Asymptotically Near-Optimal Blind Estimation of Multipath CDMA Channels

Zhengyuan Xu

***Abstract*—In this paper, correlation matching techniques are applied to estimate multipath code division multiple access (CDMA) channels. We arrange unknown multipath parameters for each of J active users in a vector. Then, the output correlation matrix is parameterized by J unknown rank one matrices, with each one formulated from the corresponding channel vector. This correlation matrix is further compared with its sample average. The resulting error can be first minimized to obtain unbiased estimates of J unknown rank one matrices in closed forms. Thus, our estimator for each channel vector is derived by singular value decomposition (SVD) on the associated rank one matrix within a scalar ambiguity. It turns out that the performance of our estimator can be improved by introducing an asymptotically optimal weighting matrix in our cost function. This weighting matrix can be estimated directly from data samples only with a small penalty on the asymptotic performance. The asymptotic covariance of our estimator is also derived and can be compared with the Cramér–Rao lower bound, both in closed forms. Simulation results show the applicability of the proposed methods and consistency with our theoretical analysis.**

***Index Terms*—CDMA channels, correlation matching, Cramér–Rao bound.**

I. INTRODUCTION

DIGITAL wireless services have recently become increasingly more demanding. Over traditional analog schemes, digital communications show much superiority in simplified modulation scheme, channel estimation, and receiver design. To detect a desired user with lower bit errors, the knowledge of the channel experienced by that particular user is often required to build a receiver. However, the channel impulse response is not *a priori* known in many applications. Therefore, a lot of research has been conducted on channel estimation [1], [12], [14].

For a typical digital communication system, channel impulse response is usually observed to have finite order. It is reasonably modeled as a digital filter with finite impulse response (FIR). This assumption facilitates the estimation process, which results in limited number of unknowns—in many cases, only a few parameters to be identified. If training symbols are not available, blind channel estimation techniques must be used based on the receiver's matched filter output only. Second-order statistics (SOS) based methods receive more attention and can provide reliable results with a relatively small number of data points [4]. Many applicable methods with good performance have been developed based on SOS information of the received signal. Subspace methods can yield good estimates for channel parameters, even for a short period of observation if the channel order is known or estimated exactly [13]. Consistent estimators can also be obtained by effective blind least-squares approaches [20], [26].

In the CDMA framework, SOS-based methods using subspace techniques [1], [19], [24], [25], constrained optimization, and beam-forming ideas [8], [22] have also been proposed recently. In this paper, we will study the applicability of correlation matching techniques to the code division multiple access (CDMA) channel estimation problem. Correlation matching is based on the SOS of the output as well. It has been successfully applied to estimate channel parameters when the spreading sequence for each user is aperiodic [27]. This method is a special case of the method of moments, which has been widely studied and successfully applied to a variety of problems such as blind identification [3], [23], detection and estimation [6], fractionally spaced equalization [4], and time-varying channel estimation [21].

As an SOS-based method, the correlation matching approach offers some appealing advantages over other typical SOS-based methods [4]. It provides a framework for developing asymptotically optimal methods and asymptotic minimum variance estimator, against which other methods can be compared. This point will also be revealed by our channel estimation results in the simulations. Moreover, a closed-form expression for the asymptotic variance of the estimator is achievable. It will be derived for the channel estimate in the current paper. There are also other advantages presented [4]. However, these advantages are established by introducing additional computations due to nonlinear optimization. Long data records are normally required for the method to approach its asymptotic performance, as shown by our experimental results and comparisons with other methods.

In order to apply correlation matching techniques to our particular problem, we collect chip rate received samples from each bit period in an output vector and collect unknown multipath parameters for different users in different channel vectors. The output correlation matrix is thus a function of those channel vectors. If we formulate a rank one matrix corresponding to each channel vector, then the correlation matrix is parameterized by rank one matrices, whereas each rank one matrix is overparameterized by associated channel vector. First, by minimizing the error between the output correlation and its sample average with respect to the unknown rank one matrices, a closed-form solution is derived. Second, from the estimate of each rank one matrix, singular value decomposition (SVD) can be performed to

Manuscript received November 1, 1999; revised May 4, 2001. The associate editor coordinating the review of this paper and approving it for publication was Dr. Sergio Barbarossa.

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

Publisher Item Identifier S 1053-587X(01)07052-0.

obtain estimates for different channel vectors, all within a scalar ambiguity.

The proposed method provides estimates for all channels simultaneously. Due to the special structure of the signature waveform of the user's information bit in a CDMA system, which is a convolution of its channel and spreading code, channel identifiability can be shown to depend on all users' spreading codes and delays. In the uplink communication, these knowledge may be known to the base station. In such a case, the identifiability condition can be checked. However, in the downlink, the mobile station may have no knowledge about other users. Thus, it imposes some limitations on the applicability of the proposed method. This will be discussed in more detail in the paper. It is true that all blind SOS-based methods suffer from the inability to estimate the phase of the channel. Our method also gives the channel estimate within that ambiguity.

To obtain our optimal estimator for the overparameterized rank-one matrix in our first step, it is suggested in [16, Ch. 3] that the optimal weighting matrix has to be considered in the matching cost function. In our context, this weighting matrix is derived as a function of true channel parameters. It can also be estimated directly from the observed data, which only results in a small penalty on the asymptotic performance. However, our estimator for channel parameters is obtained by a two-step procedure. To further investigate the near-optimal property of the proposed estimator, we also derive the asymptotic covariance for our estimator based on perturbation theory [7, Ch. 7]. For comparison, the Cramér–Rao lower bound (CRLB) of the asymptotic estimation error is also presented. Simulations are performed to verify our theoretical analysis.

The structure of the paper is as follows. In Section II, a DS-CDMA model is described. The weighted correlation matching approach is derived in Section III, whereas its performance is analyzed in Section IV. Section V presents the asymptotically near-optimal solution based on data samples only. Brief discussions on the properties of the proposed estimator and comparisons with subspace methods are made in Section VI. Simulation results for the proposed methods and comparisons with other SOS-based methods are shown in Section VII. Finally, some conclusions are made in Section VIII.

Standard notations are adopted in this paper. Nonbold symbols denote scalars, uppercase boldface letters are used for matrices, and lower boldface ones are used for vectors.

Re$(\cdot)$ and Im$(\cdot)$	real and imaginary parts of their contents and $i = \sqrt{-1}$, except as the subscript index to represent the imaginary part in some cases.
$E\{\cdot\}$	Expected value.
diag$\{\cdot\}$	Diagonal matrix (if its elements are scalars) or block diagonal matrix (if its elements are matrices).
$\mathbf{I}_k$	$k \times k$ identity matrix.
$\mathbf{0}$	Zero vector with proper dimensions in the context.
$\mathbf{O}$	zero matrix with proper dimensions in the context.
$\lfloor\cdot\rfloor$	rounding down to the nearest integer.
$\otimes$	Kronecker product (see also [10,Ch. 12]).
T	Transpose.
$*$	Complex conjugate.
H	Hermitian.
-1	Inverse.

Notation $\mathbf{C}(k_1\colon k_2, :)$ from Matlab is also employed to indicate a block of matrix $\mathbf{C}$ from the k_1th row to the k_2th row. An operation vec on a matrix will stack the columns of the matrix from the left to the right into one long vector (see [10, Ch. 12]).

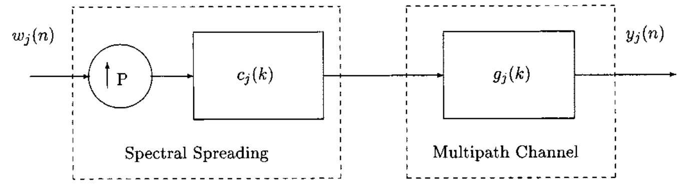

Fig. 1. DS/SS signal in multipath: discrete-time model.

II. Problem Statement

In a DS-CDMA system, let user j, $j = 1, \ldots, J$ use a spreading code $c_j(k)$, $k = 1, \ldots, P$ of length P to transmit P chips per information symbol. Let the chip sequence be transmitted through a multipath channel with a discrete-time baseband impulse response $g_j(n)$ (including the transmitter and receiver filters); then, the received discrete-time signal $y(n)$ at the chip rate receiver is a superposition of the signals from all users plus noise $v(n)$ (see [22] and Fig. 1)

$$\begin{aligned} y(n) &= \sum_{j=1}^{J} y_j(n) + v(n) \\ y_j(n) &= \sum_{l=-\infty}^{\infty} w_j(l) h_j(n - \tau_j - lP) \\ h_j(n) &= \sum_{m=-\infty}^{\infty} g_j(m) c_j(n-m) \end{aligned} \tag{1}$$

where $w_j(n)$ is zero-mean, infinite independently distributed (i.i.d.) information-bearing sequence of user j ($j = 1, \ldots, J$) with equal power[1] $\sigma_{w_j}^2 = \mathrm{E}\{\|w_j(n)\|^2\} = \sigma_w^2$, $h_j(n)$ is its signature waveform, τ_j is the delay of user j in chip periods, and $v(n)$ is assumed to be additive white Gaussian noise (AWGN) with zero-mean and variance $\sigma_v^2 = \mathrm{E}\{\|v(n)\|^2\}$. Without loss of generality, we may assume that the delay $0 \le \tau_j < P$. We will also assume that $g_j(n)$ has finite impulse response of maximum order q (typically $q \ll P$ in many applications).

If we collect $P + q$ samples of $y(n)$ in a vector $\tilde{\mathbf{y}}(n) = [y(nP+1), \ldots, y(nP+P+q)]^T$ and similarly define the noise vector $\tilde{\mathbf{v}}(n)$, then from (1), the received signal $\tilde{\mathbf{y}}(n)$ becomes

$$\tilde{\mathbf{y}}(n) = \sum_{j=1}^{J} \Big[\tilde{\mathbf{C}}_{j,1}\mathbf{g}_j w_j(n) + \tilde{\mathbf{C}}_{j,2}\mathbf{g}_j w_j(n-1) + \tilde{\mathbf{C}}_{j,3}\mathbf{g}_j w_j(n+1)\Big] + \tilde{\mathbf{v}}(n) \tag{2}$$

[1]The difference in power can be incorporated in the channel.

where interference from other symbols is taken into account, $\mathbf{g}_j = [g_j(0), \ldots, g_j(q)]^T$ is the multipath channel vector of user j, and $\tilde{\mathbf{C}}_{j,1}$, $\tilde{\mathbf{C}}_{j,2}$, $\tilde{\mathbf{C}}_{j,3}$[2] have following structures:

$$\tilde{\mathbf{C}}_j = \begin{bmatrix} c_j(1) & & \mathbf{0} \\ \vdots & \ddots & c_j(1) \\ c_j(P) & & \vdots \\ \mathbf{0} & \ddots & c_j(P) \end{bmatrix}$$

$$\tilde{\mathbf{C}}_{j,1} = \begin{bmatrix} \mathbf{0} \\ \tilde{\mathbf{C}}_j(1\colon P+q-\tau_j, :) \end{bmatrix} \tag{3}$$

$$\tilde{\mathbf{C}}_{j,2} = \begin{bmatrix} \tilde{\mathbf{C}}_j(P+1-\tau_j\colon P+q, :) \\ \mathbf{0} \end{bmatrix}$$

$$\tilde{\mathbf{C}}_{j,3} = \begin{bmatrix} \mathbf{0} \\ \tilde{\mathbf{C}}_j(1\colon q-\tau_j, :) \end{bmatrix}. \tag{4}$$

For clarity of presentation, we restrict our attention to quasisynchronous users [9], where $\tau_j \ll P$. The method can be easily extended to asynchronous CDMA systems. Consider ν data points from $\tilde{\mathbf{y}}(n)$ and formulate a partial data vector $\mathbf{y}(n) = [y(nP + P - \nu + 1), \ldots, y(nP + P)]^T$ with $P - \nu = \max(q + \tau_j)$ $(j = 1, \ldots, J)$ to eliminate an intersymbol interference (ISI) effect.[3] A similar definition is applied to the new partial noise vector $\mathbf{v}(n)$. Then, from (2), we can obtain our input/output relationship

$$\mathbf{y}(n) = \sum_{j=1}^{J} \mathbf{C}_j \mathbf{g}_j w_j(n) + \mathbf{v}(n) = \mathbf{H}\mathbf{w}(n) + \mathbf{v}(n) \tag{5}$$

where $\mathbf{C}_j = \tilde{\mathbf{C}}_{j,1}(P - \nu + 1\colon P, :)$, $\mathbf{w}(n) = [w_1(n), \ldots, w_J(n)]^T$, and

$$\mathbf{H} = [\mathbf{h}_1, \ldots, \mathbf{h}_J]_{\nu \times J}, \; \mathbf{h}_j = \mathbf{C}_j \mathbf{g}_j. \tag{6}$$

The output correlation matrix can be easily computed from (5)

$$\mathbf{R} = \sigma_w^2 \sum_{j=1}^{J} \mathbf{C}_j \mathbf{g}_j \mathbf{g}_j^H \mathbf{C}_j^H + \sigma_v^2 \mathbf{I}. \tag{7}$$

For a general discussion, we assume all code matrices, channels, inputs, and noise are complex (e.g., QPSK modulation); then, $\mathbf{y}(n)$ and $\mathbf{R}$ are also complex.[4] From (7), it is clear that the output correlation is parameterized by $\sigma_w^2 \mathbf{g}_j \mathbf{g}_j^H$ and σ_v^2. Since $\sigma_w^2 \mathbf{g}_j \mathbf{g}_j^H$ is a rank-one Hermitian matrix, it has all real eigenvalues and one largest positive one. If $\sigma_w^2 \mathbf{g}_j \mathbf{g}_j^H$ can be obtained from the knowledge of $\mathbf{R}$, then SVD can be performed on it to obtain its eigenvector associated with that unique maximum eigenvalue. This eigenvector is then our estimate of $\mathbf{g}_j$ within a scalar ambiguity. We will treat $\mathbf{R}$ as a general function of our unknowns $\mathbf{g}_j$ and σ_v^2, and focus on estimating $\sigma_w^2 \mathbf{g}_j \mathbf{g}_j^H$ and σ_v^2 next, based on correlation matching techniques.

[2] $\tilde{\mathbf{C}}_{j,3}$ will become $\mathbf{0}$ if $\tau_j \geq q$.

[3] In an asynchronous system or when ISI is present, the proposed methods can be easily extended, as explained in Section III.

[4] The case of all real numbers can be more easily treated.

