A FREQUENCY DOMAIN APPROACH TO BLIND IDENTIFICATION OF MIMO FIR SYSTEMS DRIVEN BY QUASI-STATIONARY SIGNALS

Kamran Rahbar & James P. Reilly

Electrical & Computer Eng.
McMaster University
Hamilton, Ontario, Canada
Email: reillyj@mcmaster.ca

Jonathan H. Manton

Electrical & Electronic Eng.
The University of Melbourne
Victoria 3010, Australia.
Email:jon@ee.mu.oz.au

ABSTRACT

This paper discusses the problem of blind identification of MIMO convolutive channels when the sources are quasi-stationary. No other assumptions are made about the sources; i.e., they can be temporally white or colored and they can assume any distributions. We show that by using second order statistics of the channel outputs, under some mild conditions on the non-stationarity of sources and that the channel is column-wise coprime, identification can be achieved up to a scalar ambiguity and column permutation of the original MIMO channel. We also present an efficient, two step frequency domain algorithm for identifying the channel. To demonstrate the performance of the new algorithm numerical simulations are presented.

1. INTRODUCTION

Blind identification of an N-input J-output MIMO FIR channel deals with identifying the impulse response $\mathbf{H}(n)$ of the unknown channel using only channel output data and without any knowledge about the inputs. Blind identification of MIMO channels is closely related to MIMO blind equalization and convolutive blind source separation problems, mainly because the channel estimate can be used to recover the original inputs or to separate them. In this paper we consider using second order statistics of the observed signals (channel outputs) to estimate the channel impulse response. We make this assumptions about the sources that they are quasi-stationary. More specifically, we require that only the variances of the sources, although constant over some time interval (epoch), change from one epoch to the other. This kind of quasi-stationarity for example can be induced in transmitter for blind identification (or equalization) of communication channels. Some recent methods related to our work are: In [1] a frequency domain, higher order statistics method for MIMO blind identification has been proposed but it only deals with the white source case. Both methods in [2] and [3] are based on second order statistics but they both require the sources to be colored and have different spectra. In [4] the authors propose a method for convolutive blind source separation of non-stationary signals although no identification conditions are given and the method only considers the colored sources. The contributions of this paper are: 1) We provide frequency domain blind identifiability conditions using only second order statistics with a finite number of frequency samples. We also determine the upper bound on the smallest number of frequency samples required for blind identification. 2) We propose a new, two-step algorithm in the frequency domain for MIMO blind identification. Some advantages of the new algorithm are: Unlike some other frequency domain blind identification nor source separation methods, the current algorithm has no frequency dependent permutation nor scaling ambiguity problems. Also, the new method is simple to implement, has no extra parameters to adjust and in most case has fast convergence. We demonstrate the performance of the new algorithm using numerical simulations included at the end of this paper.

2. PROBLEM FORMULATION

2.1. Notations

We use bold upper and lower case letters to show matrices and vectors respectively. The remaining notational conventions are as follows

$(\cdot)^\dagger$	Hermitian.
diag($\mathbf{a}$)	Forms a diagonal matrix from vector $\mathbf{a}$.
diag($\mathbf{A}$)	Forms a column vector from diagonal elements of $\mathbf{A}$.
vec$\{\mathbf{A}\}$	Forms a column vector by stacking the columns of $\mathbf{A}$.
$\mathbf{A}^+$	Pseudo Inverse of matrix $\mathbf{A}$.

