Blind Identification of MIMO FIR Systems Driven by Quasi-Stationary Sources Using Second Order Statistics: A Frequency Domain Approach

Kamran Rahbar, James P. Reilly+ and Jonathan H. Manton†

Abstract

This paper discusses a frequency domain method for blind identification of MIMO convolutive channels driven by white quasi-stationary sources. The sources can assume arbitrary probability distributions and in some cases they can even be all Gaussian distributed. We also show that under slightly more restrictive assumptions the algorithm can be applied to the case when the sources are colored, non-stationary signals. We demonstrate that by using the second order statistics of the channel outputs, under mild conditions on the non-stationarity of sources, and under the condition that channel is column-wise coprime, the impulse response of the MIMO channel can be identified up to an inherent scaling and permutation ambiguity. We prove that by using the new algorithm, under the stated assumptions, a uniform permutation across all frequency bins is guaranteed, and the inherent frequency dependent scaling ambiguities can be resolved. Hence no post processing is required as is the case with previous frequency domain algorithms. We further present an efficient, two step frequency domain algorithm for identifying the channel. Numerical simulations are presented to demonstrate the performance of the new algorithm[1].

I. Introduction

Blind identification of a Multi–Input, Multi–Output (MIMO) FIR system deals with identifying the impulse response of a unknown system using only the system output data, and in particular, without any (or the least amount of) knowledge about the inputs. Multichannel blind identification has been of great interest to both the communications and signal processing communities and there have been numerous publications in both societies on this subject (see [1] for a review of recent blind channel estimation and identification techniques). Extensive literature has been dedicated to a special instance of the MIMO channel identification problem, namely the single–input single–output (SISO) case [2][3]. When there is more than one output derived from a common source the problem is referred to as Single-Input, Multiple-Output (SIMO) blind identification [4] [5][2]. In the literature there are fewer works related to the more general MIMO problem. An exception is in the memoryless channel case, where MIMO blind identification is closely related to blind source separation for instantaneous mixtures. This latter problem has recently been discussed extensively in both the signal processing and neural network literature [6]. Also see [7] for indeterminacy and identification conditions in this case.

Blind identification of a MIMO channel can be used for blind signal separation and blind multichannel equalization in a convolutive environment. In most methods for blind source separation of convolutive mixtures, the sources are only separated up to a distorted (filtered) version of the original sources [8][9]. Identification of the complete MIMO channel not only permits blind signal separation, but in addition gives the potential for equalization of the outputs so that the original sources can be fully recovered. This by itself has many applications; e.g., in data communication for eliminating the ISI without knowledge of the input signal, or in speech processing for reverberation cancellation and speech enhancement, and in biomedical instrumentation for sup-

Department of Electrical & Computer Engineering
McMaster University,1280 Main St. W, Hamilton, Ontario, Canada L8S 4K1
†Department of Electrical and Electronic Engineering
The University of Melbourne, Victoria 3010, Australia.
Email: kamran@reverb.crl.mcmaster.ca, reillyj@mcmaster.ca, jon@ee.mu.oz.au
+Corresponding Author: Phone: (905)525-9140 Ext:22895 Fax: (905)521-2922

[1]Permission to publish the abstract separately is granted.

[2]In fact, for some cases SISO can often be converted into a SIMO problem by oversampling the outputs of the channel.

pressing cardio or other interfering signals from desired signals, such as electromyogram or electrogastrogram signals.

In this paper we consider the problem of blind identification of MIMO channels with finite memory. Previous work in this area can be divided into two groups. The first group uses higher order statistical (HOS) methods that exploit the higher–order moments (or higher–order spectra) of the output signals to identify the channel; e.g., [10][11]. For HOS methods, a non-Gaussian condition on the inputs is necessary. Also the main limitation of the HOS methods is their slow convergence due to large estimation variance of higher order moments. As a result of this they usually require large sample sizes for good time-averaged estimates of the higher order statistics[12].

The second group are the second–order statistical (SOS) methods that rely only on the second–order moments of the output signals to identify the channel [8][13][14]. SOS methods have the following advantages: they usually have a simple implementation, and in some cases a closed form solution even exists; also, they often require fewer data samples and they can handle Gaussian distributed inputs. In comparison to HOS methods, a significant disadvantage is they require additional assumptions on both the channel and input signals. For example, SOS methods cannot be applied to the general MIMO blind identification problem when the inputs are white and stationary. For colored inputs, some additional channel conditions are required for SOS identification [15].

The proposed method is a frequency domain approach that exploits second order non-stationarity of the input signals. Previously, for both HOS and SOS identification methods, the inputs have been assumed stationary. However, some methods have been proposed that exploit non-stationarity of the input signals. A non-stationarity assumption can be justified by realizing that most real world signals are inherently non-stationary (e.g., speech or biological signals). In communication systems non-stationarity in the form of *cyclostationarity* can be created by over-sampling the received signals [7], or as suggested in [16], cyclostationarity can be induced in the transmitted signal. Nonstationarity can also be induced by careful control of the transmit power level.

So far, most blind identification methods that exploit non–stationarity address only the SISO case. In [17] [18], methods have been proposed for blind source separation of instantaneously mixed, non-stationary (cyclostationary in the second reference) signals. References [9] and [19] consider blind source separation of colored non-stationary signals when the mixing system is convolutive.

This paper is an extended version of [20]. The proposed procedure is based on a novel formulation of the *joint diagonalization* technique[21],[22] for solving the MIMO blind identification problem. We assume that the statistics of the input signals are slowly varying with time; i.e., we assume that they are *quasi-stationary* [23]. Furthermore, we rely only on the second–order statistics of the input signal. This gives us the advantages of SOS methods, and also permits identification in situations where stationary based SOS and HOS methods fail; e.g., when the input signals are temporally white and Gaussian distributed. Although the main focus of this paper is on white, non-stationary signals, we show that under some additional conditions the same algorithm can also be applied to non-stationary colored signals.

In this paper we demonstrate sufficient identifiability conditions for blind identifiability of a MIMO system in the frequency domain under a second order non-stationarity assumption of the inputs. We also prove that a limited number of frequency samples are enough to identify the channel and to this end we derive an upper bound on the smallest number of frequency samples sufficient for blind identification of the MIMO system. This bound is lower than what has been perviously used in frequency domain blind identification or blind source separation methods ([11] [9]) and results in significant computational savings for the proposed algorithm. The advantage of using a frequency domain rather than a time domain approach is that using the frequency domain method, one can decompose a time domain estimation problem, with a large number of parameters, into multiple, independent estimation problems, with much fewer parameters to be estimated at each frequency bin. As a result of this in general the frequency domain estimation algorithms have a simpler implementation and better convergence properties. The main difficulties with frequency–domain blind identification of MIMO channels are the arbitrary column permutations and scaling ambiguities of the estimated frequency response of the channel at each frequency bin.

In this paper we exploit the quasi–stationary nature of input signals, such that the proposed algorithm

results in a *common* permutation for the estimated channel frequency response across all frequency bins. Further we demonstrate that if the channel is column-wise coprime, then the problem of arbitrary scaling factors across the frequency bins can be resolved, thus avoiding the limitations of frequency domain methods.

The organization of this paper is as follows: The problem formulation including the set of required assumptions is presented in Section 2. Section 3 establishes channel identifiability results based on only the second–order statistics and the quasi-stationarity property of the input signals. In Section 4 we present a two-stage frequency domain algorithm for blind identification of MIMO channels. Simulation results are described in Section 5. The first simulation scenario is a synthetic data case where two inputs are quasi–stationary zero–mean Gaussian noise signals. The second simulation uses colored sources, which are created by passing the white signals in the first simulation through an AR filter. Third simulation uses two speech signals as inputs. In each of these cases, the underlying channel was successfully identified. We also compare our results with those obtained using the HOS blind identification method in [11]. Conclusions and final remarks are presented in Section 6.

II. Problem statement

A. Notation

We use *plain text* bold upper and lower case letters to show matrices and vectors respectively in the time domain; e.g., $\mathbf{H}(t)$ and $\mathbf{h}(t)$. We use bold *calligraphic* upper case letters, and bold *italic* lower case letters to represent matrices and vectors respectively in the frequency and z– domains; e.g., $\boldsymbol{\mathcal{H}}(\omega)$ and $\boldsymbol{h}(\omega)$. The remaining notational conventions are listed as follows:

$(\cdot)^T$	Transpose
$(\cdot)^\dagger$	Hermitian transpose
$E\{\cdot\}$	Expectation Operator
diag($\mathbf{a}$)	Forms a diagonal matrix from the vector $\mathbf{a}$.
diag($\mathbf{A}$)	Forms a column vector from the diagonal elements of $\mathbf{A}$.
vec$\{\mathbf{A}\}$	Forms a column vector by stacking the columns of $\mathbf{A}$.
mat$\{\mathbf{a}\}$	Forms a $J \times J$ matrix from a $J^2 \times 1$ column vector $\mathbf{a}$
$\mathbf{A}^+$	Pseudo Inverse of the matrix $\mathbf{A}$.

B. Model

We consider the following N-source J-sensor MIMO linear model for the received signal for the convolutive mixing problem:

$$\mathbf{x}(t) = \sum_{l=0}^{L} \mathbf{H}(l)\mathbf{s}(t-l) + \mathbf{n}(t) \quad t \in \mathbb{Z} \tag{1}$$

where $\mathbf{x}(t) = (x_1(t), \cdots, x_J(t))^T \in \mathbb{R}^{J\times 1}$ is the vector of observed signals,
$\mathbf{s}(t) = (s_1(t), \cdots, s_N(t))^T \in \mathbb{R}^{N\times 1}$ is the vector of sources, $\mathbf{H}(t) \in \mathbb{R}^{J\times N}$ is the channel matrix with maximum element order L, and $\mathbf{n}(t) = (n_1(t), \cdots, n_J(t))^T \in \mathbb{R}^{J\times 1}$ is the additive noise vector. The objective is to estimate the $\mathbf{H}(t)$ up to a scaling and permutation factor from the observed signals $\mathbf{x}(t)$. In other words, we are interested in finding $\hat{\mathbf{H}}(t)$ such that for all $0 \le t \le L$ we have

$$\hat{\mathbf{H}}(t) = \mathbf{H}(t)\mathbf{\Pi}\mathbf{D} \tag{2}$$

where $\mathbf{D} \in \mathbb{R}^{N\times N}$ and $\mathbf{\Pi} \in \mathbb{R}^{N\times N}$ are respectively constant diagonal and permutation matrices[3]. In the frequency domain this is equivalent to finding an $\hat{\boldsymbol{\mathcal{H}}}(\omega) \in \mathbb{C}^{J\times N}$ such that:

$$\hat{\boldsymbol{\mathcal{H}}}(\omega) = \boldsymbol{\mathcal{H}}(\omega)\mathbf{\Pi}\mathbf{D} \quad \forall\ \omega \in [0, \pi) \tag{3}$$

where $\boldsymbol{\mathcal{H}}(\omega)$ is the DTFT of the $\mathbf{H}(t)$. Notice that in (3), since we assume that the elements of the channel are real numbers, we only need to estimate $\boldsymbol{\mathcal{H}}(\omega)$ over half of the frequency range; i.e., $\omega \in [0, \pi)$.