III. Correlation Matching Approach

As is well known, the correlation matrix $\mathbf{R}$ is Hermitian, where there is much redundancy. In order to reduce computational complexity, this redundancy will be removed before it is used in the matching context. The reason will become clear in later sections.[5]

A. Redundancy Removal in $\mathbf{R}$

If we write the complex matrix $\mathbf{R}$ in an explicit form as $\mathbf{R} = \mathbf{R}_1 + i\mathbf{R}_2$, where $\mathbf{R}_1$ and $\mathbf{R}_2$ are both real, then from $\mathbf{R} = \mathbf{R}^H$, we obtain $\mathbf{R}_1^T = \mathbf{R}_1$ and $\mathbf{R}_2^T = -\mathbf{R}_2$. Therefore, we consider only those elements in the lower triangular part of $\mathbf{R}_1$ together with diagonal elements. Similarly, only those elements in the lower triangular part of $\mathbf{R}_2$ will be used because $\mathbf{R}_2^T = -\mathbf{R}_2$, and all diagonal elements are zeros. In order to obtain a closed-form solution for $\sigma_w^2 \mathbf{g}_j \mathbf{g}_j^H$, we operate vec on $\mathbf{R}$ as $\tilde{\mathbf{r}} = vec(\mathbf{R})$. This makes $\tilde{\mathbf{r}} = \mathbf{r}_1 + i\mathbf{r}_2$ valid, where $\mathbf{r}_1 = vec(\mathbf{R}_1)$ and $\mathbf{r}_2 = vec(\mathbf{R}_2)$. Furthermore, to perform all computations on real numbers, $\mathbf{r}_1$ and $\mathbf{r}_2$ can be put in a new vector as [2, Ch. 4]

$$\mathbf{r} = \begin{bmatrix} \mathbf{r}_1 \\ \mathbf{r}_2 \end{bmatrix} = \begin{bmatrix} \mathrm{Re}\{vec(\mathbf{R})\} \\ \mathrm{Im}\{vec(\mathbf{R})\} \end{bmatrix}. \tag{8}$$

To remove redundancy in $\mathbf{r}_1$ and $\mathbf{r}_2$, we use two selection matrices $\mathbf{L}_1$ and $\mathbf{L}_2$ to pick up the interesting elements, respectively

$$\mathbf{z}_1 = \mathbf{L}_1 \mathbf{r}_1, \qquad \mathbf{z}_2 = \mathbf{L}_2 \mathbf{r}_2 \tag{9}$$

with

$$\mathbf{L}_1 = \mathrm{diag}\left\{\mathbf{I}_\nu, \tilde{\mathbf{L}}\right\}, \qquad \mathbf{L}_2 = [\,\tilde{\mathbf{L}} \quad \mathbf{0}_{(1/2)\nu(\nu-1)\times\nu}\,] \tag{10}$$

where

$$\tilde{\mathbf{L}} = \mathrm{diag}\left\{\tilde{\mathbf{L}}_1, \ldots, \tilde{\mathbf{L}}_{\nu-1}\right\}$$
$$\tilde{\mathbf{L}}_k = [\,\mathbf{0}_{(\nu-k)\times k}, \quad \mathbf{I}_{\nu-k}\,], \qquad k = 1, \ldots, \nu - 1. \tag{11}$$

Therefore, if $\mathbf{z}_1$ and $\mathbf{z}_2$ are also put in another vector $\mathbf{z}$

$$\mathbf{z} = \begin{bmatrix} \mathbf{z}_1 \\ \mathbf{z}_2 \end{bmatrix} \tag{12}$$

then it is related to $\mathbf{r}$ by

$$\mathbf{z} = \mathbf{L}\mathbf{r}, \qquad \mathbf{L} = \mathrm{diag}\{\mathbf{L}_1, \mathbf{L}_2\}. \tag{13}$$

In order to avoid possible confusion, dimensions of some matrices and vectors mentioned above, as well as their real or complex properties, are indicated in Table I.

It can be observed that both $\mathbf{L}_1$ and $\mathbf{L}_2$ have full row rank with each row having only one nonzero element ("1"). After this selection operation, in total, ν^2 redundant elements are removed from $\mathbf{r}$. This operation is helpful for both the identifiability of unknown parameters and the complexity reduction for

[5] To expose it in short words, the weighting matrix used in the matching cost function [see (33)] will reduce its dimension. Thus, a cost-efficient estimate for the optimal weighting matrix from data samples can be achieved [see (84) and (85)].

TABLE I
DIMENSIONS AND PROPERTIES OF SOME MATRICES AND VECTORS

Matrices or Vectors	Dimension	Properties
$\mathbf{R}$	$\nu \times \nu$	complex
$\mathbf{R}_1$, $\mathbf{R}_2$	$\nu \times \nu$	real
$\mathbf{L}_1$	$\frac{1}{2}\nu(\nu+1) \times \nu^2$	real
$\mathbf{L}_2$	$\frac{1}{2}\nu(\nu-1) \times \nu^2$	real
$\tilde{\mathbf{L}}_k$	$(\nu-k) \times \nu$	real
$\tilde{\mathbf{L}}$	$\frac{1}{2}\nu(\nu-1) \times \nu(\nu-1)$	real
$\mathbf{L}$	$\nu^2 \times 2\nu^2$	real
$\mathbf{y}(n)$	$\nu \times 1$	complex
$\mathbf{r}_1$, $\mathbf{r}_2$, $\mathbf{z}$	$\nu^2 \times 1$	real
$\mathbf{r}$	$2\nu^2 \times 1$	real
$\mathbf{z}_1$	$\frac{1}{2}\nu(\nu+1) \times 1$	real
$\mathbf{z}_2$	$\frac{1}{2}\nu(\nu-1) \times 1$	real
$\tilde{\mathbf{r}}$	$\nu^2 \times 1$	complex

the method. Hence, we can apply correlation matching techniques to match the resulting real vector $\mathbf{z}$ instead of $\mathbf{r}$. However, the derivation of the estimate for $\mathbf{z}$ is not obvious. Motivated by (13) and (8) in our framework, we will derive this estimate $\hat{\mathbf{z}}_N$ from the estimate for $\mathbf{r}$. If we adopt common practice and use the sample average $(1/N)\sum_{n=1}^{N}\mathbf{y}(n)\mathbf{y}^H(n)$ from N data windows to approximate the correlation matrix $\mathbf{R}$, then from (8), the sample average of $\mathbf{r}$ can be obtained

$$\hat{\mathbf{r}}_N = \frac{1}{N}\sum_{n=1}^{N}\hat{\mathbf{r}}_n, \qquad \hat{\mathbf{r}}_n = \begin{bmatrix}\hat{\mathbf{r}}_{1,n}\\ \hat{\mathbf{r}}_{2,n}\end{bmatrix} = \begin{bmatrix} vec\left[\mathrm{Re}\left(\hat{\mathbf{R}}_n\right)\right] \\ vec\left[\mathrm{Im}\left(\hat{\mathbf{R}}_n\right)\right]\end{bmatrix}$$
$$\hat{\mathbf{R}}_n = \mathbf{y}(n)\mathbf{y}^H(n). \tag{14}$$

Therefore, the estimate for $\mathbf{z}$ readily follows:

$$\hat{\mathbf{z}}_N = \mathbf{L}\hat{\mathbf{r}}_N. \tag{15}$$

B. Relationship Between $\mathbf{z}$ *and Channel Parameters*

Before presenting our cost function by matching $\mathbf{z}$ with $\hat{\mathbf{z}}_N$, we will first investigate the relationship between $\mathbf{z}$ and our unknowns $\mathbf{g}_j$ and σ_v^2. Let us first work on $\mathbf{r}$ because our unknowns are embedded in $\mathbf{r}$, and $\mathbf{z}$ is related to $\mathbf{r}$ by (13). In our problem, there are, in total, $K = 2J(q+1)+1$ unknown parameters. We put them in a real vector $\boldsymbol{\theta}$

$$\boldsymbol{\theta} = [\mathbf{g}_{1r}^T, \ldots, \mathbf{g}_{Jr}^T, \mathbf{g}_{1i}^T, \ldots, \mathbf{g}_{Ji}^T, \sigma_v^2]^T$$
$$\mathbf{g}_{jr} = \mathrm{Re}(\mathbf{g}_j), \ \mathbf{g}_{ji} = \mathrm{Im}(\mathbf{g}_j) \tag{16}$$

and similarly, define $\boldsymbol{\theta}_0$ to be the true parameter vector.

As we have shown in (8), $\mathbf{r}$ is determined by $\mathbf{R}$. According to the expression for $\mathbf{R}$ in (7), we have

$$\tilde{\mathbf{r}} = vec(\mathbf{R}) = \mathbf{Q}\mathbf{d} + \sigma_v^2 vec(\mathbf{I}_\nu) \tag{17}$$

where[6]

$$\mathbf{Q} = [\mathbf{Q}_1, \ldots, \mathbf{Q}_J], \qquad \mathbf{Q}_j = \mathbf{C}_j^* \otimes \mathbf{C}_j \tag{18}$$

and

$$\mathbf{d} = [\mathbf{d}_1^T, \ldots, \mathbf{d}_J^T]^T, \qquad \mathbf{d}_j = vec(\mathbf{G}_j)$$
$$\mathbf{G}_j = \sigma_w^2 \mathbf{g}_j \mathbf{g}_j^H. \tag{19}$$

Since $\mathbf{Q}$ and $\mathbf{d}$ are both complex, we partition them into real and imaginary parts as

$$\mathbf{Q} = \mathbf{Q}_R + i\mathbf{Q}_I, \qquad \mathbf{d} = \mathbf{d}_R + i\mathbf{d}_I \tag{20}$$

where $\mathbf{Q}_R$, $\mathbf{Q}_I$, and $\mathbf{d}_R$, $\mathbf{d}_I$ are all real. Then, $\mathbf{Q}\mathbf{d}$ explicitly has its real part and imaginary part

$$\mathbf{Q}\mathbf{d} = (\mathbf{Q}_R\mathbf{d}_R - \mathbf{Q}_I\mathbf{d}_I) + i(\mathbf{Q}_I\mathbf{d}_R + \mathbf{Q}_R\mathbf{d}_I). \tag{21}$$

From (8), (17), and (21), we obtain

$$\mathbf{r} = \mathbf{\Pi}\mathbf{x} \tag{22}$$

where

$$\mathbf{\Pi} = \begin{bmatrix}\mathbf{Q}_R & -\mathbf{Q}_I & vec(\mathbf{I}_\nu)\\ \mathbf{Q}_I & \mathbf{Q}_R & \mathbf{0}\end{bmatrix}, \qquad \mathbf{x} = \begin{bmatrix}\mathbf{d}_R\\ \mathbf{d}_I\\ \sigma_v^2\end{bmatrix}. \tag{23}$$

In (22), $\mathbf{\Pi}$ is a real matrix depending on code matrices and delays of all users, and $\mathbf{x}$ is a function of $\boldsymbol{\theta}$ whose relationship will be further illustrated next.

From (16), $\mathbf{g}_{jr}$, $\mathbf{g}_{ji}$, and σ_v^2 can be picked up from $\boldsymbol{\theta}$ as the following:

$$\mathbf{g}_{jr} = \mathbf{S}_{jr}\boldsymbol{\theta}, \qquad \mathbf{S}_{jr} = (\tilde{\mathbf{e}}_j^T \otimes \mathbf{I}_{q+1})[\mathbf{I}_{K-1}\mathbf{0}] \tag{24}$$
$$\mathbf{g}_{ji} = \mathbf{S}_{ji}\boldsymbol{\theta}, \qquad \mathbf{S}_{ji} = (\tilde{\mathbf{e}}_{J+j}^T \otimes \mathbf{I}_{q+1})[\mathbf{I}_{K-1}\mathbf{0}] \tag{25}$$
$$\sigma_v^2 = \mathbf{e}_K^T\boldsymbol{\theta} \tag{26}$$

where $\tilde{\mathbf{e}}_j$ and $\tilde{\mathbf{e}}_{J+j}$ are both unitary vectors of length $2J$ with jth and $(J+j)$th element equal to 1, respectively, whereas $\mathbf{e}_K$ is also a unitary vector of length K with a Kth element equal to 1. Therefore

$$\mathbf{g}_j = \mathbf{S}_{jr}\boldsymbol{\theta} + i\mathbf{S}_{ji}\boldsymbol{\theta} \tag{27}$$

from which we have

$$\mathbf{g}_j\mathbf{g}_j^H = (\mathbf{S}_{jr}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{S}_{jr}^T - \mathbf{S}_{ji}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{S}_{ji}^T) + i(\mathbf{S}_{ji}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{S}_{jr}^T + \mathbf{S}_{jr}\boldsymbol{\theta}\boldsymbol{\theta}^T\mathbf{S}_{ji}^T). \tag{28}$$

Hence, according to (19), $\mathbf{d}$ can be expressed as

$$\mathbf{d} = \mathbf{T}_r vec(\boldsymbol{\theta}\boldsymbol{\theta}^T) + i\mathbf{T}_i vec(\boldsymbol{\theta}\boldsymbol{\theta}^T) \tag{29}$$

[6] If all users are asynchronous, then $\mathbf{Q}_j$ becomes $\mathbf{C}_{j,1}^* \otimes \mathbf{C}_{j,1} + \mathbf{C}_{j,2}^* \otimes \mathbf{C}_{j,2} + \mathbf{C}_{j,3}^* \otimes \mathbf{C}_{j,3}$, where $\mathbf{C}_{j,k}$ $(k = 1, 2, 3)$ is similarly defined as $\tilde{\mathbf{C}}_{j,k}$ in (3) and (4).

where

$$\mathbf{T}_r = \sigma_w^2 \begin{bmatrix} \mathbf{S}_{1r} \otimes \mathbf{S}_{1r} - \mathbf{S}_{1i} \otimes \mathbf{S}_{1i} \\ \vdots \\ \mathbf{S}_{Jr} \otimes \mathbf{S}_{Jr} - \mathbf{S}_{Ji} \otimes \mathbf{S}_{Ji} \end{bmatrix}$$
$$\mathbf{T}_i = \sigma_w^2 \begin{bmatrix} \mathbf{S}_{1r} \otimes \mathbf{S}_{1i} + \mathbf{S}_{1i} \otimes \mathbf{S}_{1r} \\ \vdots \\ \mathbf{S}_{Jr} \otimes \mathbf{S}_{Ji} + \mathbf{S}_{Ji} \otimes \mathbf{S}_{Jr} \end{bmatrix}. \quad (30)$$

Considering (23), (26), and (29) together, $\mathbf{x}$ is related to $\boldsymbol{\theta}$ by

$$\mathbf{x} = \begin{bmatrix} \mathbf{T}_r vec(\boldsymbol{\theta}\boldsymbol{\theta}^T) \\ \mathbf{T}_i vec(\boldsymbol{\theta}\boldsymbol{\theta}^T) \\ \mathbf{e}_K^T \boldsymbol{\theta} \end{bmatrix}. \quad (31)$$

Therefore, from (13), (22), and (31), $\mathbf{z}$ is determined by $\boldsymbol{\theta}$ as the following expression:

$$\mathbf{z} = \mathbf{L\Pi x} = \mathbf{L\Pi} \begin{bmatrix} \mathbf{T}_r vec(\boldsymbol{\theta}\boldsymbol{\theta}^T) \\ \mathbf{T}_i vec(\boldsymbol{\theta}\boldsymbol{\theta}^T) \\ \mathbf{e}_K^T \boldsymbol{\theta} \end{bmatrix}. \quad (32)$$

C. Proposed Estimator

Since we have shown $\mathbf{z}$ as a function of $\boldsymbol{\theta}$ in (32), the following standard cost function can be adopted to match $\mathbf{z}$ with a positive definite weighting matrix $\mathbf{M}$

$$\tilde{J}_N = (\hat{\mathbf{z}}_N - \mathbf{z})^T \mathbf{M} (\hat{\mathbf{z}}_N - \mathbf{z}). \quad (33)$$

After substituting (15) and (32), it becomes

$$\tilde{J}_N = (\hat{\mathbf{r}}_N - \mathbf{\Pi x})^T \mathbf{L}^T \mathbf{M} \mathbf{L} (\hat{\mathbf{r}}_N - \mathbf{\Pi x}). \quad (34)$$

If we substitute $\mathbf{x}$ in (31) into (34), we observe that $\tilde{J}_N$ will be the fourth-order function of our unknown parameter vector $\boldsymbol{\theta}$. This highly nonlinear cost function will introduce much difficulty in estimation. Therefore, we propose two steps: first, to estimate $\mathbf{x}$ and then, to obtain our unknown parameters from the estimate of $\mathbf{x}$. Under this consideration, $\tilde{J}_N$ becomes a quadratic function of $\mathbf{x}$ if $\mathbf{M}$ is constant and a unique and closed-form solution to minimize (34) can be derived, which is

$$\hat{\mathbf{x}}_N = (\mathbf{\Pi}^T \mathbf{L}^T \mathbf{M} \mathbf{L} \mathbf{\Pi})^{-1} \mathbf{\Pi}^T \mathbf{L}^T \mathbf{M} \mathbf{L} \hat{\mathbf{r}}_N. \quad (35)$$

In this first step result, the weighting matrix can be any positive definite matrix. For simplicity of the algorithm, the identity matrix is usually chosen. We will discuss later its choice with asymptotically optimal performance for $\hat{\mathbf{x}}_N$. Of course, the overparameterization by $\mathbf{x}$ will cause some performance loss in estimating the channel parameters.

