

Group Decorrelation Enhanced Subspace Method for Identifying FIR MIMO Channels Driven by Unknown Uncorrelated Colored Sources

Senjian An, Yingbo Hua, *Fellow, IEEE*, Jonathan H. Manton, *Senior Member, IEEE*, and Zheng Fang

Abstract—Identification of finite-impulse-response (FIR) and multiple-input multiple-output (MIMO) channels driven by unknown uncorrelated colored sources is a challenging problem. In this paper, a group decorrelation enhanced subspace (GDES) method is presented. The GDES method uses the idea of subspace decomposition and signal decorrelation more effectively than the joint diagonalization enhanced subspace (JDES) method previously reported in the literature. The GDES method has a much better performance than the JDES method. The correctness of the GDES method is proved assuming that 1) the channel matrix is irreducible and column reduced and 2) the source spectral matrix has distinct diagonal functions. However, the GDES method has an inherent ability to trade off between the required condition on the channel matrix and that on the source spectral matrix. Simulations show that the GDES method yields good results even when the channel matrix is not irreducible, which is not possible at all for the JDES method.

Index Terms—Adaptive signal processing, blind deconvolution, blind identification, machine learning, MIMO channels, sensor array processing, source separation, system identification.

I. INTRODUCTION

BLIND identification of multiple-input-multiple-output (MIMO) and finite-impulse-response (FIR) channels driven by unknown uncorrelated colored sources is a challenging and yet fundamental signal processing problem arising from many applications. For blind identification of MIMO FIR channels, a deterministic approach is not applicable unless there is a significant amount of algebraic (known) constraints on the channel matrix and/or the sources. A statistical approach is often necessary. Due to relatively short windows of stationarity in practical data, the second-order statistics (SOS) tend to be more reliable than the higher-order statistics (HOS). Hence,

Manuscript received February 25, 2004; revised February 8, 2005. This work was supported in part by Australian Research Council, the U.S. National Science Foundation under Grant no. ECS-0219377, and the U. S. Army Research Laboratory under the CTA Program Cooperative Agreement DAAD19-01-2-0011. Part of this work was published in the *Proceedings of the IEEE International Conference for Acoustics, Speech and Signal Processing (ICASSP)*, April 6–10, 2003, Hong Kong. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Christ D. Richmond.

S. An is with Department of Computing, Curtin University of Technology, WA 6102, Australia (e-mail: senjian@cs.curtin.edu.au).

Y. Hua and Z. Fang are with Department of Electrical Engineering, University of California, Riverside, CA 92521 USA (e-mail: yhua@ee.ucr.edu, zfang@ee.ucr.edu).

J. H. Manton was with Department of Electrical and Electronic Engineering, University of Melbourne, VIC 3010, Australia. He is now with the Department of Information Engineering, Research School of Information Sciences and Engineering (RSISE), The Australian National University, Canberra ACT 0200, Australia (e-mail: j.manton@ieee.org).

Digital Object Identifier 10.1109/TSP.2005.859339

whenever applicable, the SOS methods are generally preferred to the HOS methods.

The existing methods for blind identification of MIMO FIR channels include the subspace method [5], [10], the minimum noise subspace method [1], the matrix pencil method [11], and the blind identification by decorrelating subchannels (BIDS) method [6]. If the (polynomial) channel matrix is irreducible, column reduced, and of equal column degrees, the subspace method can identify the channel matrix up to a unknown constant matrix. To determine the unknown constant matrix, a conventional approach is to use the joint diagonalization method [15], [2]. The joint diagonalization method is a signal decorrelation method that assumes and exploits that the source signals have a zero mutual correlation and are temporally colored and of diverse temporal colors. We will refer to this conventional combination of the subspace method and the joint diagonalization as joint diagonalization enhanced subspace method (JDES). The minimum noise subspace method is a computationally simplified version of the subspace method. The matrix pencil method requires a stronger condition than the JDES method. The BIDS method assumes a weaker condition on the channel matrix but a stronger condition on the source spectral matrix than the JDES method.

In this paper, we develop a new method called the group decorrelation enhanced subspace (GDES) method. Like the JDES method, the GDES method exploits the channel matrix structure via subspace decomposition (or matching) and the source spectral matrix structure via spectral decorrelation. However, the GDES method differs from the JDES method in that the GDES method exploits the subspace associated with each column of the channel matrix. Furthermore, the GDES method iteratively exploits subspace decomposition and spectral decorrelation, which provides a more effective joint exploitation of the channel matrix structure and the source spectral matrix structure. Our approach differs from the frequency-domain approach as in [13], where a special property of nonstationarity is required.

Our notational convention is as follows. We use the bold face for polynomial matrices (and vectors), and the normal face for numerical matrices (and vectors). \mathbb{R} and $\mathbb{R}[z]$ are used to denote the set of real numbers and the set of rational functions of z , respectively. A generalized Sylvester matrix associated with a $p \times m$ polynomial matrix $\mathbf{H}(z) = \sum_{k=0}^q H_k z^{-k}$ is defined as

$$\mathcal{T}_l \{\mathbf{H}(z)\} \triangleq \begin{bmatrix} H_0 & \cdots & H_q & & \\ & \ddots & & \ddots & \\ & & H_0 & \cdots & H_q \end{bmatrix} \in \mathbb{R}^{(l+1)p \times (q+l+1)m}$$

where l is an integer. A dual form of $\mathcal{T}_l\{\mathbf{H}(z)\}$ is defined as

$$\mathcal{C}_l\{\mathbf{H}(z)\} \triangleq \begin{bmatrix} H_0 & & & \\ \vdots & \ddots & & \\ H_q & & H_0 & \\ & \ddots & \vdots & \\ & & & H_q \end{bmatrix} \in \mathbb{R}^{(q+l+1)p \times (l+1)m}.$$

Note that when $l = 0$, we have $\mathcal{T}_0\{\mathbf{H}(z)\} = [H_0 \ \cdots \ H_q]$ and $\mathcal{C}_0\{\mathbf{H}(z)\} = [H_0^T \ \cdots \ H_q^T]^T$. Here, $\mathcal{T}_0\{\mathbf{H}(z)\}$ reveals the coefficient matrices of $\mathbf{H}(z)$ horizontally, and $\mathcal{C}_0\{\mathbf{H}(z)\}$ reveals the coefficient matrices of $\mathbf{H}(z)$ vertically. Conversely, given a constant matrix $\hat{H} = [H_0 \ \cdots \ H_q]$, we use $\mathcal{T}_0^{-1}(\hat{H})$ to denote the corresponding polynomial matrix $\mathbf{H}(z)$ with $\mathcal{T}_0\{\mathbf{H}(z)\} = \hat{H}$. Similarly, given a constant matrix $\bar{H} = [H_0^T \ \cdots \ H_q^T]^T$, we use $\mathcal{C}_0^{-1}(\bar{H})$ to denote the corresponding polynomial matrix $\mathbf{H}(z)$ with $\mathcal{C}_0\{\mathbf{H}(z)\} = \bar{H}$.

The generalized Sylvester matrix and its dual form can be used to transform a polynomial operation into a numerical operation, and vice versa. Specifically, if $\mathbf{C}(z) = \mathbf{A}(z)\mathbf{B}(z)$ where $\mathbf{A}(z) = \sum_{i=0}^{q_1} A_i z^{-i}$ and $\mathbf{B}(z) = \sum_{i=0}^{q_2} B_i z^{-i}$, then we have

$$\begin{aligned} \mathcal{T}_l\{\mathbf{C}(z)\} &= \mathcal{T}_l\{\mathbf{A}(z)\}\mathcal{T}_{l+q_1}\{\mathbf{B}(z)\} \\ \mathcal{C}_l\{\mathbf{C}(z)\} &= \mathcal{C}_{l+q_2}\{\mathbf{A}(z)\}\mathcal{C}_l\{\mathbf{B}(z)\}. \end{aligned}$$

These two equations will be applied frequently in this paper. Especially, if $\mathbf{C}(z) = 0$ and $l = 0$, then the rows of $\mathcal{T}_0\{\mathbf{A}(z)\}$ belong to the left null space of $\mathcal{T}_{q_1}\{\mathbf{B}(z)\}$, and the columns of $\mathcal{C}_0\{\mathbf{B}(z)\}$ belong to the right null space of $\mathcal{C}_{q_2}\{\mathbf{A}(z)\}$.

The remainder of the paper is organized as follows. Section II presents the data model. In Section III, the conventional subspace method is revisited and the JDES method is formulated. In Section IV, we develop the GDES method by first presenting the basic idea, then establishing the theoretical foundation, and finally providing a detailed algorithmic development. In Section V, we provide simulation examples to illustrate the performance of the GDES method with a comparison to the JDES method. All proofs are deferred to the appendices.

II. THE DATA MODEL

We assume that there are m unknown sources and p sensors. We also assume that there are more sensors than sources, i.e., $p > m$. The unknown sources at time n are represented by the $m \times 1$ vector $x(n)$, and the output of the sensors at time n is by the $p \times 1$ vector $y(n)$. The relationship between the sources and the sensor output is modeled by

$$y(n) = \sum_{k=0}^q H_k x(n-k) + w(n) \quad (1)$$

where $\{H_k\}_{k=0}^q$ is the channel's finite impulse response, and $w(n)$ the noise. It is clear that each matrix H_k has the dimension $p \times m$. An equivalent form of (1) is $y(n) = \mathbf{H}(z)x(n) + w(n)$, where the polynomial matrix $\mathbf{H}(z) = \sum_{k=0}^q H_k z^{-k}$ is called

the channel matrix. The channel matrix here can be viewed as a linear and time-invariant operator, where $z^{-1}[x(n)] = x(n-1)$.

The SOS of $y(n)$ is captured by the autocorrelation matrix

$$R_{yy}(\tau) \triangleq \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{n=0}^{N-1} y(n+\tau)y^T(n). \quad (2)$$

The power spectral matrix of $y(n)$ is defined as $\mathbf{S}_{yy}(z) = \sum_{\tau=-\infty}^{\infty} R_{yy}(\tau)z^{-\tau}$. The power spectral matrix of the input (the sources) $x(n)$ is similarly defined. Assuming that the noise is uncorrelated with the sources, we have from (1) that

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

Methods based on SOS must exploit the above equation (either explicitly or implicitly). Assuming that the noise is white both spatially and temporally, then $\mathbf{S}_{ww}(z) = \delta^2 I_p$, where δ^2 is the noise variance. As discussed later, the noise variance can be obtained asymptotically. For simple presentation of our theory, we will drop the noise term without loss of generality in an asymptotical sense. The effect of noise will be compensated in the algorithm development. Now, the data model in our problem is as follows:

$$y(n) = \sum_{k=0}^q H_k x(n-k) = \mathbf{H}(z)x(n) \quad (4)$$

$$\mathbf{S}_{yy}(z) = \mathbf{H}(z)\mathbf{S}_{xx}(z)\mathbf{H}^T(z^{-1}). \quad (5)$$

Our aim is to estimate the channel matrix $\mathbf{H}(z)$ using the autocorrelation matrices $R_{yy}(\tau)$. Once the channel matrix is available, the sources can be estimated in a relatively straightforward way.

III. THE JDES METHOD

We now review the spirit of the conventional subspace method as studied in [5] and [10] and then formulate the JDES method. Let $Y_l(n)$ be defined as

$$Y_l(n) = [y^T(n) \ y^T(n-1) \ \cdots \ y^T(n-l)]^T$$

and $X_l(n)$ is defined similarly. Then, from (4), we have

$$Y_l(n) = \mathcal{T}_l\{\mathbf{H}(z)\}X_{l+q}(n)$$

and therefore

$$\mathbb{E}[Y_l(n)Y_l^T(n)] = \mathcal{T}_l\{\mathbf{H}(z)\}\mathbb{E}[X_{l+q}(n)X_{l+q}^T(n)]\mathcal{T}_l^T\{\mathbf{H}(z)\}^T \quad (6)$$

where \mathbb{E} denotes the time average over n .

Assume the following:

- A1) $\mathbf{H}(z)$ is irreducible and column reduced, i.e., $\mathbf{H}(z)$ has a full column rank for every $z \in \mathbb{C}$ except $z = 0$, and its highest order coefficient matrix $[h_{1q_1}, h_{2q_2}, \dots, h_{mq_m}]$ is of full column rank, where q_i is the i th column degree and h_{iq_i} is the highest order coefficient vector of the i th column;

- A2) $l \geq \sum_{i=1}^m q_i$, where q_i is the i th column degree of the channel matrix;
- A3) the input signals are mutually uncorrelated and of distinct power spectra, i.e., the power spectral matrix $\mathbf{S}_{xx}(z)$ is diagonal, and the ratio of every two diagonal elements of $\mathbf{S}_{xx}(z)$ is not a constant;
- A4) $R_{xx}^{l+q}(0)$ is positive definite, i.e., the input signals are persistently exciting and of the order $(l + q + 1)$ ([8, p. 413]).

