Subspace-Based Blind Channel Estimation: Generalization and Performance Analysis

Wei Kang and Benoît Champagne, *Member, IEEE*

***Abstract*—In this paper, we present a systematic study of the subspace-based blind channel estimation method. We first formulate a general signal model of multiple simultaneous signals transmitted through vector channels, which can be applied to a multitude of modern digital communication systems. Based on this model, we then propose a generalized subspace-based channel estimator by minimizing a novel cost function, which incorporates the set of kernel matrices of the signals sharing the target channel via a weighted sum of projection errors. We investigate the asymptotic performance of the proposed estimator, i.e., bias, covariance, mean square error (MSE), and Cramér–Rao bound, for large numbers of independent observations. We show that the performance of the estimator can be optimized by increasing the number of kernel matrices and by using a special set of weights in the cost function. Finally, we consider the application of the proposed estimator to a downlink code division multiple access (CDMA) system operating in a frequency-selective fading channel with negligible intersymbol interference (ISI). The results of the computer simulations fully support our analytical developments.**

***Index Terms*—Blind channel estimation, DS-CDMA, performance analysis, subspace methods, wireless communications.**

I. Introduction

CHANNEL estimation has become a critical function in a variety of modern wireless communication systems, where multiple independent signals are transmitted simultaneously though vector channels. In effect, accurate channel information is important to recover the original transmitted signals by signal processing techniques, e.g., combining, deconvolution, detection, etc. [1]. Channel estimation algorithms can be roughly sorted into two basic categories: training sequence/pilot aided algorithms and blind algorithms. Recently, blind channel estimation algorithms have received considerable attention due to their advantages in terms of bandwidth efficiency [2].

Of particular interest within the family of blind algorithms are the so-called *subspace-based* blind channel estimation algorithms, which derive their properties from the second-order statistics of the received signals. In these methods, the observation space is separated into two orthogonal subspaces, namely, the signal subspace and the noise subspace, by applying eigenvalue decomposition (EVD) on the covariance matrix of the received signal. With the help of partial prior information on the structure of the transmitted signal, the vector channel of interest can be estimated by exploiting the orthogonal property between signal and noise subspaces.

During the past decade, subspace-based channel estimation algorithms have been developed for and applied to various vector channels, such as single-input multiple-output (SIMO) channels [3]–[5]; frequency-selective fading channels in direct sequence-code division multiple access (DS-CDMA) systems [6]–[8] and multicarrier (MC)-CDMA systems [9], [10]; multiple receiver antennae [6], [11] and/or multiple transmitter antennae channels [12], [13] in CDMA systems; and MC channels [14]. Although these algorithms were developed separately for certain specific transmission scenarios, the similarities among them indicate that there must exist some common features of the underlying system models, which provide for the feasibility of the subspace channel estimation. However, thus far, these common features have not been studied in the literature.

Besides, among the existing subspace-based channel estimation algorithms, a majority of them only utilize a single signal component to estimate the target channel, e.g., [6], [11], and [12]. However, in many situations of interest, the target channel is shared by multiple signal components simultaneously, as in, e.g., a typical downlink environment in cellular systems [8] or space-time block-coded channels [12]. Then, the problem of utilizing multiple signal components to estimate the target channel arises naturally. A pioneering work on this topic appeared in [3], which tackles the intersymbol interference (ISI) channel estimation problem in SIMO systems. Extension to the ISI channel in CDMA and orthogonal frequency division multiplexing (OFDM) systems can be found in [4] and [8], respectively. Thus far, there has not been a study that quantifies the effects of using multiple signal components in subspace-based blind channel estimation.

In this paper, motivated by the above considerations, we present a systematic study of the subspace-based blind channel estimation method. We first formulate a general signal model of multiple simultaneous signals transmitted through vector channels, which is applicable to a multitude of modern communication systems. Based on this model, we then propose a generalized subspace-based channel estimator by minimizing a novel cost function, which incorporates the set of kernel

Manuscript received June 25, 2003; revised February 10, 2004. This work was supported in part by grants from the Natural Sciences and Engineering Research Council of Canada and by McGill University under the Research and Development Fund. This paper was presented in part at the Vehicular Technology Conference, Orlando, FL, October 6–9, 2003. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Franz Hlawatsch.

W. Kang was with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, H3A 2A7 Canada. He is now with with the Department of Electrical and Computer Engineering, University of Maryland, College Park, MD 20742 USA (e-mail: wkang@glue.umd.edu).

B. Champagne is with the Department of Electrical and Computer Engineering, McGill University, Montreal, QC, H3A 2A7 Canada (e-mail: champagne@ece.mcgill.ca).

Digital Object Identifier 10.1109/TSP.2004.842207

matrices of the signal components sharing the target channel via a weighted sum of projection errors. The user-specified parameters in the proposed algorithms allow a generalization of previous work. Through study of the identifiability (i.e., existence and uniqueness), we find that enlarging the set of kernel matrices makes it possible to identify longer channel vectors and/or to increase the number of independent signals.

We investigate the asymptotic performance of the proposed estimator, i.e., bias, covariance, mean square error (MSE), and Cramér–Rao bound (CRB) for large numbers of independent observations. We show that the performance of the estimator can be optimized by increasing the number of kernel matrices and by using a special set of weights in the cost function. In particular, with the optimal weights and utilizing the kernel matrices of all the signal components sharing the target channel, the proposed estimator achieves both the minimum MSE and the CRB. Finally, we consider the application of the proposed estimator to a downlink CDMA system operating in a frequency-selective fading channel with negligible ISI. The results of the computer simulations fully support our analysis.

The paper provides a general mathematical framework so that the various previously developed subspace approaches may be embedded in a common formalism. However, the main contribution remains the analysis of the performance of the subspace channel estimation method, which, to the best of our knowledge, has never been thoroughly studied before from such a general angle. This analysis provides systematic tools for the design of optimal subspace estimators in any specific system that fits the proposed general model.

The paper is organized as follows. The general signal model under consideration is introduced in Section II. The proposed generalized subspace-based channel estimator is presented in Section III. The asymptotic performance properties of the proposed estimator are studied in Section IV. In Section V, we show the computer experiment results in a downlink CDMA system. This is followed by a conclusion in Section VI and Appendices that contain the proofs of theorems.

The following notations are used in this paper: $\mathbf{A}^T$, $\mathbf{A}^*$, $\mathbf{A}^H$, $\mathbf{A}^{-1}$, $\mathbf{A}^\dagger$, and Tr$[\mathbf{A}]$, respectively, denote the transpose, complex conjugate, conjugate transpose, inverse, pseudo-inverse, and trace of matrix $\mathbf{A}$. We denote the linear space spanned by the columns of $\mathbf{A}$ as Span$[\mathbf{A}]$. Let $\mathbf{A}$ be an $M \times N$ matrix with entries a_{ij}, and let $\mathbf{B}$ be an $P \times Q$ matrix. The Kronecker product of $\mathbf{A}$ and $\mathbf{B}$, which is denoted $\mathbf{A} \otimes \mathbf{B}$, is an $M \times N$ block matrix with an (i,j)th block $a_{ij}\mathbf{B}$. vec$[\mathbf{A}] \triangleq [\mathbf{a}_1^T, \ldots, \mathbf{a}_N^T]^T$, where $\mathbf{a}_i$ denotes the ith column of $\mathbf{A}$. $\mathbf{I}_L$ is the identity matrix with size $L \times L$. $\|\mathbf{v}\|$ denotes the Euclidean norm of vector $\mathbf{v}$, and diag$[\mathbf{v}]$ is a diagonal matrix with $\mathbf{v}$ on its main diagonal. $E[\cdot]$ denotes statistical expectation.

II. Problem Formulation

A. General Signal Model

Consider the following general model of an L-dimensional received signal vector in a communication system:

$$\mathbf{r} = \sum_{i=1}^{N} \gamma_i b_i \mathbf{C}_i \mathbf{h}_i + \mathbf{e} \tag{1}$$

where N is the number of individual symbols that comprise the received signal vector, γ_i is a real-valued received amplitude, which is the product of the transmitted amplitude and the channel gain, b_i is the ith information symbol, $\mathbf{h}_i$ is a normalized channel vector (i.e., $\|\mathbf{h}_i\| = 1$) with length M_i, $\mathbf{C}_i$ is defined as a kernel matrix with size $L \times M_i$, and $\mathbf{e}$ is an $L \times 1$ additive noise vector. We assume that the information symbols b_i, for $i = 1, \ldots, N$, are independent with zero mean and unit variance. The additive noise vector $\mathbf{e}$ is circularly complex Gaussian with covariance matrix $\sigma^2 \mathbf{I}_L$ and is independent from the information symbols b_i.

We define an $N \times 1$ data vector $\mathbf{b}$, an $N \times N$ amplitude matrix $\mathbf{\Gamma}$, and an $L \times N$ signature waveform matrix $\mathbf{W}$, respectively, as follows:

$$\mathbf{b} \triangleq [b_1, \ldots, b_N]^T \tag{2}$$

$$\mathbf{\Gamma} \triangleq \text{diag}[\gamma_1, \ldots \gamma_N] \tag{3}$$

$$\mathbf{W} \triangleq [\mathbf{w}_1, \ldots, \mathbf{w}_N] \tag{4}$$

where

$$\mathbf{w}_i \triangleq \mathbf{C}_i \mathbf{h}_i \quad i = 1, \ldots, N \tag{5}$$

is the effective signature waveform of the ith information symbol, i.e., combined effect of channel and kernel matrix, as seen by the receiver. Using the above matrix notations, the signal model (1) can be expressed more compactly as

$$\mathbf{r} = \mathbf{W}\mathbf{\Gamma}\mathbf{b} + \mathbf{e} = \mathbf{x} + \mathbf{e} \tag{6}$$

where we define $\mathbf{x} \triangleq \mathbf{W}\mathbf{\Gamma}\mathbf{b}$.

The system model (6) explicitly contains the random information symbol vector $\mathbf{b}$, which is convenient for subspace-based analysis (see Section III-A). However, for the purpose of deriving a channel estimation algorithm and analyzing its performance, it is more appropriate to reformulate (1) in a matrix format that explicitly displays the channel vectors.

In the sequel, we refer to the individual products $\gamma_i b_i \mathbf{C}_i \mathbf{h}_i$ $(i = 1, \ldots, N)$ in (1) as *signal components*. We assume that these N signal components experience J different channels $1 \leq J \leq N$. Then, we separate the N signal components into J groups, such that the signal components in each group share the same channel. We denote the number of signal components in the mth group as K^m, $m = 1, \ldots, J$ so that $\sum_{m=1}^{J} K^m = N$. In the mth group, we use the superscript m to denote group affiliation, as in the common channel parameter $\mathbf{h}^m$ with length M^m, and we use the superscript l to further distinguish among the K^m signal components, as in $\gamma^{m,l}$, $b^{m,l}$, $\mathbf{C}^{m,l}$, and $\mathbf{w}^{m,l}$. Finally, introducing the following quantities:

$$\mathbf{G}^m \triangleq \sum_{l=1}^{K^m} \gamma^{m,l} b^{m,l} \mathbf{C}^{m,l} \tag{7}$$

$$\mathbf{G} \triangleq [\mathbf{G}^1, \ldots, \mathbf{G}^J] \tag{8}$$

$$\underline{\mathbf{h}} \triangleq \text{vec}[\mathbf{h}^1, \ldots, \mathbf{h}^J] \tag{9}$$

the received signal vector can also be expressed as

$$\mathbf{r} = \mathbf{G}\underline{\mathbf{h}} + \mathbf{e}. \tag{10}$$

B. Discussion

The specific physical meaning of the various parameters entering the above model depends on the wireless communication system being considered. As such, the model is sufficiently general to accommodate several situations of interest, as exemplified below for different system features.