In our second step, we will obtain $\hat{\mathbf{g}}_j$ and $\hat{\sigma}_v^2$ from $\hat{\mathbf{x}}_N$. According to (20) and (23), $\hat{\mathbf{d}}$ can be reconstructed from $\hat{\mathbf{x}}_N$ as

$$\hat{\mathbf{d}} = [\mathbf{I}_{J(q+1)^2} i \mathbf{I}_{J(q+1)^2} \mathbf{0}] \hat{\mathbf{x}}_N. \quad (36)$$

From (19), we can select $\hat{\mathbf{d}}_j$ from $\hat{\mathbf{d}}$ by

$$\hat{\mathbf{d}}_j = [\mathbf{O}, \ldots, \mathbf{O I}_{(q+1)^2} \mathbf{O}, \ldots, \mathbf{O}] \hat{\mathbf{d}}$$

and therefore, $\hat{\mathbf{d}}_j$ is related to $\hat{\mathbf{x}}_N$ by

$$\hat{\mathbf{d}}_j = [\mathbf{O}, \ldots, \mathbf{O I}_{(q+1)^2} \mathbf{O}, \ldots, \mathbf{O} i \mathbf{I}_{(q+1)^2} \mathbf{O}, \ldots, \mathbf{O 0}] \hat{\mathbf{x}}_N. \quad (37)$$

As we know from (19), $\mathbf{d}_j$ is the resulting vector of the vec operation on $\mathbf{G}_j$. Hence, the reverse vec operation on $\hat{\mathbf{d}}_j$ can be performed to obtain our estimate $\hat{\mathbf{G}}_j$ for $\mathbf{G}_j$. Each column of $\hat{\mathbf{G}}_j$ is constructed by successively picking up $q+1$ elements from $\hat{\mathbf{d}}_j$

$$\hat{\mathbf{G}}_j = [\hat{\mathbf{d}}_{jm}], \qquad \hat{\mathbf{d}}_{jm} = [\mathbf{O}, \ldots, \mathbf{O I}_{q+1} \mathbf{O}, \ldots, \mathbf{O}] \hat{\mathbf{d}}_j \quad (38)$$

for $m = 1, \ldots, q+1$ columns. Substituting (37) into (38), we have

$$\hat{\mathbf{G}}_j = [\mathbf{P}_{m,j} \hat{\mathbf{x}}_N], \qquad m = 1, \ldots, q+1 \quad (39)$$

whose mth column is obtained via a selection operation $\mathbf{P}_{m,j}$ on $\hat{\mathbf{x}}_N$

$$\mathbf{P}_{m,j} = [\mathbf{e}^T_{(j-1)(q+1)+m} \otimes \mathbf{I}_{q+1} + i \mathbf{e}^T_{(J+j-1)(q+1)+m} \otimes \mathbf{I}_{q+1}, \mathbf{0}]. \quad (40)$$

Once matrix $\hat{\mathbf{G}}_j$ is obtained, our estimate for channel vector $\mathbf{g}_j$ readily follows, as explained in the sequel.

Observe that matrix $\mathbf{G}_j = \sigma_w^2 (\mathbf{g}_j \mathbf{g}_j^H)$ is a rank-one matrix. If we operate SVD on it, we can obtain

$$\sigma_w^2 \mathbf{g}_j \mathbf{g}_j^H = \lambda_j \boldsymbol{\alpha}_j \boldsymbol{\alpha}_j^H \quad (41)$$

where λ_j is its unique maximum eigenvalue and positive

$$\lambda_j = \sigma_w^2 \|\mathbf{g}_j\|^2 \quad (42)$$

and $\boldsymbol{\alpha}_j$ is the corresponding eigenvector

$$\boldsymbol{\alpha}_j = \frac{e^{i\phi_j}}{\|\mathbf{g}_j\|} \mathbf{g}_j \quad (43)$$

where ϕ_j is the phase ambiguity. Therefore, from (42) and (43), $\mathbf{g}_j$ can be determined by λ_j and $\boldsymbol{\alpha}_j$ as

$$\mathbf{g}_j = \beta_j e^{-i\phi_j} \boldsymbol{\alpha}_j, \qquad \beta_j = \sqrt{\frac{\lambda_j}{\sigma_w^2}} \quad (44)$$

within a phase ambiguity ϕ_j. Based on these analyses, our estimate of $\mathbf{g}_j$ then has the following expression:

$$\hat{\mathbf{g}}_j = \hat{\beta}_j e^{-i\hat{\phi}_j} \hat{\boldsymbol{\alpha}}_j, \qquad \hat{\beta}_j = \sqrt{\frac{\hat{\lambda}_j}{\sigma_w^2}} \quad (45)$$

where $(\hat{\lambda}_j, \hat{\boldsymbol{\alpha}}_j)$ is the maximum eigenpair of matrix $\hat{\mathbf{G}}_j$, and $\hat{\phi}_j$ is the phase ambiguity. Equation (45) tells us that $\mathbf{g}_j$ can be estimated within a phase ambiguity if σ_w^2 is known or within a scalar ambiguity if σ_w^2 is not available.

Meanwhile as a byproduct, the estimate for the noise power is obtained according to (23) by taking out the last element of $\hat{\mathbf{x}}_N$

$$\hat{\sigma}_v^2 = [0, \ldots, 0, 1]^T \hat{\mathbf{x}}_N. \quad (46)$$

IV. Performance Analysis

One may doubt that our two-step estimation procedure can provide reliable estimate for channel parameters. Thus, asymptotic estimation error for the proposed estimator needs further analysis in this section.

A. Identifiability

In the first step of our estimation, the solution obtained in (35) exists only when the square matrix $\mathbf{\Pi}^T\mathbf{L}^T\mathbf{M}\mathbf{L}\mathbf{\Pi}$ has full rank. This matrix is determined by users' spreading codes, delays, and the length of the channel, as seen by (18) and (23). Part of this information may be available to the receiver, such as codes and delays in uplink communication by the base station. Then, the identifiability condition can be checked for a particular communication environment (with some *a priori* knowledge about channel's order) before estimation. Since the proposed method is a blind method, it is not possible to determine *a priori* whether the condition is satisfied for all applications [4]. For example, in the downlink, the mobile user may have no knowledge about other users, although it can acquire its timing based on other techniques. Therefore, for the proposed method to succeed, all codes and delays have been assumed to be known/estimated. These limitations on the applicability of the proposed method are not uncommon when channels for all users are estimated simultaneously. Instead, when the channel for only one desired user is estimated, the condition has been relaxed in SOS-based approaches [1], [19], where only the spreading codes of the desired user is required while the delay is estimated.

It has been shown that the rank of $\mathbf{\Pi}^T\mathbf{L}^T\mathbf{M}\mathbf{L}\mathbf{\Pi}$ depends on all delays. The arrival time for each user may be arbitrary. For a quasisynchronous system [9], $\tau_j \ll P$. It is a plausible conjecture for typical applications that this matrix has full rank if $\mathbf{M}$ is nonsingular and positive definite and if $\mathbf{L}\mathbf{\Pi}$ has full rank, which requires $\nu^2 \geq [2J(q+1)^2+1]$. However, we are still unable to provide a proof at this point. Under these conditions, our estimator $\hat{\mathbf{x}}_N$ exists and is unique.

In order to see if $\hat{\mathbf{x}}_N$ asymptotically converges to its true value, let us denote $\mathbf{r}_0$, $\mathbf{z}_0$, and $\mathbf{x}_0$ as values of $\mathbf{r}$, $\mathbf{z}$, and $\mathbf{x}$ at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$. In our framework, the following holds:

$$\lim_{N\to\infty} \frac{1}{N}\sum_{n=1}^{N} \mathbf{y}(n)\mathbf{y}^H(n) = E\{\mathbf{y}(n)\mathbf{y}^H(n)\}.$$

Therefore, together with (8) and (14), we can conclude that as $N \to \infty$, $\hat{\mathbf{r}}_N$ will converge to $\mathbf{r}_0$. Furthermore, according to (22) and under the identifiability condition, we have

$$\lim_{N\to\infty} \hat{\mathbf{x}}_N = \mathbf{x}_0 \tag{47}$$

which indicates that $\hat{\mathbf{x}}_N$ is a consistent estimator of $\mathbf{x}_0$. Since our estimate $\hat{\mathbf{G}}_j$ is obtained from $\hat{\mathbf{x}}_N$ by linear operation, $\hat{\mathbf{G}}_j$ is unique and asymptotically consistent with $\mathbf{G}_j$. The SVD operation in our second step will also lead to our estimate where $\hat{\mathbf{g}}_j$ is unique and asymptotically consistent with $\mathbf{g}_j$ within a phase ambiguity.

B. Asymptotic Covariance of $\hat{\mathbf{z}}_N$

According to [16, Ch. 3], the asymptotic covariance of $\hat{\mathbf{z}}_N$ plays an important role in optimal estimation. Therefore, we will first focus on developing an expression for it.

The normalized asymptotic covariance of $\hat{\mathbf{z}}_N$ generated by $\boldsymbol{\theta}_0$ is defined as

$$\mathbf{\Sigma}(\boldsymbol{\theta}_0) = \lim_{N\to\infty} NE\{(\hat{\mathbf{z}}_N - \mathbf{z}_0)(\hat{\mathbf{z}}_N - \mathbf{z}_0)^T\}. \tag{48}$$

Substituting $\mathbf{z}_0$ and $\hat{\mathbf{z}}_N$ according to (13) and (15), (48) becomes

$$\mathbf{\Sigma}(\boldsymbol{\theta}_0) = \mathbf{L}\lim_{N\to\infty} NE\{(\hat{\mathbf{r}}_N - \mathbf{r}_0)(\hat{\mathbf{r}}_N - \mathbf{r}_0)^T\}\mathbf{L}^T. \tag{49}$$

Since ISI is eliminated by properly choosing data windows and $\hat{\mathbf{r}}_n$ is an independent sequence, then by using (14) and the independent assumption of $\hat{\mathbf{r}}_n$, (49) can be simplified as[7]

$$\mathbf{\Sigma}(\boldsymbol{\theta}_0) = \mathbf{L}(E\{\hat{\mathbf{r}}_n\hat{\mathbf{r}}_n^T\} - \mathbf{r}_0\mathbf{r}_0^T)\mathbf{L}^T. \tag{50}$$

Equation (50) indicates that $\mathbf{\Sigma}(\boldsymbol{\theta}_0)$ can be computed from $\boldsymbol{\theta}_0$ if $E\{\hat{\mathbf{r}}_n\hat{\mathbf{r}}_n^T\}$ can also be expressed by $\boldsymbol{\theta}_0$.

According to (14), $\hat{\mathbf{r}}_n\hat{\mathbf{r}}_n^T$ can be expanded as

$$\hat{\mathbf{r}}_n\hat{\mathbf{r}}_n^T = \begin{bmatrix} \hat{\mathbf{r}}_{1,n}\hat{\mathbf{r}}_{1,n}^T & \hat{\mathbf{r}}_{1,n}\hat{\mathbf{r}}_{2,n}^T \\ \hat{\mathbf{r}}_{2,n}\hat{\mathbf{r}}_{1,n}^T & \hat{\mathbf{r}}_{2,n}\hat{\mathbf{r}}_{2,n}^T \end{bmatrix}. \tag{51}$$

If we let (e.g., [4])

$$\mathbf{\Psi}_1 = vec\left(\hat{\mathbf{R}}_n\right)\left[vec\left(\hat{\mathbf{R}}_n\right)\right]^T \tag{52}$$

$$\mathbf{\Psi}_2 = vec\left(\hat{\mathbf{R}}_n\right)\left[vec\left(\hat{\mathbf{R}}_n\right)\right]^H \tag{53}$$

and observe that $vec\left(\hat{\mathbf{R}}_n\right) = \hat{\mathbf{r}}_{1,n} + i\hat{\mathbf{r}}_{2,n}$ from (14), then

$$\mathbf{\Psi}_1 = (\hat{\mathbf{r}}_{1,n}\hat{\mathbf{r}}_{1,n}^T - \hat{\mathbf{r}}_{2,n}\hat{\mathbf{r}}_{2,n}^T) + i(\hat{\mathbf{r}}_{1,n}\hat{\mathbf{r}}_{2,n}^T + \hat{\mathbf{r}}_{2,n}\hat{\mathbf{r}}_{1,n}^T) \tag{54}$$

$$\mathbf{\Psi}_2 = (\hat{\mathbf{r}}_{1,n}\hat{\mathbf{r}}_{1,n}^T + \hat{\mathbf{r}}_{2,n}\hat{\mathbf{r}}_{2,n}^T) + i(\hat{\mathbf{r}}_{2,n}\hat{\mathbf{r}}_{1,n}^T - \hat{\mathbf{r}}_{1,n}\hat{\mathbf{r}}_{2,n}^T). \tag{55}$$

From (51), (54), and (55), we obtain

$$E\{\hat{\mathbf{r}}_n\hat{\mathbf{r}}_n^T\} = \frac{1}{2}\begin{bmatrix} \mathrm{Re}(E\{\mathbf{\Psi}_1\} + E\{\mathbf{\Psi}_2\}) & \mathrm{Im}(E\{\mathbf{\Psi}_1\} - E\{\mathbf{\Psi}_2\}) \\ \mathrm{Im}(E\{\mathbf{\Psi}_1\} + E\{\mathbf{\Psi}_2\}) & \mathrm{Re}(E\{\mathbf{\Psi}_2\} - E\{\mathbf{\Psi}_1\}) \end{bmatrix}. \tag{56}$$

In (56), $E\{\mathbf{\Psi}_1\}$ and $E\{\mathbf{\Psi}_2\}$ are not difficult to obtain because $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$ are both explicit functions of $\hat{\mathbf{R}}_n$ or, equivalently, $\mathbf{y}(n)\mathbf{y}^H(n)$. We will present this computation result by the following Lemma.