Under A4), the left null space of $\mathcal{T}_l\{\mathbf{H}(z)\}$ is the same as that of $\mathbb{E}[Y_l(n)Y_l^T(n)]$, and therefore an orthogonal basis matrix, denoted by U_0 , of this space can be computed from the eigenvalue decomposition of $\mathbb{E}[Y_l(n)Y_l^T(n)]$.

If $\mathbf{H}(z)$ is of an identical column degree q , under A1) and A2), U_0 determines $\mathbf{H}(z)$ up to a right nonsingular constant matrix Q [10]. Let $\mathbf{G}_0(z) = \mathcal{T}_0^{-1}(U_0^T)$. Then, $\mathbf{H}(z)$ can be estimated, up to a right nonsingular constant matrix Q , from the following relationships:

$$\begin{aligned} U_0^T \mathcal{T}_l\{\mathbf{H}(z)\} = 0 &\Leftrightarrow \mathbf{G}_0(z)\mathbf{H}(z) = 0 \\ &\Leftrightarrow \mathcal{C}_q\{\mathbf{G}_0(z)\}\mathcal{C}_0\{\mathbf{H}(z)\} = 0. \end{aligned} \quad (7)$$

If $\mathbf{H}(z)$ is not of an identical column degree, U_0 determines $\mathbf{H}(z)$ up to an upper triangular polynomial matrix $\mathbf{Q}(z)$. This ambiguity polynomial matrix $\mathbf{Q}(z)$ can be further reduced to a constant matrix by a technique developed in [5].

To determine the unknown constant matrix Q , we proceed as follows. Denote by $\mathbf{H}_e(z)$ the estimate of the channel matrix $\mathbf{H}(z)$, then $\mathbf{H}(z) = \mathbf{H}_e(z)Q$. Since $\mathbf{H}(z)$ is irreducible, so is $\mathbf{H}_e(z)$, and hence one can compute a filter $\mathbf{E}_H(z) = \sum_{k=0}^L E_H(k)z^{-k}$ of finite degree L such that

$$\mathbf{E}_H(z)\mathbf{H}_e(z) = I_m. \quad (8)$$

As shown in [3] and [5], under A1), such $\mathbf{E}_H(z)$ exists for any $L \geq \sum_{i=1}^m q_i - 1$. One way to obtain the coefficient matrix $\mathcal{T}_0\{\mathbf{E}_H(z)\}$ is to observe that

$$\mathcal{T}_0\{\mathbf{E}_H(z)\}[\mathcal{T}_L\{\mathbf{H}_e(z)\}] = [I_m, 0, \dots, 0]$$

and therefore

$$\mathcal{T}_0\{\mathbf{E}_H(z)\} = [I_m, 0, \dots, 0][\mathcal{T}_L\{\mathbf{H}_e(z)\}]^\# \quad (9)$$

where (#) stands for the pseudoinverse. Applying $\mathbf{E}_H(z)$ to (4) yields

$$\begin{aligned} v(n) &\triangleq \mathbf{E}_H(z)y(n) \Rightarrow v(n) = Qx(n) \\ \mathbf{S}_{vv}(z) &\triangleq \mathbf{E}_H(z)\mathbf{S}_{yy}(z)\mathbf{E}_H^T(z^{-1}) \Rightarrow \mathbf{S}_{vv}(z) = Q\mathbf{S}_{xx}(z)Q^T. \end{aligned} \quad (10)$$

Following the joint diagonalization method [2], Q can be identified up to a columnwise scaling and permutation if A3) holds.

One major weakness of the JDES method is that it heavily relies on the initial estimate of the channel matrix $\mathbf{H}(z)$ to determine the ambiguity matrix Q . A poor initial estimate of the channel matrix makes a poor channel equalization (i.e., from

$y(n)$ to $v(n)$), and hence the resulting estimate of the matrix Q is not reliable. In fact, the subspace method fails miserably if the channel matrix is only weakly irreducible.

IV. THE GDES METHOD

A. The Idea

The principal idea of our method is to estimate each column $\mathbf{h}(z)$ of the channel matrix $\mathbf{H}(z)$ from the left null space of $\mathcal{T}_l\{\mathbf{h}(z)\}$. We wish to know $\mathbf{G}_1(z)$ of degree l such that $\mathbf{G}_1(z)\mathbf{h}(z) = 0$. In numerical form, this equation means that $\mathcal{T}_0\{\mathbf{G}_1(z)\}\mathcal{T}_l\{\mathbf{h}(z)\} = 0$. Assuming that $\mathbf{h}(z)$ is irreducible and of the column degree q , $\mathcal{T}_l\{\mathbf{h}(z)\}$ has the full column rank $l + q + 1$ provided $l \geq q$. If $\mathcal{T}_0\{\mathbf{G}_1(z)\}$ has $n_1 = p(l + 1) - (l + q + 1)$ independent rows, then the column span of $\mathcal{T}_0\{\mathbf{G}_1(z)\}^T$ is the orthogonal complement of the column span of $\mathcal{T}_l\{\mathbf{h}(z)\}$. Note that $\mathcal{T}_0\{\mathbf{G}_1(z)\}\mathcal{T}_l\{\mathbf{h}(z)\} = 0$ is equivalent to $\mathcal{C}_q\{\mathbf{G}_1(z)\}\mathcal{C}_0\{\mathbf{h}(z)\} = 0$. Therefore, $\mathcal{C}_0\{\mathbf{h}(z)\}$ is uniquely determined up to a constant scalar by $\mathcal{C}_q\{\mathbf{G}_1(z)\}$, or equivalently, $\mathbf{h}(z)$ is uniquely determined up to a scalar by $\mathbf{G}_1(z)$.

Removing the column $\mathbf{h}(z)$ from $\mathbf{H}(z)$, we have $\mathbf{H}_c(z)$ as the resulting submatrix. If $\mathbf{H}_c(z)$ is irreducible and column reduced and $l \geq \sum_{i=1}^{m-1} q_i$, where q_i is the i th column degree of $\mathbf{H}_c(z)$, then it is known that the rank of $\mathcal{T}_l\{\mathbf{H}_c(z)\}$ is $\sum_{i=1}^{m-1} (l + q_i + 1)$. Therefore, there is a $\mathbf{G}_2(z)$ of degree l such that $\mathbf{G}_2(z)\mathbf{H}_c(z) = 0$ and $\mathcal{T}_0\{\mathbf{G}_2(z)\}$ has $n_2 = p(l + 1) - \sum_{i=1}^{m-1} (l + q_i + 1)$ independent rows.

Given the required property of $\mathbf{G}_1(z)$ and $\mathbf{G}_2(z)$, we know that $\mathbf{G}_1(z)y(n)$ and $\mathbf{G}_2(z)y(n)$ are two separated groups of sources. The two separated groups must also be uncorrelated since the sources are assumed to be mutually uncorrelated. The uncorrelation implies that $\mathbf{G}_1(z)\mathbf{S}_{yy}(z)\mathbf{G}_2^T(z^{-1}) = 0$. This equation can be achieved or approximately achieved when the data $y(n)$ is available. However, a fundamental question now is, Does this equation imply the required property of $\mathbf{G}_1(z)$ and $\mathbf{G}_2(z)$? This question is answered next.

B. Theoretical Foundation

Theorem 1: Assume A1)–A3). Let k be an integer, $1 \leq k \leq m$. Define two matrix filters $\mathbf{G}_j(z) = \sum_{i=0}^l G_{ji}z^{-i} \in \mathbb{R}[z]^{n_j \times p}$, $j = 1, 2$, where $\text{rank}[\mathcal{T}_0\{\mathbf{G}_1(z)\}] = n_1 = p(l + 1) - (l + q_k + 1)$ and $\text{rank}[\mathcal{T}_0\{\mathbf{G}_2(z)\}] = n_2 = p(l + 1) - \sum_{i=1, i \neq k}^m (l + q_i + 1)$. Then, there are such $\mathbf{G}_1(z)$ and $\mathbf{G}_2(z)$ to make the following hold:

$$\mathbf{G}_1(z)\mathbf{S}_{yy}(z)\mathbf{G}_2^T(z^{-1}) = 0. \quad (11)$$

Furthermore, if (11) holds, then

$$\mathbf{G}_1(z)\mathbf{h}(z) = 0 \quad (12)$$

$$\mathbf{G}_2(z)\mathbf{H}_c(z) = 0 \quad (13)$$

where $\mathbf{h}(z)$ is a column of the channel matrix $\mathbf{H}(z)$ and with degree q_k , and $\mathbf{H}_c(z)$ is the remaining submatrix of $\mathbf{H}(z)$ without $\mathbf{h}(z)$.

Proof: See Appendix A.

C. Algorithm Development

1) *Main Structure of $\mathbf{G}_1(z)$ and $\mathbf{G}_2(z)$* : A key problem now is to find $\mathbf{G}_1(z)$ and $\mathbf{G}_2(z)$ such that the group decorrelation property (11) holds. Denote the degree of $\mathbf{S}_{yy}(z)$ by l_y , which can be infinity in theory if any of the diagonal elements is a rational function. However, we will only need to choose a value of l_y corresponding to the dominant coefficients in $\mathbf{S}_{yy}(z)$. Define

$$\Phi(z) \triangleq \mathbf{G}_1(z)\mathbf{S}_{yy}(z)\mathbf{G}_2^T(z^{-1}) = \sum_{\tau=-l_y-l}^{l_y+l} \Phi(\tau)z^{-\tau} \quad (14)$$

where (easy to verify)

$$\Phi(\tau) = \mathcal{T}_0\{\mathbf{G}_1(z)\}P(\tau)\mathcal{T}_0\{\mathbf{G}_2(z)\}^T$$

$$P(\tau) \triangleq \begin{bmatrix} R_{yy}(\tau) & \cdots & R_{yy}(\tau+l) \\ \vdots & \ddots & \vdots \\ R_{yy}(\tau-l) & \cdots & R_{yy}(\tau) \end{bmatrix} \in \mathbb{R}^{p(l+1) \times p(l+1)} \quad (15)$$

$$R_{yy}(\tau) = \mathbb{E}[y(n)y(n-\tau)^T]. \quad (16)$$

Here, $R_{yy}(-\tau) = R_{yy}(\tau)^T$ and $\mathcal{T}_0\{\mathbf{G}_i(z)\} \in \mathbb{R}^{n_i \times p(l+1)}$.

To solve (11), the following cost function is a natural choice:

$$\begin{aligned} \mathcal{J}_G\{\mathbf{G}_1(z), \mathbf{G}_2(z)\} &= \|\mathcal{T}_0\{\Phi(z)\}\|_F^2 \\ &= \sum_{\tau=-l_y-l}^{l_y+l} \|\mathcal{T}_0\{\mathbf{G}_1(z)\}P(\tau)\mathcal{T}_0\{\mathbf{G}_2(z)\}^T\|_F^2. \end{aligned} \quad (17)$$

To consider the effect of noise, let us observe that $P(0) = \mathcal{T}_l\{\mathbf{H}(z)\}R_{xx}^{l+q}(0)\mathcal{T}_l\{\mathbf{H}(z)\}^T$. By A1)–A3), in the absence of noise, $\text{rank}[P(0)] = \text{rank}[\mathcal{T}_l\{\mathbf{H}(z)\}] = r = \sum_{i=1}^m (l+q_i+1)$. With white noise, $P(0)$ contains an additional component $\delta^2 I$, and hence the noise variance δ^2 equals the $p(l+1) - r$ smallest eigenvalues of $P(0)$. With a finite set of data, the distribution of the $p(l+1) - r$ smallest eigenvalues of $P(0)$ tends to spread, and δ^2 can be simply chosen to be the average of these smallest eigenvalues of $P(0)$. When available, δ^2 can and should be removed from $R_{yy}(0)$, and hence from all $P(\tau)$ used in (17).

Because of the effects of finite sample size, the cost function (17) should be further revised. Let the eigenvalue decomposition of $P(0)$ be given by

$$\begin{bmatrix} U_1 & U_2 \end{bmatrix} \begin{bmatrix} S_1 & \\ & S_2 \end{bmatrix} \begin{bmatrix} U_1 & U_2 \end{bmatrix}^T$$

where $U_1 \in \mathbb{R}^{p(l+1) \times r}$, $U_2 \in \mathbb{R}^{p(l+1) \times [p(l+1)-r]}$, S_1 contains the r largest eigenvalues, and S_2 contains the remaining eigenvalues, which are zero in theory.