2.2. Model

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

$$\mathbf{x}(k) = \sum_{l=0}^{L_{max}} \mathbf{H}(l)\mathbf{s}(k-l) + \mathbf{n}(k) \qquad k \in \mathcal{Z} \tag{1}$$

where $\mathbf{x}(k) = (x_1(k), \cdots, x_J(k))^T$ is the vector of observed signals, $\mathbf{s}(k) = (s_1(k), \cdots, s_N(k))^T$ is the vector of sources, $\mathbf{H}(k)$ is the $J \times N$ channel matrix and $\mathbf{n}(k) = (n_1(k), \cdots, n_J(k))^T$ is the additive noise vector. The objective is to estimate $\mathbf{H}(l)$ up to a scaling and permutation factor from the observed signals $\mathbf{x}(k)$. In other words we are interested to find a $\hat{\mathbf{H}}(l)$ such that for all $0 \leq l \leq L_{max}$ we have

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

where $\mathbf{D}$ and $\mathbf{\Pi}$ are respectively constant diagonal and permutation matrices. In frequency domain this is equivalent to finding a $\hat{\mathbf{H}}(\omega)$ such that:

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

2.3. Assumptions

We make the following assumptions on the model:

A0: $J \geq N$; i.e, we can have more sensors than sources.

A1: The sources $\mathbf{s}(k)$ are real, zero mean, quasi-stationary (we assume that only the variance of sources change with time) and the cross–spectral density matrices of the sources $\mathbf{P}_s(\omega, m)$ are diagonal for all ω and m where ω denotes frequency and m denotes time epoch.

A2: Let $\alpha_i(m)$ denotes the variance of the i_{th} sources at epoch m, we assume the matrix $\mathbf{A}$ given by:

$$\mathbf{A} = \begin{pmatrix} \alpha_1(0) & . & . & . & \alpha_1(M-1) \\ . & & & & . \\ . & & & & . \\ \alpha_N(0) & & & & \alpha_N(M-1) \end{pmatrix} \quad (4)$$

has full row rank.

A3: $\mathbf{H}(k) = [\mathbf{h}_1(k), ..., \mathbf{h}_N(k)]$ is real, causal and it does not change over the entire observation interval. Also $\mathbf{H}(\omega)$, the DTFT of $\mathbf{H}(k)$, has full column rank for all $\omega \in [0, 2\pi)$.

A4: The noise $\mathbf{n}(k)$ is zero mean, *iid* across sensors, with unknown power σ^2 and is independent of the sources.

A5: $\mathbf{H}(z)$, the z-transform of $\mathbf{H}(k)$, is column wise coprime, i.e. the elements in each column of $\mathbf{H}(z)$ do not share common zeros. Also, the degree of each of the columns of $\mathbf{H}(z)$, denoted by $\mathbf{L}_i, \quad i = 1, ..., N,$, is assumed known.

3. BLIND IDENTIFIABILITY

In this section we present frequency domain blind identifiability results based on assumptions A1...A5 and using only second order statistics of the observed signals. Let $\mathbf{P}_x(\omega, m)$ represent the cross-spectral density of the observed signal at epoch m. We have:

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

where $\mathbf{P}_s(\omega, m)$ by assumption is diagonal for all ω and m. For $J > N$, σ^2 can be estimated from the smallest eigenvalue of the matrix $\mathbf{P}_x(\omega, m)$; so for now, we assume the following noise free case

$$\mathbf{P}_x(\omega_k, m) = \mathbf{H}(\omega_k)\mathbf{P}_s(\omega_k, m)\mathbf{H}^\dagger(\omega_k), \quad (6)$$

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

Theorem 1 *Given the finite set of matrices $\mathbf{P}_x(\omega_k, m)$ defined by (6) where $k = 0, .., K-1, \quad m = 0, .., M-1$ and $K > \max(L_i)$ and assuming A0,..,A3 are all satisfied, $\mathbf{H}(\omega)$ can be uniquely identified up to a frequency dependent scaling and constant permutation factor.* □