Lemma: Under the following assumptions:

AS1) The noise $v_k(n)$ $(k = 1, \ldots, \nu)$ is circular complex AWGN with zero mean, variance σ_v^2, and independent of $w_j(n)$, $E\{v_k^2(n)\} = 0$.

AS2) $w_j(n)$ is a zero mean, i.i.d. sequence with circularly symmetric complex constellation and variance $E\{|w_j(n)|^2\} = \sigma_w^2$. This means $E\{w_j(n)\} = 0$, $E\{w_j^2(n)\} = 0$, and $E\{w_j^3(n)\} = 0$.

AS3) The fourth-order moment of $w_j(n)$ is $E\{|w_j(n)|^4\} = m_{4w}$.

[7] There will be more terms if $\hat{\mathbf{r}}_n$ is not independent.

$E\{\boldsymbol{\Psi}_1\}$ and $E\{\boldsymbol{\Psi}_2\}$ are functions of $\mathbf{H}$ and σ_v^2 as follows:[8]

$$\begin{aligned}E\{\boldsymbol{\Psi}_1\} =& \sigma_w^4 vec(\mathbf{H}\mathbf{H}^H)vec^T(\mathbf{H}\mathbf{H}^H)\\ &+ (m_{4w} - 2\sigma_w^4)(\mathbf{H}^* \otimes \mathbf{H})\mathbf{B}(\mathbf{H}^H \otimes \mathbf{H}^T)\\ &+ \sigma_w^4(\mathbf{H}^* \otimes \mathbf{H})\mathbf{A}_1(\mathbf{H}^H \otimes \mathbf{H}^T)\\ &+ \sigma_v^4 vec(\mathbf{I}_\nu)vec^T(\mathbf{I}_\nu) + \sigma_v^4\mathbf{A}_3\\ &+ \sigma_w^2\sigma_v^2[vec(\mathbf{I}_\nu)vec^T(\mathbf{H}\mathbf{H}^H)\\ &+ vec(\mathbf{H}\mathbf{H}^H)vec^T(\mathbf{I}_\nu)]\\ &+ \sigma_w^2\sigma_v^2[(\mathbf{H}^* \otimes \mathbf{I}_\nu)\mathbf{A}_2(\mathbf{I}_\nu \otimes \mathbf{H}^T)\\ &+ (\mathbf{I}_\nu \otimes \mathbf{H})\mathbf{A}_2^T(\mathbf{H}^H \otimes \mathbf{I}_\nu)] \end{aligned}\quad (57)$$

$$\begin{aligned}E\{\boldsymbol{\Psi}_2\} =& \sigma_w^4 vec(\mathbf{H}\mathbf{H}^H)vec^H(\mathbf{H}\mathbf{H}^H)\\ &+ (m_{4w} - 2\sigma_w^4)(\mathbf{H}^* \otimes \mathbf{H})\mathbf{B}(\mathbf{H}^T \otimes \mathbf{H}^H)\\ &+ \sigma_w^4(\mathbf{H}^*\mathbf{H}^T) \otimes (\mathbf{H}\mathbf{H}^H)\\ &+ \sigma_v^4 vec(\mathbf{I}_\nu)vec^T(\mathbf{I}_\nu) + \sigma_v^4\mathbf{I}_{\nu^2}\\ &+ \sigma_w^2\sigma_v^2[vec(\mathbf{I}_\nu)vec^H(\mathbf{H}\mathbf{H}^H)\\ &+ vec(\mathbf{H}\mathbf{H}^H)vec^T(\mathbf{I}_\nu)]\\ &+ \sigma_w^2\sigma_v^2[(\mathbf{H}^*\mathbf{H}^T) \otimes \mathbf{I}_\nu + \mathbf{I}_\nu \otimes (\mathbf{H}\mathbf{H}^H)] \end{aligned}\quad (58)$$

where

$$\mathbf{A}_1 = \left[\mathbf{A}_{k,l}^{(1)}\right]_{J\times J}, \qquad \mathbf{A}_{k,l}^{(1)} = \mathbf{s}_l\mathbf{s}_k^T,$$

$$\mathbf{s}_l^T = \Big[0, \ldots, 0, \underbrace{1}_{l\text{th element}}, 0, \ldots, 0\Big]_{1\times J}$$

$$\mathbf{A}_2 = \left[\mathbf{A}_{k,l}^{(2)}\right]_{J\times \nu}, \qquad \mathbf{A}_{k,l}^{(2)} = \mathbf{u}_l\mathbf{s}_k^T,$$

$$\mathbf{u}_l^T = \Big[0, \ldots, 0, \underbrace{1}_{l\text{th element}}, 0, \ldots, 0\Big]_{1\times \nu}$$

$$\mathbf{A}_3 = \left[\mathbf{A}_{k,l}^{(3)}\right]_{\nu\times \nu}, \qquad \mathbf{A}_{k,l}^{(3)} = \mathbf{u}_l\mathbf{u}_k^T,$$

$$\mathbf{B} = \text{diag}\{\mathbf{s}_1\mathbf{s}_1^T, \ldots, \mathbf{s}_J\mathbf{s}_J^T\}.$$

Proof: See the Appendix. □

Since $\mathbf{H}$ is a function of true channel parameters, $E\{\boldsymbol{\Psi}_1\}$ in (57) and $E\{\boldsymbol{\Psi}_2\}$ in (58) can be determined by $\boldsymbol{\theta}_0$. According to these two results and (56), we can obtain $E\{\hat{\mathbf{r}}_n\hat{\mathbf{r}}_n^T\}$. Therefore, the normalized asymptotic covariance $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ of $\hat{\mathbf{z}}_N$ in (50) is explicitly determined by true channel parameters and noise power. It will be used to evaluate the theoretical performance against which experimental results can be compared.

C. *Performance Analysis of Asymptotic Covariance of the Proposed Estimator*

We have shown that the estimator $\hat{\mathbf{g}}_j$ and $\hat{\sigma}_v^2$ are obtained by our two-step approach. One may wonder about the performance loss for our solution. In this part, we will analyze the asymptotic covariance of our estimator for both the true channel vector $\mathbf{g}_{j0}$ of user j and noise power. We first derive the asymptotic covariance of $\hat{\mathbf{x}}_N$ from our first step approach (35) from which $\hat{\mathbf{G}}_j$ in (39) is obtained, and then, we apply perturbation theory to eigenvectors of matrix $\hat{\mathbf{G}}_j$ and illustrate the relation between the perturbation $\hat{\mathbf{g}}_j - \mathbf{g}_{j0}$ and $\hat{\mathbf{x}}_N - \mathbf{x}_0$. Combining these results, we can achieve our goal.

First, let us define the normalized asymptotic covariance of $\hat{\mathbf{x}}_N$

$$\mathbf{COV}_\mathbf{x} = \lim_{N\to\infty} NE\{\Delta\mathbf{x}\Delta\mathbf{x}^T\}, \qquad \Delta\mathbf{x} = \hat{\mathbf{x}}_N - \mathbf{x}_0. \quad (59)$$

Because $\hat{\mathbf{x}}_N$ is determined by $\hat{\mathbf{r}}_N$ in (35), $\mathbf{COV}_\mathbf{x}$ can be expressed as a function of the normalized asymptotic covariance of $\hat{\mathbf{r}}_N$

$$\mathbf{COV}_\mathbf{x} = \mathbf{W}\lim_{N\to\infty} NE\{(\hat{\mathbf{r}}_N - \mathbf{r}_0)(\hat{\mathbf{r}}_N - \mathbf{r}_0)^T\}\mathbf{W}^T \quad (60)$$

where

$$\mathbf{W} = (\boldsymbol{\Pi}^T\mathbf{L}^T\mathbf{M}\mathbf{L}\boldsymbol{\Pi})^{-1}\boldsymbol{\Pi}^T\mathbf{L}^T\mathbf{M}\mathbf{L}. \quad (61)$$

According to (49) and (50), (60) is simplified to

$$\mathbf{COV}_\mathbf{x} = \mathbf{W}(E\{\hat{\mathbf{r}}_n\hat{\mathbf{r}}_n^T\} - \mathbf{r}_0\mathbf{r}_0^T)\mathbf{W}^T. \quad (62)$$

Utilizing the result for $E\{\hat{\mathbf{r}}_n\hat{\mathbf{r}}_n^T\} - \mathbf{r}_0\mathbf{r}_0^T$ in Section IV-B, $\mathbf{COV}_\mathbf{x}$ can be computed from the true parameters in $\boldsymbol{\theta}_0$.

Our estimate of channel vectors are from the eigenvectors of matrix $\hat{\mathbf{G}}_j$ in (39), which is the estimate of true value $\mathbf{G}_{j0} = \sigma_w^2\mathbf{g}_{j0}\mathbf{g}_{j0}^H$. Let us similarly define the error

$$\Delta\mathbf{G}_j = \hat{\mathbf{G}}_j - \mathbf{G}_{j0}$$

in estimating $\mathbf{G}_{j0}$. Then, based on (39), $\Delta\mathbf{G}_j$ can be derived from $\Delta\mathbf{x}$

$$\Delta\mathbf{G}_j = [\mathbf{P}_{1,j}\Delta\mathbf{x}, \ldots, \mathbf{P}_{q+1,j}\Delta\mathbf{x}]. \quad (63)$$

Next, we will investigate how the error in estimating $\mathbf{G}_{j0}$ propagates to the estimate of its eigenvectors.

According to (42) and (43), $\mathbf{G}_{j0}$ has a unique nonzero eigenvalue ξ_1

$$\xi_1 = \sigma_w^2\|\mathbf{g}_{j0}\|^2 \quad (64)$$

and the corresponding eigenvector $\boldsymbol{\beta}_1$ can be chosen as

$$\boldsymbol{\beta}_1 = \frac{\mathbf{g}_{j0}}{\|\mathbf{g}_{j0}\|} \quad (65)$$

whereas all other q eigenvalues are zero with associated eigenvectors $\boldsymbol{\beta}_m$ for $m = 2, \ldots, q+1$. We only focus on $\Delta\boldsymbol{\beta}$, the perturbation on $\boldsymbol{\beta}_1$ corresponding to the maximum eigenvalue ξ_1. According to [16, Ch. 3], $\Delta\boldsymbol{\beta}$ will be asymptotically Gaussian. Therefore, we will only investigate its first-order perturbation. Based on perturbation theory (e.g. [7, Ch. 7]), the first-order perturbation of $\boldsymbol{\beta}_1$ is given by the following expression:

$$\Delta\boldsymbol{\beta} = \frac{1}{\xi_1}\sum_{m=2}^{q+1}(\boldsymbol{\beta}_m^H\Delta\mathbf{G}_j\boldsymbol{\beta}_1)\boldsymbol{\beta}_m. \quad (66)$$

If we denote $\boldsymbol{\beta}_1$ as $\boldsymbol{\beta}_1 = [\beta_{1l}]$, $l = 1, \ldots, q+1$, then from (63), the scalar $\boldsymbol{\beta}_m^H\Delta\mathbf{G}_j\boldsymbol{\beta}_1$ in (66) can be computed

$$\boldsymbol{\beta}_m^H\Delta\mathbf{G}_j\boldsymbol{\beta}_1 = \sum_{l=1}^{q+1}\beta_{1l}(\boldsymbol{\beta}_m^H\mathbf{P}_{l,j}\Delta\mathbf{x}). \quad (67)$$

[8]Here, $\mathbf{H}$ and σ_v^2 refer to their true values. $\mathbf{H}$ is obtained according to (6). For the sake of notational convenience, subscript "o" is omitted.

Hence, $\Delta\boldsymbol{\beta}$ can be derived from $\Delta\mathbf{x}$ by

$$\Delta\boldsymbol{\beta} = \frac{1}{\xi_1}\sum_{m=2}^{q+1}\sum_{l=1}^{q+1}\beta_{1l}(\boldsymbol{\beta}_m^H\mathbf{P}_{l,j}\Delta\mathbf{x})\boldsymbol{\beta}_m. \qquad (68)$$

However, $\Delta\boldsymbol{\beta}$ is complex with its real part $\Delta\boldsymbol{\beta}_R$ and imaginary part $\Delta\boldsymbol{\beta}_I$. We need to construct a new corresponding vector with all real elements

$$\Delta\tilde{\boldsymbol{\beta}}_j = [\Delta\boldsymbol{\beta}_R^T\Delta\boldsymbol{\beta}_I^T]^T$$

and define the normalized asymptotic covariance of $\Delta\tilde{\boldsymbol{\beta}}_j$ as

$$\mathbf{COV}_{\mathbf{g}_j} = \lim_{N\to\infty} NE\left\{\Delta\tilde{\boldsymbol{\beta}}_j\Delta\tilde{\boldsymbol{\beta}}_j^T\right\}.$$

Using the same technique as in Section IV-B, $\mathbf{COV}_{\mathbf{g}_j}$ can be obtained from $\Delta\boldsymbol{\beta}$

$$\mathbf{COV}_{\mathbf{g}_j} = \frac{1}{2}\begin{bmatrix}\mathrm{Re}(\boldsymbol{\Omega}_1+\boldsymbol{\Omega}_2) & \mathrm{Im}(\boldsymbol{\Omega}_1-\boldsymbol{\Omega}_2)\\ \mathrm{Im}(\boldsymbol{\Omega}_1+\boldsymbol{\Omega}_2) & \mathrm{Re}(\boldsymbol{\Omega}_2-\boldsymbol{\Omega}_1)\end{bmatrix} \qquad (69)$$

where

$$\boldsymbol{\Omega}_1 = \lim_{N\to\infty} NE\{\Delta\boldsymbol{\beta}\Delta\boldsymbol{\beta}^T\},$$
$$\boldsymbol{\Omega}_2 = \lim_{N\to\infty} NE\{\Delta\boldsymbol{\beta}\Delta\boldsymbol{\beta}^H\}. \qquad (70)$$

From (68), both $\boldsymbol{\Omega}_1$ and $\boldsymbol{\Omega}_2$ are not difficult to obtain from $\mathbf{COV}_\mathbf{x}$, which gives us

$$\boldsymbol{\Omega}_1 = \frac{1}{\xi_1^2}\sum_{m_1,m_2=2}^{q+1}\sum_{l_1,l_2=1}^{q+1} \cdot[\boldsymbol{\beta}_{m_1}\boldsymbol{\beta}_{m_2}^T\beta_{1l_1}\beta_{1l_2}(\boldsymbol{\beta}_{m_1}^H\mathbf{P}_{l_1,j})\mathbf{COV}_\mathbf{x}(\mathbf{P}_{l_2,j}^T\boldsymbol{\beta}_{m_2}^*)] \qquad (71)$$

$$\boldsymbol{\Omega}_2 = \frac{1}{\xi_1^2}\sum_{m_1,m_2=2}^{q+1}\sum_{l_1,l_2=1}^{q+1} \cdot[\boldsymbol{\beta}_{m_1}\boldsymbol{\beta}_{m_2}^H\beta_{1l_1}\beta_{1l_2}^*(\boldsymbol{\beta}_{m_1}^H\mathbf{P}_{l_1,j})\mathbf{COV}_\mathbf{x}(\mathbf{P}_{l_2,j}^T\boldsymbol{\beta}_{m_2})]. \qquad (72)$$

Therefore, based on the previous result for $\mathbf{COV}_\mathbf{x}$, $\mathbf{COV}_{\mathbf{g}_j}$ can be directly obtained from true parameters $\boldsymbol{\theta}_0$.

