Subspace-Based Blind Channel Estimation and Tracking for MIMO-OFDM Systems

Chao-Cheng Tu

Department of Electrical & Computer Engineering
McGill University
Montreal, Canada

July 2010

A thesis submitted to McGill University in partial fulfillment of the requirements for the degree of Doctor of Philosophy.

2010/07/30

Sommaire

Le multiplexage par répartition en fréquences orthogonales (orthogonal frequency division multiplexing, soit OFDM) à entrées et à sorties multiples (multiple-input multiple-output, soit MIMO) est maintenant généralement considéré comme une technologie à préconiser pour les nouveaux systèmes sans fil et ceux des générations ultérieures. Le MIMO-OFDM vise à augmenter la limite de capacité Shannon en combinant l'utilisation d'antennes multiples et la modulation orthogonale multiporteuse. Bien que la possibilité d'atteindre cette limite soit possible grâce à l'invention de techniques d'encodage et de décodage atteignant la capacité, en réalité, cette perspective d'avenir se base en grande partie sur l'existence et l'utilisation de techniques d'estimation de voie avancées. Pour faciliter l'estimation de voie rapide et fiable dans les systèmes MIMO-OFDM, on songe habituellement à l'insertion de symboles pilotes; cependant, la capacité de la voie est grandement réduite par leur insertion. L'utilisation d'estimation de voie aveugle fiable et à convergence rapide pour les MIMO-OFDM semble donc être une solution attrayante pour les futurs systèmes sans fil.

À cette fin, l'estimation de voie aveugle basée sur des statistiques de deuxième ordre, au lieu des statistiques d'ordre supérieur, est généralement considérée comme une candidate acceptable. Parmi les approches aveugles basées sur les statistiques de deuxième ordre, l'estimation basée dans le sous-espace est attrayante, puisque des estimations fiables peuvent souvent être obtenues de façon simple en optimisant une fonction de coût quadratique. Néanmoins, la performance des estimateurs de voie aveugles basés dans le sous-espace peut être gravement dégradée dans des conditions instationnaires. Ce problème peut habituellement rendre la performance globalement insatisfaisante, surtout dans les systèmes MIMO-OFDM avec un nombre de sous-porteuses élevé. Afin de compenser cette restriction et d'utiliser avec succès l'estimation de voie basée dans le sous-espace avec les systèmes MIMO-OFDM, il est essentiel de minimiser la longueur de la période de calcul des moyennes sous-jacente.

Dans la présente thèse, nous proposons un nouvel estimateur de voie aveugle basé dans le sous-espace qui ne nécessite qu'une période de calcul des moyennes relativement courte. Nous envisageons la conception d'un tel estimateur directement dans le domaine des fréquences, par opposition à la majorité des modèles existants où les estimateurs sont conçus dans le domaine temporel. Notre première contribution est de proposer et d'étudier un estimateur sous-espace innovateur avec un calcul des moyennes réduit en exploitant la

corrélation des fréquences au sein de sous-porteuses adjacentes résidant dans la largeur de bande de cohérence des voies à large bande des scénarios MIMO-OFDM typiques. Afin de réduire la grande complexité informatique créée par la décomposition des valeurs propres et la matrice d'ambigüité associée, notre seconde contribution est de mettre au point une version adaptative améliorée de l'estimateur pour augmenter sa capacité dans des conditions de variation temporelle MIMO. Nous réussissons cela en employant une forme modifiée de l'itération orthogonale permettant un repérage suffisant dans le sous-espace ainsi qu'une technique de précodage permettant une réduction de la taille de la matrice d'ambigüité. Les expériences numériques démontrent que les techniques proposées peuvent en effet avoir une meilleure performance que plusieurs des estimateurs de référence dans divers scénarios pratiques.

Abstract

Multiple-input multiple-output (MIMO) orthogonal frequency division multiplexing (OFDM) is now widely considered as a favored technology for emerging and future generation wireless systems. MIMO-OFDM aims to achieve increased channel capacity limit by exploiting the use of multiple antennas in combination with multi-carrier orthogonal modulation. While the possibility of achieving this limit is bestowed on the invention of capacity-achieving coding and decoding techniques, in reality, this prospect relies heavily on the existence and use of advanced channel estimation techniques. To facilitate fast and reliable channel estimation in MIMO-OFDM systems, pilot symbol insertion is usually considered; however, the channel capacity is greatly reduced by inserting those pilot symbols. Therefore, employing fast-converging and reliable *blind* channel estimation for MIMO-OFDM seems to be an attractive solution for future wireless systems.

To this end, blind channel estimation based on second order statistics (SOS), instead of higher order statistics (HOS), has been widely considered as a suitable candidate. Amid SOS-based blind approaches, subspace-based estimation is attractive since reliable estimates can often be obtained in a simple form by optimizing a quadratic cost function. Nonetheless, the performance of the subspace-based blind channel estimators may still be seriously degraded under time-varying conditions. This problem can generally make overall performance unsatisfactory, especially in MIMO-OFDM systems whose number of subcarriers is large. In order to overcome this limitation and successfully employ subspace-based channel estimation in MIMO-OFDM systems, it is essential to minimize the required length of the underlying time averaging period.

In this thesis, we propose a new subspace-based blind channel estimator that requires only a comparably short time averaging period. We consider the design of such an estimator directly in the frequency domain, as opposed to the majority of existing designs in which estimators are developed in the time domain. Our first contribution is to propose and investigate a novel subspace-based estimator with reduced time averaging, by exploiting the frequency correlation among adjacent subcarriers, residing within the coherence bandwidth of the broadband channels in typical MIMO-OFDM scenarios. To reduce the high computational complexity incurred by the eigenvalue decomposition and the associated ambiguity matrix, our second contribution is to develop an improved, adaptive version of the estimator for enhancing its capability under MIMO time-varying conditions. This

is achieved by employing a modified form of the orthogonal iteration for efficient subspace tracking along with a precoding technique that allows a reduction in the size of ambiguity matrix. Numerical experiments demonstrate that the proposed techniques can indeed outperform several benchmark estimators in various practical scenarios.

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Prof. Benoît Champagne for his guidance, encouragement, and support through the course of my Ph.D. studies. I have no doubt that my thesis could not have been completed without his help and encouragement. I am also grateful for his financial support via research grants from the Natural Sciences and Engineering Research Council of Canada (NSERC). I also acknowledge Prof. Mladen Glavinović from the Department of Physiology via the McGill work-study program, and l'Aide financière aux études of the Québec government, for providing financial support to complete the research.