[3]Note that frequency independent quantities are denoted using plain text.

C. Main Assumptions

A0*:* $J \geq N \geq 2$; i.e, we have at least as many sensors as sources and number of the sources are at least two.
A1*:* The sources $\mathbf{s}(t)$ are zero mean, second order quasi-stationary white signals. The cross–spectral density matrices of the sources $\boldsymbol{\mathcal{P}}_s(\omega, m)$ are diagonal for all ω and m where ω denotes frequency and m is the time epoch index.
A2*:* Let $\lambda_i(m)$ denote the variance of the i_{th} source at epoch m. We assume the matrix $\boldsymbol{\Gamma}$ given by:

$$\boldsymbol{\Gamma} = \begin{pmatrix} \lambda_1(0) & . & . & . & \lambda_1(M-1) \\ . & & & & . \\ . & & & & . \\ \lambda_N(0) & . & . & . & \lambda_N(M-1) \end{pmatrix} \in \mathbb{R}^{N \times M} \quad M > N \tag{4}$$

has full row rank where M is the total number of epochs, available from the observed data.
A3*:* Channel is modelled by a causal FIR system of the form $\mathbf{H}(t) = [\mathbf{h}_1(t), ..., \mathbf{h}_N(t)]$ and does not change over the entire observation interval. Also $\boldsymbol{\mathcal{H}}(\omega)$, the DTFT of $\mathbf{H}(t)$, has full column rank for all $\omega \in [0, 2\pi)$.
A4*:* The noise $\mathbf{n}(t)$ is zero mean, *iid* across sensors, with power σ^2. The noise is assumed independent of the sources.
A5*:* $\boldsymbol{\mathcal{H}}(z)$, the z-transform of $\mathbf{H}(t)$, is column–wise coprime, i.e. the elements in each column of $\boldsymbol{\mathcal{H}}(z)$ do not share common zeros.
Assumption $A1$ is the core assumption. As is shown later, this non-stationarity assumption enables us to identify a MIMO channel using only the second order statistics of the observed signal. Although in our assumptions we consider white signals, the identifiability results and the algorithm can be extended to the colored signal case under the condition that the spectra of the sources stays constant over the observation interval and only their variances change between epochs. As is shown later, this condition will guarantee a uniform permutation across all frequency bins. The reason behind imposing assumptions $A2, \ldots, A4$ will become clear when we explain the identifiability proof. Notice that assumptions $A1, \ldots, A4$ are sufficient to identify the frequency response of a MIMO channel up to a constant permutation but a frequency dependent scaling ambiguity. This means that if we use the estimated channel to recover the sources, the outputs correspond to a separated but filtered version of the original sources. Assumption A5 enables us to remove the frequency dependent scaling ambiguity so the channel can be identified up to a constant scaling and permutation ambiguity which is the best that can be achieved in MIMO blind identification problems.

III. Blind Identifiability

In this section we present frequency domain blind identifiability results based on the above assumptions using only the second order statistics of the observed signals. Let $\boldsymbol{\mathcal{P}}_x(\omega, m)$ represent the cross-spectral density matrix of the observed signal at frequency ω and time epoch m. Using $A1, A3$ and $A4$ we have:

$$\boldsymbol{\mathcal{P}}_x(\omega, m) = \boldsymbol{\mathcal{H}}(\omega)\boldsymbol{\mathcal{P}}_s(\omega, m)\boldsymbol{\mathcal{H}}^{\dagger}(\omega) + \sigma^2\mathbf{I} \tag{5}$$

where $\boldsymbol{\mathcal{P}}_s(\omega, m)$ by assumption is diagonal for all ω and m. Notice that for white sources we have $\boldsymbol{\mathcal{P}}_s(\omega, m) = \boldsymbol{\Lambda}(m)$ where $\boldsymbol{\Lambda}(m) \in \mathbb{R}^{N \times N}$ is a diagonal matrix for each m and its i_{th} diagonal value, $\lambda_i(m)$, represents the variance of the i_{th} source at epoch m. Based on assumption $A2$ we can immediately see that the vectors $\text{diag}\{\boldsymbol{\Lambda}(m)\}$, $m = 0, \ldots, M-1$, span $\mathbb{R}^N$. For identifiability purposes, we assume that σ^2 is known although for $J > N$, σ^2 can be estimated from the smallest eigenvalue of the matrix $\boldsymbol{\mathcal{P}}_x(\omega, m)$; so for now we consider the following noise free case

$$\boldsymbol{\mathcal{P}}_x(\omega_k, m) = \boldsymbol{\mathcal{H}}(\omega_k)\boldsymbol{\Lambda}(m)\boldsymbol{\mathcal{H}}^{\dagger}(\omega_k) \tag{6}$$

where $\omega_k = (2\pi k)/K$ is the discretized version of ω and K is the number of frequency samples.

Theorem 1: Consider the cross spectral density matrices

$$\boldsymbol{\mathcal{P}}_x(\omega_k, m) = \boldsymbol{\mathcal{H}}(\omega_k)\boldsymbol{\Lambda}(m)\boldsymbol{\mathcal{H}}^{\dagger}(\omega_k) \tag{7}$$

for $k = 0, \ldots, K-1$ and $m = 0, \ldots, M-1$. Under the assumptions that the $\boldsymbol{\mathcal{H}}(\omega_k) \in \mathbb{C}^{J\times N}$ has full column rank and the vectors $\text{diag}\{\boldsymbol{\Lambda}(m)\} \in \mathbb{R}^N$, $m = 0, \ldots, M-1$, span $\mathbb{R}^N$, if there exist matrices $\boldsymbol{\mathcal{B}}(\omega_k) \in \mathbb{C}^{J\times N}$ and $\tilde{\boldsymbol{\Lambda}}(m) \in \mathbb{R}^{N\times N}$, with $\tilde{\boldsymbol{\Lambda}}(m)$ diagonal, such that

$$\boldsymbol{\mathcal{P}}_x(\omega_k, m) = \boldsymbol{\mathcal{B}}(\omega_k)\tilde{\boldsymbol{\Lambda}}(m)\boldsymbol{\mathcal{B}}^{\dagger}(\omega_k) \tag{8}$$

then there exists a permutation matrix $\boldsymbol{\Pi} \in \mathbb{R}^{N\times N}$ and diagonal matrices $\boldsymbol{\mathcal{S}}_k \in \mathbb{R}^{N\times N}$ such that:

$$\boldsymbol{\mathcal{B}}(\omega_k) = \boldsymbol{\mathcal{H}}(\omega_k)\boldsymbol{\Pi}\mathbf{D}e^{-j\boldsymbol{\mathcal{S}}_k} \tag{9}$$

where $\mathbf{D} \in \mathbb{R}^{N\times N}$ is a non-singular diagonal matrix.

Proof: See Appendix A.

In (9) $\mathbf{D}e^{-j\boldsymbol{\mathcal{S}}_k}$ represents a frequency dependent diagonal matrix where the magnitudes of the diagonal values are constant and only their phase varies with k. An important implication of Theorem 1 is that under the assumptions $A0, \ldots, A4$, equation (6) can be used to estimate the channel up to a constant permutation, but a frequency dependent phase ambiguity, across all frequency bins. Thus, the commonly–experienced difficulty with frequency domain approaches to blind identification problems of ensuring a constant permutation over all frequency bins can be alleviated with the proposed approach.

We can easily extend the above theorem to colored sources under a more restrictive assumption. More specifically we assume that only the scale of the power spectral density of each source changes with time. In other words we have:

$$\boldsymbol{\mathcal{P}}_s(\omega_k, m) = \boldsymbol{\Lambda}_1(\omega_k)\boldsymbol{\Lambda}(m) \tag{10}$$

where $\boldsymbol{\Lambda}_1(\omega_k)$ and $\boldsymbol{\Lambda}(m)$ are diagonal matrices for ω_k and m. Based on this, the power spectral density of the observed signals can be written as:

$$\boldsymbol{\mathcal{P}}_x(\omega_k, m) = \boldsymbol{\mathcal{H}}(\omega_k)\boldsymbol{\Lambda}_1(\omega_k)\boldsymbol{\Lambda}(m)\boldsymbol{\mathcal{H}}^{\dagger}(\omega_k). \tag{11}$$

Define $\boldsymbol{\mathcal{H}}_1(\omega_k) = \boldsymbol{\mathcal{H}}(\omega_k)\boldsymbol{\Lambda}_1^{\frac{1}{2}}(\omega_k)$. Then $\boldsymbol{\mathcal{P}}_x(\omega_k, m) = \boldsymbol{\mathcal{H}}_1(\omega_k)\boldsymbol{\Lambda}(m)\boldsymbol{\mathcal{H}}_1^{\dagger}(\omega_k)$ and based on Theorem 1 we have, for any $\boldsymbol{\mathcal{B}}(\omega_k)$ satisfying (8)

$$\begin{aligned}\boldsymbol{\mathcal{B}}(\omega_k) &= \boldsymbol{\mathcal{H}}_1(\omega_k)\boldsymbol{\Pi}\mathbf{D}e^{-j\boldsymbol{\mathcal{S}}_k} \\ &= \boldsymbol{\mathcal{H}}(\omega_k)\boldsymbol{\Lambda}_1^{\frac{1}{2}}(\omega_k)\boldsymbol{\Pi}\mathbf{D}e^{-j\boldsymbol{\mathcal{S}}_k}.\end{aligned} \tag{12}$$

Note that (12) in its general form can be written as

$$\boldsymbol{\mathcal{B}}(\omega_k) = \boldsymbol{\mathcal{H}}(\omega_k)\boldsymbol{\Pi}\boldsymbol{\mathcal{D}}(\omega_k) \tag{13}$$

where $\boldsymbol{\mathcal{D}}(\omega_k) = \boldsymbol{\Lambda}_1^{\frac{1}{2}}(\omega_k)\mathbf{D}e^{-j\boldsymbol{\mathcal{S}}_k}$ is diagonal for all ω_k. In other words, when the sources are colored and (10) is satisfied, $\boldsymbol{\mathcal{H}}(\omega)$ can be identified up to a constant permutation and a frequency dependent scaling factor of it's columns. Note that (9) can be considered as a special case of (13).