Similarly, from (46), the normalized asymptotic covariance of the estimate of noise variance can be derived as

$$\mathbf{COV}_{\sigma_v^2} = [0, \ldots, 0, 1]^T\mathbf{COV}_\mathbf{x}[0, \ldots, 0, 1]. \qquad (73)$$

This concludes the derivation for the normalized asymptotic covariance of all parameter estimates and will serve as a performance measure of the proposed estimator.

D. CRB of Asymptotic Covariance

Our estimator is from a two-step approach. For comparison purpose, we will derive the CRB (e.g., [18]) of asymptotic covariance of optimal estimator. Let us define $\hat{\boldsymbol{\theta}}_{o,N}$ to be the optimal solution to true parameters $\boldsymbol{\theta}_0$ and the asymptotic covariance of $\hat{\boldsymbol{\theta}}_{o,N}$ as

$$\boldsymbol{\Phi}(\boldsymbol{\theta}_0) = \lim_{N\to\infty} NE\left\{\left(\hat{\boldsymbol{\theta}}_{o,N}-\boldsymbol{\theta}_0\right)\left(\hat{\boldsymbol{\theta}}_{o,N}-\boldsymbol{\theta}_0\right)^T\right\}. \qquad (74)$$

Then, it is shown in [16, Ch. 3] that it is determined by

$$\boldsymbol{\Phi}(\boldsymbol{\theta}_0) = \mathbf{U}(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)\mathbf{U}^T(\boldsymbol{\theta}_0) \qquad (75)$$

where, in our context with cost function (33)

$$\mathbf{U}(\boldsymbol{\theta}_0) = [\mathbf{F}^T(\boldsymbol{\theta}_0)\mathbf{M}\mathbf{F}(\boldsymbol{\theta}_0)]^{-1}\mathbf{F}^T(\boldsymbol{\theta}_0)\mathbf{M} \qquad (76)$$
$$\mathbf{F}(\boldsymbol{\theta}) = [\nabla_{\theta_1}\mathbf{z}, \ldots, \nabla_{\theta_K}\mathbf{z}], \qquad \mathbf{F}(\boldsymbol{\theta}_0) = \mathbf{F}(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}. \qquad (77)$$

$\theta_1 \sim \theta_K$ are K successive elements in $\boldsymbol{\theta}$. It is also established in [16, Ch. 3] that when $\mathbf{M} = \boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)$, $\boldsymbol{\Phi}(\boldsymbol{\theta}_0)$ achieves its lower bound $\mathbf{B}(\boldsymbol{\theta}_0)$

$$\mathbf{B}(\boldsymbol{\theta}_0) = [\mathbf{F}^T(\boldsymbol{\theta}_0)\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{F}(\boldsymbol{\theta}_0)]^{-1}. \qquad (78)$$

To obtain the expression for $\mathbf{B}(\boldsymbol{\theta}_0)$, $\mathbf{F}(\boldsymbol{\theta}_0)$ needs to be evaluated, which requires the knowledge of $\nabla_{\theta_j}\mathbf{z}$. These derivatives are not hard to evaluate because $\mathbf{z}$ in (32) is an explicit function of $\boldsymbol{\theta}$. Since

$$vec(\boldsymbol{\theta}\boldsymbol{\theta}^T) = [\theta_1\boldsymbol{\theta}^T, \ldots, \theta_K\boldsymbol{\theta}^T]^T$$

we have

$$\begin{aligned}\nabla_{\theta_j} vec(\boldsymbol{\theta}\boldsymbol{\theta}^T) &= [\theta_1\mathbf{e}_j^T, \ldots, \theta_{j-1}\mathbf{e}_j^T, \boldsymbol{\theta}^T\\ &\quad + \theta_j\mathbf{e}_j^T, \theta_{j+1}\mathbf{e}_j^T, \ldots, \theta_K\mathbf{e}_j^T]^T\\ &= [\mathbf{0}^T, \ldots, \mathbf{0}^T, \boldsymbol{\theta}^T, \mathbf{0}^T, \ldots, \mathbf{0}^T]^T\\ &\quad + vec[\theta_1\mathbf{e}_j^T, \ldots, \theta_K\mathbf{e}_j^T]\\ &= \mathbf{e}_j\otimes\boldsymbol{\theta} + vec(\boldsymbol{\theta}^T\otimes\mathbf{e}_j). \end{aligned} \qquad (79)$$

Therefore

$$\nabla_{\theta_j}\mathbf{z} = \mathbf{L}\boldsymbol{\Pi}\begin{bmatrix}\mathbf{T}_r & \mathbf{O}\\ \mathbf{T}_i & \mathbf{O}\\ \mathbf{0}^T & \mathbf{e}_K^T\end{bmatrix}\begin{bmatrix}\mathbf{e}_j\otimes\boldsymbol{\theta} + vec(\boldsymbol{\theta}^T\otimes\mathbf{e}_j)\\ \mathbf{e}_j\end{bmatrix}. \qquad (80)$$

From (77) and (80), we obtain

$$\mathbf{F}(\boldsymbol{\theta}) = \mathbf{L}\boldsymbol{\Pi}\mathbf{T}(\boldsymbol{\theta}) \qquad (81)$$

where

$$\mathbf{T}(\boldsymbol{\theta}) = \begin{bmatrix}\mathbf{T}_r & \mathbf{O}\\ \mathbf{T}_i & \mathbf{O}\\ \mathbf{0}^T & \mathbf{e}_K^T\end{bmatrix}\begin{bmatrix}\mathbf{I}_K\otimes\boldsymbol{\theta} + \boldsymbol{\theta}\otimes\mathbf{I}_K)\\ \mathbf{I}_K\end{bmatrix}. \qquad (82)$$

Hence

$$\mathbf{B}(\boldsymbol{\theta}_0) = [(\mathbf{L}\boldsymbol{\Pi}\mathbf{T}(\boldsymbol{\theta}_0))^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)\mathbf{L}\boldsymbol{\Pi}\mathbf{T}(\boldsymbol{\theta}_0)]^{-1}. \qquad (83)$$

Based on our result for $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$ in Section IV-B, the CRB in (83) can be evaluated directly from true parameters $\boldsymbol{\theta}_0$.

As far as the lower bound for a particular channel vector or noise power is concerned, we can partition $\mathbf{B}(\boldsymbol{\theta}_0)$ into $(2J+1)\times(2J+1)$ blocks with the (m, l)th block entry as $\mathbf{B}_{ml}$ for $m, l = 1, \ldots 2J+1$. According to our definition of parameter vector $\boldsymbol{\theta}$, the entry $\mathbf{B}_{ml}$ has dimension $(q+1)\times(q+1)$ for $m, l = 1, \ldots 2J$ and is a scalar for $m = l = 2J+1$. Then, for the interest of channel estimate for user j, the bound for its real and imaginary parts can be obtained by taking out the (j, j)th and $(J+j, J+j)$th block matrices of $\mathbf{B}(\boldsymbol{\theta}_0)$, respectively.

Similarly, the last element of the last column of $\mathbf{B}(\boldsymbol{\theta}_0)$ will be the lower bound for the estimate of noise variance.

V. Asymptotically Near-Optimal Estimator

As shown in [16, Ch. 3], if the weighting matrix $\mathbf{M}$ in our cost function (33) is equal to the inverse of asymptotic covariance matrix of $\hat{\mathbf{z}}_N$ ($\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$), then the asymptotic optimality of our estimator for $\boldsymbol{\theta}_0$ can be achieved by considering the cost function as a function of $\boldsymbol{\theta}$ instead of $\mathbf{x}$. Now, due to the high nonlinearity of this cost function with respect to $\boldsymbol{\theta}$, difficulty will be encountered in obtaining a closed-form solution for $\boldsymbol{\theta}_0$. One can apply the gradient decent method to iteratively search the minimizer of (33), as discussed in [4]. This adaptive implementation requires further investigation, which is beyond the scope of the current paper.

It is established in [16, Ch. 3] that the same asymptotic performance can be achieved when

$$\mathbf{M} = \hat{\boldsymbol{\Sigma}}^{-1}(\boldsymbol{\theta}_0) \tag{84}$$

where $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)$ is the consistent estimate of $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$. The asymptotically optimal estimator for $\mathbf{x}_0$ is then determined by (35). For the current problem, however, our approach to estimate channel vectors is still near-optimal due to the fact of two-step procedure, but substantial improvement can be achieved by choosing a weighting matrix as (84), as will be seen in our simulation results. This requires the sample estimate $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)$, which is not hard to derive.

In (50), $\mathbf{r}_0 = E\{\hat{\mathbf{r}}_n\}$. We can use sample averages to approximate $E\{\hat{\mathbf{r}}_n\hat{\mathbf{r}}_n^T\}$ and $\mathbf{r}_0$. Then, it becomes an easier task to obtain $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)$ from samples, which gives

$$\begin{aligned}\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0) = &\mathbf{L}\left[\frac{1}{N}\sum_{n=1}^{N}(\hat{\mathbf{r}}_n\hat{\mathbf{r}}_n^T)\right]\mathbf{L}^T\\ &- \mathbf{L}\left(\frac{1}{N}\sum_{n_1=1}^{N}\hat{\mathbf{r}}_{n_1}\right)\left(\frac{1}{N}\sum_{n_2=1}^{N}\hat{\mathbf{r}}_{n_2}^T\right)\mathbf{L}^T.\end{aligned} \tag{85}$$

It is easy to show that $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)$ in (85) is asymptotically consistent with $\boldsymbol{\Sigma}(\boldsymbol{\theta}_0)$. From (14), $\hat{\mathbf{r}}_n$ is determined by the observed data $\mathbf{y}(n)$ as follows:

$$\hat{\mathbf{r}}_n = \begin{bmatrix} vec[\mathrm{Re}(\mathbf{y}(n)\mathbf{y}^H(n))] \\ vec[\mathrm{Im}(\mathbf{y}(n)\mathbf{y}^H(n))] \end{bmatrix}. \tag{86}$$

Substituting (86) in (85), we arrive at our estimated asymptotic covariance matrix $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)$ directly from samples $\mathbf{y}(n)$.

Since the inverse of $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)$ is required for $\mathbf{M}$ according to (84), if the matrix $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)$ is ill conditioned for some data points, replacing its negative eigenvalues with zero and using pseudo-inverse are highly recommended [4].

VI. Discussions

We have developed a correlation-based channel estimation method. Its advantages and disadvantages should be clarified. In addition, comparisons with existing methods need to be made.

As an SOS-based method, our method provides an asymptotically minimum variance estimator if the weighting matrix is optimal. A closed-form expression for the asymptotic variance of the channel estimator has been derived. It can be evaluated for a given communication system and can be used to compare with both other methods and the experimental performance of the proposed method, as will be seen in our simulations. Since the optimal weighting matrix $\boldsymbol{\Sigma}^{-1}(\boldsymbol{\theta}_0)$ depends on true values of the unknown parameters, it is not *a priori* known. However, this matrix has been shown to be directly estimated from received data samples, achieving the same asymptotic performance. Thus, if there is sufficiently large amount of data ($N \to \infty$), then compared with the optimal estimator, performance loss from the proposed estimator becomes negligible due to the consistency of estimate for the weighting matrix. It is not surprising that significant performance gain over subspace approaches [1], [19] is observed in our simulations when N is large. However, long data records are required for the method to approach its asymptotic performance, as shown by our experimental results and comparisons with subspace methods.

As a tradeoff, a certain amount of data needs to be accumulated before estimation in order for the matrix $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)$ to be full rank. In addition, the computation is expensive since the inverse of $\hat{\boldsymbol{\Sigma}}(\boldsymbol{\theta}_0)$ is not easily updated and thus has to be computed at each time. This matrix has dimension $\nu^2 \times \nu^2$, where ν is the length of the data vector. Thus, the complexity is about $O(\nu^6)$. Compared with the subspace method, it is tremendous where the estimate of the noise subspace requires only computations in $O(\nu^3)$. The complexity of our method also depends on the number of users in the system. However, this cannot be avoided if all channels are estimated simultaneously. The proposed method is more desirable if performance instead of implementation cost is of particular concern.

The current paper aims at multiuser channel estimation. It requires more information for all channels to be identifiable than those in subspace methods [1], [19] with the interest of only a particular channel at a time. We have assumed that the system is quasisynchronous [9] and that all users' spreading codes and delays are known. In [19], the delay is assumed to be multiples of chip period (thus chip delay) and estimated in the initial synchronization by extensive searching from a set with finite elements. It can be seen from [1] that even a single delay is very difficult to estimate if it can take an arbitrary value. There is no doubt that estimation for delays for all users is still an interesting topic in the literature. Joint estimation of delays and channels based on correlation matching techniques need further investigation. In addition, channel order also plays an important role in our estimation, similar to the discussion in [19]. With this information, our blind estimator can be obtained under identifiability conditions.

VII. Simulations

Our experiment is performed to test the proposed methods in Sections III and V and make comparisons both with our theoretical analysis in Section IV and with other SOS-based methods such as subspace methods [1], [19]. The performance of these methods is measured in terms of the mean square error (MSE) of estimators.

Five users in a DS-CDMA system are simulated to transmit 1000 i.i.d. $\{\pm 1 \pm 1i\}$ symbols with equal power by QPSK modulation. The 15-dB AWGN is added to the input. These inputs are spread by independently and randomly generated binary short codes $\{\pm 1\}$ with spreading factor $P = 16$. Different chip sequences pass through different multipath channels with order $q = 2$ for five users. Channel vectors of all users are arbitrarily selected and normalized as follows:

$$\begin{aligned} \mathbf{g}_1 =& [0.3669 - 0.5007i,\ -0.3791 + 0.3631i \\ & -0.1423 - 0.5647i]^T \\ \mathbf{g}_2 =& [-0.6351 + 0.3604i,\ -0.0777 - 0.3158i \\ & 0.5959 + 0.0770i]^T \\ \mathbf{g}_3 =& [0.0303 + 0.3791i,\ 0.4449 + 0.0545i \\ & 0.6081 - 0.5335i]^T \\ \mathbf{g}_4 =& [-0.4922 + 0.0284i,\ -0.6260 + 0.4380i \\ & -0.3944 + 0.1331i]^T \\ \mathbf{g}_5 =& [-0.4131 - 0.3677i,\ -0.5045 + 0.4436i \\ & 0.1501 - 0.4694i]^T. \end{aligned}$$

Assume all signals from different users arrive at the matched filter simultaneously. Upon its output, we collect $\nu = 14$ samples for each symbol period, as described in Section II. Without loss of generality, we restrict our interest in presenting only the result for one of five users (e.g., channel 1 for user 1). We express $\mathbf{g}_1$ in a complex form $\mathbf{g}_1 = \mathbf{g}_{1r} + i\mathbf{g}_{1i}$ with real part $\mathbf{g}_{1r} = [g_{r0},\ g_{r1},\ g_{r2}]^T$ and imaginary part $\mathbf{g}_{1i} = [g_{i0},\ g_{i1},\ g_{i2}]^T$. Consider the MSE of them according to

$$\text{MSE}(\mathbf{g}_{1r},\ N) = \frac{1}{K_1} \sum_{k=1}^{K_1} \|\hat{\mathbf{g}}_{1r}(N,\ k) - \mathbf{g}_{1r}\|^2$$
$$\text{MSE}(\mathbf{g}_{1i},\ N) = \frac{1}{K_1} \sum_{k=1}^{K_1} \|\hat{\mathbf{g}}_{1i}(N,\ k) - \mathbf{g}_{1i}\|^2$$

for N received data vectors and total K_1 Monte Carlo runs, where $\hat{\mathbf{g}}_{1r}(N,\ k)$ and $\hat{\mathbf{g}}_{1i}(N,\ k)$ are the estimates in the kth run. In the experiment, $K_1 = 50$. As discussed in Section III, $\mathbf{g}_1$ can only be estimated within a phase ambiguity ϕ_1 as $e^{i\phi_1}\boldsymbol{\alpha}_1$. In order to obtain a reasonable comparison between the channel and its estimate, we use the first coefficient of $\mathbf{g}_1$ $(g_{r0} + ig_{i0})$ as a phase reference after the consideration that three estimated elements in $\mathbf{g}_1$ may have different phase ambiguity.