Furthermore, I would like to thank Prof. Harry Leib, Prof. Ioannis Psaromiligkos and Prof. Frank Ferrie, members of my Ph.D. committee, for their feedback and suggestions. I owe a great many thanks to Prof. James Nemes, Prof. Jane Everett, my external examiner, Prof. Saeed Gazor from Queen's University, Kingston, and the members of my oral defense committee, namely: Pro-Dean, Prof. Amir Shmuel, Prof. Jan Bajcsy, Prof. Mark Coates, Prof. Zetian Mi, and Prof. Sonia Aïssa from Institut National de la Recherche Scientifique - Énergie, Matériaux et Télécommunications (INRS-EMT), Université du Québec. I am also grateful for many dedicated faculty and staff at McGill University, who have contributed to teachings and holding various workshops, making this experience fruitful and memorable. I am particularly grateful to my fellow graduate students in Telecommunications and Signal Processing (TSP) Laboratory and in McGill Taiwanese Graduate Student Association (MTGSA), for their companionship during my studies at McGill University. My special thanks go to Mr. Kar Lun (Clarence) Wong, Mr. Yongteng (David) Ma, Ms. Zhengyan (Stella) Shi, Ms. Isabel Deslauriers, and Dr. Ping Luo from TSP, and Ms. Pi-Yu Chiang and Kao-Shen Chung from MTGSA, who encouraged me in any respect during the completion of this thesis.

I am forever indebted to my parents, sister, and other family members for their love, support, and encouragement throughout my life. Last but not least, my deepest gratitude goes to my feline children, kiki and lulumi, for their companionship and emotional support.

Contents

List of Figures

List of Tables

List of Acronyms

3GPP	Third Generation Partnership Project
3GPP2	Third Generation Partnership Project 2
4G	Fourth Generation Mobile Communication System
AIE	Air Interface of Evolution
BER	Bit Error rate
CCSA	China Communication Standardization Association
CP	Cyclic Prefix
CSI	Channel State Information
E3G	Enhanced 3G
EVD	Eigenvalue Decomposition
FFT	Fast Fourier Transform
GSM	Global System for Mobile communications
HSDPA	High Speed Downlink Packet Access
ITU	International Telecommunication Union
IFFT	Inverse Fast Fourier Transform
LTE	Long Term Evolution
LORAF	Low Rank Adaptive Filter
MIMO	Multiple Input Multiple Output
MITF	Mobile IT Forum
NGMC	Next-Generation Mobile Committee
NIC	Novel Information Criterion
NRMSE	Normalized Root Mean Square Error

NMSE	Normalized Mean Square Error
OFDM	Orthogonal Frequency Division Multiplexing
OPAST	Orthogonal Projection Approximation and Subspace Tracking
RF	Radio Frequency
SC	Single Carrier
SIMO	Single Input Multiple Output
SISO	Single Input Single Output
SNR	Signal to Noise Ratio
SOS	Second Order Statistics
SVD	Singular Value Decomposition
TI	Time Invariant
TV	Time Varying
UE	User Equipment
UMB	Untramobile Broadband
UMTS	Universal Mobile Telecommunication System
VC	Virtual Carriers
WCDMA	Wideband Code Division Multiple Access
WWRF	Wireless World Research Forum

Notation

$\otimes$	Kronecker product
$\oslash$	element-wise division
$\odot$	Hadamard product
$\text{vec}\,(\cdot)$	Vec operator
$E[x]$	expected value of the random variable x
$\mathbf{I}_n$	$n \times n$ identity matrix
$\text{tr}(\mathbf{X})$	trace of the square matrix $\mathbf{X}$
$\|\mathbf{X}\|$	any consistent matrix norm
$\|\mathbf{x}\|_p$	p-norms of the vector $\mathbf{x}$
$\|\mathbf{X}\|_p$	p-norms of the matrix $\mathbf{X}$
$\|\mathbf{X}\|_F$	Frobenius norm of the matrix $\mathbf{X}$
$\mathbf{X}^\dagger$	pseudo-inverse of the matrix $\mathbf{X}$
$\mathbf{X} = \text{diag}(\mathbf{x})$	a diagonal matrix whose main diagonal is constructed from the entries of the vector $\mathbf{x}$
$\mathfrak{R}(\mathbf{X})$	range space of the matrix $\mathbf{X}$
$\mathfrak{N}(\mathbf{X})$	null space of the matrix $\mathbf{X}$
$\mathfrak{D}_r(\mathbf{X})$	the subspace spanned by the eigenvectors corresponding to the r largest eigenvalues of the matrix $\mathbf{X}$
$\lambda_r(\mathbf{X})$	the rth largest eigenvalue of the matrix $\mathbf{X}$
$\gamma_r(\mathbf{X})$	the rth smallest eigenvalue of the matrix $\mathbf{X}$
$\sigma_r(\mathbf{X})$	the rth largest singular value of the matrix $\mathbf{X}$

Chapter 1

Introduction

Since the first demonstration of radio signaling by Marconi in the late nineteenth century, the ability to communicate with people on the move has changed dramatically. Particularly during the past decade, the mobile communication technologies, fueled by various technical achievements, including digital and radio frequency (RF) circuit fabrication, very large-scale circuit integration, and digital switching techniques, have made the portable mobile devices more affordable and reliable [1–3].

The trend toward a more reliable and affordable portable device not only stimulated the rapidly growing number of users but also brought about a fundamental change on the design of wireless systems and networks [4]. For example, the traditionally voice-centered services has been gradually replaced by data-centered ones [5]. Besides, the data transmission rate has also increased tremendously, from 9.6 kbps in 1995 on a GSM system [2] to 2Mbps in 2005 on a WCDMA system [5, 6]; this represents more than 200 times increase in data rate within this ten year period. To date, the so-called "Super-3G" or "Beyond-3G" wireless systems and networks with a peak data transmission rate that can reach as high as 500Mbps have been demonstrated in the very recent field trials [7]. Even more ambitious 4G

wireless systems and networks which are targeting a peak data transmission rate of approximately 1Gbps, are also enthusiastically investigated by various national and international organizations [4, 8], such as the International Telecommunication Union (ITU), European Commission Framework Program (FP), Wireless World Research Forum (WWRF), Korean Next-Generation Mobile Committee (NGMC), Japanese Mobile IT Forum (MITF), and China Communication Standardization Association (CCSA). International standards organizations are currently working on the standardization of the Enhanced 3G (E3G) and the 4th Generation Mobile Communication System (4G), including the Long Term Evolution (LTE) plan for the 3rd Generation Partnership Project (3GPP) and the air interface of evolution/ultramobile broadband (AIE/UMB) plan of 3GPP2 [9, 10].

