $\bar{\mathbf{G}}_i$ are required to ensure that the global minimum of E_i is achieved with a high likelihood. If there is some prior estimate of $\mathbf{H}(z)$, it should be used to initialize $\bar{\mathbf{G}}_i$.

A related idea shown in [22] was recently brought to our attention, but the work shown there was incomplete and was not continued since then (according to its authors). Theorem 3 in [22] is incorrect, a counter result of which is Theorem 2, which will be shown later in this paper.

B. Conditions for Decorrelation Implying Separation

We now go back to (6) and discuss the conditions under which a decorrelation of $\mathbf{u}_i(n)$ leads to a separation of $\mathbf{x}(n)$. We can rewrite (6) as $\mathbf{u}_i(n) = \mathbf{G}_i(z)\mathbf{y}_i(n) = \mathbf{G}_i(z)\mathbf{H}_i(z)\mathbf{x}(n)$, where $\mathbf{H}_i(z) = \mathbf{S}_i\mathbf{H}(z)$. We show next that with a proper choice of $\deg(\mathbf{G}_i(z))$, the product $\mathbf{G}_i(z)\mathbf{H}_i(z)$ is diagonalizable by $\mathbf{G}_i(z)$.

Lemma 1: Provided $\deg(\mathbf{G}_i(z)) \geq (I-1)\deg(\mathbf{H}_i(z))$, there exists a $\mathbf{G}_i(z)$ such that $\mathbf{G}_i(z)\mathbf{H}_i(z)$ is diagonal.

Proof: It is easy to verify that under the condition of the lemma, the number of linear and scalar equations corresponding to all off-diagonal polynomial elements of $\mathbf{G}_i(z)\mathbf{H}_i(z)$ is no larger than the total number of the scalar parameters in $\mathbf{G}_i(z)$ minus I. The extra degree I of freedom here corresponds to the number of nonzero rows of $\mathbf{G}_i(z)$. Hence, there exists a $\mathbf{G}_i(z)$ (with all nonzero rows) such that $\mathbf{G}_i(z)\mathbf{H}_i(z)$ is diagonal.

If $\mathbf{G}_i(z)\mathbf{H}_i(z)$ is diagonal, then each element of $\mathbf{u}_i(n)$ corresponds to a distinct element of the input signal $\mathbf{x}(n)$ and, hence, $\mathbf{G}_i(z)$ is said to be a "signal separator." It is such separators $\mathbf{G}_i(z)$ that will be further exploited by the BIDS method. However, a decorrelator is not necessarily a separator unless a diversity condition on $\mathbf{S}_{\mathbf{xx}}(z)$ is satisfied. This condition will be detailed later after we introduce the definition of spectral diversity.

Definition of Diversity: Given two polynomials, we say that the distinction of one polynomial from the other is the number of *distinct* zeros of the first polynomial that are not shared by the second polynomial. The diversity of two polynomials is defined to be the larger distinction of the two polynomials. The diversity of two diagonal functions of a power spectral matrix is defined to be half the diversity of the two functions. The diversity of a power spectral matrix is defined to be the minimum diversity between any two diagonal functions of the matrix. The diversity of $\mathbf{S}_{\mathbf{xx}}(z)$ will be denoted by $\text{div}(\mathbf{S}_{\mathbf{xx}}(z))$. If $\mathbf{S}_{\mathbf{xx}}(z)$ is rational, we first write $\mathbf{S}_{\mathbf{xx}}(z) = (1/q(z)q(z^{-1}))\mathbf{S}'_{\mathbf{xx}}(z)$, where $q(z)q(z^{-1})$ is the least common multiple of the denominators of the diagonal elements of $\mathbf{S}_{\mathbf{xx}}(z)$, and $\mathbf{S}'_{\mathbf{xx}}(z)$ is a polynomial power spectral matrix. We then define $\text{div}\,(\mathbf{S}_{\mathbf{xx}}(z)) = \text{div}\,(\mathbf{S}'_{\mathbf{xx}}(z))$.

The notion of diversity is a key for the study of the identifiability conditions of the BIDS method as shown below.

Theorem 2: Let $C(z)$ be a nonsingular (i.e., full rank almost everywhere) $I \times I$ polynomial matrix. The diagonalization of $\mathbf{C}(z)\mathbf{S}_{\mathbf{xx}}(z)\mathbf{C}(z^{-1})^T$ implies the diagonalization of $C(z)$ up to a row permutation if

$$\text{div}\,(\mathbf{S}_{\mathbf{xx}}(z)) > (I-1)\deg(\mathbf{C}(z)). \tag{7}$$

Proof: See Appendix A.

Remark: The necessity of an alternative condition $div(\mathbf{S}_{\mathbf{xx}}(z)) > \deg(\mathbf{C}(z))$ [as opposed to (7)] was discussed and established in [10], but the sufficiency of $div(\mathbf{S}_{\mathbf{xx}}(z)) > \deg(\mathbf{C}(z))$ has not been proved or disproved.[7]

Corollary 1: Provided that $\mathbf{G}_i(z)\mathbf{H}_i(z)$ is nonsingular and

$$\begin{aligned}(I-1)\deg(\mathbf{H}_i(z)) &\leq \deg(\mathbf{G}_i(z)) \\ &< \frac{\text{div}\,(\mathbf{S}_{\mathbf{xx}}(z))}{I-1} - \deg(\mathbf{H}_i(z))\end{aligned} \tag{8}$$

there exists $\mathbf{G}_i(z)$ such that $\mathbf{S}_{\mathbf{u}_i\mathbf{u}_i}(z)$ is diagonal, and the diagonalization of $\mathbf{S}_{\mathbf{u}_i\mathbf{u}_i}(z)$ also implies that $\mathbf{G}_i(z)\mathbf{H}_i(z)$ is diagonal up to a row permutation.

Proof: The proof follows from Lemma 1 and Theorem 2.

Note that $\mathbf{G}_i(z)\mathbf{H}_i(z)$ is nonsingular if and only if $\mathbf{S}_{\mathbf{u}_i\mathbf{u}_i}(z)$ is nonsingular (since $\mathbf{S}_{\mathbf{xx}}(z)$ is nonsingular). The nonsingularity condition on

$$\mathbf{g}_{i,k}^T \hat{\mathbf{R}}_{\bar{\mathbf{y}}_i\bar{\mathbf{y}}_i}(0)\mathbf{g}_{i,k} = 1$$

can be ensured in the BIDS method with the constraint.

[7]An error in the previous report [23] was later discovered.

$$\hat{\mathbf{R}}_{\bar{\mathbf{y}}_i\bar{\mathbf{y}}_i}(\tau) = \begin{bmatrix} \hat{\mathbf{R}}_{\mathbf{y}_i\mathbf{y}_i}(\tau) & \hat{\mathbf{R}}_{\mathbf{y}_i\mathbf{y}_i}(\tau+1) & \cdots & \hat{\mathbf{R}}_{\mathbf{y}_i\mathbf{y}_i}(\tau+L_G) \\ \hat{\mathbf{R}}_{\mathbf{y}_i\mathbf{y}_i}(\tau-1) & \hat{\mathbf{R}}_{\mathbf{y}_i\mathbf{y}_i}(\tau) & \cdots & \hat{\mathbf{R}}_{\mathbf{y}_i\mathbf{y}_i}(\tau+L_G-1) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathbf{R}}_{\mathbf{y}_i\mathbf{y}_i}(\tau-L_G) & \hat{\mathbf{R}}_{\mathbf{y}_i\mathbf{y}_i}(\tau-L_G+1) & \cdots & \hat{\mathbf{R}}_{\mathbf{y}_i\mathbf{y}_i}(\tau) \end{bmatrix}$$

and

$$\hat{\mathbf{R}}_{\mathbf{y}_i\mathbf{y}_i}(\tau) = \mathbf{S}_i\hat{\mathbf{R}}_{\mathbf{yy}}(\tau)\mathbf{S}_i^T$$

C. Sorting Out Subchannel Decorrelators

Assuming that the condition of Corollary 1 is met, we can now find such a $\mathbf{G}_i(z)$ by diagonalizing $\mathbf{S}_{\mathbf{u}_i\mathbf{u}_i}(z)$ that $\mathbf{G}_i(z)\mathbf{H}_i(z)$ is diagonal up to an unknown row permutation. The row permutations in the computed $\mathbf{G}_i(z)$ for different i are generally different. In this subsection, we show how to find these permutations needed to sort out $\mathbf{G}_i(z)$.