We now show that under the additional assumption $A5$, $\boldsymbol{\mathcal{D}}(\omega_k)$ in equation (13) is constant for all frequency bins.

Theorem 2: Let $\boldsymbol{\mathcal{H}}(\omega) = \sum_{t=0}^{L} \mathbf{H}(t)e^{-j\omega t} \in \mathbb{C}^{J\times N}$ be the transfer function of a MIMO FIR channel of order L. Similarly let $\boldsymbol{\mathcal{B}}(\omega) \in \mathbb{C}^{J\times N}$ be the transfer function of a MIMO FIR channel of unknown order. Assume that $\boldsymbol{\mathcal{B}}(\omega)$ and $\boldsymbol{\mathcal{H}}(\omega)$, evaluated at K uniformly spaced samples, satisfy

$$\boldsymbol{\mathcal{B}}(\omega_k) = \boldsymbol{\mathcal{H}}(\omega_k)\boldsymbol{\Pi}\boldsymbol{\mathcal{D}}(\omega_k), \quad \omega_k = \frac{2\pi k}{K}, \quad k = 0, \ldots, K-1 \tag{14}$$

for some permutation matrix $\boldsymbol{\Pi} \in \mathbb{R}^{N\times N}$ and non-singular diagonal matrices $\boldsymbol{\mathcal{D}}(\omega_k) \in \mathbb{C}^{N\times N}$. If $K \geq 2L+1$ and $\boldsymbol{\mathcal{B}}(z)$ and $\boldsymbol{\mathcal{H}}(z)$, the corresponding z-transforms of $\boldsymbol{\mathcal{B}}(\omega)$ and $\boldsymbol{\mathcal{H}}(\omega)$, are column-wise coprime then (14) implies

$$\mathbf{B}(t) = \mathbf{H}(t)\boldsymbol{\Pi}\mathbf{D}, \quad t = 0, \ldots, L \tag{15}$$

for some non-singular diagonal matrix $\mathbf{D} \in \mathbb{R}^{N\times N}$.

Proof: See Appendix B.

An important implication of Theorem 2 is that under the further assumption that the columns of $\boldsymbol{\mathcal{H}}(z)$ are coprime we can remove the frequency dependent scaling ambiguity $\boldsymbol{\mathcal{D}}(\omega_k)$ given in equation (13). Another important result is that the number of frequency bins required for identification need only be at least $2L+1$. This number is significantly less than what was used in previous frequency domain approaches [11][9]; hence, significant computational savings can be realized with the proposed identification procedure.

The proof of the theorem is motivated by the fact that all elements of the ith column of the estimated channel $\boldsymbol{b}_i(\omega_k)$ (from Theorem I) are subject to the same frequency dependent scale ambiguity, due to the fact they are common with the ith source. Thus, $\boldsymbol{h_i}(\omega_k) = \boldsymbol{b_i}(\omega_k)d_i(\omega_k)$, where $d_i(\omega_k)$ is the ith diagonal element of $\boldsymbol{\mathcal{D}}(\omega_k)$. Therefore we have $\boldsymbol{h_i}(z) = \boldsymbol{b_i}(z)d_i(z)$, and we see that the the presence of a frequency dependent scaling ambiguity introduces common zeros amongst the elements of the columns of $\boldsymbol{\mathcal{B}}(z)$. Thus, the presence of a scale ambiguity indicates the channel estimate is not column–wise coprime. This leads to the idea that column–wise coprimeness of the true channel $\boldsymbol{\mathcal{H}}$ is sufficient for identification of the unknown scale parameters $d_i(\omega_k)$.

IV. The Algorithm

In this section we propose a two step algorithm. The first step estimates the channel up to a frequency dependent scaling ambiguity and constant permutation factor. In other words, the first step finds a $\boldsymbol{\mathcal{B}}(\omega_k)$ such that:

$$\boldsymbol{\mathcal{B}}(\omega_k) = \boldsymbol{\mathcal{H}}(\omega_k)\mathbf{\Pi}\boldsymbol{\mathcal{D}}(\omega_k) \tag{16}$$

where $\mathbf{\Pi}$ is a permutation matrix and $\boldsymbol{\mathcal{D}}(\omega_k)$ represents the frequency dependent scaling ambiguity. The second step removes the frequency dependent scaling ambiguity $\boldsymbol{\mathcal{D}}(\omega_k)$ by exploiting the column-wise coprimeness of the channel $\boldsymbol{\mathcal{H}}(z)$.

A. Step I

For the first part of the algorithm we propose to estimate $\boldsymbol{\mathcal{B}}(\omega_k)$ via the following weighted least squares criterion:

$$\min_{\boldsymbol{\mathcal{B}}(\omega_k),\mathbf{\Lambda}(m)} \sum_{k=0}^{K-1}\sum_{m=0}^{M-1} W_k||\hat{\boldsymbol{\mathcal{P}}}_x(\omega_k,m) - \boldsymbol{\mathcal{B}}(\omega_k)\mathbf{\Lambda}(m)\boldsymbol{\mathcal{B}}^{\dagger}(\omega_k)||_F^2 \tag{17}$$

where $\hat{\boldsymbol{\mathcal{P}}}_x(\omega_k,m)$ is a sample estimate of the observed signal cross spectral density matrix at frequency bin ω_k and time epoch m, $\mathbf{\Lambda}(m)$ is a diagonal matrix, representing the unknown cross-spectral density matrix of the sources at epoch m, and $W_k, k = 0,\ldots,K-1$ are positive scalars. The rational for introducing the weight factor W_k into the optimization criterion is to emphasize the contribution of those frequency bins that are known to give a more reliable estimate of the channel. In case where such prior information is not available we set $W_k = 1$ for all $k = 0,\ldots,K-1$.

To minimize (17) we propose an alternating least squares method (ALS). The basic idea behind ALS is that in the optimization process we divide the parameter space into multiple sets. At each iteration of the algorithm we minimize the criterion with respect to one set conditioned on the previously estimated sets of the parameters. The newly estimated set is then used to update the remaining sets. This process continues until convergence is achieved. Notice that the convergence of ALS is guaranteed because at each iteration we either improve or maintain the value of the cost function [24]. Alternating least squares methods have been used for blind source separation of finite alphabet signals in [25] and [26] and *parallel factor analysis* (PARAFAC) in [24]. The advantage of using ALS (rather than gradient based optimization methods) is that it is simple to implement and there are no parameters to adjust. One disadvantage, shared by most non-linear optimization techniques, is that unless it is properly initialized, it can fall into a local minimum. Later on in this section we introduce a procedure for initializing the algorithm to diminish this possibility. The quantity $\boldsymbol{\mathcal{B}}(\omega_k)\mathbf{\Lambda}(m)\boldsymbol{\mathcal{B}}^{\dagger}(\omega_k)$ in (17) can be written as $\sum_{i=1}^{N}\lambda_i(m)\boldsymbol{b}_i(\omega_k)\boldsymbol{b}_i(\omega_k)^{\dagger}$, where $\boldsymbol{b}_i(\omega_k)$ is the i_{th} column of $\boldsymbol{\mathcal{B}}(\omega_k)$.

From this, equation (17) can be written as:

$$\min_{\boldsymbol{g}_i(\omega_k)\in\Omega,\mathbf{d}(m)} \sum_{k=0}^{K-1}\sum_{m=0}^{M-1} W_k||\hat{\boldsymbol{p}}_x(\omega_k,m)-\boldsymbol{\mathcal{G}}(\omega_k)\mathbf{d}(m)||_2^2 \tag{18}$$

where $\hat{\boldsymbol{p}}_x(\omega_k,m)=\text{vec}\{\hat{\boldsymbol{\mathcal{P}}}_x(\omega_k,m)\}$ is a $J^2\times 1$ column vector, $\boldsymbol{g}_i(\omega_k)$ is the i_{th} column of $\boldsymbol{\mathcal{G}}(\omega_k)=[\text{vec}\{\boldsymbol{b}_1(\omega_k)\boldsymbol{b}_1^\dagger(\omega_k)\},.$ which is a $J^2\times N$ tall matrix, and $\mathbf{d}(m)=\text{diag}(\boldsymbol{\Lambda}(m))$ is an $N\times 1$ column vector. Since there is an inherent scaling ambiguity in calculating $\boldsymbol{b}_i(\omega_k)$ form (17), without loss of generality we can assume $||\boldsymbol{b}_i(\omega_k)||_2^2=1$. Also the constraint set $\Omega\subset\mathbb{C}^{J^2\times 1}$ is defined as:

$$\Omega=\{\text{vec}\{\boldsymbol{\Phi}\}|\boldsymbol{\Phi}=\boldsymbol{\nu}\boldsymbol{\nu}^\dagger,\ \boldsymbol{\nu}\in\mathbb{C}^{J\times 1},\ ||\boldsymbol{\nu}||_2^2=1\}. \tag{19}$$

Following the ALS procedure, we can first minimize (18) with respect to $\boldsymbol{g}_i(\omega_k)$ conditioned on $\hat{\mathbf{d}}(m)$, the previously estimated values of $\mathbf{d}(m)$. To do this we form the matrices $\boldsymbol{\mathcal{T}}(\omega_k)=[\hat{\boldsymbol{p}}(\omega_k,0),...,\hat{\boldsymbol{p}}(\omega_k,M-1)]$ and $\mathbf{F}=[\hat{\mathbf{d}}(0),...,\hat{\mathbf{d}}(M-1)]$ and we write equation (17) as:

$$\min_{\boldsymbol{g}_i(\omega_k)\in\Omega} \sum_{k=0}^{K-1} W_k||\boldsymbol{\mathcal{T}}(\omega_k)-\boldsymbol{\mathcal{G}}(\omega_k)\mathbf{F}||_F^2. \tag{20}$$

Note that (20) is a constrained least squares problem. One simple, although approximate, way to minimize (20) is to first find the unconstrained least squares minimizer of (20) by setting

$$\tilde{\boldsymbol{\mathcal{G}}}(\omega_k)=\boldsymbol{\mathcal{T}}(\omega_k)\mathbf{F}^+. \tag{21}$$

We then project each column of $\tilde{\boldsymbol{\mathcal{G}}}(\omega_k)$ onto Ω; i.e.,

$$\hat{\boldsymbol{g}}_i(\omega_k)=\text{proj}_{\,\Omega}[\tilde{\boldsymbol{g}}_i(\omega_k)] \tag{22}$$

where $\tilde{\boldsymbol{g}}_i(\omega_k)$ is the i_{th} column of $\tilde{\boldsymbol{\mathcal{G}}}(\omega_k)$.