The radio spectrum still being a scarce and limited resource, high spectral efficiency is crucial to support the demand of high transmission rate from future mobile users. The LTE physical layer is targeted to provide improved radio interface capabilities between the base station and user equipment (UE), as compared to previous cellular technologies like Universal Mobile Telecommunications System (UMTS) [5] or High-Speed Downlink Packet Access (HSDPA) [11]. According to the initial requirements defined by the 3GPP (3GPP 25.913) [12], the LTE physical layer should support peak data rates of more than 100Mb/s over the downlink and 50Mb/s over the uplink [13]. A flexible transmission bandwidth ranging from 1.25 to 20MHz will provide support for users with different capabilities [14]. These requirements will be fulfilled by employing new technologies for cellular environments, such as orthogonal frequency division multiplexing (OFDM) and multiple-input multiple-output (MIMO) schemes (3GPP 36.201) [4, 15].

1.1 Channel estimation for wireless communications

A MIMO system deploys multiple antennas at both ends of a wireless link to increase the channel capacity and to mitigate adverse effects of the wireless channel [3, 16], while OFDM aims to provide high spectral efficiency and to eliminate the need for high-complexity equalization algorithms by decomposing the broadband radio channel into a set of orthogonal parallel sub-channels [14]. Therefore, MIMO-OFDM, which combines multiple transmit and receive antennas and the OFDM modulation, has become a practical alternative to either the single-carrier (SC) or the single-input single-output (SISO) transmission [17].

Three main techniques can be applied to fully exploit the special structures of the MIMO-OFDM systems, namely: diversity [18, 19], beamforming [5, 6, 16], and spatial multiplexing [4, 20]. Diversity techniques improve the average signal-to-noise ratio (SNR) on the receiver side by exploiting independent fading channels, as seen from either space, time, frequency, polarization, or a combination of these domains. In beamforming, signals are coherently combined (either in the transmitter or receiver) so as to enhance the antenna array response in a preferred direction. Spatial multiplexing offers a linear increase in the signaling rate by exploiting the parallel transmission of different information stream from different antennas. However, all these techniques can only achieve the desired improvement in performance when the channel state information (CSI) is perfectly known [21].

There are mainly two ways to obtain the required CSI, i.e., pilot-based versus blind estimation. In pilot-based channel estimation, known symbols are transmitted to assist the receiver in determining the CSI. Clearly this can only be done at the expense of a lower bandwidth efficiency. So far, pilot-based channel estimation is still considered to support most communication systems due to its reliability and low complexity. Nevertheless, with the ever increasing mobile speed requirement, pilot symbols have to be sent frequently

to cope with the time-variations of the wireless channels. Hence, if we insist in using the pilot-based channel estimation, this will greatly reduce the bandwidth efficiency and inevitably lessen the advantages brought by these MIMO technologies. In summary, there is an urgent need for a new blind channel estimation technique in MIMO-OFDM systems, and the research results on this topic will definitely influence the development of the future mobile communication technologies.

1.1.1 Subspace-based blind channel estimation

Among recent studies of MIMO-OFDM, blind channel estimation has received great attention and has become a vital area of research. Existing blind methods can broadly be categorized as statistical or deterministic: The former methods rely on assumptions on the statistics of the input sequence [22–24] while the latter make no such assumptions [25, 26]. In the first category, i.e., statistical approaches, blind channel estimation using second order statistics (SOS) can potentially achieve superior estimation performance for a given time averaging interval than approaches using higher order statistics (HOS) [27–31]. The second category, i.e., deterministic methods, is generally favored when the input statistics are unknown, or there may not be sufficient time samples to obtain the channel estimate. To date, several interesting deterministic methods have been developed by, e.g., employing the maximum likelihood approach [32, 33], exploiting null guard intervals [34, 35], exploiting zeros of the channel impulse response [36], or by using fractional sampling as well as interpolation [37]; however, most of them are exclusively for SISO or single-carrier transmissions.

Amid SOS-based blind approaches, subspace-based estimation is attractive since estimates can often be obtained in a simple form by optimizing a quadratic cost function [38]. Without employing any precoding at the transmitter, a subspace-based method is proposed

for OFDM systems by utilizing the redundancy introduced by the cyclic prefix (CP) [39, 40], and it is further extended for MIMO-OFDM systems in [41]. Virtual carriers (VCs) are subcarriers that are set to zero with no any information being transmitted. The presence of VCs provides another useful resource that can be used for channel estimation. Such a scheme is proposed for OFDM systems [42], and it is further extended to MIMO-OFDM systems in [43, 44].

The aforementioned approaches primarily exploit the separability of the noise and signal subspaces by applying the eigenvalue decomposition (EVD) to the correlation matrix of the received signals. In practice, the correlation matrix can only be estimated by averaging over multiple time samples, given the wireless channel is time-invariant during this averaging period. Since the quadratic cost function is constructed from the eigenvectors of the noise subspace obtained from the EVD, the accuracy of the eigenvectors obtained from the sampled correlation matrix dominates the performance of the estimation. Hence, in a time-invariant environment, the more time samples are averaged, the better the estimation performance is.

Considering that radio propagation conditions can only be invariant over a limited time interval (related to fading conditions, user mobility, etc.), it is legitimate to wonder how many samples are sufficient to obtain a sampled correlation matrix meeting a certain level of accuracy in the channel estimate. A basic rule is to assure that the number of the time samples must be no less than the dimension of the correlation matrix to make it full rank or invertible. Thus, to achieve desired estimation accuracy in the presence of noise, the required number of time samples for the CP and VC approaches may become prohibitive. For example, simulation results have shown that at least 500 OFDM symbols are required in order to achieve a normalized root mean square error (NRMSE) of 10^{-2} on the channel estimate, when we consider the number of subcarriers in OFDM is 16 and SNR = 20dB

[43]. If the number of subcarriers in OFDM is increased to 64, the number of required OFDM symbols would increase up to thousands for time averaging [44], making these subspace-based blind approaches impractical.