In this section, we use $\mathbf{G}_i(z)$ to denote the ideal decorrelator such that $\mathbf{G}_i(z)\mathbf{H}_i(z)$ is exactly diagonal. A permuted version of $\mathbf{G}_i(z)$ is given by $\hat{\mathbf{G}}_i(z) = \mathbf{P}_i\mathbf{G}_i(z)$, where $\mathbf{P}_i$ is an unknown permutation matrix. Consider the cross spectral matrix between $\mathbf{u}_1$ and $\mathbf{u}_i$:

$$\begin{aligned}\mathbf{S}_{\mathbf{u}_1\mathbf{u}_i}(z)&=\mathbf{G}_1(z)\mathbf{H}_1(z)\mathbf{S}_{\mathbf{xx}}(z)\mathbf{H}_i(z^{-1})^T\mathbf{G}_i(z^{-1})^T\\&=\mathbf{P}_1^T\hat{\mathbf{G}}_1(z)\mathbf{H}_1(z)\mathbf{S}_{\mathbf{xx}}(z)\mathbf{H}_i(z^{-1})^T\hat{\mathbf{G}}_i(z^{-1})^T\mathbf{P}_i\\&=\mathbf{P}_1^T\hat{\mathbf{G}}_1(z)\mathbf{S}_1\mathbf{S}_{\mathbf{yy}}(z)\mathbf{S}_i^T\hat{\mathbf{G}}_i(z^{-1})^T\mathbf{P}_i\\&=\mathbf{P}_1^T\hat{\mathbf{S}}_{\mathbf{u}_1\mathbf{u}_i}(z)\mathbf{P}_i\end{aligned} \tag{9}$$

where $\hat{\mathbf{S}}_{\mathbf{u}_1\mathbf{u}_i}(z) = \hat{\mathbf{G}}_1(z)\mathbf{S}_1\mathbf{S}_{\mathbf{yy}}(z)\mathbf{S}_i^T\hat{\mathbf{G}}_i(z^{-1})^T$, which is computable from given data. Note that the (ideal) cross spectral matrix $\mathbf{S}_{\mathbf{u}_1\mathbf{u}_i}(z)$ is diagonal, given that $\mathbf{S}_{\mathbf{xx}}(z)$ is diagonal. The time-domain expression of (9) is

$$\mathbf{R}_{\mathbf{u}_1\mathbf{u}_i}(\tau) = \mathbf{P}_1^T\hat{\mathbf{R}}_{\mathbf{u}_1\mathbf{u}_i}(\tau)\mathbf{P}_i \tag{10}$$

where $\mathbf{R}_{\mathbf{u}_1\mathbf{u}_i}(\tau)$ is ideal and diagonal, and

$$\hat{\mathbf{R}}_{\mathbf{u}_1\mathbf{u}_i}(\tau) = \int_{-0.5}^{0.5}\hat{\mathbf{S}}_{\mathbf{u}_1\mathbf{u}_i}(e^{j2\pi f})e^{j2\pi f\tau}df \tag{11a}$$

$$\hat{\mathbf{S}}_{\mathbf{u}_1\mathbf{u}_i}(z) = \sum_{\tau=-\infty}^{\infty}\hat{\mathbf{R}}_{\mathbf{u}_1\mathbf{u}_i}(\tau)z^{-\tau}. \tag{11b}$$

Note that the above definitions of the autocorrelation matrices are for the asymptotical case, i.e., $N = \infty$, but the actual implementations use a finite N. Define the operator η as follows:

$$\eta(x) = \begin{cases}1, & |x| > \varepsilon \geq 0\\ 0, & \text{otherwise}\end{cases} \tag{12}$$

$$(\eta(\mathbf{A}))_{i,j} = \eta(\mathbf{A}_{i,j}). \tag{13}$$

In the ideal situation where the decorrelation is perfect, ϵ is zero. Therefore, ideally, we have [from (10)]

$$\mathbf{P}_{1i}\hat{=}\eta\left(\sum_{\tau}\left|\hat{\mathbf{R}}_{\mathbf{u}_1\mathbf{u}_i}(\tau)\right|^2\right) = \mathbf{P}_1\mathbf{P}_i^T \tag{14}$$

or equivalently

$$\mathbf{P}_{1i}\hat{=}\eta\left(\int_{-0.5}^{0.5}\left|\hat{\mathbf{S}}_{\mathbf{u}_1\mathbf{u}_i}(e^{j2\pi f})\right|^2 df\right) = \mathbf{P}_1\mathbf{P}_i^T \tag{15}$$

where $|\mathbf{A}|^2$ is a matrix of $\mathbf{A}$ with all elements replaced by their squared magnitudes. Then, we have

$$\mathbf{P}_{1i}\hat{\mathbf{G}}_i(z) = \mathbf{P}_1\mathbf{G}_i(z). \tag{16}$$

We see that the matrix $\mathbf{P}_{1i}\hat{\mathbf{G}}_i(z)$ for any i is equal to the corresponding ideal decorrelator $\mathbf{G}_i(z)$ up to a (unknown) permutation matrix independent of i. Therefore, the jth row of $\mathbf{P}_{1i}\hat{\mathbf{G}}_i(z)$ for all i is associated with a distinct input signal (common for all i) or, equivalently, a distinct column of $\mathbf{H}(z)$.

Without loss of generality, we can next ignore $\mathbf{P}_1$ in (16) and assume that $\mathbf{G}_i(z)$ is available such that $\mathbf{G}_i(z)\mathbf{H}_i(z)\hat{=}\mathbf{C}_i(z) =$ diagonal.

D. Properties of Subchannel Decorrelators

We now show some properties of the subchannel matrices and subchannel decorrelators. These will be useful for establishing the BIDS algorithms. Let $\mathbf{H}_i(z)$, $i = 1, 2, \ldots, M$ be all of the $I \times I$ square submatrices of $\mathbf{H}(z)$. Assuming that $\mathbf{H}(z)$ has a full column rank for almost all z, we have Lemmas 2–5.

Lemma 2: Each (nonzero) row of $\mathbf{H}(z)$ must be a row of an $I \times I$ square submatrix $\mathbf{H}_i(z)$ that is nonsingular (i.e., with full rank for almost all z).

Proof: We know that $\mathbf{H}(z)$ must have at least one $I \times I$ square submatrix $\mathbf{H}_i(z)$ that is nonsingular. Now, choose any row $\mathbf{r}(z)$ from the rest of the rows of $\mathbf{H}(z)$. $\mathbf{r}(z)$ must belong to the span of $K(\leq I)$ rows of $\mathbf{H}_i(z)$ for almost all z. Substituting $\mathbf{r}(z)$ for any of these K rows, we form another nonsingular submatrix.

Lemma 3: All the submatrices $\mathbf{H}_i(z)$ that are nonsingular are "chained" together in the sense that every two nonsingular submatrices share a common row either within the two submatrices or with another.

Proof: Suppose $\mathbf{H}_i(z)$ and $\mathbf{H}_j(z)$ are nonsingular and that they do not have a common row. Then, we know that not all rows of $\mathbf{H}_i(z)$ belong to (for almost all z) the span of any chosen $I-1$ rows of $\mathbf{H}_j(z)$. Then, for a row of $\mathbf{H}_i(z)$ that does not belong to (for almost all z) the span of the chosen $I-1$ rows of $\mathbf{H}_j(z)$, we can form another nonsingular submatrix $\mathbf{H}_k(z)$ that shares one row with $\mathbf{H}_i(z)$ and $I-1$ rows with $\mathbf{H}_j(z)$.

Lemma 4: If $\mathbf{H}_i(z)$ is nonsingular, so is the corresponding $\mathbf{G}_i(z)$.

Proof: If $\mathbf{H}_i(z)$ is nonsingular, we know that $\mathbf{G}_i(z)$ is such that $\mathbf{S}_{\mathbf{u}_i\mathbf{u}_i}(z) = \mathbf{G}_i(z)\mathbf{H}_i(z)\mathbf{S}_{\mathbf{xx}}(z)\mathbf{H}_i(z^{-1})^T\mathbf{G}_i(z^{-1})^T$ is nonsingular diagonal, and hence, $\mathbf{G}_i(z)$ is nonsingular.

We now let $\mathbf{G}_i(z)$, $i = 1, 2, \ldots, M'$ (where $M \geq M' \geq J$) is all of the nonsingular decorrelators. Furthermore, we can assume that each row of $\mathbf{G}_i(z)$ is coprime. If it is not, one can extract out the greatest common divisor (GCD) of each row of $\mathbf{G}_i(z)$ by following the method shown in [16] or [24]. We can now define that for $i = 1, 2, \ldots, M'$

$$\begin{aligned}\mathbf{G}_i(z)\mathbf{H}_i(z)\hat{=}\mathbf{C}_i(z) &= \text{diag}\{\, c_{i,1}(z) \quad c_{i,2}(z) \quad \cdots \quad c_{i,I}(z)\,\}\\ \mathbf{c}_j(z) &= [\, c_{1,j}(z) \quad c_{2,j}(z) \quad \cdots \quad c_{M',j}(z)\,]^T.\end{aligned}$$

Lemma 5: For any $j = 1, 2, \ldots, I$, any zero of $\mathbf{c}_j(z)$ is a zero of $\mathbf{H}(z)$.

Proof: Suppose α is a zero of $\mathbf{c}_j(z)$, i.e., $\mathbf{c}_j(\alpha) = 0$. Then, the jth diagonal element of the diagonal matrix $\mathbf{G}_i(\alpha)\mathbf{H}_i(\alpha)$ must be zero for all $i = 1, 2, \ldots, M'$. Since none of the rows of $\mathbf{G}_i(\alpha)$ for any i is zero, the I rows of $\mathbf{H}_i(\alpha)$ must be linearly dependent for all $i = 1, 2, \ldots, M'$. Now, by Lemma 4,

$\mathbf{H}_i(\alpha)$ is singular for all $i = 1, 2, \ldots, M$. (Note that for any $i = M' + 1, \ldots, M$, $\mathbf{H}_i(z)$ does not have full rank for any z.) Hence, $H(\alpha)$ has a rank less than I, i.e., α is a zero of $H(z)$.

Corollary 2: If $\mathbf{H}(z)$ is irreducible, so is $\mathbf{c}_j(z)$ for all $j = 1, 2, \ldots, M'$.

Proof: It follows from Lemma 5.

