Subspace-based Blind Channel Estimation: Generalization, Performance Analysis and Adaptive Implementation

Wei Kang

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

September 2003

A thesis submitted to the Faculty of Graduate Studies and Research in partial fulfillment of the requirements for the degree of Master of Engineering.

2003/09/17

Abstract

In this thesis, 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 Cramer-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.

We also propose a novel adaptive implementation of the generalized subspace channel estimator. The low-complexity and numerical robustness of this adaptive implementation make it suitable for online estimation of time-varying channels over long observation periods.

Finally, we consider the application of the proposed estimator to a down-link CDMA system operating in frequency selective fading channel with negligible ISI. The results of the computer simulations fully support our analytical developments.

Sommaire

Dans ce mémoire, une étude systématique des méthodes d'estimation aveugle de canaux basées sur les sous-espaces est présentée.

Nous formulons d'abord un modèle général du signal reçu, formé de multiples transmissions simultanées de signaux sur des canaux vectoriels. Ce modèle peut être utilisé par une multitude de systèmes de communications numériques modernes. À partir de ce modèle, nous proposons un estimateur de canaux généralisé basé sur les sous-espaces et minimisant une nouvelle fonction de coût. Cette dernière incorpore l'ensemble des matrices noyaux des signaux partageant le canal cible via une somme pondérée de projections d'erreurs.

Nous étudions les comportements asymptotiques de l'estimateur en question, c'est-à-dire le biais, la covariance, l'erreur quadratique moyenne et la borne Cramér-Rao, pour un grand nombre d'observations indépendantes. Nous démontrons que le rendement de l'estimateur peut être optimisé en augmentant le nombre des matrices noyaux et en utilisant un ensemble spécial de poids dans la fonction de coût.

Nous proposons également une nouvelle version adaptive de l'estimateur de canaux par sous-espaces généralisés. La faible complexité et la robustesse numérique de cette version adaptative rendent possible l'estimation de canaux instationnaires en temps réel sur de longues périodes de temps.

Finalement, nous examinons l'application de l'estimateur proposé à la liaison descendante d'un système d'accès multiple par répartition en code (CDMA) opérant dans un canal à évanouissement sélectif de fréquences avec un brouillage intersymbole négligeable. Les résultats des simulations par ordinateur supportent entièrement nos développements analytiques.

Acknowledgments

I would like to thank my supervisor, Prof. Benoît Champagne, for his guidance and support during my graduate studies at McGill University. I am deeply grateful for the invaluable advice and the research freedom he has given me. Working with him has been a great learning experience.

I would like to thank Prof. Ioannis Psaromiligkos and Prof. Jan Bajcsy for their helpful suggestions and encouragement in both my course studies and the research practice. I have gained plenty of knowledge and fun from the discussions with them.

I would like to thank Prof. M. Reza Soleymani from Concordia University for reviewing this thesis. His comments and suggestions help me improve this thesis.

I would like to thank Prof. Xiao-Wen Chang from the School of Computer Science for his useful inputs regarding the computation of subspace intersections.

I would like to thank my friends in Telecommunications and Signal Processing Laboratory for their friendship and help; they are Mr. Benoît Pelletier, Mr. François Duplessis-Beaulieu, Mr. Jean-François Marceau, Mr. Xiaojian Lu and Dr. Jian Mao.

I would like to thank Miss Nan Liu for her love, which makes everyday meaningful.

Finally, I would like to thank my mother and father. It is them who made me who I am today.

Contents

List of Figures

List of Tables

Chapter 1

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. [25].

Channel estimation algorithms can be roughly sorted into two basic categories: data aided algorithm (training sequence or pilot aided algorithms) and non-data aided algorithm (blind algorithms). Recently, blind channel estimation algorithms have received considerable attention due to their advantages in terms of bandwidth efficiency. Indeed, unlike the training sequence/pilot aided methods, blind algorithms only rely on the received signal to carry out the channel estimation and therefore make more efficient use of the available bandwidth (i.e. higher data rates) [36].

Of particular interest within the family of blind algorithms are the so-called *subspace-based* blind channel 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, or by applying the Singular Value Decomposition (SVD) on the received signal matrix. 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 orthogonality property between signal and noise subspaces.

In the following sections, we present the motivation, contribution and organization of the thesis, which is devoted to the study of subspace-based blind channel estimation algorithms. We first review the literature in the area of subspace channel estimation in Section 1.1 and then summarize the main thesis contribution in Section 1.2. The organization of the thesis is presented in Section 1.3 followed by a list of mathematical notations in Section 1.4.

1.1 Literature Survey

During the past decade, subspace based channel estimation algorithms have been developed for and applied to various vector channels.

One of the earliest propositions of applying the subspace method to channel estimation problems can be traced back to the work by Moulines *et al.* in 1995 (see [22]), which focuses on identifying time dispersive channel (modelled as an FIR filter) in Time Division Multiple Access (TDMA) system with oversampling in time and/or space domain by using subspace methods.

With the popularity of Code Division Multiple Access (CDMA) communication systems, several works on the estimation of multipath channels in CDMA system by subspace methods have been reported in [2, 19, 33]. Among them, Liu and Xu's work in [19] deserves extra attention. In this work, the authors study the identifiability problem of subspace channel estimation for the first time. Moreover, they provide a closed form expression of the asymptotic performance of their estimator by using a first-order perturbation analysis.

Since then, several blind subspace channel estimation methods have been proposed and applied to different scenarios, such as: Single Input Multiple Output (SIMO) channels [39, 29, 42], frequency selective fading channels in DS-CDMA systems [37, 1, 39] and Multi-Carrier CDMA (MC-CDMA) systems [41, 16], multiple receiver antennae [19, 6] and multiple transmitter antennae channels [28, 34] in CDMA systems, multi-carrier channels [23], etc.

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, so 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. [19, 6, 28]. 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 [39], time dispersive channels [39] or space-time block coded channels [28]. Then the problem of utilizing multiple signal components to estimate the target channel arises naturally. A pioneering work on this topic appeared in [22], which tackles the Inter-Symbol 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 [39, 29], respectively. So far, there has not been a study that quantifies the effects of using multiple signal components in subspace-based blind channel estimation.

Finally, most subspace channel estimation algorithms (e.g. [19]) were originally proposed for time-invariant channels. In order to apply it to on-line estimation of time-varying channels, some hybrid (adaptive-batch) algorithms have been proposed. The hybrid algorithms use subspace tracking techniques to adaptively generate the subspace information; however, the batch method is still used to estimate the target channel. Therefore the hybrid algorithms still require a considerable amount of computations (e.g. [40]). Recently, a low-complexity, adaptive channel estimation algorithm based on the Orthogonal Projection Approximation Subspace Tracking (OPAST) method was proposed in [6]. However, OPAST only provides a basis of the signal subspace and cannot track the individual dominant eigenvectors and eigenvalues, which are needed in advanced subspaced-based multi-user detection. Furthermore, some numerical stability problems have been observed with OPAST when running it over a large number of time iterations.

1.2 Thesis Contribution

In this thesis, motivated by the above considerations, we present a systematical study of the subspace-based blind channel estimation method[1].

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. In this model, individual signals are characterized by a normalized channel vector and a kernel matrix whose specific structure depends on the transmis-

[1]Parts of this thesis appear in [15, 14, 13]

sion system under consideration. Based on this model, we then propose a generalized subspace-based channel estimator based on the minimization of a novel cost function, which incorporates the set of kernel matrices of the signal components sharing the target channel via a weighted sum of projection errors. Through study of the identifiability (i.e. existence and uniqueness) of the proposed estimator, we find that enlarging the set of kernel matrices, i.e. utilizing more independent signals in the estimator, makes it possible to identify longer channel vectors and/or to increase the number of independent system users.

We investigate the asymptotic performance of the proposed estimator as the number of independent observations increases. We derive its bias, covariance and mean square error (MSE), as well as the associated Cramer-Rao bound. We show that the performance of the estimator can be optimized by incorporating the maximum 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.

We also propose a fully adaptive implementation of the proposed generalized subspace channel estimator. The new adaptive implementation is derived by exploiting common structural properties of plane rotation-based subspace trackers (e.g. [5, 27]), whose numerical robustness has been well established. The main advantage of the proposed adaptive implementation is its low-complexity and numerical robustness over long periods of utilization, an essential requirement for practical operation in wireless radio applications.

Finally, we consider a down-link synchronous CDMA system operating in frequency selective fading channel with negligible ISI, where a new channel estimator is obtained by applying the generalized estimator to this specific scenario. The performance of the proposed estimator (MSE) is studied via computer simulations. In the simulations for the batch estimator, we consider a locally stationary environment, where the channel does not change during T samples. We use the average value of the square errors of the proposed estimator in a large number of independent experiments to approximate the mean square error. We find that all the experimental results match the theoretical results well, especially in the case of high SNR and large T. In the simulations for the adaptive estimator, we compare the performance of the proposed adaptive algorithm and the hybrid algorithm in time-varying channels, where the channel is generated by a first-order AR model. The plane rotations based EVD tracker used in the simulations is PROTEUS-2 [5]. Five simulation experiments were conducted with different parameter setting. We use an average of the square errors over a large number of iterations to approximate the mean square error of the estimator. The simulation results show that in all the cases the performance of the fully adaptive implementation is still comparable or superior to that of the hybrid scheme while its computational complexity has been greatly reduced.

1.3 Organization of Thesis

The rest of the thesis is organized as follows. In Chapter 2, we formulate the problem of blind channel estimation within the framework of a general system model. In Chapter 3, we propose a generalized subspace-based blind channel estimation algorithm and derive the asymptotic performance properties of the proposed estimator. In Chapter 4, we propose a novel adaptive implementation of the generalized subspace channel estimator. In Chapter 5, we present and discuss the results of the computer experiments for the down-link CDMA system. In Chapter 6, we make a brief summary of the thesis and present some unsolved open problems. This is followed by Appendices that contain the proofs of various theorems.

1.4 Notation

$\mathbf{A}^T$		the transpose of matrix $\mathbf{A}$
$\mathbf{A}^*$		the conjugate of matrix $\mathbf{A}$
$\mathbf{A}^H$		the conjugate transpose of matrix $\mathbf{A}$
$\mathbf{A}^{-1}$		the inverse of non-singular matrix $\mathbf{A}$
$\mathbf{A}^{\dagger}$		the pseudo-inverse of singular matrix $\mathbf{A}$
$E[\mathbf{A}]$		the expectation of matrix $\mathbf{A}$
$Tr[\mathbf{A}]$		the trace of matrix $\mathbf{A}$
$\text{Span}[\mathbf{A}]$		the space spanned by the columns of $\mathbf{A}$
$\hat{\mathbf{A}}$		the estimate of $\mathbf{A}$
$\mathbf{A}(t)$		the sample of time varying variable $\mathbf{A}$ at the t-th time iteration
$\|\|\mathbf{v}\|\|$		the norm of vector $\mathbf{v}$
$\mathbf{I}_L$		the identity matrix with size $L \times L$
$\text{vec}[\mathbf{A}]$	$\triangleq$	$[\mathbf{a}_1^T, \ldots, \mathbf{a}_N^T]^T$, where $\mathbf{A} = [\mathbf{a}_1, \ldots, \mathbf{a}_N]$
$\text{diag}[\mathbf{v}]$	$\triangleq$	$\begin{bmatrix} v_1 & & \\ & \ddots & \\ & & v_M \end{bmatrix}$, where $\mathbf{v} = [v_1, \ldots, v_M]$
$\mathbf{A} \otimes \mathbf{B}$	$\triangleq$	$\begin{bmatrix} a_{1,1}\mathbf{B} & \cdots & a_{1,N}\mathbf{B} \\ \vdots & \ddots & \vdots \\ a_{M,1}\mathbf{B} & \cdots & a_{M,N}\mathbf{B} \end{bmatrix}$, where $\mathbf{A} = \begin{bmatrix} a_{1,1} & \cdots & a_{1,N} \\ \vdots & \ddots & \vdots \\ a_{M,1} & \cdots & a_{M,N} \end{bmatrix}$
$\dfrac{\partial}{\partial \mathbf{v}}$	$\triangleq$	$\dfrac{1}{2} \begin{bmatrix} \frac{\partial}{\partial x_1} - j\frac{\partial}{\partial y_1} \\ \vdots \\ \frac{\partial}{\partial x_M} - j\frac{\partial}{\partial y_M} \end{bmatrix}$, where $\mathbf{v} = \begin{bmatrix} x_1 + jy_1 \\ \vdots \\ x_M + jy_M \end{bmatrix}$
$\delta_{i,j}$	$\triangleq$	$\begin{cases} 0 & i \neq j \\ 1 & i = j \end{cases}$

Chapter 2

Problem Formulation

During the past decade, subspace based channel estimation algorithms have been developed for and applied to various vector channels. Each of these algorithms focuses on the channel estimation problem in a specific communication system. However, the similarities among these algorithms indicate that there must exist some common features of the underlying system models. In this chapter, we intend to formulate a general signal model which is applicable to a multitude of communication systems. In this general model, we incorporate the common features of the applicable systems and leave their unique distinguishing features as user-specified parameters.

In this chapter, we first review the signal models of three existing subspace channel estimation algorithms in Section 2.1. We then propose a general signal model in Section 2.2. The connection between this general model and different specific systems is discussed in Section 2.3. At last, in Section 2.4, we formulate the problem of subspace channel estimation within the framework of the suggested general model.

2.1 Background Review

In this Section, we review three signal models of existing subspace channel estimation algorithms. In Subsection 2.1.1, we study the signal model of CDMA system operating in a frequency selective fading channel with negligible ISI; in Subsection 2.1.2, we consider the signal model of space time block coded CDMA system; in Subsection 2.1.3, we review the signal model of SIMO channels in TDMA system. Several features common among the above discussed systems are summarized in Subsection 2.1.4.

2.1.1 Synchronous CDMA System in a Frequency Selective Fading Channel with Negligible ISI

A synchronous CDMA system operating in a frequency selective fading channel has been studied in [19]. In a direct sequence-spread spectrum (DS-SS) CDMA system, information symbols are modulated by pre-assigned signature waveforms of length L_c. For the q-th user, the normalized signature waveform is represented by $\mathbf{c}^q = [c_1^q, \cdots, c_{L_c}^q]^T$. At time t, the spread transmitted signal of the q-th user $\mathbf{x}^q(t)$ may be represented in vector form as

$$\mathbf{x}^q(t) = \gamma_1^q(t) b^q(t) \mathbf{c}^q \tag{2.1}$$

where $\gamma_1^q(t)$ is the amplitude of the q-th user signal and $b^q(t)$ is the corresponding information bit.

The frequency selective fading channel can be modelled as a FIR filter. When

$\mathbf{x}^q(t)$ is input into the channel, the output signal can be represented as follows:

$$
\begin{aligned}
\mathbf{y}'^q(t) &= \gamma_2(t)[\mathbf{x}^q(t) * \mathbf{h}^q(t)] \\
&= \gamma^q(t)b^q(t)[\mathbf{c}^q * \mathbf{h}^q(t)] \\
&= [y_1^q(t), \cdots, y_{L_c+M-1}^q(t)]^T
\end{aligned}
\tag{2.2}
$$

where $\gamma_2(t)$ is the channel propagation gain, $\mathbf{h}^q(t)$ is an $M \times 1$ discrete normalized channel impulse response vector, $\gamma^q(t) \triangleq \gamma_2(t)\gamma_1^q(t)$ represents the received amplitude, and the operator $*$ is defined as follows:

$$
\mathbf{c}^q * \mathbf{h}^q(t) = \underbrace{\begin{bmatrix}
c_1^q(t) & & & \\
c_2^q(t) & c_1^q(t) & & \\
\vdots & c_2^q(t) & \ddots & \\
c_{L_c}^q(t) & \vdots & \ddots & c_1^q(t) \\
 & c_{L_c}^q(t) & \vdots & c_2^q(t) \\
 & & \ddots & \vdots \\
 & & & c_{L_c}^q(t)
\end{bmatrix}}_{(L_c+M-1)\times M} \mathbf{h}^q(t)
\tag{2.3}
$$

Since the length of $\mathbf{y}'^q(t)$ is $L_c + M - 1$ and the symbol duration is L_c chips, there exists an $(M-1)$-chip overlap, or Inter-Symbol Interference (ISI), between $\mathbf{y}'^q(t-1)$ and $\mathbf{y}'^q(t)$ at the receiver. We may assume $M \ll L_c$ in the case that the time delay spread of the channel is much smaller than the symbol period [19]. To avoid the ambiguity caused by ISI, we consider $\mathbf{y}^q(t)$, the ISI-free section of $\mathbf{y}'^q(t)$:

$$
\mathbf{y}^q(t) \triangleq [y_M^q(t), \cdots, y_{L_c}^q(t)]^T \tag{2.4}
$$

Denote $L \triangleq L_c - M + 1$ as the length of the vector $\mathbf{y}^q(t)$. Then $\mathbf{y}^q(t)$ can be expressed as

$$\mathbf{y}^q(t) = \gamma^q(t) b^q(t) \mathbf{C}^q \mathbf{h}^q(t) \tag{2.5}$$

where matrix $\mathbf{C}^q$ comprises the M-th to the L_c-th rows of the matrix in (2.3):

$$\mathbf{C}^q \triangleq \underbrace{\begin{bmatrix} c_M^q(t) & \cdots & c_1^q(t) \\ c_{M+1}^q(t) & \cdots & c_2^q(t) \\ \vdots & \vdots & \vdots \\ c_{L_c}^q(t) & \cdots & c_L^q(t) \end{bmatrix}}_{L \times M} \tag{2.6}$$

In the synchronous system, the received noisy signal has the following form

$$\mathbf{r}(t) = \sum_{q=1}^{N} \mathbf{y}^q(t) + \mathbf{e}(t) = \sum_{q=1}^{N} \gamma^q(t) b^q(t) \mathbf{C}^q \mathbf{h}^q(t) + \mathbf{e}(t) \tag{2.7}$$

where N is the number of users and $\mathbf{e}(t)$ is a zero-mean white Gaussian noise vector.

In an up-link system, the signals of the q-th user experience a unique channel $\mathbf{h}^q(t)$. In a conventional down-link system[1], the signals of all the users share the same channel, so that

$$\mathbf{h}(t) \triangleq \mathbf{h}^1(t) = \cdots = \mathbf{h}^N(t) \tag{2.8}$$

2.1.2 Space-Time Block Coded CDMA

The channel estimation algorithm in space-time block coded CDMA system has been studied in [28, 34].

[1] Here conventional down-link system refers to that where the base station does not equipped by antenna array.

In a CDMA system, information symbols of multiple users are transmitted through the wireless connection simultaneously. We consider a connection between the transmitter, which is equipped with B transmitting antennae, to a receiver, which has 1 receiving antenna. To obtain the diversity gain of multiple transmitter antennae, the information symbols are firstly encoded by a space-time block encoder. Without loss of generality, we assume that the encoder for each user is identical. If we input F continuous information symbols of the i-th user, $s^{i,1}, \dots, s^{i,F}$, into a space-time block encoder, the output is an $F \times B$ matrix $\mathbf{Y}^i$, where the (j, l)-th element of $\mathbf{Y}^i$ will be transmitted at the j-th symbol duration by the l-th antenna. Here we consider code matrices $\mathbf{Y}^i$ satisfying the real orthogonal design [35]. Recall that a real orthogonal design of size n is an $n \times n$ orthogonal matrix with entries $\pm s^{i,1}, \dots, \pm s^{i,F}$; such an orthogonal design exists only if $n = 2, 4, 8$. Thus we conclude that $\mathbf{Y}^i$ is a square matrix and

$$\mathbf{Y}^i(\mathbf{Y}^i)^H = c\mathbf{I}_F \tag{2.9}$$

where $F = B \in \{2, 4, \text{ or } 8\}$ and c is a constant. The orthogonal design matrix has the following property [9]:

$$\mathbf{Y}^i = \sum_{j=1}^{F} \mathbf{A}_j s^{i,j} \tag{2.10}$$

where $\mathbf{A}_j$ is a matrix with all the entries being 0, +1 or −1.

A unique signature waveform $\mathbf{c}^{i,l} = [c^{i,l}(1), \dots, c^{i,l}(L_c)]$ is assigned to spread the information bits of the i-th user transmitted by the l-th antenna, where L_c is the processing gain. After spreading, the code matrix $\mathbf{Y}^i$ turns into a $L_cF \times F$ matrix $\mathcal{Y}^i$

$$\mathcal{Y}^i = (\mathbf{Y}^i \otimes \mathbf{I}_{L_c})\mathbf{O}^i = \sum_{j=1}^{F} (\mathbf{A}_j \otimes \mathbf{I}_{L_c})\mathbf{O}^i s^{i,j} \tag{2.11}$$

where $\mathbf{O}^i \triangleq \mathrm{diag}[\mathbf{c}^{i,1}, \ldots, \mathbf{c}^{i,F}]$

Assume that the signals of different users are synchronized, then the received signal can be expressed as

$$\mathbf{r} = \sum_{i=1}^{D} \sum_{j=1}^{F} \gamma^i \mathbf{C}(i,j) \mathbf{h}^i s^{i,j} + \mathbf{e} \tag{2.12}$$

where $\mathbf{C}(i,j) \triangleq (\mathbf{A}_j \otimes \mathbf{I}_{L_c}) \mathbf{O}^i$, and where γ^i and $\mathbf{h}^i$ are the channel gain and normalized channel vector of the i-th user, respectively.

Similar to the case in the last subsection, in an up-link system, the signals of the q-th user experiences a unique channel $\mathbf{h}^q(t)$ and in a down-link system, the signals of all the users share the same channels.

2.1.3 SIMO Channel in TDMA System

The Single Input Multiple Output (SIMO) Channel estimation problem has been studied in [22].

We consider a single user time dispersive channel:

$$r(t) = \sum_{m=-\infty}^{\infty} h(t - mT) b_m + e(t) \tag{2.13}$$

where b_m denotes the transmitted symbol at time mT, $e(t)$ is an independent Gaussian noise process, and $h(t)$ is the channel response.