Direction of propagation:

- Conventional downlink: All the signals share the same channel so that there is only one group, as in, e.g., [8].
- Uplink: The signals from the same remote user share the same channel and the number of groups is equal to the number of remote users, e.g., [6] and [8].

Nature of information symbols:

- Intersymbol Interference (ISI): The entries of vector $\mathbf{b}$ represent consecutive information bits in the data stream, e.g., [3], [4], and [8].
- Multiple Access Interference (MAI): Vector $\mathbf{b}$ contains the simultaneous information bits of the different users, e.g., [6]–[8].
- Space-Time Block Codes (STBCs): Vector $\mathbf{b}$ contains the input symbols of an STBC encoder, e.g., [12] and [13].

Nature of the kernel matrix:

- CDMA: The kernel matrix $\mathbf{C}^{m,i}$ is a function of the signature waveform, or spreading code, of the ith user. This is the case for instance in DS-CDMA with time spreading, e.g., [6], [7], [11], and [12], and in MC-CDMA with frequency spreading, e.g., [9], [10], and [13].
- Oversampling in TDMA: This is used in ISI channels where the kernel matrix $\mathbf{C}^{m,l} = [\mathbf{0}_{M\times j}, \mathbf{I}_M, \mathbf{0}_{M\times k}]^T$ with $j + k + M = L$, e.g., [3]–[5].

Nature of the channel vectors:

- Dispersive Channel: The mth multipath channel is modeled as a tapped-delay-line [15] with tap coefficient vector $\mathbf{h}^m$ e.g., [6]–[8], and [12].
- MIMO Channel: Vector $\mathbf{h}^m$ is a concatenation of the various channel impulse responses (or gains) between the multiple transmit antennae and multiple receiver antennae, e.g., [6], and [11]–[13].

The above list is far from exhaustive but, to some degree, demonstrates the generality of the proposed signal model. The detailed example in Section V may be helpful to improve the understanding of the general model considered in this work. Further examples can be found in [16] as well as in the above cited references.

C. Blind Channel Estimation

Within the above framework, the goal of blind channel estimation is to determine one or more target channel vectors $\mathbf{h}^m$, $m = 1, \ldots, J$, using T observations of the received signal vector in (1), say, $\mathbf{r}_j \ (j = 1, \ldots, T)$. In this paper, and without loss of generality, we only consider the problem of estimating one target channel vector. As we explain in Section IV, the problem of estimating multiple target channel vectors can be uncoupled into multiple independent estimation problems, without any loss in theoretical achievable performance.

In blind channel estimation, the transmitted information symbols, as represented by vector $\mathbf{b}$ (2), are unknown. To estimate the target channel vector $\mathbf{h}^m$, at least one kernel matrix in the mth group needs to be known by the estimating algorithm. In practice, the specific available knowledge of the kernel matrices depends on the particular system under consideration. In this paper, we also assume that the length of the target channel vector, i.e., M^m is known or well estimated (the same as in [6] and [8]). For convenience of notation, we simply denote this parameter as $M \equiv M^m$ in the sequel.

Below, we formulate a generalized cost function for subspace-based blind channel estimation, which incorporates the set of kernel matrices of the signal components sharing the target channel. We then investigate the asymptotic performance of the estimator when the number of observations T is large.

III. Generalized Blind Subspace Channel Estimation

A. Theoretical Foundation

Let $\mathbf{R}$ denote the covariance matrix of received signal vector $\mathbf{r}$ in (1):

$$\mathbf{R} = E[\mathbf{r}\mathbf{r}^H] = \mathbf{W}\mathbf{\Gamma}^2\mathbf{W}^H + \sigma^2\mathbf{I}_L. \tag{11}$$

Blind subspace methods exploit the special structure of $\mathbf{R}$ to estimate the channel parameters. Specifically, let us express the EVD of $\mathbf{R}$ in the form

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \tag{12}$$

where $\mathbf{\Lambda} = \mathrm{diag}[\lambda_1, \ldots, \lambda_L]$ denotes the eigenvalue matrix, with the eigenvalues in a nonincreasing order, and $\mathbf{U}$ is a unitary matrix that contains the corresponding eigenvectors. Since the rank of matrix $\mathbf{W}\mathbf{\Gamma}^2\mathbf{W}^H$ (11) is N and the rank of matrix $\mathbf{R}$ is generically L due to the additive white noise component in (1), it follows that

$$\lambda_1 \geq \cdots \geq \lambda_N > \lambda_{N+1} = \cdots = \lambda_L = \sigma^2. \tag{13}$$

Thus, the eigenvalues can be separated into two distinct groups—the signal eigenvalues and the noise eigenvalues—respectively, represented by matrices

$$\mathbf{\Lambda}_s \triangleq \mathrm{diag}[\lambda_1, \ldots, \lambda_N] \tag{14}$$

$$\mathbf{\Lambda}_n \triangleq \mathrm{diag}[\lambda_{N+1}, \ldots, \lambda_L]. \tag{15}$$

Accordingly, the eigenvectors can be separated into the signal and noise eigenvectors, as represented by matrices $\mathbf{U}_s$ and $\mathbf{U}_n$ with dimensions $L \times N$ and $L \times (L - N)$, respectively. With these notations, the EVD in (12) can be expressed in the form

$$\mathbf{R} = [\,\mathbf{U}_s \quad \mathbf{U}_n\,] \begin{bmatrix} \mathbf{\Lambda}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_n \end{bmatrix} \begin{bmatrix} \mathbf{U}_s^H \\ \mathbf{U}_n^H \end{bmatrix}. \tag{16}$$

The columns of $\mathbf{U}_s$ span the so-called signal subspace with dimension N, whereas those of $\mathbf{U}_n$ span its orthogonal complement, i.e., the noise subspace. The signal subspace is indeed equal to the space spanned by the columns of $\mathbf{W}$:

$$\mathrm{Span}[\mathbf{W}] = \mathrm{Span}[\mathbf{U}_s] \perp \mathrm{Span}[\mathbf{U}_n]. \tag{17}$$

To estimate the target channel vector $\mathbf{h}^m$, which is shared by the signal components in the mth group, we select $1 \leq P \leq K^m$ effective signature waveforms from the mth group, say, $\mathbf{w}^{m,j}$ $(j = 1, \ldots, P)$ without loss of generality, and construct a matrix $\bar{\mathbf{W}} \triangleq [\mathbf{w}^{m,1}, \ldots, \mathbf{w}^{m,P}]$. Since the columns of $\bar{\mathbf{W}}$ form a subset of the columns of $\mathbf{W}$, it follows that

$$\text{Span}[\bar{\mathbf{W}}] \subseteq \text{Span}[\mathbf{W}] \perp \text{Span}[\mathbf{U}_n]. \tag{18}$$

Consequently

$$\mathbf{U}_n^H \bar{\mathbf{W}} = \mathbf{0}. \tag{19}$$

Defining

$$\mathcal{U}_s \triangleq \mathbf{I}_P \otimes \mathbf{U}_s \tag{20}$$

$$\mathcal{U}_n \triangleq \mathbf{I}_P \otimes \mathbf{U}_n \tag{21}$$

$$\mathcal{C}^T \triangleq [(\mathbf{C}^{m,1})^T, \ldots, (\mathbf{C}^{m,P})^T] \tag{22}$$

and applying vectorization operation on $\mathbf{U}_n^H \bar{\mathbf{W}}$, we obtain

$$\text{vec}\left[\mathbf{U}_n^H \bar{\mathbf{W}}\right] = \mathcal{U}_n^H \text{vec}[\bar{\mathbf{W}}] = \mathcal{U}_n^H \mathcal{C} \mathbf{h}^m = \mathbf{0}. \tag{23}$$

B. Identifiability

The first problem related to the identifiability issue is that there exists a phase ambiguity in the unit-norm solution of (23), i.e., the unit-norm solution of (23) is $\phi \mathbf{h}^m$, where ϕ is an arbitrary phase factor $|\phi| = 1$. The phase ambiguity indeed exists in all kinds of blind channel estimators and can be remedied by introducing extra constraints, e.g., by using differentially encoded information bits (more details about this problem can be found in [17]). Thus, we assume that the phase factor is known exactly in the rest of this paper.

Equation (23) can be interpreted as searching for a unit-norm vector $\mathbf{h}^m$ such that the vector $\mathcal{C}\mathbf{h}^m$ is orthogonal to $\mathcal{U}_n$ or, equivalently, within the subspace Span$[\mathcal{U}_s]$. Thus, (23) has a unique unit-norm solution if and only if both of the following conditions are satisfied.

1) The intersection space of Span$[\mathcal{C}]$ and Span$[\mathcal{U}_s]$ have dimension one (see [6]).
2) Matrix $\mathcal{C}$ has full column rank, i.e., its rank is equal to M.

Condition 1) depends on the channel vectors $\mathbf{h}^1, \ldots, \mathbf{h}^J$ because the subspace Span$[\mathcal{U}_s]$ = Span$[\mathbf{I}_P \otimes \mathbf{W}]$ is a function of $\mathbf{h}^1, \ldots, \mathbf{h}^J$. Condition 2) can be satisfied in practice by a judicious choice of the kernel matrix. To help understanding the above identifiability conditions, [6] can be consulted for additional explanation. Moreover, a study of identifiability condition in SIMO case can be found in [18]. In this work, we assume that the above two conditions are satisfied.

The dimension of Span$[\mathcal{U}_s]$ is NP, while under identifiability condition 2) above, the dimension of Span$[\mathcal{C}]$ is M. Clearly, the dimension of Span$[\mathcal{C}]\cup$Span$[\mathcal{U}_s]$ cannot exceed LP: the number of rows of matrices $\mathcal{C}$ and $\mathcal{U}_s$. Thus, the identifiability condition implies

$$M + NP - 1 \leq LP \iff N \leq L - \frac{M-1}{P} \iff M \leq (L-N)P + 1. \tag{24}$$

From the above inequalities, we conclude that increasing P will make it possible to allow more independent signals in the system (i.e., a larger N) and/or enable the estimator to identify a longer channel (i.e., a larger M).

C. Algorithm

In practice, the covariance matrix of the received signal $\mathbf{R} = E[\mathbf{r}\mathbf{r}^H]$ is usually unknown and must be estimated from the observed data via time averaging. Assuming a locally stationary environment, one such estimate based on a rectangular window of T samples is given by

$$\hat{\mathbf{R}} = \frac{1}{T} \sum_{j=1}^{T} \mathbf{r}_j \mathbf{r}_j^H \tag{25}$$

where $\mathbf{r}_j$ now denotes the received signal vector at the jth time instant (with similar modifications for other quantities of interest in (6)–(8): $\mathbf{b} \to \mathbf{b}_j, \mathbf{e} \to \mathbf{e}_j, \mathbf{G}^m \to \mathbf{G}_j^m$, and $\mathbf{G} \to \mathbf{G}_j$), for $j = 1, \ldots, T$.