More recently, variants of the statistics-based methods have been proposed, e.g., by inserting zero-padding instead of CP for each OFDM block [45] or by introducing the so-called repetition index [46] and re-modulation [47] on the received signal. However, the number of required time samples is still implicitly proportional to the number of subcarriers (or the size of the IFFT) in the OFDM modulator. We also note that deterministic approaches still need to accumulate data samples in order to algebraically obtain channel estimates, and their performance in noise improves as the number of samples increases. Therefore, as the dimension of the parameter space is increased in the MIMO-OFDM context, the number of samples required for deterministic methods to achieve an acceptable level of performance will also inevitably be increased.

1.1.2 Tracking time-varying channels by subspace updating

Tracking time-varying (TV) channels with a large Doppler spread is a critical task, regardless of whether a non-blind or blind approach is used [48]. A non-blind approach in general requires to employ pilots more frequently since the channel estimate becomes obsolete shortly after the training period ends. On the contrary, a blind approach eliminates the need of large amount of pilot symbols and therefore is favored if complexity is not the main concern.

There are mainly two categories of *blind* approaches for tracking rapidly time-varying channels: (1) using a block processing approach to estimate the unknown parameters of an underlying time-varying channel model, and (2) using an adaptive processing algorithm that is sufficiently fast to track the channel variations. Among various blind approaches

in the first category, a basis expansion model has been proposed to convert a TV-SISO channel into a time-invariant (TI) single input multiple output (SIMO) channel, followed by a standard SOS-based subspace method for blind channel estimation [49]. The idea of basis expansion was further extended for a TV-SIMO channel [50–52], and a generalized OFDM system over a TV-SISO channel [53]. Similarly, an interpolation model was also proposed to convert a TV-SISO channel into fixed parameters in a long-code code division multiple access system [54].

Recently, there has been much interest in the search of new adaptive algorithms in the second category. A zero padding SISO-OFDM system associated with either recursive least squares or least mean squares method for blind adaptive channel estimation was considered in [55]. It was reported that for an IFFT size of 64 and padding length of 16, the relative channel estimation error converges to -27dB in 500 symbols when the maximum Doppler shift is limited to 100Hz and the SNR is 20dB. By properly choosing the so-called repetition index, a cyclic prefixing SISO-OFDM system was also proposed in [46], where it was reported that for a 64-point IFFT with a cyclic prefix length of 16, the bit error rate (BER) can reach a level of 10^{-2} within 12 received blocks when the maximum Doppler shift is 50Hz and the SNR $\geq$ 25dB.

While the above adaptive approaches offer interesting capabilities in tracking time-varying channels with high spectral efficiency, they may not be adequate for applications in future generation of broadband mobile wireless systems, in which there is a need to provide high-rate transmission, e.g., a real-time video stream, between a user terminal and an access node whose relative position may vary rapidly over time [56]. For example, the 3GPP LTE specifications call for high performance broadband transmission with mobile speed up to 120km/h, corresponding to a maximum Doppler shift of 220Hz, and additional provision to support much higher speeds up to e.g. 350km/h (high speed train) [12]. In

the 15 - 120km/h speed range, the targeted data rates are 50Mbps for the uplink, up to 100Mbps for the downlink, with increased spectral efficiency of 2-3 times of Release 6 enhanced uplink and 3-4 times of Release 6 HSDPA, respectively.

Based on these considerations, there is a need to further push the capability of adaptive channel tracking for even faster time-varying channels. In addition, this goal should not result in the loss of any bandwidth efficiency, or placing restrictions on the number of transmit or receive antennas. We notice that there also exist algorithms for channel estimation in mobile MIMO-OFDM systems with large Doppler shifts [57–59]; however, they all require proper preambles or training sequence to work.

1.2 Research objectives and methodology

The applicability of traditional subspace-based blind estimators and trackers is severely limited by the requirement of a large time averaging period. The main objective of this research work is to develop new subspace-based blind estimators and trackers for MIMO-OFDM systems to relieve such a constraint. This is mainly achieved by exploiting the frequency correlation among adjacent subcarriers in OFDM transmissions through subcarrier grouping [60], for which some supportive field measurements can also be found in [61].

In light of the discussions given in Section 1.1, it can be concluded that a number of elements in existing subspace-based algorithms for blind channel estimation in MIMO-OFDM systems can be considered for possible improvements. In particular, the following topics are studied in detail in this work:

1. *Time-invariant block-based channel estimation:*

   Subspace-based blind channel estimators are characterized by good performance and

simple structures; however, their requirement of a large time averaging period makes them less appealing for practical uses. In this thesis, we seek to develop a more efficient estimation algorithm in order to reduce the time averaging period while maintaining the same performance. In addition, we will investigate the performance of the newly developed algorithms to demonstrate their usefulness in practical scenarios.

2. *Time-varying adaptive channel tracking:*

   Orthogonal iteration and its variants have been considered for subspace tracking to a large extent. On the basis of the orthogonal iteration, we aim to extend the concept of the aforementioned block-based MIMO-OFDM channel estimator to an adaptive channel tracker in time-varying scenarios. Although the convergence properties of orthogonal iteration in stationary cases are well understood, the corresponding properties in non-stationary ones are not. Hence, it requires a better study of different aspect of the convergence properties of orthogonal iteration in various non-stationary scenarios. With a better understanding of those properties in non-stationary scenarios, we can develop an orthogonal-iteration-based subspace trackers for MIMO-OFDM systems over time-varying channels with large Doppler spread.

The proposed subspace-based blind estimator and tracker will be studied and compared using well-proven analytical and simulation approaches. The simulation experiments will focus on measuring the NMSE of the proposed estimators and the corresponding BER when the latter are embedded in a complete MIMO-OFDM transmission chain.