Since the columns of U_2 are “almost” in the null space of $P(\tau)$, we choose the rows of U_2^T as part of the rows in each of $\mathcal{T}_0\{\mathbf{G}_1(z)\}$ and $\mathcal{T}_0\{\mathbf{G}_2(z)\}$. To find the remaining rows of the two matrices, we first define the following weighting matrix:

$$W = (S_1 - \delta^2 I)^{-1/2} U_1^T \quad (18)$$

and then construct the following cost function with respect to X_1 and X_2 :

$$\mathcal{J}_X(X_1, X_2) = \sum_{\tau=-l_y-l}^{l_y+l} \|X_1^T W P(\tau) W^T X_2\|_F^2 \quad (19)$$

where $X_i \in \mathbb{R}^{r \times N_i}$, $X_i^T X_i = I_{N_i}$, $N_i = n_i - p(l+1) + r$, and $i = 1, 2$. Note that the weighting matrix used here tends to redistribute evenly the signal power in $P(\tau)$. The final pair of the group decorrelators $\mathbf{G}_1(z)$ and $\mathbf{G}_2(z)$ is constructed as

$$\mathcal{T}_0\{\mathbf{G}_i(z)\} = \begin{bmatrix} U_2^T \\ X_i^T W \end{bmatrix}, i = 1, 2. \quad (20)$$

2) *Alternating Projection*: It is clear that with a fixed X_1 , the cost (19) is a quadratic function of X_2 . The same is true if we reverse the order of X_1 and X_2 . We can minimize the cost with respect to X_2 with a fixed X_1 and then minimize the cost with respect to X_1 with a fixed X_2 . This process can be repeated until convergence. We refer to this procedure as alternating projection (AP). The minimization with respect to each of X_1 and X_2 has an established solution ([14, pp. 262–263]) as given next. Define

$$Q_1 = \sum_{\tau=-l_y-l}^{l_y+l} W P(\tau) W^T X_2 X_2^T W P(\tau)^T W^T \quad (21)$$

$$Q_2 = \sum_{\tau=-l_y-l}^{l_y+l} W P(\tau)^T W^T X_1 X_1^T W P(\tau) W^T \quad (22)$$

and then $\mathcal{J}_X(X_1, X_2) = \text{tr}(X_1^T Q_1 X_1) = \text{tr}(X_2^T Q_2 X_2)$. Let the eigenvalue decompositions of Q_i be

$$Q_i = V_i \Sigma_i V_i^T, i = 1, 2$$

where $V_i^T V_i = I_r$, and $\Sigma_i = \text{diag}(\sigma_{i,1}, \sigma_{i,2}, \dots, \sigma_{i,r})$ with $\sigma_{i,1} \geq \sigma_{i,2} \geq \dots \geq \sigma_{i,r} \geq 0$. With a fixed X_1 , the optimal X_2 is simply given by the last N_2 columns of V_2 , and with a fixed X_2 , the optimal X_1 is simply given by the last N_1 columns of V_1 . At each step, the minimum of the cost is $\varepsilon_i = \sum_{k=r-N_i+1}^r \sigma_{i,k}$.

The cost function (19) is nonlinear and nonquadratic of the joint unknowns $\mathbf{G}_1(z)$ and $\mathbf{G}_2(z)$. The algorithm is likely to be stuck at a local minimum of (19). A proper initialization of the AP procedure is needed. Next, we provide an initialization algorithm to overcome the local minima problem.

3) *Initialization Algorithm*: Let

$$\mathcal{J}_0(b) = \sum_{\tau=-l_y-l, \tau \neq 0}^{l_y+l} (b^T W P(\tau) W^T b)^2 \quad (23)$$

where $b \in \mathbb{R}^r$ and $b^T b = 1$. Assume the following:

A3') the power spectral matrix $\mathbf{S}_{xx}(z)$ is diagonal, and every two diagonal elements $s_i(z)$ and $s_j(z)$ of $\mathbf{S}_{xx}(z)$ are such that one of the two elements has a root of an odd repeated number, and this root is not a root of the other element (note that with probability one, there are no repeated roots in each element, and in this case the repeated number is defined to be one).

Then, the following result holds.

Theorem 2: Assume A1)–A3'). Let b_* be a local maximizer of $\mathcal{J}_0(b)$ subject to $b^T b = 1$. Then

$$b_*^T W \mathcal{T}_l \{\mathbf{h}(z)\} \neq 0, b_*^T W \mathcal{T}_l \{\mathbf{H}_c(z)\} = 0 \quad (24)$$

where $\mathbf{h}(z)$ is a column of the channel matrix $\mathbf{H}(z)$, and $\mathbf{H}_c(z)$ is the remaining submatrix of $\mathbf{H}(z)$ without $\mathbf{h}(z)$. Furthermore, using $b_*^T W \mathcal{T}_l \{\mathbf{H}_c(z)\} = 0$ and $U_2^T \mathcal{T}_l \{\mathbf{H}_c(z)\} = 0$, $\mathbf{H}_c(z)$ can be identified up to a right invertible polynomial matrix.

Proof: See Appendix B.

After a local maximizer of $\mathcal{J}_0(b)$ is found, using $b_*^T W \mathcal{T}_l \{\mathbf{H}_c(z)\} = 0$ and $U_2^T \mathcal{T}_l \{\mathbf{H}_c(z)\} = 0$, we can apply the subspace method (see Section III or [5]) to identify $\mathbf{H}_c(z)$ up to a right invertible constant matrix (or a polynomial matrix if $\mathbf{H}_c(z)$ is of nonidentical column degrees) and then compute an orthogonal basis, say N , of the left null space of $W \mathcal{T}_l \{\hat{\mathbf{H}}_c(z)\}$, where $\hat{\mathbf{H}}_c(z)$ is the estimate of $\mathbf{H}_c(z)$. Note that $W \mathcal{T}_l \{\hat{\mathbf{H}}_c(z)\}$ and $W \mathcal{T}_l \{\mathbf{H}_c(z)\}$ share the same left null space. We can then use N as an initial value of X_2 to minimize $\mathcal{J}_X(X_1, X_2)$ by the AP algorithm.

Next, we provide an algorithm to find a local maximizer of (23) under the constraint $b^T b = 1$. This constraint is a special form of a unitary matrix constraint. Hence, we can apply the modified steepest descent (MSD) method on Grassmann manifold as in [12] to find a local maximizer. This method requires that the derivative of the cost function $\mathcal{J}_0(b)$ is available. This derivative can be calculated directly from (23). That is

$$D_b \doteq \frac{d\mathcal{J}_0(b)}{db} = 4 \sum_{\tau=-l_y-l, \tau \neq 0}^{l_y+l} (b^T W P(\tau) W^T b) W P(\tau) W^T b. \quad (25)$$

The MSD algorithm [12] is summarized as follows.

- 1) Choose b such that $b^T b = 1$. Set the step size $\gamma := 1$.
- 2) Compute D_b , which is the derivative of $\mathcal{J}_0(b)$ at b .
- 3) Compute the descent direction $Z := (I - bb^T)D_b$.
- 4) Evaluate $\|Z\| = \sqrt{Z^T Z}$. If it is sufficiently small, then stop.
- 5) If $\mathcal{J}_0((b + 2\gamma Z)/\|b + 2\gamma Z\|) - \mathcal{J}_0(b) \geq \gamma \|Z\|^2$, then set $\gamma := 2\gamma$ and repeat Step 5).
- 6) If $\mathcal{J}_0((b + \gamma Z)/\|b + \gamma Z\|) - \mathcal{J}_0(b) < (1/2)\gamma \|Z\|^2$, then set $\gamma := (1/2)\gamma$ and repeat Step 6).
- 7) Set $b = (b + \gamma Z)/\|b + \gamma Z\|$. Go to Step 2).

In general, Steps 5) and 6) of MSD may cause a slow rate of convergence. Fortunately, however, for the cost function given here, an optimal step size γ can be calculated to speed up the convergence. This is discussed in detail next.

First, we normalize Z such that $Z^T Z = 1$ and define $f(\gamma) = \mathcal{J}_0((b + \gamma Z)/\|b + \gamma Z\|)$. Note that $b^T b = 1$, $Z^T Z = 1$, $b^T Z = 0$, by direct calculation, we have

$$f(\gamma) = \sum_{\tau=-l_y-l, \tau \neq 0}^{l_y+l} \left(\frac{\alpha_{\tau,0} + \alpha_{\tau,1}\gamma + \alpha_{\tau,2}\gamma^2}{1 + \gamma^2} \right)^2 \quad (26)$$

where $\alpha_{\tau,0} = b^T W P(\tau) W^T b$, $\alpha_{\tau,1} = Z^T W P(\tau) W^T b + b^T W P(\tau) W^T Z$, and $\alpha_{\tau,2} = Z^T W P(\tau) W^T Z$.

The derivative of $f(\gamma)$ is then

$$f'(\gamma) = \frac{2}{(1 + \gamma^2)^3} \sum_{\tau=-l_y-l, \tau \neq 0}^{l_y+l} \alpha(\gamma)\beta(\gamma)$$

$$\alpha(\gamma) = a_{\tau,0} + a_{\tau,1}\gamma + a_{\tau,2}\gamma^2$$

$$\beta(\gamma) = a_{\tau,1} + 2(a_{\tau,2} - a_{\tau,0})\gamma - a_{\tau,1}\gamma^2. \quad (27)$$

A necessary condition for γ to achieve the maximum is that $f'(\gamma) = 0$. Since $f'(\gamma)$ is a polynomial of degree 4, its real roots can be computed by available formula. Alternatively, one can compute all its roots, including the complex ones, say r_1, r_2, r_3, r_4 , and let γ_i be the real part of r_i . Then by comparing $f(\gamma_i)$, one can find the optimal step size γ that maximizes $f(\gamma)$. Using this optimal step size to replace Steps 5) and 6), the above algorithm will converge faster.

4) *Group Decorrelation and Channel Estimation:* Following the previously discussed initialization and AP, a group decorrelator is now found. By Theorem 1 and the subspace method, one column of the channel matrix is now identified.

Suppose that k columns of the channel matrix have been found, we now show how to estimate another column of the channel matrix. We denote the k estimated columns by $\mathbf{h}_i(z), i = 1, \dots, k$ and denote $\mathbf{H}_k(z) = [\mathbf{h}_1(z), \dots, \mathbf{h}_k(z)]$. Removing $\mathbf{H}_k(z)$ from $\mathbf{H}(z)$ results in a submatrix $\bar{\mathbf{H}}_k(z)$. Correspondingly, let $x_k(n)$ and $\bar{x}_k(n)$ denote the source vectors associated with $\mathbf{H}_k(z)$ and $\bar{\mathbf{H}}_k(z)$, respectively. Then, $y(n) = \mathbf{H}_k(z)x_k(n) + \bar{\mathbf{H}}_k(z)\bar{x}_k(n)$. Denote $y_1(n) = \mathbf{H}_k(z)x_k(n)$ and $y_2(n) = \bar{\mathbf{H}}_k(z)\bar{x}_k(n)$. Then, $R_{yy}(\tau) = R_{y_1 y_1}(\tau) + R_{y_2 y_2}(\tau)$ and therefore $P(\tau) = P_1(\tau) + P_2(\tau)$, where $P_i(\tau)$ is a matrix with the same structure as $P(\tau)$ in (15), with $R_{yy}(\tau)$ being replaced by $R_{y_i y_i}(\tau)$.

We can compute an orthogonal basis, say N_k , of the left null space of $W \mathcal{T}_l \{\mathbf{H}_k(z)\}$, where $N_k \in \mathbb{R}^{r \times r_k}$ and $r_k = r - \text{Rank}[\mathcal{T}_l \{\mathbf{H}_k(z)\}]$. Let $W_k = N_k^T W$ and

$$\mathcal{J}_k(b) = \sum_{\tau=-l_y-l, \tau \neq 0}^{l_y+l} (b^T W_k P(\tau) W_k^T b)^2 \quad (28)$$

where $b \in \mathbb{R}^{r_k}$ and $b^T b = 1$.