In other words, we want to prove that for any matrix $\mathbf{B}(\omega_k)$ which satisfies assumptions A0,...,A3 as well as the equation

$$\mathbf{P}_x(\omega_k, m) = \mathbf{B}(\omega_k)\tilde{\mathbf{P}}_s(\omega_k, m)\mathbf{B}^\dagger(\omega_k), \quad (7)$$

where $\tilde{\mathbf{P}}_s(\omega_k, m)$ is diagonal for all ω_k and m, then $\mathbf{B}(\omega)$ is related to $\mathbf{H}(\omega)$ as:

$$\mathbf{B}(\omega_k) = \mathbf{H}(\omega_k)\mathbf{\Pi}\mathbf{D}(\omega_k), \qquad k = 0, .., K-1, \quad (8)$$

where $\mathbf{\Pi}$ is a permutation matrix and $\mathbf{D}(\omega_k)$ is diagonal for each ω_k. It can be further proved that when the source signals are white we have

$$\mathbf{D}(\omega_k) = \mathbf{D}_c e^{-j\mathbf{S}(\omega_k)} \quad (9)$$

where $\mathbf{S}(\omega_k)$ is diagonal for all ω_k and $\mathbf{D}_c$ is constant diagonal matrix.
Proof : Refer to [5].

Theorem 2 *Using the further assumption A5, $\mathbf{H}(n)$ can be identified up to a constant diagonal and permutation factor using $K > 2L_{max} - 1$ frequency samples where $L_{max} = \max\{L_1, ..., L_N\}$.*

This means that any matrix $\mathbf{B}(n)$ which satisfies A5 and whose DTFT satisfies equation (8) for all $k = 0, .., K-1$ where $K > 2L_{max} - 1$ and also whose columns have the same degree as the columns of $\mathbf{H}(n)$ then it must satisfies

$$\mathbf{B}(n) = \mathbf{H}(n)\mathbf{\Pi}\mathbf{D}. \quad (10)$$

Proof: Refer to [5]. Notice that in practice to estimate the cross power spectral density matrices (CPSD) using an FFT method, the number of FFT points should be greater than roughly ten times the length of the channel. This makes the total number of estimated CPSDs at each epoch to be at least $10L_{max}$. Based on Theorem 2 we only need to use more than $2L_{max}$ of these CPSDs to estimate the channel. This implies a great reduction in the total computational cost.

4. ALGORITHM

In this section we propose a two steps algorithm. The first step estimates $\mathbf{H}(\omega)$ up to a frequency dependent scaling ambiguity and constant permutation factor as given by (8), while the second step removes the scaling ambiguity by putting a constraint on the degrees of the columns of $\mathbf{H}(\omega)$ in the time domain.

4.1. Step I

For the first part of the algorithm we propose the following least squares criterion:

$$\min_{\hat{\mathbf{H}}(\omega_k), \mathbf{\Lambda}(m)} \sum_{k=0}^{K-1} \sum_{m=0}^{M-1} ||\hat{\mathbf{P}}_x(\omega_k, m) - \hat{\mathbf{H}}(\omega_k)\mathbf{\Lambda}(m)\hat{\mathbf{H}}^\dagger(\omega_k)||_F^2 \quad (11)$$

where $\mathbf{\Lambda}(m)$ is diagonal and constant over all ω_k. It can be verified easily that this selection for $\mathbf{\Lambda}(m)$ prevents from arbitrary permutation across frequency bins. This selection also justified by noticing that when the sources are white, at each epoch their CPSDs are constant over all frequency range, For non-white sources, as long as only the power of source change with time, we can write $\mathbf{P}_s(\omega_k, m) = \mathbf{D}_s^{\frac{1}{2}}(\omega_k)\mathbf{\Lambda}(m)\mathbf{D}_s^{\frac{1}{2}}(\omega_k)$ and without loss of generality we can consider $\mathbf{D}_s^{\frac{1}{2}}(\omega_k)$ as part of frequency dependent scaling ambiguity of the columns of $\mathbf{H}(\omega_k)$. As a result (11) can be used both for white and non-white sources. To minimize (11) we use 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 and at each iteration 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 and this process continues until convergence is achieved. The advantage of using ALS (rather than

other gradient based optimization methods) is that it is simple to implement and there are no parameters to adjust. The main disadvantage is unless it is properly initialized it can fall into a local minima. Later on in this section we will introduce a procedure for initializing the algorithm. Equation (11) can be written as:

$$\min_{\mathbf{G}(\omega_k),\mathbf{d}(m)} \sum_{k=0}^{K-1}\sum_{m=0}^{M-1} ||\mathbf{p}_x(\omega_k,m) - \mathbf{G}(\omega_k)\mathbf{d}(m)||_2^2 \qquad (12)$$

where $\mathbf{p}_x(\omega_k,m) = \text{vec}\{\mathbf{P}_x(\omega_k,m)\}$ is a $J^2 \times 1$ column vector, $\mathbf{G}(\omega_k) = [\text{vec}\{\mathbf{h}_1(\omega_k)\mathbf{h}_1^\dagger(\omega_k)\}, ..., \text{vec}\{\mathbf{h}_N(\omega_k)\mathbf{h}_N^\dagger(\omega_k)\}]$ is a $J^2 \times N$ tall matrix, $\mathbf{h}_i(\omega_k)$ is the i_{th} column of $\mathbf{H}(\omega_k)$ and $\mathbf{d}(m) = diag(\mathbf{\Lambda}(m))$ is an $N \times 1$ column vector. Following what is said above, we can first minimize (12) with respect to $\mathbf{G}(\omega_k)$ conditioned on the previously estimated values of $\mathbf{d}(m)$. To do this we form the matrices $\mathbf{T}(\omega_k) = [\mathbf{p}(\omega_k,0), ..., \mathbf{p}(\omega_k, M-1)]$ and $\hat{\mathbf{F}} = [\hat{\mathbf{d}}(0), ..., \hat{\mathbf{d}}(M-1)]$ and we write equation (11) as:

$$\min_{\mathbf{G}(\omega_k)} \sum_{k=0}^{K-1} ||\mathbf{T}(\omega_k) - \mathbf{G}(\omega_k)\hat{\mathbf{F}}||_F^2. \qquad (13)$$

In order to recover $\mathbf{h}_i(\omega_k)$, it is necessary to constrain the the columns of $\mathbf{G}(\omega_k)$ so that if we form a matrix from each column, the resulting matrix will have rank 1. To minimize (12) with respect to $\mathbf{G}(\omega_k)$ we first calculate:

$$\tilde{\mathbf{G}}(\omega_k) = \mathbf{T}(\omega_k)\hat{\mathbf{F}}^+, \qquad (14)$$

and then we project each column of $\tilde{\mathbf{G}}(\omega_k)$ into space of vectors that have rank 1 matrix forms. This corresponds to setting each column of $\hat{\mathbf{G}}(\omega_k)$, the constrained minimizer of (12), to:

$$\hat{\mathbf{g}}_i(\omega_k) = \text{vec}\{\mathbf{v}_{max}^i \mathbf{v}_{max}^{i\dagger}\}, \qquad (15)$$

where $\mathbf{v}_{max}^i(\omega_k)$ is the dominant eigenvector of the matrix formed from $\tilde{\mathbf{g}}_i(\omega_k)$, the i_{th} column of $\tilde{\mathbf{G}}(\omega_k)$. Therefore the estimated columns of $\hat{\mathbf{H}}(\omega_k)$ are given as:

$$\hat{\mathbf{h}}_i(\omega_k) = \mathbf{v}_{max}^i(\omega_k). \qquad (16)$$

To minimize (11) with respect to $\mathbf{d}(m)$ conditioned on the previous estimate of $\mathbf{H}(\omega_k)$ we concatenate the vectors $\mathbf{p}(\omega_k,m)$ and the matrices $\mathbf{G}(\omega_k)$ for all values of $k = 0, ..., K-1$. For each m we have:

$$\min_{\mathbf{d}(m)} \left\| \begin{bmatrix} \mathbf{p}(\omega_0,m) \\ \cdot \\ \cdot \\ \mathbf{p}(\omega_{K-1},m) \end{bmatrix} - \begin{bmatrix} \mathbf{G}(\omega_0) \\ \cdot \\ \cdot \\ \mathbf{G}(\omega_{K-1}) \end{bmatrix} \mathbf{d}(m) \right\|_2^2 \qquad (17)$$