1.3 Contributions and claim of originality

In this work, new subspace-based blind channel estimation and tracking algorithms for MIMO-OFDM systems are proposed. Through analytical studies and numerical simulations, the proposed algorithms are shown to achieve superior performance compared to existing algorithms in the literature. The main original contributions of this research work can be summarized as follows:

- Generalization of the subspace-based blind estimation for XIXO-OFDM systems over a quasi-stationary wireless channel (where XIXO denotes the abbreviation of SIMO, MISO, and MIMO configurations).
- Development of a novel subspace-based blind estimator for MIMO-OFDM systems, reducing the number of required data samples for time averaging.
- Proof of identifiability conditions of the proposed subspace-based blind estimator.
- Derivation of asymptotical performance bound and Cramer-Rao bound of the proposed subspace-based blind estimator.
- Development of a novel subspace tracking algorithm for precoded MIMO-OFDM systems over a rapidly time-varying wireless channel.
- Extension of the convergence analysis of the orthogonal iteration to include the non-stationary cases.
- Derivation of a fundamental limitation on the use of orthogonal iteration when it is applied to fast time-varying channels.

These contributions have led to a number of publications in peer-reviewed journals and refereed conferences, as listed below:

Journal papers

J-1) C. C. Tu and B. Champagne, "Performance analysis of blind subspace-based MIMO-OFDM channel estimator exploiting frequency correlation," (in preparation).

J-2) C. C. Tu and B. Champagne, "Blind subspace tracking of wideband time-varying MIMO channels with large Doppler spread," submitted to *IEEE Trans. Wireless Commun.*, Feb. 2010.

J-3) C. C. Tu and B. Champagne, "Subspace-based blind channel estimation for MIMO-OFDM systems with reduced time averaging," *IEEE Trans. on Veh. Technol.*, vol. 59, No. 3, pp. 1539-1544, March 2010.

Conference papers

C-1) C.-C. Tu and B. Champagne, "On convergence properties of subspace trackers based on orthogonal iteration," in *Proc. IEEE Pacific Rim Conf. on Commun., Comput. and Signal Process.*, Aug. 2009, pp. 65-70.

C-2) C.-C. Tu and B. Champagne, "Subspace tracking of fast time-varying channels in precoded MIMO-OFDM systems," in *Proc. IEEE Int. Conf. on Acoust., Speech and Signal Process.*, Apr. 2009, pp. 2565-2568.

C-3) C.-C. Tu and B. Champagne, "Subspace blind MIMO-OFDM channel estimation with short averaging periods: performance analysis," in *Proc. IEEE Wireless Commun. Networking Conf.*, Mar. 2008, pp. 24-29.

C-4) C.-C. Tu and B. Champagne, "Subspace-based blind channel estimation for MIMO-OFDM systems: reducing the time averaging interval of the correlation matrix," in *Proc. IEEE Int. Symp. on Pers. Indoor Mobile Radio Commun.*, Sept. 2007, pp. 1-5.

1.4 Thesis organization

An introduction to MIMO-OFDM systems is given in Chapter 2, along with an overview of various radio propagation effects in the mobile radio channels. In order to characterize the channel parameters to be estimated, a brief review of channel models that are largely used to generate channel parameters for numerical experiments is addressed in a later section.

In Chapter 3, the motivation for channel estimation in a wireless system is recapitulated. This is accompanied by a discussion on its alternatives, namely the non-coherent and differential techniques, in which the CSI is not required at the receiver side. Then various channel estimation methods, including pilot- and blind-based approaches for MIMO-OFDM systems are enumerated and discussed in detail. This chapter is concluded by addressing the state of the art of blind channel estimation in terms of its performance and associated limitations.

In Chapter 4, we propose a new subspace-based estimation scheme to improve the aforementioned deficiencies. This therefore leads to the first contribution of the thesis, in which

a new subspace-based algorithm relieving the limitation of a subspace-based estimation is proposed. The identifiability condition, asymptotical performance bound and Cramer-Rao bound of the proposed estimator are also presented.

Due to the high computational complexity incurred in the eigenvalue decomposition and the need to identify time-varying channels, the proposed algorithm is further extended to incorporate the idea of subspace tracking, in which the subspace information is updated rather than recomputed for a new channel estimate. In chapter 5, we focus on analyzing the convergence properties of the so-called orthogonal iteration method, which will be employed for tracking the subspace of interest. Then we propose a new time-varying channel tracking algorithm in Chapter 6, leading to the second contribution of the thesis. Specifically, the assumption of the wireless channel considered is further relieved to be only quasi-stationary within one OFDM symbol. Numerical results of the proposed algorithms are then presented in Chapter 7, and conclusions are drawn in Chapter 8.

Chapter 2

Background: MIMO-OFDM system and channel models

In this chapter, we first introduce the MIMO-OFDM system model under consideration. Then, an overview of various radio propagation effects and their modeling is given. This includes a discussion on the concepts of channel delay profile and channel coefficient characterization, which play a central role in the channel estimation problem. The chapter ends with a brief review of some important wireless channel models that are widely used nowadays for numerical experiments.

2.1 Introduction to MIMO-OFDM transmission systems

Demands for higher capacity in wireless communications, driven by high data rate applications and multimedia services, are never ceasing [56]. However, the available frequency spectrum is limited and the high capacity needs of these new applications cannot be fulfilled without a significant increase in the communication spectral efficiency [62].

With the advances in channel coding schemes such as turbo codes [63] or low density

parity check (LDPC) codes [64], Shannon capacity can be approached with single antenna [18]. However, the use of MIMO systems can further push the fundamental capacity limits with the same SNR, compared to the single antenna systems [62]. MIMO systems are created by deploying multiple antennas at both ends of a wireless link, i.e. transmitter and receiver sides. Using multiple antennas in wireless communications has been proposed to increase channel capacity; research results [65, 66] have shown that channel capacity increases linearly with the number of antennas deployed at both the transmitter and receiver sides under ideal propagation scenarios, where channel coefficients are modeled as independent complex Gaussian random variables. Hence, by suitable coding designs, large gains in capacity over wireless communications are feasible in MIMO systems as compared to traditional SISO systems; however, if we consider a broadband communication system, then conventional SC modulation will inevitably incur a high complexity for the MIMO detection.

OFDM provides a good solution to this problem. Indeed, by combining the advantages offered by MIMO and OFDM, the broadband frequency selective fading MIMO channel can be treated as a collection of multiple independent flat fading MIMO sub-channels, which greatly reduces the complexity of the MIMO detection. To date, OFDM has been widely adopted into various standards, including the European digital audio broadcast (DAB) [67] as well as the digital video broadcast (DVB) scheme [68]. It was also selected as the high performance local area network Type 2 standard (HiperLAN/2) [69] as well as part of the IEEE 802.11a/b/g wireless local area network (WLAN) standard [56]. Furthermore, it has been included in the Super-3G mobile radio standards, and has also being in the standardization process of 3GPP LTE [15]. In the following, we introduce the OFDM system model for a SISO channel first, and then extend the concept to MIMO channels.