E. BIDS-1 Algorithm

We are now ready to present the BIDS-1 algorithm. Recall $\mathbf{u}_i(n) = \mathbf{G}_i(z)\mathbf{y}_i(n) = \mathbf{G}_i(z)\mathbf{H}_i(z)\mathbf{x}(n)$. Define that for $i = 1, 2, \ldots, M'$

$$\mathbf{u}_i(n) = [\, u_{i,1}(n) \quad u_{i,2}(n) \quad \cdots \quad u_{i,I}(n)\,]^T$$
$$\mathbf{v}_j(n) = [\, u_{1,j}(n) \quad u_{2,j}(n) \quad \cdots \quad u_{M',j}(n)\,]^T.$$

It follows that $\mathbf{v}_j(n) = \mathbf{c}_j(z)x_j(n)$, which is an FIR SIMO system. If $\mathbf{H}(z)$ is irreducible (and, hence, so is $\mathbf{c}_j(z)$), one can apply any of the SIMO system methods [1]–[3] to retrieve $x_j(n)$ from $\mathbf{v}_j(n)$ for each j. If $\mathbf{H}(z)$ is not irreducible, we can write $\mathbf{c}_j(z) = \bar{\mathbf{c}}_j(z)c_j(z)$ where all zeros of $c_j(z)$ belong to the zeros of $\mathbf{H}(z)$, and $\bar{\mathbf{c}}_j(z)$ is coprime. The SIMO methods can then be used to retrieve $c_j(z)x_j(n)$ from $\mathbf{v}_j(n)$. The SIMO system method shown [1] will be used in conjunction with the BIDS-1 algorithm in the simulation. If $\mathbf{H}(z)$ is not irreducible, $c_j(z)$ is unknown, and hence, $c_j(z)x_j(n)$ would be an unknown distortion of the desired signal $x_j(n)$.

Once the input signals $x_j(n)$ for all $j = 1, 2, \ldots, J$ are estimated, one can then use both the input and output signals to estimate $\mathbf{H}(z)$. A least square fitting method would suffice.

The BIDS-1 algorithm not only shows some interesting insights into the FIR MIMO channel but also has some useful features. In particular, the BIDS-1 algorithm suggests that one can obtain the input signals before the channel response is obtained, and the relatively mature methods for SIMO systems can be readily applied here.

Summary of BIDS-1 Algorithm

Step 1) Compute $\hat{\mathbf{R}}_{\mathbf{yy}}(\tau) = (1/N)\sum_{n=0}^{N-1} \mathbf{y}(n)\mathbf{y}^T(n-\tau)$ for $\tau = 0, 1, 2, \ldots, L_R$, where L_y is such that $L_y \ll N$ and $\hat{\mathbf{R}}_{\mathbf{yy}}(\tau) \cong 0$ for $|\tau| > L_y$. Note that $\hat{\mathbf{R}}_{\mathbf{yy}}(\tau) = \left(\hat{\mathbf{R}}_{\mathbf{yy}}(-\tau)\right)^T$.

Step 2) Remove the noise contribution from $\hat{\mathbf{R}}_{\mathbf{yy}}(\tau)$ and then construct $\hat{\mathbf{R}}_{\bar{\mathbf{y}}_i\bar{\mathbf{y}}_i}(\tau)$ for each (square) subchannel i.

Step 3) Carry out the following minimization for each subchannel i:

$$\min \frac{1}{L_y I(I-1)} \sum_{\tau=0}^{L_y} \sum_{j \neq k} \left(\mathbf{g}_{i,k}^T \hat{\mathbf{R}}_{\bar{\mathbf{y}}_i\bar{\mathbf{y}}_i}(\tau)\mathbf{g}_{i,j}\right)^2$$
$$\text{subject to } \mathbf{g}_{i,k}^T \hat{\mathbf{R}}_{\bar{\mathbf{y}}_i\bar{\mathbf{y}}_i}(0)\mathbf{g}_{i,k} = 1.$$

Step 4) Construct the decorrelator $\mathbf{G}_i(z)$ from $\bar{\mathbf{G}}_i^T = [\, \mathbf{g}_{i,1} \quad \mathbf{g}_{i,2} \quad \cdots \quad \mathbf{g}_{i,I}\,]$ for each subchannel i.

Step 5) Compute the sorting matrix $\mathbf{P}_{1i}$ for each subchannel i.

Step 6) Multiply the decorrelator output vector $\mathbf{u}_i(n) = \mathbf{G}_i(z)[\mathbf{y}_i(n)]$ by $\mathbf{P}_{1i}$ (from left) for each subchannel i.

Step 7) Construct the vector sequence $\mathbf{v}_j(n) = [\, u_{1,j}(n) \quad u_{2,j}(n) \quad \cdots \quad u_{M',j}(n)\,]^T$ for each input signal j.

Step 8) For each j, apply a SIMO system technique [18], [19] to estimate $x_j(n)$ based on the model $\mathbf{v}_j(n) = \mathbf{c}_j(z)\{x_j(n)\}$, where $\mathbf{c}_j(z)$ is an unknown FIR channel vector.

F. BIDS-2 Algorithm

The BIDS-2 algorithm first estimates the channel matrix $\mathbf{H}(z)$ from the decorrelators $\mathbf{G}_i(z)$ and then estimates the input signals using the estimated $\mathbf{H}(z)$. Recall that we can obtain the row-wise coprime nonsingular matrices $\mathbf{G}_i(z)$ for $i = 1, 2, \ldots, M'$ such that $\mathbf{G}_i(z)\mathbf{H}_i(z)$ for $i = 1, 2, \ldots, M'$ are diagonal. Furthermore, given the assumption that $\mathbf{H}(z)$ has full column rank for almost all z, we have Lemmas 2–4. We now define

$$\mathbf{H}_i(z) = (\, \mathbf{h}_{i,1}(z) \quad \mathbf{h}_{i,2}(z) \quad \cdots \quad \mathbf{h}_{i,I}(z)\,)$$
$$\mathbf{G}_{i,p}(z) = \mathbf{G}_i(z) \text{ without its } p\text{th row.}$$

Then, we know that for $i = 1, 2, \ldots, M'$ and $p = 1, 2, \ldots, M$

$$\mathbf{G}_{i,p}(z)\mathbf{h}_{i,p}(z) = 0. \tag{17}$$

Since $\mathbf{G}_{i,p}(z)$ has the rank $I-1$ for almost all z, the solution to $\mathbf{G}_{i,p}(z)\hat{\mathbf{h}}_{i,p}(z) = 0$ for each i is $\hat{\mathbf{h}}_{i,p}(z) = \mathbf{h}_{i,p}(z)f_{i,p}(z)$, where $f_{i,p}(z)$ is a scalar polynomial. From Lemma 3, we know that $\mathbf{h}_{i,p}(z)$ for $i = 1, 2, \ldots, M'$ are "chained" together through shared elements. Hence, the solution of $\hat{\mathbf{h}}_{i,p}(z)$ to

$$\mathbf{G}_{i,p}(z)\hat{\mathbf{h}}_{i,p}(z) = 0 \text{ for all } i = 1, 2, \ldots, M' \tag{18}$$

where $\hat{\mathbf{h}}_{i,p}(z)$ has the same overlapping (or chained) pattern as $\mathbf{h}_{i,p}(z)$, is then $\hat{\mathbf{h}}_{i,p}(z) = \mathbf{h}_{i,p}(z)f_p(z)$, where $f_p(z)$ is independent of i. In other words, (18) yields the pth column of the channel matrix $\mathbf{H}(z)$ up to a common polynomial. This leads to the following lemma.

Lemma 6: If $\mathbf{H}(z)$ is nonsingular (full column rank for almost all z) and column-wise coprime (each column is a coprime vector), then each column of $\mathbf{H}(z)$ can be found uniquely (up to scaling) from (18).

Proof: With the column-wise coprimeness, the common factor $f_p(z)$ in the solution of (18) can be removed by following the method in [16] or [24].

Equation (18) can be made more explicit once all the nonsingular $\mathbf{H}_i(z)$, or equivalently all $\mathbf{G}_i(z)$, are detected. For example, consider a channel matrix $\mathbf{H}(z)$ of the dimension $J \times I$ with $J = I+1$. Let $\mathbf{H}_i(z)$ be $\mathbf{H}(z)$ without its ith row. If $\mathbf{H}_i(z)$, for all $i = 1, 2, \ldots, J$, are nonsingular, then (18) can be written as

$$\bar{\mathbf{G}}_p(z)\hat{\mathbf{h}}_p(z) = 0 \tag{19}$$

with

$$\bar{\mathbf{G}}_p(z) = \begin{bmatrix} \mathbf{G}_{1,p}(z)' \\ \mathbf{G}_{2,p}(z)' \\ \vdots \\ \mathbf{G}_{J,p}(z)' \end{bmatrix}$$

where $\mathbf{G}_{i,p}(z)'$ is $\mathbf{G}_{i,p}(z)$ with a zero column inserted before its ith column, and $\hat{\mathbf{H}}(z) \hat{=} (\,\hat{\mathbf{h}}_1(z) \quad \hat{\mathbf{h}}_2(z) \quad \cdots \quad \hat{\mathbf{h}}_I(z)\,)$ is the estimate of $\mathbf{H}(z)$.

To describe (19) in more detail, let $L_H \geq \max_i \deg(\mathbf{H}_i(z))$ and $L_G = \max_i \deg(\mathbf{G}_i(z))$, which must meet the condition of Corollary 1. Note that we do not need to know the exact degree (but an upper bound) of $\mathbf{H}(z)$. Then, we write $\hat{\mathbf{h}}_p(z) = \sum_{i=0}^{L_H} \hat{\mathbf{h}}_p(i) z^{-i}$ and $\bar{\mathbf{G}}_p(z) = \sum_{i=0}^{L_G} \bar{\mathbf{h}}_p(i) z^{-i}$. Then, the polynomial matrix (19) can be reformulated into the numerical matrix equation

$$\bar{\mathbf{G}}_p \hat{\mathbf{h}}_p = 0 \tag{20}$$

where

$$\hat{\mathbf{h}}_p = \begin{bmatrix} \hat{\mathbf{h}}_p(0) \\ \hat{\mathbf{h}}_p(1) \\ \vdots \\ \hat{\mathbf{h}}_p(L_H) \end{bmatrix}$$

$$\bar{\mathbf{G}}_p = \begin{bmatrix} \bar{\mathbf{G}}_p(0) & & & \\ \vdots & \bar{\mathbf{G}}_p(0) & & \\ \bar{\mathbf{G}}_p(L_G) & \vdots & \ddots & \\ & \bar{\mathbf{G}}_p(L_G) & \ddots & \bar{\mathbf{G}}_p(0) \\ & & \ddots & \vdots \\ & & & \bar{\mathbf{G}}_p(L_G) \end{bmatrix}.$$

The least square solution to (20) is given by the least eigenvector of $\bar{\mathbf{G}}_p^T \bar{\mathbf{G}}_p$. Any common factor in $\hat{\mathbf{h}}_p(z)$ can be removed, and hence, the complete channel matrix $\mathbf{H}(z)$ can be estimated even if $\mathbf{H}(z)$ has zeros under the condition of Lemma 6.