Multiple outputs of the received signal can be obtained by oversampling the received signal in the time domain by a factor L_c. Assuming the channel has a finite time-span, the received signal can be represented as

$$r_n^i = \sum_{m=0}^{M} b_{n-m} h(t + iT/L_c + mT) + e_n^i \tag{2.14}$$

where

$$\begin{aligned} r_n^i &\triangleq r(t + iT/L_c + nT) && (2.15) \\ e_n^i &\triangleq e(t + iT/L_c + nT) && (2.16) \end{aligned}$$

The received signal sequence r_n^i depends on a discrete channel impulse response $\mathbf{h}^i$ defined as follows:

$$\begin{aligned} \mathbf{h}^i &\triangleq [h_0^i, \ldots, h_M^i]^T && (2.17) \\ &\triangleq [h(t + iT/L_c), \ldots, h(t + MT + it/L_c)]^T && (2.18) \end{aligned}$$

We stack signals within N consecutive symbol durations and obtain the sequence

$$\mathbf{r}_n^i \triangleq [r_n^i, \ldots, r_{n-N+1}^i] \tag{2.19}$$

Then

$$\mathbf{r}_n^i = \mathbf{H}_N^i \mathbf{b}_n + \mathbf{e}_n^i \tag{2.20}$$

where

$$\begin{aligned} \mathbf{b}_n &\triangleq [b_n, \ldots, b_{n-N+1}] && (2.21) \\ \mathbf{e}_n &\triangleq [e_n^i, \ldots, e_{n-N+1}^i] && (2.22) \end{aligned}$$

$$\mathbf{H}_N^i \triangleq \underbrace{\begin{bmatrix} h_0^i & \cdots & h_M^i & 0 & \cdots & \cdots & 0 \\ 0 & h_0^i & \cdots & h_M^i & 0 & \cdots & 0 \\ \vdots & & & & & & \vdots \\ 0 & \cdots & \cdots & 0 & h_0^i & \cdots & h_M^i \end{bmatrix}}_{N\times(N+M)} \quad (2.23)$$

Consequently

$$\underline{\mathbf{r}}_n = \mathcal{H}_N \mathbf{b}_n + \underline{\mathbf{e}}_n \quad (2.24)$$

where

$$\underline{\mathbf{r}}_n \triangleq \mathrm{vec}[\mathbf{r}_n^0, \ldots, \mathbf{r}_n^{L_c-1}] \quad (2.25)$$

$$\underline{\mathbf{e}}_n \triangleq \mathrm{vec}[\mathbf{e}_n^0, \ldots, \mathbf{e}_n^{L_c-1}] \quad (2.26)$$

$$\mathcal{H}_N \triangleq [(\mathbf{H}_N^0)^T, \ldots, (\mathbf{H}_N^{L_c-1})^T]^T \quad (2.27)$$

Multiple outputs of the received signal can also be obtained by using oversampling in spacial domain, i.e. using L_c sensors at the receiver. Thus the received signal at the i-th sensor may be represented as

$$r^i(t) = \sum_{m=0}^{M} b_{n-m} h^i(t + mT) + e^i(t) \quad (2.28)$$

In this case, we denote the discrete channel response vector as

$$\mathbf{h}^i \triangleq [h_0^i, \ldots, h_M^i]^T \quad (2.29)$$

$$\triangleq [h^i(t), \ldots, h^i(t + MT)]^T \quad (2.30)$$

Proceeding as in a similar way, the problem naturally takes the same form as (2.24).

We note that the matrix $\mathcal{H}$ has the entries h_m^i, for $i = 0, \ldots, L_c - 1$ and $m = 0, \ldots, M$. By defining the channel vector $\mathbf{h}$ as

$$\mathbf{h} \triangleq \mathrm{vec}[\mathbf{h}^0, \ldots, \mathbf{h}^{L_c-1}] \tag{2.31}$$

and denoting the k-th column of $\mathcal{H}$ as $\bar{\mathbf{h}}_k$, we have

$$\bar{\mathbf{h}}_k = \mathbf{C}_k \mathbf{h} \tag{2.32}$$

where $\mathbf{C}_k$ is a matrix with entries 0 and 1, where there is at most one 1 in each row or column.

Based on the above definition, the signal model in (2.24) can be reformed as

$$\underline{\mathbf{r}}_n = \sum_{k=1}^{N+M} \mathbf{C}_k \mathbf{h} b_{n-k+1} + \underline{\mathbf{e}}_n \qquad n = \infty, \ldots, 0, \ldots, \infty \tag{2.33}$$

2.1.4 Common Features

Although the above signal models are formulated for different communication systems, there exist some common features among them, e.g.

1. The received signals have multiple outputs, which are represented as a vector;

2. The received signals contain multiple simultaneous signal components, which carrier different information symbols;

3. Each information symbol is spread by an effective signature waveform, which is modelled as a product of a user-specified kernel matrix and a channel vector.

In the next section, we propose a general system model applicable to a multitude of communication systems, such as those presented in Subsections 2.1.1 to 2.1.3. This general model preserves the above common features found in these examples, while allowing the treatment of their distinguishing features via user-specified parameters, e.g. kernel matrices. This way, our further work within the framework of the general model, such as the proposition of a new channel estimator, performance analysis of the estimator, etc., will be applicable to any particular system satisfying the general model by properly specifying its parameters.

2.2 General System Model

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

$$\mathbf{r} = \sum_{i=1}^{N} \gamma_i b_i \mathbf{C}_i \mathbf{h}_i + \mathbf{e} \tag{2.34}$$

where N is the number of independent informatin symbols. γ_i is a real-value channel gain, b_i is the i-th information symbol, $\mathbf{C}_i$ is defined as a kernel matrix with size $L \times M$, $\mathbf{h}_i$ is an $M \times 1$ normalized channel vector, 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 and identically distributed with zero mean and equal variance. The additive noise vector $\mathbf{e}$ is zero mean circularly complex Gaussian with covariance matrix $\sigma^2 \mathbf{I}_L$ and is independent from the transmitted symbols b_i. That is, if we define the i-th element of the vector $\mathbf{e}$ as e_i, then

$$\begin{aligned} E[e_i] &= 0 \qquad\qquad i = 1, \ldots, L && (2.35) \\ E[e_i e_j^*] &= \sigma^2 \delta_{i,j} \qquad i, j = 1, \ldots, L && (2.36) \\ E[e_i e_j] &= 0 \qquad\qquad i, j = 1, \ldots, L && (2.37) \end{aligned}$$

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:

$$\begin{aligned} \mathbf{b} &\triangleq [b_1, \ldots, b_N]^T && (2.38) \\ \mathbf{\Gamma} &\triangleq \mathrm{diag}[\gamma_1, \ldots \gamma_N] && (2.39) \\ \mathbf{W} &\triangleq [\mathbf{w}_1, \ldots, \mathbf{w}_N] && (2.40) \end{aligned}$$

where

$$\mathbf{w}_i \triangleq \mathbf{C}_i \mathbf{h}_i \qquad i = 1, \ldots, N \tag{2.41}$$

is the effective signature waveform of the i-th information symbol, i.e. combined effect of channel and kernel matrix as seen by the receiver. Using the above matrix notations, the signal model (2.34) can be expressed more compactly as

$$\mathbf{r} = \mathbf{W}\mathbf{\Gamma}\mathbf{b} + \mathbf{e} = \mathbf{x} + \mathbf{e} \tag{2.42}$$

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

The system model Equation (2.42) explicitly contains the random information symbol vector $\mathbf{b}$, which is convenient for subspace-based analysis (see Section 3.2). However, for the purpose of deriving a channel estimation algorithm and analyzing its performance, it is more appropriate to reformulate (2.34) in a matrix format that explicitly displays the channel vectors.

We assume that the N independent signals (in the sequel, the term *signal* refers to the variables $\gamma_i b_i \mathbf{C}_i \mathbf{h}_i$) experience J different channels, $1 \leq J \leq N$. Then we separate the N signals into J groups, such that the signals in each group share the same channel. We denote the number of independent signals in each group as K^m ($m = 1, \ldots, J$), so that $\sum_{m=1}^{J} K^m = N$. In the m-th group, we use the superscript m to denote group affiliation, as in the common channel parameters γ^m and $\mathbf{h}^m$, and we use the superscript l to further distinguish among the K^m independent signals, as in b^{ml}, $\mathbf{C}^{m,l}$ and $\mathbf{w}^{m,l}$. Finally, introducing the following quantities

$$\mathbf{G}^m \triangleq \gamma^m \sum_{l=1}^{K^m} b^{m,l} \mathbf{C}^{m,l} \tag{2.43}$$

$$\mathbf{G} \triangleq [\mathbf{G}^1, \ldots, \mathbf{G}^J] \tag{2.44}$$

$$\underline{\mathbf{h}}^T \triangleq \mathrm{vec}[\mathbf{h}^1, \ldots, \mathbf{h}^J] \tag{2.45}$$

the received signal vector can also be expressed as

$$\mathbf{r} = \mathbf{G}\underline{\mathbf{h}} + \mathbf{e} \tag{2.46}$$

2.3 Connection to Existing Works

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.

1. Direction of propagation:

    - Conventional downlink: All the signals share the same channel so that there is only one group, as in e.g. [39].
    - 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. [19, 39].

2. Nature of information bits:

    - Inter-symbol Interference (ISI): The entries of vector $\mathbf{b}$ represent consecutive information bits in the data stream, e.g. [22, 29, 39].
    - Multiple Access Interference (MAI): Vector $\mathbf{b}$ contains the simultaneous information bits of the different users, e.g. [19, 2, 39].
    - Space-Time Block Codes (STBC): Vector $\mathbf{b}$ contains the input symbols of a STBC encoder, e.g. [28, 34].
    - Orthogonal transmission: Vector $\mathbf{b}$ is generated by applying a unitary transformation on the original symbols, e.g. [29] where the unitary transformation is IFFT.

3. Nature of the kernel matrix:

    - CDMA: The kernel matrix $\mathbf{C}^i$ is a function of the signature waveform, or spreading code, of the i-th user. This is the case for instance in DS-CDMA with time spreading, e.g. [19, 2, 6, 28], and in MC-CDMA with frequency spreading, e.g. [41, 16, 34].
    - Oversampling in TDMA: This is used in ISI channels where the kernel matrix $\mathbf{C}^i = [\mathbf{0}_{M\times l}, \mathbf{I}_M, \mathbf{0}_{M\times k}]^T$ with $k + l + M = L$, e.g. [22, 29, 42].

4. Nature of the channel vectors:

   - Dispersive Channel: The mth multi-path channel is modelled as a tapped-delay-line [26] with tap coefficient vector $\mathbf{h}^m$ e.g. [19, 2, 39, 28].
   - 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. [28, 34, 19, 6].
   - Multiple carrier: Vector $\mathbf{h}^m$ contains the gains of the different sub-carriers, e.g. [23].

The above list is far from exhaustive, but in some degree it demonstrates the generality of the proposed signal model.

2.4 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 (2.34), say $\mathbf{r}_j$ ($j = 1, \ldots, T$). In this thesis, and without loss of generality, we only consider the problem of estimating one target channel vector. As we explain in Section 3.4, the problem of estimating multiple target channel vectors is equivalent to multiple independent estimation problems for the case of a single channel vector.

We define the estimation error vector as

$$\Delta\mathbf{h}^m \triangleq \hat{\mathbf{h}}^m - \mathbf{h}^m \tag{2.47}$$

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

$$\begin{aligned}
\text{Bias} &\triangleq E[\Delta\mathbf{h}^m] && (2.48)\\
\text{Cov} &\triangleq E[(\Delta\mathbf{h}^m - E[\Delta\mathbf{h}^m])(\Delta\mathbf{h}^m - E[\Delta\mathbf{h}^m])^H] && (2.49)\\
\text{MSE} &\triangleq E[||\Delta\mathbf{h}^m||^2] && (2.50)
\end{aligned}$$

A "good" estimator is one that is unbiased (i.e. Bias $= \mathbf{0}$) and minimizes the covariance and the mean square error.

For blind channel estimation, the transmitted symbols $\mathbf{b}$ is unknown. This is the main difference between blind and training sequence/pilot aided algorithms. To guarantee the estimation of the target channel vector $\mathbf{h}^m$, at least one kernel matrix in the m-th 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 the next Chapter, we formulate a generalized cost function for subspace-based blind channel estimation, which incorporates the set of kernel matrices of the signals sharing the target channel. We then investigate the asymptotic performance of the estimator, including bias, covariance and mean square error (MSE), as the number of independent observations T increases. In particular, we shall show that the performance of the estimator can be improved by increasing the number of kernel matrices.

2.5 Chapter Summary

In this chapter, we discussed the signal model of different applications of subspace channel estimation methods. We began by reviewing the signal models of three specific communication systems. By incorporating the common features of the different signal models, we then proposed a general signal model, which is applicable to a multitude of communication systems. This was demonstrated by establishing the connections between the above general model and some specific examples of communication systems. At last, we formulated the problem of blind subspace channel estimation within the framework of the proposed general model.

Chapter 3

Generalized Blind Subspace Channel Estimation

In this chapter, we propose a generalized blind subspace channel estimator within the framework of the general signal model formulated in the Section 2.2. We also study the asymptotic performance of the proposed estimator with a large number of observed data.

This chapter is organized as follows. In Section 3.1, we firstly review the concept of subspace decomposition of the covariance matrix. Based on this concept, we formulate the theoretical foundation of subspace channel estimation algorithm and study the identifiability (i.e. existence and uniqueness of the solutions) of the subspace channel estimation problem in Section 3.2. In Section 3.3, we 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. In Section 3.4, we investigate the asymptotic performance of the proposed estimator, i.e. bias, covariance, mean square error (MSE) and Cramer-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.

3.1 Subspace Decomposition

Let $\mathbf{R}$ denote the covariance matrix of received signal vector $\mathbf{r}$ in (2.34):

$$\mathbf{R} = E[\mathbf{r}\mathbf{r}^H] = \mathbf{W}\mathbf{\Gamma}^2\mathbf{W}^H + \sigma^2\mathbf{I}_L \tag{3.1}$$

Blind subspace methods exploit the special structure of $\mathbf{R}$ to estimate the channel parameters. Specifically, let us express the EigenValue Decomposition (EVD) (see [10]) of $\mathbf{R}$ in the form

$$\mathbf{R} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H \tag{3.2}$$

where $\mathbf{\Lambda} = \mathrm{diag}[\lambda_1, \ldots, \lambda_L]$ denotes the eigenvalue matrix, with the eigenvalues in a non-increasing order, and $\mathbf{U}$ is a unitary matrix that contains the corresponding eigenvectors. Since the rank of matrix $\mathbf{W\Gamma^2W}^H$ in (3.1) is N, it follows that

$$\lambda_1 \geq \cdots \geq \lambda_N > \lambda_{N+1} = \cdots = \lambda_L = \sigma^2 \tag{3.3}$$

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

$$\mathbf{\Lambda}_s \triangleq \mathrm{diag}[\lambda_1, \ldots, \lambda_N] \tag{3.4}$$

$$\mathbf{\Lambda}_n \triangleq \mathrm{diag}[\lambda_{N+1}, \ldots, \lambda_L] \tag{3.5}$$

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 this notation, the EVD in (3.2) can be expressed in the form

$$\mathbf{R} = \begin{bmatrix} \mathbf{U}_s & \mathbf{U}_n \end{bmatrix} \begin{bmatrix} \mathbf{\Lambda}_s & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_n \end{bmatrix} \begin{bmatrix} \mathbf{U}_s^H \\ \mathbf{U}_n^H \end{bmatrix} \tag{3.6}$$

The columns of $\mathbf{U}_s$ span the so-called signal subspace with dimension N, while those of $\mathbf{U}_n$ span its orthogonal complement, i.e. the noise subspace.

The problem of computing EVD, or more generally Singular Value Decomposition (SVD), has already been deeply studied. The emphasis of the research on this topic is to reduce the huge computational complexity. Currently, the most efficient algorithms may calculate the eigenvalues and eigenvectors with the complexity as $O(L^3)$ [10].

3.2 Theoretical Foundation

The special structure of the covariance matrix of the received signal in (3.1) provides the following important property: the signal subspace is indeed equal to the space spanned by the columns of $\mathbf{W}$:

$$\text{Span}[\mathbf{W}] = \text{Span}[\mathbf{U}_s] \tag{3.7}$$

From the orthogonality between the signal subspace and the noise subspace, it follows that

$$\text{Span}[\mathbf{W}] \perp \text{Span}[\mathbf{U}_n] \tag{3.8}$$

To estimate the target channel vector $\mathbf{h}^m$, which is shared by the signal components in the m-th group, we select $1 \leq P \leq K^m$ effective signature waveforms from the m-th group, say $\mathbf{w}^{m,j}$ $(j = 1, \ldots, P)$ without loss of generality, and construct a matrix

$$\bar{\mathbf{W}} \triangleq [\mathbf{w}^{m,1}, \ldots, \mathbf{w}^{m,P}] \tag{3.9}$$

From the above definition, it follows that $\text{Span}[\bar{\mathbf{W}}] \subseteq \text{Span}[\mathbf{W}]$, Thus according to (3.8), we have

$$\text{Span}[\bar{\mathbf{W}}] \perp \text{Span}[\mathbf{U}_n] \tag{3.10}$$

Consequently

$$\mathbf{U}_n^H \bar{\mathbf{W}} = \mathbf{0} \tag{3.11}$$

Define

$$\mathcal{U}_s \triangleq \mathbf{I}_P \otimes \mathbf{U}_s \tag{3.12}$$

$$\mathcal{U}_n \triangleq \mathbf{I}_P \otimes \mathbf{U}_n \tag{3.13}$$

$$\mathcal{C}^T \triangleq [(\mathbf{C}^{m,1})^T, \ldots, (\mathbf{C}^{m,P})^T] \tag{3.14}$$

and note from (2.41) that

$$\text{vec}[\bar{\mathbf{W}}] = \mathcal{C}\mathbf{h}^m \tag{3.15}$$

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

$$\text{vec}[\mathbf{U}_n^H \bar{\mathbf{W}}] = \mathcal{U}_n^H \text{vec}[\bar{\mathbf{W}}] = \mathcal{U}_n^H \mathcal{C}\mathbf{h}^m = \mathbf{0} \tag{3.16}$$

Theoretically, we may determine the target channel vector by solving (3.16). However, there still exist several practical difficulties in doing that, such as the identifiability of the solution, which will be discussed in the following subsection, and the noise perturbation in $\mathcal{U}_n$ as discussed in Section 3.3.

3.2.1 Identifiability

To determine $\mathbf{h}^m$ as a unique non-trivial solution of (3.16), the matrix $\mathcal{U}_s$ (3.12) and the matrix $\mathcal{C}$ (3.14) must satisfy the identifiability condition, that is: the intersection space of Span[$\mathcal{C}$] and Span[$\mathcal{U}_s$] should have dimension one (see [19]). To understand this condition, we note that (3.16) implies that vector $\mathcal{C}\mathbf{h}^m$ belongs to the subspace Span[$\mathcal{U}_s$], which is orthogonal to Span[$\mathcal{U}_n$]. Thus, if Span[$\mathcal{C}$]$\cup$Span[$\mathcal{U}_s$] has dimension 0, there is no non-trivial solution of (3.16); if the dimension Span[$\mathcal{C}$] $\cup$ Span[$\mathcal{U}_s$] is greater than 1, the solution of (3.16) is not unique (i.e. dimensions of solution space > 1).

Here, the dimensions [1] of Span[$\mathcal{C}$] and Span[$\mathcal{U}_s$] are respectively given by M and NP. Clearly, the dimension of Span[$\mathcal{C}$]$\cup$Span[$\mathcal{U}_s$] cannot exceeds LP, the number of rows of matrices $\mathcal{C}$ and $\mathcal{U}_s$. Thus, the identifiability condition implies that

$$\begin{aligned} M + NP - 1 \leq LP \quad &\Longleftrightarrow \quad N \leq L - \frac{M-1}{P} \\ &\Longleftrightarrow \quad M \leq (L-N)P + 1 \end{aligned} \tag{3.17}$$

From the above inequalities, we conclude that increasing P will 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 large M).

Another problem related to the identifiability issue is that there exists a phase ambiguity in the solution of (3.16), i.e. the solution of (3.16) is $\phi\mathbf{h}^m$, where ϕ is

[1]To simplify the discussion, we assume that matrix $\mathcal{C}$ has full column rank.

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 [7]). Thus, we assume that the phase factor is known exactly in the rest of this Chapter.

3.3 The Algorithm

Ideally, the target vector channel $\mathbf{h}^m$ can be determined exactly by solving (3.16) when the covariance matrix $\mathbf{R} = E[\mathbf{r}\mathbf{r}^H]$, and thus the noise eigenvector matrix $\mathbf{U}_n$ are known accurately. In practice, however, $\mathbf{R}$ 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

$$\hat{\mathbf{R}} = \frac{1}{T}\sum_{j=1}^{T} \mathbf{r}_j \mathbf{r}_j^H \tag{3.18}$$

where $\mathbf{r}_j$ now denotes the received signal vector at the j-th time instant for $j = 1, \ldots, T$ (with similar modifications for other quantities of interest in (2.42)-(2.44), $\mathbf{b} \to \mathbf{b}_j$, $\mathbf{e} \to \mathbf{e}_j$, $\mathbf{G}^m \to \mathbf{G}_j^m$ and $\mathbf{G} \to \mathbf{G}_j$).

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

$$\hat{\mathbf{R}} = \begin{bmatrix} \hat{\mathbf{U}}_s & \hat{\mathbf{U}}_n \end{bmatrix} \begin{bmatrix} \hat{\mathbf{\Lambda}}_s & \mathbf{0} \\ \mathbf{0} & \hat{\mathbf{\Lambda}}_n \end{bmatrix} \begin{bmatrix} \hat{\mathbf{U}}_s^H \\ \hat{\mathbf{U}}_n^H \end{bmatrix} \tag{3.19}$$

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 $\mathcal{U}_s$ (3.12) and $\mathcal{U}_n$ (3.13) are respectively defined as

$$\hat{\mathcal{U}}_n \triangleq \mathbf{I}_P \otimes \hat{\mathbf{U}}_n \tag{3.20}$$

$$\hat{\mathcal{U}}_s \triangleq \mathbf{I}_P \otimes \hat{\mathbf{U}}_s \tag{3.21}$$

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

$$\hat{\mathbf{h}}^m = \arg \min_{||\mathbf{t}||=1} \mathbf{t}^H \mathcal{C}^H \hat{\mathcal{U}}_n \hat{\mathcal{U}}_n^H \mathcal{C} \mathbf{t} \tag{3.22}$$

$$= \arg \min_{||\mathbf{t}||=1} \mathbf{t}^H [\sum_{i=1}^{P} (\mathbf{C}^{m,i})^H \hat{\mathbf{U}}_n \hat{\mathbf{U}}_n^H \mathbf{C}^{m,i}] \mathbf{t} \tag{3.23}$$

Ideally, if $\hat{\mathcal{U}}_n = \mathcal{U}_n$ and the identifiability condition is satisfied, all the eigenvalues of $\mathcal{C}^H \hat{\mathcal{U}}_n \hat{\mathcal{U}}_n^H \mathcal{C}$ in the above criterion are positive except the smallest one, which is equal to 0. 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, (3.16) does not have a (non-trivial) solution, but the target channel vector still can be estimated by minimizing the cost function in (3.22). Thus, we conclude that the optimization criterion in (3.22) is more robust to the perturbation of $\mathcal{U}_n$ than (3.16).