We now discuss a convenient method of performing the projection operation. The projection operation can be effected by the following minimization:

$$\min_{\boldsymbol{g}_i(\omega_k)\in\Omega} ||\tilde{\boldsymbol{g}}_i(\omega_k)-\boldsymbol{g}_i(\omega_k)||_2^2. \tag{23}$$

Since $\boldsymbol{g}_i(\omega_k)=\text{vec}\{\boldsymbol{b}_i(\omega_k)\boldsymbol{b}_i^\dagger(\omega_k)\}$, by defining $\boldsymbol{\mathcal{Y}}_i(\omega_k)=\text{mat}\{\tilde{\boldsymbol{g}}_i(\omega_k)\}$ we can write the above equation as:

$$\begin{aligned}
&\min_{||\boldsymbol{b}_i(\omega_k)||_2=1} ||\boldsymbol{\mathcal{Y}}_i(\omega_k)-\boldsymbol{b}_i(\omega_k)\boldsymbol{b}_i^\dagger(\omega_k)||_F^2 = \\
&\min_{||\boldsymbol{b}_i(\omega_k)||_2=1} (\boldsymbol{b}_i^\dagger(\omega_k)\boldsymbol{b}_i(\omega_k))^2+\text{Trace}\{\boldsymbol{\mathcal{Y}}_i^\dagger(\omega_k)\boldsymbol{\mathcal{Y}}_i(\omega_k)\}-2\boldsymbol{b}_i^\dagger(\omega_k)\boldsymbol{\mathcal{Y}}_i(\omega_k)\boldsymbol{b}_i(\omega_k) = \\
&\min_{||\boldsymbol{b}_i(\omega_k)||_2=1} C-2\boldsymbol{b}_i^\dagger(\omega_k)\boldsymbol{\mathcal{Y}}_i(\omega_k)\boldsymbol{b}_i(\omega_k)
\end{aligned} \tag{24}$$

where $C=1+\text{Trace}\{\boldsymbol{\mathcal{Y}}_i^\dagger(\omega_k)\boldsymbol{\mathcal{Y}}_i(\omega_k)\}$ is a constant term. The above minimization can be done easily by choosing $\hat{\boldsymbol{b}}_i(\omega_k)$, the estimated i_{th} column of $\boldsymbol{\mathcal{B}}(\omega_k)$, to be the dominant eigenvector of $\boldsymbol{\mathcal{Y}}_i(\omega_k)$. To find the dominant eigenvector of a matrix we can use the power iteration method described in [27]. Since an initial estimate of $\boldsymbol{b}_i(\omega_k)$ is available (as is explained later), $\boldsymbol{\mathcal{Y}}_i(\omega_k)$ is nearly a rank one matrix. Hence, the ratio of the largest eigenvalue of $\boldsymbol{\mathcal{Y}}_i(\omega_k)$ to the second-largest (this ratio determines the convergence of the power method), is large. Hence, we need to apply only few iterations of the power method to minimize (24) [4].

[4]In our simulations we use only one power iteration per ALS iteration. Increasing the number of iterations beyond one did not noticeably improve the convergence nor the performance of the algorithm.

To minimize (17) with respect to $\mathbf{d}(m)$ conditioned on the previous estimate of $\boldsymbol{\mathcal{B}}(\omega_k)$ we concatenate the vectors $\hat{\boldsymbol{p}}(\omega_k, m)$ and the matrices $\hat{\boldsymbol{\mathcal{G}}}(\omega_k) = [\text{vec}\{\hat{\boldsymbol{b}}_1(\omega_k)\hat{\boldsymbol{b}}_1^{\dagger}(\omega_k)\}, \ldots, \text{vec}\{\hat{\boldsymbol{b}}_N(\omega_k)\hat{\boldsymbol{b}}_N^{\dagger}(\omega_k)\}]$ for all values of $k = 0, ..., K-1$. For each m we have:

$$\min_{\mathbf{d}(m)} \left\| \begin{bmatrix} \sqrt{W_0}\hat{\boldsymbol{p}}(\omega_0, m) \\ \cdot \\ \cdot \\ \sqrt{W_{K-1}}\hat{\boldsymbol{p}}(\omega_{K-1}, m) \end{bmatrix} - \begin{bmatrix} \sqrt{W_0}\hat{\boldsymbol{\mathcal{G}}}(\omega_0) \\ \cdot \\ \cdot \\ \sqrt{W_{K-1}}\hat{\boldsymbol{\mathcal{G}}}(\omega_{K-1}) \end{bmatrix} \mathbf{d}(m) \right\|_2^2 \tag{25}$$

Minimizing (25) with respect to $\mathbf{d}(m)$ we get:

$$\hat{\mathbf{d}}(m) = \begin{bmatrix} \sqrt{W_0}\hat{\boldsymbol{\mathcal{G}}}(\omega_0) \\ \cdot \\ \cdot \\ \sqrt{W_{K-1}}\hat{\boldsymbol{\mathcal{G}}}(\omega_{K-1}) \end{bmatrix}^{+} \begin{bmatrix} \sqrt{W_0}\hat{\boldsymbol{p}}(\omega_0, m) \\ \cdot \\ \cdot \\ \sqrt{W_{K-1}}\hat{\boldsymbol{p}}(\omega_{K-1}, m) \end{bmatrix} \quad m = 0, ..., M-1. \tag{26}$$

Using equation (21), (22) and (26) we can repeatedly update the values of $\mathbf{d}(m)$ and $\boldsymbol{\mathcal{G}}(m)$ until convergence is achieved.

As mentioned previously, to avoid being trapped in local minima, we need to properly initialize the algorithm. One simple way of doing this is to use the following simple, closed form algorithm for joint diagonalization of two matrices [28]:

B. Initialization

To initialize the algorithm we can select two matrices $\hat{\boldsymbol{\mathcal{P}}}(\omega_k, m_1)$, $\hat{\boldsymbol{\mathcal{P}}}(\omega_k, m_2)$, $m_1 \neq m_2$. We then choose the initial estimate for $\hat{\boldsymbol{\mathcal{B}}}(\omega_k)$ to be the matrix consisting of the N dominant generalized eigenvectors of the matrix couple $(\hat{\boldsymbol{\mathcal{P}}}(\omega_k, m_1), \hat{\boldsymbol{\mathcal{P}}}(\omega_k, m_2))$. Although no optimum selection for m_1 and m_2 can be given at this stage, in our simulations we choose $\hat{\boldsymbol{\mathcal{P}}}(\omega_k, m_1)$ and $\hat{\boldsymbol{\mathcal{P}}}(\omega_k, m_2)$ such that their non-zero generalized eigenvalues are not all repeated [28].

C. Step II

Step II removes the frequency dependent scaling ambiguity by exploiting A5 via Theorem 2. Let the frequency–domain quantity $\boldsymbol{b}_i(\omega_k)$ denote the i_{th} column of $\boldsymbol{\mathcal{B}}(\omega_k)$ obtained in Step I of the algorithm. Without loss of generality we can assume the permutation matrix $\boldsymbol{\Pi}$ in (16) is an identity matrix and we can write:

$$\boldsymbol{b}_i(\omega_k) = \boldsymbol{h}_i(\omega_k) d_{ii}(\omega_k) \tag{27}$$

where $\boldsymbol{h}_i(\omega_k)$ is the i_{th} column of $\boldsymbol{\mathcal{H}}(\omega_k)$ and $d_{ii}(\omega_k)$ is the i_{th} diagonal element of $\boldsymbol{\mathcal{D}}(\omega_k)$. Assumption A5 states the elements of $\boldsymbol{h}_i(z)$ are coprime; i.e., they do not share common zeros. In the time domain this corresponds to the matrix

$$\mathcal{S}_Q(\mathbf{h}_i) = \begin{pmatrix} \mathbf{h}_i(0) & . & . & . & \mathbf{h}_i(L_i) & \mathbf{0} & . & . & \mathbf{0} \\ & \cdot & & & & \cdot & & & \\ & & . & & & & . & & \\ \mathbf{0} & . & . & \mathbf{0} & \mathbf{h}_i(0) & . & . & . & \mathbf{h}_i(L_i) \end{pmatrix} \in \mathbb{R}^{J(Q+1)\times(L_i+Q+1)} \tag{28}$$

having full column rank for $Q \geq L_i - 1$ where L_i is the maximum order of the elements of $\mathbf{h}_i(t)$ [29],[30]. To remove the frequency dependent scaling ambiguities, $d_{ii}(\omega_k)$, we use the following steps. In the time domain (27) can be expressed as the circular convolution of $d_{ii}(t)$, the K-point IDFT of $d_{ii}(\omega_k)$, with $\mathbf{h}_i(t)$, the K-point IDFT of $\boldsymbol{h}_i(\omega_k)$. Assuming that $K > L_i$, (27) can therefore be written as:

$$(\mathbf{b}_i(0), ..., \mathbf{b}_i(K-1)) = (\mathbf{h}_i(0), ..., \mathbf{h}_i(L_i), \mathbf{0}_{J\times(K-L_i-1)})\mathbf{D}_C^i \tag{29}$$

where

$$\mathbf{D}_C^i = \begin{pmatrix} d_i(0) & d_i(1) & . & . & d_i(K-1) \\ d_i(K-1) & d_i(0) & . & . & d_i(K-2) \\ . & . & . & . & . \\ d_i(1) & . & . & d_i(K-1) & d_i(0) \end{pmatrix} \in \mathbb{R}^{K\times K} \tag{30}$$

is a circulant matrix. To remove the scaling ambiguities $d_{ii}(\omega_k)$ we need to find a circulant matrix $\mathbf{\Phi}_C^i$ given as:

$$\mathbf{\Phi}_C^i = \begin{pmatrix} \phi_i(0) & \phi_i(1) & . & . & \phi_i(K-1) \\ \phi_i(K-1) & \phi_i(0) & . & . & \phi_i(K-2) \\ . & . & . & . & . \\ \phi_i(1) & . & . & \phi_i(K-1) & \phi_i(0) \end{pmatrix} \in \mathbb{R}^{K\times K} \tag{31}$$

such that

$$\mathbf{D}_C^i \mathbf{\Phi}_C^i = \alpha \mathbf{I} \tag{32}$$

where α is a constant scalar. Having found such a $\mathbf{\Phi}_C^i$ we can then calculate $\hat{\mathbf{h}}_i(t)$, the estimated i_{th} column of $\mathbf{H}(t)$, by setting

$$\hat{\mathbf{h}}_i(t) = \mathbf{b}_i(t) \circledast \phi_i(t), \quad t = 0, ..., K-1 \tag{33}$$

where $\circledast$ represents the circular convolution operation. Note that in general $\mathbf{D}_C^i$ is unknown so we cannot find $\mathbf{\Phi}_C^i$ from (32). To calculate $\mathbf{\Phi}_C^i$ we exploit the full column rank property of the matrix $\mathcal{S}_Q(\mathbf{h}_i)$ for $Q \geq L_i - 1$. For $K = L_i + Q + 1$, from equations (28) and (29) we can write:

$$\mathbf{B}_C^i = [\mathcal{S}_{Q-1}(\mathbf{h}_i) \quad \mathbf{0}_{JQ\times 1}]\mathbf{D}_C^i \tag{34}$$

where

$$\mathbf{B}_C^i = \begin{pmatrix} \mathbf{b}_i(0) & \mathbf{b}_i(1) & . & . & \mathbf{b}_i(K-1) \\ \mathbf{b}_i(K-1) & \mathbf{b}_i(0) & . & . & \mathbf{b}_i(K-2) \\ . & . & . & . & . \\ \mathbf{b}_i(K-Q+1) & . & . & . & \mathbf{b}_i(K-Q) \end{pmatrix} \in \mathbb{R}^{JQ\times K}. \tag{35}$$