In practice, the EVD is applied to $\hat{\mathbf{R}}$, resulting in

$$\hat{\mathbf{R}} = [\,\hat{\mathbf{U}}_s \quad \hat{\mathbf{U}}_n\,] \begin{bmatrix} \hat{\mathbf{\Lambda}}_s & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{\Lambda}}_n \end{bmatrix} \begin{bmatrix} \hat{\mathbf{U}}_s^H \\ \hat{\mathbf{U}}_n^H \end{bmatrix} \tag{26}$$

where $\hat{\mathbf{U}}_s, \hat{\mathbf{U}}_n, \hat{\mathbf{\Lambda}}_s$, and $\hat{\mathbf{\Lambda}}_n$ are noisy estimates of $\mathbf{U}_s, \mathbf{U}_n, \mathbf{\Lambda}_s$, and $\mathbf{\Lambda}_n$, respectively. Consequently, the noisy estimates of $\mathcal{U}_s$ (20) and $\mathcal{U}_n$(21) are, respectively, defined as

$$\hat{\mathcal{U}}_n \triangleq \mathbf{I}_P \otimes \hat{\mathbf{U}}_n \tag{27}$$

$$\hat{\mathcal{U}}_s \triangleq \mathbf{I}_P \otimes \hat{\mathbf{U}}_s. \tag{28}$$

Ideally, if $\hat{\mathcal{U}}_n = \mathcal{U}_n$ and the identifiability condition is satisfied, all the eigenvalues of $\mathcal{C}^H \hat{\mathcal{U}}_n \hat{\mathcal{U}}_n^H \mathcal{C}$ are positive except the smallest one, which is equal to 0. In this case, the channel vector $\mathbf{h}^m$ can be estimated exactly by solving the (23). However, in practice, the estimation error in $\hat{\mathcal{U}}_n$ may result in a positive perturbation in the smallest eigenvalue so that the matrix $\mathcal{C}^H \hat{\mathcal{U}}_n \hat{\mathcal{U}}_n^H \mathcal{C}$ is positive definite. In this case, (23) does not have a (nontrivial) solution. Thus, we conclude that estimating $\mathbf{h}^m$ by solving (23) is not robust to the perturbation of $\mathcal{U}_n$.

In this work, we consider the following optimization criterion for the blind estimation of channel vector $\mathbf{h}^m$:

$$\hat{\mathbf{h}}^m = \arg \min_{\|\mathbf{t}\|=1} \mathbf{t}^H \left[\sum_{i=1}^{P} \alpha^i (\mathbf{C}^{m,i})^H \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H \mathbf{C}^{m,i} \right] \mathbf{t} \tag{29}$$

where α^i for $i = 1, \ldots, P$ are user-specified positive weight parameters. We define

$$\mathcal{A} \triangleq \text{diag}\left[\sqrt{\alpha^1}, \ldots, \sqrt{\alpha^P}\right] \otimes \mathbf{I}_{L-N} \tag{30}$$

so that criterion (29) can be expressed in matrix form as

$$\hat{\mathbf{h}}^m = \arg \min_{\|\mathbf{t}\|=1} \mathbf{t}^H \mathcal{C}^H \hat{\mathcal{U}}_n \mathcal{A} \mathcal{A}^H \hat{\mathcal{U}}_n^H \mathcal{C} \mathbf{t}. \tag{31}$$

The choice of kernel matrices $\mathbf{C}^{m,i}$ included in the proposed criterion (29) is specified by the user, allowing a generalization of previous work. For example, the single signal algorithm in [6] can be obtained as a special case of (29) with $P = K^m = 1$, whereas the multiple signals algorithms in [3] correspond to $P = K^m$. Here, any value of P between 1 and K^m can be used.

We point out that the proposed criterion (29) is novel in that it allows for the introduction of user-specified weights α^i for $i = 1, \ldots, P$ in the estimation. This is motivated by the consid-

TABLE I
GENERALIZED BLIND SUBSPACE CHANNEL ESTIMATION ALGORITHM

Step	Complexity
$\hat{\mathbf{R}} = \frac{1}{T}\sum_{j=1}^{T}\mathbf{r}_j\mathbf{r}_j^H$	$\frac{1}{2}TL^2$
$\hat{\mathbf{R}} = [\,\hat{\mathbf{U}}_s\,\hat{\mathbf{U}}_n] \begin{bmatrix}\hat{\Lambda}_s & \mathbf{0}\\ \mathbf{0} & \hat{\Lambda}_n\end{bmatrix}\begin{bmatrix}\hat{\mathbf{U}}_s^H\\ \hat{\mathbf{U}}_n^H\end{bmatrix}$	$\mathrm{O}(L^3)$ [1]
P is user specified	
$\hat{\mathcal{U}}_n \triangleq \mathbf{I}_P \otimes \hat{\mathbf{U}}_n$	
$\mathcal{C}^T \triangleq [(\mathbf{C}^{m,1})^T, \ldots, (\mathbf{C}^{m,P})^T]$	
$\alpha^i \ \triangleq\ i = 1,\ldots,P$, are user specified	
$\mathbf{A} \ \triangleq\ \mathrm{diag}[\sqrt{\alpha^1},\ldots,\sqrt{\alpha^P}]$	
$\mathcal{A} \ \triangleq\ \mathbf{A}\otimes\mathbf{I}_N$	
Construct the matrix $\mathcal{C}^H\hat{\mathcal{U}}_n\mathcal{A}\mathcal{A}\hat{\mathcal{U}}_n^H\mathcal{C}$ or $\sum_{i=1}^{P}\alpha^i(\mathbf{C}^{m,i})^H\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^H\mathbf{C}^{m,i}$	$P[(L-N)ML + \frac{1}{2}(L-N)M^2 + \frac{1}{2}M^2]$
Find eigenvector $\hat{\mathbf{h}}^m$ corresponding to smallest eigenvalue of $\mathcal{C}^H\hat{\mathcal{U}}_n\mathcal{A}\mathcal{A}\hat{\mathcal{U}}_n^H\mathcal{C}$	$\mathrm{O}(M^3)$

[1] The exact figure for the computational complexity of the EVD depends on which specific algorithm is being used (see e.g. [19]).

eration of performance, which will be discussed in Sections IV and V in detail.

From an algorithmic viewpoint, the solution $\hat{\mathbf{h}}^m$ of (31) can be calculated as the eigenvector corresponding to the smallest eigenvalue of $\mathcal{C}^H\hat{\mathcal{U}}_n\mathcal{A}\mathcal{A}^H\hat{\mathcal{U}}_n^H\mathcal{C}$. The resulting estimation algorithm is summarized in Table I; we call it the *generalized* blind subspace channel estimation algorithm. Not only is this algorithm based on a general signal model, but it also gives the freedom to choose the kernel matrices and specify the weight parameters so as to optimize performance (see below).

IV. Asymptotic Performance Analysis

In this section, we investigate the asymptotic performance of the proposed generalized blind subspace channel estimation algorithm (see Table I). We define the estimation error as

$$\Delta\mathbf{h}^m \triangleq \hat{\mathbf{h}}^m - \mathbf{h}^m \tag{32}$$

where $\hat{\mathbf{h}}^m$ and $\mathbf{h}^m$, respectively, denote the estimated and true target channel vector for the mth group. The performance criteria of interest are the bias, covariance, and MSE of the proposed estimator, respectively, which are defined as

$$\mathrm{Bias} \triangleq E[\Delta\mathbf{h}^m] \tag{33}$$
$$\mathrm{Cov} \triangleq E\left[(\Delta\mathbf{h}^m - E[\Delta\mathbf{h}^m])\,(\Delta\mathbf{h}^m - E[\Delta\mathbf{h}^m])^H\right] \tag{34}$$
$$\mathrm{MSE} \triangleq E[\|\Delta\mathbf{h}^m\|^2]. \tag{35}$$

We study these performance measures under the assumption that the number of time samples T in (25) is large. According to weak law of large numbers [20], as $T\to\infty$

$$\frac{1}{T}\sum_{j=1}^{T}\mathbf{b}_j\mathbf{b}_j^H \longrightarrow E[\mathbf{b}\mathbf{b}^H] = \mathbf{I}_N,\ \text{in probability.} \tag{36}$$

Accordingly, the algorithm performance shall not depend on the specific sequence of information symbols $\{\mathbf{b}_j\}$ being transmitted.

Theorem 1: The proposed generalized estimator $\hat{\mathbf{h}}^m$ is asymptotically unbiased (i.e., Bias $=\mathbf{0}$) with the covariance

$$\mathrm{Cov} = \frac{\sigma^2}{T}\left[\left(\mathcal{C}^H\mathcal{U}_n\mathcal{A}\right)^{\dagger}\right]^H \mathcal{A}^H\boldsymbol{\Upsilon}^{-2}\mathcal{A}\left(\mathcal{C}^H\mathcal{U}_n\mathcal{A}\right)^{\dagger} \tag{37}$$

and MSE

$$\mathrm{MSE} = \frac{\sigma^2}{T}Tr\left[(\boldsymbol{\Upsilon}^{-1}\mathcal{A})\left(\mathcal{A}^H\mathcal{U}_n^H\mathcal{C}\mathcal{C}^H\mathcal{U}_n\mathcal{A}\right)^{\dagger}(\boldsymbol{\Upsilon}^{-1}\mathcal{A})^H\right] \tag{38}$$

where $\boldsymbol{\Upsilon} \triangleq \mathrm{diag}[\gamma^{m,1},\ldots,\gamma^{m,P}]\otimes\mathbf{I}_{L-N}$.

Proof: See Appendix A.

Theorem 1 indicates that the performance of the proposed estimator depends on the user-specified parameters, i.e., the weight matrix $\mathcal{A}$ (30) and the compounded kernel matrix $\mathcal{C}$ (22), which is determined by the set of kernel matrices utilized in the estimator, i.e., $S \triangleq \{\mathbf{C}^{m,1},\ldots,\mathbf{C}^{m,P}\}$.

We next investigate the optimal choice of parameters $\mathcal{A}$ and S that minimizes the MSE and the covariance of the estimator. To this end, it is convenient to explicitly indicate the functional dependence of these measures on $\mathcal{A}$ and S, i.e., MSE$(\mathcal{A},S)$ and Cov$(\mathcal{A},S)$. We begin with the minimization of MSE$(\mathcal{A},S)$, which proceeds in two steps. First, we minimize this measure by adjusting the weight matrix $\mathcal{A}$ such that for each fixed set S, we can obtain an optimal weight matrix, say, $\mathcal{A}^o(S)$; second, we search for a best choice of S to minimize MSE$(\mathcal{A}^o(S),S)$, i.e., when the optimal weight matrix for this set, as determined in the first step, is used. Then, the resulting choice of the parameters $\mathcal{A}$ and S minimizes MSE$(\mathcal{A},S)$.