The choice of kernel matrices $\mathbf{C}^{m,i}$ included in the proposed criterion (3.22) is specified by the user, allowing a generalization of previous work. For example, the single signal algorithm in [19] can be obtained as a special case of (30) with $P = K^m = 1$, while the multiple signals algorithms in [22] corresponds to $P = K^m$. Here any value of P between 1 and K^m can be used.

A further modification to the above criterion is motivated by the consideration of performance (see Sections 3.4). Specifically, we shall allow the assignment of different weights to the different terms $(\mathbf{C}^{m,i})^H\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^H\mathbf{C}^{m,i}$, i.e.

$$\hat{\mathbf{h}}^m = \arg\min_{||\mathbf{t}||=1} \mathbf{t}^H[\sum_{i=1}^{P}\alpha^i(\mathbf{C}^{m,i})^H\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^H\mathbf{C}^{m,i}]\mathbf{t} \tag{3.24}$$

where $\alpha^i > 0$ are user-specified weight parameters. We define

$$\mathcal{A} \triangleq \mathrm{diag}[\sqrt{\alpha^1},\ldots,\sqrt{\alpha^P}] \otimes \mathbf{I}_{L-N} \tag{3.25}$$

so that criterion (3.24) can be expressed in matrix form as:

$$\hat{\mathbf{h}}^m = \arg\min_{||\mathbf{t}||=1} \mathbf{t}^H\mathcal{C}^H\hat{\mathcal{U}}_n\mathcal{A}\mathcal{A}^H\hat{\mathcal{U}}_n^H\mathcal{C}\mathbf{t} \tag{3.26}$$

From an algorithmic viewpoint, the solution $\hat{\mathbf{h}}^m$ of (3.26) 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 3.1.

The computational complexity under consideration in this thesis refers to the number of complex multiplications. The computational complexity of each step in the proposed algorithm is explained as follows. In the first step, each product $\mathbf{r}_j\mathbf{r}_j^H$ requires $L^2/2$ complex multiplications. In the second step, the EVD requires $O(L^3)$ complex multiplications as we mentioned before. In the step 9, for each of the P terms of $\alpha^i(\mathbf{C}^{m,i})^H\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^H\mathbf{C}^{m,i}$, we first calculate the matrix $\mathcal{B} \triangleq (\mathbf{C}^{m,i})^H\hat{\mathbf{U}}_n$ which requires $(L-N)ML$ complex multiplications; then we calculate the product $\mathcal{B}\mathcal{B}^H$ with complexity $(L-N)M^2$; at last, we apply the coefficient α^i to the Hermitian matrix $\mathcal{B}\mathcal{B}^H$, which needs $M^2/2$ complex multiplications. The complexity of the EVD in the last step is $O(M^3)$.

We call the proposed estimator in Table 3.1 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 Section 3.4).

Table 3.1 Generalized Blind Subspace Channel Estimation Algorithm

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

[1] The exact figure for the computational complexity of the EVD depend on which specific algorithm is being used (see e.g. [10]).

3.4 Asymptotic Performance Analysis

In this section, we investigate the asymptotic performance of the proposed generalized blind subspace channel estimation algorithm (see Table 3.1). The performance criteria under consideration here are the bias, the covariance and the mean square error of the proposed estimator, as defined in (2.48), (2.49) and (2.50), respectively. We study these performance measures under the assumption that the number of time samples T in (3.18) is large, so that

$$\frac{1}{T}\sum_{j=1}^{T}\mathbf{b}_j\mathbf{b}_j^H \approx \mathbf{I}_N \tag{3.27}$$

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

The main results take the form of Theorems that describe the specific performance properties of the proposed algorithm. The proofs of these theorems are relatively tedious. To simplify the presentation, we have found it appropriate to collect these proofs in Appendix A at the end of the thesis.

Theorem 1 *The proposed generalized estimator $\hat{\mathbf{h}}^m$ is asymptotically unbiased, that is $Bias = \mathbf{0}$, with the covariance*

$$Cov = \frac{\sigma^2}{T}[(\mathcal{C}^H\mathcal{U}_n\mathcal{A})^{\dagger}]^H\mathcal{A}^H\mathbf{\Upsilon}^{-2}\mathcal{A}(\mathcal{C}^H\mathcal{U}_n\mathcal{A})^{\dagger} \tag{3.28}$$

and mean square error

$$MSE = \frac{\sigma^2}{T}Tr[(\mathbf{\Upsilon}^{-1}\mathcal{A})(\mathcal{A}^H\mathcal{U}_n^H\mathcal{C}\mathcal{C}^H\mathcal{U}_n\mathcal{A})^{\dagger}(\mathbf{\Upsilon}^{-1}\mathcal{A})^H] \tag{3.29}$$

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

Proof: See Appendix A Section A.1.

Theorem 1 indicates that the performance of the proposed estimator depends on the user specified parameters, i.e. the weight matrix $\mathcal{A}$ (3.25) and the compounded kernel matrix $\mathcal{C}$ (3.14), 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 mean square error 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. $\mathrm{MSE}(\mathcal{A}, S)$ and $\mathrm{Cov}(\mathcal{A}, S)$.

We begin with the minimization of $\mathrm{MSE}(\mathcal{A}, S)$, which proceeds in two steps. Firstly we minimize this measure by adjusting the weight matrix $\mathcal{A}$ in the case of a fixed set S; secondly we search for a best choice of S to minimize $\mathrm{MSE}(\mathcal{A}, S)$ when the optimal weight matrix determined in the first step is used. Then the resulting choice on the parameters $\mathcal{A}$ and S minimizes $\mathrm{MSE}(\mathcal{A}, S)$.

Theorem 2 $\mathcal{A}_{opt} = c\boldsymbol{\Upsilon}$ *is the optimal weight matrix in the sense of minimizing* $MSE(\mathcal{A}, S)$ *for a fixed set of kernel matrices* S*:*

$$
\begin{aligned}
MSE^o(S) &\triangleq \min_{\mathcal{A}} MSE(\mathcal{A}, S) \\
&= MSE(c\boldsymbol{\Upsilon}, S) = \frac{\sigma^2}{T} Tr[\mathcal{Q}^\dagger]
\end{aligned}
\tag{3.30}
$$

where c *is an arbitrary constant and*

$$
\mathcal{Q} \triangleq \mathcal{C}^H \mathcal{U}_n \boldsymbol{\Upsilon}^2 \mathcal{U}_n^H \mathcal{C} \tag{3.31}
$$

Proof: See Appendix A Section A.2.

Theorem 2 shows that the optimal weights α^i $(i = 1, \ldots, P)$ are proportional to the corresponding received powers $(\gamma^{m,i})^2$.

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 m-th 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 non-empty subsets as S_q $(q = 1, \ldots, Q)$:

$$\bigcup_{q=1}^{Q} S_q = S \tag{3.32}$$

$$S_p \cap S_q = \emptyset \quad \text{for any } p \neq q \tag{3.33}$$

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

$$MSE^o(S) < MSE^o(S_q) \tag{3.34}$$

Proof: See Appendix A Section A.3.

The above theorem implies qualitatively that enlarging the set of kernel matrices S in the estimator will decrease its mean square error. Consequently the minimum mean square error is achieved when the estimator utilizes the kernel matrices of all the signal components in this group, i.e. $S = U$:

$$\begin{aligned} \mathrm{MSE}^o &\triangleq \min_{S \subseteq U} \mathrm{MSE}^o(S) = \mathrm{MSE}^o(U) \\ &= \frac{\sigma^2}{T} Tr[(\mathcal{Q}^m)^\dagger] \end{aligned} \tag{3.35}$$

where

$$\mathcal{Q}^m \triangleq (\mathcal{C}^m)^H \mathcal{U}_n^m (\mathbf{\Upsilon}^m)^2 (\mathcal{U}_n^m)^H \mathcal{C}^m \tag{3.36}$$

$$\mathcal{C}^m \triangleq [(\mathbf{C}^{m,1})^T, \ldots, (\mathbf{C}^{m,K^m})^T]^T \tag{3.37}$$

$$\mathcal{U}_n^m \triangleq \mathbf{I}_{K^m} \otimes \mathbf{U}_n \tag{3.38}$$

$$\mathbf{\Upsilon}^m \triangleq \mathrm{diag}[\gamma^{m,1}, \ldots, \gamma^{m,K^m}] \otimes \mathbf{I}_{L-N} \tag{3.39}$$

The next theorem provides a better indication of the relationship between the achievable mean square error $\mathrm{MSE}^o(S)$ and the individual mean square errors $\mathrm{MSE}^o(S_q)$ for the subsets of kernel matrices forming the partition in (3.32)-(3.33).

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

$$MSE^o(S) \leq \frac{\sum_{q=1}^{Q} c_q^2 MSE^o(S_q)}{(\sum_{q=1}^{Q} c_q)^2} \tag{3.40}$$

Proof: See Appendix A Section A.4.

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 mean square error of the single signal estimator (e.g. [19]) applied on $\mathbf{C}^{m,q}$. Thus according to Theorem 4

$$\mathrm{MSE}^o(S) \leq \frac{1}{P^2} \sum_{q=1}^{P} \mathrm{MSE}^o(S_q) = \frac{\overline{\mathrm{MSE}}}{P} \tag{3.41}$$

where $\overline{\mathrm{MSE}} \triangleq \frac{1}{P} \sum_{q=1}^{P} \mathrm{MSE}^o(S_q)$ denotes the average mean square error of single signal estimator 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 mean square errors.

Based on the above considerations, we suggest the following principles for minimizing the mean square error of the proposed estimator:

1. *Choose the weights proportional to the received powers;*

2. *Include the maximum possible number of kernel matrices.*

Clearly, as can be seen from Table 1, increasing the number of kernel matrices will entail an additional computational cost.

We now turn our attention to the optimization of the covariance of the proposed estimator, as defined in (2.49). So 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 Cramer-Rao bound (CRB) to judge the optimality of the parameter choice obtained in the case of mean square error. That is: if the covariance matrix with the parameters $\mathcal{A} = c\boldsymbol{\Upsilon}$ 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. [38]) is no longer applicable. The CRB for parameter estimation under constraints was recently given in [31], where a so-called *constrained* CRB is derived which depends on the specific algebraic constraints imposed on the estimated parameters. In [7], 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*

$$CRB_{C,\underline{\mathbf{h}}} = \frac{\sigma^2}{T} diag[(\mathcal{Q}^1)^\dagger, \ldots, (\mathcal{Q}^J)^\dagger] \tag{3.42}$$

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

$$CRB_{C,\mathbf{h}^m} = \frac{\sigma^2}{T} (\mathcal{Q}^m)^\dagger \tag{3.43}$$

Proof: See Appendix A Section A.5.

From the previously derived expression (3.28) 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 mean square error, but also minimizes the covariance of the estimator. Finally, we note that the minimal constrained CRB in (3.42) 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.

3.5 Chapter Summary

In this chapter, we proposed the generalized blind subspace channel estimation algorithm. We first introduced the subspace concepts and properties. Based on them, we formulated the theoretical foundation of the subspace channel estimation algorithm. Through the study of the identifiability, we showed that by enlarging the set of the kernel matrices, we may enable the estimator to identify a longer channel and/or to be used in a system having larger number of independent signals. With the consideration of robustness to the error in the data correlation matrix, we proposed a generalized blind subspace channel estimator by minimizing a novel cost function, which incorporates a set of kernel matrices via a weighted projection error. Through an asymptotic performance analysis, we showed that the performance of the proposed estimator can be optimized by utilizing the kernel matrices of all the signal components sharing the target channel and by using a special set of weights in the cost function.

Chapter 4

Adaptive Implementation

In the previous chapter, the proposed generalized channel estimation algorithm (see Table 3.1) and its performance analysis were based on the assumption that the channel does not vary within an observation window of T time samples (i.e. block stationary). However, in some practical applications, the systems of interest operate in dynamic signal environments, where the target channel varies from time to time. To estimate the time varying target channel in real time (i.e. updating the estimation on each time iteration), we consider the adaptive implementation of the proposed generalized subspace channel estimation algorithm in this chapter. The most important issue under consideration here is the computational complexity of the adaptive implementation, which greatly impacts its practicality for real-time applications.

This chapter is organized as follows. In Section 4.1, we review a class of subspace trackers, which can provide the required subspace information with low complexities. In Section 4.2, based on this class of subspace trackers, we propose a novel low-complexity adaptive implementation of the generalized subspace channel estimation algorithm presented in Chapter 3.

4.1 Spherical Subspace Tracking Techniques

In dynamic signal environments, the subspace information needs to be updated when the new data becomes available. In such cases, the estimation method of the covariance matrix of the received signal based on a rectangular window (3.18) is no longer applicable. Instead, the Exponentially Weighted Moving Averaging (EWMA) window is commonly used to estimate the covariance matrix as:

$$\hat{\mathbf{R}}(t) = \alpha\hat{\mathbf{R}}(t-1) + (1-\alpha)\mathbf{r}(t)\mathbf{r}(t)^H \tag{4.1}$$

where $\hat{\mathbf{R}}(t)$ denotes the estimate of the $L \times L$ covariance matrix at the t-th time iteration, $\mathbf{r}(t)$ is the $L \times 1$ received signal vector at the t-th time iteration, and α is the forgetting factor used to deemphasize the effect of the past data, which takes the value in the interval $(0, 1)$.

The subspace information can be generated by applying an EVD to $\hat{\mathbf{R}}(t)$ at each time iteration. However, the computational complexity of each EVD is $O(L^3)$, which is too expensive for the online algorithms. To update the subspace information cheaply, several subspace tracking algorithms with much less complexity have been developed (see e.g. [5, 27, 8, 43, 30, 32, 24]), among which a group of so-called spherical subspace trackers are of special interest (e.g. [5, 27, 8, 24]).

The spherical subspace trackers utilize the property that the noise eigenvalues are equal to each other due to the whiteness of the noise vector (see (3.3)). Thus, in practice, the spherical subspace trackers only track one noise eigenvalue, say $\hat{\lambda}_A(t)$, which can be calculated as the average value of the power lying along each noise eigenvector [4]. As a result, the estimated noise subspace can be replaced by a spherical subspace, where the orthogonal basis of this subspace, say $\hat{\mathbf{U}}_n(t)$, can be selected arbitrarily. When new data arrive, it is possible to make the component of the new data vector within the noise subspace lie on a single noise eigenvector by wisely choosing $\hat{\mathbf{U}}_n(t)$. Then, the dimension of the EVD updating is reduced from L to $N+1$ and this updating can be achieved with a complexity as low as $O(NL)$.

Given the estimated signal eigenvector matrix $\hat{\mathbf{U}}_s(t-1)$, estimated signal eigenvalue matrix $\hat{\mathbf{\Lambda}}_s(t-1)$, estimated noise eigenvalue $\hat{\lambda}_A(t-1)$ and the incoming data vector $\mathbf{r}(t)$, the spherical subspace tracker usually update the subspace information as follows [4]:

Data projection:

$$\begin{aligned} \mathbf{y}_s(t) &= \hat{\mathbf{U}}_s(t-1)^H \mathbf{r}(t) && (4.2)\\ \mathbf{r}_n(t) &= \mathbf{r}(t) - \hat{\mathbf{U}}_s(t-1)\mathbf{y}_s(t) && (4.3)\\ y_A(t) &= ||\mathbf{r}_n(t)|| && (4.4)\\ \hat{\mathbf{u}}_n^b(t) &= \mathbf{r}_n(t)/y_A(t) && (4.5) \end{aligned}$$

where $\hat{\mathbf{u}}_n^b(t)$ is the noise eigenvector before updating.

Eigenvector updating:

$$[\hat{\mathbf{U}}_s(t), \hat{\mathbf{u}}_n^a(t)] = \mathcal{T}[\hat{\mathbf{U}}_s(t-1), \hat{\mathbf{u}}_n^b(t)] \tag{4.6}$$

where $\hat{\mathbf{u}}_n^a(t)$ is the noise eigenvector after updating.

Eigenvalue updating:

$$\hat{\lambda}_i(t) = \alpha\hat{\lambda}_i(t-1) + (1-\alpha)|y_i(t)|^2 \tag{4.7}$$

$$\hat{\lambda}_A(t) = \alpha\hat{\lambda}_A(t-1) + (1-\alpha)\frac{|y_A(t)|^2}{L-N} \tag{4.8}$$

where $\hat{\lambda}_i(t)$ is the i-th estimated signal eigenvalue at time iteration t and $y_i(t)$ is the i-th element of the vector $\mathbf{y}_s(t)$. Note that the noise eigenvalue $\hat{\lambda}_A(t)$ is updated by the average value of the power lying in the noise subspace.

The exact form of the eigenvector updating function $\mathcal{T}[\cdot]$ in (4.6) depends on the specific spherical subspace tracker being used. Of special interest here are the so-called plane rotation-based subspace/EVD trackers, including PROTEUS [5], RO-FST [27], etc. The common feature of these trackers is that the signal subspace eigenvectors are updated as

$$[\hat{\mathbf{U}}_s(t), \hat{\mathbf{u}}_n^a(t)] = [\hat{\mathbf{U}}_s(t-1), \hat{\mathbf{u}}_n^b(t)] \prod_i \bar{\mathbf{G}}_i \tag{4.9}$$

where $\{\bar{\mathbf{G}}_i\}$ is a sequence of $O(N)$ plane (or Givens) rotations which depend on the specific subspace/EVD tracker being used.

Plane rotation matrix $\bar{\mathbf{G}}_i$ here is defined as a matrix similar to an $(N+1)\times(N+1)$ identity matrix except for four elements: the (j,j)-th element is c_i, the (j,k)-th element is s_i, the (k, j)-th element is $-s_i$ and the (k, k)-th element is c_i, namely:

$$\bar{\mathbf{G}}_i = \begin{bmatrix} 1 & \cdots & 0 & \cdots & 0 & \cdots & 0 \\ \vdots & \ddots & \vdots & & \vdots & & \vdots \\ 0 & \cdots & c_i & \cdots & s_i & \cdots & 0 \\ \vdots & & \vdots & \ddots & \vdots & & \vdots \\ 0 & \cdots & -s_i & \cdots & c_i & \cdots & 0 \\ \vdots & & \vdots & & \vdots & \ddots & \vdots \\ 0 & \cdots & 0 & \cdots & 0 & \cdots & 1 \end{bmatrix} \tag{4.10}$$

where c_i and s_i are real parameters respectively defined as

$$c_i = \cos\theta_i \tag{4.11}$$

$$s_i = \sin\theta_i \tag{4.12}$$

As a result, it follows that the matrix $\bar{\mathbf{G}}_i$ is unitary, i.e.

$$\bar{\mathbf{G}}_i \bar{\mathbf{G}}_i^H = \mathbf{I}_{N+1} \tag{4.13}$$

The angle parameter θ_i and the position parameters j and k are functions of the subspace information before updating, i.e. $\hat{\mathbf{U}}_s(t-1)$, $\hat{\mathbf{\Lambda}}_s(t-1)$, $\hat{\lambda}_A(t-1)$ and the new data vector $\mathbf{r}(t)$. It is convenient to introduce

$$\bar{\mathbf{G}} \triangleq \prod_i \bar{\mathbf{G}}_i \tag{4.14}$$

as a compact notation for the product of notations in (4.9). We then have

$$\bar{\mathbf{G}}\bar{\mathbf{G}}^H = \mathbf{I}_{N+1} \tag{4.15}$$

4.2 Adaptive Subspace Channel Estimation

Here, we consider the signal model (2.34) in a dynamic signal environment, where the $M \times 1$ channel vector $\mathbf{h}^m(t)$ is assumed to be a time varying parameter.

Define the following variables for convenience:

$$\bar{\mathbf{A}} \triangleq \mathrm{diag}[\sqrt{\alpha^1}, \ldots, \sqrt{\alpha^P}] \otimes \mathbf{I}_L \tag{4.16}$$

$$\bar{\mathbf{w}}(t) \triangleq \bar{\mathbf{A}}\mathrm{vec}[\bar{\mathbf{W}}(t)] \tag{4.17}$$

$$\bar{\mathcal{C}} \triangleq \bar{\mathbf{A}}\mathcal{C} \tag{4.18}$$

where $\bar{\mathbf{A}}$ is a $PL \times PL$ matrix, $\bar{\mathbf{W}}(t)$ is the time varying version of $\bar{\mathbf{W}}$ in (3.9) with size $L \times P$, $\mathcal{C}$ is defined in (3.14) with size $PL \times M$ and $\bar{\mathcal{C}}$ has the same size as $\mathcal{C}$.

The proposed subspace channel estimation algorithm in Table 3.1 estimates the target channel by calculating the eigenvector corresponding to the smallest eigenvalue of $\mathcal{C}^H\hat{\mathcal{U}}_n\mathcal{A}\mathcal{A}\hat{\mathcal{U}}_n^H\mathcal{C}$, which can be reformed as follows:

$$\begin{aligned}\mathcal{C}^H\hat{\mathcal{U}}_n\mathcal{A}\mathcal{A}\hat{\mathcal{U}}_n^H\mathcal{C} &= \sum_{i=1}^{P}\alpha^i(\mathbf{C}^{m,i})^H\hat{\mathbf{U}}_n\hat{\mathbf{U}}_n^H\mathbf{C}^{m,i} \\ &= \bar{\mathcal{C}}^H\hat{\mathcal{U}}_n\hat{\mathcal{U}}_n^H\bar{\mathcal{C}} \end{aligned} \tag{4.19}$$

Here we begin from a different viewpoint. From (3.9), (3.10), (3.12) and (4.17), it can be verified that

$$\bar{\mathbf{w}}(t) \in \text{Span}[\mathbf{I}_P \otimes \bar{\mathbf{W}}(t)] \subseteq \text{Span}[\mathcal{U}_s(t)]. \tag{4.20}$$

and from (3.15), (4.17) and (4.18), we find

$$\bar{\mathbf{w}}(t) = \bar{\mathbf{A}}\text{vec}[\bar{\mathbf{W}}(t)] = \bar{\mathbf{A}}\mathcal{C}\mathbf{h}^m = \bar{\mathcal{C}}\mathbf{h}^m \in \text{Span}[\bar{\mathcal{C}}] \tag{4.21}$$

Thus, $\bar{\mathbf{w}}(t)$ belongs to the intersection of Span$[\bar{\mathcal{C}}]$ and Span$[\mathcal{U}_s(t)]$.

$$\bar{\mathbf{w}}(t) \in \text{Span}[\bar{\mathcal{C}}] \cap \text{Span}[\mathcal{U}_s(t)] \tag{4.22}$$

Thus, we may uniquely determine the vector $\bar{\mathbf{w}}(t)$ as the intersection of Span$[\bar{\mathcal{C}}]$ and Span$[\mathcal{U}_s(t)]$ when this intersection space has dimension one [19].