2.1.1 SISO-OFDM

OFDM belongs to a family of transmission schemes called multi-carrier modulation, and is very attractive for broadband wireless systems which encounter large delay spread [4]. Its main idea is to divide the serial input data stream into a number of parallel streams and then modulate each stream on separate carriers, and transmit these low-rate parallel streams simultaneously. By doing so, the delay spread is only a small fraction of the symbol duration, which considerably simplify the task of channel equalization. A cyclic prefix (CP) can be applied to each OFDM symbol to remove intersymbol interference (ISI) with a small penalty in channel capacity [70]. In addition, these separate carriers, often called subcarriers, are allowed to overlap in the frequency domain by maintaining orthogonality of their corresponding time domain waveforms over the symbol duration. As a result, the bandwidth efficiency of OFDM is very high. Moreover, the CP enables the use of fast Fourier transform (FFT) for OFDM implementation, and thus greatly reduces the hardware complexity [7].

In the following, we describe more details regarding the operation of a OFDM transmission chain. To this end, we assume that the channel impulse response of the SISO channel under consideration is represented as $h(l)$, $l = 0, 1, \cdots, L-1$, where L denotes the channel order. In addition, we assume $\mathbf{x} \stackrel{\text{def}}{=} [\, x[0]\ x[1]\ \cdots\ x[N_C - 1]\,]^T$ denotes a data vector to be transmitted over the SISO channel, with $E[|x[k]|^2] = 1$. The OFDM modulation illustrated in Fig. 2.1 can be described as follows:

The transmitter first performs an inverse fast Fourier transform (IFFT) operation on the data vector $\mathbf{x}$. Thus, the output of the IFFT operation can be denoted as $\tilde{\mathbf{x}} = \mathbf{D}^H \mathbf{x}$,

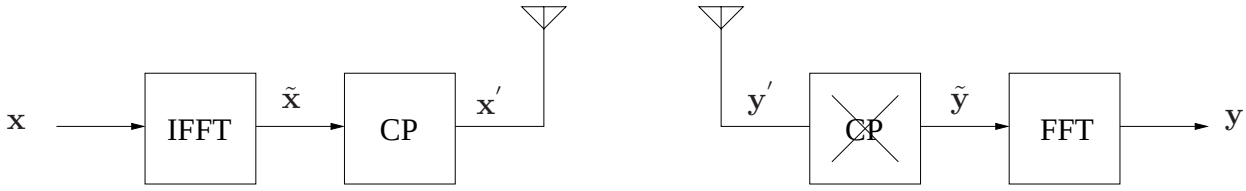


Fig. 2.1 A simple illustration of the OFDM modulation.

where $\mathbf{D}$ is an $N_C \times N_C$ matrix with the (m,n)th entry defined as

$$[\mathbf{D}]_{m,n} = \frac{1}{\sqrt{N_C}} \exp(-j2\pi(m-1)(n-1)/N_C). \tag{2.1}$$

In this thesis, the entries of the matrix $\mathbf{D}$ are normalized such that the latter is unitary, i.e. $\mathbf{D}\mathbf{D}^H = \mathbf{I}$. Then a new data vector $\mathbf{x}'$ is constructed by appending the CP of length $N_{cp} \geq L-1$, which consists of the last N_{cp} symbols of $\tilde{\mathbf{x}}$, to the data vector $\tilde{\mathbf{x}}$ itself. Hence, $\mathbf{x}'$ can be written as $\mathbf{x}' \stackrel{\text{def}}{=} [\,\tilde{x}[N_C - N_{cp}] \;\cdots\; \tilde{x}[N_C-1]\; \tilde{x}[0] \;\cdots\; \tilde{x}[N_C-1]\,]^T$; this vector is serially transmitted beginning with the symbol $\tilde{x}[N_C - N_{cp}]$.

Assuming perfect symbol synchronization, the received data vector $\mathbf{y}'$, corresponding to the transmitted sequence $\mathbf{x}'$, is of length $N_C + N_{cp} + L - 1$, as a result of the convolution of the transmitted sequence convolved with the channel impulse sequence of length L. The receiver's first task is to strip off the CP from $\mathbf{y}'$ and then collects N_C samples of the received signal to construct a new vector $\tilde{\mathbf{y}}$. The latter satisfies

$$\tilde{\mathbf{y}} = \tilde{\mathbf{G}}\mathbf{x}' + \tilde{\mathbf{n}}, \tag{2.2}$$

where $\tilde{\mathbf{n}}$ is the zero mean circularly symmetric complex Gaussian (ZMCSCG) noise vector

with covariance matrix $\sigma_n^2 \mathbf{I}_{N_C}$, and $\tilde{\mathbf{G}}$ is an $N_C \times (N_C + N_{cp} - 1)$ Toeplitz matrix defined as

$$\tilde{\mathbf{G}} = \begin{bmatrix} 0 & \cdots & 0 & h(L-1) & \cdots & h(0) & 0 & 0 & \cdots & 0 \\ 0 & 0 & \cdots & 0 & h(L-1) & \cdots & h(0) & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & 0 & \cdots & 0 & h(L-1) & \cdots & h(0) & 0 \\ 0 & \cdots & 0 & 0 & \cdots & 0 & 0 & h(L-1) & \cdots & h(0) \end{bmatrix}, \tag{2.3}$$

where the first $N_{cp} - L + 1$ columns are zeros vectors, resulting from our choice of the CP such that $N_{cp} \geq L - 1$. Since the first N_{cp} samples of the data vector $\mathbf{x}'$ are identical to its last N_{cp} samples, (2.2) can be re-written as $\tilde{\mathbf{y}} = \tilde{\mathbf{G}}_c \tilde{\mathbf{x}} + \tilde{\mathbf{n}}$, where