Multiplying both sides of equation (34) by $\mathbf{\Phi}_C^i$ yields:

$$\mathbf{B}_C^i \mathbf{\Phi}_C^i = [\mathcal{S}_{Q-1}(\mathbf{h}_i) \quad \mathbf{0}_{JQ\times 1}]\mathbf{D}_C^i \mathbf{\Phi}_C^i. \tag{36}$$

Define $\hat{\mathbf{H}}_C^i = \mathbf{B}_C^i \mathbf{\Phi}_C^i$. We now show for $Q \geq L_i$, if we find a non-zero $\mathbf{\Phi}_C^i$ that makes the last column of $\hat{\mathbf{H}}_C^i$ equal to zero, then it also satisfies equation (32); i.e., $\hat{\mathbf{H}}_C^i$ becomes a scaled version of $\mathbf{B}_C^i$. Assign $\mathbf{\Lambda}_C^i = \mathbf{D}_C^{i^T} \mathbf{\Phi}_C^i$, which is also a circulant matrix written as

$$\mathbf{\Lambda}_C^i = \begin{pmatrix} \lambda_i(0) & \lambda_i(1) & . & . & \lambda_i(K-1) \\ \lambda_i(K-1) & \lambda_i(0) & . & . & \lambda_i(K-2) \\ . & . & . & . & . \\ \lambda_i(1) & . & . & \lambda_i(K-1) & \lambda_i(0) \end{pmatrix} \in \mathbb{R}^{K\times K}. \tag{37}$$

Then the last column of $\hat{\mathbf{H}}_C^i$ is equal to:

$$\begin{pmatrix} \hat{\mathbf{h}}_i(K-1) \\ . \\ . \\ \hat{\mathbf{h}}_i(K-Q) \end{pmatrix} = \mathcal{S}_{Q-1}(\mathbf{h}_i) \begin{pmatrix} \lambda_i(K-1) \\ . \\ . \\ \lambda_i(1) \end{pmatrix}. \tag{38}$$

For $Q \geq L_i$, $\mathcal{S}_{Q-1}(\mathbf{h}_i)$ has full column rank. Therefore, if the vector on the LHS of (38) is to have all zero elements, then $\lambda(K-1), ..., \lambda(1)$ must also be all zeros; i.e., the matrix $\mathbf{\Lambda}_C^i$ is diagonal with all diagonal elements equal to $\lambda(0)$, and the proof is complete. Following what was said above we need to choose $\mathbf{\Phi}_C^i$ such that the

elements of last column of $\hat{\mathbf{H}}_C^i$ become all zeros. In other words, we find the vector $\boldsymbol{\phi}_i = (\phi_i(K-1), .., \phi_i(0))^T$, the last column of the circulant matrix $\boldsymbol{\Phi}_C^i$, such that

$$\mathbf{B}_C^i \boldsymbol{\phi}_i = \mathbf{0}. \tag{39}$$

To find $\boldsymbol{\phi}_i$ we minimize the quantity $||\mathbf{B}_C^i \boldsymbol{\phi}_i||_2^2$, or equivalently:

$$\min_{||\boldsymbol{\phi}_i||_2=1} \boldsymbol{\phi}_i^T \mathbf{B}_C^{i^T} \mathbf{B}_C^i \boldsymbol{\phi}_i. \tag{40}$$

The solution is given by choosing $\boldsymbol{\phi}_i$ to be the eigenvector of $\mathbf{B}_C^{i^T} \mathbf{B}_C^i$ corresponding to it's minimum eigenvalue. The last step is to compute $\hat{\mathbf{h}}_i(t)$ from (33).

V. Simulation Results

A. Example I, White Sources

For the first simulation, the sources are two independent white Gaussian signals, multiplied by slowly varying sine and cosine signals to create the desired quasi–stationary effect. The purpose of this example is to show that the algorithm is capable of identifying the channel even when the sources are white and Gaussian. Note that none of the previous SOS and HOS methods can identify the channel in this case because they require the sources to be colored in the case of SOS methods and Non-Gaussian in the case of HOS methods. We choose the channel to be a 3×2 system whose impulse response $\mathbf{H}(t)$ is given in Table 1. The epoch size

TABLE I
Impulse Response of the MIMO system for Examples I & II & III

t	0	1	2	3	4	5	6	7
$H_{11}(t)$	-0.528	-0.153	0.631	0.942	-0.221	-0.701	0.274	-0.681
$H_{12}(t)$	0.696	1.952	0.234	-0.938	0.856	1.347	0.341	0.213
$H_{21}(t)$	0.963	-0.927	-0.085	0.322	-0.963	0.049	-0.614	0.000
$H_{22}(t)$	0.675	0.056	-0.143	0.180	1.054	0.230	1.704	0.704
$H_{31}(t)$	0.719	0.538	-1.070	-1.351	0.105	-1.493	0.224	0.144
$H_{32}(t)$	0.774	0.047	-0.147	-0.381	0.287	-0.047	0.649	0.147

is kept constant at 500 and the data length was varied between 10000 and 50000 samples, corresponding to M, the number of epochs, ranging between 20 to 100 epochs. White Gaussian noise was added to the output of the system at a level corresponding to the desired value of averaged SNR over all epochs[5]. At each epoch, 128–point FFTs, applied to time segments overlapping by 50%, weighted by Hanning windows, were used to estimate the cross-spectral density matrices. At each epoch, only $K = 16$ cross-spectral density matrices, evaluated at uniformly spaced frequency samples, were calculated as input to the algorithm. We also choose the W_k in (17) to be:

$$W_k = \frac{1}{\sum_{m=0}^{M-1} ||\boldsymbol{\mathcal{P}}_x(\omega_k, m)||_F^2}. \tag{41}$$

Note that by this choice of W_k we put more emphasis on those frequency bins where average norm of the cross spectral density matrices is small. For white sources this corresponds to those frequency bins where channel parameters have small values and as a result are harder to estimate. Compared to a case when all W_k are set to ones, our simulations show that this choice of W_k improves the overall estimation error. To measure the estimation error, since a scaling ambiguity exists in the final results, we use the following measure for mean–squared error (MSE) based on a method suggested in [31] for evaluating the estimated

[5]The power of the noise was kept constant at all epochs.

impulse responses[6]

$$MSE = 1 - \frac{1}{JNM_c} \sum_{k=1}^{M_c} \sum_{j=1}^{J} \sum_{i=1}^{N} \left(\frac{\mathbf{h}_{ij}^T \hat{\mathbf{h}}_{ij}^k}{\|\mathbf{h}_{ij}\| \|\hat{\mathbf{h}}_{ij}^k\|} \right)^2 \tag{42}$$

where $\mathbf{h}_{ij} = (h_{ij}(0), ..., h_{ij}(L_{ij}))^T$ is the true ij_{th} impulse response of the channel and $\hat{\mathbf{h}}_{ij}^k$ is the estimated response at Monte Carlo run k. The quantity M_c is the total number of Monte Carlo runs.

Table 2 shows the MSE (42) for different SNR's and varying M, using $M_c = 50$ Monte-Carlo runs. As can be seen from Table 2, by increasing the number of epochs, which corresponds to increasing the data length, the MSE decreases. To get a visual impression of the results in Table 2, Figs. 2 and 3 illustrate the corresponding time domain impulse responses and frequency domain responses of the estimated and true channel for SNR=25dB, $M = 20$ epochs, and a data length of 10000 samples. In all of the simulations the algorithm converges in between 7 and 17 iterations.

Further, in Fig. 1, we show the results for varying K. As can be seen from the figure, for $K = 15$ we have a sudden drop in the estimation error. These results match our theoretical bound, derived in Theorem 2, which states that $K \geq 2L + 1$ frequency samples are required for identifiability of $\mathbf{H}(t)$. Since $L = 7$, in theory the number of frequency samples should be $K \geq 14 + 1 = 15$.