A standard method to compute the intersection of two subspaces is given in [10].

1. A QR decomposition is applied to $\bar{\mathcal{C}}$, i.e.

$$\bar{\mathcal{C}} = \bar{\mathcal{Q}}\bar{\mathcal{R}} \tag{4.23}$$

   where the columns of the $LP \times M$ matrix $\bar{\mathcal{Q}}$ form an orthonormal basis of Span$[\mathcal{C}]$ and the matrix $\bar{\mathcal{R}}$ is an $M \times M$ upper-triangular matrix.

2. Find $\mathbf{h}_R^m(t)$ as the the dominant left singular vector of $\bar{\mathcal{Q}}^H\mathcal{U}_s(t)$.

3. Vector $\mathbf{w}(t)$ is given by

$$\bar{\mathbf{w}}(t) = \bar{\mathcal{Q}}\mathbf{h}_R^m(t) \tag{4.24}$$

In most situations, what is of interests is the target channel vector $\mathbf{h}^m$. From

(4.21) and (4.24), it follows that

$$\bar{\mathbf{w}}(t) = \bar{\mathcal{C}}\mathbf{h}^m = \bar{\mathcal{Q}}\bar{\mathcal{R}}\mathbf{h}^m = \bar{\mathcal{Q}}\mathbf{h}_R^m(t) \tag{4.25}$$

Since the $LP \times M$ matrix $\bar{\mathcal{Q}}$ is of full column rank and the square matrix $\bar{\mathcal{R}}$ is non-singular, the target channel can be determined by

$$\mathbf{h}^m = \bar{\mathcal{R}}^{-1}\mathbf{h}_R^m(t) \tag{4.26}$$

4.2.1 Hybrid Algorithm

In practice, only $\hat{\mathcal{U}}_s(t)$, a noisy estimate of $\mathcal{U}_s(t)$, is available. Let $\hat{\mathbf{w}}(t)$, $\hat{\mathbf{h}}_R^m(t)$ and $\hat{\mathbf{h}}^m(t)$ respectively denote the noisy estimates of $\bar{\mathbf{w}}(t)$, $\mathbf{h}_R^m(t)$ and $\mathbf{h}^m(t)$ obtained from the above calculations when $\hat{\mathcal{U}}_s(t)$ instead of $\mathcal{U}_s(t)$ is being used.

A channel estimation algorithm based on the above derivation is presented in Table 4.1. We refer to this algorithm as an hybrid (adaptive-batch) approach because the signal subspace eigenvectors are updated adaptively with a subspace tracker but the subspace intersection (4.22) is computed using an exact (batch) Singular Value Decomposition (SVD).

In practice, the length of channel vector M is usually smaller than the dimension of the observation space L. In this case, the computational bottleneck in the this algorithm is that of constructing the matrix $\bar{\mathcal{Q}}^H\hat{\mathcal{U}}_s(t)$ in the 3-rd step. If $M > L$, we may apply the power method [10] to iteratively update the estimate of the channel vector $\hat{\mathbf{h}}^m(t)$ with the complexity $O(PNM)$ (details of the power method is discussed below). Thus the 3-rd step is still the computational bottleneck of the hybrid algorithm.

Table 4.1 Hybrid Channel Estimation Algorithm

Step	**Complexity**
1. $[\bar{\mathcal{Q}}, \bar{\mathcal{R}}] =$ QR decomposition of $\bar{\mathcal{C}}$	Negligible[a]
FOR $t = 1, 2, \ldots$	
2. $\hat{\mathbf{U}}_s(t)$ is given by the subspace tracker	$O(NL)$[b]
3. Construct the matrix $\bar{\mathcal{Q}}^H \hat{\mathcal{U}}_s(t)$	$PNML$
4. Find $\hat{\mathbf{h}}_R^m(t)$ the dominant left singular vector of $\bar{\mathcal{Q}}^H \hat{\mathcal{U}}_s(t)$	$O(PNM^2)$[c]
5. $\hat{\mathbf{h}}^m(t) = \bar{\mathcal{R}}^{-1} \hat{\mathbf{h}}_R^m(t)$	M^2
END	

[a] This step can be calculated off-line.
[b] The exact computational complexity depends on which specific subspace tracker is being used.
[c] The exact computational complexity depends on which specific SVD algorithm is being used.

4.2.2 Adaptive Algorithm

Here, we propose a new adaptive channel estimation algorithm by utilizing the special updating form of plane rotation-based subspace/EVD trackers in (4.9),

Note that the dominant left singular vector of $\bar{\mathcal{Q}}^H \hat{\mathcal{U}}_s(t)$ is identical to the dominant eigenvector of the following $M \times M$ matrix

$$D(t) \triangleq \bar{\mathcal{Q}}^H \hat{\mathcal{U}}_s(t) \hat{\mathcal{U}}_s(t)^H \bar{\mathcal{Q}} \tag{4.27}$$

By defining

$$\begin{aligned} \bar{\mathcal{Q}}^T &\triangleq [\bar{\mathcal{Q}}_1^T, \ldots, \bar{\mathcal{Q}}_P^T] &(4.28)\\ D_i(t) &\triangleq \bar{\mathcal{Q}}_i^H \hat{\mathbf{U}}_s(t) \hat{\mathbf{U}}_s(t)^H \bar{\mathcal{Q}}_i &(4.29)\end{aligned}$$

we have

$$D(t) = \sum_{i=1}^{P} D_i(t) \tag{4.30}$$

To construct $D(t)$ directly requires a huge computational complexity $(PNML + PNM^2/2)$. By using the special updating form in (4.9) and the property of plane rotations in (4.15), we may develop a simple recursive update for $D(t)$. We consider the matrix $D_i(t)$

$$\begin{aligned}
D_i(t) &= \bar{\mathcal{Q}}_i^H \begin{bmatrix} \hat{\mathbf{U}}_s(t) & \mathbf{u}_n^a(t) \end{bmatrix} \begin{bmatrix} \hat{\mathbf{U}}_s(t)^H \\ \mathbf{u}_n^a(t)^H \end{bmatrix} \bar{\mathcal{Q}}_i \\
&\quad -\bar{\mathcal{Q}}_i^H \mathbf{u}_n^a(t)\mathbf{u}_n^a(t)^H \bar{\mathcal{Q}}_i \\
&= \bar{\mathcal{Q}}_i^H \begin{bmatrix} \hat{\mathbf{U}}_s(t-1) & \mathbf{u}_n^b(t) \end{bmatrix} \bar{\mathbf{G}}\bar{\mathbf{G}}^H \begin{bmatrix} \hat{\mathbf{U}}_s(t-1)^H \\ \mathbf{u}_n^b(t)^H \end{bmatrix} \bar{\mathcal{Q}}_i \\
&\quad -\bar{\mathcal{Q}}_i^H \mathbf{u}_n^a(t)\mathbf{u}_n^a(t)^H \bar{\mathcal{Q}}_i \\
&= \bar{\mathcal{Q}}_i^H \hat{\mathbf{U}}_s(t-1)\hat{\mathbf{U}}_s(t-1)^H \bar{\mathcal{Q}}_i + \bar{\mathcal{Q}}_i^H \mathbf{u}_n^b(t)\mathbf{u}_n^b(t)^H \bar{\mathcal{Q}}_i \\
&\quad -\bar{\mathcal{Q}}_i^H \mathbf{u}_n^a(t)\mathbf{u}_n^a(t)^H \bar{\mathcal{Q}}_i \\
&= D_i(t-1) + \mathbf{v}_i^b(t)\mathbf{v}_i^b(t)^H - \mathbf{v}_i^a(t)\mathbf{v}_i^a(t)^H \qquad (4.31)
\end{aligned}$$

where

$$\mathbf{v}_i^b(t) \triangleq \bar{\mathcal{Q}}_i^H \mathbf{u}_n^b(t) \qquad (4.32)$$

$$\mathbf{v}_i^a(t) \triangleq \bar{\mathcal{Q}}_i^H \mathbf{u}_n^a(t) \qquad (4.33)$$

are vectors with size $M \times 1$. Then $D(t)$ can be updated recursively as:

$$D(t) = D(t-1) + \sum_{i=1}^{P} [\mathbf{v}_i^b(t)\mathbf{v}_i^b(t)^H - \mathbf{v}_i^a(t)\mathbf{v}_i^a(t)^H] \qquad (4.34)$$

By using (4.32)-(4.34), the complexity of updating $D(t)$ is reduced from $PNML + PNM^2/2$ to $2PML + PM^2$.

A further reduction in complexity can be achieved by using the power method to search the dominant eigenvector of $D(t)$ iteratively [10]. In the present application, where the true channel estimate is varying slowly over time, a single iteration of the power method can be applied at each time step to update $\hat{\mathbf{h}}_R^m(t)$, the dominant eigenvector of $D(t)$. Specifically:

$$\hat{\mathbf{h}}_R^m(t) = D(t)\hat{\mathbf{h}}_R^m(t-1) \tag{4.35}$$

$$\hat{\mathbf{h}}_R^m(t) = \frac{\hat{\mathbf{h}}_R^m(t)}{||\hat{\mathbf{h}}_R^m(t)||} \tag{4.36}$$

In a stationary environment, the power method will converge to the dominant eigenvector of $D \equiv D(t)$ when the initialization is not orthogonal to the eigenvector [10].

Based on (4.34), (4.35) and (4.36), a fully adaptive subspace channel estimation algorithm is developed in Table 4.2. The complexity of this adaptive algorithm at each iteration (not including that of subspace tracker) is $2PLM + (P+2)M^2 + 2M$, which is an order of magnitude lower than that of the hybrid algorithm. A comparison of the complexity for both algorithms is presented in Table 4.2.2 for reference. We note that the power method can also be applied to the 4-th step of the hybrid algorithm (see Table 4.1). By doing so, the complexity of this step can be reduced from $O(PNM^2)$ to $O(PNM)$. However, in this case the complexity of the 3-rd step ($O(PNML)$) becomes the computation bottleneck of the algorithm, which is still an order of magnitude larger than that of the adaptive algorithm (Table 4.2). Thus we conclude that the power method can not reduce the computational complexity of the hybrid algorithm to the same order as that of the adaptive algorithm.

Table 4.2 Adaptive Channel Estimation Algorithm

Step	**Complexity**
1. $[\bar{\mathcal{Q}}, \bar{\mathcal{R}}] =$ QR decomposition of $\bar{\mathcal{Q}}$	Negligible
2. $\hat{\mathbf{U}}_s(0)$ is given by the subspace tracker	Negligible
3. $\hat{\mathbf{h}}_R^m(0)$: an arbitrary vector	
4. $D(0) = \bar{\mathcal{Q}}^H \hat{\mathcal{U}}_s(0) \hat{\mathcal{U}}_s(0)^H \bar{\mathcal{Q}}$	Negligible
FOR $t = 1, 2, \ldots$	
5. $\hat{\mathbf{U}}_s(t)$, $\mathbf{u}_n^b(t)$ and $\mathbf{u}_n^a(t)$ are updated by the subspace tracker	$O(NL)$
6. $\mathbf{v}_i^b(t) = \bar{\mathcal{Q}}_i^H \mathbf{u}_n^b(t)$, $i = 1, \ldots, P$	PML
7. $\mathbf{v}_i^a(t) = \bar{\mathcal{Q}}_i^H \mathbf{u}_n^a(t)$, $i = 1, \ldots, P$	PML
8. $D(t) = D(t-1) + \sum_{i=1}^{P} [\mathbf{v}_i^b(t)\mathbf{v}_i^b(t)^H - \mathbf{v}_i^a(t)\mathbf{v}_i^a(t)^H]$	PM^2
9. $\hat{\mathbf{h}}_R^m(t) = D(t)\hat{\mathbf{h}}_R^m(t-1)$	M^2
10. $\hat{\mathbf{h}}_R^m(t) = \hat{\mathbf{h}}_R^m(t)/\|\|\hat{\mathbf{h}}_R^m(t)\|\|$	$2M$
11. $\hat{\mathbf{h}}^m(t) = \bar{\mathcal{R}}^{-1}\hat{\mathbf{h}}_R^m(t)$	M^2
END	

Table 4.3 Complexity Comparison

Algorithm	Complexity
Hybrid Algorithm	$PNML + O(PNM^2) + M^2 + O(NL)$
Adaptive Algorithm	$2PML + PM^2 + M^2 + 2M + O(NL)$

4.3 Chapter Summary

In this chapter, we have studied the adaptive implementation of the proposed generalized subspace blind channel estimation algorithm. We firstly reviewed the spherical subspace tracking methods, which can update the desired subspace information with low complexity. Then we proposed a novel low-complexity subspace implementation of subspace channel estimator by utilizing the special updating feature of plane rotation based spherical subspace trackers. The complexity of the resulting adaptive estimator is an order of magnitude lower than the hybrid estimator, i.e. the adaptive subspace tracker followed by batch estimation at each time step.

Chapter 5

Computer Experiments

In this Chapter, we present and discuss the simulation results of the proposed channel estimator in a down-link synchronous CDMA system operating in a frequency selective fading channel with negligible ISI, which satisfies the general model formulated in 2.2. In Section 5.1, we describe the specific signal model of the system under consideration. In Section 5.2, we show the simulation results of the generalized estimator proposed in Chapter 3 (batch algorithm). In Section 5.3, we show the simulation results of the adaptive implementation proposed in Chapter 4.

5.1 System Model

Here, we briefly review the signal model of the down-link CDMA system operating in frequency selective fading channel with negligible ISI, as previously introduced in Subsection 2.1.1.

Consider a synchronous down-link DS-CDMA connection from a base station to N remote users. The information bit to the i-th user b_i is spread by a unique spreading code $\mathbf{c}^i \triangleq [c_1^i, \dots, c_{L_c}^i]^H$, where L_c is the processing gain. The frequency-selective channel is modelled as an FIR filter. The normalized coefficient vector of the filter is represented by vector $\mathbf{h}$ with size $M \times 1$. The kernel matrix of the i-th user $\mathbf{C}^i$ is an $(L_c - M + 1) \times M$ Toeplitz matrix with the first column equal to $[c_M^i, \dots, c_{L_c}^i]^T$ and the first row equal to $[c_M^i, \dots, c_1^i]$ [19]. Assuming the received amplitude of the i-th user is γ^i and the signal of all the users are synchronized, the received signal can be represented as

$$\mathbf{r} = (\sum_{i=1}^{N} \gamma^i \mathbf{C}^i b^i)\mathbf{h} + \mathbf{e} \tag{5.1}$$

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

5.2 Simulations for Batch Algorithm

Computer experiments are conducted to verify the theoretical performance results of the batch algorithm derived in Section 3.4. In the simulations, we consider a locally stationary environment, where within T samples, the channel vector $\mathbf{h}(t)$ and the received amplitudes $\gamma^i(t)$, $i = 1, \dots, N$ do not change, i.e. $\mathbf{h} \triangleq \mathbf{h}(1) = \dots, \mathbf{h}(T)$ and $\gamma^i \triangleq \gamma^i(1) = \dots, \gamma^i(T)$.

In the simulations, the following parameter values are used: the information bits are BPSK modulated (± 1), 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 mean square error.

According to the analysis in Section 3.4, the asymptotic MSE performance of the proposed estimator has the following properties: From (3.30), it follows that

$$\mathrm{MSE}^o(S^4) = \mathrm{MSE}(\boldsymbol{\Upsilon}, S^4) \leq \mathrm{MSE}(\mathbf{I}_{4(L-N)}, S^4) \tag{5.2}$$

where $\boldsymbol{\Upsilon} \triangleq \mathrm{diag}[\gamma^1, \ldots, \gamma^4] \otimes \mathbf{I}_{4(L-N)}$ and $L = L_c - M + 1$. From (3.34), it follows that

$$\mathrm{MSE}^o(S^4) < \mathrm{MSE}^o(S^3) < \mathrm{MSE}^o(S^2) < \mathrm{MSE}^o(S^1) \tag{5.3}$$

Finally, from (3.41), we have

$$\mathrm{MSE}^o(S^4) \leq \frac{1}{4^2} \sum_{i=1}^{4} \mathrm{MSE}^o(\{\mathbf{C}^i\}) \tag{5.4}$$

The simulation results are presented in Fig. 5.1 to 5.6. In all the figures, both theoretical and experimental results are illustrated, where the theoretical results are calculated according to (3.29) and the experimental results are given by the Monte-Carlo simulations. Fig. 5.1 to 5.3, respectively, show the MSEs in (5.2), (5.3) and (5.4) plotted for as a function of SNR, with a number of observed samples $T = 10^4$. Fig. 5.4 to 5.6 show the MSEs in (5.2), (5.3) and (5.4) plotted as a function the number of observed samples T, with the SNR set to 10dB.

Clearly, the theoretical performance properties in (5.2), (5.3) an (5.4) are verified in our simulations. From Fig. 5.1 and Fig. 5.4, we find that the estimator with optimal weights greatly outperforms the estimator with equal weights. Fig. 5.2 and Fig. 5.5 show that the performance of the estimators with optimal weights will be improved when the set of kernel matrices utilized is enlarged. Fig. 5.3 and Fig. 5.6 indicate that the mean square error of the estimator using all the kernel matrices in set S^4 is much smaller than one quarter of the average mean square error of the estimator with single kernel matrix over the set S^4.

Generally, we find that there is a very good match between all the experimental results and the theoretical results derived in Section 3.4, 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 sample T.

Our results thus support the performance analysis in the general model derived in Section 3.4.

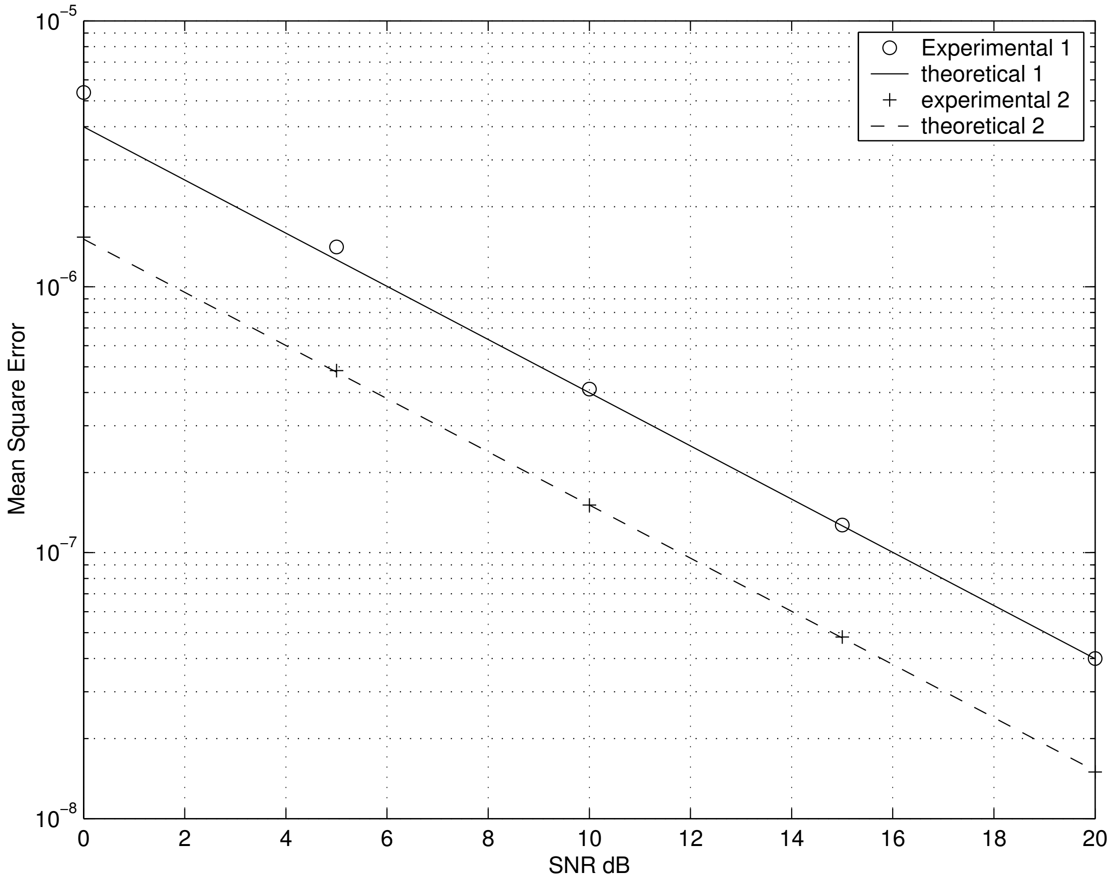

Fig. 5.1 Comparison in (5.2): 1: $\mathrm{MSE}(\mathbf{I}_{4(L-N)}, S^4)$, 2: $\mathrm{MSE}^o(S^4)$