$$\tilde{\mathbf{G}}_c = \begin{bmatrix} h(0) & 0 & \cdots & 0 & 0 & h(L-1) & \cdots & h(1) \\ h(1) & h(0) & 0 & \cdots & 0 & 0 & \ddots & \vdots \\ \vdots & h(1) & h(0) & 0 & 0 & \ddots & 0 & h(L-1) \\ h(L-1) & \vdots & h(1) & \ddots & 0 & \ddots & 0 & 0 \\ 0 & h(L-1) & \vdots & \ddots & h(0) & \ddots & \ddots & 0 \\ \vdots & 0 & h(L-1) & \ddots & h(1) & h(0) & 0 & 0 \\ \vdots & \vdots & 0 & \ddots & \vdots & \ddots & \ddots & 0 \\ 0 & 0 & \cdots & 0 & h(L-1) & \cdots & h(1) & h(0) \end{bmatrix}, \tag{2.4}$$

where $\tilde{\mathbf{G}}_c$ is now an $N_C \times N_C$ matrix whose circular structure plays an important role in the study of OFDM system. Indeed, we first observe that $\tilde{\mathbf{G}}_c$ being circulant, its EVD may be expressed as $\tilde{\mathbf{G}}_c = \mathbf{D}^H \mathbf{\Omega} \mathbf{D}$, where $\mathbf{\Omega} \stackrel{\text{def}}{=} \text{diag}\,(\, h[0]\; h[1]\; \cdots\; h[N_C - 1]\,)$ and

$$h[k] = \sum_{l=0}^{L-1} h(l) \exp(-j2\pi lk/N_C), \quad k = 0, 1, 2, \cdots, N_C - 1, \tag{2.4}$$

with $h[k]$ being the sampled frequency response of the kth sub-channel.

In order to recover the transmitted symbols, the receiver performs a fast Fourier transform (FFT) on the data vector $\tilde{\mathbf{y}}$. Thus, the output of the FFT operation can be denoted as $\mathbf{y} = \mathbf{D}\tilde{\mathbf{y}}$, and re-written by

$$\mathbf{y} = \mathbf{D}(\tilde{\mathbf{G}}_c\tilde{\mathbf{x}} + \tilde{\mathbf{n}}) = \mathbf{D}\underbrace{\mathbf{D}^H\mathbf{\Omega}\mathbf{D}}_{=\tilde{\mathbf{G}}_c}\underbrace{\mathbf{D}^H\mathbf{x}}_{=\tilde{\mathbf{x}}} + \mathbf{D}\tilde{\mathbf{n}} = \mathbf{\Omega}\mathbf{x} + \mathbf{n}, \tag{2.5}$$

where we have used the fact that $\mathbf{D}\mathbf{D}^H = \mathbf{I}_{N_C}$, and $\mathbf{n} \stackrel{\text{def}}{=} \mathbf{D}\tilde{\mathbf{n}}$. Note that $\mathbf{n}$ is still a ZMCSCG noise vector with covariance matrix $\sigma_n^2\mathbf{I}_{N_C}$.

Since $\mathbf{\Omega}$ is a diagonal matrix, the kth entry of the vector $\mathbf{y}$ in (2.5) can be expressed by

$$y[k] = h[k]x[k] + n[k], \quad k = 0, 1, 2, \cdots, N_C - 1, \tag{2.6}$$

where $n[k]$ denotes the kth entry of the vector $\mathbf{n}$. Therefore, we observe that the use of a CP in conjunction with the IFFT and FFT operations at the transmitter and the receiver, respectively, decouples the wireless channel into N_C parallel sub-channels.

Having discussed the basic principle of OFDM modulation for SISO channels, we now explain how this modulation schemes may be extended to MIMO channels.

2.1.2 MIMO-OFDM

We now extend the above discussion to a MIMO channel with N_T transmit antennas, N_R receive antennas. The MIMO channel is depicted in Fig. 2.2. Let the symbol to be transmitted at the qth transmit antenna over the kth subcarrier be $x_q[k]$, $q = 1, 2, \cdots, N_T$. Similar to the case of SISO-OFDM, the data block to be transmitted over each transmit antenna is first subject to an IFFT operation followed by the CP insertion. At each of the

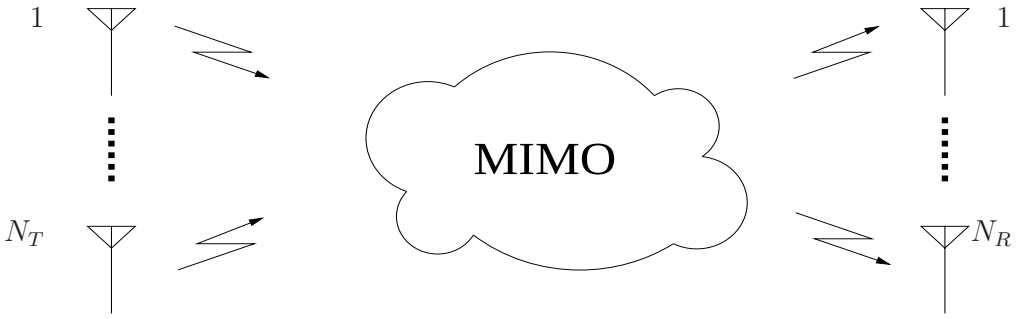


Fig. 2.2 A simple illustration of an $N_T \times N_R$ multiple-input multiple-output channel.

receive antenna, the CP is discarded and then followed by an FFT operation. Therefore, for the considered MIMO-OFDM system, the signal received at the pth receive antenna over the kth subcarrier is given by

$$y_p[k] = \sum_{q=1}^{N_T} h_{p,q}[k]x_q[k] + n_p[k], \quad p = 1, 2, \cdots, N_R, \tag{2.7}$$

where $h_{p,q}[k]$ denotes the channel gain between the qth transmit antenna and the pth receive antenna over the kth tone, and $n_p[k]$ represents the ZMCSCG noise at the pth receive antenna over the kth subcarrier.