Theorem 2: $\mathcal{A}^o(s) = c\boldsymbol{\Upsilon}$ is the optimal weight matrix minimizing MSE$(\mathcal{A},S)$ for a fixed set S:

$$\begin{aligned}\mathrm{MSE}^o(S) &\triangleq \min_{\mathcal{A}}\mathrm{MSE}(\mathcal{A},S)\\ &= \mathrm{MSE}(c\boldsymbol{\Upsilon},S) = \frac{\sigma^2}{T}\,\mathrm{Tr}[\mathcal{Q}^{\dagger}]\end{aligned} \tag{39}$$

where c is an arbitrary constant, and

$$\mathcal{Q} \triangleq \mathcal{C}^H\mathcal{U}_n\boldsymbol{\Upsilon}^2\mathcal{U}_n^H\mathcal{C}. \tag{40}$$

Proof: See Appendix B.

Theorem 2 shows that the optimal weights α^i $(i=1,\ldots,P)$ are proportional to the corresponding received powers $(\gamma^{m,i})^2$. Since all the signal components in the mth group share the same channel, the received powers of these components are proportional to their respective transmitted powers. Thus, the optimal weights are proportional to the transmitted powers as well. Theorem 2 also shows that introducing off-diagonal terms in the weight matrix $\mathcal{A}$ will not improve the performance but will increase the computational complexity of the estimator.

Next, we consider the set of kernel matrices $S = \{\mathbf{C}^{m,1},\ldots,\mathbf{C}^{m,P}\}$ utilized in the estimator with optimal weight matrix. We define the universal set of kernel matrices in the mth group as $U \triangleq \{\mathbf{C}^{m,1},\ldots,\mathbf{C}^{m,K^m}\}$ so that $S\subseteq U$. We also consider an arbitrary partition of S into Q nonempty subsets as $S_q (q=1,\ldots,Q)$:

$$\bigcup_{q=1}^{Q} S_q = S \tag{41}$$

$$S_p\cap S_q = \emptyset \text{ for any} p\neq q. \tag{42}$$

Theorem 3: For any proper subset S_q of S, that is $S_q \subset S \subseteq U$, we have

$$\mathrm{MSE}^o(S) < \mathrm{MSE}^o(S_q). \tag{43}$$

Proof: See Appendix C.

The above theorem implies qualitatively that enlarging the set of kernel matrices S in the estimator will decrease its MSE. Clearly, increasing the number of kernel matrices will entail an additional computational cost (see Table I). Thus, there is a tradeoff between computational complexity and estimation performance.

A direct result of Theorem 3 is that the minimum MSE of the proposed estimator is achieved when the estimator utilizes the kernel matrices of all the signal components in this group, i.e., $S = U$:

$$\begin{aligned}\mathrm{MSE}^o &\triangleq \min_{S\subseteq U} \mathrm{MSE}^o(S) = \mathrm{MSE}^o(U)\\ &= \frac{\sigma^2}{T}\,\mathrm{Tr}[(\mathcal{Q}^m)^\dagger]\end{aligned} \tag{44}$$

where

$$\mathcal{Q}^m \triangleq (\mathcal{C}^m)^H \mathcal{U}_n^m (\mathbf{\Upsilon}^m)^2 (\mathcal{U}_n^m)^H \mathcal{C}^m \tag{45}$$

$$\mathcal{C}^m \triangleq \left[(\mathbf{C}^{m,1})^T, \ldots, (\mathbf{C}^{m,K^m})^T\right]^T \tag{46}$$

$$\mathcal{U}_n^m \triangleq \mathbf{I}_{K^m} \otimes \mathbf{U}_n \tag{47}$$

$$\mathbf{\Upsilon}^m \triangleq \mathrm{diag}[\gamma^{m,1}, \ldots, \gamma^{m,K^m}] \otimes \mathbf{I}_{L-N}. \tag{48}$$

The next theorem provides a better indication of the relationship between the achievable $\mathrm{MSE}^o(S)$ and the individual $\mathrm{MSE}^o(S_q)$ for the subsets of kernel matrices forming the partition in (41) and (42).

Theorem 4: For arbitrary positive integers c_q, $q = 1, \ldots, Q$

$$\mathrm{MSE}^o(S) \le \frac{\sum_{q=1}^{Q} c_q^2 \mathrm{MSE}^o(S_q)}{\left(\sum_{q=1}^{Q} c_q\right)^2}. \tag{49}$$

Proof: See Appendix C.

As a special case of Theorem 4, assume that for $q = 1, \ldots, Q$, $c_q = 1$, and subset S_q only has one element, i.e., $S_q = \{\mathbf{C}^{m,q}\}$, and consequently, $Q = P$. Then, $\mathrm{MSE}^o(S_q)$ represents the MSE of the single signal estimator (e.g., [6]) applied on $\mathbf{C}^{m,q}$. Thus, according to Theorem 4

$$\mathrm{MSE}^o(S) \le \frac{1}{P^2} \sum_{q=1}^{P} \mathrm{MSE}^o(S_q) = \frac{\overline{\mathrm{MSE}}}{P} \tag{50}$$

where $\overline{\mathrm{MSE}} \triangleq (1/P)\sum_{q=1}^{P} \mathrm{MSE}^o(S_q)$ denotes the average MSE of single signal estimators over the set S. This result provides an easy way to roughly evaluate the performance gain of a multiple signal estimator over the single signal estimator without calculating their MSE.

Based on the above considerations, we suggest the following principles for minimizing the MSE of the proposed estimator.

1) Choose the weights proportional to the received powers.
2) Include maximum number of kernel matrices.

We now turn our attention to the optimization of the covariance of the proposed estimator, as defined in (34). Thus far, we have not been able to extend the results of Theorems 2 to 4 to the covariance matrix so that they remain valid in the form of matrix inequalities. Fortunately, we can use the Cramér–Rao bound (CRB) to judge the optimality of the parameter choice obtained in the case of MSE. That is, if the covariance matrix with the parameters $\mathcal{A} = c\mathbf{\Upsilon}$ and $S = U$ achieves the CRB, this parameter setting is considered the optimal one to minimize the covariance of the estimator.

As mentioned before, some constraints are usually imposed on the estimated channel vector, e.g., unit norm and/or known phase factor. In this case, the traditional CRB (see, e.g., [21]) is no longer applicable. The CRB for parameter estimation under constraints was recently given in [22], where a so-called *constrained* CRB is derived that depends on the specific algebraic constraints imposed on the estimated parameters. In [17], the concept of *minimal constrained* CRB is further introduced, which corresponds to the CRB matrix with the smallest trace (i.e., MSE) among the various constrained CRB matrices within the constraint class.

Theorem 5: The minimal constrained CRB of the vector $\underline{\mathbf{h}}$, which contains all the channel vectors, is given by

$$\mathrm{CRB}_{C,\underline{\mathbf{h}}} = \frac{\sigma^2}{T} \mathrm{diag}[(\mathcal{Q}^1)^\dagger, \ldots, (\mathcal{Q}^J)^\dagger] \tag{51}$$

and the minimal constrained CRB for the channel vector of interest is given by

$$\mathrm{CRB}_{C,\mathbf{h}^m} = \frac{\sigma^2}{T} (\mathcal{Q}^m)^\dagger. \tag{52}$$

Proof: See Appendix D.

From the previously derived expression (37) for the covariance matrix of the target channel, we find that the proposed generalized subspace estimator $\hat{\mathbf{h}}^m$ achieves the minimal constrained CRB when $\mathcal{A} = c\mathbf{\Upsilon}$ and $S = U$. Therefore, we conclude that the choice of parameters $\mathcal{A} = c\mathbf{\Upsilon}$ and $S = U$ not only minimizes the MSE but also minimizes the covariance of the estimator. Finally, we note that the minimal constrained CRB in (51) is block diagonal, providing a justification for our earlier statement that joint multiple channel estimation can be uncoupled into several independent single channel estimation problems, without any loss in theoretical achievable performance.

V. Illustrative Application With Computer Experiments

In this Section, we consider the problem of channel estimation for a downlink DS-CDMA system through a frequency-selective fading channel with negligible ISI, which satisfies the general model introduced in Section II.

Consider a downlink DS-CDMA connection from a base station to N remote users. The information bit to the ith user b_i is spread by a unique spreading code $\mathbf{c}^i \triangleq \left[c_1^i, \ldots, c_{L_c}^i\right]^H$, where L_c is the processing gain. The frequency-selective channel is modeled as a finite impulse response (FIR) filter of length M. The normalized coefficient vector of the filter is represented by $\mathbf{h}$ with size $M \times 1$. The kernel matrix of the ith user $\mathbf{C}^i$ is an $(L_c - M + 1) \times M$ Toeplitz matrix with the first column equal to $\left[c_M^i, \ldots, c_{L_c}^i\right]^T$ and the first row equal to $\left[c_M^i, \ldots, c_1^i\right]$ [6].

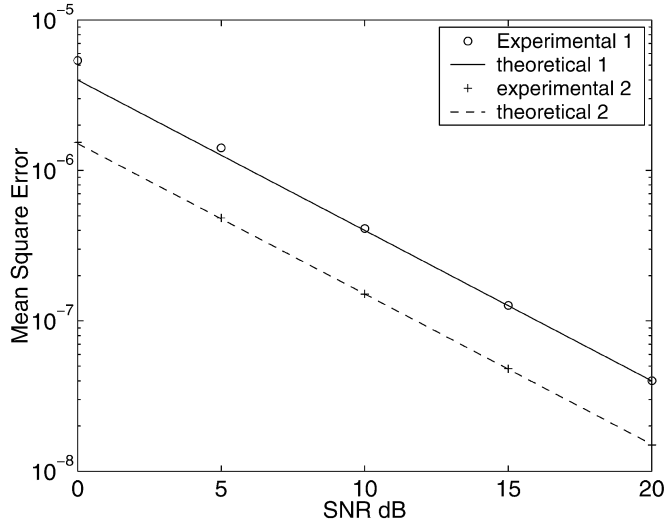

Fig. 1. Comparison in (54): 1) MSE$(\mathbf{I}_{4(L-N)}, S^4)$. 2) MSE$^o(S^4)$.

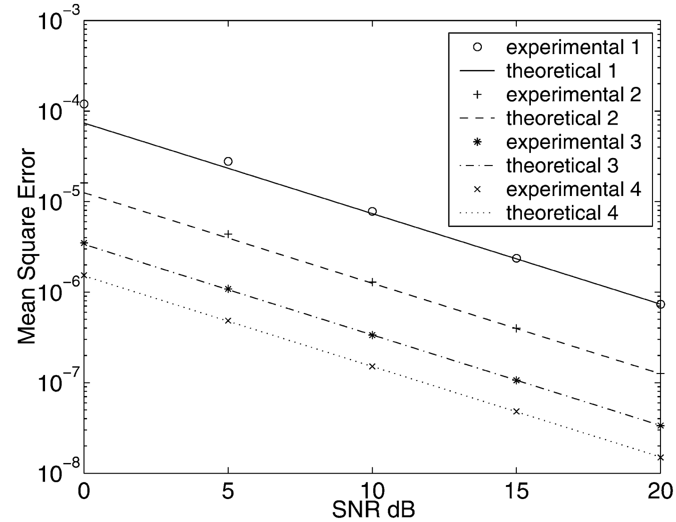

Fig. 2. Comparison in (55): 1) MSE$^o(S^1)$. 2) MSE$^o(S^2)$. 3) MSE$^o(S^3)$. 4) MSE$^o(S^4)$.