To estimate the input signals, we can reformulate the data model (1) into a slide-window channel model as follows:

$$\mathbf{y}_{sw}(n) = \mathbf{H}_{sw} \mathbf{x}_{sw}(n) + \mathbf{w}_{sw}(n) \tag{21}$$

where

$$\mathbf{y}_{sw}(n) = \begin{bmatrix} \mathbf{y}(n) \\ \mathbf{y}(n-1) \\ \vdots \\ \mathbf{y}(n-W+1) \end{bmatrix}$$

$$\mathbf{H}_{sw} = \begin{bmatrix} \mathbf{H}(0) & \cdots & \cdots & \mathbf{H}(L_H) & & & \\ & \mathbf{H}(0) & \cdots & \cdots & \mathbf{H}(L_H) & & \\ & & \ddots & & & \ddots & \\ & & & \mathbf{H}(0) & \cdots & \cdots & \mathbf{H}(L_H) \end{bmatrix} \in R^{WJ \times (W+L_H)I}$$

$$\mathbf{x}_{sw}(n) = \begin{bmatrix} \mathbf{x}(n) \\ \mathbf{x}(n-1) \\ \vdots \\ \mathbf{x}(n-W-L_H+1) \end{bmatrix}.$$

It is known [17], [20], [25] that the generalized Sylvestor matrix $\mathbf{H}_{sw}$ has full column rank if $W \geq IL_H$ and $\mathbf{H}(z)$ is irreducible, column-reduced, and of equal column degrees. Under these conditions, (21), with any given n, has the unique solution $\mathbf{x}_{sw}(n)$ in the absence of noise.

A more general result is shown here.

Lemma 7: If $W > IL_H$ and $\mathbf{H}(z)$ is irreducible (but is not necessarily column-reduced or of equal column degrees), then in the absence of noise, the $((W+L_H)I) \times 1$ solution $\mathbf{x}_{sw}(n)$ to (21) with a fixed n contains the unique $x(n)$ (although $\mathbf{x}_{sw}(n)$ is not necessarily unique in itself), i.e., the first $I \times 1$ subvector $x(n)$ of $\mathbf{x}_{sw}(n)$ is unique.

Proof: It is known [7] that if $\mathbf{H}(z)$ has the degree L_H and the dimension $J \times I (J > I)$ and is irreducible, then there is a $\mathbf{G}(z)$ of degree $L_G \leq IL_H$ such that $\mathbf{G}(z)\mathbf{H}(z) = I$, which means that $\mathbf{G}(z)[\mathbf{y}(n)] = \mathbf{G}(z)\mathbf{H}(z)[\mathbf{x}(n)] = \mathbf{x}(n)$. It also means that for $W > IL_H \geq L_G$

$$\begin{aligned} \mathbf{G}(z)\,[\mathbf{y}(n)] &= [\,\mathbf{G}(0) \quad \mathbf{G}(1) \quad \cdots \quad \mathbf{G}(L_G) \quad \cdots\,]\,\mathbf{y}_{sw}(n) \\ &= \mathbf{x}(n) \end{aligned}$$

where $\mathbf{G}(z) = \sum_{i=0}^{L_G} \mathbf{G}(i) z^{-i}$. Hence, $x(n)$ must be unique given $\mathbf{y}_{sw}(n)$.

Lemma 7 implies that except for $((W+L_H)I) - I$ vector samples of the vector sequence $\mathbf{x}(n)$, all other vector samples of $\mathbf{x}(n)$ can be obtained from the vector sequence $\mathbf{y}(n)$ if $\mathbf{H}(z)$ is irreducible. When the number of available samples is large, this number $((W+L_H)I) - I$ can be negligible.

Therefore, like the BIDS-1 algorithm, the BIDS-2 algorithm yields the exact input signals in the absence of noise if $\mathbf{H}(z)$ is irreducible and the subchannel decorrelation is ideal. However, unlike the BIDS-1 algorithm, the BIDS-2 algorithm yields an estimate of $\mathbf{H}(z)$ that can reveal whether or not $\mathbf{H}(z)$ is irreducible. To do that, one can form a vector $\mathbf{h}_H(z)$ where each element is a determinant of an $I \times I$ square submatrix of $\mathbf{H}(z)$. Let the vector $\mathbf{h}_H(z)$ contain all such polynomials. It follows that any zero of $\mathbf{h}_H(z)$ is a zero of $\mathbf{H}(z)$ and vice versa. One can apply the method shown in [16] or [24] to find all zeros of $\mathbf{h}_H(z)$. Furthermore, unlike the BIDS-1 algorithm, the estimation of $\mathbf{H}(z)$ by the BIDS-2 algorithm does not require $\mathbf{H}(z)$ to be irreducible, and hence, the estimate of $\mathbf{H}(z)$ by the BIDS-2 algorithm is generally more accurate than that by the BIDS-1 algorithm.

Summary of BIDS-2 Algorithm

Steps 1–6) Same as BIDS-1 algorithm.

Step 7) For each $p = 1, 2, \ldots, M-1$, construct $\bar{\mathbf{G}}_p(z)$ (and, hence, $\bar{\mathbf{G}}_p$) from $\mathbf{G}_i(z), i = 1, \ldots, M'$.

Step 8) For each $p = 1, 2, \ldots, M-1$, solve $\bar{\mathbf{G}}_p \hat{\mathbf{h}}_p = 0$ in least squares for $\hat{\mathbf{h}}_p$, and hence, $\hat{\mathbf{h}}_p(z)$.

Step 9) Remove the GCD from $\hat{\mathbf{h}}_p(z)$ for each p to obtain the estimate of the p th column of $\mathbf{H}(z)$.

Step 10) For each n, solve $\mathbf{y}_{sw}(n) = \mathbf{H}_{sw}\mathbf{x}_{sw}(n)$ in least squares for $\mathbf{x}_{sw}(n)$, and hence, $\mathbf{x}(n)$. Alternatively, follow a discussion shown below.

The accuracy of Step 10 depends on the condition of the generalized Sylvestor matrix $\mathbf{H}_{sw}$. This matrix does not necessarily have a full column rank even if $\mathbf{H}(z)$ is irreducible. As an alternative of Step 10, one can use the following approach to estimate $\mathbf{x}(n)$ once $\mathbf{G}_i(z)$ and $\mathbf{H}_i(z)$ are available. Without

loss of generality, we can assume that the common factor of each row of $\mathbf{G}_i(z)$ is removed. Then, we can assume that

$$\deg(\mathbf{G}_i(z)) = (I-1)\deg(\mathbf{H}_i(z))$$
and
$$\mathbf{G}_i(z)\mathbf{H}_i(z) = \det(\mathbf{H}_i(z))\,\mathbf{I}.$$

It is then easy to show that for each $j = 1, 2, \ldots, I$, the SIMO system defined for BIDS-1 algorithm becomes

$$\mathbf{v}_j(n) = \mathbf{h}(z)\{x_j(n)\}$$

where

$$\mathbf{h}(z) = \begin{bmatrix} \det(\mathbf{H}_1(z)) \\ \det(\mathbf{H}_2(z)) \\ \vdots \\ \det(\mathbf{H}_{M'}(z)) \end{bmatrix}.$$

Note that this polynomial vector $\mathbf{h}(z)$ is coprime if and only if $\mathbf{H}(z)$ is irreducible. Therefore, the generalized Sylvestor matrix of $\mathbf{h}(z)$ is of full column rank if and only if $\mathbf{H}(z)$ is irreducible. The inversion of $\mathbf{v}_j(n) = \mathbf{h}(z)\lfloor x_j(n)\rfloor$ is hence more robust than the previous approach of Step 10. Namely, in Step 10, we can replace the MIMO slide-window model $\mathbf{y}_{sw}(n) = \mathbf{H}_{sw}\mathbf{x}_{sw}(n)$ by the SIMO slide-window model $\mathbf{v}_{j,sw}(n) = \mathbf{H}_{h,sw}x_{j,sw}(n)$ for $j = 1, 2, \ldots, I$ and compute the input by

$$\mathbf{x}_{j,sw}(n) = \left(\mathbf{H}_{h,sw}^T\mathbf{H}_{h,sw}\right)^{-1}\mathbf{H}_{h,sw}^T\mathbf{v}_{j,sw}(n)$$

where

$$\mathbf{x}_{j,sw}(n) = \begin{bmatrix} x_j(n) \\ x_j(n-1) \\ \\ x_j(n-W+1) \end{bmatrix}.$$

Another modification is that we can choose $x_j(n-l)$ from $\mathbf{x}_{j,sw}(n)$ if the $(l+1)$th row of the pseudoinverse $(\mathbf{H}_{h,sw}^T\mathbf{H}_{h,sw})^{-1}\mathbf{H}_{h,sw}^T$ has the minimum norm among all rows. This choice of input estimation has a further improvement of robustness against noise.