Fig. 2 shows the MSE of estimate for the real part $\mathbf{g}_{1r}$ versus number of received data vectors N. In this figure, we consider three cases with different weighting matrices: optimal $\mathbf{M} = \mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_0)$ computed from (50), estimated $\mathbf{M} = \hat{\mathbf{\Sigma}}^{-1}(\boldsymbol{\theta}_0)$ obtained from (85), and $\mathbf{M} = \mathbf{I}$. In computing $\mathbf{\Sigma}(\boldsymbol{\theta}_0)$, results from the Lemma in Section IV-B are required. The corresponding simulation results are presented as a dashed line, a solid line, and a dashed–dotted line, respectively. In addition, we compute the theoretical limits for $\mathbf{M} = \mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_0)$ and $\mathbf{M} = \mathbf{I}$ cases according to (69) and show them by circles and stars, respectively, after being scaled by a factor $1/N$. The CRB is found from

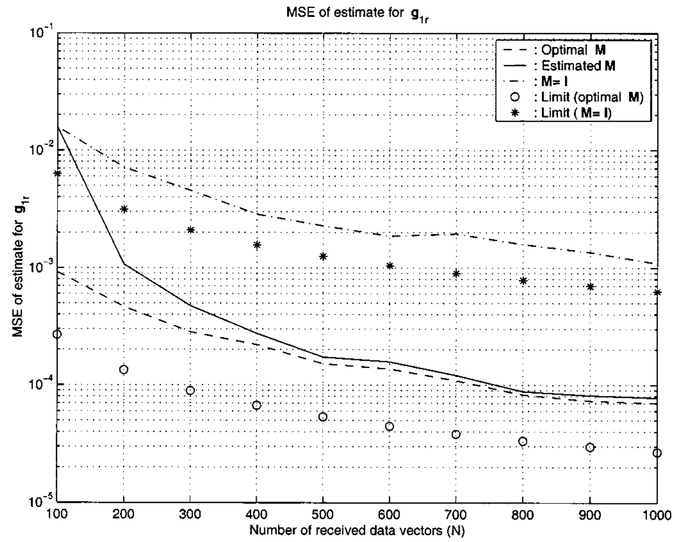

Fig. 2. MSE of estimate for the real part of one channel (channel 1).

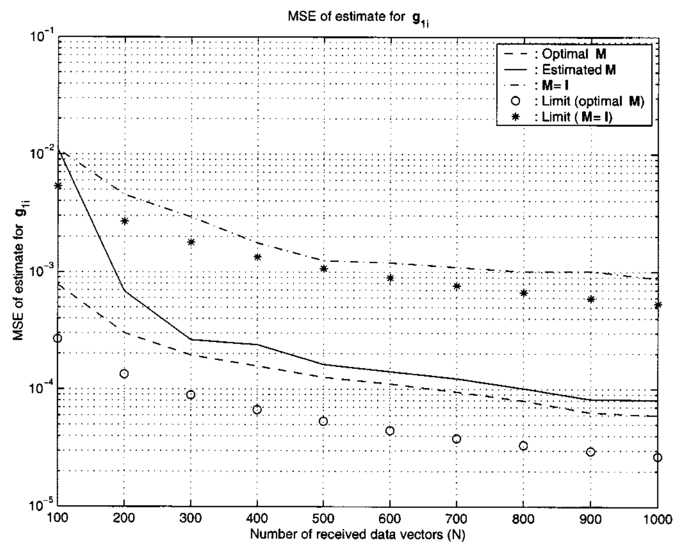

Fig. 3. MSE of estimate for the imaginary part of one channel (channel 1).

(83) as the trace of a corresponding submatrix and computed as 2.6×10^{-5}. It can be seen that the experimental MSEs are very close to their limits. The MSE for $\mathbf{M} = \mathbf{I}$ reaches 9×10^{-3} after 1000 bit periods, which is very close to its limit 5×10^{-3}. However, they are both much higher than the MSE for other two cases. The MSE for $\mathbf{M} = \hat{\mathbf{\Sigma}}^{-1}(\boldsymbol{\theta}_0)$ reaches 8×10^{-5}, which results in only a little penalty compared with the case $\mathbf{M} = \mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_0)$ at 6×10^{-5} level. It indicates that satisfactory performance is achieved by using sample estimate $\hat{\mathbf{\Sigma}}(\boldsymbol{\theta}_0)$. As expected, errors for these two latter cases are still above their limit but are not so significant. For example, at $N = 1000$, the limit is 3×10^{-5}. Similar results can be observed in Fig. 3 for the estimate of the imaginary part $\mathbf{g}_{1i}$. The lower bound is found to be 2.2×10^{-5}. The error levels are also comparable with those in Fig. 2 for corresponding cases. Since we can obtain the estimate of noise variance simultaneously, we also present the MSE of its estimate in Fig. 4. Definitions of lines and symbols in this figure are same as in Fig. 2. This time, the experimental results

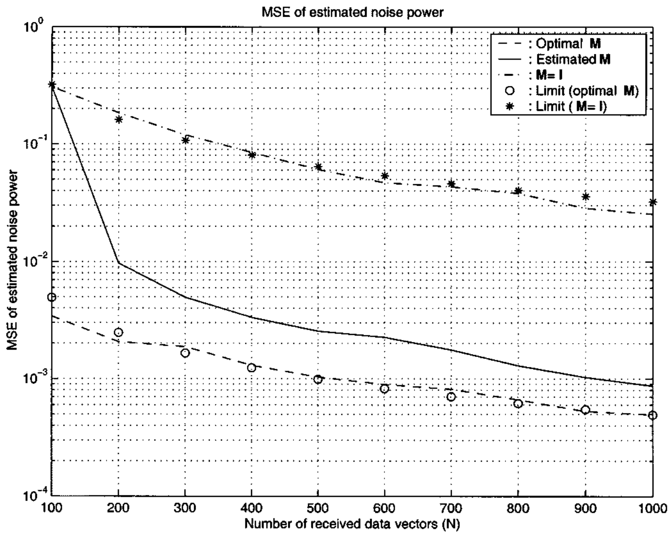

Fig. 4. MSE of estimated noise power.

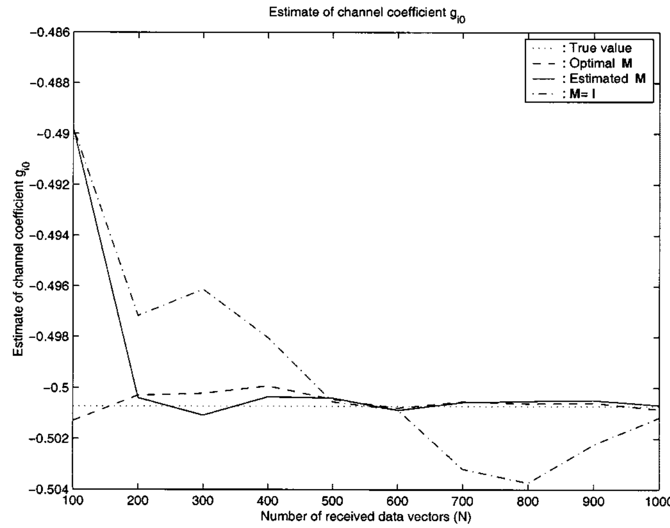

Fig. 6. Estimate of the imaginary part of the first channel coefficient g_{1i}.

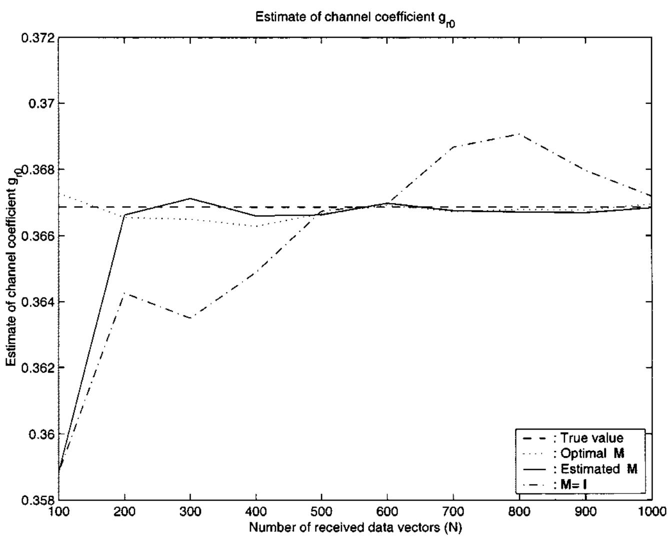

Fig. 5. Estimate of the real part of the first channel coefficient g_{1r}.

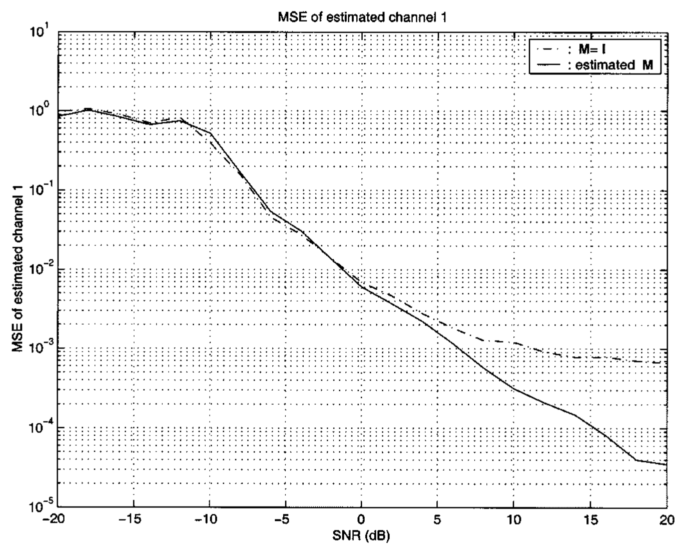

Fig. 7. Effect of noise on the channel estimation error.

with $\mathbf{M} = \mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_0)$ and $\mathbf{M} = \mathbf{I}$ approaches their limits [computed from (73)], respectively, but compared with channel estimation errors in previous figures, the overall MSE level in this figure is higher. The best one at $N = 1000$ is still 5×10^{-4} when $\mathbf{M} = \mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_0)$. The reason behind this is that to estimate a smaller value (the noise variance in the current context) usually results in a large variance in estimation.

We have shown that our estimator is unbiased. The estimate will asymptotically converge to the true value of the parameter. To demonstrate this point, we present the estimate for one channel coefficient (e.g., the first element in $\mathbf{g}_1$). Fig. 5 shows the average value of the estimate for its real part g_{r0}. Its true value of 0.3669 is plotted by a dashed line, the estimate with optimal $\mathbf{M} = \mathbf{\Sigma}^{-1}(\boldsymbol{\theta}_0)$ is represented by a dotted line, with $\mathbf{M} = \hat{\mathbf{\Sigma}}^{-1}(\boldsymbol{\theta}_0)$ represented by a solid line, and $\mathbf{M} = \mathbf{I}$ by a dashed–dotted line. We can see that when an optimal weighting matrix is used, the estimate has a very small error. When $\mathbf{M}$ is estimated from data samples, this average estimate falls far away from its true value at the beginning due to the effect of small number of samples. As more data samples are collected, it gets closer to its true value because $\hat{\mathbf{\Sigma}}(\boldsymbol{\theta}_0)$ converges to $\mathbf{\Sigma}(\boldsymbol{\theta}_0)$. Therefore, with increasingly more data vectors available, the estimation precess behaves similarly to the case with optimal weighting matrix. Both cases will provide reliable estimates of channel parameters after 500 bit periods. However, when $\mathbf{M} = \mathbf{I}$, the estimate oscillates around the dashed line with larger variations. The same conclusion can be made from the result for the first imaginary coefficient g_{i0} in Fig. 6, which is not illustrated in detail.

In order to see how the AWGN noise in the communication system affects the performance of our estimator, we plot in Fig. 7 the MSEs ($N = 1000$) of estimated channel $\hat{\mathbf{g}}_1$ for a large range of signal-to-noise ratios (SNRs) for two cases: $\mathbf{M} = \hat{\mathbf{\Sigma}}^{-1}(\boldsymbol{\theta}_0)$ and $\mathbf{M} = \mathbf{I}$. It is observed that when SNR $>$ -10 dB, both estimators are improved with increased SNR. At SNR $=$ 0 dB, the MSEs are less than 10^{-2}. After the point SNR $=$ 0 dB, the importance of the weighting matrix becomes

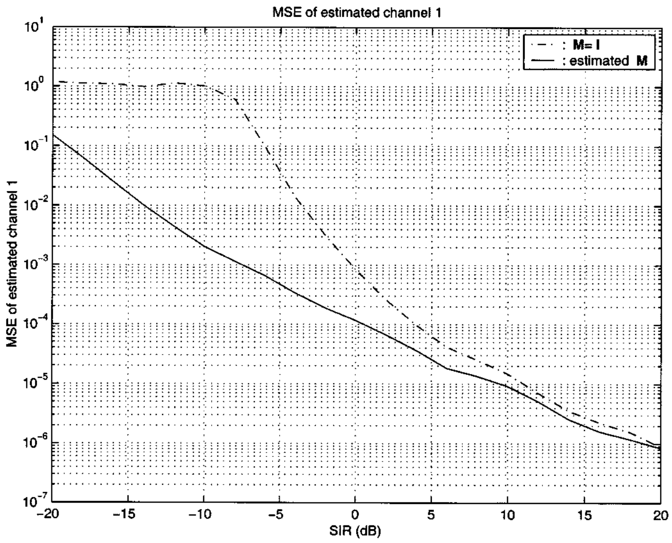

Fig. 8. Effect of multiuser interference on the channel estimation error.