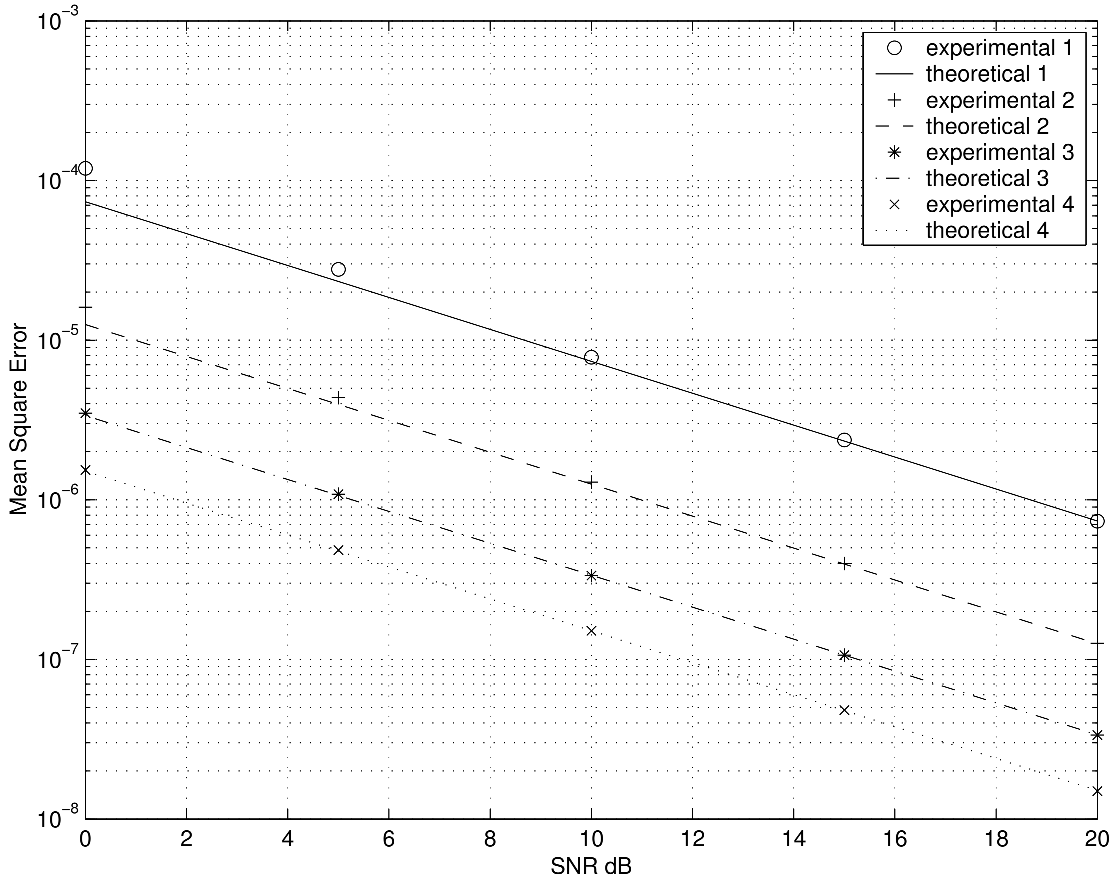

Fig. 5.2 Comparison in (5.3): 1: $\mathrm{MSE}^o(S^1)$, 2: $\mathrm{MSE}^o(S^2)$, 3: $\mathrm{MSE}^o(S^3)$, 4: $\mathrm{MSE}^o(S^4)$,

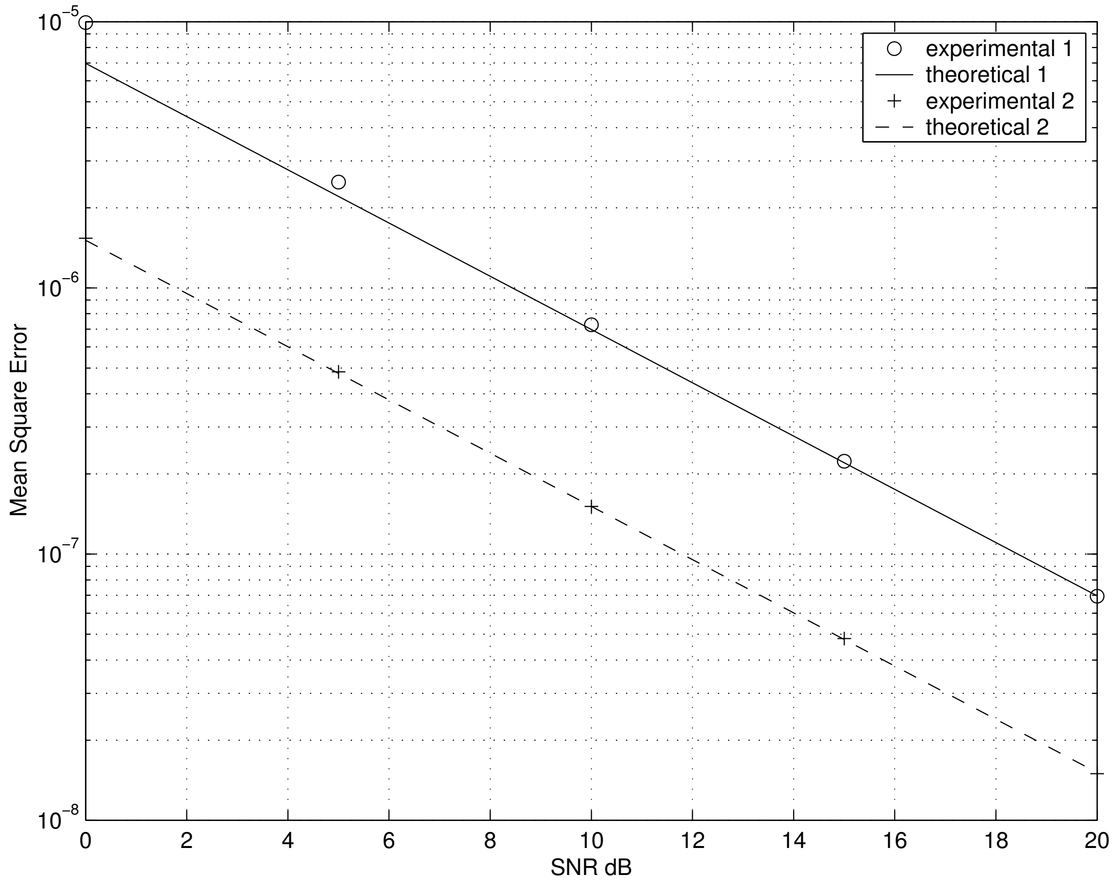

Fig. 5.3 Comparison in (5.4): 1: $\frac{1}{4^2}\sum_{i=1}^{4}\mathrm{MSE}^o(\{\mathbf{C}^i\})$, 2: $\mathrm{MSE}^o(S^4)$

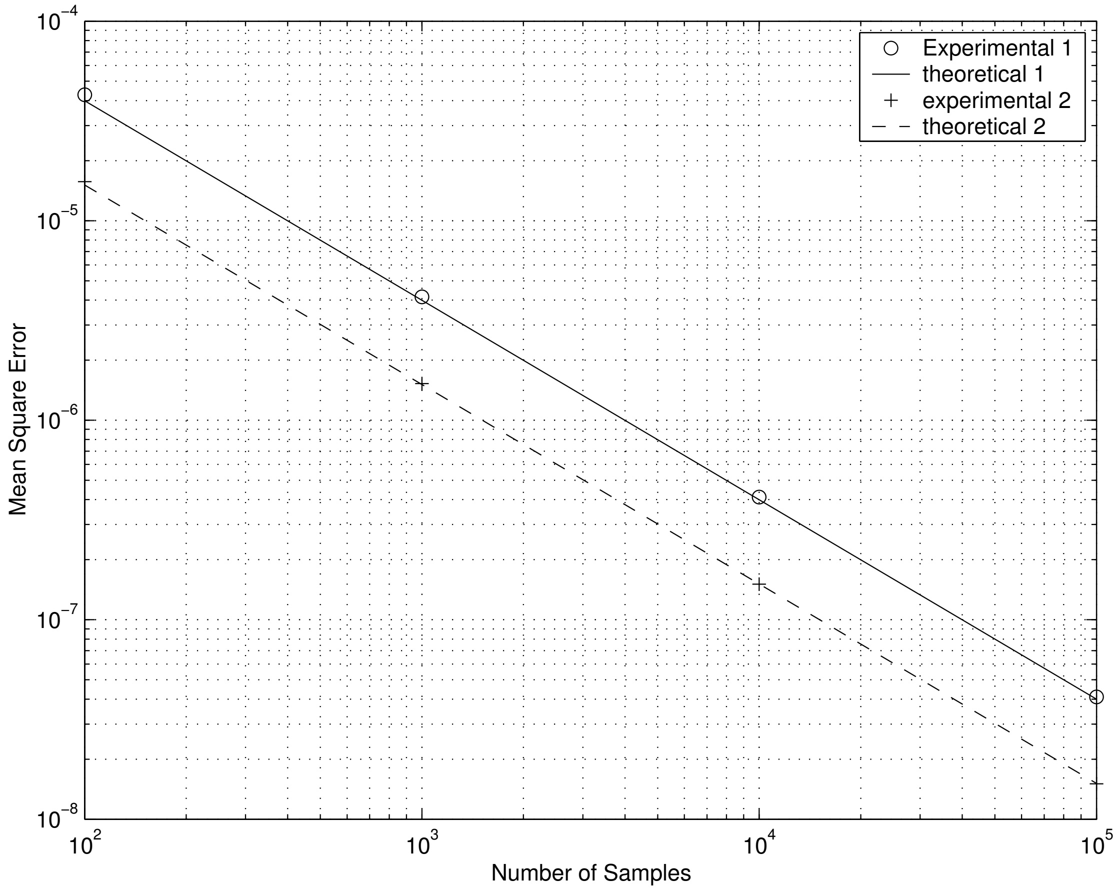

Fig. 5.4 Comparison in (5.2): 1: $\mathrm{MSE}(\mathbf{I}_{4(L-N)}, S^4)$, 2: $\mathrm{MSE}^o(S^4)$

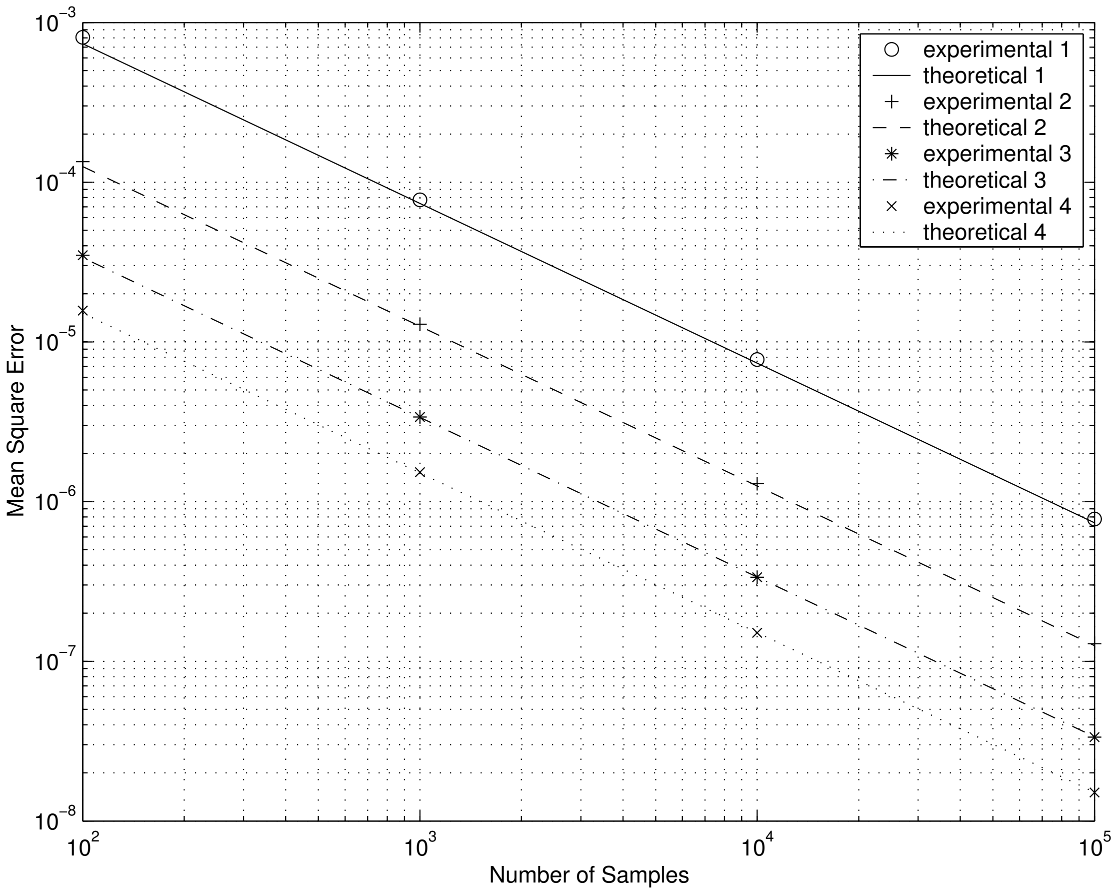

Fig. 5.5 Comparison in (5.3): 1: $\mathrm{MSE}^o(S^1)$, 2: $\mathrm{MSE}^o(S^2)$, 3: $\mathrm{MSE}^o(S^3)$, 4: $\mathrm{MSE}^o(S^4)$,

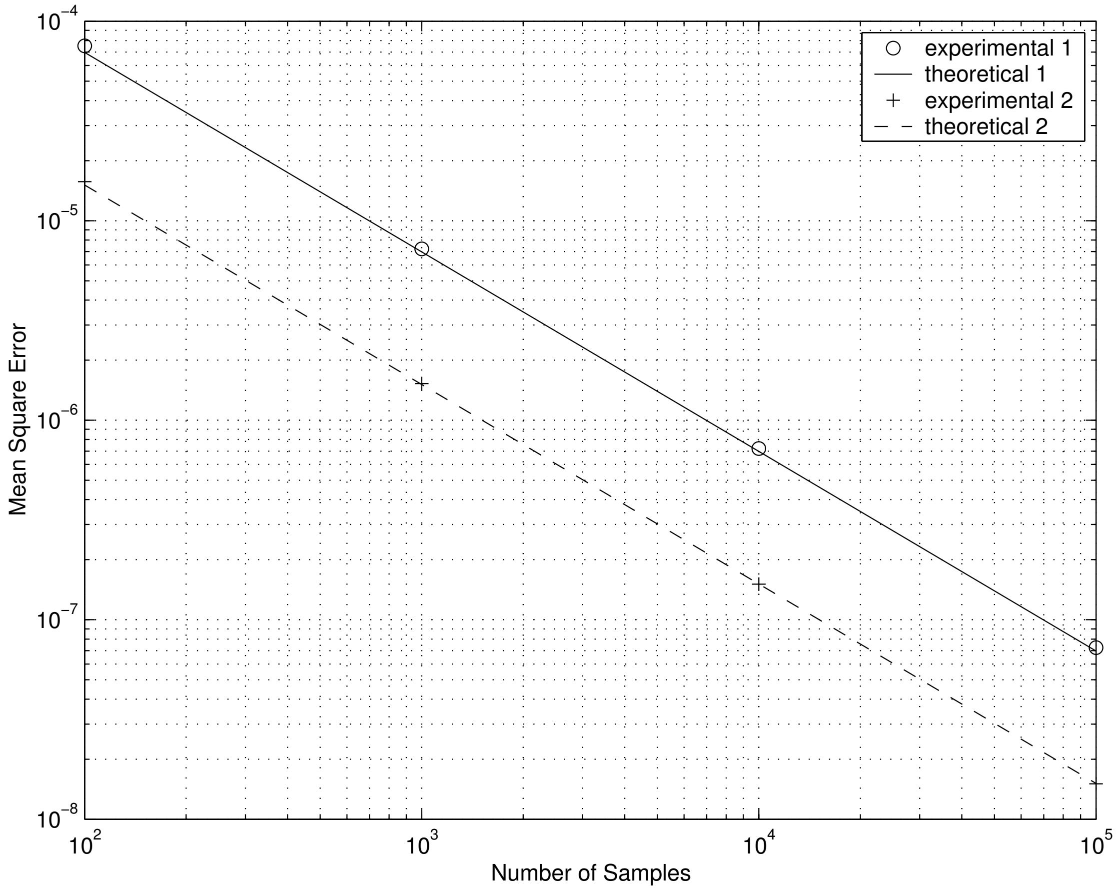

Fig. 5.6 Comparison in (5.4): 1: $\frac{1}{4^2}\sum_{i=1}^{4} \mathrm{MSE}^o(\{\mathbf{C}^i\})$, 2: $\mathrm{MSE}^o(S^4)$

5.3 Simulations for Adaptive Algorithm

In this section, we show the simulation results related to the performance evaluation of the proposed adaptive algorithm (Table 4.2) and the hybrid algorithm (Table 4.1) in time-varying channels. We use the same parameter settings as those in Section 5.2 except for the following differences: The channel is assumed to be a first-order AR model, i.e.

$$\mathbf{h}(t) = \beta\mathbf{h}(t-1) + (1-\beta)\mathbf{f}(t) \tag{5.5}$$

where $\mathbf{f}(t)$ is an i.i.d complex white Gaussian source and parameter β is used to control the rate of change of the radio channel. The plane rotations based EVD tracker used in the simulations is PROTEUS-2 [5], with forgetting factor α. We use an average of the square errors over $T = 10^5$ iterations to approximate the mean square error of the estimators.

The computational complexities of the hybrid algorithm (Table 4.1) and the adaptive algorithm (Table 4.2) in our simulations are calculated here with the parameter setting: $L = 9$, $M = 4$, $N = 6$, $P = 4$. In this case, the computational complexity of the hybrid algorithm is $880 + O(384)$ while that of the adaptive algorithm is $392 + O(36)$.

Five simulation experiments were conducted to test the performance of the two algorithms in the case of different Signal-to-Noise Ratio (SNR), different number N of users, different number P of kernel matrices utilized in the estimator, different forgetting factor α, and different rate of channel variation β. The corresponding results are presented in Fig. 5.7 to 5.11, respectively. The simulations show that the performance of both hybrid and adaptive estimator are effected by the above parameters. Generally, a better performance will be obtained in the case of a larger

SNR, a smaller number of users N, a larger number of kernel matrices P, a properly selected forgetting factor α and/or a smaller rate of channel variation β. In all the above cases, the performance of the fully adaptive algorithm is comparable or superior to that of the hybrid scheme, while the computational complexity of the adaptive algorithm is an order of magnitude lower than that of the hybrid algorithm. Especially we note, from Fig. 5.8, 5.9 and 5.10, that in the case of large N, small P and/or large α, the proposed adaptive algorithm performs better than the hybrid algorithm.

Fig. 5.12 (top) compares the time evolution of the square error produced by both algorithms during a single run in the case $SNR = 10dB$, $N = 6$, $P = 1$ and $\alpha = \beta = 0.005$; Fig. 5.12 (bottom) shows the time evolution of the first two dominant eigenvalues of $D(t)$ during the same experiment. In general, we find that the proposed adaptive algorithm is more robust than the hybrid algorithm in case when the second eigenvalue approaches 1. In this case, the identifiability condition is no longer valid and then the batch estimation in the hybrid algorithm (i.e. step 3 to 5 in Table 4.1) can not provide an accurate estimate. We note that this kind of errors usually last only for a short time duration. However, the power method in the adaptive algorithm, which tracks the channel vector adaptively, limits the speed of the divergence from the the "true" value of the channel vector, thus making the algorithm robust to this kind of short time errors.

We also investigate the performance of the estimators in the case when there is a user entering/exiting the system, which corresponds to the increase/decrease in the rank of the signal subspace. The time evolution of the various quantities of interest are showed in Fig. 5.13. We find that the adaptive algorithm is robust to the rank change and after a short transient period, the channel estimate re-converges to the desired channel vector.

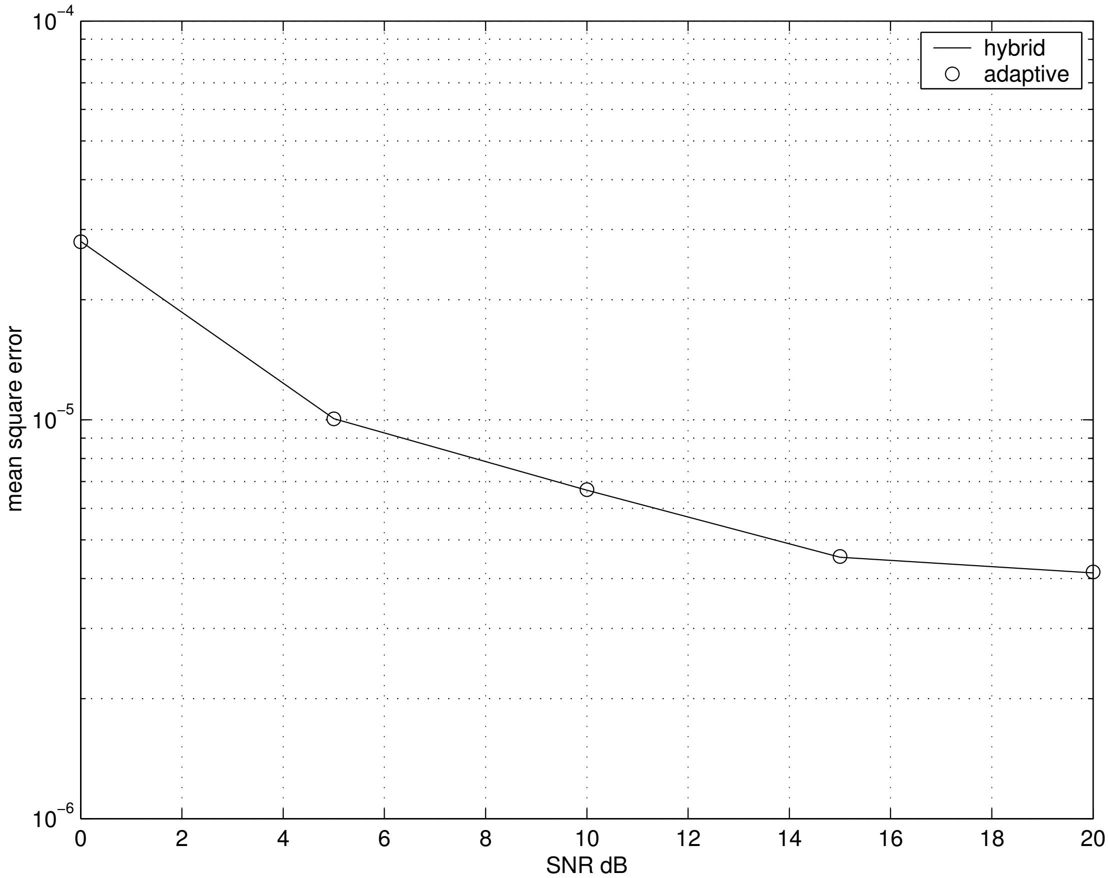

Fig. 5.7 MSE vs. SNR, $N = 6$, $P = 4$, $\alpha = 10^{-3}$, $\beta = 10^{-4}$

5.4 Chapter Summary

In this chapter, we showed the results of computer simulations in a down-link CDMA system operating in frequency selective fading channel with negligible ISI. The simulations in the stationary environments confirmed our performance analysis in the Section 3.4. The simulations in the dynamic signal environment showed that the proposed adaptive algorithm in Section 4.2 keeps a comparable performance to the