Assuming the received amplitude of the ith user is γ^i and the signal of all the users are synchronized, the received signal can be represented as

$$\mathbf{r} = \left(\sum_{i=1}^{N} \gamma^i \mathbf{C}^i b^i\right) \mathbf{h} + \mathbf{e} \tag{53}$$

where $\mathbf{e}$ is a white Gaussian noise vector.

The algorithm in Table I was specialized to this situation, resulting in a novel blind channel estimator that utilizes multiple signal components. Computer experiments were then conducted to verify the theoretical performance results derived in the last Section.

In the simulations, the following parameter values are used: number of active users $N = 4$, processing gain $L_c = 12$, and length of the channel vector $M = 4$. The binary spreading codes were randomly generated and stored for later use. We assume that some power control technique is applied so that the received amplitudes $[\gamma^1, \gamma^2, \gamma^3, \gamma^4]$ are proportional to $[1, 2, 3, 4]$, respectively. The following sets of kernel matrices were considered in the evaluation: $S^1 = \{\mathbf{C}^1\} \subset S^2 = \{\mathbf{C}^1, \mathbf{C}^2\} \subset S^3 = \{\mathbf{C}^1, \mathbf{C}^2, \mathbf{C}^3\} \subset S^4 = \{\mathbf{C}^1, \mathbf{C}^2, \mathbf{C}^3, \mathbf{C}^4\}$. We use the average value of the square error in 10^4 independent experiments to approximate the MSE.

According to the analysis in Section IV, the asymptotic MSE performance of the proposed estimator has the following properties. From (39), it follows that

$$\mathrm{MSE}^o(S^4) = \mathrm{MSE}(\boldsymbol{\Upsilon}, S^4) \leq \mathrm{MSE}(\mathbf{I}_{4(L-N)}, S^4) \tag{54}$$

where $\boldsymbol{\Upsilon} \triangleq \mathrm{diag}[\gamma^1, \ldots, \gamma^4] \otimes \mathbf{I}_{4(L-N)}$, and $L = L_c - M + 1$. From (43), it follows that

$$\mathrm{MSE}^o(S^4) < \mathrm{MSE}^o(S^3) < \mathrm{MSE}^o(S^2) < \mathrm{MSE}^o(S^1). \tag{55}$$

Finally, from (50), we have

$$\mathrm{MSE}^o(S^4) \leq \frac{1}{4^2} \sum_{i=1}^{4} \mathrm{MSE}^o(\{\mathbf{C}^i\}). \tag{56}$$

The simulation results are presented in Figs. 1–4. Figs. 1–3, respectively, show the MSEs in (54)–(56) plotted for both theoretical and experimental results as a function of SNR, with a number of observed samples $T = 10^4$. Fig. 4 shows the MSEs

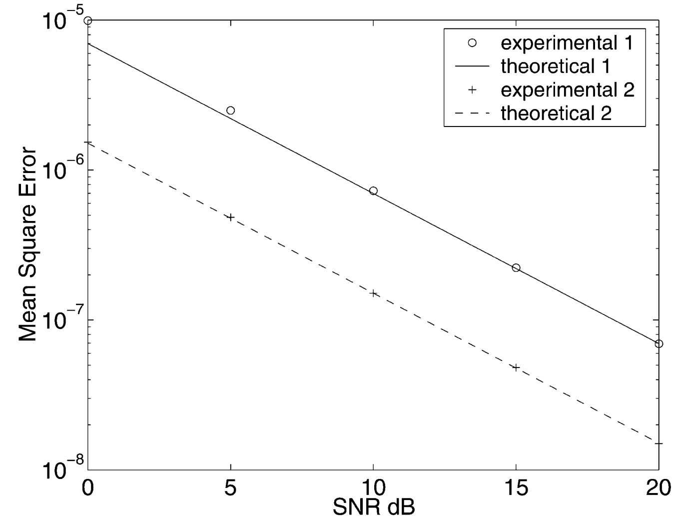

Fig. 3. Comparison in (56). 1) $(1/4^2)\sum_{i=1}^{4}$ MSE$^o(\{\mathbf{C}^i\})$. 2)MSE$^o(S^4)$.

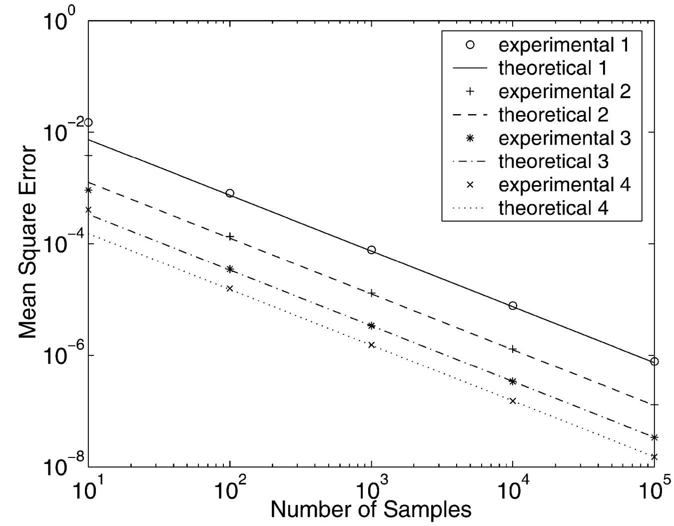

Fig. 4. Comparison in (55). 1) MSE$^o(S^1)$. 2) MSE$^o(S^2)$. 3) MSE$^o(S^3)$. 4) MSE$^o(S^4)$.

in (55) plotted for both theoretical and experimental results as a function of the number of observed samples T, with the SNR set to 10 dB. Clearly, the theoretical performance properties in

(54)–(56) are verified in our simulations. Generally, we find that all the experimental results match the theoretical results well, especially in the case of high SNR and large T. The former is because our theoretical results are derived on the basis of a first-order perturbation analysis, which is accurate in the case of small perturbations (i.e., high SNR region); the latter is because our asymptotic analysis is based on the assumption of a large number of samples T. Our results thus support the performance analysis in the general model derived in Section IV.

VI. Conclusion

We presented a systematic study of the subspace-based blind channel estimation method. We first introduced a general signal model of multiple simultaneous signals transmitted through vector channels, which can be applied to a multitude of modern digital communication systems. Based on this model, we formulated a generalized cost function for the purpose of subspace-based blind channel estimation, which incorporates the set of kernel matrices of the signals sharing the target channel via a weighted sum of projection errors. We investigated the asymptotic bias, covariance, MSE, and CRB of the proposed estimator when the number of observations is large. We showed that the performance of the estimator can be optimized by using the maximum number of available kernel matrices and a special set of weights in the cost function. The results of the computer simulations fully support our analysis.

Appendix A
Proof of Theorem 1

We begin by defining the following variables:

$$\mathbf{X} \triangleq [\mathbf{x}_1, \ldots, \mathbf{x}_T] \tag{57}$$
$$\mathbf{E} \triangleq [\mathbf{e}_1, \ldots, \mathbf{e}_T] \tag{58}$$
$$\bar{\mathbf{E}} \triangleq \mathbf{U}_n^H \mathbf{E} = [\bar{\mathbf{e}}_1, \ldots, \bar{\mathbf{e}}_T] \tag{59}$$
$$\mathcal{T} \triangleq \mathcal{C}^H \mathcal{U}_n \mathcal{A} \tag{60}$$
$$\mathcal{X} \triangleq \mathbf{I}_P \otimes \mathbf{X}^\dagger \tag{61}$$
$$\mathcal{E} \triangleq \mathbf{I}_P \otimes \bar{\mathbf{E}} = [\mathcal{E}_1, \ldots, \mathcal{E}_P] \tag{62}$$
$$\mathcal{E}_i \triangleq \iota_i \otimes \bar{\mathbf{E}} \quad i = 1, \ldots, P \tag{63}$$
$$\iota_i \triangleq \text{the } i\text{th column of the identity matrix} \tag{64}$$
$$\mathbf{z}^{m,i} \triangleq \mathbf{X}^\dagger \mathbf{w}^{m,i} = \mathbf{X}^\dagger \mathbf{C}^{m,i} \mathbf{h}^m = \left[z_1^{m,i}, \ldots, z_T^{m,i}\right]^T \tag{65}$$
$$\mathbf{z} \triangleq \text{vec}[\mathbf{z}^{m,1}, \ldots, \mathbf{z}^{m,P}] = \mathcal{X}\mathcal{C}\mathbf{h}. \tag{66}$$

According to [6, Lemma 1], the first-order perturbation of $\mathbf{U}_n$ is

$$\Delta \mathbf{U}_n = -(\mathbf{X}^\dagger)^H \mathbf{E}^H \mathbf{U}_n = -(\mathbf{X}^\dagger)^H \bar{\mathbf{E}}^H \tag{67}$$

and consequently, the perturbations of $\mathcal{U}_n$ and $\mathcal{T}$ are

$$\Delta \mathcal{U}_n = \mathbf{I}_P \otimes \Delta \mathbf{U}_n = -\mathcal{X}^H \mathcal{E}^H \tag{68}$$
$$\Delta \mathcal{T} = \mathcal{C}^H \Delta \mathcal{U}_n \mathcal{A} = -(\mathcal{A}^H \mathcal{E} \mathcal{X} \mathcal{C})^H. \tag{69}$$

According to [6, eq. (A.4)], the corresponding estimation error in $\hat{\mathbf{h}}^m$ is

$$\Delta \mathbf{h}^m = -(\mathcal{T}^\dagger)^H (\Delta \mathcal{T})^H \mathbf{h}^m = (\mathcal{T}^\dagger)^H \mathcal{A}^H \mathcal{E} \mathbf{z}. \tag{70}$$

Since $\bar{\mathbf{E}}$ is a zero-mean matrix, the bias of the estimator is

$$\text{Bias} = E[\Delta \mathbf{h}^m] = E[(\mathcal{T}^\dagger)^H \mathcal{A} \mathcal{E} \mathbf{z}] = \mathbf{0}. \tag{71}$$

The covariance of $\hat{\mathbf{h}}^m$ is

$$\text{Cov} = E[\Delta \mathbf{h}^m (\Delta \mathbf{h}^m)^H] = (\mathcal{T}^\dagger)^H \mathcal{A}^H E[\mathcal{E} \mathbf{z} \mathbf{z}^H \mathcal{E}^H] \mathcal{A} \mathcal{T}^\dagger. \tag{72}$$

According to [23], $E[\mathcal{E}\mathbf{z}\mathbf{z}^H\mathcal{E}^H]$ can be expressed as

$$E[\mathcal{E}\mathbf{z}\mathbf{z}^H\mathcal{E}^H] = E\left[\sum_{i=1}^{P}\sum_{p=1}^{T} z_p^{m,i}(\iota_i \otimes \bar{\mathbf{e}}_p) \times \sum_{j=1}^{P}\sum_{q=1}^{T} (z_q^{m,j})^* (\iota_j \otimes \bar{\mathbf{e}}_q)^H\right]. \tag{73}$$

To calculate $E[\mathcal{E}\mathbf{z}\mathbf{z}^H\mathcal{E}^H]$, we need to study $\mathbf{z}$ and $\mathcal{E}$, respectively. First, according to the result in [24] and the asymptotic property in (36), we have

$$(\mathbf{z}^{m,i})^H \mathbf{z}^{m,j} = \frac{1}{\gamma^{m,i}\gamma^{m,j} T} \delta_{i,j} \tag{74}$$

where the Kronecker delta function $\delta_{i,j} = 1$ for $i = j$ and 0 otherwise. Second, we derive the following results based on the definition in (62) and (64):

$$E[(\iota_i \otimes \bar{\mathbf{e}}_p)(\iota_j \otimes \bar{\mathbf{e}}_q)^H] = \sigma^2 (\iota_{i,j} \otimes I_{(L-N)}) \delta_{p,q} \tag{75}$$

where $\iota_{i,j}$ is a $P \times P$ matrix with all zero elements, except that the (i, j)th element is equal to one.