G. Finding the Number of Input Signals

We now discuss the issue of finding the number of input signals. It is known that a nonsingular (full rank almost everywhere) polynomial channel matrix $\mathbf{H}(z)$ can be decomposed into the following form [17]:

$$\mathbf{H}(z) = \mathbf{Q}(z)\mathbf{R}(z)$$

where $\mathbf{Q}(z)$ is irreducible and column-reduced, and $\mathbf{R}(z)$ is a nonsingular square polynomial matrix. In fact, $\mathbf{R}(z)$ is a greatest common divisor (GCD) of $\mathbf{H}(z)$. Then, the data model (2) becomes

$$\mathbf{y}(z) = \mathbf{Q}(z)\mathbf{z}(n) + \mathbf{w}(n)$$

where $\mathbf{z}(n) = \mathbf{R}(z)\mathbf{x}(n)$. Similar to (21), we can reformulate the above into a slide-window form:

$$\mathbf{y}_{sw}(n) = \mathbf{Q}_{sw}\mathbf{z}_{sw}(n) + \mathbf{w}_{sw}(n)$$

where $\mathbf{Q}_{sw}$ and $\mathbf{z}_{sw}(n)$ are defined in the same way as $\mathbf{H}_{sw}$ and $\mathbf{x}_{sw}(n)$. It is known [17], [20], [25] that if $W \geq IL_Q \geq D_Q$ (a weaker condition is shown in [25]), then

$$r_H \hat{=} \text{rank}(\mathbf{Q}_{sw}) = IW + D_Q$$

where $L_Q = \deg(\mathbf{Q}(z)) \leq \deg(\mathbf{H}(z))$, and D_Q is the sum of the column degrees of $\mathbf{Q}(z)$. Note that D_Q is independent of W. Therefore, if r_H is given for two or more values of W, then I can be found (via linear regression). r_H can be estimated from the distribution of the eigenvalues of the following covariance matrix:

$$\bar{\mathbf{R}}_{\mathbf{yy}}(0) \hat{=} \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}_{sw}(n)\mathbf{y}_{sw}(n)^T.$$

For large N, we have

$$\bar{\mathbf{R}}_{\mathbf{yy}}(0) = \mathbf{Q}_{sw}\bar{\mathbf{R}}_{\mathbf{zz}}(0)\mathbf{Q}_{sw}^T + \bar{\mathbf{R}}_{\mathbf{ww}}(0)$$

where $\bar{\mathbf{R}}_{\mathbf{zz}}(0)$ and $\bar{\mathbf{R}}_{\mathbf{ww}}(0)$ are the covariance matrices of $\mathbf{z}_{sw}(n)$ and $\mathbf{w}_{sw}(n)$, respectively. For large N, if $\bar{\mathbf{R}}_{\mathbf{ww}}(0)$ is proportional to identity (i.e., white noise) and $\bar{\mathbf{R}}_{\mathbf{zz}}(0)$ is nonsingular, then the first r_H eigenvalues of $\bar{\mathbf{R}}_{\mathbf{yy}}(0)$ are strictly larger than the rest of its eigenvalues, and all the rest of the eigenvalues are equal. Various statistical criteria are available to estimate r_H from the eigenvalues of $\bar{\mathbf{R}}_{\mathbf{yy}}(0)$ [21], but we will not pursue this topic further here.

IV. Simulation Examples

In this section, we present some examples to illustrate the performance of the BIDS-1 and BIDS-2 algorithms. We also compare the BIDS algorithms with the matrix pencil algorithm developed in [9] and the subspace method developed in [8], [25], and [26]. The subspace method yields an estimate of $\mathbf{H}(z)$ up to a constant matrix factor if $\mathbf{H}(z)$ is irreducible, column-reduced, and of equal column degrees. To retrieve the constant matrix (up to a column-wise scaling and permutation), one needs to use additional methods. The methods shown in [5] and [6] are good choices for incorporation into the subspace method for blind identification of the FIR MIMO system driven by colored signals.

The performance measure considered here is the mean-squared-error (MSE) of the channel estimation and the MSE of the signal estimation, i.e., we have the equation at the bottom of the next page, where R is the total number of runs ($R = 600$), N the total number of estimated samples of the input as well as (approximately) the number of available output samples, Λ_r is a diagonal matrix, and $\mathbf{P}_r$ is a permutation matrix. $\hat{\mathbf{H}}_r(l)$ and $\hat{\mathbf{s}}_r(n)$ are the estimates of $\mathbf{H}(l)$ and $\mathbf{s}(n)$, respectively, at each run.

At each run, the input sequences and the noise sequences are all randomly selected. The noise sequences are white Gaussian and of variance σ^2. The signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = 10\log_{10}\frac{\frac{1}{N}\sum_{n=1}^{N}\|\mathbf{y}(n)\|^2 - J\sigma^2}{J\sigma^2}.$$

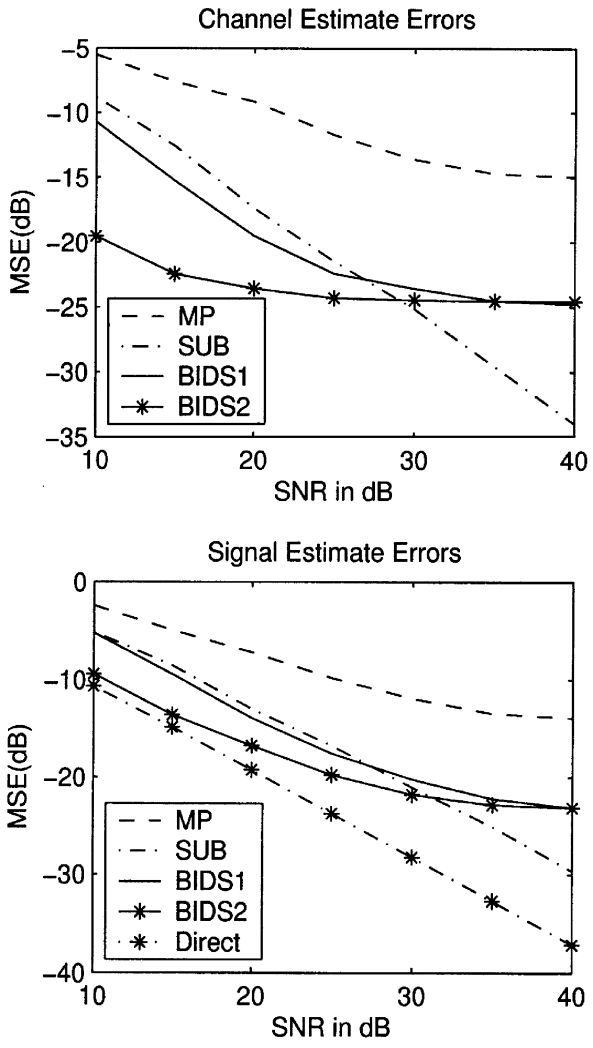


Fig. 2. MSE versus SNR of the BIDS-1 and BIDS-2 algorithms and two other algorithms for randomly chosen channels. The "Direct" algorithm (nonblind) uses the exact channel parameters to estimate the input signal. $N = 10\,000$.

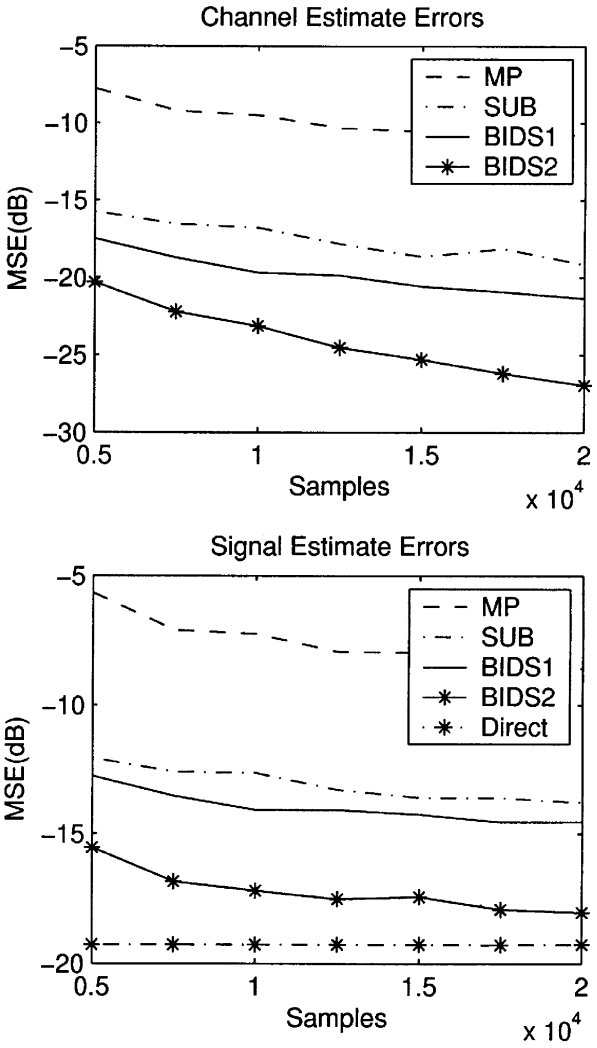


Fig. 3. MSE versus sample size N of the BIDS-1 and BIDS-2 algorithms and two other algorithms for randomly chosen channels. SNR $= 20$ dB.

The 3×1 input sequences are generated as follows:

$$\mathbf{x}(n) = \text{diag}\left(\frac{1}{p_1(z)}, \frac{1}{p_2(z)}, \frac{1}{p_3(z)}\right)[\mathbf{q}(n)]$$

where each entry in $\mathbf{q}(n)$ is a randon selection of white noise from $N(0,1)$, and

$$p_i(z) = 1 + 2r\cos\theta_i z^{-1} + r^2 z^{-2}$$

where $r = 0.96$, $\theta_1 = 2\pi/3$, $\theta_2 = \pi/3$, and $\theta_3 = \pi/2$. These input signals are all autoregressive (AR) random processes and have distinct power spectral centers.