Note that we can write $P_1(\tau) = \mathcal{T}_l \{\mathbf{H}_k(z)\} R_1(\tau) [\mathcal{T}_l \{\mathbf{H}_k(z)\}]^T$, where $R_1(\tau)$ is a matrix of proper dimensions, and hence $W_k P_1(\tau) W_k^T = 0$. Therefore

$$\mathcal{J}_k(b) = \sum_{\tau=-l_y-l, \tau \neq 0}^{l_y+l} (b^T W_k P_2(\tau) W_k^T b)^2.$$

Clearly, $\mathcal{J}_k(b)$ is similar as $\mathcal{J}_0(b)$, where $\mathbf{H}(z)$ and $x(n)$ are replaced by $\bar{\mathbf{H}}_k(z)$ and $\bar{x}_k(n)$, respectively. Thus, one can apply MSD to find a local maximizer (say, b_*) of $\mathcal{J}_k(b)$ subject to $b^T b = 1$. Then, by Theorem 2 and that $b_*^T W_k \mathcal{T}_l \{\mathbf{h}_i(z)\} = 0$ for $i = 1, 2, \dots, k$, we have

$$b_*^T W_k \mathcal{T}_l \{\mathbf{h}(z)\} \neq 0, b_*^T W_k \mathcal{T}_l \{\mathbf{H}_c(z)\} = 0 \quad (29)$$

where $\mathbf{h}(z)$ is a column of $\mathbf{H}(z)$ but not of $\mathbf{H}_k(z)$, and $\mathbf{H}_c(z)$ is the remaining submatrix of $\mathbf{H}(z)$ without $\mathbf{h}(z)$. Hence, following the similar initialization and AP procedures, a new group

decorrelator is found, and a new column of the channel matrix can be identified. Sequentially, all the columns of the channel matrix will be found.

In summary, the proposed GDES method is as follows.

- 1) Initialize $k = 0$ and $W_0 = W$.
- 2) Apply MSD to find a local maximizer b_* of $\mathcal{J}_k(b)$.
- 3) Use the subspace method to estimate $\mathbf{H}_c(z)$ via $b_*^T W_k \mathcal{T}_l \{\mathbf{H}_c(z)\} = 0$ and $U_2^T \mathcal{T}_l \{\mathbf{H}_c(z)\} = 0$.
- 4) Compute a basis matrix of the left null space of $W \mathcal{T}_l \{\mathbf{H}_c(z)\}$ based on the estimate of $\mathbf{H}_c(z)$ and set it as X_2 .
- 5) Perform AP to find a minimizer (X_1, X_2) of $\mathcal{J}_X(X_1, X_2)$ and evaluate $\mathbf{G}_1(z)$ using (20).
- 6) Estimate $\mathbf{h}_{k+1}(z)$ using $\mathbf{G}_1(z)\mathbf{h}_{k+1}(z) = 0$ by the subspace method.
- 7) If $k + 1 = m$, stop. Otherwise, set $k = k + 1$.
- 8) Compute a basis N_k of the left null space of $W \mathcal{T}_l \{\mathbf{H}_k(z)\}$ where $\mathbf{H}_k(z) = [\mathbf{h}_1(z), \dots, \mathbf{h}_k(z)]$. Set $W_k = N_k^T W$ and go to Step 2).

Remarks: We have developed the above algorithm under the assumption that the column degrees are known, and hence the rank of $P(0)$ is known. The subspace method in [5] can be used to estimate the column degrees. There are also many existing methods for estimating the “effective” rank of a matrix from its eigenvalues ([7, Ch. 1]). Also note that a less complex and more heuristic version of the GDES method is shown in Appendix C, which has a comparable performance as the above algorithm.

V. SIMULATIONS

In this section, we present simulation examples to evaluate the performance of the GDES method with comparison to the JDES method.

Recall the data model $y(n) = \mathbf{H}(z)x(n) + w(n)$, $n = 0, 1, \dots, N - 1$, where $\mathbf{H}(z) = \sum_{i=0}^q H_i z^{-i}$ with the dimension $p \times m$ and the degree q . The entries of the coefficient matrices $\{H_i\}_{i=0}^q$ are randomly selected from a normal distribution $N(0, \sigma^2)$ with mean zero and variance one. The noise sequences $w(n)$ are randomly generated from $N(0, \sigma^2)$. The sequences in the input $x(n)$ are generated as follows:

$$x_i(n) = s_i(z)[q_i(n)], \quad i = 1, 2, \dots, m$$

where each sample in the sequence $q_i(n)$ is randomly selected from $N(0, 1)$, and each $s_i(z)$ is a polynomial (of degree 6) with coefficients randomly selected from $N(0, 1)$. Note that the choice of $s_i(z)$ for $i = 1, 2, \dots, m$ determines the source spectral matrix. The signal-to-noise ratio (SNR) of the observed data $y(n)$ is defined as

$$\text{SNR} = 10 \log_{10} \frac{\frac{1}{N} \sum_{n=1}^N \|y(n)\|_2^2 - p\sigma^2}{p\sigma^2}.$$

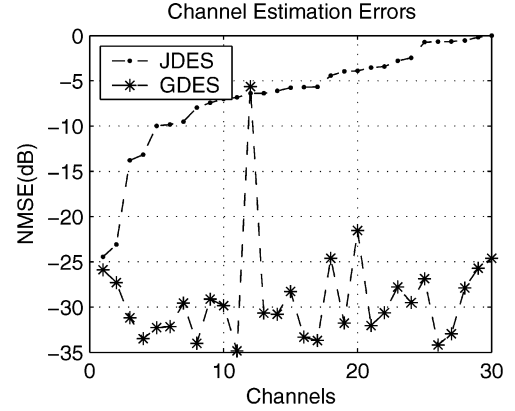


Fig. 1. Performance of the GDES method and the JDES method over 30 randomly selected channels. (The one outlier shown here of the GDES method did not happen when the algorithm in Appendix C was applied.)

The performance measure of the estimated channel matrices is chosen to be the normalized mean-squared errors (NMSEs):

$$\text{NMSE} = 10 \log_{10} \left(\frac{\sum_{k=1}^R \min_{P_k} \sum_{i=0}^q \|\hat{H}_i^{(k)} P_k - H_i\|_F^2}{R \sum_{i=0}^q \|H_i\|_F^2} \right)$$

where R is the number of the Monte Carlo runs, P_k is a permutation matrix (with entries equal to 0, 1 or -1), $\hat{H}_i^{(k)}$ is the estimated value of H_i , and each column of the coefficient matrices $C_0\{\mathbf{H}(z)\}$ and $C_0\{\hat{\mathbf{H}}^{(k)}(z)\}$ is normalized to have unit norm.

A. Example 1: Performance Over Different Channels

We use this example to show that some channels for which the irreducible-and-column-reduced condition is only weakly satisfied can still be estimated effectively by GDES but not by JDES. For a channel matrix $\mathbf{H}(z)$, the inverse condition number of $\mathcal{T}_l\{\mathbf{H}(z)\}$, defined as the ratio of its least singular value over and its largest singular value, can be used to indicate how weak the irreducible-and-column-reduced condition is.

We consider 30 randomly selected channel matrices with inverse condition number less than 0.1. Each channel matrix is of the dimension 4×3 and the degree 1. For each channel matrix, a source spectral matrix is independently generated. With probability one, the assumptions A1) and A3) should be met by the channel matrices and the source spectral matrices. We choose SNR = 20 dB, the sample size $N = 5000$. We will refer to “a given channel matrix and a given spectral matrix of the sources” as a channel.

Fig. 1 compares the performance of the GDES method and the JDES method for 30 independent channels as described previously. The figure is sorted in the increasing order of the NMSE of the JDES method. This figure shows that GDES can identify most of the randomly selected channels. On the other hand, very few of them can be identified effectively by JDES. It suggests that, if the inverse condition number of the channel matrix is less than 0.1, the channel is not identifiable by JDES but is still mostly identifiable by GDES. For randomly selected channels

with degree 1 and the dimension 4×3 , about half of them have the inverse condition number less than 0.1.

B. Example 2: Performance Versus SNR and Sample Size

Now we choose a channel matrix with degree 2, as follows:

$$\mathbf{H}(z) = \begin{bmatrix} 0.6832 & -0.7209 & 1.4418 \\ -1.1710 & -0.2594 & -1.2567 \\ 0.4911 & -0.4657 & -0.0467 \\ 0.6525 & 0.2404 & 0.1820 \end{bmatrix} + \begin{bmatrix} -0.0548 & 0.8197 & 0.1566 \\ -1.8212 & -1.0386 & 1.7413 \\ -0.1826 & 0.8229 & -1.7660 \\ 0.1237 & -0.8503 & -0.0142 \end{bmatrix} z^{-1} + \begin{bmatrix} 0.2423 & -0.0298 & -0.7317 \\ 0.7915 & 0.3410 & -1.3594 \\ 0.3609 & -0.9887 & 2.1230 \\ -0.8512 & -0.7442 & -0.3500 \end{bmatrix} z^{-2}.$$

We also choose the following polynomials that govern the source spectral matrix:

$$\begin{aligned} s_1(z) &= 0.3872 + 0.5131z^{-1} - 0.2080z^{-2} - 0.6971z^{-3} \\ &\quad - 0.1929z^{-4} - 0.1396z^{-5} - 0.0295z^{-6}; \\ s_2(z) &= -0.1335 + 0.1462z^{-1} + 0.1015z^{-2} - 0.0673z^{-3} \\ &\quad - 0.4151z^{-4} - 0.3606z^{-5} + 0.8023z^{-6}; \\ s_3(z) &= 0.4475 + 0.7156z^{-1} - 0.2586z^{-2} + 0.0041z^{-3} \\ &\quad + 0.3188z^{-4} + 0.2715z^{-5} + 0.2132z^{-6}. \end{aligned}$$

For each pair of SNR and the sample size N , 50 Monte Carlo runs are conducted. The two plots in Fig. 2 show the performances of the GDES method and the JDES method versus SNR and N , respectively. An important observation is that the JDES method performs better than the GDES method only when SNR is very high. At a very high SNR, most of the irreducible column-reduced equal-column-degree channel matrices (corresponding to full-rank generalized Sylvester matrices) make even the smallest eigenvalue of the covariance matrix $\mathcal{T}_l\{\mathbf{H}(z)\}R_{xx}^{l+q}(0)\mathcal{T}_l\{\mathbf{H}(z)\}^T$ much larger than the noise variance, and hence the subspace method alone can yield a reliable channel matrix (up to a constant matrix) without use of decorrelation. However, at a moderate or lower SNR, the smallest eigenvalue of $\mathcal{T}_l\{\mathbf{H}(z)\}R_{xx}^{l+q}(0)\mathcal{T}_l\{\mathbf{H}(z)\}^T$ becomes insignificant to the noise variance, and hence a joint exploitation of subspace and decorrelation inherent in the GDES method becomes necessary to improve the performance. In fact, as shown in the next example, even when the channel matrix is not irreducible and hence the matrix $\mathcal{T}_l\{\mathbf{H}(z)\}$ does not have a full-column rank, the GDES method still yields good results.

C. Example 3: Performance for Nonirreducible Channel Matrix

A nonirreducible channel matrix can be constructed by $\mathbf{H}(z) = \mathbf{H}_{\text{irr}}(z)\mathbf{Q}(z)$, where $\mathbf{H}_{\text{irr}}(z)$ is randomly selected with dimension 4×3 , and degree 2, and $\mathbf{Q}(z)$ is also randomly

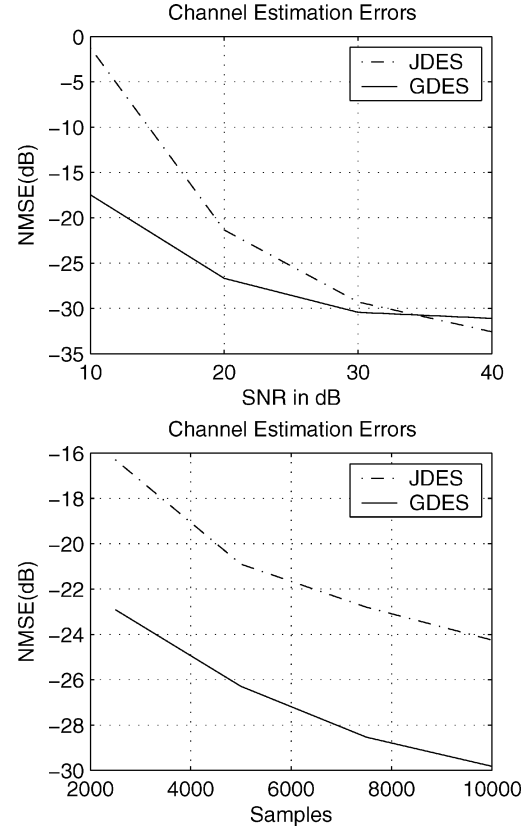


Fig. 2. Performances of the GDES method and the JDES method versus SNR and the sample size. For the upper plot, $N = 5000$. For the lower plot, SNR = 20 dB.