Based on (74) and (75), we simplify (73) as

$$\begin{aligned} E[\mathcal{E}\mathbf{z}\mathbf{z}^H\mathcal{E}^H] &= \sum_{i=1}^{P}\sum_{j=1}^{P} \sigma^2 (\iota_{i,j} \otimes I_{(L-N)}) (\mathbf{z}^{m,j})^H \mathbf{z}^{m,i} \\ &= \frac{\sigma^2}{T} \sum_{i=1}^{P} \frac{1}{(\gamma^{m,i})^2} (\iota_{i,i} \otimes I_{(L-N)}) \\ &= \frac{\sigma^2}{T} \boldsymbol{\Upsilon}^{-2}. \end{aligned} \tag{76}$$

By substituting (76) into (72), we have

$$\begin{aligned} \text{Cov} &= \frac{\sigma^2}{T} (\mathcal{T}^\dagger)^H \mathcal{A}^H \boldsymbol{\Upsilon}^{-2} \mathcal{A} \mathcal{T}^\dagger \\ &= \frac{\sigma^2}{T} [(\mathcal{C}^H \mathcal{U}_n \mathcal{A})^\dagger]^H \mathcal{A}^H \boldsymbol{\Upsilon}^{-2} \mathcal{A} (\mathcal{C}^H \mathcal{U}_n \mathcal{A})^\dagger. \end{aligned} \tag{77}$$

Finally

$$\begin{aligned} \text{MSE} &= E[(\Delta \mathbf{h}^m)^H \Delta \mathbf{h}^m] = \text{Tr}[\text{Cov}] \\ &= \frac{\sigma^2}{T} Tr \left[(\boldsymbol{\Upsilon}^{-1} \mathcal{A}) \left(\mathcal{A}^H \mathcal{U}_n^H \mathcal{C} \mathcal{C}^H \mathcal{U}_n \mathcal{A} \right)^\dagger \times (\boldsymbol{\Upsilon}^{-1} \mathcal{A})^H \right]. \end{aligned} \tag{78}$$

APPENDIX B
PROOF OF THEOREM 2

Lemma 1—[25]: If $\mathbf{A} \in \mathbb{C}^{M\times N}$ and $\mathbf{B} \in \mathbb{C}^{N\times P}$, then

$$(\mathbf{AB})^{\dagger} = (\mathbf{P}_{R(\mathbf{A}^H)}\mathbf{B})^{\dagger}(\mathbf{AP}_{R(\mathbf{B})})^{\dagger} \tag{79}$$

where $R(\mathbf{A})$ denotes the linear span of the columns of $\mathbf{A}$, and $\mathbf{P}_{R(\mathbf{A})}$ denotes the orthogonal projector onto $R(\mathbf{A})$.

We may express MSE (38) in the following form:

$$\mathrm{MSE}(\mathcal{A}, S) = \frac{\sigma^2}{T}\mathrm{Tr}[\bar{\mathcal{A}}(\bar{\mathcal{A}}^H\mathbf{Q}\bar{\mathcal{A}})^{\dagger}\bar{\mathcal{A}}^H] \tag{80}$$

where $\bar{\mathcal{A}} \triangleq \mathbf{\Upsilon}^{-1}\mathcal{A}$, and $\mathbf{Q} \triangleq \mathbf{\Upsilon}^H\mathcal{U}_n^h\mathcal{C}\mathcal{C}^H\mathcal{U}_n\mathbf{\Upsilon}$. According to Lemma 1, we have

$$(\bar{\mathcal{A}}^H\mathbf{Q}\bar{\mathcal{A}})^{\dagger} = [\mathbf{P}_{R[(\bar{\mathcal{A}}^H\mathbf{Q})^H]}\bar{\mathcal{A}}]^{\dagger}[\mathbf{P}_{R(\bar{\mathcal{A}})}\mathbf{Q}\mathbf{P}_{R(\bar{\mathcal{A}})}]^{\dagger} \times[\bar{\mathcal{A}}^H\mathbf{P}_{R[\mathbf{QP}_{R(\bar{\mathcal{A}})}]}]^{\dagger}. \tag{81}$$

Here, $\bar{\mathcal{A}}$ is full-rank, and referring to (23), $\mathbf{Q}$ is Hermitian and semi-positive definite. Thus

$$\mathbf{P}_{R(\bar{\mathcal{A}})} = \mathbf{I}_{P(L-N)} \tag{82}$$

$$\mathbf{P}_{R[\mathbf{QP}_{R(\bar{\mathcal{A}})}]} = \mathbf{P}_{R(\mathbf{Q})} = \mathbf{P}_{R(\mathbf{Q}^H)} = \mathbf{P}_{R[(\bar{\mathcal{A}}^H\mathbf{Q})^H]}. \tag{83}$$

Define

$$\mathbf{P} \triangleq \mathbf{P}_{R(\mathbf{Q})} \tag{84}$$

$$\mathbf{P}^{\perp} \triangleq \mathbf{I}_{P(L-N)} - \mathbf{P}. \tag{85}$$

Then

$$(\bar{\mathcal{A}}^H\mathbf{Q}\bar{\mathcal{A}})^{\dagger} = [\mathbf{P}\bar{\mathcal{A}}]^{\dagger}\mathbf{Q}^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}]^{\dagger}. \tag{86}$$

Consequently

$$\begin{aligned}\mathrm{Tr}[\bar{\mathcal{A}}(\bar{\mathcal{A}}^H\mathbf{Q}\bar{\mathcal{A}})^{\dagger}\bar{\mathcal{A}}^H] &= Tr\{\bar{\mathcal{A}}[\mathbf{P}\bar{\mathcal{A}}]^{\dagger}\mathbf{Q}^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}]^{\dagger}\bar{\mathcal{A}}^H\}\\ &= Tr\{[(\mathbf{P}+\mathbf{P}^{\perp})\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger}\mathbf{Q}^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}]^{\dagger}\\ &\quad\times[\bar{\mathcal{A}}^H(\mathbf{P}+\mathbf{P}^{\perp})]\}\\ &= Tr\{[\mathbf{P}\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger}\mathbf{Q}^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}]^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}]\}\\ &\quad+ Tr\{[\mathbf{P}^{\perp}\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger}\mathbf{Q}^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}]^{\dagger}\\ &\quad\times[\bar{\mathcal{A}}^H\mathbf{P}^{\perp}]\}\end{aligned} \tag{87}$$

where two terms in above equation have been cancelled because of the property $\mathbf{PP}^{\perp} = \mathbf{0}$. We separately study the two trace terms left in (87). In the first term

$$[\mathbf{P}\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger} = \mathbf{P}_{R(\mathbf{P}\bar{\mathcal{A}})} = \mathbf{P}. \tag{88}$$

Then

$$\begin{aligned}Tr\{[\mathbf{P}\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger}\mathbf{Q}^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}]^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}]\} &= \mathrm{Tr}[\mathbf{PQ}^{\dagger}\mathbf{P}]\\ &= \mathrm{Tr}[\mathbf{Q}^{\dagger}].\end{aligned} \tag{89}$$

In the second term, since $\mathbf{Q}$ is semi-positive definite, we have

$$Tr\{[\mathbf{P}^{\perp}\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger}\mathbf{Q}^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}]^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}^{\perp}]\} \geq 0. \tag{90}$$

Finally, we observe that when $\bar{\mathcal{A}} = c\mathbf{I}_{P(L-N)}$, i.e., $\mathcal{A} = c\mathbf{\Upsilon}$

$$\mathrm{Tr}\{[\mathbf{P}^{\perp}\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger}\mathbf{Q}^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}]^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}^{\perp}]\} = 0. \tag{91}$$

Therefore

$$\mathrm{MSE}^o(S) = \mathrm{MSE}(c\mathbf{\Upsilon}, S) = \frac{\sigma^2}{T}\mathrm{Tr}[\mathbf{Q}^{\dagger}]. \tag{92}$$

According to Lemma 1

$$\begin{aligned}\mathbf{Q}^{\dagger} &= (\mathbf{\Upsilon}^H\mathcal{U}_n^H\mathcal{C}\mathcal{C}^H\mathcal{U}_n\mathbf{\Upsilon})^{\dagger}\\ &= (\mathcal{C}^H\mathcal{U}_n\mathbf{\Upsilon})^{\dagger}(\mathbf{\Upsilon}^H\mathcal{U}_n^H\mathcal{C})^{\dagger}\end{aligned} \tag{93}$$

$$\begin{aligned}\mathcal{Q}^{\dagger} &= (\mathcal{C}^H\mathcal{U}_n\mathbf{\Upsilon}\mathbf{\Upsilon}^H\mathcal{U}_n^H\mathcal{C})^{\dagger}\\ &= (\mathbf{\Upsilon}^H\mathcal{U}_n^H\mathcal{C})^{\dagger}(\mathcal{C}^H\mathcal{U}_n\mathbf{\Upsilon})^{\dagger}.\end{aligned} \tag{94}$$

Therefore

$$\mathrm{Tr}[\mathbf{Q}^{\dagger}] = \mathrm{Tr}[\mathcal{Q}^{\dagger}] \tag{95}$$

and

$$\mathrm{MSE}^o(S) = \frac{\sigma^2}{T}\mathrm{Tr}[\mathbf{Q}^{\dagger}] = \frac{\sigma^2}{T}\mathrm{Tr}[\mathcal{Q}^{\dagger}]. \tag{96}$$

APPENDIX C
PROOF OF THEOREMS 3 AND 4

Definition 1: [26] Let $x_1, \ldots, x_n$ and $y_1, \ldots, y_n$ be real numbers. A vector $\mathbf{y} = [y_1, \ldots, y_n]$ is said to be majorized by a vector $\mathbf{x} = [x_1, \ldots, x_n]$, in symbols $\mathbf{x} \succ \mathbf{y}$ or $\mathbf{y} \prec \mathbf{x}$, if, after possible reordering of its components so that

$$x_1 \geq \cdots \geq x_n, \text{ and } y_1 \geq \cdots \geq y_n \tag{97}$$

we have

$$\sum_{i=1}^{k} x_i \geq \sum_{i=1}^{k} y_i, \text{ for } k = 1, \ldots, n-1 \tag{98}$$