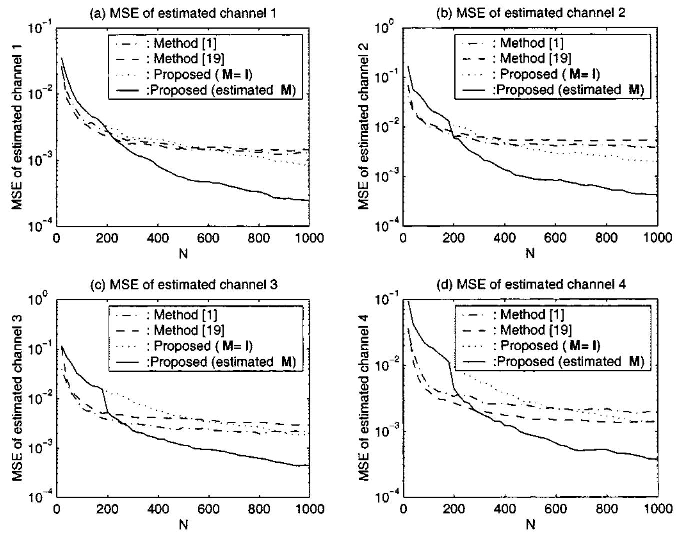

Fig. 9. Comparison of the proposed method with subspace methods [1], [19].

clear since the difference between performance of these two estimators becomes larger. With an estimated optimal weighting matrix, better performance is achieved. The proposed estimator is also tested in different near-far scenarios. The MSE is plotted in Fig. 8 as a function of signal-to-interference ratio (SIR). User 1 is assumed to be the desired user. The interference is from other four equally powered users. SNR $= 15$ dB, and $N = 1000$. It can be seen that the estimator with estimated optimal weighting matrix gives a better result than the estimator with $\mathbf{M} = \mathbf{I}$. When the desired user is 10 dB weaker, the MSE is still satisfactory at a level of 2×10^{-3}. When this user has a higher power (e.g., 10 dB stronger), the MSE becomes much smaller (10^{-5}). This figure shows that although the MSE depends on the SIR, satisfactory results are obtained for a large range of SIRs.

Finally, we compare our method with subspace methods [1], [19], which are also SOS based. We test the effect of amount of data on the channel estimation error. The spreading factor is 16. Each of four users with equal power in the system experiences individual frequency-selective channels (randomly generated)

$$\begin{aligned}\mathbf{g}_1 &= [0.2622 - 0.2606i,\ -0.6233 + 0.4933i\\ &\quad 0.0946 + 0.4718i]^T\\ \mathbf{g}_2 &= [0.0045 + 0.4133i,\ -0.1198 - 0.6520i\\ &\quad -0.3151 - 0.5389i]^T\\ \mathbf{g}_3 &= [0.5625 + 0.5805i,\ -0.2751 + 0.4191i\\ &\quad 0.2837 - 0.1216i]^T\\ \mathbf{g}_4 &= [0.3804 + 0.5792i,\ 0.3975 - 0.0067i\\ &\quad -0.3929 - 0.4554i]^T.\end{aligned}$$

The noise level is SNR $= 10$ dB. In implementing [19], the smoothing factor is set to be 3. The MSE's for these four channels with a different number of data vectors are plotted in Fig. 9(a)–(d), respectively. Dashed–dotted lines are based on [1], dashed lines are obtained from [19], dotted lines are the proposed with $\mathbf{M} = \mathbf{I}$, and solid lines are for the proposed method with $\mathbf{M} = \hat{\mathbf{\Sigma}}^{-1}(\boldsymbol{\theta}_0)$. Some conclusions can be easily made from this figure. First, two subspace methods show similar performance. They provide better estimates than the proposed methods for a shorter data records ($N < 200$). Second, when $\mathbf{M} = \mathbf{I}$, the proposed method gives as good as [Fig. 9(c)–(d)] or even better results [Fig. 9(a)–(b)] than the subspace methods after some time periods ($N > 600$). The advantage of the proposed method with the estimated asymptotically optimal weighting matrix over all other methods can be observed for a large number of data samples. Therefore, asymptotically superior performance is obtained.

VIII. CONCLUSIONS

The weighted correlation matching techniques have been applied to estimate multipath channel parameters in a multiuser CDMA system. The weighting matrix can be arbitrarily chosen as long as it is positive definite. The simplest way is to set it equal to identity matrix, but without any optimality. Therefore, we investigate other possible choices with improved performance. Due to overparameterization, our estimator for channel vectors achieves near-optimal performance with an asymptotically near-optimal weighting matrix. This weighting matrix is further derived as a function of true channel parameters and noise variance in a closed form and is also able to be estimated directly from output samples. Meanwhile, the asymptotic covariance of our estimator is derived, and its lower bound is presented. It turns out that our simulation results are highly consistent with our performance analysis.

APPENDIX
PROOF OF LEMMA

It is very straightforward to consider $\hat{\mathbf{R}}_n$ first since it is present in both $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$ according to (52), (53). Based on the system's input/output relationship (5) and the definition in (14), it can be expressed as

$$\begin{aligned}\hat{\mathbf{R}}_n =& \mathbf{H}\mathbf{w}(n)\mathbf{w}^H(n)\mathbf{H}^H + \mathbf{v}(n)\mathbf{w}^H(n)\mathbf{H}^H\\ &+ \mathbf{H}\mathbf{w}(n)\mathbf{v}^H(n) + \mathbf{v}(n)\mathbf{v}^H(n).\end{aligned} \tag{87}$$

Then, the vec operation is performed on it (see [10, Ch. 12]) to meet the requirement in $\mathbf{\Psi}_1$ and $\mathbf{\Psi}_2$

$$\begin{aligned} vec\left(\hat{\mathbf{R}}_n\right) &= (\mathbf{H}^* \otimes \mathbf{H}) vec(\mathbf{w}(n)\mathbf{w}^H(n)) + vec(\mathbf{v}(n)\mathbf{w}^H(n)\mathbf{H}^H) \\ &\quad + vec(\mathbf{H}\mathbf{w}(n)\mathbf{v}^H(n)) + vec(\mathbf{v}(n)\mathbf{v}^H(n)). \end{aligned} \tag{88}$$

According to definitions of $\mathbf{\Psi}_1$ in (52), $\mathbf{\Psi}_2$ in (53), and assumptions on $w_j(n)$ and $\mathbf{v}(n)$ in the Lemma, only following terms survive in $E\{\mathbf{\Psi}_1\}$ and $E\{\mathbf{\Psi}_2\}$

$$\begin{aligned} E\{\mathbf{\Psi}_1\} &= (\mathbf{H}^* \otimes \mathbf{H}) E\{vec(\mathbf{w}(n)\mathbf{w}^H(n)) \\ &\qquad \cdot vec^T(\mathbf{w}(n)\mathbf{w}^H(n))\}(\mathbf{H}^H \otimes \mathbf{H}^T) \\ &\quad + (\mathbf{H}^* \otimes \mathbf{H}) E\{vec(\mathbf{w}(n)\mathbf{w}^H(n)) vec^T(\mathbf{v}(n)\mathbf{v}^H(n))\} \\ &\quad + E\{vec(\mathbf{v}(n)\mathbf{w}^H(n)\mathbf{H}^H) vec^T(\mathbf{H}\mathbf{w}(n)\mathbf{v}^H(n))\} \\ &\quad + E\{vec(\mathbf{H}\mathbf{w}(n)\mathbf{v}^H(n)) vec^T(\mathbf{v}(n)\mathbf{w}^H(n)\mathbf{H}^H)\} \\ &\quad + E\{vec(\mathbf{v}(n)\mathbf{v}^H(n)) vec^T(\mathbf{w}(n)\mathbf{w}^H(n))\}(\mathbf{H}^H \otimes \mathbf{H}^T) \\ &\quad + E\{vec(\mathbf{v}(n)\mathbf{v}^H(n)) vec^T(\mathbf{v}(n)\mathbf{v}^H(n))\} \end{aligned} \tag{89}$$

$$\begin{aligned} E\{\mathbf{\Psi}_2\} &= (\mathbf{H}^* \otimes \mathbf{H}) E\{vec(\mathbf{w}(n)\mathbf{w}^H(n)) \\ &\qquad \cdot vec^H(\mathbf{w}(n)\mathbf{w}^H(n))\}(\mathbf{H}^T \otimes \mathbf{H}^H) \\ &\quad + (\mathbf{H}^* \otimes \mathbf{H}) E\{vec(\mathbf{w}(n)\mathbf{w}^H(n)) vec^H(\mathbf{v}(n)\mathbf{v}^H(n))\} \\ &\quad + E\{vec(\mathbf{v}(n)\mathbf{w}^H(n)\mathbf{H}^H) vec^H(\mathbf{v}(n)\mathbf{w}^H(n)\mathbf{H}^H)\} \\ &\quad + E\{vec(\mathbf{H}\mathbf{w}(n)\mathbf{v}^H(n)) vec^H(\mathbf{H}\mathbf{w}(n)\mathbf{v}^H(n))\} \\ &\quad + E\{vec(\mathbf{v}(n)\mathbf{v}^H(n)) vec^H(\mathbf{w}(n)\mathbf{w}^H(n))\}(\mathbf{H}^T \otimes \mathbf{H}^H) \\ &\quad + E\{vec(\mathbf{v}(n)\mathbf{v}^H(n)) vec^H(\mathbf{v}(n)\mathbf{v}^H(n))\} \end{aligned} \tag{90}$$

while all other terms are zeroed out.

The following step is to evaluate each term on the right-hand side of (89) and (90). In order to simplify notations, we successively denote those six matrices inside the expectation symbol "E" on the right-hand side of (89) as $\mathbf{T}_1$ to $\mathbf{T}_6$. Similarly, $\overline{\mathbf{T}}_1$ to $\overline{\mathbf{T}}_6$ for (90). Let us first focus on $\mathbf{T}_1$ and $\overline{\mathbf{T}}_1$. As we can see, the expansion of $vec(\mathbf{w}(n)\mathbf{w}^H(n))$ is a prerequisite

$$vec(\mathbf{w}(n)\mathbf{w}^H(n)) = \begin{bmatrix} w_1^*(n)\mathbf{w}(n) \\ \vdots \\ w_J^*(n)\mathbf{w}(n) \end{bmatrix}. \tag{91}$$

Then, the (j, j)th diagonal block of $\mathbf{T}_1$ has a typical form $w_j^*(n)w_j^*(n)\mathbf{w}(n)\mathbf{w}^T(n)$, whereas its typical (k, l)th off-diagonal block $(k \neq l)$ is $w_k^*(n)w_l^*(n)\mathbf{w}(n)\mathbf{w}^T(n)$. According to our assumptions, their expected values can be obtained as

$$\begin{aligned} E\{w_j^*(n)w_j^*(n)\mathbf{w}(n)\mathbf{w}^T(n)\} &= m_{4w}\mathbf{s}_j\mathbf{s}_j^T, \\ E\{w_k^*(n)w_l^*(n)\mathbf{w}(n)\mathbf{w}^T(n)\} &= \sigma_w^4(\mathbf{s}_k\mathbf{s}_l^T + \mathbf{s}_l\mathbf{s}_k^T). \end{aligned}$$

If we write $m_{4w}\mathbf{s}_j\mathbf{s}_j^T$ as $(m_{4w} - 2\sigma_w^4)\mathbf{s}_j\mathbf{s}_j^T + 2\sigma_w^4\mathbf{s}_j\mathbf{s}_j^T$, we arrive at

$$E\{\mathbf{T}_1\} = (m_{4w} - 2\sigma_w^4)\mathbf{B} + \sigma_w^4\mathbf{A}_1 + \sigma_w^4\mathbf{X} \tag{92}$$

where

$$\mathbf{X} = [\mathbf{X}_{k,l}]_{J \times J}, \qquad \mathbf{X}_{k,l} = \mathbf{s}_k\mathbf{s}_l^T.$$

Since $vec^T(\mathbf{I}_J) = [\mathbf{s}_1^T, \ldots, \mathbf{s}_J^T]$, then $\mathbf{X} = vec(\mathbf{I}_J)vec^T(\mathbf{I}_J)$. Therefore

$$E\{\mathbf{T}_1\} = (m_{4w} - 2\sigma_w^4)\mathbf{B} + \sigma_w^4\mathbf{A}_1 + \sigma_w^4 vec(\mathbf{I}_J)vec^T(\mathbf{I}_J). \tag{93}$$

Similarly, the (j, j)th diagonal block of $\overline{\mathbf{T}}_1$ is $w_j(n) w_j^*(n)\mathbf{w}(n)\mathbf{w}^H(n)$, and its typical (k, l)th off-diagonal block $(k \neq l)$ is $w_k^*(n)w_l(n)\mathbf{w}(n)\mathbf{w}^H(n)$. Their expected values are

$$\begin{aligned} E\{w_j(n)w_j^*(n)\mathbf{w}(n)\mathbf{w}^H(n)\} &= (m_{4w} - \sigma_w^4)\mathbf{s}_j\mathbf{s}_j^T + \sigma_w^4\mathbf{I}_J \\ E\{w_k^*(n)w_l(n)\mathbf{w}(n)\mathbf{w}^H(n)\} &= \sigma_w^4\mathbf{s}_k\mathbf{s}_l^T \end{aligned}$$

respectively. If we write $(m_{4w} - \sigma_w^4)\mathbf{s}_j\mathbf{s}_j^T$ as $(m_{4w} - 2\sigma_w^4)\mathbf{s}_j\mathbf{s}_j^T + \sigma_w^4\mathbf{s}_j\mathbf{s}_j^T$, we get

$$E\left\{\overline{\mathbf{T}}_1\right\} = (m_{4w} - 2\sigma_w^4)\mathbf{B} + \sigma_w^4\mathbf{I}_{J^2} + \sigma_w^4 vec(\mathbf{I}_J)vec^T(\mathbf{I}_J). \tag{94}$$

By employing the property of "$\otimes$" [10, Ch. 12] and using (93), the first term on the right-hand side of (89) becomes

$$\begin{aligned} &\text{first term of } E\{\mathbf{\Psi}_1\} \\ &\quad = \sigma_w^4 vec(\mathbf{H}\mathbf{H}^H)vec^T(\mathbf{H}\mathbf{H}^H) \\ &\qquad + \sigma_w^4(\mathbf{H}^* \otimes \mathbf{H})\mathbf{A}_1(\mathbf{H}^H \otimes \mathbf{H}^T) \\ &\qquad + (m_{4w} - 2\sigma_w^4)(\mathbf{H}^* \otimes \mathbf{H})\mathbf{B}(\mathbf{H}^H \otimes \mathbf{H}^T). \end{aligned} \tag{95}$$

Substituting (94) into the right-hand side of (90), its first term is

$$\begin{aligned} &\text{first term of } E\{\mathbf{\Psi}_2\} \\ &\quad = \sigma_w^4 vec(\mathbf{H}\mathbf{H}^H)vec^H(\mathbf{H}\mathbf{H}^H) \\ &\qquad + \sigma_w^4(\mathbf{H}^*\mathbf{H}^T) \otimes (\mathbf{H}\mathbf{H}^H) \\ &\qquad + (m_{4w} - 2\sigma_w^4)(\mathbf{H}^* \otimes \mathbf{H})\mathbf{B}(\mathbf{H}^T \otimes \mathbf{H}^H). \end{aligned} \tag{96}$$