selected with the dimension 3×3 and the degree 1. A random selection of $\mathbf{H}_{\text{irr}}(z)$ and $\mathbf{Q}(z)$ yields

$$\mathbf{H}(z) = \begin{bmatrix} -2.6632 & 2.5426 & -1.3756 \\ -0.8858 & 0.7646 & 2.1288 \\ -1.1610 & 1.1227 & 0.3201 \\ -2.1989 & -2.7145 & -0.2484 \end{bmatrix} + \begin{bmatrix} 3.2681 & 4.5355 & 1.8888 \\ 3.9594 & -2.9482 & -0.5839 \\ 1.4927 & 1.0550 & 0.8164 \\ 1.1168 & -0.1365 & 1.1735 \end{bmatrix} z^{-1} + \begin{bmatrix} -2.0554 & -4.6226 & -1.8164 \\ -2.6092 & 0.0701 & 1.6969 \\ 0.9021 & -2.1156 & 0.3722 \\ -1.3207 & -1.7483 & -2.6016 \end{bmatrix} z^{-2} + \begin{bmatrix} 0.3813 & 0.9575 & 1.6625 \\ 2.7311 & -1.1591 & 0.4015 \\ 0.4254 & -1.6611 & 0.1720 \\ -3.3242 & 1.2801 & 0.3513 \end{bmatrix} z^{-3}.$$

Another random selection of the source spectral matrix is determined by

$$\begin{aligned} s_1(z) &= 0.2581 + 0.7615z^{-1} - 0.3135z^{-2} + 0.1692z^{-3} \\ &\quad - 0.2827z^{-4} - 0.2789z^{-5} + 0.2624z^{-6}; \\ s_2(z) &= -0.5188 + 0.0548z^{-1} - 0.2202z^{-2} - 0.5570z^{-3} \\ &\quad + 0.1063z^{-4} - 0.0748z^{-5} - 0.5935z^{-6}; \\ s_3(z) &= 0.6408 + 0.1114z^{-1} + 0.3050z^{-2} - 0.5452z^{-3} \\ &\quad + 0.2012z^{-4} + 0.2871z^{-5} + 0.2524z^{-6}. \end{aligned}$$

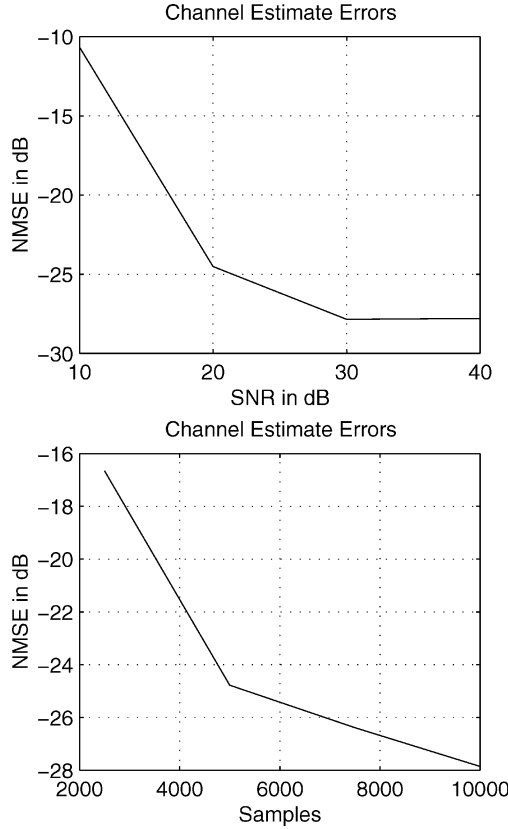


Fig. 3. Performance of the GDES method for a nonirreducible channel matrix. For the upper plot, $N = 5000$. For the lower plot, SNR = 20 dB.

Fig. 3 shows the performance of the GDES method for the previously defined channel, which is based on $R = 50$ runs. At a moderate SNR and a moderate N , the performance of the GDES method is reasonably good. The JDES method does not apply here at all due to the nonirreducible nature of the channel matrix.

Since we do not have a valid initialization algorithm for nonirreducible channels, the initialization of the GDES method has to be done differently. For this example, the i th column $\mathbf{h}_i(z)$ of the channel $\mathbf{H}(z)$ was initialized as $\hat{\mathbf{h}}_i(z) = \mathbf{h}_i(z) + 0.1\mathbf{h}(z)$, where $\mathbf{h}(z)$ was selected at random. The coefficient matrix of $\hat{\mathbf{h}}_i(z)$ has the same norm as that of $\mathbf{h}(z)$. For a nonirreducible channel matrix, the GDES method (or any other existing method) is incomplete without a proper initialization. However, this example provides an important evidence to support that the GDES method is more robust than the JDES method in general.

VI. CONCLUSION

In this paper, we have studied the problem of blind identification of finite-impulse-response (FIR) multiple-input multiple-output (MIMO) channels driven by uncorrelated colored sources. The group decorrelation enhanced subspace (GDES) method that we have developed in this paper has the best performance for this problem among all methods known to date. Although having roots in the subspace method [10] and the BIDS [6] method, the GDES method represents a new and major step toward a complete understanding of this challenging problem.

APPENDIX A PROOF OF THEOREM 1

First, we briefly introduce the rational vector space theory [4], which is fundamental for our proofs. Let $\text{span}\{\mathbf{F}(z)\}$ denote the m -dimensional rational vector space spanned by the column vectors of a $p \times m$ polynomial matrix $\mathbf{F}(z)$ with a normal full-column rank. A polynomial matrix $\mathbf{B}(z) = [\mathbf{b}_1(z), \mathbf{b}_2(z), \dots, \mathbf{b}_m(z)]$ is said to be a *polynomial basis* of $\text{span}\{\mathbf{F}(z)\}$ if $\text{span}\{\mathbf{B}(z)\} = \text{span}\{\mathbf{F}(z)\}$. The *order* of $\mathbf{B}(z)$ is defined as $\sum_{i=1}^m \deg(\mathbf{b}_i(z))$. The matrix $\mathbf{B}(z)$ is said to be a *minimal polynomial basis* if its order is minimum over the set of all polynomial bases of $\text{span}\{\mathbf{F}(z)\}$. The ordered column degrees $L_1 \leq L_2 \leq \dots \leq L_m$ of a minimal polynomial basis are called the Kronecker indices of $\text{span}\{\mathbf{F}(z)\}$. The dual space of $\text{span}\{\mathbf{F}(z)\}$, denoted by $\text{span}\{\mathbf{F}(z)\}^\perp$, is defined as the $(p - m)$ -dimensional subspace of all $p \times 1$ rational vectors $\mathbf{g}(z)$ satisfying $\mathbf{g}^T(z)\mathbf{f}(z) = 0$ for all $\mathbf{f}(z) \in \text{span}\{\mathbf{F}(z)\}$. We denote by $L_1^\perp \leq L_2^\perp \leq \dots \leq L_{p-m}^\perp$ the Kronecker indices of $\text{span}\{\mathbf{F}(z)\}^\perp$. It is known that

$$\sum_{i=1}^m L_i = \sum_{j=1}^{p-m} L_j^\perp. \quad (30)$$

It is also known [9] that if $\mathbf{F}(z)$ has a normal full-column rank m , then

$$\text{Rank}[\mathcal{T}_l\{\mathbf{F}(z)\}] = (l+1)p - \sum_{j=1}^{p-m} \max\{l+1 - L_j^\perp, 0\} \quad (31)$$

or equivalently

$$\text{Rank}[\mathcal{T}_l\{\mathbf{F}(z)\}] = (l+1)m + \sum_{j=1}^{p-m} \min\{l+1, L_j^\perp\}. \quad (32)$$

Before proving Theorem 1, we present the following technical Lemma.

Lemma 1: Let l and n_i for $i = 1, 2$ be positive integers. Also let $n_1 > (p-2)(l+1)$ and $n_2 > (p-m-2)(l+1)$. Define two matrix filters $\mathbf{G}_i(z) = \sum_{k=0}^l G_{ik}z^{-k} \in \mathbb{R}[z]^{n_i \times p}$ with $\text{rank}[\mathcal{T}_0\{\mathbf{G}_i(z)\}] = n_i$. If (11) holds, then the output power spectra matrix $\mathbf{S}_{yy}(z)$ has the decomposition

$$\begin{aligned} \mathbf{S}_{yy}(z) &= [\mathbf{b}_1(z) \quad \mathbf{B}_2(z)] \begin{bmatrix} \lambda_1(z) & 0_{1 \times (m-1)} \\ 0_{(m-1) \times 1} & \Lambda_2(z) \end{bmatrix} \begin{bmatrix} \mathbf{b}_1(z^{-1})^T \\ \mathbf{B}_2(z^{-1})^T \end{bmatrix} \end{aligned} \quad (33)$$

where $\lambda_1(z)$ is a nonzero rational function, $\Lambda_2(z) \in \mathbb{R}[z]^{(m-1) \times (m-1)}$ is a nonsingular rational matrix, $\mathbf{b}_1(z)$ is an irreducible polynomial vector, $\mathbf{B}_2(z) \in \mathbb{R}[z]^{p \times (m-1)}$ is an irreducible and column-reduced polynomial matrix, and

$$\mathbf{G}_1(z)\mathbf{b}_1(z) = 0, \quad \mathbf{G}_2(z)\mathbf{B}_2(z) = 0. \quad (34)$$

Proof: From (11), we have

$$\mathbf{G}_1(z)\mathbf{H}(z)\mathbf{S}_{xx}(z)\mathbf{H}^T(z^{-1})\mathbf{G}_2^T(z^{-1}) = 0$$

or equivalently

$$\mathbf{G}_1(z)\mathbf{B}_1(z) = 0 \quad (35)$$

where $\mathbf{B}_1(z) \in \mathbb{R}[z]^{p \times m_1}$ is any minimal polynomial basis matrix of $\text{span}\{\mathbf{H}(z)\mathbf{S}_{xx}(z)\mathbf{H}^T(z^{-1})\mathbf{G}_2^T(z^{-1})\}$. Then, using $\mathcal{T}_0\{\mathbf{G}_1(z)\}\mathcal{T}_l\{\mathbf{B}_1(z)\} = 0$, we have

$$\text{rank}[\mathcal{T}_l\{\mathbf{B}_1(z)\}] \leq p(l+1) - \text{rank}[\mathcal{T}_0\{\mathbf{G}_1(z)\}] = p(l+1) - n_1. \quad (36)$$

By the rank formula (32), $m_1 = 1$ since $n_1 > (p-2)(l+1)$. $\mathbf{B}_1(z)$ is then a polynomial vector, and we use $\mathbf{b}_1(z)$ to denote it thereafter.

Since $\text{span}\{\mathbf{H}(z)\mathbf{S}_{xx}(z)\mathbf{H}^T(z^{-1})\mathbf{G}_2^T(z^{-1})\} \subset \text{span}\{\mathbf{H}(z)\}$, a rational vector $\mathbf{x}_1(z) \in \mathbb{R}[z]^m$ exists such that

$$\mathbf{b}_1(z) = \mathbf{H}(z)\mathbf{x}_1(z). \quad (37)$$

Since $\mathbf{H}(z)\mathbf{x}_1(z)$ is a minimal polynomial basis of $\text{span}\{\mathbf{H}(z)\mathbf{S}_{xx}(z)\mathbf{H}^T(z^{-1})\mathbf{G}_2^T(z^{-1})\}$, there exists a rational vector $\mathbf{a}_1(z) \in \mathbb{R}[z]^{n_2}$ such that $\mathbf{H}(z)\mathbf{S}_{xx}(z)\mathbf{H}^T(z^{-1})\mathbf{G}_2^T(z^{-1}) = \mathbf{H}(z)\mathbf{x}_1(z)\mathbf{a}_1^T(z)$, or equivalently

$$\mathbf{S}_{xx}(z)\mathbf{H}^T(z^{-1})\mathbf{G}_2^T(z^{-1}) = \mathbf{x}_1(z)\mathbf{a}_1^T(z).$$

Let $\mathbf{x}_1^\perp(z) \in \mathbb{R}[z]^{m \times (m-1)}$ consist of the basis vectors of the orthogonal complement of $\text{span}\{\mathbf{x}_1(z)\}$. Then, we have

$$\mathbf{G}_2(z)\mathbf{H}(z)\mathbf{S}_{xx}(z)\mathbf{x}_1^\perp(z) = 0 \quad (38)$$

or equivalently

$$\mathbf{G}_2(z)\mathbf{B}_2(z) = 0 \quad (39)$$

where $\mathbf{B}_2(z) \in \mathbb{R}[z]^{p \times (m-1)}$ is any minimal polynomial basis matrix of $\text{span}\{\mathbf{H}(z)\mathbf{S}_{xx}(z)\mathbf{x}_1^\perp(z)\}$.