In the figures to be shown, the channel matrix $\mathbf{H}(z)$ is chosen to have the dimension $J \times I = 4 \times 3$ and the degree $L_H = 1$. Higher degrees of $\mathbf{H}(z)$ were also considered in our experiment, but the number of local minima of the cost function used in Step 3 of the BIDS method increases significantly as the degree of $\mathbf{H}(z)$ increases. The global minimum of the cost function is

$$\text{MSE}_{\text{channel}}(\text{dB}) = 10\log_{10}\left(\frac{\sum_{r=1}^{R}\min_{\Lambda_r,\mathbf{P}_r}\sum_{l=0}^{L_H}\left\|\hat{\mathbf{H}}_r(l)\Lambda_r\mathbf{P}_r - \mathbf{H}(l)\right\|_F^2}{R\sum_{l=0}^{L_H}\|\mathbf{H}(l)\|_F^2}\right)$$

$$\text{MSE}_{\text{signal}}(\text{dB}) = 10\log_{10}\left(\frac{\sum_{r=1}^{R}\min_{\Lambda_r,\mathbf{P}_r}\sum_{n=0}^{N-1}\|\mathbf{P}_r\Lambda_r\hat{\mathbf{x}}(n) - \mathbf{x}(n)\|_F^2}{R\sum_{n=0}^{N-1}\|\mathbf{x}(n)\|_F^2}\right)$$

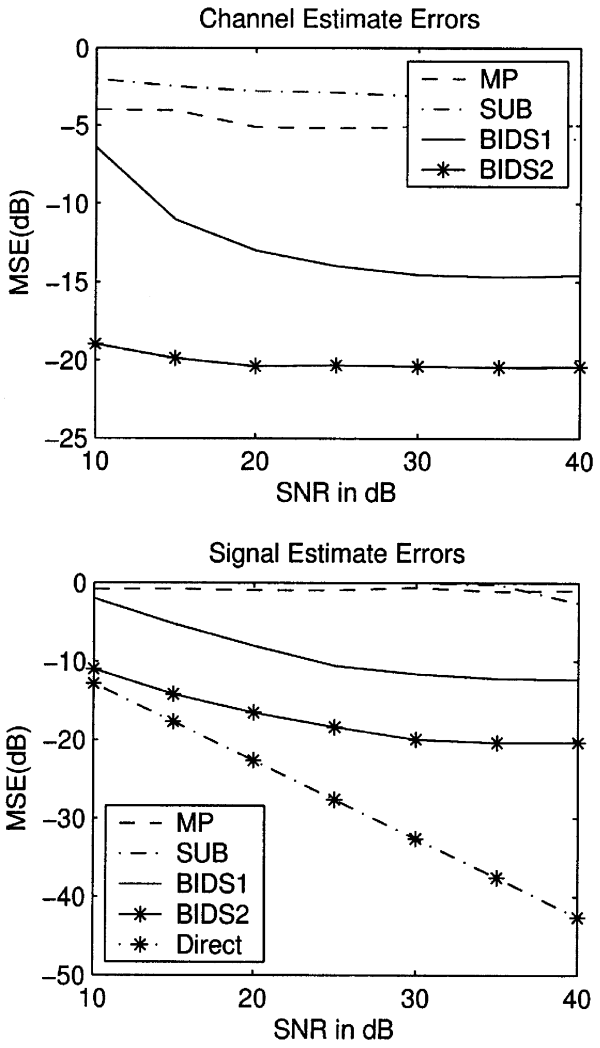


Fig. 4. MSE versus SNR of the BIDS-1 and BIDS-2 algorithms and two other algorithms for a fixed irreducible (but not column-reduced) channel. $N = 10\,000$. (Most MSEs of the SUB method are above 0 dB and, hence ,not visible here.)

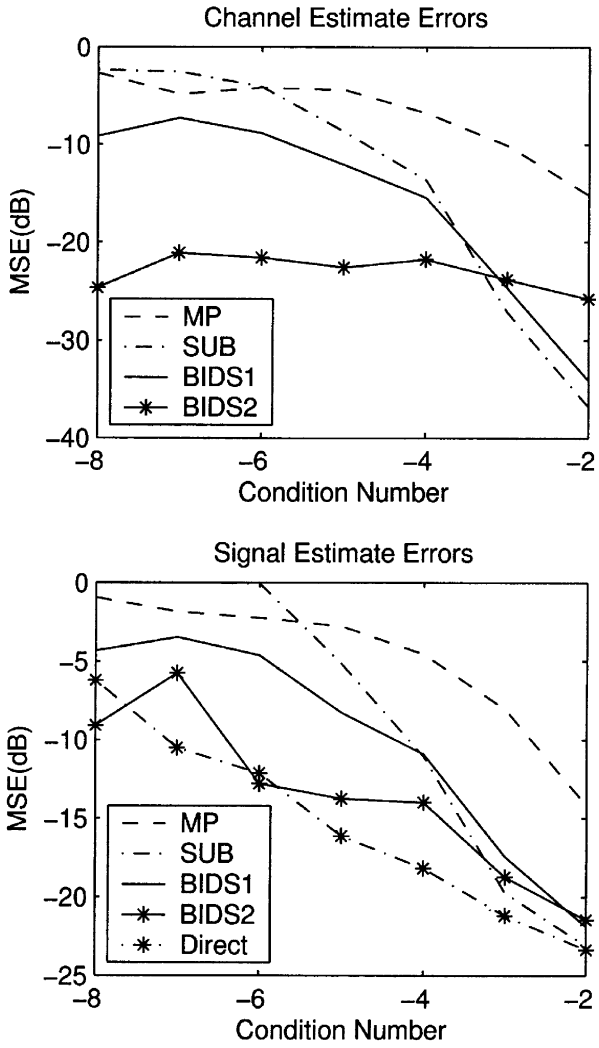


Fig. 5. MSE versus channel condition number of the BIDS-1 and BIDS-2 algorithms and two other algorithms. SNR $= 20$ dB. $N = 10\,000$. The condition number distribution of 600 randomly selected channels is shown in Table I.

difficult to find for higher degrees of $\mathbf{H}(z)$ unless a good initial estimate of $\mathbf{H}(z)$ is available.

Fig. 2 compares the MSEs versus SNR of the four algorithms: BIDS-1, BIDS-2, MP (matrix pencil) and SUB (subspace), where all parameters of $\mathbf{H}(z)$ were randomly chosen from $N(0,1)$ (i.e., Gaussian distribution of zero mean and unit deviation) at each run. The channel estimation error is shown in the upper part of the figure, and the signal estimation error is shown in the lower part of the figure. The "direct" algorithm shown for signal estimation uses the exact channel parameters. The number of samples is $N = 10\,000$.

Fig. 3 compares the MSEs versus the sample size of the four algorithms. The channel parameters of $\mathbf{H}(z)$ are the same as in Fig. 2. The SNR is 20 dB. The upper part is for the channel estimation error, and the lower part is for the signal estimation error.

From Figs. 2 and 3, we see that the BIDS-2 algorithm is the most robust in the low SNR region and that the SUB algorithm yields the highest accuracy in the higher SNR region (above 30 dB). This is because that a randomly selected channel with a fixed degree is almost surely irreducible and column-reduced, and hence, at high SNR, the SUB method enjoys an ideal data model, but at lower SNR, a weakly irreducible channel matrix does not stand strong against noise. The BIDS algorithms require a weaker condition on the channel matrix and, hence, perform better at lower SNR. The "direct" algorithm has a constant MSE of the estimated input signals versus the sample size. This is because the number of input samples increases in (one-to-one) proportion with the number of output samples.

Fig. 4 illustrates the MSEs versus SNR of the four algorithms with $N = 10\,000$ and a fixed channel matrix of degree 1. This channel matrix is irreducible but not column reduced. In this case, both the MP and SUB methods fail miserably, as expected. Note that most part of the MSEs (in decibels) of the SUB method is higher than 0 dB and, hence, is outside the (vertical) range of the figure.

Fig. 5 shows the MSEs of the four methods as a function of the channel condition number. This number is defined to

TABLE I
DISTRIBUTION OF THE CONDITION NUMBERS OF 600 RANDOMLY SELECTED CHANNEL MATRICES, WHERE C DENOTES A RANGE OF CONDITION NUMBERS, AND D DENOTES THE NUMBER OF CHANNELS WITHIN A GIVEN RANGE OF CONDITION NUMBERS. EACH MSE VALUE SHOWN IN FIG. 5 WAS AN AVERAGE OF MSES CORRESPONDING TO THE ABOVE RANGES OF CONDITION NUMBERS

C	(-9, -8)	(-8, -7)	(-7, -6)	(-6, -5)	(-5, -4)	(-4, -3)	(-3, -2)
D	7	9	21	48	125	232	158

be the natural logarithm of the smallest singular value of the Sylvester matrix of the channel matrix.[8] The Sylvester matrix $\mathbf{H}_{sw}$ is defined as in (21) with $W = 3$. The sample size is $N = 10\,000$. The SNR is 20 dB. It is clear from this figure that the performance of all methods depends significantly on the condition number (and, in fact, for a wide range of SNRs and sample sizes[9]). We see that the SUB method performs poorly if the condition number is poor (small). On the other hand, the BIDS-2 algorithm always performs the best among all the methods when the condition number is poor (small). Table I shows the distribution of the condition numbers of 600 randomly selected channel matrices of degree 1.

Finally, we remark that the BIDS method depends critically on its Step 3—the source separation step. Provided that this step is successful, all other steps are computationally efficient and statistically robust. Step 3, however, involves a minimization of a nonquadratic function that potentially has many local minima. Developing a more effective approach of Step 3 is an important future direction.