Minimizing (17) with respect to $\mathbf{d}(m)$ we get:

$$\hat{\mathbf{d}}(m) = \begin{bmatrix} \hat{\mathbf{G}}(\omega_0) \\ \cdot \\ \cdot \\ \hat{\mathbf{G}}(\omega_{K-1}) \end{bmatrix}^+ \begin{bmatrix} \mathbf{p}(\omega_0,m) \\ \cdot \\ \cdot \\ \mathbf{p}(\omega_{K-1},m) \end{bmatrix}. \qquad (18)$$

Using equation (14), (16) and (18) we can repeatedly update the values of $\mathbf{d}(m)$ and $\mathbf{G}(m)$ until convergence is achieved. As mentioned before to avoid the local minima we need to properly initialize the algorithm. One simple but not the optimum way of doing this is to use the following exact joint diagonalization algorithm.

4.2. Initialization

To initialize the algorithm we can select two matrices $\mathbf{P}(\omega_k, m_1)$, $\mathbf{P}(\omega_k, m_2)$ and form the matrix

$$\begin{aligned} \mathbf{M} &= \mathbf{P}(\omega_k,m_1)\mathbf{P}^+(\omega_k,m_2) \\ &= \mathbf{H}(\omega_k)\mathbf{\Lambda}(m_1)\mathbf{\Lambda}^{-1}(m_2)\mathbf{H}^+(\omega_k) \end{aligned} \qquad (19)$$

We then set the matrix $\mathbf{H}_{init}(\omega_k) = \mathbf{U}$ where $\mathbf{U}$ is a $J \times N$ matrix whose columns are the N dominant generalized eigenvectors of the matrix pencil $\mathbf{P}(\omega_k, m_1)$ and $\mathbf{P}(\omega_k, m_2)$.

4.3. Step II

The second step of the algorithm removes the frequency dependent scaling ambiguity by constraining the degrees of each column of $\hat{\mathbf{H}}(\omega_k)$ in the time domain. Let $\hat{\mathbf{h}}_i(n)$ represent the i_{th} column of $\mathbf{H}(n)$, the time domain counterpart of $\hat{\mathbf{H}}(\omega_k)$. We have:

$$\hat{\mathbf{h}}(n) = \frac{1}{K}\sum_{k=0}^{K-1} \hat{\mathbf{h}}_i(\omega_k) e^{j\omega_k n}. \qquad (20)$$

Based on Theorem 2, we can adjust the scaling of the $\hat{\mathbf{h}}_i(\omega_k)$ such that its time domain version $\hat{\mathbf{h}}(n)$ will have the correct length. To do this let's assume:

$$\tilde{\mathbf{h}}_i(n) = \frac{1}{K}\sum_{k=0}^{K-1} \hat{\mathbf{h}}_i(\omega_k)\alpha_i(\omega_k) e^{j\omega_k n}. \qquad (21)$$

We adjust $\alpha(\omega_k)$ to maximize $\sum_{n=0}^{L_i} \|\tilde{\mathbf{h}}_i(n)\|_2^2$ under the constraint that the total energy $\sum_{n=0}^{K-1} \|\tilde{\mathbf{h}}_i(n)\|_2^2$ stays constant. This way we are actually minimizing the energy under the tail of the $\tilde{\mathbf{h}}(n)$. The above leads to the following optimization problem:

$$\begin{aligned} \max_{\alpha_i(\omega_k)} \quad & \sum_{n=0}^{L_i} \tilde{\mathbf{h}}_i^T(n)\tilde{\mathbf{h}}_i(n) \\ \text{subject to} \quad & \sum_{n=0}^{K-1} \tilde{\mathbf{h}}_i^T(n)\tilde{\mathbf{h}}_i(n) = 1 \end{aligned} \qquad (22)$$