Similar to the previous discussion on $\mathbf{b}_1(z)$, we have

$$\mathbf{B}_2(z) = \mathbf{H}(z)\mathbf{X}_2(z) \quad (40)$$

where $\mathbf{X}_2(z)$ is a rational matrix in $\mathbb{R}[z]^{m \times (m-1)}$ and of the rank $m-1$. There is a rational matrix $\mathbf{A}_2(z) \in \mathbb{R}[z]^{(m-1) \times (m-1)}$ such that $\mathbf{H}(z)\mathbf{S}_{xx}(z)\mathbf{x}_1^\perp(z) = \mathbf{H}(z)\mathbf{X}_2(z)\mathbf{A}_2(z)$, or equivalently, $\mathbf{S}_{xx}(z)\mathbf{x}_1^\perp(z) = \mathbf{X}_2(z)\mathbf{A}_2(z)$. Let $\mathbf{X}_2^\perp(z) \in \mathbb{R}[z]^m$ be the basis vector of the orthogonal complement of $\text{span}\{\mathbf{X}_2(z)\}$. Then, we have

$$(\mathbf{x}_1^\perp(z))^T \mathbf{S}_{xx}(z) \mathbf{X}_2^\perp(z) = 0 \quad (41)$$

and therefore

$$\begin{bmatrix} \{\mathbf{X}_2^\perp(z)\}^T \\ \{\mathbf{x}_1^\perp(z)\}^T \end{bmatrix} \begin{bmatrix} \mathbf{x}_1(z) & \mathbf{X}_2(z) \end{bmatrix} = \begin{bmatrix} \alpha_1(z) & \mathbf{0}_{1 \times (m-1)} \\ \mathbf{0}_{(m-1) \times 1} & \mathbf{R}_2(z) \end{bmatrix} \quad (42)$$

i.e.,

$$\begin{bmatrix} \mathbf{x}_1(z) & \mathbf{X}_2(z) \end{bmatrix}^{-1} = \begin{bmatrix} \alpha_1^{-1}(z) [\mathbf{X}_2^\perp(z)]^T \\ \mathbf{R}_2^{-1}(z) [\mathbf{x}_1^\perp(z)]^T \end{bmatrix} \quad (43)$$

where $\alpha_1(z)$ is a rational function and $\mathbf{R}_2(z)$ is a $(m-1) \times (m-1)$ rational matrix.

By (37) and (40), $\mathbf{H}(z) = [\mathbf{b}_1(z), \mathbf{B}_2(z)][\mathbf{x}_1(z), \mathbf{X}_2(z)]^{-1}$. Then, by letting

$$\begin{aligned} \boldsymbol{\Lambda}_1(z) &= \alpha_1^{-1}(z) [\mathbf{X}_2^\perp(z)]^T \mathbf{S}_{xx}(z) \mathbf{X}_2^\perp(z^{-1}) \alpha_1^{-1}(z^{-1}) \\ \boldsymbol{\Lambda}_2(z) &= \mathbf{R}_2^{-1}(z) [\mathbf{x}_1^\perp(z)]^T \mathbf{S}_{xx}(z) \mathbf{x}_1^\perp(z^{-1}) \mathbf{R}_2^{-T}(z^{-1}) \end{aligned}$$

we obtain (33) and then (34) from (35) and (39). This completes the proof. \square

Proof of Theorem 1: Let $\mathbf{h}(z)$ be a column of $\mathbf{H}(z)$ with degree q_k . Then, $\mathcal{T}_l\{\mathbf{h}(z)\}$ has a left null space of dimension n_1 ,

and $\mathcal{T}_l\{\mathbf{H}_c(z)\}$ has a left null space of dimension n_2 . There are $\mathcal{T}_0\{\mathbf{G}_1(z)\}$ and $\mathcal{T}_0\{\mathbf{G}_2(z)\}$ such that $\mathcal{T}_0\{\mathbf{G}_1(z)\}\mathcal{T}_l\{\mathbf{h}(z)\} = 0$ and $\mathcal{T}_0\{\mathbf{G}_2(z)\}\mathcal{T}_l\{\mathbf{H}_c(z)\} = 0$. This implies (12) and (13) and hence (11).

Now, suppose that (11) is true. Under A2), $n_1 > (p-2)(l+1)$ and $n_2 > (p-m-2)(l+1)$. From Lemma 1, $\mathbf{S}_{yy}(z)$ has a decomposition with the form (33) and

$$\mathbf{G}_1(z)\mathbf{b}_1(z) = 0, \quad \mathbf{G}_2(z)\mathbf{B}_2(z) = 0 \quad (44)$$

which implies that

$$\text{Rank}[\mathcal{T}_l\{\mathbf{b}_1(z)\}] \leq p(l+1) - n_1 = l + q_k + 1,$$

$$\text{Rank}[\mathcal{T}_l\{\mathbf{B}_2(z)\}] \leq p(l+1) - n_2 = \sum_{i=1, i \neq k}^m (l + q_i + 1). \quad (45)$$

Denote $\mathbf{B}(z) = [\mathbf{b}_1(z), \mathbf{B}_2(z)]$. Then, $\text{Rank}[\mathcal{T}_l\{\mathbf{B}(z)\}] \leq \sum_{i=1}^m (l + q_i + 1)$. By the rank formula (32) and A2), the order of $\mathbf{B}(z)$ is $\sum_{i=1}^m q_i$, the same as that of $\mathbf{H}(z)$. Note that $\text{span}\{\mathbf{B}(z)\} = \text{span}\{\mathbf{H}(z)\}$. $\mathbf{B}(z)$ is also a minimal polynomial basis of $\text{span}\{\mathbf{H}(z)\}$ as well as $\mathbf{H}(z)$. Without loss of generality, we assume that the column degrees of $\mathbf{H}(z)$ are non-decreasing and denote $\mathbf{H}(z) = [\mathbf{H}_1(z), \mathbf{H}_2(z), \dots, \mathbf{H}_v(z)]$, where $\mathbf{H}_i(z)$ is of dimension $p \times m_i$ and of identical column degree μ_i , where $\mu_1 < \mu_2 < \dots < \mu_v$ are v different column degrees of the channel matrix. Since $\mathbf{B}(z)$ and $\mathbf{H}(z)$ are minimal polynomial bases of the same rational vector space, there is a permutation matrix P to make $\mathbf{B}(z)P$ have the same structure as $\mathbf{H}(z)$, i.e., $\mathbf{B}(z) = [\bar{\mathbf{B}}_1(z), \bar{\mathbf{B}}_2(z), \dots, \bar{\mathbf{B}}_v(z)]$. Furthermore, it is known [5] that there exists an upper triangular polynomial matrix $\mathbf{Q}(z)$ such that

$$\mathbf{H}(z) = \mathbf{B}(z)P\mathbf{Q}(z) \quad (46)$$

where

$$\mathbf{Q}(z) = \begin{bmatrix} Q_{11} & \mathbf{Q}_{12}(z) & \cdots & \mathbf{Q}_{1v}(z) \\ & Q_{22} & \cdots & \mathbf{Q}_{2v}(z) \\ & & \ddots & \vdots \\ & & & Q_{vv} \end{bmatrix} \quad (47)$$

with Q_{ii} being the nonsingular constant matrices of dimension $m_i \times m_i$, and $\mathbf{Q}_{ij}(z)$ being the polynomial matrices of dimension $m_i \times m_j$ and of a degree $\leq (\mu_j - \mu_i)$. From (46) and (33), it follows that

$$P\mathbf{Q}(z)\mathbf{S}_{xx}(z)\mathbf{Q}^T(z^{-1})P^T = \begin{bmatrix} \boldsymbol{\Lambda}_1(z) & \mathbf{0}_{1 \times (m-1)} \\ \mathbf{0}_{(m-1) \times 1} & \boldsymbol{\Lambda}_2(z) \end{bmatrix}. \quad (48)$$

To coincide with the structure of $\mathbf{Q}(z)$, we denote

$$\mathbf{S}_{xx}(z) = \begin{bmatrix} \mathbf{S}_1(z) & & & \\ & \mathbf{S}_2(z) & & \\ & & \ddots & \\ & & & \mathbf{S}_v(z) \end{bmatrix} \quad (49)$$

where $\mathbf{S}_i(z)$ is of dimension $m_i \times m_i$ and is obviously diagonal.

Suppose $q_k = \mu_j$. Without losing generality, we can assume that $\mathbf{b}_1(z)$ is the first column of $\bar{\mathbf{B}}_j(z)$. Let $L = \sum_{i=1}^{j-1} m_i + 1$. Then, the L th entry of the first row of P is 1, and other entries of this row are zero.

Let $\mathbf{a}^T(z) = [0_{1 \times (L-1)}, a_j^T, \mathbf{a}_{j+1}^T(z), \dots, \mathbf{a}_v^T(z)]$ denote the L th row of $\mathbf{Q}(z)$, where $a_j \in \mathbb{R}^{m_j}$ and $\mathbf{a}_i(z) \in \mathbb{R}[z]^{m_i}$, $i = j+1, \dots, v$. By (48), we have $\mathbf{a}_v^T(z)\mathbf{S}_v(z)Q_{vv}^T = 0$, which

implies that $\mathbf{a}_v^T(z) = 0$. Using $\mathbf{a}_v^T(z) = 0$ and (48), we have $\mathbf{a}_{v-1}^T(z)\mathbf{S}_{v-1}(z)Q_{v-1,v-1}^T = 0$ and therefore $\mathbf{a}_{v-1}^T(z) = 0$. Sequentially, we find that $\mathbf{a}_i^T(z) = 0$ for any $i \geq j+1$. Thus, also from (48), $a_j^T \mathbf{S}_j(z)c_i = 0, i = 2, \dots, m_j$, where c_i is the i th column of Q_{jj}^T . Note that a_j^T is the first row of Q_{jj} . We have

$$Q_{jj}\mathbf{S}_j(z)Q_{jj}^T = \begin{bmatrix} \boldsymbol{\alpha}_1(z) & 0_{1 \times (m_j-1)} \\ 0_{(m_j-1) \times 1} & \mathbf{Q}_2(z) \end{bmatrix} \quad (50)$$

i.e.,

$$Q_{jj}\mathbf{S}_j(z) = \begin{bmatrix} \boldsymbol{\alpha}_1(z) & 0_{1 \times (m_j-1)} \\ 0_{(m_j-1) \times 1} & \mathbf{Q}_2(z) \end{bmatrix} Q_{jj}^{-T}. \quad (51)$$

Since every two diagonals of $\mathbf{S}_j(z)$ are not equal up to a (constant) scaling, the first row of Q_{jj} , i.e., a_j^T , must have only one nonzero entry. Without losing generality, we assume that the first entry is not zero. We have shown that the L th row of $\mathbf{Q}(z)$ has only one nonzero element, which is the L th one. Using this property and apply (48) again, we conclude that the L th column of $\mathbf{Q}(z)$ also has only one nonzero element, which is obviously the L th one. Hence, by proper row and column permutations P_1 and P_2 , we get

$$P_1\mathbf{Q}(z)P_2 = \begin{bmatrix} \boldsymbol{\alpha} & 0_{1 \times (m-1)} \\ 0_{(m-1) \times 1} & \mathbf{Q}_2(z) \end{bmatrix}. \quad (52)$$

Then, from (46)

$$\mathbf{H}(z) = \mathbf{B}(z)PP_1^T \begin{bmatrix} \boldsymbol{\alpha} & 0_{1 \times (m-1)} \\ 0_{(m-1) \times 1} & \mathbf{Q}_2(z) \end{bmatrix} P_2^T \quad (53)$$

which implies

$$\mathbf{B}(z) = \mathbf{H}(z)P_2 \begin{bmatrix} \alpha^{-1} & 0_{1 \times (m-1)} \\ 0_{(m-1) \times 1} & \mathbf{Q}_2^{-1}(z) \end{bmatrix} P_1 P^T. \quad (54)$$

Hence, $\mathbf{b}_1(z)$ is equal to a column $\mathbf{h}(z)$ of $\mathbf{H}(z)$ up to scaling, and $\mathbf{B}_2(z)$ is a polynomial basis of $\text{span}\{\mathbf{H}_c(z)\}$, where $\mathbf{H}_c(z)$ is the remaining submatrix of $\mathbf{H}(z)$ without $\mathbf{h}(z)$. Recall (44). Equations (12) and (13) hold, and the proof is completed. \square

APPENDIX B

PROOF OF THEOREM 2

The proof of Theorem 2 is based on a color property of non-white signals. Let $r_{xx}(\tau) = \mathbb{E}\{x(n)x(n-\tau)\}, \tau = 0, 1, \dots, K$ denote the autocovariances of $x(n)$. $x(n)$ is a white noise if and only if $r_{xx}(\tau) = 0$ for all $\tau \neq 0$. We use

$$Cr(x) = \frac{1}{r_{xx}(0)}[r_{xx}(1), r_{xx}(2), \dots, r_{xx}(K)]$$

to describe the color of a signal $x(n)$ and call it the color vector.