V. CONCLUSIONS

In this paper, we have shown a sufficient and "necessary" condition of blind identification of FIR MIMO channels. This condition sets a theoretical limit of blind identification, although thus far, there exists no working algorithm under this condition. We have also developed the BIDS method that performs far better than the previously developed methods: the subspace method and the matrix pencil method (in a lower SNR region and/or with a poor channel matrix condition). A key feature of the BIDS method is that it exploits a bank of source separators/decorrelators. As long as the source separators are well constructed, the BIDS method is robust. The BIDS method provides a new framework of blind identification of MIMO channels where (subchannel) source separation plays a partial and yet critical role. Among different possible implementations of the BIDS method, the BIDS-1 and BIDS-2 algorithms are presented in detail. The BIDS-2 algorithm has the most promising performance. Through the development of the BIDS-1 and BIDS-2, a great deal of insight into the FIR MIMO channels has been revealed. We hope that these insights will be useful for future development in the field of blind system identification.

[8]A more conventional definition of condition number is the ratio of the largest singular value over the smallest singular value, which is different from ours.

[9]For other values of SNR and N, the patterns of all figures look similar and, hence, are omitted.

APPENDIX

In this Appendix, we prove Theorem 2. We begin with Lemma A1.

Lemma A1

Let $\mathbf{Q} = [\mathbf{q}_1, \ \mathbf{q}_2, \ \cdots, \ \mathbf{q}_m]$ be an m-column constant matrix, and let $\mathbf{Q}_i$ be $\mathbf{Q}$ with the ith column deleted. If $\mathbf{Q}_1$ has a full column rank and $\mathbf{Q}_i$ for any $i \geq 2$ does not have a full column rank, then $\mathbf{q}_1 = \mathbf{0}$.

Proof: Since $\mathbf{Q}_1$ has a full column rank, $\mathbf{q}_i$ for $i \geq 2$ are linearly independent vectors. Since $\mathbf{Q}_2$ does not have a full column rank, $\mathbf{q}_1$ (unless $\mathbf{q}_1 = \mathbf{0}$) must then be a linear combination of $\mathbf{q}_i$ for $i \geq 3$, i.e.,

$$\mathbf{q}_1 = \sum_{i=2}^{m} k_{1,i}\mathbf{q}_i$$

where $k_{1,2} = 0$. Since $\mathbf{Q}_3$ does not have a full rank, $\mathbf{q}_1$ (unless $\mathbf{q}_1 = \mathbf{0}$) must then be given by the above expression, where $k_{1,2} = 0$, and $k_{1,3} = 0$. By continuing this reasoning, we see that $k_{1,i} = 0$ for all $i \geq 2$, i.e., $\mathbf{q}_1 = \mathbf{0}$.

Proof of Theorem 2: We denote $c_{i,j}(z) = (\mathbf{C}(z))_{i,j}$ and $s_i(z) = (\mathbf{S}(z))_{i,i}$. Without loss of generality, we assume that $\deg(s_1(z)) \leq \deg(s_2(z)) \leq 1 \cdots \leq \deg(s_I(z))$. Let us consider the pth column of the diagonal matrix $\mathbf{C}(z)\mathbf{S}(z)\mathbf{C}^T(z^{-1})$. The $I-1$ zero entries of the pth column imply

$$\mathbf{C}_p(z)\mathbf{D}_p(z)\mathbf{s}(z) = \mathbf{0} \tag{A.1}$$

where

$$\begin{aligned}
\mathbf{C}_p(z) &= \mathbf{C}(z) \text{ with the } p\text{th row deleted} \\
\mathbf{D}_p(z) &= \text{diag}\left(c_{p,1}\left(z^{-1}\right), \ c_{p,2}\left(z^{-1}\right), \ \cdots, \ c_{p,I}\left(z^{-1}\right)\right) \\
\mathbf{s}(z) &= [s_1(z), \ s_2(z), \ \cdots, \ s_I(z)]^T.
\end{aligned}$$

Note that the diagonal elements of $\mathbf{D}_p(z)$ correspond to the pth row of $\mathbf{C}(z^{-1})$. We will show later that $\mathbf{Q}_{(p)}(z) \hat{=} \mathbf{C}_p(z)\mathbf{D}_p(z) \equiv \mathbf{0}$. At the moment, we assume that it is true.

For $p = 1$, this implies that

$$\begin{aligned}
\mathbf{Q}_{(1)}(z) \hat{=} \mathbf{C}_1(z)\mathbf{D}_1(z) &= \begin{bmatrix} c_{2,1}(z) & c_{2,2}(z) & \cdots & c_{2,I}(z) \\ c_{3,1}(z) & c_{3,2}(z) & \cdots & c_{3,I}(z) \\ \vdots & \vdots & \ddots & \vdots \\ c_{I,1}(z) & c_{I,2}(z) & \cdots & c_{I,I}(z) \end{bmatrix} \\
&\quad \cdot \begin{bmatrix} c_{1,1}(z^{-1}) & & & \\ & c_{1,2}(z^{-1}) & & \\ & & \ddots & \\ & & & c_{1,I}(z^{-1}) \end{bmatrix} \equiv \mathbf{0}.
\end{aligned}$$

Observing each column of the above equation, we see that there are two possibilities for each column of $\mathbf{C}(z)$: 1) All elements except the first one of that column are zero, or 2) the first element of that column is zero. However, since $\mathbf{C}(z)$ is nonsingular, one of the elements in the first row of $\mathbf{C}(z)$ must be nonzero. Corresponding to that nonzero element, all other elements in that corresponding column must be zero.

For $p = 2$, we can similarly observe that another column of $\mathbf{C}(z)$ must be of all zero elements except for one. This nonzero element must be in a different row from that for the $p = 1$ case since $\mathbf{C}(z)$ is nonsingular.

By considering $\mathbf{Q}_{(p)}(z) \hat{=} \mathbf{C}_p(z)\mathbf{D}_p(z) \equiv \mathbf{0}$ for all p, we see that $\mathbf{C}(z)$ must be diagonal up to a row permutation.

We now show that $\mathbf{Q}_{(p)}(z) \hat{=} \mathbf{C}_p(z)\mathbf{D}_p(z) \equiv \mathbf{0}$. Let $\Omega_{1,k} = \{\, z: \quad s_1(z) \neq 0; \quad s_k(z) = 0 \,\}$. For all z in $\Omega_{1,k}$, where $k > 1$, we know from (A.1) that

$$\mathbf{Q}_{(p),k}(z)\mathbf{s}_k(z) = \mathbf{0} \tag{A.2}$$

where

$$\begin{aligned} \mathbf{Q}_{(p),k}(z) &= \mathbf{Q}_{(p)}(z) \text{ with the } k\text{th column deleted} \\ \mathbf{s}_k(z) &= \mathbf{s}(z) \text{ with the } k\text{th element deleted.} \end{aligned}$$

Since $\mathbf{s}_k(z) \neq \mathbf{0}$ for any z in $\Omega_{1,k}$, (A.2) implies that $\mathbf{Q}_{(p),k}(z)$ does not have full column rank for any z in $\Omega_{1,k}$, or equivalently, $\det(\mathbf{Q}_{(p),k}(z)) = 0$ for all z in $\Omega_{1,k}$. Note that by the definition of diversity, the number of (distinct) entries in $\Omega_{1,k}$ is no less than 2 div$(\mathbf{S}(z))$. Hence, given that $\deg(\det(\mathbf{Q}_{(p),k}(z))) < 2\,\text{div}(\mathbf{S}(z))$, which is equivalent to $\text{div}(\mathbf{S}(z)) > (I-1)\deg(\mathbf{C}(z))$, we have $\det(\mathbf{Q}_{(p),k}(z)) \equiv 0$, which implies that the columns of $\mathbf{Q}_{(p),k}(z)$ are linearly dependent of each other for all z and all $k > 1$.

As for $\mathbf{Q}_{(p),1}(z)$, there are two possibilities: 1) It has full column rank for almost all z, or 2) it does not have full column rank for any z. For the first case, Lemma A1 implies that the first column of $\mathbf{Q}_{(p)}(z) \hat{=} \mathbf{C}_p(z)\mathbf{D}_p(z)$ is identical to zero. For the second case, every $I-1$ columns of $\mathbf{Q}_{(pt)}(z)$ are linearly dependent of each other for all z, and hence, $\text{rank}(\mathbf{Q}_{(p)}(z)) = \text{rank}(\mathbf{C}_p(z)\mathbf{D}_p(z)) < I-1$ for all z, but we know that $\mathbf{C}_p(z)$ have the rank $I-1$ for almost all z. Then, $\mathbf{D}_p(z)$ must be singular for (almost) all z. This implies that one of the diagonal elements of $\mathbf{D}_p(z)$ is zero, and hence, one of the columns of $\mathbf{Q}_{(p)}(z) \hat{=} \mathbf{C}_p(z)\mathbf{D}_p(z)$ is zero.

By deleting the zero column from $\mathbf{Q}_{(p)}(z) \hat{=} \mathbf{C}_p(z)\mathbf{D}_p(z)$, we have a corresponding matrix $\mathbf{Q}_{(p)}(z)' \hat{=} \mathbf{C}_p(z)'\mathbf{D}_p(z)'$, which has the dimension $(I-1) \times (I-1)$, and then, (A.1) becomes

$$\mathbf{C}_p(z)'\mathbf{D}_p(z)'\mathbf{s}(z)' = \mathbf{0}. \tag{A.3}$$

Without loss of generality, we can assume that the degrees of the elements in $\mathbf{s}(z)'$ are arranged in ascending order. We can then consider (A.3) for all z in $\Omega'_{1,k}$ (associated with the elements of $\mathbf{s}(z)'$), where $k > 1$. In a manner applied before, one can show that the $I-2$ columns of $\mathbf{Q}_{(p),k}(z)'$ for any $k > 1$ are linearly dependent for all z in $\Omega'_{1,k}$. Given that $\text{div}(\mathbf{S}(z)) > (I-1)\deg(\mathbf{C}(z))$ (which is sufficient but more than necessary at this step), the $I-2$ columns of $\mathbf{Q}_{(p),k}(z)'$ for $k > 1$ are linearly dependent for all z. [That is, the determinant of every $(I-2)\times(I-2)$ submatrix of $\mathbf{Q}_{(p),k}(z)'$ for $k > 1$ is identical to zero under $\text{div}(\mathbf{S}(z)) > (I-1)\deg(\mathbf{C}(z))$.]