Since $w_j(n)$ and $\mathbf{v}(n)$ are independent, the computation for $E\{\mathbf{T}_2\}$ becomes easier

$$\begin{aligned} E\{\mathbf{T}_2\} &= vec(E\{\mathbf{w}(n)\mathbf{w}^H(n)\})vec^T(E\{\mathbf{v}(n)\mathbf{v}^H(n)\}) \\ &= \sigma_w^2\sigma_v^2 vec(\mathbf{I}_J)vec^T(\mathbf{I}_\nu). \end{aligned}$$

Similarly, $E\{\overline{\mathbf{T}}_2\}$ can be proved to be equal to $E\{\mathbf{T}_2\}$. Then, the second term on the right-hand side of (89) or (90) becomes

$$\begin{aligned} &\text{second term of } E\{\mathbf{\Psi}_1\} \text{ or } E\{\mathbf{\Psi}_2\} \\ &\quad = \sigma_w^2\sigma_v^2 vec(\mathbf{H}\mathbf{H}^H)vec^T(\mathbf{I}_\nu). \end{aligned} \tag{97}$$

In order to compute the third term of (89) based on our assumptions, we apply the property of "$\otimes$" to combine $\mathbf{w}(n)$, which yields

$$\begin{aligned} E\{\mathbf{T}_3\} &= E\{(\mathbf{H}^* \otimes \mathbf{v}(n))vec(\mathbf{w}^H(n)) \\ &\qquad \cdot vec^T(\mathbf{w}(n))(\mathbf{v}^H(n) \otimes \mathbf{H}^T)\}. \end{aligned}$$

Due to the independent assumption, we may first compute $E\{vec(\mathbf{w}^H(n))vec^T(\mathbf{w}(n))\}$, which is equal to $\sigma_w^2\mathbf{I}_J$, and then deal with the left. Hence

$$E\{\mathbf{T}_3\} = \sigma_w^2 E\{(\mathbf{H}^* \otimes \mathbf{v}(n))(\mathbf{v}^H(n) \otimes \mathbf{H}^T)\}.$$

To further combine $\mathbf{v}(n)$, we write

$$\mathbf{H}^* \otimes \mathbf{v}(n) = (\mathbf{H}^* \otimes \mathbf{I}_\nu)(\mathbf{I}_J \otimes \mathbf{v}(n))$$

and

$$\mathbf{v}^H(n) \otimes \mathbf{H}^T = (\mathbf{v}^H(n) \otimes \mathbf{I}_J)(\mathbf{I}_\nu \otimes \mathbf{H}^T).$$

Then

$$\begin{aligned} E\{\mathbf{T}_3\} =& \sigma_w^2(\mathbf{H}^* \otimes \mathbf{I}_\nu) \\ &\cdot E\{(\mathbf{I}_J \otimes \mathbf{v}(n))(\mathbf{v}^H(n) \otimes \mathbf{I}_J)\}(\mathbf{I}_\nu \otimes \mathbf{H}^T). \end{aligned}$$

It is not difficult to show that $E\{(\mathbf{I}_J \otimes \mathbf{v}(n))(\mathbf{v}^H(n) \otimes \mathbf{I}_J)\} = \sigma_v^2\mathbf{A}_2$. Thus, the third term of (89) is

$$\text{third term of } E\{\mathbf{\Psi}_1\} = \sigma_w^2\sigma_v^2(\mathbf{H}^* \otimes \mathbf{I}_\nu)\mathbf{A}_2(\mathbf{I}_\nu \otimes \mathbf{H}^T). \quad (98)$$

Observing that its transpose is the fourth term of (89), then

$$\text{fourth term of } E\{\mathbf{\Psi}_1\} = \sigma_w^2\sigma_v^2(\mathbf{I}_\nu \otimes \mathbf{H})\mathbf{A}_2^T(\mathbf{H}^H \otimes \mathbf{I}_\nu). \quad (99)$$

As for the third term of (90), it can be treated in the same way

$$\begin{aligned} E\left\{\hat{\mathbf{T}}_3\right\} =& \sigma_w^2 E\{(\mathbf{H}^* \otimes \mathbf{v}(n))(\mathbf{H}^T \otimes \mathbf{v}^H(n))\} \\ =& \sigma_w^2(\mathbf{H}^*\mathbf{H}^T) \otimes E\{\mathbf{v}(n)\mathbf{v}^H(n)\}. \end{aligned}$$

This can be simplified as

$$\text{third term of } E\{\mathbf{\Psi}_2\} = \sigma_w^2\sigma_v^2(\mathbf{H}^*\mathbf{H}^T) \otimes \mathbf{I}_\nu \quad (100)$$

because $E\{\mathbf{v}(n)\mathbf{v}^H(n)\} = \sigma_v^2\mathbf{I}_\nu$. After reorganizing and using the fact that

$$E\{vec(\mathbf{w}(n))vec^H(\mathbf{w}(n))\} = \mathbf{I}_J$$

the fourth term of (90) becomes

$$\begin{aligned} E\{\hat{\mathbf{T}}_4\} =& \sigma_w^2 E\{(\mathbf{v}^*(n) \otimes \mathbf{H})(\mathbf{v}^T(n) \otimes \mathbf{H}^H(n))\} \\ =& \sigma_w^2 E\{\mathbf{v}^*(n)\mathbf{v}^T(n)\} \otimes (\mathbf{H}\mathbf{H}^H). \end{aligned}$$

Noticing that $E\{\mathbf{v}^*(n)\mathbf{v}^T(n)\} = \sigma_v^2\mathbf{I}_\nu$, we have

$$\text{fourth term of } E\{\mathbf{\Psi}_2\} = \sigma_w^2\sigma_v^2\mathbf{I}_\nu \otimes (\mathbf{H}\mathbf{H}^H). \quad (101)$$

The fifth term of (89) is the transpose of its second term [see (97)]

$$\text{fifth term of } E\{\mathbf{\Psi}_1\} = \sigma_w^2\sigma_v^2 vec(\mathbf{I}_\nu)vec^T(\mathbf{H}\mathbf{H}^H) \quad (102)$$

whereas the fifth term of (90) is the Hermitian of the corresponding second term

$$\text{fifth term of } E\{\mathbf{\Psi}_2\} = \sigma_w^2\sigma_v^2 vec(\mathbf{I}_\nu)vec^H(\mathbf{H}\mathbf{H}^H). \quad (103)$$

Following the line of derivation of $E\{\mathbf{T}_1\}$, similarly, we can obtain $E\{\mathbf{T}_6\}$ from (93)

$$E\{\mathbf{T}_6\} = (m_{4v} - 2\sigma_v^4)\tilde{\mathbf{B}} + \sigma_v^4\mathbf{A}_3 + \sigma_v^4 vec(\mathbf{I}_\nu)vec^T(\mathbf{I}_\nu) \quad (104)$$

where

$$\tilde{\mathbf{B}} = \left[\tilde{\mathbf{B}}_{k,\,l}\right]_{\nu\times\nu}, \qquad \tilde{\mathbf{B}}_{k,\,l} = \mathbf{u}_l\mathbf{u}_k^T$$

and m_{4v} is the fourth-order moment of $\mathbf{v}(n)$. As we know, for a zero mean complex AWGN process $x = x_1 + ix_2$ with variance $E\{|x|^2\} = a^2$, its independent real and imaginary parts both have variance

$$a_1^2 = a_2^2 = E\{x_1^2\} = E\{x_2^2\} = \tfrac{1}{2}\,a^2.$$

Then, the fourth-order moment of x_1 or x_2 is $E\{x_1^4\} = E\{x_2^4\} = 3a_1^4$ (see [15, Ch. 5]). Therefore , the fourth-order moment of x will be

$$\begin{aligned} E\{|x|^4\} =& E\{(x_1^2 + x_2^2)^2\} \\ =& E\{x_1^4\} + 2E\{x_1^2\}E\{x_2^2\} + E\{x_2^4\} \\ =& 8a_1^4 = 2a^4. \end{aligned}$$

Thus, $m_{4v} = 2\sigma_v^4$, which makes the first term in (104) zero. From (104), the last term of (89), which is $E\{\mathbf{T}_6\}$, becomes

$$\text{sixth term of } E\{\mathbf{\Psi}_1\} = \sigma_v^4\mathbf{A}_3 + \sigma_v^4 vec(\mathbf{I}_\nu)vec^T(\mathbf{I}_\nu). \quad (105)$$

Similarly, according to (94), $E\{\overline{\mathbf{T}}_6\}$ can be evaluated as

$$E\{\overline{\mathbf{T}}_6\} = (m_{4v} - 2\sigma_v^4)\tilde{\mathbf{B}} + \sigma_v^4\mathbf{I}_{\nu^2} + \sigma_v^4 vec(\mathbf{I}_\nu)vec^T(\mathbf{I}_\nu). \quad (106)$$

After zeroing out the first term in (106), the last term of (90) is achieved

$$\text{sixth term of } E\{\mathbf{\Psi}_2\} = \sigma_v^4\mathbf{I}_{\nu^2} + \sigma_v^4 vec(\mathbf{I}_\nu)vec^T(\mathbf{I}_\nu). \quad (107)$$

By combining (95), (97)–(99), (102), and (105), (58) follows, whereas considering (96), (97), (100), (101), (103), and (107) together, we obtain (58), which concludes our proof. □

REFERENCES

[1] S. E. Bensley and B. Aazhang, "Subspace-based channel estimation for code division multiple access communication systems," *IEEE Trans. Commun.*, vol. 45, pp. 1009–1020, Aug. 1996.

[2] D. R. Brillinger, *Time Series: Data Analysis and Theory*. New York: McGraw-Hill, 1981.

[3] B. Friedlander and B. Porat, "Asymptotically optimal estimation of MA and ARMA parameters of non-Gaussian processes from higher-order moments," *IEEE Trans. Automat. Contr.*, vol. 35, pp. 27–35, Jan. 1990.

[4] G. B. Giannakis and S. D. Halford, "Asymptotically optimal blind fractionally spaced channel estimation and performance analysis," *IEEE Trans. Signal Processing*, vol. 45, pp. 1815–1830, July 1997.

[5] G. B. Giannakis and J. M. Mendal, "Identification of nonminimum phase systems via high order statistics," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 38, pp. 360–377, Mar. 1989.

[6] G. B. Giannakis and M. K. Tsatsanis, "A unifying maximum-likelihood view of cumulant and polyspectral measures for non-Gaussian signal classification and estimation," *IEEE Trans. Inform. Theory*, vol. 38, pp. 386–406, Mar. 1992.

[7] G. H. Golub and C. F. Van Loan, *Matrix Computations*, 2nd ed. Baltimore, MD: Johns Hopkins Univ. Press, 1989.

[8] M. Honig, U. Madhow, and S. Verdu, "Blind adaptive multiuser detection," *IEEE Trans. Inform. Theory*, vol. 41, pp. 944–960, July 1995.

[9] R. A. Iltis and L. Mailaender, "Multiuser detection of quasisynchronous CDMA signals using linear decorrelators," *IEEE Trans. Commun.*, vol. 44, pp. 1561–1571, Nov. 1996.

[10] P. Lancaster and M. Tismenetsky, *The Theory of Matrices*, 2nd ed. San Diego, CA: Academic, 1985.

[11] L. Ljung, *System Identification: Theory for the User*. Englewood Cliffs, NJ: Prentice-Hall, 1987.

[12] Y. Li and K. J. R. Liu, “On blind MIMO channel identification using second order statistics,” in *Proc. 30th Conf. Inform. Sci., Syst.*, vol. 2, Princeton Univ., Princeton, NJ, May 1996, pp. 1166–1171.
[13] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, “Subspace methods for the blind identification of multichannel FIR filters,” *IEEE Trans. Signal Processing*, vol. 43, pp. 516–525, Feb. 1995.
[14] B. C. Ng, M. Cedervall, and A. Paulraj, “A structured channel estimator for maximum-likelihood sequence detection,” *IEEE Commun. Lett.*, vol. 1, pp. 52–55, Mar. 1997.
[15] A. Papoulis, *Probability, Random Variables and Stochastic Process*, 3rd ed. New York: McGraw-Hill, 1991.
[16] B. Porat, *Digital Processing of Random Signals: Theory & Methods*. Englewood Cliffs, NJ: Prentice-Hall, 1994.
[17] B. Porat and B. Friedlander, “Performance analysis of parameter estimation algorithms based on high-order moments,” *Int. J. Adapt. Contr. Signal Process.*, vol. 3, pp. 1191–1229, 1989.
[18] H. W. Sorenson, *Parameter Estimation*. New York: Marcel Dekker, 1980.
[19] M. Torlak and G. Xu, “Blind multiuser channel estimation in asynchronous CDMA systems,” *IEEE Trans. Signal Processing*, vol. 45, pp. 137–147, Jan. 1997.
[20] 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.
[21] M. K. Tsatsanis and G. B. Giannakis, “Equalization of rapidly fading channels: Self-recovering methods,” *IEEE Trans. Commun.*, vol. 44, pp. 619–630, May 1996.
[22] M. K. Tsatsanis and Z. Xu, “Performance analysis of minimum variance CDMA receivers,” *IEEE Trans. Signal Processing*, vol. 46, pp. 3014–3022, Nov. 1998.
[23] J. K. Tugnait, “Identification of linear stochastic systems via second order and fourth-order cumulant matching,” *IEEE Trans. Inform. Theory*, vol. IT-33, pp. 393–407, May 1987.
[24] X. Wang and H. V. Poor, “Blind multiuser detection: A subspace approach,” *IEEE Trans. Inform. Theory*, vol. 44, pp. 677–690, Mar. 1998.
[25] X. Wang and H. V. Poor, “Blind equalization and multiuser detection in dispersive CDMA channels,” *IEEE Trans. Commun.*, vol. 46, pp. 91–103, Jan. 1998.
[26] G. Xu, H. Liu, L. Tong, and T. Kailath, “A least-squares approach to blind channel identification,” *IEEE Trans. Signal Processing*, vol. 43, pp. 2982–2993, Dec. 1995.
[27] Z. Xu and M. K. Tsatsanis, “Blind channel estimation for long code multiuser CDMA systems,” *IEEE Trans. Signal Processing*, vol. 48, pp. 988–1001, Apr. 2000.

Zhengyuan Xu received both the B.E. and M.E. degrees in electronic engineering from Tsinghua University, Beijing, China, in 1989 and 1991, respectively, and the Ph.D. degree in electrical engineering from Stevens Institute of Technology, Hoboken, NJ, in 1999.

From 1991 to 1996, he worked as a System Engineer and Department Manager at the Tsinghua Unisplendour Group Corp. of Tsinghua University. In August 1996, when he entered Stevens Institute, he was a Research Assistant working on signal processing for wireless communications, especially for multiuser CDMA systems. He then worked as a Research Associate and then joined the Department of Electrical Engineering, University of California, Riverside, as an Assistant Professor. His current research interests include advanced signal processing, multirate communication, multiuser detection, and system identification.

Dr. Xu received the Outstanding Student Award and the Motorola Scholarship from Tsinghua University. He also received the Peskin Award for his outstanding performance at Stevens Institute.