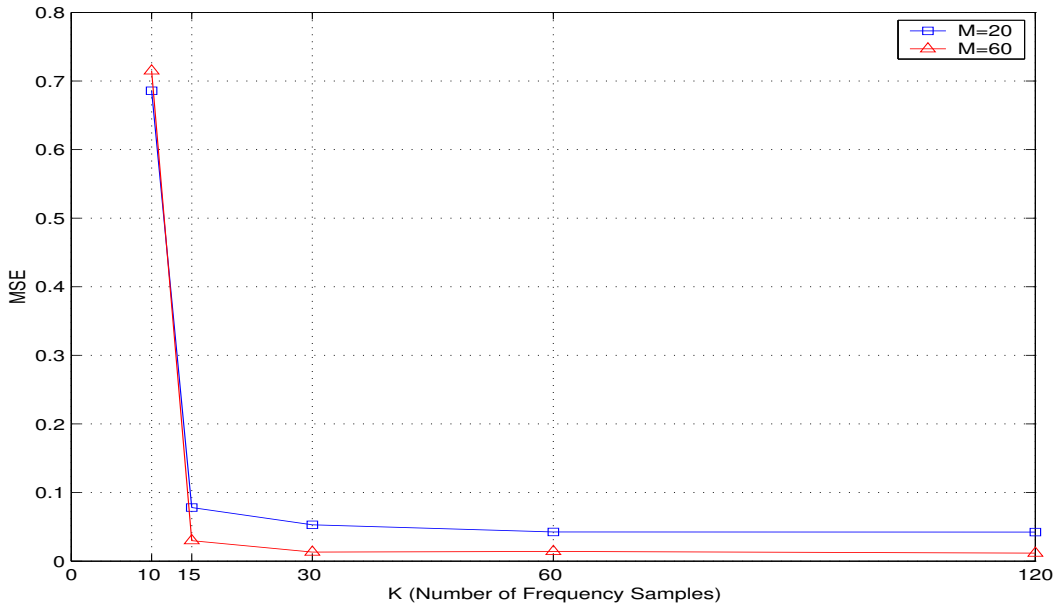

Fig. 1. MSE versus K, number of frequency samples, for M=20 and M=60 and for SNR=10dB using $M_c = 50$ Monte Carlo

TABLE II

EXAMPLE I, SHOWING MSE FOR DIFFERENT SNR'S AND VARYING M USING $M_c = 50$ MONTE CARLO RUNS.

M	20	40	60	100
SNR=30 dB	0.0423	0.0212	0.0148	0.0068
SNR=15 dB	0.0531	0.0282	0.0180	0.0114
SNR=10 dB	0.0714	0.0324	0.0235	0.0153

[6]Also refer to [32] where the author proves that the mathematically correct way of measuring errors when scale ambiguity is present is to use a distance function on the complex projective space. In fact, (42) is one of a number of well known distance functions on the complex projective space.

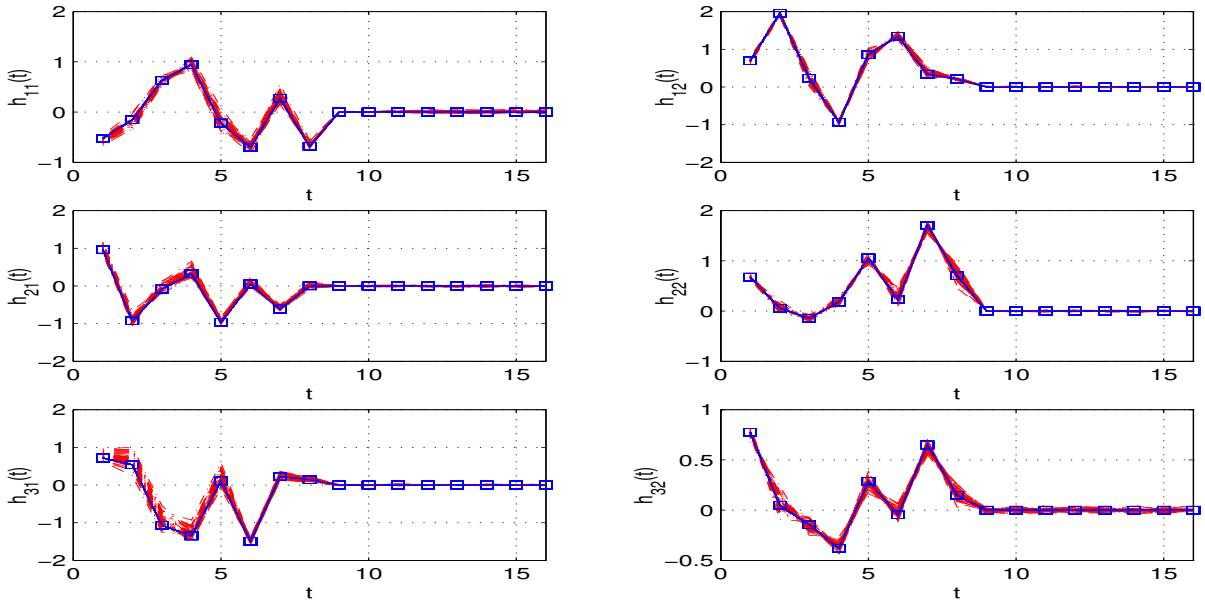

Fig. 2. 50 superimposed independently estimated time-domain responses $\hat{\mathbf{H}}(t)$ shown as dot-dashed lines, along with the true $\mathbf{H}(t)$, shown by the solid lines with squares at the true data points. The horizontal axis is the time index. $M = 60$, SNR=30dB, and $K = 16$.

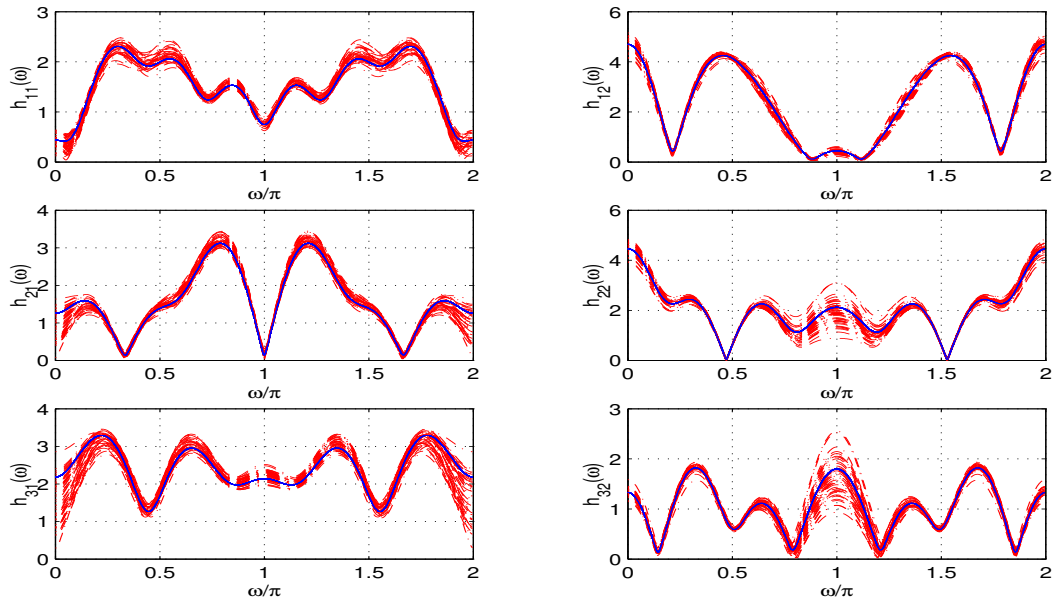

Fig. 3. Same as Fig. 2, except the impulse responses are shown in the frequency domain.

B. Example II, Colored Sources

In this example we show the performance of the algorithm when the sources are colored signals. For this simulation, to create the colored sources we pass the source signals in the previous example through a first order AR filter with the pole equal to 0.8. Notice that contrary to some second order statistics methods which require the sources to have distinct color [15], here the color of the sources can be identical. The sources are mixed through the same channel as example I and we apply the algorithm using the same parameters used in the previous example. The results are shown in Table 3. It can be seen the results are close to what were obtained for the white source case.

TABLE III

EXAMPLE II, SHOWING MSE FOR DIFFERENT SNR'S AND VARYING M USING $M_c = 50$ MONTE CARLO RUNS.

M	20	40	60	100
SNR=30 dB	0.0566	0.0231	0.0154	0.0124
SNR=15 dB	0.0695	0.0298	0.0236	0.0160
SNR=10 dB	0.0983	0.0774	0.0511	0.0358

C. Example III, Speech Signals

The purpose of this next example is to show the performance of the algorithm when the sources are speech signals. Note that for speech signals not only the power but the whole spectrum of the signal changes with time. This violates (10) which we require for colored signals; therefore, we can expect degradation in performance. We use the same channel given in example I; however, for the sources, we use two female speech sequences sampled at 8.0 KHz with a total duration of 2.0 seconds. White Gaussian noise is added to the output of the system commensurate with the specified value of SNR. In a manner similar to Example I, we use 128–point FFTs, with the CPSD matrices being computed at only 16 FFT points for each epoch. Table 4 shows the computed MSE for $M = 50$ epochs for varying values of SNR. As can be seen from the results, the performance has been degraded somewhat compared to the previous example, especially at low values of SNR. However, for high signal to noise ratios it can be observed that the channel can be identified to within a reasonable error.

TABLE IV

EXAMPLE II, SHOWING THE ESTIMATION MSE FOR VARYING SNR, FOR $M_c = 50$ MONTE CARLO RUNS IN THE CASE OF SPEECH SOURCES, WITH $M = 50$.

SNR (dB)	30	15	10
	0.1291	0.4478	0.5195

D. Example IV

In this example we compare the performance of our method with other existing MIMO channel identification approaches. Since most of the current methods for MIMO blind identification assume stationary sources, a direct comparison with our method, which explicitly exploits the non-stationarity of the sources, is not possible. The closest method to the proposed approach is the one recently proposed by Chen and Petropulu in [11]. Their method is also a frequency domain approach with the major difference that they use higher order statistics of the observed signal to identify the channel. Also their method can only be applied to white non-Gaussian signals while the proposed method can be directly extended to colored signals under assumption (10). To compare our results we use the same channel, data length, signal to noise ratio and number of FFT points used in example (1) of [11]. The only exceptions are the sources; in [11] the sources are non-Gaussian stationary signals while for the proposed method we use non-stationary white Gaussian sources. Also to measure the estimation error we use the same performance measure given by equation (53) in [11]. The comparative results are shown in Fig. (4), where we have used the mean square error data in Table 1 of [11] to compare with our results. As can be seen from the figure, the performance of the two methods are very close. For shorter data lengths the method in [11] has a slightly better performance over the proposed method while for a higher number of data samples (more epochs), specially at low SNR, the proposed method has the advantage over the method in [11].