Then, if $\mathbf{Q}_{(p),1}(z)'$ has full column rank for almost all z, we know by Lemma A1 that the first column of $\mathbf{Q}_{(p)}(z)' \hat{=} \mathbf{C}_p(z)'\mathbf{D}_p(z)'$ is zero. If $\mathbf{Q}_{(p),1}(z)'$ does not have a full column rank for any z, then we know that $\text{rank}(\mathbf{Q}_{(p)}(z)') = \text{rank}(\mathbf{C}_p(z)'\mathbf{D}_p(z)') < I-2$ for all z. Since $\mathbf{C}_p(z)'$ has a rank no less than $I-2$ for almost all z, $\mathbf{D}_p(z)'$ must be singular for all z. Therefore, one of the diagonal elements of $\mathbf{D}_p(z)'$ must be zero, and hence, one of the columns of $\mathbf{Q}_{(p)}(z)' \hat{=} \mathbf{C}_p(z)'\mathbf{D}_p(z)'$ is zero.

By continuing the above reasoning of (B.1), we can see that all columns of $\mathbf{Q}_{(p)}(z) \hat{=} \mathbf{C}_p(z)\mathbf{D}_p(z)$ must be zero, and hence, $\mathbf{C}(z)$ must be diagonal up to a row permutation.

REFERENCES

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

[2] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-square approach to blind channel identification," *IEEE Trans. Signal Processing*, vol. 43, pp. 2982–2993, Dec. 1995.

[3] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Processing*, vol. 44, pp. 661–672, Mar. 1996.

[4] T. L. Tong, R.-W. Liu, V. C. Soon, and Y.-F Huang, "Indeterminacy and identifiability of blind identification," *IEEE Trans. Circuits Syst.*, vol. 38, pp. 499–509, May 1991.

[5] 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. 434–443, Feb. 1997.

[6] Y. Xiang, K. Abed-Meraim, and Y. Hua, "Adaptive blind source separation by second order statistics and natural gradient," in *IEEE ICASSP*, Phoenix, AZ, Mar. 1999.

[7] J. Tugnait and B. Huang, "On a whitening approach to partial channel estimation and blind equalization of FIR/IIR multiple-input multiple-output channels," *IEEE Trans. Signal Processing*, vol. 48, pp. 832–845, Mar. 2000.

[8] 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.

[9] T. Ma, Z. Ding, and S. F. Yau, "A two-stage algorithm for MIMO blind deconvolution of colored input signals," *IEEE Trans. Signal Processing*, vol. 48, pp. 1187–1192, Apr. 2000.

[10] Y. Hua, "Blind identification and equalization of channels driven by colored signals," in *Signal Processing Advances in Wireless Communications*, G. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. Englewood Cliffs, NJ: Prentice-Hall, 2000, ch. 4.

[11] Y. Hua, Y. Xiang, and K. Abed-Meraim, "Blind identification of colored signals distorted by FIR channels," in *Proc. IEEE ICASSP*, Istanbul, Turkey, June 2000.

[12] Y. Hua and Y. Xiang, "The BID for blind equalization of FIR MIMO channels," in *Proc. IEEE Workshop Neural Network Signal Process.*, Sydney, Australia, Nov. 2000.

[13] U. A. Lindgren and H. Broman, "Source separation using a criterion based on second-order statistics," *IEEE Trans. Signal Processing*, vol. 46, pp. 1837–1850, July 1998.

[14] K. I. Diamanraras, A. P. Petropulu, and B. Chen, "Blind two-input-two-output FIR channel identification based on second-order statistics," *IEEE Trans. Signal Processing*, vol. 48, pp. 534–542, Feb. 2000.

[15] Y. Hua and J. Tugnait, "Blind identifiability of FIR-MIMO systems with colored input using second order statistics," *IEEE Signal Processing Letters*, vol. 7, pp. 348–350, Dec. 2000.

[16] W. Qiu, Y. Hua, and K. Abed-Meraim, "A subspace method for the computation of the GCD of polynomials," *Automatica*, vol. 33, no. 4, pp. 741–743, Apr. 1997.

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

[18] K. Abed-Meraim, W. Qiu, and Y. Hua, "Blind system identification," *Proc. IEEE*, vol. 85, pp. 1310–1322, Aug. 1997.

[19] R.-W Liu and L. Tong, "Special issue of blind identification," *Proc. the IEEE*, vol. 86, Oct. 1998.
[20] R. R. Bitmead, S.-Y. Kung, B. D. O. Anderson, and T. Kailath, "Greatest common division via generalized Sylvester and Bezout matrices," *IEEE Trans. Automat. Contr.*, vol. 23, pp. 1043–1047, Dec. 1978.
[21] B. Porat, *Digital Processing of Random Signals—Theory and Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1994, pp. 91–91.
[22] L. Tong, "Blind asynchronous near-far resistant FIR receivers for CDMA communications," in *Proc. 29th Asilomar Conf. Signals, Syst., Comput.*, Monterey, CA, Nov. 1996, pp. 167–171.
[23] Y. Hua and S. An, "Blind identifiability conditions of FIR MIMO channels driven by colored signals," in *Proc. 34th Annu. Asilomar Conf. Signals, Syst., Comput.*, Monterey, CA, Nov. 1, 2000.
[24] P. Stoica and T. Soderstrom, "Common factor detection and estimation," *Automatica*, vol. 33, pp. 985–989, 1997.
[25] P. Loubaton and E. Moulines, "On blind multiuser forward link channel estimation by the subspace method: Identifiability results," *IEEE Trans. Signal Processing*, vol. 48, pp. 2366–2376, Aug. 2000.
[26] P. Loubaton, E. Moulines, and P. Regalia, "Subspace method for blind identification and deconvolution," in *Signal Processing Advances for Wireless and Mobile Communications*, G. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. New York: Wiley, 2001, vol. 1, ch. 3, pp. 63–112.
[27] J. Xavier, V. Barroso, and J. Moura, "Closed-form correlative coding (CFC2) blind identification of MIMO channels: Isometry fitting to second order statistics," *IEEE Trans. Signal Processing*, vol. 49, pp. 1073–1086, May 2001.
[28] I. Bradaric, A. Petropulu, and K. Diamantaras, "On blind identifiability of FIR-MIMO systems with cyclostationary inputs using second order statistics," in *Proc. IEEE ICASSP*, vol. 2, May 2002, pp. 1745–1748.

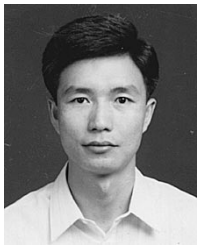
Yingbo Hua (S'86–M'88–SM'92–F'02) was born in China in 1960. He received the B.E. degree from Nanjing Institute of Technology (currently Southeast University), Nanjing, China, in 1982 and the M.S. and Ph.D. degrees from Syracuse University, Syracuse, NY, in 1983 and 1988, respectively.

He has been a Professor of electrical engineering with the University of California, Riverside, since Feb 2001. He was with the University of Melbourne, Parkville, Australia, from 1990 to 2000 and was promoted to Senior Lecturer in 1992 and Reader in 1995. He was a Project Leader of Statistical Signal Processing with the Australian Cooperative Research Center for Sensor Signal and Information Processing from 1993 to 2000. He was Visiting Faculty with the Hong Kong University of Science and Technology from 1999 to 2000. He is the author/coauthor of 80 journal articles, five book chapters, and 140 conference papers in the fundamental areas of estimation, detection, system identification, and fast algorithms, with applications in communications, remote sensing, and medical data analysis. He is a coeditor of *Signal Processing Advances in Wireless and Mobile Communications* (Englewood Cliffs, NJ: Prentice-Hall, 2001).

Dr. Hua won a Chinese Government Scholarship for Overseas Graduate Study from 1983 to 1984 and Syracuse University Graduate Fellowship from 1985 to 1986. He has served as Associate Editor of the IEEE TRANSACTIONS ON SIGNAL PROCESSING from 1994 to 1997 and again from 2001 to 2002, and was Associate Editor of the IEEE SIGNAL PROCESSING LETTERS from 1998 to 2002. He was an elected member of the IEEE Signal Processing Society's Technical Committees for Underwater Acoustic Signal Processing from 1997 to 1998, Sensor Array and Multichannel Signal Processing from 1998 to 2001, and again from 2002 to 2004, and Signal Processing for Communications from 2002 to 2004.

Senjian An was born in Shandong, China, in 1966. He received the B.S. degree from Shandong University in 1989, the M.S. degree from the Chinese Academy of Sciences, Beijing, China, in 1992, and the Ph.D. degree from Peking University, Beijing, in 1996.

From 1996 to 1998, he was a Postdoctoral Research Fellow at Institute of Systems Science, Chinese Academy of Sciences. In August 1998, he joined the Beijing Institute of Technology as an Associate Professor. Since April 2001, he has been a Research Fellow with the Department of Electrical Engineering, The University of Melbourne, Parkville, Australia. His research interests include blind source separation, blind system identification, and robust control theory.



Yong Xiang received the B.Eng. and M.Eng. degrees from the University of Electronic Science and Technology of China, Chengdu, in 1983 and 1989, respectively, and the Ph.D. degree from the University of Melbourne, Parkville, Australia, in 2003.

He was with the Southwest Institute of Electronic Equipment of China, Chengdu, from 1983 to 1986. In 1989, he joined the University of Electronic Science and Technology of China, where he was a Lecturer from 1989 to 1991 and an Associate Professor from 1992 to 1997. He was a Senior Communications Engineer with Bandspeed Pty Ltd., Melbourne, Australia, from 2000 to 2002. He is currently a faculty member with Deakin University, Victoria, Australia.