A. Color Maximization Under Instantaneous Mixtures

Lemma 1: Let $x_i(n), i = 1, 2, \dots, m$, be m independent time signals. Define $x(n) = [x_1(n), x_2(n), \dots, x_m(n)]^T$. Let $b = [b_1, b_2, \dots, b_m]^T$ be a nonzero vector in \mathbb{R}^m . Then, the color vector of $b^T x(n)$ is a convex combination of $Cr(x_i)$, i.e.,

$$Cr(b^T x) = \sum_{i=1}^m \lambda_i Cr(x_i), \exists \lambda_i \geq 0, \sum_{i=1}^m \lambda_i = 1. \quad (55)$$

Furthermore, if each of the two signals $x_i(n)$ and $x_j(n)$ have different color vectors, i.e., $Cr(x_i) \neq Cr(x_j)$, then each local maximizer b^* of $\|Cr(b^T x)\|$ must have only one nonzero entry.

Proof: By direct calculation, the k th entry of $Cr(b^T x)$ is

$$\{Cr(b^T x)\}_k = \frac{\sum_{i=1}^m b_i^2 r_{x_i x_i}(k)}{\sum_{i=1}^m b_i^2 r_{x_i x_i}(0)} = \frac{\sum_{i=1}^m a_i^2 \{Cr(x_i)\}_k}{\sum_{i=1}^m a_i^2}$$

where $a_i = b_i \sqrt{r_{x_i x_i}(0)}$. Hence, we have

$$Cr(b^T x) = \frac{\sum_{i=1}^m a_i^2 Cr(x_i)}{\sum_{i=1}^m a_i^2}$$

which leads to (55) by letting $\lambda_i = a_i^2 / \sum_{k=1}^m a_k^2$.

By (55), $Cr(b^T x)$ is in the convex hull of $\{Cr(x_i)\}_{i=1}^m$. Hence, if $Cr(x_i) \neq Cr(x_j)$ for all $i \neq j$, then each local maximum of $\|Cr(b^T x)\|$ is achieved by some vertex point $Cr(x_i)$. In other words, each local maximizer b^* of $\|Cr(b^T x)\|$ must have only one nonzero entry. \square

B. Color Maximization Under Convolutional Filters

Let $a(z) = \sum_{i=0}^q a_i z^{-i}$ be a polynomial operator. Then, $[a(z)]x(n) = \sum_{i=0}^q a_i x(n-i)$. An interesting property concerning the color of $[a(z)]x(n)$ is that the maximization (locally) of its color implies $a(z) = \pm z^{-q} a(z^{-1})$. This means $a = \pm J a$, where a is the coefficient vector of $a(z)$, and J is a $(q+1) \times (q+1)$ matrix that has ones along the antidiagonal and zeros everywhere else. Note that JQ reverses the rows of Q , QJ reverses the columns of Q , and JQJ reverses both the rows and columns of Q .

Lemma 2: Let $x(n)$ be a signal of time n and $a^*(z)$ be a local maximizer of $\|Cr\{[a(z)]x(n)\}\|$ under the constraint that $\deg[a(z)] \leq q$. Then, $a^*(z) = \pm z^{-q} a^*(z^{-1})$.

Proof: We only need to show that if $a(z) \neq \pm z^{-q} a(z^{-1})$, then $a(z)$ is not a local maximizer. Let $a_+(z) = a(z) + z^{-q} a(z^{-1})$, $a_-(z) = a(z) - z^{-q} a(z^{-1})$, and $b(z, \lambda) = \lambda a_+(z) + (1-\lambda)a_-(z)$. Obviously, $b(z, \lambda)$ equals $a(z)$ when $\lambda = 1/2$. We will show that $\lambda = 1/2$ is not a local maximizer of $\|Cr\{[b(z, \lambda)]x\}\|$.

By direct calculation, the k th entry of $Cr\{[b(z, \lambda)]x\}$ is

$$\{Cr\{[b(z, \lambda)]x\}\}_k = \frac{[\lambda a_+ + (1-\lambda)a_-]^T X(k) [\lambda a_+ + (1-\lambda)a_-]}{[\lambda a_+ + (1-\lambda)a_-]^T X(0) [\lambda a_+ + (1-\lambda)a_-]} \quad (56)$$

where a_+, a_- are the coefficient vector of polynomials $a_+(z)$ and $a_-(z)$, respectively, and

$$X(k) = \begin{bmatrix} r_{xx}(k) & \cdots & r_{xx}(k+q) \\ \vdots & \ddots & \vdots \\ r_{xx}(k-q) & \cdots & r_{xx}(k) \end{bmatrix} \in \mathbb{R}^{(q+1) \times (q+1)}. \quad (57)$$

Since $a_+ = J a_+$, $a_- = -J a_-$, and $X^T(k) = J X(k) J$, we have $a_-^T X(k) a_+ = -a_-^T X^T(k) a_+ = -a_+^T X(k) a_-$, i.e., $a_-^T X(k) a_+ + a_+^T X(k) a_- = 0$. Following (56)

$$\{Cr\{[b(z, \lambda)]x\}\}_k = \frac{\lambda^2 a_+^T X(k) a_+ + (1-\lambda)^2 a_-^T X(k) a_-}{\lambda^2 a_+^T X(0) a_+ + (1-\lambda)^2 a_-^T X(0) a_-} \quad (58)$$

and then

$$Cr\{[b(z, \lambda)]x\} = \alpha(\lambda) Cr\{[a_+(z)]x\} + [1-\alpha(\lambda)] Cr\{[a_-(z)]x\}$$

where

$$\alpha(\lambda) = \frac{\lambda^2 a_+^T X(0) a_+}{\lambda^2 a_+^T X(0) a_+ + (1-\lambda)^2 a_-^T X(0) a_-}.$$

Note that $\alpha(\lambda) \in [0, 1]$ is a monotonous function when $\lambda \in [0, 1]$. The norm of the color vector $Cr\{[b(z, \lambda)]x\}$ has no local maxima in $(0, 1)$ since it is a convex combination of two different vectors $Cr\{[a_+(z)]x\}$ and $Cr\{[a_-(z)]x\}$. In particular, $\lambda = 1/2$ is not a local maximizer of $\|Cr\{[b(z, \lambda)]x\}\|$, and the proof is then completed. \square

C. Color Maximization Under Convolutive Mixtures of Multiple Sources

Lemma 3: Let $x_i(n), i = 1, 2, \dots, m$, be m independent time signals, $x(n) = [x_1(n), \dots, x_m(n)]^T$ and $\mathbf{b}^*(z) = [b_1^*(z), b_2^*(z), \dots, b_m^*(z)]^T$ be a local maximizer of $\|Cr\{[\mathbf{b}^T(z)]x\}\|$ under the constraint that $\deg[b_i(z)] \leq q_i$. Then, each nonzero $b_i^*(z)$ is a local maximizer of $\|Cr\{[b_i(z)]x_i\}\|$, and all nonzero $[b_i^*(z)]x_i$ have the same color vector.

Proof: Suppose that the first r entries of $\mathbf{b}^*(z)$ are nonzero. By Lemma 1, all $Cr\{[b_i^*(z)]x_i\}$ must be equal, i.e., $Cr\{[b_i(z)]x_i\} = c$ for any $i = 1, 2, \dots, r$, where c is a nonzero constant vector. Now we show that $b_i^*(z)$ is also a local maximizer of $\|Cr\{[b_i(z)]x_i\}\|$ under the constraint that $\deg[b_i(z)] \leq q_i$. Proceeding by contradiction, assume that $b_1^*(z)$ is not a local maximizer of $\|Cr\{[b_1(z)]x_1\}\|$; then, there exists $d(z)$ and a small $\gamma_0 > 0$ such that $\|Cr\{[b_1^*(z) + \gamma d(z)]x_1\}\| > \|Cr\{[b_1^*(z)]x_1\}\| = \|c\|$ for any $\gamma \in (0, \gamma_0]$. If γ_0 is sufficiently small, $Cr\{[b_1^*(z) + \gamma d(z)]x_1\}$ will be very close to c and then $\|\lambda Cr\{[b_1^*(z) + \gamma d(z)]x_1\} + (1-\lambda)c\| > \|c\|$ for any $\lambda \in (0, 1], \gamma \in (0, \gamma_0)$.

By Lemma 1, $Cr\{[b_1^*(z) + \gamma d(z)]x_1 + \sum_{i=2}^r [b_i^*(z)]x_i\}$ is a convex combination of $Cr\{[b_1^*(z) + \gamma d(z)]x_1\}$ and c . Therefore, $\|Cr\{[b_1^*(z) + \gamma d(z)]x_1 + \sum_{i=2}^r [b_i^*(z)]x_i\}\| > \|c\|$ for any $\gamma \in (0, \gamma_0]$. This implies that $\mathbf{b}^*(z)$ is not a local maximizer and then contradicts the assumption. Hence, $b_1^*(z)$ must be a local maximizer of $\|Cr\{[b_1(z)]x_1\}\|$. Similarly, any nonzero $b_i(z)$ is a local maximizer of $\|Cr\{[b_i(z)]x_i\}\|$. \square

D. Proof of Theorem 2

Denote $b_i^T = b^T W \mathcal{T}_l \{\mathbf{h}_i(z)\} \in \mathbb{R}^{l+q_i+1}$, where $\mathbf{h}_i(z)$ is the i th column of $\mathbf{H}(z)$ and of degree q_i . Let $b_i(z) = \sum_{k=0}^{l+q_i} b_{i,k} z^{-k}$, where $b_{i,k}$ is the k th entry of b_i . Since the sources are uncorrelated

$$P(\tau) = \sum_{i=1}^m \mathcal{T}_l \{\mathbf{h}_i(z)\} R_{ii}(\tau) [\mathcal{T}_l \{\mathbf{h}_i(z)\}]^T$$

where

$$R_{ii}(\tau) = \begin{bmatrix} r_{x_i x_i}(\tau) & \cdots & r_{x_i x_i}(\tau + l + q_i) \\ \vdots & \ddots & \vdots \\ r_{x_i x_i}(\tau - l - q_i) & \cdots & r_{x_i x_i}(\tau) \end{bmatrix} \in \mathbb{R}^{(l+q_i+1) \times (l+q_i+1)} \quad (59)$$

and $r_{x_i x_i}(\tau) = \mathbb{E}\{x_i(n)x_i(n-\tau)\}$.

Then under the constraint $b^T b = 1$

$$\mathcal{J}_0(b) = \left\| Cr \left\{ \sum_{i=1}^m [b_i(z)]x_i(n) \right\} \right\|^2 - 1. \quad (60)$$

Note that $W[\mathcal{T}_l \{\mathbf{h}_1(z)\}, \mathcal{T}_l \{\mathbf{h}_2(z)\}, \dots, \mathcal{T}_l \{\mathbf{h}_m(z)\}]$ is a nonsingular square matrix. If b is a local maximizer of $\mathcal{J}_0(b)$, then $[b_1(z), b_2(z), \dots, b_m(z)]^T$ is a local maximizer of $\|Cr\{\sum_{i=1}^m [b_i(z)]x_i(n)\}\|$ under the constraint that $\deg[b_i(z)] \leq l + q_i + 1$. By Lemma 3, all nonzero $Cr\{[b_i(z)]x_i(n)\}$ are equal, and therefore all nonzero $b_i(z)b_i(z^{-1})s_i(z)$ are equal up to scaling. Next, we prove that only one $b_i(z)$ is not zero. Proceeding by contradiction, assume that $b_1(z)$ and $b_2(z)$ are not zero. Then, $b_1(z)s_1(z)b_1(z^{-1}) = \beta b_2(z)s_2(z)b_2(z^{-1})$ for some real number $\beta \neq 0$. By A3'), $s_1(z)$ has a root of odd repeated number, and this root is not shared by $s_2(z)$. Then, this root should be a root of $b_1(z)b_1(z^{-1})$. However, by Lemma 2, $b_1(z) = z^{-(l+q_1+1)}b_1(z^{-1})$, i.e., $b_1(z)$ shares all its roots with $b_1(z^{-1})$ and therefore any root of $b_1(z)b_1(z^{-1})$ is of even repeated number. This introduces contradiction. Thus, only one $b_i(z)$ is nonzero, i.e., (24) is true for $\mathbf{h}(z)$ being a column of $\mathbf{H}(z)$ and $\mathbf{H}_c(z)$ being the submatrix of $\mathbf{H}(z)$ without $\mathbf{h}(z)$.