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i. \tag{99}$$

Lemma 2: [27] If $\mathbf{H}$ is an $n \times n$ Hermitian matrix with diagonal elements $h_1, \ldots, h_n$ and eigenvalues $\lambda_1, \ldots, \lambda_n$, then

$$[\lambda_1, \ldots, \lambda_n] \succ [h_1, \ldots, h_n]. \tag{100}$$

Lemma 3 (Majorization Inequality): [26] $[x_1, \ldots, x_n]$ majorizes $[y_1, \ldots, y_n]$ iff for every convex function f

$$\sum_{i=1}^{n} f(x_i) \geq \sum_{i=1}^{n} f(y_i). \tag{101}$$

According to Theorem 2

$$\mathrm{MSE}^o(S) = \frac{\sigma^2}{T}\mathrm{Tr}[\mathcal{Q}^{\dagger}] \tag{102}$$

$$\mathrm{MSE}^o(S_q) = \frac{\sigma^2}{T}\mathrm{Tr}[\mathcal{Q}_q^{\dagger}], \quad q = 1, \ldots, Q \tag{103}$$

where

$$\mathcal{Q} = \sum_{\mathbf{C}^{m,i} \in S} (\gamma^{m,i})^2 (\mathbf{C}^{m,i})^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{C}^{m,i} \tag{104}$$

$$\mathcal{Q}_q \triangleq \sum_{\mathbf{C}^{m,i} \in S_q} (\gamma^{m,i})^2 (\mathbf{C}^{m,i})^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{C}^{m,i}. \tag{105}$$

Clearly, $\mathcal{Q} = \sum_{q=1}^{Q} \mathcal{Q}_q$. Apply EVD on $\mathcal{Q}$ and $\mathcal{Q}_q$, respectively

$$\mathcal{Q} = \mathbf{V} \mathbf{\Psi} \mathbf{V}^H \tag{106}$$

$$\mathcal{Q}_q = \mathbf{V}_q \mathbf{\Psi}_q \mathbf{V}_q^H, \quad q = 1, \ldots, Q \tag{107}$$

where $\mathbf{\Psi} \triangleq \text{diag}[\psi^1, \ldots, \psi^M]$, and $\mathbf{\Psi}_q \triangleq \text{diag}\left[\psi_q^1, \ldots, \psi_q^M\right]$ are the eigenvalue matrices of $\mathcal{Q}$ and $\mathcal{Q}_q$, respectively. Using a well-known property of matrix trace [28], we have

$$\text{Tr}[\mathcal{Q}^\dagger] = \sum_{j=1}^{m-1} \frac{1}{\psi^j} \tag{108}$$

$$\text{Tr}\left[\mathcal{Q}_q^\dagger\right] = \sum_{j=1}^{m-1} \frac{1}{\psi_q^j}, \quad q = 1, \ldots, Q. \tag{109}$$

For convenience of comparing $\text{Tr}[\mathcal{Q}^\dagger]$ and $\text{Tr}\left[\mathcal{Q}_q^\dagger\right]$, we define

$$\bar{\mathbf{V}}_q \triangleq \mathbf{V}^H \mathbf{V}_q \tag{110}$$

$$\bar{\mathbf{\Psi}}_q \triangleq \bar{\mathbf{V}}_q \mathbf{\Psi}_q \bar{\mathbf{V}}_q^h = \mathbf{V}^H \mathcal{Q}_q \mathbf{V} \tag{111}$$

where the (j,j)th element of $\bar{\mathbf{\Psi}}_q$ is denoted as $\bar{\psi}_q^j$.

On one hand, from (106) and (111) and the property $\mathcal{Q} = \sum_{q=1}^{Q} \mathcal{Q}_q$, it follows that

$$\mathbf{\Psi} = \sum_{q=1}^{Q} \bar{\mathbf{\Psi}}_q \tag{112}$$

and consequently

$$\psi^j = \sum_{q=1}^{Q} \bar{\psi}_q^j > \bar{\psi}_q^j > 0 \quad j = 1, \ldots, M-1. \tag{113}$$

Thus

$$\sum_{j=1}^{M-1} \frac{1}{\bar{\psi}_q^j} > \sum_{j=1}^{M-1} \frac{1}{\psi^j}. \tag{114}$$

On the other hand, since $\mathbf{V}$ in (111) is a unitary matrix, $\bar{\mathbf{\Psi}}_q$ and $\mathcal{Q}_q$ have the same eigenvalues, i.e., $\mathbf{\Psi}_q$ are the eigenvalue matrix of $\bar{\mathbf{\Psi}}_q$. According to Lemma 2

$$\left[\psi_q^1, \ldots, \psi_q^M\right] \succ \left[\bar{\psi}_q^1, \ldots, \bar{\psi}_q^M\right] \tag{115}$$

we know the function x^{-1} is only strictly convex in the interval $(0, \infty)$. To satisfy the condition in Lemma 3, we need to check if elements of $\left[\psi_q^1, \ldots, \psi_q^M\right]$ and $\left[\bar{\psi}_q^1, \ldots, \bar{\psi}_q^M\right]$ are within $(0, \infty)$. From the previous discussion, we know $\psi_q^j > 0$ for $j = 1, \ldots, M-1$ and $\psi_q^M = 0$. Since $\mathcal{Q}_q$ is semi-positive definite, $\bar{\psi}_q^j \geq 0$, $j = 1, \ldots, M$. Since $\psi^M = \sum_{q=1}^{Q} \bar{\psi}_q^M = 0$ [see (112)], it follows that $\bar{\psi}_q^M = 0$. Define ψ_q^o as the minimum of $\left[\psi_q^1, \ldots, \psi_q^{M-1}\right]$ and $\bar{\psi}_q^o$ as the minimum of $\left[\bar{\psi}_q^1, \ldots, \bar{\psi}_q^{M-1}\right]$. According to (98) and (99), $\bar{\psi}_q^o \geq \psi_q^o$. Since $\psi_q^o > 0$, then $\bar{\psi}_q^o > 0$. Consequently, we conclude that $\bar{\psi}_q^j > 0 \quad J = 1, \ldots, M-1$. Applying Lemma 3 to $\left[\psi_q^1, \ldots, \psi_q^{M-1}\right]$ and $\left[\bar{\psi}_q^1, \ldots, \bar{\psi}_q^{M-1}\right]$ with $f(x) = x^{-1}$, then

$$\sum_{j=1}^{M-1} \frac{1}{\psi_q^j} \geq \sum_{j=1}^{M-1} \frac{1}{\bar{\psi}_q^j}. \tag{116}$$

Based on (114) and (116), we have

$$\text{Tr}\left[\mathcal{Q}_q^\dagger\right] > \text{Tr}[\mathcal{Q}^\dagger] \tag{117}$$

and thus, we conclude that $\text{MSE}^o(S_q) > \text{MSE}^o(S)$ for $S_q \subset S$.

To prove Theorem 4, we consider the following property [26]: For any finite sequence of positive numbers $a_1, \ldots, a_n$

$$\frac{a_1 + \cdots + a_n}{n^2} \geq \frac{1}{\frac{1}{a_1} + \cdots + \frac{1}{a_n}}. \tag{118}$$

Then, we have

$$\begin{aligned}
\text{Tr}[\mathcal{Q}^\dagger] &= \sum_{j=1}^{M-1} \frac{1}{\psi^j} = \sum_{j=1}^{M-1} \frac{1}{\sum_{q=1}^{Q} \bar{\psi}_q^j} = \sum_{j=1}^{M-1} \frac{1}{\sum_{q=1}^{Q} c_q \frac{\bar{\psi}_q^j}{c_q}} \\
&\leq \sum_{j=1}^{M-1} \frac{\sum_{q=1}^{Q} c_q \frac{c_q}{\bar{\psi}_q^j}}{\left(\sum_{q=1}^{Q} c_q\right)^2} = \frac{\sum_{q=1}^{Q} c_q^2 \sum_{j=1}^{M-1} \frac{1}{\bar{\psi}_q^j}}{\left(\sum_{q=1}^{Q} c_q\right)^2} \\
&\leq \frac{\sum_{q=1}^{Q} c_q^2 \sum_{j=1}^{M-1} \frac{1}{\psi_q^j}}{\left(\sum_{q=1}^{Q} c_q\right)^2} = \frac{\sum_{q=1}^{Q} c_q^2 \text{Tr}[\mathcal{Q}_q^\dagger]}{\left(\sum_{q=1}^{Q} c_q\right)^2}.
\end{aligned} \tag{119}$$

The first two equalities in (119) follow from (108) and (113), respectively. The inequality on the second line follows from (118). Here, the sequence $\{a_i\}$ consists of the positive numbers $(c_1/\bar{\psi}_1^j), \ldots, (c_Q/\bar{\psi}_Q^j)$, where each number $(c_q/\bar{\psi}_q^j)$ is repeated c_q times: $q = 1, \ldots, Q$. The length of this sequence thus is $\sum_{q=1}^{Q} c_q$. The second inequality directly comes from (116). Therefore

$$\text{MSE}^o(S) \leq \frac{\sum_{q=1}^{Q} c_q^2 \text{MSE}^o(S_q)}{\left(\sum_{q=1}^{Q} c_q\right)^2}. \tag{120}$$

APPENDIX D
PROOF OF THEOREM 5

We define the following variables for convenience:

$$\underline{\mathbf{r}} \triangleq \text{vec}[\mathbf{r}_1, \ldots, \mathbf{r}_T] \tag{121}$$

$$\underline{\mathbf{x}} \triangleq \text{vec}[\mathbf{x}_1, \ldots, \mathbf{x}_T] \tag{122}$$

$$\underline{\mathbf{e}} \triangleq \text{vec}[\mathbf{e}_1, \ldots, \mathbf{e}_T] \tag{123}$$

$$\underline{\mathbf{b}} \triangleq \text{vec}[\mathbf{b}_1, \ldots, \mathbf{b}_T] \tag{124}$$

$$\underline{\mathbf{y}}^T \triangleq [\underline{\mathbf{h}}^T, \underline{\mathbf{b}}^T] \tag{125}$$

$$\mathcal{G}^T \triangleq \left[\mathbf{G}_1^T, \ldots, \mathbf{G}_T^T\right] \tag{126}$$

$$\mathcal{W} \triangleq \mathbf{I}_T \otimes (\mathbf{W}\mathbf{\Gamma}). \tag{127}$$

Concatenated received signal vector $\underline{\mathbf{r}}$ can be expressed as

$$\underline{\mathbf{r}} = \underline{\mathbf{x}} + \underline{\mathbf{e}} = \mathcal{W}\underline{\mathbf{b}} + \underline{\mathbf{e}} = \mathcal{G}\underline{\mathbf{h}} + \underline{\mathbf{e}}. \tag{128}$$

The Fisher Information Matrix (FIM) of estimating $\underline{\mathbf{y}}$ is derived as [29]

$$\mathbf{J} = \frac{1}{\sigma^2}\begin{bmatrix}\mathcal{G}^H \\ \mathcal{W}^H\end{bmatrix}[\mathcal{G} \quad \mathcal{W}]. \tag{129}$$

Lemma 4 (Constrained CRB): [22] Let $\hat{\mathbf{y}}$ be an unbiased estimator of a parameter vector $\mathbf{y}$ satisfying a constraint $f(\mathbf{y}) = 0$. Define $\mathbf{F}(\mathbf{y}) \triangleq \partial f(\mathbf{y})/\partial \mathbf{y}^T$, and hence, there exists a matrix $\mathbf{U}_0$ whose columns form an orthonormal basis for the null space of $\mathbf{F}(\mathbf{y})$. If $\mathbf{U}_0^H\mathbf{J}\mathbf{U}_0$ is nonsingular, then the constrained Cramér–Rao bound

$$\mathrm{CRB}_C = \mathbf{U}_0\left(\mathbf{U}_0^H\mathbf{J}\mathbf{U}_0\right)^{-1}\mathbf{U}_0^H. \tag{130}$$

Lemma 5 (Minimal Constrained CRB): [17] If $\mathrm{Span}[\mathbf{U}_0] = \mathrm{Span}[\mathbf{J}]$ and $\mathbf{U}_0$ has full column rank, then $\mathbf{U}_0^H\mathbf{J}\mathbf{U}_0$ is nonsingular, and the constrained CRB is

$$\mathrm{CRB}_c = \mathbf{J}^\dagger. \tag{131}$$

This is a particular constrained CRB: Among all sets of constraints, $\mathrm{CRB}_c = \mathbf{J}^\dagger$ yields the lowest value for $\mathrm{Tr}[\mathrm{CRB}_c]$.