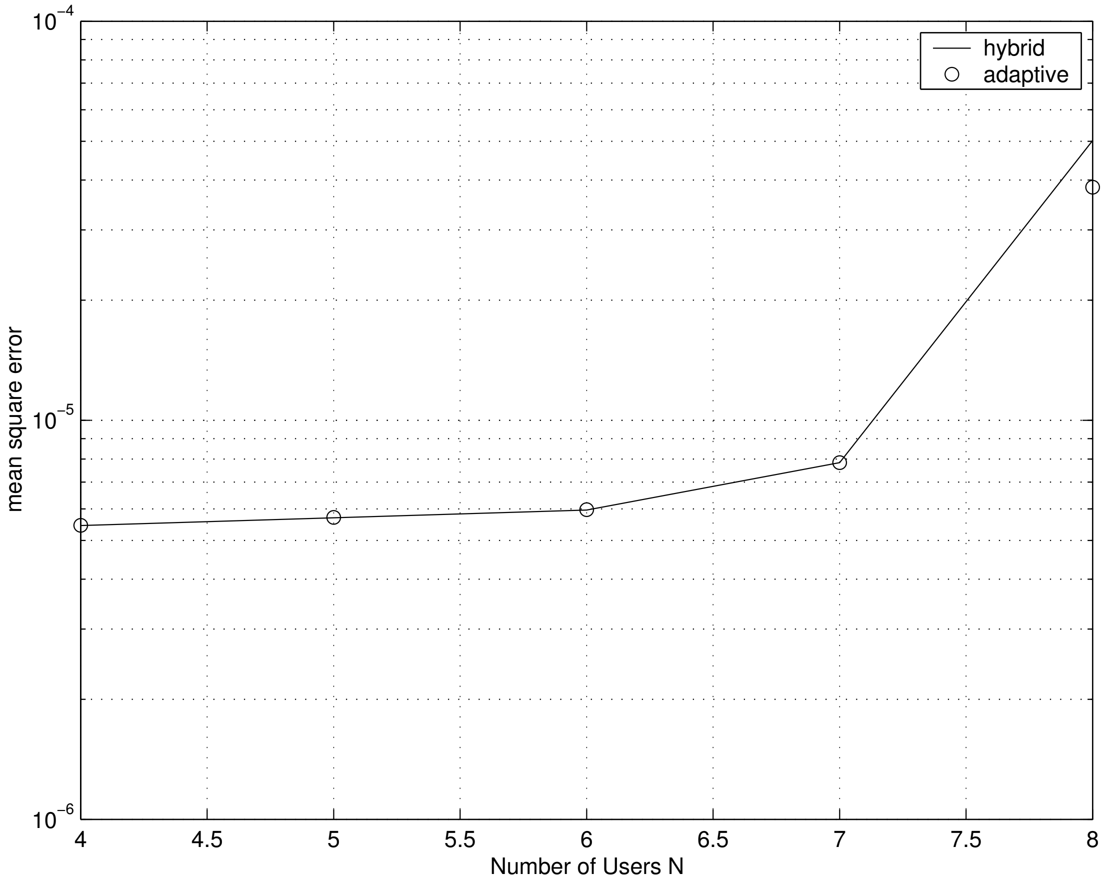

Fig. 5.8 MSE vs. N, SNR$= 10dB$, $P = 4$, $\alpha = 10^{-3}$, $\beta = 10^{-4}$

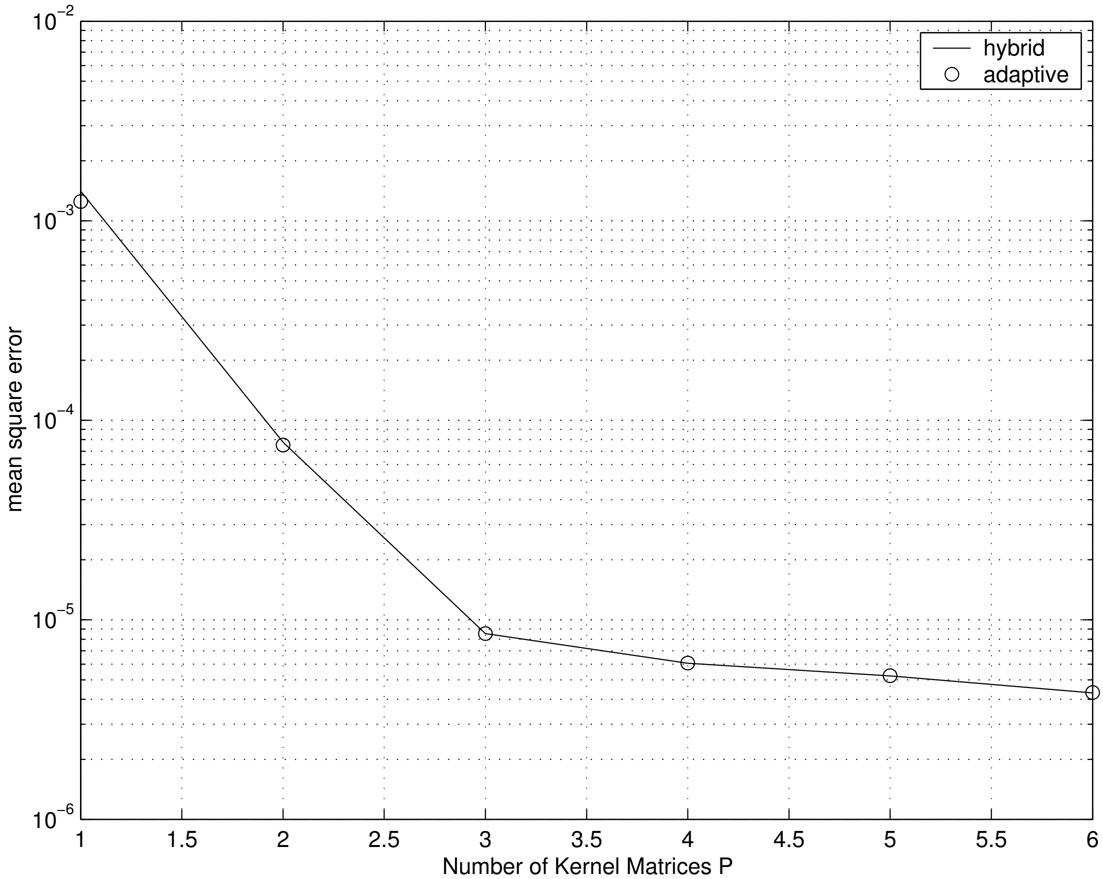

Fig. 5.9 MSE vs. P, SNR$= 10dB$, $N = 6$, $\alpha = 10^{-3}$, $\beta = 10^{-4}$

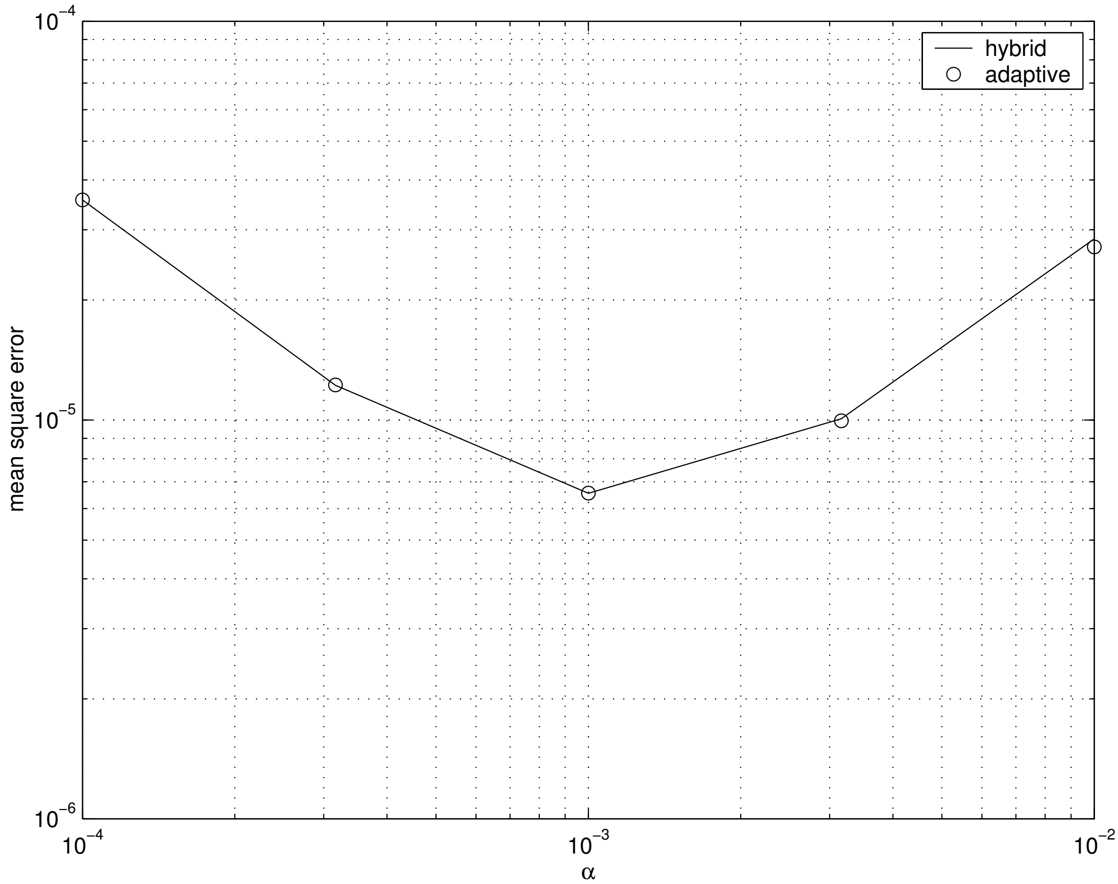

Fig. 5.10 MSE vs. α, SNR$= 10dB$, $N = 6$, $P = 4$, $\beta = 10^{-4}$

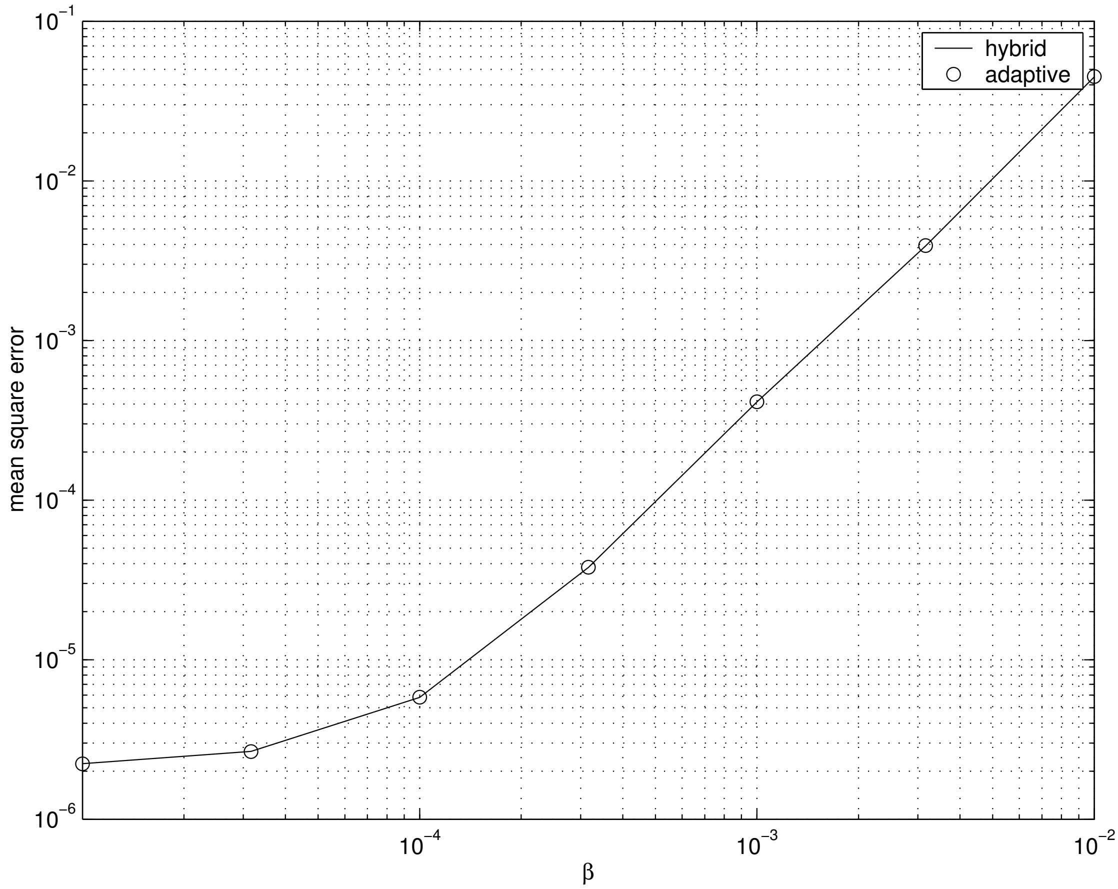

Fig. 5.11 MSE vs. β, SNR= $10dB$, $N=6$, $P=4$, $\alpha=10^{-3}$

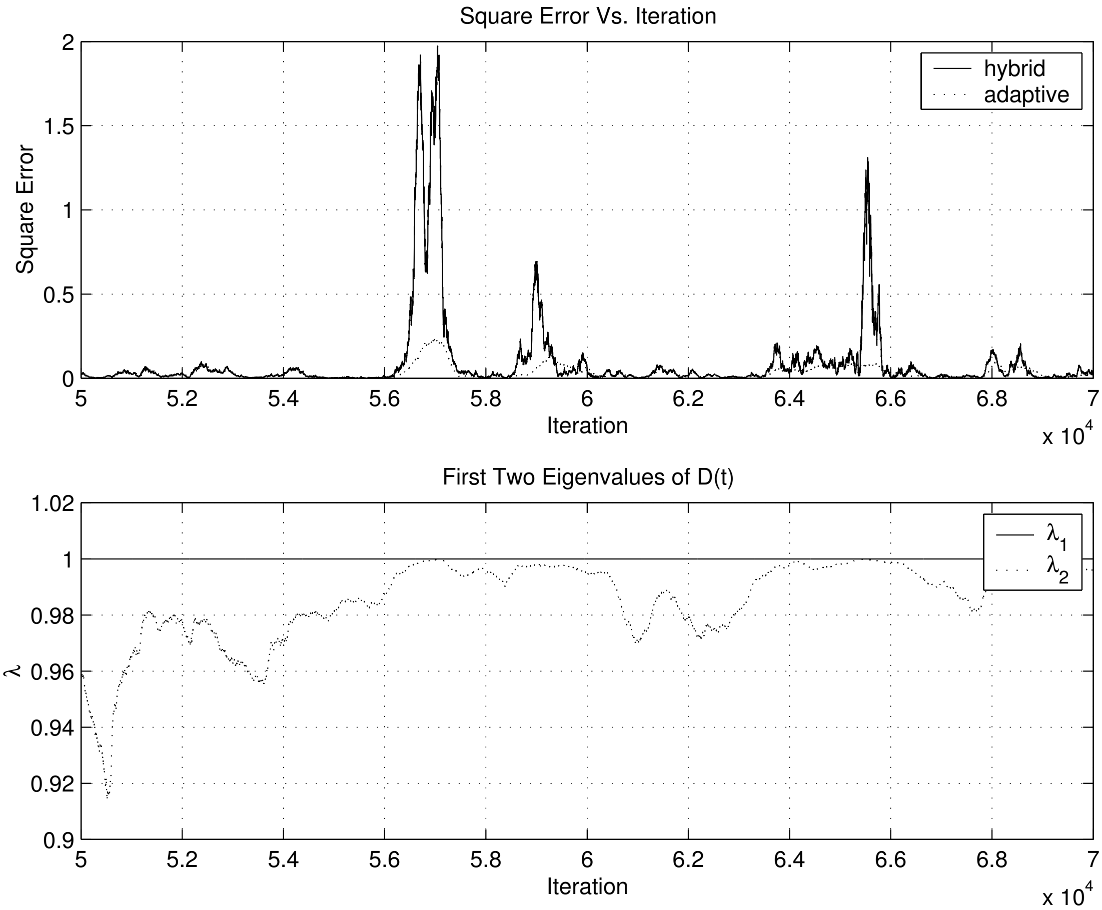

Fig. 5.12 Performance Comparison, SNR$= 10dB$, $N = 6$, $P = 1$, $\alpha = \beta = 0.005$

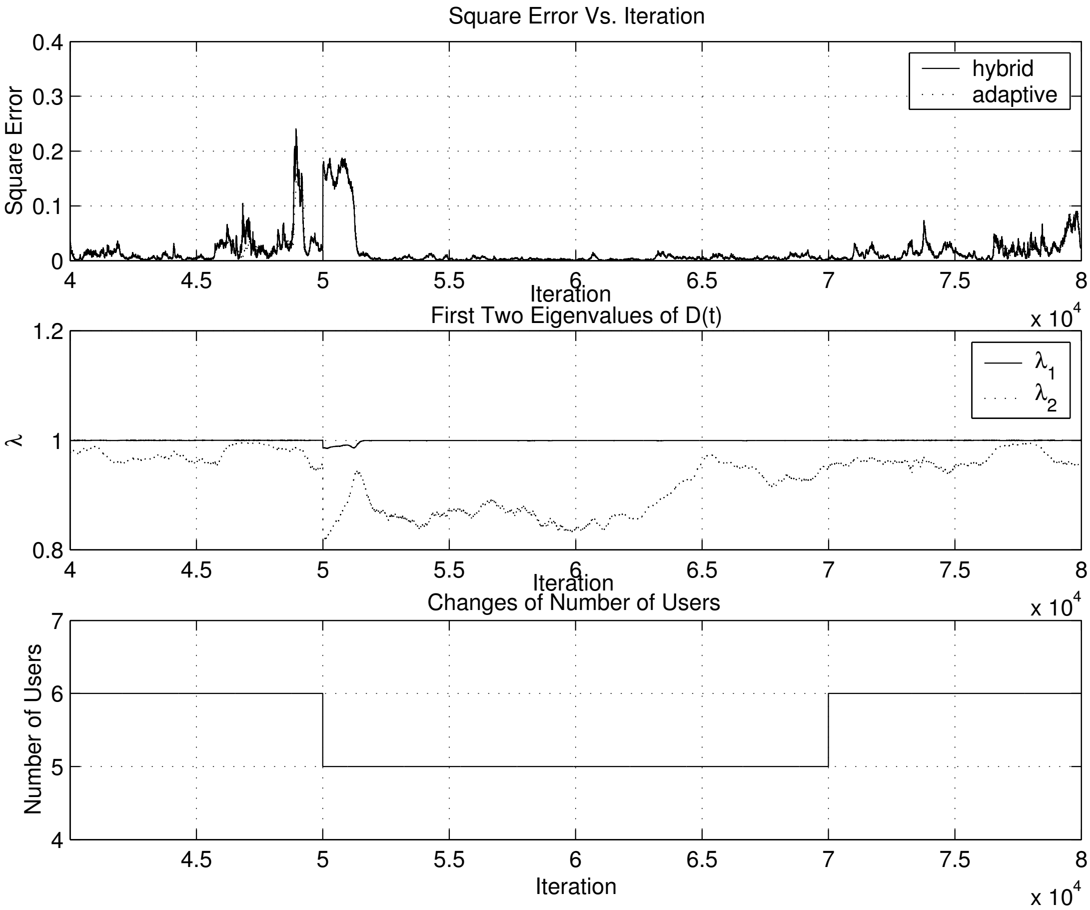

Fig. 5.13 Time Evolution with Rank Change, SNR= $10dB$, $N = 6$, $P = 1$, $\alpha = \beta = 0.005$

hybrid algorithm while greatly reducing the computational complexity.

Chapter 6

Conclusion

In this thesis, we presented a systematic study of the subspace-based blind channel estimation method.

We first discussed the signal model of different applications of subspace channel estimation methods. We began by reviewing the signal models of three specific communication systems. By incorporating the common features of the different signal models, we then proposed a general signal model, which is applicable to a multitude of communication systems. Within the framework of the proposed general model, we formulated the problem of blind subspace channel estimation.

Based on the proposed general signal model, we proposed the generalized blind subspace channel estimation algorithm. We formulated the theoretical foundation of the subspace channel estimation algorithm based on the orthogonality property between the signal and noise subspace. Through the study of the identifiability, we showed that by enlarging the set of the kernel matrices, we may enable the estimator to identify a longer channel and/or to be used in a system having a larger number of independent signals. With the consideration of robustness to the error in the data correlation matrix, we proposed a generalized blind subspace channel estimator by minimizing a novel cost function, which incorporates a set of kernel matrices via a weighted projection error.

We investigated the asymptotic performance of the proposed estimator when the number of independent observations is large. We derive its bias, covariance and mean square error (MSE), as well as the associated Cramer-Rao bound. We showed 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.

We also studied the adaptive implementation of the proposed generalized subspace blind channel estimation algorithm. We firstly reviewed the spherical subspace tracking methods, which can update the desired subspace information with low complexity. Then we proposed a novel low-complexity subspace implementation of the subspace channel estimator by utilizing the special updating feature of plane rotation based spherical subspace trackers. The complexity of the resulting adaptive estimator is an order of magnitude lower than the hybrid estimator, that is the adaptive subspace tracker followed by batch estimation at each time step.

Finally, we showed the results of computer simulations in a down-link CDMA system operating in frequency selective fading channel with negligible ISI. The simulations in the stationary environments confirmed our performance analysis for the batch estimator in the Section 3.4. The simulations in the dynamic signal environment showed that the performance of the proposed adaptive algorithm in Section 4.2 was comparable to that of the hybrid algorithm while greatly reducing the computational complexity.

In this thesis, there are still several problems, which remain open for future research. These are described briefly below.

First, according to the identifiability condition discussed in Section 3.2.1, the dimension of $\mathrm{Span}[\mathcal{C}] \cap \mathrm{Span}[\mathcal{U}_s]$ must be 1. However, the subspace $\mathrm{Span}[\mathcal{U}_s] = \mathrm{Span}[\mathbf{I}_P \otimes \mathbf{W}]$ depends on the realization of the random channel vector $\underline{\mathbf{h}}$, which is unknown to the transmitter. Therefore this identifiability condition can not provide a useful, definite constraint for the configuration of the transmitter. To overcome this problem, we would need to find an alternative identifiability condition, where the randomness of the vector $\underline{\mathbf{h}}$ is incorporated into the development. That is, we should find the constraints on the parameters controllable at the transmitter, e.g. number of independent symbols, structure of kernel matrices, etc., which are sufficient and necessary to ensure that the dimension of $\mathrm{Span}[\mathcal{U}_s] = \mathrm{Span}[\mathbf{I}_P \otimes \mathbf{W}]$ is one with probability one.

Second, as explained in Section 3.4, we have not been able to extend Theorem 2 to 4 to the case of the covariance of the proposed estimator in the form of matrix inequalities. A possible way to prove the validity of the extension of Theorem 2 is by checking the constrained CRB with partial knowledge of kernel matrices. That is, we assume that only the kernel matrices in the set S are available and those in the set $U - S$ are unknown. In this case, we can derive the constrained CRB of the joint estimation of the channel vector $\underline{\mathbf{h}}$, transmitted symbol $\underline{\mathbf{b}}$ and the kernel matrices in the set $U - S$. If the resulting constrained CRB can be achieved by the proposed estimator with some parameter $\mathcal{A}$, we view this parameter as the optimal one minimizing the covariance of the estimator.