We can write (22) in matrix form as follows:

$$\begin{aligned} \max_{\boldsymbol{\alpha}_i} \quad & \boldsymbol{\alpha}_i^\dagger \mathbf{\Psi}_i \boldsymbol{\alpha}_i \\ \text{subject to} \quad & \boldsymbol{\alpha_i}^\dagger \mathbf{\Omega}_i \boldsymbol{\alpha}_i = 1 \end{aligned} \qquad (23)$$

where $\boldsymbol{\alpha}_i = [\alpha_i(\omega_0), ..., \alpha_i(\omega_{K-1})]^T$, $\mathbf{\Psi}_i$ and $\mathbf{\Omega}_i$ are $K \times K$ matrices given by

$$\begin{aligned} \mathbf{\Psi}_i &= \frac{1}{K^2}\sum_{n=0}^{L_i} \hat{\mathcal{H}}_i(n)\hat{\mathcal{H}}_i^\dagger(n) \\ \mathbf{\Omega}_i &= \frac{1}{K^2}\sum_{n=0}^{K-1} \hat{\mathcal{H}}_i(n)\hat{\mathcal{H}}_i^\dagger(n) \qquad K \geq 2L_i - 1 \end{aligned} \qquad (24)$$

where,

$$\hat{\mathcal{H}}_i(n) = \begin{bmatrix} \hat{h}_{1i}(\omega_0)e^{j\omega_0 n} & . & . & . & \hat{h}_{Ji}(\omega_0)e^{j\omega_0 n} \\ \hat{h}_{1i}(\omega_1)e^{j\omega_1 n} & . & . & . & . \\ . & . & . & . & . \\ . & . & . & . & . \\ \hat{h}_{1i}(\omega_{K-1})e^{j\omega_{K-1} n} & . & . & . & \hat{h}_{Ji}(\omega_{K-1})e^{j\omega_{K-1} n} \end{bmatrix}_{K \times J} \qquad (25)$$

To solve the optimization problem (23) we can use Lagrangian multipliers to rewrite it in the following unconstrained form:

$$\max_{\boldsymbol{\alpha},\lambda_i} \quad \boldsymbol{\alpha}_i^\dagger \boldsymbol{\Psi}_i \boldsymbol{\alpha}_i + \lambda_i(1 - \boldsymbol{\alpha}_i^\dagger \boldsymbol{\Omega}_i \boldsymbol{\alpha}_i). \tag{26}$$

Taking the derivative of above equation with respect to $\boldsymbol{\alpha}_i$ and equating it to zero we get:

$$\boldsymbol{\Psi}_i \boldsymbol{\alpha}_i = \lambda_i \boldsymbol{\Omega}_i \boldsymbol{\alpha}_i \tag{27}$$

Thus the optimum $\boldsymbol{\alpha}_i$ is the generalized eigenvector of the matrix pencil $(\boldsymbol{\Psi}_i, \boldsymbol{\Omega}_i)$ corresponding to the maximum generalized eigenvalue.

5. SIMULATION RESULTS

In this section we present numerical simulations to demonstrate the performance of the algorithm we create quasi-stationary sources $s_i(k)$ by setting:

$$s_i(n) = \alpha_i(\lfloor n/N_E \rfloor)\eta_i(n) \tag{28}$$

where $\alpha_i(n)$ and $\eta(n)$ are mutually independent *i.i.d*, Gaussian distributed signals, the integer N_E represents the duration over which stationarity can be assumed and the symbol $\lfloor\rfloor$ represents the integer part of a fraction. We pass the sources through a 3×2, non-minimum phase MIMO channel with $L_i = 5$ adopted from [1]. To include the effect of noise we then add white Gaussian noise to the output. To estimate the cross power spectral density matrices we use the method described in [4]. It's important to know that in this simulation to estimate the CPSD matrices themselves we use 128 point FFTs while to estimate $\hat{\mathbf{H}}(n)$ we only use 16 uniformly spaced frequency samples($K = 16 > 2L_i - 1$) out of these 128 points. To measure the estimation error, since we have a scaling ambiguity, we use the following measure based on a method suggested in [6] for evaluating the estimated impulse responses when the scaling factor is ambiguous:

$$MSE = 1 - \frac{1}{JN}\sum_{j=1}^{J}\sum_{i=1}^{N}\left(\frac{\mathbf{h}_{ij}^T\hat{\mathbf{h}}_{ij}}{\|\mathbf{h}_{ij}\|\|\hat{\mathbf{h}}_{ij}\|}\right)^2 \tag{29}$$

where $\mathbf{h}_{ij} = (h_{ij}(0), ..., h_{ij}(L_{ij}))^T$ is the true ij_{th} impulse response of the channel and $\hat{\mathbf{h}}_{ij}$ is the estimated one. Table (1) shows the averaged estimation error, measured by (29), for different SNR's and different number of epochs and using 50 Monte-Carlo runs. It also shows the average number of iterations (Itr_{ave}) for the first part of the algorithm to converge. Also, to get a visual impression of the results Figures (1) and (2) illustrate the corresponding time domain impulse responses and frequency domain responses of the estimated and true channels for SNR=20dB and $N_E = 20$. As can be seen from the figure at some rare instances the estimation results are poor. This could be due to the first part of the algorithm not converging to the global minimum.

	M=20	M=40	M=100
MSE	0.0532	0.0387	0.0149
Itr_{ave}	7	7	7

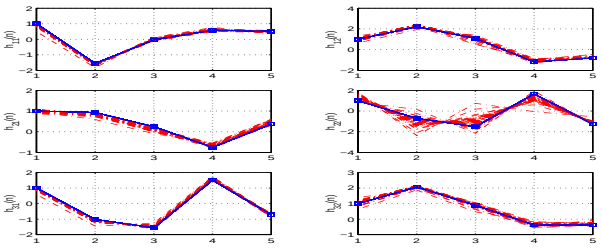

Fig. 1. 50 independently estimated $\hat{\mathbf{H}}(n)$ (red) and true $\mathbf{H}(n)$(blue). Data length=10000,M=20,SNR=20dB,K=16.

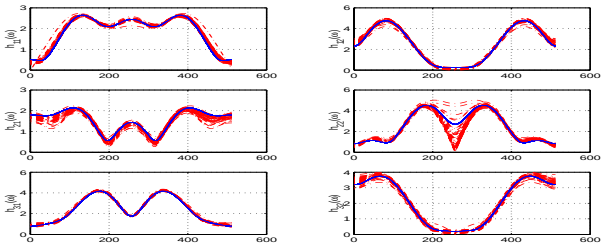

Fig. 2. 50 independently estimated $\hat{\mathbf{H}}(\omega)$ (red) and true $\mathbf{H}(\omega)$(blue). Data length=10000,M=20,SNR=20dB,K=16.

6. 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.

7. REFERENCES

[1] B. Chen and A. Petropulu, "Frequency Domain Blind MIMO System Identification Based on Second and Higher Order Statistics," *IEEE Transactions on Signal Processing*, vol. 49, pp. 1677–1688, no. 8 2001.

[2] Y. Hua, S. An, and Y. Xiang, "Blind Identification and Equalization of FIR MIMO Channels By BIDS ," in *Proc. ICASSP2001*, May 2001.

[3] C. T. Ma, Z. Ding, and S. F. Yau, "A Two-stage Algorithm for MIMO Blind Deconvolution of Nonstationary Colored Signals," *IEEE Transactions on Signal Processing*, vol. 48, pp. 1187–1192, no. 4 2000.

[4] K. Rahbar and J. Reilly, "Blind source separation of convolved sources by joint approximate diagonalization of cross-spectral density matrices," in *Proc. ICASSP2001*, May 2001.

[5] K. Rahbar, J. Reilly, and J. Manton, "Frequency Domain Blind Identification of MIMO Channels Driven by Quasi-Stationary Signals," *To be submitted to IEEE Transaction on Signal Processing*.

[6] D. R. Morgan, J. Benesty, and M. M. Sondhi, "On the Evaluation of Estimated Impulse Responses," *IEEE Signal Processing Letters*, vol. 5, pp. 174–176, no. 7 1998.