VI. CONCLUSIONS

This paper derived sufficient conditions for a MIMO system, driven by white quasi-stationary sources, to be identified in the frequency domain using only second order statistics of the observed signals. We also showed that the same results can be directly extended to quasi stationary colored sources when only the power of the signal is slowly varying with time. We also proposed a two stage algorithm. The first stage estimates

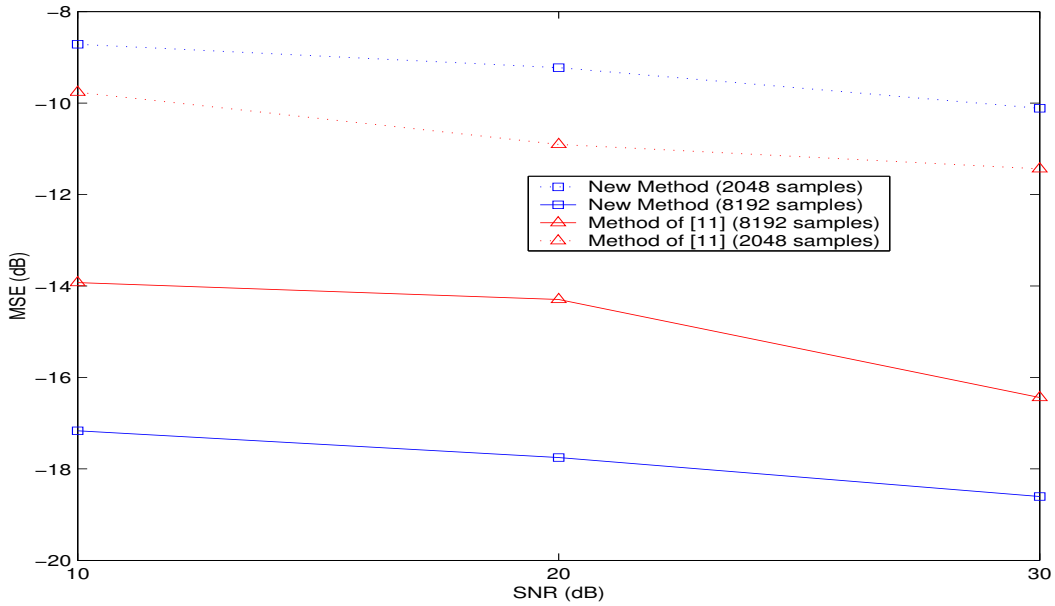

Fig. 4. Comparison of the new proposed algorithm with the method in [11] .

the channel parameters up to a constant permutation and frequency dependent scaling factor, based on a alternating least squares method, while the second stage removes the frequency dependent scaling ambiguity using a closed form algorithm. The results of applying the new algorithm to white and colored sources under the stated assumptions verifies the identifiability conditions, as well as the performance of the algorithm. Application of the proposed algorithm to speech signals was also demonstrated.

VII. Acknowledgements

The first two authors wish to acknowledge the support of following institutions: Mitel Corporation, Canada; The Centre for Information Technology Ontario (CITO); and the Natural Sciences and Engineering Research Council of Canada (NSERC). The last author acknowledges the support of the Australian Research Council and the Special Research Center for Ultra-Broadband Information Networks.

Appendix

I. Proof of Theorem 1

It must be shown that

$$\boldsymbol{\mathcal{B}}(\omega_k)\tilde{\boldsymbol{\Lambda}}(m)\boldsymbol{\mathcal{B}}^{\dagger}(\omega_k) = \boldsymbol{\mathcal{H}}(\omega_k)\boldsymbol{\Lambda}(m)\boldsymbol{\mathcal{H}}^{\dagger}(\omega_k) \tag{43}$$

implies (9). For any sequence of scalars $a = (a_0, ..., a_{M-1})$, define the diagonal matrices

$$\boldsymbol{\Sigma}_a = \sum_{m=0}^{M-1} a_m \boldsymbol{\Lambda}(m), \quad \tilde{\boldsymbol{\Sigma}}_a = \sum_{m=0}^{M-1} a_m \tilde{\boldsymbol{\Lambda}}(m), \quad a_m \in \mathbb{R}. \tag{44}$$

Therefore, an arbitrary linear combination of (43) over different epochs can be written as

$$\boldsymbol{\mathcal{B}}(\omega_k)\tilde{\boldsymbol{\Sigma}}_a\boldsymbol{\mathcal{B}}^{\dagger}(\omega_k) = \boldsymbol{\mathcal{H}}(\omega_k)\boldsymbol{\Sigma}_a\boldsymbol{\mathcal{H}}^{\dagger}(\omega_k). \tag{45}$$

Since the vectors $\text{diag}\{\boldsymbol{\Lambda}(m)\}$ $m = 0, \ldots, M-1$ span $\mathbb{R}^N$, $\boldsymbol{\Sigma}_a$ can be made equal to any real valued diagonal matrix by an appropriate choice of a. In this instance, choose a such that $\boldsymbol{\Sigma}_a$ is the identity matrix. Then since by assumption $\boldsymbol{\mathcal{H}}(\omega_k)$ has full column rank for all $k = 0, ..., K-1$, the RHS of (45) has rank N implying $\boldsymbol{\mathcal{B}}(\omega_k)$ has full column rank for all k. In particular, $\boldsymbol{\mathcal{B}}^{+}(\omega_k)\boldsymbol{\mathcal{B}}(\omega_k)$ is the identity matrix. Thus $\boldsymbol{\mathcal{B}}(\omega_k)$ can be cancelled from the LHS of (45), giving

$$\tilde{\boldsymbol{\Sigma}}_a = \boldsymbol{\mathcal{C}}_k\boldsymbol{\Sigma}_a\boldsymbol{\mathcal{C}}_k^{\dagger}, \quad \boldsymbol{\mathcal{C}}_k = \boldsymbol{\mathcal{B}}^{+}(\omega_k)\boldsymbol{\mathcal{H}}(\omega_k). \tag{46}$$

Observe that if $\boldsymbol{\Sigma}_a$ is the identity matrix then (45) implies $\tilde{\boldsymbol{\Sigma}}_a$ has full rank and thus (46) implies $\boldsymbol{\mathcal{C}}_k$ is invertible for all k.

It is first shown that

$$\boldsymbol{\mathcal{C}}_k^{-1} = \boldsymbol{\Pi}\mathbf{D}e^{-j\boldsymbol{\mathcal{S}}_k} \tag{47}$$

where $\boldsymbol{\Pi}$ is a permutation matrix and $\mathbf{D}$ and $\boldsymbol{\mathcal{S}}_k$ are diagonal matrices. For any i, choose a such that all elements of $\boldsymbol{\Sigma}_a$ are zero except for the i_{th} diagonal element which is unity. Then $\boldsymbol{\mathcal{C}}_k\boldsymbol{\Sigma}_a\boldsymbol{\mathcal{C}}_k^\dagger = \boldsymbol{c}_i(k)\boldsymbol{c}_i(k)^\dagger$ where $\boldsymbol{c}_i(k)$ is the i_{th} diagonal element of $\boldsymbol{\mathcal{C}}_k$. Moreover, since the LHS of (46) is diagonal, all the off-diagonal elements of $\boldsymbol{c}_i(k)\boldsymbol{c}_i^\dagger(k)$ are zero. Because $\boldsymbol{c}_i(k)\boldsymbol{c}_i^\dagger(k)$ has rank one at most, it can have at most one non-zero diagonal element. It follows immediately that every column of $\boldsymbol{\mathcal{C}}_k$ has precisely one non-zero element, and moreover, because $\boldsymbol{\mathcal{C}}_k$ is invertible, every row has precisely one non-zero element too. Clearly the same is true for $\boldsymbol{\mathcal{C}}_k^{-1}$; i.e.,

$$\boldsymbol{\mathcal{C}}_k^{-1} = \boldsymbol{\Pi}_k\boldsymbol{\mathcal{D}}_k \tag{48}$$

where $\boldsymbol{\mathcal{D}}_k$ are non-singular diagonal matrices for all k. Because the LHS of (46) is independent of k, $\boldsymbol{\mathcal{C}}_k\boldsymbol{\mathcal{C}}_k^\dagger$ and thus $\boldsymbol{\mathcal{D}}_k\boldsymbol{\mathcal{D}}_k^\dagger$ is independent from k too; this means only the phase and not the magnitude of the elements of $\boldsymbol{\mathcal{D}}_k$ change with k. Therefore $\boldsymbol{\Pi}$ in (48) must be independent of k and

$$\boldsymbol{\mathcal{D}}_k = \mathbf{D}e^{-j\boldsymbol{\mathcal{S}}_k}. \tag{49}$$

Therefore (47) is proved.

We now substitute $\boldsymbol{\mathcal{C}}_k = \boldsymbol{\mathcal{B}}^+(\omega_k)\boldsymbol{\mathcal{H}}(\omega_k)$ into (48). Rearranging gives

$$\boldsymbol{\mathcal{B}}(\omega_k) = \boldsymbol{\mathcal{B}}(\omega_k)\boldsymbol{\mathcal{B}}^+(\omega_k)\boldsymbol{\mathcal{H}}(\omega_k)\boldsymbol{\Pi}\mathbf{D}e^{-j\boldsymbol{\mathcal{S}}_k}. \tag{50}$$

Note that $\boldsymbol{\mathcal{B}}(\omega_k)\boldsymbol{\mathcal{B}}^+(\omega_k)$ is a projector matrix onto the range space of $\boldsymbol{\mathcal{B}}(\omega_k)$. Choosing $\boldsymbol{\Sigma}_a$ to be the identity matrix in (45) reveals that the range space of $\boldsymbol{\mathcal{B}}(\omega_k)$ must contain the range space of $\boldsymbol{\mathcal{H}}(\omega_k)$. Therefore,

$$\boldsymbol{\mathcal{B}}(\omega_k)\boldsymbol{\mathcal{B}}^+(\omega_k)\boldsymbol{\mathcal{H}}(\omega_k) = \boldsymbol{\mathcal{H}}(\omega_k) \tag{51}$$

and (9) follows immediately from (51) and (50). □

II. Proof of Theorem 2

Let $\boldsymbol{b}(\omega_k)$ be an arbitrary column of $\boldsymbol{\mathcal{B}}(\omega_k)$ and let $\boldsymbol{h}(\omega_k)$ be the corresponding column of $\boldsymbol{\mathcal{H}}(\omega)\boldsymbol{\Pi}$. It assumed that elements of $\boldsymbol{b}(z)$, the corresponding z-transform of $\boldsymbol{b}(\omega)$, are coprime, as are the elements of $\boldsymbol{h}(z)$, the corresponding z-transform of $\boldsymbol{h}(\omega)$. It will be proved that

$$\boldsymbol{b}(\omega_k) = d_k\boldsymbol{h}(\omega_k), \quad \omega_k = \frac{2\pi k}{K}, \;\; d_k \in \mathbb{C}, \;\; d_k \neq 0, \;\; k = 0, \ldots, K-1 \tag{52}$$

implies $\mathbf{b}(t) = \alpha\mathbf{h}(t), \;\; t = 0, \ldots, L$, for some non-zero $\alpha \in \mathbb{C}$ provided $K \geq 2L+1$, where L is the order of $\mathbf{h}(z)$. The theorem then follows immediately.