Appendix A

Proofs of Theorems

A.1 Proof of Theorem 1

For convenience of further derivation, we define the following variables:

$$\mathbf{X} \triangleq [\mathbf{x}_1, \ldots, \mathbf{x}_T] \tag{A.1}$$

$$\mathbf{E} \triangleq [\mathbf{e}_1, \ldots, \mathbf{e}_T] \tag{A.2}$$

$$\bar{\mathbf{E}} \triangleq \mathbf{U}_n^H \mathbf{E} = [\bar{\mathbf{e}}_1, \cdots, \bar{\mathbf{e}}_T] \tag{A.3}$$

$$\mathcal{T} \triangleq \mathcal{C}^H \mathcal{U}_n \mathcal{A} \tag{A.4}$$

$$\mathcal{X} \triangleq \mathbf{I}_P \otimes \mathbf{X}^\dagger \tag{A.5}$$

$$\mathcal{E} \triangleq \mathbf{I}_P \otimes \bar{\mathbf{E}} = [\mathcal{E}_1, \ldots, \mathcal{E}_P] \tag{A.6}$$

$$\mathcal{E}_i \triangleq \iota_i \otimes \bar{\mathbf{E}} \quad i = 1, \ldots, P \tag{A.7}$$

$$\iota_i \triangleq \text{the } i\text{-th column of the identity matrix} \tag{A.8}$$

$$\mathbf{z}^{m,i} \triangleq \mathbf{X}^\dagger \mathbf{w}^{m,i} = \mathbf{X}^\dagger \mathbf{C}^{m,i} \mathbf{h} = [z_1^{m,i}, \ldots, z_T^{m,i}]^T \tag{A.9}$$

$$\mathbf{z} \triangleq \mathrm{vec}[\mathbf{z}^{m,1}, \ldots, \mathbf{z}^{m,P}] = \mathcal{X} \mathrm{vec}[\bar{\mathbf{W}}] \tag{A.10}$$

$$\mathbf{Z} \triangleq \mathbf{X}^\dagger \mathbf{W} \tag{A.11}$$

where $\mathbf{z}^{m,i}$, for $i = 1, \ldots, K^m$ and $m = 1, \ldots, J$, comprises the column of $\mathbf{Z}$.

According to [19], the first order perturbation of $\mathbf{U}_n$ is

$$\Delta \mathbf{U}_n = -(\mathbf{X}^\dagger)^H \mathbf{E}^H \mathbf{U}_n = -(\mathbf{X}^\dagger)^H \bar{\mathbf{E}}^H \tag{A.12}$$

Consequently the perturbations of $\mathcal{U}_n$ and $\mathcal{T}$ are:

$$\begin{aligned} \Delta \mathcal{U}_n &= \mathbf{I}_P \otimes \Delta \mathbf{U}_n = \mathbf{I}_P \otimes [-(\mathbf{X}^\dagger)^H \bar{\mathbf{E}}^H] \\ &= -[\mathbf{I}_P \otimes (\mathbf{X}^\dagger)^H][\mathbf{I}_P \otimes \bar{\mathbf{E}}^H] = -\mathcal{X}^H \mathcal{E}^H \quad \text{(A.13)} \\ \Delta \mathcal{T} &= \mathcal{C}^H \Delta \mathcal{U}_n \mathcal{A} = -(\mathcal{A}^H \mathcal{E} \mathcal{X} \mathcal{C})^H \quad \text{(A.14)} \end{aligned}$$

According to [19], the estimation error in the target channel $\hat{\mathbf{h}}$ is

$$\begin{aligned} \Delta \mathbf{h} &= -(\mathcal{T}^\dagger)^H (\Delta \mathcal{T})^H \mathbf{h} = (\mathcal{T}^\dagger)^H \mathcal{A}^H \mathcal{E} \mathcal{X} \mathcal{C} \mathbf{h} \\ &= (\mathcal{T}^\dagger)^H \mathcal{A}^H \mathcal{E} \mathcal{X} \mathrm{vec}[\bar{\mathbf{W}}] = (\mathcal{T}^\dagger)^H \mathcal{A}^H \mathcal{E} \mathbf{z} \end{aligned} \tag{A.15}$$

It is known that $E[\mathcal{E}] = \mathbf{0}$. Then the bias of the proposed estimator is

$$\begin{aligned} \mathrm{Bias} &= E[\Delta \mathbf{h}] = E[(\mathcal{T}^\dagger)^H \mathcal{A} \mathcal{E} \mathbf{z}] \\ &= (\mathcal{T}^\dagger)^H \mathcal{A} E[\mathcal{E}] \mathbf{z} = \mathbf{0} \end{aligned} \tag{A.16}$$

Therefore, we conclude that the proposed estimator $\hat{\mathbf{h}}$ is unbiased. Accordingly, the covariance of $\hat{\mathbf{h}}$ can be expressed as

$$\mathrm{Cov} = E[\Delta \mathbf{h} \Delta \mathbf{h}^H] = (\mathcal{T}^\dagger)^H \mathcal{A}^H E[\mathcal{E} \mathbf{z} \mathbf{z}^H \mathcal{E}^H] \mathcal{A} \mathcal{T}^\dagger \tag{A.17}$$

According to [18], $E[\mathcal{E}\mathbf{z}\mathbf{z}^H\mathcal{E}^H]$ can be expressed as

$$E[\mathcal{E}\mathbf{z}\mathbf{z}^H\mathcal{E}^H] = E[\sum_{i=1}^{P}\sum_{p=1}^{T} z_p^{m,i}(\iota_i \otimes \bar{\mathbf{e}}_p)\sum_{j=1}^{P}\sum_{q=1}^{T}(z_q^{m,j})^*(\iota_j \otimes \bar{\mathbf{e}}_q)^H] \tag{A.18}$$

To calculate $E[\mathcal{E}\mathbf{z}\mathbf{z}^H\mathcal{E}^H]$, we need to study $\mathbf{z}$ and $\mathcal{E}$ respectively.

Firstly, according to the result in [17] and the asymptotic property in (3.27), we have

$$\mathbf{Z}^H\mathbf{Z} = \frac{1}{T}\mathbf{\Gamma}^{-2} \tag{A.19}$$

Thus

$$(\mathbf{z}^{m,i})^H\mathbf{z}^{m,j} = \frac{1}{\gamma^{m,i}\gamma^{m,j}T}\delta_{i,j} \tag{A.20}$$

where the Kronecker delta function $\delta_{i,j}$ is defined as $\delta_{i,j} = 1$ for $i = j$ and otherwise equal to 0.

Secondly, since the i-th column of $\mathcal{E}_p$ is $\iota_i \otimes \bar{\mathbf{e}}_p$, we derive the following results based on the definition in (A.6) and (A.8)

$$E[(\iota_i \otimes \bar{\mathbf{e}}_p)(\iota_j \otimes \bar{\mathbf{e}}_q)^H] = \sigma^2(\iota_{i,j} \otimes I_{(L-K)})\delta_{p,q} \tag{A.21}$$

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

Based on the results in (A.20) and (A.21), we simplify (A.18) as

$$\begin{aligned} E[\mathcal{E}\mathbf{z}\mathbf{z}^H\mathcal{E}^H] &= \sum_{i=1}^{P}\sum_{j=1}^{P}\sum_{p=1}^{T}\sum_{q=1}^{T}[z_p^{m,i}(z_q^{m,j})^*]E[(\iota_i \otimes \bar{\mathbf{e}}_p)(\iota_j \otimes \bar{\mathbf{e}}_q)^H] \\ &= \sum_{i=1}^{P}\sum_{j=1}^{P}\sum_{p=1}^{T}\sum_{q=1}^{T}[z_p^{m,i}(z_q^{n,j})^*]\sigma^2(\iota_{i,j} \otimes I_{(L_c-K)})\delta_{p,q} \\ &= \sum_{i=1}^{P}\sum_{j=1}^{P}\sigma^2(\iota_{i,j} \otimes I_{(L_c-K)})\sum_{p=1}^{T} z_p^{m,i}(z_p^{m,j})^* \\ &= \sum_{i=1}^{P}\sum_{j=1}^{P}\sigma^2(\iota_{i,j} \otimes I_{(L_c-K)})(\mathbf{z}^{m,j})^H\mathbf{z}^{m,i} \\ &= \sum_{i=1}^{P}\sum_{j=1}^{P}\frac{\sigma^2}{\gamma^{m,i}\gamma^{m,j}T}(\iota_{i,j} \otimes I_{(L_c-K)})\delta_{i,j} \\ &= \frac{\sigma^2}{T}\sum_{i=1}^{P}\frac{1}{(\gamma^{m,i})^2}(\iota_{i,i} \otimes I_{(L-K)}) \\ &= \frac{\sigma^2}{(\gamma^m)^2T}\boldsymbol{\Upsilon} \end{aligned} \tag{A.22}$$

By substituting (A.22) into (A.17), we have

$$\begin{aligned} \mathrm{Cov} &= \frac{\sigma^2}{T}(\mathcal{T}^\dagger)^H\mathcal{A}^H\boldsymbol{\Upsilon}^{-2}\mathcal{A}\mathcal{T}^\dagger \\ &= \frac{\sigma^2}{T}[(\mathcal{C}^H\mathcal{U}_n\mathcal{A})^\dagger]^H\mathcal{A}^H\boldsymbol{\Upsilon}^{-2}\mathcal{A}(\mathcal{C}^H\mathcal{U}_n\mathcal{A})^\dagger \end{aligned} \tag{A.23}$$

Finally,

$$\begin{aligned} \mathrm{MSE} &= E[\Delta\mathbf{h}^H\Delta\mathbf{h}] = Tr[\mathrm{Cov}] \\ &= \frac{\sigma^2}{T}Tr[(\boldsymbol{\Upsilon}^{-1}\mathcal{A})(\mathcal{A}^H\mathcal{U}_n^H\mathcal{C}\mathcal{C}^H\mathcal{U}_n\mathcal{A})^\dagger(\boldsymbol{\Upsilon}^{-1}\mathcal{A})^H] \end{aligned} \tag{A.24}$$

A.2 Proof of Theorem 2

Lemma 1 *[3] If* $\mathbf{A} \in \mathbb{C}^{m\times n}$ *and* $\mathbf{B} \in \mathbb{C}^{n\times p}$*, then*

$$(\mathbf{AB})^{\dagger} = (\mathbf{P}_{R(\mathbf{A}^H)}\mathbf{B})^{\dagger}(\mathbf{AP}_{R(\mathbf{B})})^{\dagger} \tag{A.25}$$

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* $R(\mathbf{A})$*.*

According to the definition of $\bar{\mathcal{A}} \triangleq \mathbf{\Upsilon}^{-1}\mathcal{A}$, we may express the mean square error in the following form

$$\begin{aligned}\mathrm{MSE}(\mathcal{A}, S) &= \frac{\sigma^2}{T}Tr[(\mathbf{\Upsilon}^{-1}\mathcal{A})(\mathcal{A}^H\mathcal{U}_n^H\mathcal{C}\mathcal{C}^H\mathcal{U}_n\mathcal{A})^{\dagger}(\mathbf{\Upsilon}^{-1}\mathcal{A})^H] \\ &= \frac{\sigma^2}{T}Tr[\bar{\mathcal{A}}(\bar{\mathcal{A}}^H\mathbf{Q}\bar{\mathcal{A}})^{\dagger}\bar{\mathcal{A}}^H]\end{aligned} \tag{A.26}$$

According to Lemma 1, the term $(\bar{\mathcal{A}}^H\mathbf{Q}\bar{\mathcal{A}})^{\dagger}$ can be expressed as

$$\begin{aligned}(\bar{\mathcal{A}}^H\mathbf{Q}\bar{\mathcal{A}})^{\dagger} &= [\mathbf{P}_{R[(\bar{\mathcal{A}}^H\mathbf{Q})^H]}\bar{\mathcal{A}}]^{\dagger}[\bar{\mathcal{A}}^H\mathbf{QP}_{R(\bar{\mathcal{A}})}]^{\dagger} \\ &= [\mathbf{P}_{R[(\bar{\mathcal{A}}^H\mathbf{Q})^H]}\bar{\mathcal{A}}]^{\dagger}[\mathbf{P}_{R(\bar{\mathcal{A}})}\mathbf{QP}_{R(\bar{\mathcal{A}})}]^{\dagger} \\ &\quad [\bar{\mathcal{A}}^H\mathbf{P}_{R[\mathbf{QP}_{R(\bar{\mathcal{A}})}]}]^{\dagger}\end{aligned} \tag{A.27}$$

Here $\bar{\mathcal{A}}$ is full-rank and referring to (3.16), $\mathbf{Q}$ is Hermitian and semi-positive definite.

Thus

$$\mathbf{P}_{R(\bar{\mathcal{A}})} = \mathbf{I}_{P(L_c-K)} \tag{A.28}$$

$$\mathbf{P}_{R[\mathbf{Q}\mathbf{P}_{R(\bar{\mathcal{A}})}]} = \mathbf{P}_{R(\mathbf{Q})} \tag{A.29}$$

$$\mathbf{P}_{R[(\bar{\mathcal{A}}^H\mathbf{Q})^H]} = \mathbf{P}_{R(\mathbf{Q}^H)} \tag{A.30}$$

Define

$$\mathbf{P} \triangleq \mathbf{P}_{R(\mathbf{Q}^H)} = \mathbf{P}_{R(\mathbf{Q})} \tag{A.31}$$

$$\mathbf{P}^{\perp} \triangleq \mathbf{I}_{P(L_c-K)} - \mathbf{P} \tag{A.32}$$

Then

$$(\bar{\mathcal{A}}^H\mathbf{Q}\bar{\mathcal{A}})^{\dagger} = [\mathbf{P}\bar{\mathcal{A}}]^{\dagger}\mathbf{Q}^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}]^{\dagger} \tag{A.33}$$

Consequently

$$\begin{aligned} Tr[\bar{\mathcal{A}}(\bar{\mathcal{A}}^H\mathbf{Q}\bar{\mathcal{A}})^{\dagger}\bar{\mathcal{A}}^H] &= Tr[\bar{\mathcal{A}}[\mathbf{P}\bar{\mathcal{A}}]^{\dagger}\mathbf{Q}^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}]^{\dagger}\bar{\mathcal{A}}^H] \\ &= Tr[(\mathbf{P}+\mathbf{P}^{\perp})\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger}\mathbf{Q}^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}]^{\dagger}[\bar{\mathcal{A}}^H(\mathbf{P}+\mathbf{P}^{\perp})] \\ &= Tr[\mathbf{P}\mathcal{A}][\mathbf{P}\mathcal{A}]^{\dagger}\mathbf{Q}^{\dagger}[\mathcal{A}^H\mathbf{P}]^{\dagger}[\mathcal{A}^H\mathbf{P}] \\ &\quad +Tr[\mathbf{P}^{\perp}\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger}[\mathbf{Q}]^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}]^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}] \\ &\quad +Tr[\mathbf{P}\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger}[\mathbf{Q}]^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}]^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}^{\perp}] \\ &\quad +Tr[\mathbf{P}^{\perp}\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger}\mathbf{Q}^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}]^{\dagger}[\bar{\mathcal{A}}^H\mathbf{P}^{\perp}] \end{aligned} \tag{A.34}$$

We study on the trace of the four terms in (A.34), separately. In the first term

$$[\mathbf{P}\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger} = \mathbf{P}_{R(\mathbf{P}\bar{\mathcal{A}})} = \mathbf{P} \tag{A.35}$$

Then

$$Tr[[\mathbf{P}\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger}\mathbf{Q}^{\dagger}[\bar{\mathcal{A}}^{H}\mathbf{P}]^{\dagger}[\bar{\mathcal{A}}^{H}\mathbf{P}]] = Tr[\mathbf{P}\mathbf{Q}^{\dagger}\mathbf{P}] = Tr[\mathbf{Q}^{\dagger}] \tag{A.36}$$

The second term

$$Tr[[\mathbf{P}^{\perp}\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger}[\mathbf{Q}]^{\dagger}[\bar{\mathcal{A}}^{H}\mathbf{P}]^{\dagger}[\bar{\mathcal{A}}^{H}\mathbf{P}]] = Tr[[\bar{\mathcal{A}}^{H}\mathbf{P}\mathbf{P}^{\perp}\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger}[\mathbf{Q}]^{\dagger}[\bar{\mathcal{A}}^{H}\mathbf{P}]^{\dagger}] = 0 \tag{A.37}$$

Similarly, the third term

$$Tr[[\mathbf{P}\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger}[\mathbf{Q}]^{\dagger}[\bar{\mathcal{A}}^{H}\mathbf{P}]^{\dagger}[\bar{\mathcal{A}}^{H}\mathbf{P}^{\perp}]] = 0 \tag{A.38}$$

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

$$Tr[[\mathbf{P}^{\perp}\bar{\mathcal{A}}][\mathbf{P}\bar{\mathcal{A}}]^{\dagger}[\mathbf{Q}]^{\dagger}[\bar{\mathcal{A}}^{H}\mathbf{P}]^{\dagger}[\bar{\mathcal{A}}^{H}\mathbf{P}^{\perp}]] \geq 0 \tag{A.39}$$

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

$$Tr[[\mathbf{P}^{\perp}\mathcal{A}][\mathbf{P}\mathcal{A}]^{\dagger}[\mathbf{Q}]^{\dagger}[\mathcal{A}^{H}\mathbf{P}]^{\dagger}[\mathcal{A}^{H}\mathbf{P}^{\perp}]] = 0 \tag{A.40}$$

Therefore

$$\mathrm{MSE}^{o}(S) = \mathrm{MSE}(c\boldsymbol{\Upsilon}, S) = \frac{\sigma^2}{T}Tr[\mathbf{Q}^{\dagger}] \tag{A.41}$$

According to Lemma 1,

$$\mathbf{Q}^\dagger = (\boldsymbol{\Upsilon}^H \mathcal{U}_n^H \mathcal{C}\mathcal{C}^H \mathcal{U}_n \boldsymbol{\Upsilon})^\dagger = (\mathcal{C}^H \mathcal{U}_n \boldsymbol{\Upsilon})^\dagger (\boldsymbol{\Upsilon}^H \mathcal{U}_n^H \mathcal{C})^\dagger \tag{A.42}$$

$$\mathcal{Q}^\dagger = (\mathcal{C}^H \mathcal{U}_n \boldsymbol{\Upsilon}\boldsymbol{\Upsilon}^H \mathcal{U}_n^H \mathcal{C})^\dagger = (\boldsymbol{\Upsilon}^H \mathcal{U}_n^H \mathcal{C})^\dagger (\mathcal{C}^H \mathcal{U}_n \boldsymbol{\Upsilon})^\dagger \tag{A.43}$$

Therefore

$$Tr[\mathbf{Q}^\dagger] = Tr[\mathcal{Q}^\dagger] \tag{A.44}$$

and

$$\mathrm{MSE}^o(S) = \frac{\sigma^2}{T} Tr[\mathbf{Q}^\dagger] = \frac{\sigma^2}{T} Tr[\mathcal{Q}^\dagger] \tag{A.45}$$

A.3 Proof of Theorem 3

Definition 1 *[21] 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*

$$x_1 \geq \cdots \geq x_n, \quad \textit{and} \quad y_1 \geq \cdots \geq y_n \tag{A.46}$$

we have

$$\sum_{i=1}^{k} x_i \geq \sum_{i=1}^{k} y_i \quad \textit{for} \quad k = 1, \ldots, n-1 \tag{A.47}$$

$$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \tag{A.48}$$

Lemma 2 *[20] $\mathbf{H}$ is an $n \times n$ Hermitian matrix with diagonal elements $h_1, \ldots, h_n$*

and eigenvalues $\lambda_1, \ldots, \lambda_n$*, then*

$$[\lambda_1, \ldots, \lambda_n] \succ [h_1, \ldots, h_n] \tag{A.49}$$

Lemma 3 (Majorization Inequality) *[21]* $[x_1, \ldots, x_n]$ *majorizes* $[y_1, \ldots, y_n]$ *iff for every convex function* f

$$\sum_{i=1}^{n} f(x_i) \geq \sum_{i=1}^{n} f(y_i) \tag{A.50}$$

According to Theorem 2,

$$\begin{aligned} \mathrm{MSE}^o(S) &= \frac{\sigma^2}{(\gamma^m)^2 T} Tr[\mathcal{Q}^\dagger] && \text{(A.51)} \\ \mathrm{MSE}^o(S_q) &= \frac{\sigma^2}{(\gamma^m)^2 T} Tr[\mathcal{Q}_q^\dagger] \quad q = 1, \ldots, Q && \text{(A.52)} \end{aligned}$$

where

$$\begin{aligned} \mathcal{Q} &= \mathcal{C}^H \mathcal{U}_n \mathbf{\Upsilon}^2 \mathcal{U}_n^H \mathcal{C} \\ &= \sum_{\mathbf{C}^{m,i} \in S} (\gamma^{m,i})^2 (\mathbf{C}^{m,i})^H \mathbf{U}\mathbf{U}^H \mathbf{C}^{m,i} && \text{(A.53)} \\ \mathcal{Q}_q &\triangleq \sum_{\mathbf{C}^{m,i} \in S_q} (\gamma^{m,i})^2 (\mathbf{C}^{m,i})^H \mathbf{U}\mathbf{U}^H \mathbf{C}^{m,i} && \text{(A.54)} \end{aligned}$$

Clearly

$$\mathbf{Q} = \sum_{q=1}^{Q} \mathbf{Q}_q \tag{A.55}$$

Apply EVD on $\mathcal{Q}$ and $\mathcal{Q}_q$ respectively

$$\begin{aligned} \mathcal{Q} &= \mathbf{V}\mathbf{\Psi}\mathbf{V}^H && \text{(A.56)} \\ \mathcal{Q}_q &= \mathbf{V}_q \mathbf{\Psi}_q \mathbf{V}_q^H \quad q = 1, \ldots, Q && \text{(A.57)} \end{aligned}$$

where $\mathbf{\Psi} \triangleq \mathrm{diag}[\psi^1, \ldots, \psi^M]$ and $\mathbf{\Psi}_q \triangleq \mathrm{diag}[\psi_q^1, \ldots, \psi_q^M]$ are the eigenvalue matrices of $\mathcal{Q}$ and $\mathcal{Q}_q$, respectively. According to the property of the trace of matrix [12], we have

$$
\begin{aligned}
Tr[\mathcal{Q}^\dagger] &= \sum_{j=1}^{M-1} \frac{1}{\psi^j} && \text{(A.58)} \\
Tr[\mathcal{Q}_q^\dagger] &= \sum_{j=1}^{M-1} \frac{1}{\psi_q^j} \quad q = 1, \ldots, Q && \text{(A.59)}
\end{aligned}
$$