Corollary 1: [17] Suppose $\mathbf{y}^T = \left[\mathbf{y}_1^T, \mathbf{y}_2^T\right]$, and the FIM is

$$\mathbf{J} = \begin{bmatrix}\mathbf{J}_{\mathbf{Y}_1\mathbf{Y}_1} & \mathbf{J}_{\mathbf{Y}_1\mathbf{Y}_2} \\ \mathbf{J}_{\mathbf{Y}_2\mathbf{Y}_1} & \mathbf{J}_{\mathbf{Y}_2\mathbf{Y}_2}\end{bmatrix}. \tag{132}$$

Assume $\mathbf{J}$ is singular but that $\mathbf{J}_{\mathbf{y}_2\mathbf{y}_2}$ is nonsingular. Then, the minimal constrained CRB for $\mathbf{y}_1$ separately is

$$\mathrm{CRB}_{C,\mathbf{y}_1} = \left[\mathbf{J}_{\mathbf{y}_1\mathbf{y}_1} - \mathbf{J}_{\mathbf{y}_1\mathbf{y}_2}\mathbf{J}_{\mathbf{y}_2\mathbf{y}_2}^{-1}\mathbf{J}_{\mathbf{y}_2\mathbf{y}_1}\right]^\dagger. \tag{133}$$

Applying the above corollary in our problem, we have

$$\mathrm{CRB}_{C,\underline{\mathbf{h}}} = \sigma^2[\mathcal{G}^H\mathcal{G} - \mathcal{G}^H\mathcal{W}(\mathcal{W}^H\mathcal{W})^{-1}\mathcal{W}^H\mathcal{G}]^\dagger. \tag{134}$$

Due to the definition in (126) and (127) and the property $(\mathbf{W\Gamma})[(\mathbf{W\Gamma})^H(\mathbf{W\Gamma})]^{-1}(\mathbf{W\Gamma})^H = \mathbf{U}_s\mathbf{U}_s^H$, we have

$$\mathrm{CRB}_{C,\underline{\mathbf{h}}} = \sigma^2\left[\sum_{j=1}^{T}\mathbf{G}_j^H\mathbf{U}_n\mathbf{U}_n^H\mathbf{G}_j\right]^\dagger. \tag{135}$$

From (8), $\mathbf{G}_j = \left[\mathbf{G}_j^1, \ldots, \mathbf{G}_j^J\right]$ so that the CRB (135) turns into a $J \times J$ block matrix, where the (m,n)th block is $\sigma^2\sum_{j=1}^T\left(\mathbf{G}_j^m\right)^H\mathbf{U}_n\mathbf{U}_n^H\mathbf{G}_j^n$. According to the definition of $\mathbf{G}_j^m$ in (7) and the asymptotic property in (36), we have

$$\sum_{j=1}^{T}\left(\mathbf{G}_j^m\right)^H\mathbf{U}_n\mathbf{U}_n^H\mathbf{G}_j^n = \begin{cases} T\sum_{l=1}^{K^m}(\gamma^{m,l})^2(\mathbf{C}^{m,l})^H\mathbf{U}_n\mathbf{U}_n^H\mathbf{C}^{m,l}, & m = n \\ \mathbf{0}, & m \neq n. \end{cases} \tag{136}$$

Considering the property $\sum_{l=1}^{K^m}(\gamma^{m,l})^2(\mathbf{C}^{m,l})^H\mathbf{U}_n\mathbf{U}_n^H\mathbf{C}^{m,l} = \mathcal{Q}^m$ [see (45)], we obtain a closed-form expression for the minimal constrained CRB as

$$\mathrm{CRB}_{C,\underline{\mathbf{h}}} = \frac{\sigma^2}{T}\mathrm{diag}[(\mathcal{Q}^1)^\dagger, \ldots, (\mathcal{Q}^j)^\dagger]. \tag{137}$$

Consequently

$$\mathrm{CRB}_{C,\mathbf{h}^m} = \frac{\sigma^2}{T}(\mathcal{Q}^m)^\dagger. \tag{138}$$

References

[1] H. V. Poor and G. W. Wornell, Eds., *Wireless Communications: Signal Processing Perspectives*. Englewood Cliffs, NJ: Prentice-Hall, 1998.

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

[3] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue, “Subspace methods for the blind identification of multichannel FIR filters,” *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 527–534, Feb. 1995.

[4] S. Roy and C. Li, “A subspace blind channel estimation method for OFDM systems without cyclic prefix,” *IEEE Trans. Wireless Commun.*, vol. 1, pp. 572–579, Oct. 2002.

[5] Z. Xu, “On the second-order statistics of the weighted sample covariance matrix,” *IEEE Trans. Signal Process.*, vol. 51, pp. 516–525, Feb. 2003.

[6] H. Liu and G. Xu, “A subspace method for signature waveform estimation in synchronous CDMA systems,” *IEEE Trans. Commun.*, vol. 44, pp. 1346–1354, Oct. 1996.

[7] S. E. Bensley and B. Aazhang, “Subspace-based channel estimation for code division multiple access communication systems,” *IEEE Trans. Commun.*, vol. 44, pp. 1009–1020, Aug. 1996.

[8] 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.

[9] J. Wu, Y. Wang, and K. K. M. Cheng, “Blind channel estimation based on subspace for multicarrier CDMA,” in *Proc. Veh. Technol. Conf.*, vol. 4, Rhodes, Greece, 2001, pp. 2374–2378.

[10] C. Li and S. Roy, “Subspace based detection in MC-CDMA over dispersive channel,” in *Proc. IEEE Int. Conf. Pers. Wireless Commun.*, Hyderabad, India, 2000, pp. 98–103.

[11] A. Chkeif, K. Abed-Meraim, G. Kawas-Kaleh, and Y. Hua, “Spatio-temporal blind adaptive multiuser detection,” *IEEE Trans. Commun.*, vol. 48, no. 5, pp. 729–732, May 2000.

[12] D. Reynolds, X. Wang, and H. V. Poor, “Blind adaptive space-time multiuser detection with multiple transmitter and receiver antennas,” *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1261–1276, Jun. 2002.

[13] W. Sun and H. Li, “Blind channel identification for multicarrier CDMA systems with transmit diversity,” in *Proc. IEEE Int. Conf. Commun.*, vol. 2, New York, 2002, pp. 727–731.

[14] J. Namgoong, T. F. Wong, and J. S. Lehnert, “Subspace multiuser detection for multicarrier DS-CDMA,” *IEEE Trans. Commun.*, vol. 48, no. 11, pp. 1897–1908, Nov. 2000.

[15] J. G. Proakis, *Digital Communications*, Fourth ed. New York: McGraw-Hill, 2001.

[16] W. Kang, “Subspace-Based Blind Channel Estimation: Generalization, Performance Analysis, and Adaptive Implementation,” M.Eng. thesis, McGill Univ., Montreal, QC, Canada, 2003.

[17] E. de Carvalho, J. Cioffi, and D. Slock, “Cramer-Rao bounds for blind multichannel estimation,” in *Proc. Global Telecommun. Conf.*, vol. 2, San Francisco, CA, 2000, pp. 1036–1040.

[18] P. Loubaton and E. Moulines, “On blind multiuser forward link channel estimation by the subspace method: Identifiability results,” *IEEE Trans. Signal Process.*, vol. 48, no. 8, pp. 2366–2376, Aug. 2000.

[19] G. H. Golub and C. F. V. Loan, *Matrix Computation*. Baltimore, MD: Johns Hopkins Univ. Press, 1996.

[20] A. Papoulis, *Probability, Random Variables, and Stochastic Processes*, Third ed. New York: McGraw-Hill, 1991.

[21] H. L. V. Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, 1968, pt. I.

[22] P. Stoica and B. C. Ng, “On the Cramer-Rao bound under parametric constraints,” *IEEE Signal Process. Lett.*, vol. 5, no. 7, pp. 177–179, Jul. 1998.

[23] F. Li and R. J. Vaccaro, "Analysis of min-norm and MUSIC with arbitrary array geometry," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 26, pp. 976–985, Nov. 1990.

[24] F. Li, H. Liu, and R. J. Vaccaro, "Performance analysis for DOA estimation algorithms: Unification, simplification, and observations," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, pp. 1170–1184, Oct. 1993.

[25] S. L. Campbell and C. D. Meyer, *Generalized Inverses of Linear Transformations*. New York: Dover, 1979.

[26] D. S. Mitrinovic, *Analytic Inequalities*. New York: Springer-Verlag, 1970.

[27] I. O. A. W. Marshall, *Inequalities: Theory of Majorization and Its Applications*. New York: Academic, 1979.

[28] R. A. Horn and C. R. Johnson, *Matrix Analysis*. Cambridge, U.K: Cambridge Univ. Press, 1985.

[29] S. S. Haykin, *Adaptive Filter Theory*, Fourth ed. Englewood Cliffs, NJ: Prentice-Hall, 2002.

Wei Kang received the B.Eng. degree in electrical engineering from Beijing University of Posts and Telecommunications, Beijing, China, in 2001 and the M.Eng. degree in electrical engineering from McGill University, Montreal, QC, Canada in 2003. He is currently pursuing the Ph.D. degree with the Department of Electrical and Computer Engineering, University of Maryland, College Park.

His research interests include information theory and signal processing for communications.

Benoît Champagne (M'89) was born in Joliette, QC, Canada in 1961. He received the B.Ing. degree in engineering physics from the Ecole Polytechnique of Montreal, Montreal, QC, in 1983, the M.Sc. degree in physics from the University of Montreal in 1985, and the Ph.D. degree in electrical engineering from the University of Toronto, Toronto, ON, Canada, in 1990.

From 1990 to 1999, he was an Assistant and then Associate Professor at INRS-Télécommunications, Université du Québec, Montreal, where he is currentl a Visiting Professor. In September 1999, he joined McGill University, Montreal, as an Associate Professor within the Department of Electrical and Computer Engineering; he is currently acting as Associate Chairman of Graduate Studies. His research interests lie in the area of statistical signal processing, including signal/parameter estimation, sensor array processing, and adaptive filtering and applications thereof to communications systems.