Finally, we prove that \mathbf{H}_c can be identified, up to a right nonsingular polynomial matrix, from $b_*^T W \mathcal{T}_l \{\mathbf{H}_c(z)\} = 0$ and $U_2^T \mathcal{T}_l \{\mathbf{H}_c(z)\} = 0$. Let $\hat{\mathbf{H}}_c$ be an irreducible and column-reduced polynomial matrix meeting the above requirement. Since U_2 determines $\mathbf{H}(z)$ up to an right nonsingular polynomial matrix, there exists a polynomial matrix $\mathbf{A}(z) \in \mathbb{R}[z]^{m \times (m-1)}$ such that $\hat{\mathbf{H}}_c(z) = [\mathbf{h}(z), \mathbf{H}_c(z)]\mathbf{A}(z)$. Since $b_*^T W \mathcal{T}_l \{\mathbf{h}(z)\} \neq 0$, the first row of $\mathbf{A}(z)$ must be a zero row. That is, $\hat{\mathbf{H}}_c$ equals \mathbf{H}_c up to a right nonsingular polynomial matrix. \square

E. A Brief Discussion

From the proof of Theorem 2, we know that the local maximization of the proposed cost function implies the local maximization of $\|Cr\{[\mathbf{b}^T(z)]x\}\|$. We would hope that this local maximization implies that only one $b_i(z)$ is not zero. By Lemma 3, we see that this is true if the following condition holds: for any $i \neq j$, the local maximization of $\|Cr\{[b_i(z)]x_i\}\|$ and $\|Cr\{[b_j(z)]x_j\}\|$ does not result in the same maximal color vector, i.e., $Cr\{[b_i(z)]x_i\} \neq Cr\{[b_j(z)]x_j\}$. This condition is true for almost all $x_i(n)$ and $x_j(n)$ having different color vectors. The assumption A3'), though not very restrictive, is only a sufficient condition.

APPENDIX C

AN ALTERNATIVE FORM OF THE GDES METHOD

The local minimum problem of the cost function (19) can also be handled quite effectively by the following algorithm.

A. Local Refinement of $\mathbf{H}(z)$

Assume that we have an initial estimate $\hat{\mathbf{H}}(z)$ of $\mathbf{H}(z)$. Such an estimate can be found by using the conventional subspace method. To refine this estimate, we can do the following. For each column, say $\hat{\mathbf{h}}_i(z)$, of $\hat{\mathbf{H}}(z)$, we perform the AP with X_1^T initialized by a basis of the left null space of $W \mathcal{T}_l \{\hat{\mathbf{h}}_i(z)\}$. At

convergence, we obtain the minimizer $(X_{i,1}, X_{i,2})$ of the cost and then a pair of the group decorrelators $(\mathbf{G}_{1,i}(z), \mathbf{G}_{2,i}(z))$ from (20). The refined i th column $\mathbf{h}_i(z)$ of the channel matrix then follows from the equation $\mathbf{G}_{1,i}(z)\mathbf{h}_i(z) = 0$ (i.e., $C_0\{\mathbf{h}_i(z)\}$ is the least right singular vector of $C_0\{\mathbf{G}_{1,i}(z)\}$). We repeat the above for each column of $\hat{\mathbf{H}}(z)$ to obtain a refined estimate of $\mathbf{H}(z)$.

B. Global Refinement of $\mathbf{H}(z)$

The global refinement shown next is a heuristic approach to yield more robust results than the local refinement. The basic idea of the global refinement is as follows. Given an initialization of $\mathbf{H}(z)$, the local refinement is carried out to estimate all columns of $\mathbf{H}(z)$. But then, only the best estimated column is selected (based on the values of the cost (19)), and the rest is discarded. To find a new column, a new process of the local refinement is carried out with a new initialization of $\mathbf{H}(z)$ without the previously estimated columns.

Let $\mathbf{H}_1(z)$ be an initial estimate of the channel matrix obtained by the subspace method without the joint diagonalization, and hence $\mathbf{H}_1(z)$, at its best, is no better than the exact channel matrix $\mathbf{H}(z)$ with a right-multiplicative constant matrix.

We will denote by $\mathbf{H}_k(z)$ an estimate of $\mathbf{H}(z)$ with $k - 1$ columns removed. Clearly, $\mathbf{H}_k(z)$ has $m - k + 1$ columns. For each value of k , only one column in $\mathbf{H}(z)$ is estimated. The algorithm is as follows with the initial index $k = 1$:

- 1) For the i th column, say $\mathbf{h}_{k,i}(z)$, of $\mathbf{H}_k(z)$, perform the AP with X_1^T initialized by a basis of the left null space of $WT_l\{\mathbf{h}_{k,i}(z)\}$. At the convergence of the AP, we then obtain the pair $(X_{i,1}, X_{i,2})$ with the cost ε_i and then a candidate pair of group decorrelators $(\mathbf{G}_{1,k,i}(z), \mathbf{G}_{2,k,i}(z))$ from (20). Among all $(\mathbf{G}_{1,k,i}(z), \mathbf{G}_{2,k,i}(z))$ for $i = 1, 2, \dots, m - k + 1$, select the pair $(\mathbf{G}_{1,k}(z), \mathbf{G}_{2,k}(z))$ with the minimal cost.
- 2) Estimate the k th column $\mathbf{h}^{(k)}(z)$ of the channel matrix via $\mathbf{G}_{1,k}(z)\mathbf{h}^{(k)}(z) = 0$. If $k = m$, stop. Otherwise, set $k = k + 1$.
- 3) Stack $\{\mathbf{G}_{2,i}(z)\}_{i=1}^{k-1}$ as $\mathbf{G}_k(z) = [\mathbf{G}_{2,1}^T(z) \ \dots \ \mathbf{G}_{2,k-1}^T(z)]^T$, compute $\mathbf{H}_k(z)$ using $\mathbf{G}_k(z)\mathbf{H}_k(z) = 0$ and then go to Step 1).

Based on our experiment, the global refinement is much more likely to yield the global minimization of the cost (19) than the local refinement. In fact, among 1000 randomly selected channel matrices with dimension 4×3 and with degree 1, we only had one case where the global refinement did not yield the global minimum of (19). Although less complex and more heuristic, this algorithm has a comparable performance as the one shown in Section IV. However, the global convergence is not guaranteed especially when the channel matrix is of relatively high degrees.

ACKNOWLEDGMENT

The authors would like to thank J. P. Reilly and the reviewers for their comments, which were helpful to improve the presentation of this paper.

REFERENCES

- [1] K. Abed-Meraim and Y. Hua, "Blind identification of multi-input multi-output system using minimum noise subspace," *IEEE Trans. Signal Process.*, vol. 45, no. 1, pp. 254–258, Jan. 1997.
- [2] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Trans. Signal Process.*, vol. 45, no. 2, pp. 434–443, Feb. 1997.
- [3] R. Bitmead, S. Kung, B. Anderson, and T. Kailath, "Greatest common division via generalized Sylvester and Bezout matrices," *IEEE Trans. Autom. Control*, vol. 23, no. 6, pp. 1043–1047, Dec. 1978.
- [4] G. D. Forney, "Minimal bases of rational vector spaces, with applications to multivariable linear systems," *SIAM J. Control*, vol. 13, no. 3, pp. 493–520, 1975.
- [5] A. Gorokhov and P. Loubaton, "Subspace-based techniques for blind separation of convolutive mixtures with temporally correlated sources," *IEEE Trans. Circuits Syst. I, Fundam. Theory Appl.*, vol. 44, no. Sep., pp. 813–820, 1997.
- [6] Y. Hua, S. An, and Y. Xiang, "Blind identification of FIR MIMO channels by decorrelating subchannels," *IEEE Trans. Signal Process.*, vol. 51, no. 5, pp. 1143–1155, May 2003.
- [7] *High-Resolution and Robust Signal Processing*, Y. Hua, A. Gershman, and Q. Cheng, Eds., Marcel Dekker, New York, 2003.
- [8] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall PTR, 1999.
- [9] P. Loubaton and E. Moulines, "On blind multiuser forward link channel estimation by the subspace method: Identifiability results," *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2366–2376, Aug. 2000.
- [10] P. Loubaton, E. Moulines, and P. Regalia, "Subspace method for blind identification and deconvolution," in *Signal Processing Advances in Wireless and Mobile Communications*, G. B. Giannakis, Y. Hua, P. Stoica, and L. Tong, Eds. Englewood Cliffs, NJ: Prentice-Hall, 2001, vol. 1, Trends in channel estimation and equalization, ch. 3, pp. 63–112.
- [11] T. Ma, Z. Ding, and S. F. Yau, "A two-stage algorithm for MIMO blind deconvolution of colored input signals," *IEEE Trans. Signal Process.*, vol. 48, no. 4, pp. 1187–1192, Apr. 2000.
- [12] J. H. Manton, "Optimization algorithms exploiting unitary constraints," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 635–650, Mar. 2002.
- [13] K. Rahbar, J. P. Reilly, and J. H. Manton, "Blind identification of MIMO FIR channels driven by quasistationary sources using second-order statistics: A frequency domain approach," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 406–417, Feb. 2004.
- [14] P. Stoica and R. Moses, *Introduction to Spectra Analysis*. Englewood Cliffs, NJ: Prentice-Hall, 1997.
- [15] L. Tong, V. C. Soon, R. Liu, and Y. Huang, "AMUSE: A new blind identification algorithm," presented at the IEEE Int. Symp. on Circuits Systems (ISCAS), New Orleans, LA, 1990.



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

He was with the Institute of Systems Science, Chinese Academy of Sciences, Beijing, where he was a Postdoctoral Research Fellow from 1996 to 1998. In 1998, he joined the Beijing Institute of Technology, Beijing, where he was an Associate Professor from 1998 to 2001. From 2001 to 2004, he was a Research Fellow with The University of Melbourne, Parkville, Australia. He joined Curtin University of Technology, Perth, WA, Australia, in 2004, where he is currently a Postdoctoral Research Fellow. His research interests include signal processing, robust control, and machine learning.



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

From 1988 to 1989, he was a Research Fellow at Syracuse, consulting for Syracuse Research, New York, and Aeritalia, Italy. He was with the University of Melbourne, Australia, as a Lecturer from 1990 to 1992, a Senior Lecturer from 1993 to 1995, and a Reader and Associate Professor from 1996 to 2001.

He served as a Visiting Professor with the Hong Kong University of Science and Technology from 1999 to 2000 and consulted for Microsoft Research, WA, in summer 2000. Since February 2001, he has been Professor of Electrical Engineering with the University of California, Riverside. He is a Co-Editor of *Signal Processing Advances in Wireless and Mobile Communications* (Prentice-Hall, 2001) and *High-Resolution and Robust Signal Processing* (Marcel Dekker, 2003). He is also an author/coauthor of more than 250 articles in journals, conference proceedings, and books. His publications span the fields of high-resolution signal processing, blind system identification, wireless communications, and sensor networks.

Dr. Hua received a Chinese Government Scholarship for Overseas Graduate Study from 1983 to 1984 and a Syracuse University Graduate Fellowship from 1985 to 1986. He served as Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING and the IEEE SIGNAL PROCESSING LETTERS from 1994 to 2002. He has been a Member of the IEEE Signal Processing Society (SPS) Technical Committees for Underwater Acoustic Signal Processing from 1997 to 1998; Sensor Array and Multichannel Signal Processing from 1999 to 2001, 2002 to 2004, and 2005 to present; and Signal Processing in Communications from 2002 to 2004 and 2005 to present.



Jonathan H. Manton (S'95–A'98–M'02–SM'03) was born in April 1973. He received the B.S. degree in mathematics and the B.Eng. degree (electrical) in 1995 and the Ph.D. degree in 1998, all from the University of Melbourne, Australia.

From 1998 to 2004, he was with the Department of Electrical and Electronic Engineering at the University of Melbourne. During that time, he held a Post-doctoral Research Fellowship and then subsequently a Queen Elizabeth II Fellowship, both from the Australian Research Council. In 2005, he became a Full Professor in the Department of Information Engineering, Research School of Information Sciences and Engineering (RSISE), The Australian National University. His research interests range from pure mathematics (e.g., commutative algebra, algebraic geometry, and differential geometry) to engineering (e.g., signal processing and wireless communications).

Dr. Manton was an Associate Editor for the conference editorial board, IEEE Control and Systems Society, and currently is an Associate Editor for the IEEE TRANSACTIONS ON SIGNAL PROCESSING. He is also on the IEEE Signal Processing for Communications Technical Committee.



Zheng Fang received the B.Sc. degree from the University of Science and Technology of China (USTC), Hefei, China, in 2001 and the M.Sc. degree from University of California, Riverside (UCR), in 2003, both in electrical engineering. He is currently working toward the Ph.D. degree with the Department of Electrical Engineering, UCR.

His current research interests lie in the areas of signal processing and communications, including blind channel identification, blind source separation, and MIMO systems.