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

$$
\begin{aligned}
\bar{\mathbf{V}}_q &\triangleq \mathbf{V}^H \mathbf{V}_q && \text{(A.60)} \\
\bar{\mathbf{\Psi}}_q &\triangleq \bar{\mathbf{V}}_q \mathbf{\Psi}_q \bar{\mathbf{V}}_q^H = \mathbf{V}^H \mathcal{Q}_q \mathbf{V} && \text{(A.61)}
\end{aligned}
$$

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

On one hand, since $\mathbf{V}$ in (A.53) is a unitary matrix, $\bar{\mathbf{\Psi}}_q$ and $\mathcal{Q}_q$ have the same eigenvalues [12], i.e. $\mathbf{\Psi}_q$ are the eigenvalue matrix of $\bar{\mathbf{\Psi}}_q$. According to Lemma 2,

$$
[\psi_q^1, \ldots, \psi_q^M] \succ [\bar{\psi}_q^1, \ldots, \bar{\psi}_q^M] \tag{A.62}
$$

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 the ranges of elements of $[\psi_q^1, \ldots, \psi_q^M]$ and $[\bar{\psi}_q^1, \ldots, \bar{\psi}_q^M]$ 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 $\psi^M = \sum_{q=1}^{Q} \bar{\psi}_q^M = 0$, it follows that $\bar{\psi}_q^M = 0$. According to the definition of majorization, we may modify (A.62) as

$$[\psi_q^1, \ldots, \psi_q^{M-1}] \succ [\bar{\psi}_q^1, \ldots, \bar{\psi}_q^{M-1}] \tag{A.63}$$

Define ψ_q^o as the minimum of $[\psi_q^1, \ldots, \psi_q^{M-1}]$ and $\bar{\psi}_q^o$ as the minimum of $[\bar{\psi}_q^1, \ldots, \bar{\psi}_q^{M-1}]$. According to (A.47) and (A.48), $\bar{\psi}_q^o \geq \psi_q^o$. 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 $[\psi_q^1, \ldots, \psi_q^{M-1}]$ and $[\bar{\psi}_q^1, \ldots, \bar{\psi}_q^{M-1}]$ with $f(x) = x^{-1}$, then

$$\sum_{j=1}^{M-1} \frac{1}{\psi_q^j} \geq \sum_{j=1}^{M-1} \frac{1}{\bar{\psi}_q^j} \tag{A.64}$$

On the other hand, from $\mathcal{Q} = \sum_{q=1}^{Q} \mathcal{Q}_q$, it follows that

$$\begin{aligned} \mathbf{\Psi} &= \mathbf{V}^H \mathcal{Q} \mathbf{V} = \mathbf{V}^H (\sum_{q=1}^{Q} \mathcal{Q}_q) \mathbf{V} = \mathbf{V}^H (\sum_{q=1}^{Q} \mathbf{V}_q \mathbf{\Psi}_q \mathbf{V}_q^H) \mathbf{V} \\ &= \sum_{q=1}^{Q} \bar{\mathbf{V}}_q \mathbf{\Psi}_q \bar{\mathbf{V}}_q^H = \sum_{q=1}^{Q} \bar{\mathbf{\Psi}}_q \end{aligned} \tag{A.65}$$

and consequently

$$\psi^j = \sum_{q=1}^{Q} \bar{\psi}_q^j > \bar{\psi}_q^j > 0 \quad j = 1, \ldots, M-1 \tag{A.66}$$

Thus

$$\sum_{j=1}^{M-1} \frac{1}{\bar{\psi}_q^j} > \sum_{j=1}^{M-1} \frac{1}{\psi^j} \tag{A.67}$$

Based on (A.64) and (A.67), we have

$$Tr[\mathcal{Q}_q^\dagger] > Tr[\mathcal{Q}^\dagger] \tag{A.68}$$

Therefore, we conclude that $\mathrm{MSE}^o(S_q) > \mathrm{MSE}^o(S)$ for any $S_q \subset S$.

A.4 Proof of Theorem 4

Lemma 4 (AM-GM-HM Inequality) *[21] For any finite sequence of positive numbers $\mathbf{a} = [a_1, \ldots, a_n]$, we have*

$$AM = \frac{a_1 + \cdots + a_n}{n} \geq GM = (a_1 \cdots a_n)^{1/n} \geq HM = \frac{n}{\frac{1}{a_1} + \cdots + \frac{1}{a_n}} \tag{A.69}$$

with equality if and only if $a_1 = \cdots = a_n$

According to (A.64) and (A.69), we have

$$\begin{aligned} Tr[\mathcal{Q}^\dagger] &= \sum_{j=1}^{M-1} \frac{1}{\sum_{q=1}^{Q} \bar{\psi}_q^j} = \sum_{j=1}^{M-1} \frac{1}{\sum_{q=1}^{Q} c_q \frac{\bar{\psi}_q^j}{c_q}} \\ &\leq \sum_{j=1}^{M-1} \frac{\sum_{q=1}^{Q} \frac{c_q^2}{\bar{\psi}_q^j}}{(\sum_{q=1}^{Q} c_q)^2} = \frac{\sum_{q=1}^{Q} c_q^2 \sum_{j=1}^{M-1} \frac{1}{\bar{\psi}_q^j}}{(\sum_{q=1}^{Q} c_q)^2} \\ &\leq \frac{\sum_{q=1}^{Q} c_q^2 \sum_{j=1}^{M-1} \frac{1}{\psi_q^j}}{(\sum_{q=1}^{Q} c_q)^2} = \frac{\sum_{q=1}^{Q} c_q^2 Tr[\mathcal{Q}_q^\dagger]}{(\sum_{q=1}^{Q} c_q)^2} \end{aligned} \tag{A.70}$$

Therefore

$$\mathrm{MSE}^o(S) \leq \frac{\sum_{q=1}^{Q} c_q^2 \mathrm{MSE}^o(S_q)}{(\sum_{q=1}^{Q} c_q)^2} \tag{A.71}$$

A.5 Proof of Theorem 5

We define the following variables for convenience:

$$\underline{\mathbf{r}}^T \triangleq [\mathbf{r}_1^T, \ldots, \mathbf{r}_T^T] \tag{A.72}$$

$$\underline{\mathbf{x}}^T \triangleq [\mathbf{x}_1^T, \ldots, \mathbf{x}_T^T] \tag{A.73}$$

$$\underline{\mathbf{e}}^T \triangleq [\mathbf{e}_1^T, \ldots, \mathbf{e}_T^T] \tag{A.74}$$

$$\underline{\mathbf{b}}^T \triangleq [\mathbf{b}_1^T, \ldots, \mathbf{b}_T^T] \tag{A.75}$$

$$\underline{\mathbf{y}}^T \triangleq [\underline{\mathbf{h}}^T, \underline{\mathbf{b}}^T] \tag{A.76}$$

$$\mathcal{G}^T \triangleq [\mathbf{G}_1^T, \ldots, \mathbf{G}_T^T] \tag{A.77}$$

$$\mathcal{W} \triangleq \mathbf{I}_T \otimes (\mathbf{W}\boldsymbol{\Gamma}) \tag{A.78}$$

where $\underline{\mathbf{r}}$, $\underline{\mathbf{x}}$, $\underline{\mathbf{e}}$ are vectors with size $LT \times 1$; $\underline{\mathbf{b}}$ and $\underline{\mathbf{y}}$ are vectors with size $NT \times 1$ and $(JM + NT) \times 1$, respectively; $\mathcal{G}$ and $\mathcal{W}$ are matrix with size $LT \times JM$ and $LT \times NT$, respectively.

The received signal in T time iterations can be expressed as

$$\underline{\mathbf{r}} = \underline{\mathbf{x}} + \underline{\mathbf{e}} = \mathcal{W}\underline{\mathbf{b}} + \underline{\mathbf{e}} = \mathcal{G}\underline{\mathbf{h}} + \underline{\mathbf{e}} \tag{A.79}$$

The probability density function of noise vector $\underline{\mathbf{e}}$ is

$$f(\underline{\mathbf{e}}) = \frac{1}{(\pi\sigma^2)^{LT}} \exp[-\frac{1}{\sigma^2}[\underline{\mathbf{r}} - \underline{\mathbf{x}}]^H[\underline{\mathbf{r}} - \underline{\mathbf{x}}]] \tag{A.80}$$

The Fisher's Information Matrix (FIM) is defined as [11]

$$\mathbf{J} \triangleq E[(\frac{\partial f(\underline{\mathbf{e}})}{\partial \underline{\mathbf{y}}^*})(\frac{\partial f(\underline{\mathbf{e}})}{\partial \underline{\mathbf{y}}^H})] \tag{A.81}$$

It is known [11]

$$\frac{\partial f(\underline{\mathbf{e}})}{\partial \underline{\mathbf{y}}^*} = \begin{bmatrix} \frac{\partial f(\underline{\mathbf{e}})}{\partial \underline{\mathbf{h}}^*} \\ \frac{\partial f(\underline{\mathbf{e}})}{\partial \underline{\mathbf{b}}^*} \end{bmatrix} = \frac{1}{\sigma^2} \begin{bmatrix} \mathcal{G}^H \\ \mathcal{W}^H \end{bmatrix} \underline{\mathbf{e}} \tag{A.82}$$

Therefore, the FIM is

$$\mathbf{J} = \frac{1}{\sigma^2} \begin{bmatrix} \mathcal{G}^H \\ \mathcal{W}^H \end{bmatrix} \begin{bmatrix} \mathcal{G} & \mathcal{W} \end{bmatrix} \tag{A.83}$$

Lemma 5 (Constrained CRB) *[31] Let $\hat{\mathbf{y}}$ be an unbiased estimator of a parameter vector $\mathbf{y}$ satisfying a constraint $f(\mathbf{y}) = 0$. Define $\mathbf{F}(\mathbf{y}) \triangleq \frac{\partial f(\mathbf{y})}{\partial \mathbf{y}^T}$ 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 Cramer-Rao bound*

$$CRB_C = \mathbf{U}_0(\mathbf{U}_0^H \mathbf{J} \mathbf{U}_0)^{-1} \mathbf{U}_0^H \tag{A.84}$$

Lemma 6 (Minimal Constrained CRB) *[7] If Span$[\mathbf{U}_0]$ = 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*

$$CRB_C = \mathbf{J}^\dagger \tag{A.85}$$

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

Corollary 1 *[7] Suppose* $\mathbf{y}^T = [\mathbf{y}_1^T, \mathbf{y}_2^T]$ *and the FIM is*

$$\mathbf{J} = \begin{bmatrix} \mathbf{J}_{\mathbf{y}_1\mathbf{y}_1} & \mathbf{J}_{\mathbf{y}_1\mathbf{y}_2} \\ \mathbf{J}_{\mathbf{y}_2\mathbf{y}_1} & \mathbf{J}_{\mathbf{y}_2\mathbf{y}_2} \end{bmatrix} \tag{A.86}$$

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

$$CRB_{C,\mathbf{y}_1} = [\mathbf{J}_{\mathbf{y}_1\mathbf{y}_1} - \mathbf{J}_{\mathbf{y}_1\mathbf{y}_2}\mathbf{J}_{\mathbf{y}_2\mathbf{y}_2}^{-1}\mathbf{J}_{\mathbf{y}_2\mathbf{y}_1}]^\dagger \tag{A.87}$$

Applying the above corollary in our problem, we have

$$CRB_{C,\underline{\mathbf{h}}} = \sigma^2[\mathcal{G}^H\mathcal{G} - \mathcal{G}^H\mathcal{W}(\mathcal{W}^H\mathcal{W})^{-1}\mathcal{W}^H\mathcal{G}]^\dagger \tag{A.88}$$

Recall from our definitions of $\mathcal{G}$ in (A.77) and $\mathcal{W}$ in (A.78), we have

$$\mathcal{G}^H\mathcal{G} = \sum_{j=1}^{T} \mathbf{G}_j^H\mathbf{G}_j \tag{A.89}$$

and

$$\mathcal{G}^H\mathcal{W}(\mathcal{W}^H\mathcal{W})^{-1}\mathcal{W}^H\mathcal{G} = \sum_{j=1}^{T} \mathbf{G}_j^H(\mathbf{W\Gamma})[(\mathbf{W\Gamma})^H(\mathbf{W\Gamma})]^{-1}(\mathbf{W\Gamma})^H\mathbf{G}_j \tag{A.90}$$

And we know that $(\mathbf{W\Gamma})[(\mathbf{W\Gamma})^H(\mathbf{W\Gamma})]^{-1}(\mathbf{W\Gamma})^H$ is the orthogonal projector on the subspace Span$[\mathbf{W\Gamma}]$, which is equal to the signal subspace. Thus

$$(\mathbf{W\Gamma})[(\mathbf{W\Gamma})^H(\mathbf{W\Gamma})]^{-1}(\mathbf{W\Gamma})^H = \mathbf{U}_s\mathbf{U}_s^H \tag{A.91}$$

Therefore

$$CRB_{C,\underline{\mathbf{h}}} = \sigma^2[\sum_{j=1}^{T} \mathbf{G}_j^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{G}_j]^\dagger \tag{A.92}$$

Recall from the definition of $\mathbf{G}_j = [\mathbf{G}_j^1, \ldots, \mathbf{G}_j^J]$ in (2.44), we have

$$\sum_{j=1}^{T} \mathbf{G}_j^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{G}_j = \begin{bmatrix} \sum_{j=1}^{T} (\mathbf{G}_j^1)^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{G}_j^1 & \cdots & \sum_{j=1}^{T} (\mathbf{G}_j^1)^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{G}_j^J \\ \vdots & \ddots & \vdots \\ \sum_{j=1}^{T} (\mathbf{G}_j^J)^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{G}_j^1 & \cdots & \sum_{j=1}^{T} (\mathbf{G}_j^J)^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{G}_j^J \end{bmatrix} \tag{A.93}$$

According to the definition of $\mathbf{G}_j^m = \sum_{l=1}^{K^m} \gamma^{m,l} b_j^{m,l} \mathbf{C}^{m,l}$ in (2.43) and the asymptotic property in (3.27), we have, for any $1 \leq m, n \leq J$,

$$\begin{aligned} \sum_{j=1}^{T} (\mathbf{G}_j^m)^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{G}_j^n &= \gamma^m \gamma^n \sum_{l=1}^{K^m} \sum_{k=1}^{K^n} (\mathbf{C}^{m,l})^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{C}^{n,k} \sum_{j=1}^{T} b_j^{m,l} b_j^{n,k} \\ &= \gamma^m \gamma^n \sum_{l=1}^{K^m} \sum_{k=1}^{K^n} (\mathbf{C}^{m,l})^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{C}^{n,k} T \delta_{m,n} \delta_{l,k} \\ &= \begin{cases} (\gamma^m)^2 T \sum_{l=1}^{K^m} (\mathbf{C}^{m,l})^H \mathbf{U}_n \mathbf{U}_n^H \mathbf{C}^{m,l} & m = n \\ \mathbf{0} & m \neq n \end{cases} \end{aligned} \tag{A.94}$$

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 (3.36)), we obtain the close form of the minimal constrained CRB as

$$CRB_{C,\underline{\mathbf{h}}} = \frac{\sigma^2}{T} \text{diag}[(\mathcal{Q}^1)^\dagger, \ldots, (\mathcal{Q}^J)^\dagger] \tag{A.95}$$

Consequently

$$CRB_{C,\mathbf{h}^m} = \frac{\sigma^2}{T} (\mathcal{Q}^m)^\dagger \tag{A.96}$$

References

[1] E. Aktas and U. Mitra. Complexity reduction in subspace-based blind channel identification for DS/CDMA systems. *IEEE Trans. Commun.*, 48:1392 –1404, Aug. 2000.

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

[3] S. L. Campbell and C. D. Meyer. *Generalized Inverses of Linear Transformations*. Dover, 1979.

[4] B. Champagne, W. Kang, and H. C. Tam. Improved strategy for adaptive rank estimation with sperical subspace trackers. In *Proc. Canadian Conf. on Electrical and Computer Engineering (CCECE 2003)*, Montreal, QC, Canada, 2003.

[5] B. Champagne and Q. Liu. Plane rotation-based evd updating schemes for efficient subspace tracking. *IEEE Trans. Signal Processing*, 46:1886–1900, Jul. 1998.

[6] A. Chkeif, K. Abed-Meraim, G. Kawas-Kaleh, and Y. Hua. Spatio-temporal blind adaptive multiuser detection. *IEEE Trans. Commun.*, 48:729–732, May. 2000.

[7] E. de Carvalho, J. Cioffi, and D. Slock. Cramer-Rao bounds for blind multichannel estimation. In *Proc. Global Telecommunications Conf. (GLOBECOM 2000)*, volume 2, pages 1036 –1040, San Francisco, CA , USA, 2000.

[8] R. D. DeGroat and R. A. Roberts. Efficient, numerically stabilized rank-one eigenstructure updating. *IEEE Trans. Signal Processing*, 38:301–316, Feb. 1990.

[9] A. V. Geramita and J. Seberry. *Orthogonal Design, Quadratic Forms and Hadamard Matrices.* Marcel Dekker, 1979.

[10] G. H. Golub and C. F. V. Loan. *Matrix Computation.* Johns Hopkins University, 1996.

[11] S. S. Haykin. *Adaptive filter theory.* Prentice Hall, 4 edition, 2002.

[12] R. A. Horn and C. R. Johnson. *Matrix Analysis.* Cambridge, 1985.

[13] W. Kang and B. Champagne. Subspace-base blind channel estimation: generalization and performance analysis. *Submitted to IEEE Trans. Signal Processing.*

[14] W. Kang and B. Champagne. Generalized blind subspace channel estimation. *To appear in Proc. Vehicular Technology Conference, 2003 (VTC 2001 Fall)*, 2003.

[15] W. Kang and B. Champagne. A low-complexity adaptive blind subspace channel estimation algorithm. In *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2003)*, volume 4, pages 321–324, Hong Kong, China, 2003.

[16] C. Li and S. Roy. Subspace based detection in MC-CDMA over dispersive channel. In *Proc. IEEE Int. Conf. on Personal Wireless Communications*, pages 98 –103, Hyderabad, India, 2000.

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

[18] F. Li and R. J. Vaccaro. Analysis of Min-Norm and MUSIC with arbitrary array geometry. *IEEE Trans. Aerosp. Electron. Syst.*, 26:976 –985, Nov. 1990.

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

[20] A. W. Marshall and I. Olkin. *Inequalities: Theory of Majorization and Its Applications*. Academic Press, 1979.

[21] D. S. Mitrinovic. *Analytic Inequalities*. Springer-Verlag, 1970.

[22] E. Moulines, P. Duhamel, J. F. Cardoso, and S. Mayrargue. Subspace methods for the blind identification of multichannel FIR filters. *IEEE Trans. Signal Processing*, 43:527–534, Feb. 1995.

[23] J. Namgoong, T. F. Wong, and J. S. Lehnert. Subspace multiuser detection for multicarrier DS-CDMA. *IEEE Trans. Commun.*, 48:1897 –1908, Nov. 2000.

[24] P. Pango and B. Champagne. On the efficient use of givens rotations in svd-based subspace tracking algorithms. *Signal Processing*, 74:252–277, 1999.

[25] H. V. Poor and G. W. Wornell, editors. *Wireless communications: signal processing perspectives*. Prentice Hall, 1998.

[26] J. G. Proakis. *Digital Communications.* McGraw-Hill, 4 edition, 2001.

[27] D. J. Rabideau. Fast, rank adaptive subspace tracking and applications. *IEEE Trans. Signal Processing*, 44:2229–2244, Sep. 1996.

[28] D. Reynolds, X. Wang, and H. V. Poor. Blind adaptive space-time multiuser detection with multiple transmitter and receiver antennas. *IEEE Trans. Signal Processing*, 50:1261–1276, Jun. 2002.

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

[30] G. W. Stewart. An updating algorithm for subspace tracking. *IEEE Trans. Signal Processing*, 40:1535–1541, Jun. 1992.

[31] P. Stoica and B. C. Ng. On the Cramer-Rao bound under parametric constraints. *IEEE Signal Processing Lett.*, 5:177 –179, Jul. 1998.

[32] P. Strobach. Low-rank adaptive filters. *IEEE Trans. Signal Processing*, 44:2932–2947, Dec. 1996.

[33] E. G. Strom, S. Parkvall, S. L. Miller, and B. E. Ottersten. Propagation delay estimation in asynchronous direct-sequence code-division multiple access systems. *IEEE Trans. Commun.*, 44:84 –93, Jan. 1996.

[34] W. Sun and H. Li. Blind channel identification for multicarrier CDMA systems with transmit diversity. In *Proc. IEEE Int. Conf. on Communications (ICC 2002)*, volume 2, pages 727 –731, New York, NY, USA, 2002.

[35] V. Tarokh, H. Jafarkhani, and A. R. Calderbank. Space-time block codes from orthogonal designs. *IEEE Trans. Inform. Theory*, 45:1456 –1467, Jul. 1999.

[36] L. Tong and S. Perreau. Multichannel blind identification: from subspace to maximum likelihood methods. *Proc. IEEE*, 86:1951–1968, Oct. 1998.

[37] M. Torlak and G. Xu. Blind multiuser channel estimation in asynchronous CDMA systems. *IEEE Trans. Signal Processing*, 45:137 –147, Jan. 1997.

[38] H. L. V. Trees. *Detection, Estimation, and Modulation Theory, Part I.* John Wiley and Sons, 1968.

[39] X. Wang and H. V. Poor. Blind equalization and multiuser detection in dispersive CDMA channels. *IEEE Trans. Commun.*, 46:91 –103, Jan. 1998.

[40] Y. Wang, H. Zhang, and W. Wu. A simple blind adaptive receiver in multipath cdma channels. volume 1, pages 654 –656, 1999.

[41] J. Wu, Y. Wang, and K. K. M. Cheng. Blind channel estimation based on subspace for multicarrier CDMA. In *Proc. Vehicular Technology Conference, 2001 (VTC 2001 Spring)*, volume 4, pages 2374 –2378, Rhodes, Greece, 2001.

[42] Z. Xu. On the second-order statistics of the weighted sample covariance matrix. *IEEE Trans. Signal Processing*, 51:516 –525, Feb. 2003.

[43] B. Yang. Projection approximation subspace tracking. *IEEE Trans. Signal Processing*, 43:95–107, Jan. 1995.