Let $\mathbf{b}(t)$ and $\mathbf{h}(t)$ be the impulse responses of $\boldsymbol{b}(\omega)$ and $\boldsymbol{h}(\omega)$ respectively, so that

$$\boldsymbol{b}(\omega) = \sum_{t=0}^{\tilde{L}} \mathbf{b}(t)e^{-j\omega t}, \qquad \boldsymbol{h}(\omega) = \sum_{t=0}^{L} \mathbf{h}(t)e^{-j\omega t} \tag{53}$$

where the order $\tilde{L}$ is unknown but finite. The proof below repeatedly uses the fact that, for any $Q \geq \tilde{L}-1$, a SIMO FIR channel of order $\tilde{L}$ having impulse response $\mathbf{b}(0), ..., \mathbf{b}(\tilde{L}) \in \mathbb{R}^J$ is coprime if and only if the block Sylvester matrix

$$\mathcal{S}_Q(\mathbf{b}) = \begin{pmatrix} \mathbf{b}(0) & \mathbf{b}(1) & \ldots & \mathbf{b}(\tilde{L}) & \mathbf{0} & \ldots & \mathbf{0} \\ \mathbf{0} & \mathbf{b}(0) & \ldots & \mathbf{b}(\tilde{L}-1) & \mathbf{b}(\tilde{L}) & \ldots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ldots & \ddots & \ddots & \vdots \\ \mathbf{0} & \ldots & \mathbf{0} & \mathbf{b}(0) & \mathbf{b}(1) & \ldots & \mathbf{b}(\tilde{L}) \end{pmatrix} \in \mathbb{R}^{J(Q+1)\times(\tilde{L}+Q+1)} \tag{54}$$

has full column rank [29].

It is first proved that $\tilde{L} \leq L$. Assume to the contrary that $\tilde{L} > L$. By substituting (53) into (52), it follows that, for the choice $Q = K - \tilde{L} - 1$,

$$\mathcal{S}_Q(\mathbf{b})\mathbf{F} = [\mathcal{S}_Q(\mathbf{h}) \;\; \mathbf{0}_{J(Q+1)\times(\tilde{L}-L)}]\mathbf{F}\boldsymbol{\Delta} \tag{55}$$

where $\mathcal{S}_Q(\mathbf{h})$ is a $J(Q+1) \times (L+Q+1)$ matrix having the same form as $\mathcal{S}_Q(\mathbf{b})$, $\mathbf{F}$ is the non-singular DFT matrix

$$\mathbf{F} = \begin{pmatrix} 1 & 1 & \dots & 1 \\ 1 & e^{-j2\pi/K} & \dots & e^{-j2\pi(K-1)/K} \\ 1 & e^{-j4\pi/K} & \dots & e^{-j4\pi(K-1)/K} \\ \vdots & \vdots & \dots & \vdots \\ 1 & e^{-j2\pi(K-1)/K} & \dots & e^{-j2\pi(K-1)^2/K} \end{pmatrix} \tag{56}$$

and $\boldsymbol{\Delta} = \mathrm{diag}\{d_0, \dots, d_{K-1}\}$. The LHS of (55) has full column rank because $\boldsymbol{b}(z)$ is coprime and $Q = K - \tilde{L} - 1 \geq L - 1$ by assumption that $K \geq 2\tilde{L} + 1$ and $\tilde{L} > L$. However, the RHS of (55) clearly does not have full column rank, a contradiction.

This time, choose $Q = K - L - 1$. Analogous to (55), but this time because it is known $\tilde{L} \geq L$,

$$[\mathcal{S}_Q(\mathbf{b}) \;\; \mathbf{0}_{J(Q+1)\times(L-\tilde{L})}]\mathbf{F} = \mathcal{S}_Q(\mathbf{h})\mathbf{F}\boldsymbol{\Delta}. \tag{57}$$

Define $\mathbf{C} = \mathbf{F}\boldsymbol{\Delta}\mathbf{F}^{-1}$; since $\mathbf{F}$ is a DFT matrix, $\mathbf{C}$ is circulant:

$$\mathbf{C} = \begin{pmatrix} c(0) & c(1) & \dots & c(K-1) \\ c(K-1) & c(0) & \dots & c(K-2) \\ \vdots & \ddots & \dots & \vdots \\ c(1) & c(2) & \dots & c(0) \end{pmatrix}. \tag{58}$$

Then

$$[\mathcal{S}_Q(\mathbf{b}) \;\; \mathbf{0}_{J(Q+1)\times(L-\tilde{L})}] = \mathcal{S}_Q(\mathbf{h})\mathbf{C}. \tag{59}$$

Even if $\tilde{L} = L$, the first JQ elements of the last column of the LHS of (59) are zero. Therefore,

$$\mathcal{S}_{Q-1}(\mathbf{h})\mathbf{c} = \mathbf{0}, \tag{60}$$

where $\mathbf{c} \in \mathbb{R}^{K-1}$ is the vector formed from the first $K-1$ elements of the last column of $\mathbf{C}$. Because $Q - 1 = K - L - 2 \geq L - 1$ by assumption that $K \geq 2L + 1$, $\mathcal{S}_{Q-1}(\mathbf{h})$ has full column rank, and in particular, $\mathbf{c} = \mathbf{0}$. Since $\mathbf{C}$ is circulant, $\mathbf{c} = \mathbf{0}$ implies $\mathbf{C} = \alpha\mathbf{I}$ for some $\alpha \in \mathbb{R}$. It follows from (59) that $\mathbf{b}(t) = \alpha\mathbf{h}(t)$ for $t = 0, ..., L$. Notice that $\alpha \neq 0$ for otherwise the coprimeness of the elements of $\boldsymbol{b}(z)$ would be contradicted. The theorem then follows. □

References

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

[2] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inform. Theory*, vol. 40, pp. 340–349, Mar. 1994.

[3] G. Giannakis and J. Mendel, "Identification of nonminimum phase systems using higher order statistics," *IEEE Trans. Acoust., Speech, and Signal Processing*, vol. 37, pp. 7360–7377, Mar. 1989.

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

[5] Y. Hua and M. Wax, "Strict identifiability of multiple FIR channels driven by an unkown arbitrary sequence," *IEEE Trans. Signal Processing*, vol. 44, pp. 756–759, Mar. 1996.

[6] J. Cardoso, "Blind Signal Separation: Statistical Principles," *Proc. IEEE*, vol. 86, pp. 2009–2025, Oct. 1998.

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

[8] H. Sahlin and H. Broman, "MIMO signal separation for FIR channels: a criterion and performance analysis," *IEEE Trans. Signal Processing*, vol. 48, pp. 642–649, Mar. 2000.

[9] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech Audio Processing*, vol. 8, pp. 320–327, May. 2000.

[10] J. K. Tugnait, "Adaptive blind separation of convolutive mixtures of independent linear signals," *Signal Processing*, vol. 73, pp. 139–152, 1999.

[11] B. Chen and A. Petropulu, "Frequency Domain Blind MIMO System Identification Based on Second and Higher Order Statistics," *IEEE Trans. Signal Processing*, vol. 49, pp. 1677–1688, Aug. 2001.

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

[13] Y. Hua, S. An, and Y. Xiang, "Blind Identification and Equalization of FIR MIMO Channels By BIDS ," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing*, May 2001.

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

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

[16] E. Serpedin and G. Giannakis, "Blind channel identification and equalization with modulation-induced cyclostationarity," *IEEE Trans. Signal Processing*, vol. 46, pp. 1930–1944, Jul. 1998.

[17] D. T. Pham and J. Cardoso, "Blind source separation of instantaneous mixtures of nonstationary sources," *IEEE Trans. Signal Processing*, vol. 49, pp. 1837–1848, Sept. 2001.

[18] K. Abed_Meraim, Y. Xiang, J. H. Manton, and Y. Hua, "Blind Source Separation Using Second-Order Cyclostationary Statistics," *IEEE Trans. Signal Processing*, vol. 49, pp. 694–701, Apr. 2001.

[19] K. Rahbar and J. Reilly, "Blind source separation of convolved sources by joint approximate diagonalization of cross-spectral density matrices," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing*, May 2001.

[20] K. Rahbar, J. Reilly, and J. Manton, "A Frequency Domain Approach to Blind Identification of MIMO FIR Systems Driven By Quasi-Stationary Signals," in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Processing*, May 2002.

[21] J. Cardoso and A. S. Soulomiac, "Jacobi angles for simultaneous diagonalization," *SIAM Journal on Matrix Analysis and Applications*, vol. 17, pp. 161–164, Jan. 1996.

[22] A. Yeredor, "Approximate joint diagonalization using non-orthogonal matrices," in *IEEE Workshop on Independent Component Analysis*, June 2000.

[23] A. Papoulis, *Probability, Random Variables and Stochastic Processes*. Singapore: McGraw-Hill, 1984.

[24] N. Sidiropoulos, G. Giannakis, and R. Bro, "Blind PARAFAC Receivers for DS-CDMA Systems," *IEEE Trans. Signal Processing*, vol. 48, pp. 810–823, Mar. 2000.

[25] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array-Part I: Algorithms," *IEEE Trans. Signal Processing*, vol. 44, pp. 1184–1197, May 1996.

[26] T. Li and N. Sidiropoulos, "Blind Digital Signal Separation Using Successive Interference Cancellation Iterative Least Squares," *IEEE Trans. Signal Processing*, vol. 48, pp. 3146–3152, Nov. 2000.

[27] G. Golub and C. VanLoan, *Matrix Computations*. Baltimore and London: John Hopkins, third ed., 1996.

[28] C. T. Ma, Z. Ding, and S. F. Yau, "A Two-stage Algorithm for MIMO Blind Deconvolution of Nonstationary Colored Signals," *IEEE Trans. Signal Processing*, vol. 48, pp. 1187–1192, Apr. 2000.

[29] E. Serpedin and G. Giannakis, "A simple proof of a known blind channel identifiability result," *IEEE Trans. Signal Processing*, vol. 47, pp. 591–593, Feb. 1999.

[30] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind Channel Identification based on second order statistics: A frequency domain approach," *IEEE Trans. Inform. Theory*, vol. 41, pp. 329–334, Jan. 1995.

[31] D. R. Morgan, J. Benesty, and M. M. Sondhi, "On the Evaluation of Estimated Impulse Responses," *IEEE Signal Processing Lett.*, vol. 5, pp. 174–176, Jul. 1998.

[32] J. H. Manton, "A Packet Based Channel Identification Algorithm For Wireless Multimedia Communications," in *IEEE International Conference on Multimedia and Expo*, (Tokyo, Japan), August 2001.