Let $\mathbf{x}[k] \stackrel{\text{def}}{=} [\, x_1[k] \;\; x_2[k] \;\; \cdots \;\; x_{N_T}[k] \,]^T$, and $\mathbf{x} \stackrel{\text{def}}{=} \left[\mathbf{x}[0]^T \;\; \mathbf{x}[1]^T \;\; \cdots \;\; \mathbf{x}[N_C-1]^T\right]^T$ be a block of data transmitted over this MIMO channel at a given symbol epoch. It follows that the received data block $\mathbf{y} \stackrel{\text{def}}{=} \left[\mathbf{y}[0]^T \;\; \mathbf{y}[1]^T \;\; \cdots \;\; \mathbf{y}[N_C-1]^T\right]^T$, with $\mathbf{y}[k] \stackrel{\text{def}}{=} [\, y_1[k] \;\; y_2[k] \;\; \cdots \;\; y_{N_R}[k] \,]^T$, can be related to $\mathbf{x}$ according to

$$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{n}, \tag{2.8}$$

where $\mathbf{n} \stackrel{\text{def}}{=} \left[\mathbf{n}[0]^T \;\; \mathbf{n}[1]^T \;\; \cdots \;\; \mathbf{n}[N_C-1]^T\right]^T$ with $\mathbf{n}[k] \stackrel{\text{def}}{=} [\, n_1[k] \;\; n_2[k] \;\; \cdots \;\; n_{N_R}[k] \,]^T$, and

$\mathbf{H} \stackrel{\text{def}}{=} \text{diag}\,(\,\mathbf{H}[0]\;\mathbf{H}[1]\;\cdots\;\mathbf{H}[N_C-1]\,)$, with diagonal blocks defined as

$$\mathbf{H}[k] = \begin{bmatrix} h_{1,1}[k] & h_{1,2}[k] & \cdots & h_{1,N_T}[k] \\ h_{2,1}[k] & h_{2,2}[k] & \cdots & h_{2,N_T}[k] \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_R,1}[k] & h_{N_R,2}[k] & \cdots & h_{N_R,N_T}[k] \end{bmatrix} \in \mathbb{C}^{N_R \times N_T}. \tag{2.9}$$

If we choose the subcarrier spacing judiciously, MIMO-OFDM decomposes the frequency selective channel of bandwidth B into N_C frequency flat fading MIMO sub-channels, each with bandwidth B/N_C. However, due to the CP insertion, (MIMO-)OFDM transmission in average incurs a loss in spectral efficiency of $N_{cp}/(N_C + N_{cp})$; e.g., in IEEE 802.11a standard for WLAN, the efficiency loss due to the CP insertion is no more than 20% [71]. Nevertheless, this loss is negligible if $N_C \gg N_{cp}$.

To describe the input-output relationship for the SISO-OFDM system in (2.6) and MIMO-OFDM system in (2.8), we have assumed that the channel impulse responses or their corresponding channel gains, i.e., $h[k]$'s and $h_{p,q}[k]$'s, are readily available. Nevertheless, to formulate these quantitative terms from a practical wireless channel encompassing various propagation effects are never trivial. In the following, we give an overview of the most prominent wireless propagation effects, and then present how suitable models of channel impulse responses or their corresponding channel gains can be developed on this basis.

2.2 The wireless propagation channel

In wireless communication systems, information is transmitted from the emitter to the receiver in the form of radio waves propagating through the so-called *wireless propagation channel*. The channel distorts the transmitted signal in an unpredictable way, and thus

places fundamental limitations on the performance of the wireless communication systems. Hence, the receiver can only recover the information by employing channel estimation or equalization techniques.

To characterize the distortion imposed on the transmitted signal, we need to understand the aforementioned propagation environment, which can be suitably described as an idealized phenomenon in which a signal propagating through the wireless channel arrives at the destination along a number of different paths. Except for the simple line-of-sight, propagation along these paths is severely obstructed by buildings, mountains, and foliage, which results into scattering, reflection and diffraction of the radiated energy by objects in the environment or refraction in the medium, and hence can influence path loss and fading models differently. In addition, as a mobile station moves in space, the speed of the mobile impacts significantly on how rapidly the signal level fades, and therefore contribute to an extremely random mechanism that do not offer easy analysis.

2.2.1 Propagation mechanism

In general, propagation models can be broadly categorized as large scale or small scale. A large scale model mainly focuses on predicting the average received signal strength at a given distance from the transmitter, and thus is useful for estimating the radio coverage area of a transmitter. On the other hand, a small-scale model describes the rapid fluctuations of the received signal strength over very short travel distances or short time duration, and thus is useful for evaluating the performance of a transceiver chain. In the following, we introduce the large scale and small scale models respectively.

Large scale model

Large scale model designates the average signal power attenuation over large areas. The total signal attenuation caused by wave propagation along a transmission path is often referred to as *path loss*. In a free-space model, the power received by a receive antenna that is separated from a transmit antenna by a certain distance d can be calculated by the Friis free-space equation [72], which described as follows

$$P_r = P_t \left(\frac{\lambda}{4\pi d}\right)^2 G_t G_r, \tag{2.10}$$

where P_t, P_r denote the transmitted and the received powers respectively, λ is the wavelength, G_t, G_r represents the antenna gains of the transmit and receive antennas respectively. The Friis free-space equation establishes that the received power is reduced as a function of the square of the distance between the transmitter and the receiver. However, in a typical cellular wireless channel, free-space propagation seldom happens. Besides free-space propagation loss, *reflection*, *diffraction*, and *scattering* also affect propagation. These three mechanisms contribute to the large scale signal attenuation in different degrees.

Reflection, diffraction, and scattering are main propagation mechanisms which impact propagation in a mobile communication system. Reflection occurs when a propagating electromagnetic wave hits an object which has very large dimensions when compared to the wavelength of the propagating wave, as in the case of a radio wave bouncing from the surface of the earth, buildings, walls, and so on. Reflection is the major cause of multipath effect in wireless communication channels.

Diffraction occurs when the electromagnetic wave propagation path is obstructed by a surface that has of many sharp irregularities or edges. The secondary waves resulting from the obstructing surface are present throughout the space and sometimes behind the

obstacle, giving rise to a bending of waves around the obstacle, even when a line-of-sight path does not exist between transmitter and receiver. Diffraction takes place less often when carrier frequency is higher, as the electromagnetic waves behave more like particles rather than waves.

Scattering occurs when the medium through which the wave travels consists of objects with dimensions that are small when compared to the wavelength of the propagating wave, and where the number of obstacles per unit volume is relatively large. Scattered waves are usually generated by rough surfaces, small objects, or other irregularities in the radio channel. In practice, foliage, street sign poles and lamp posts induce scattering in a radio propagation channel.

Even if line-of-sight transmission does occur, the combination of the signal components from the propagation mechanisms above, will prevent the received signal from obeying the free-space propagation law and will eventually resulting in a larger attenuation than what is predicted by the Friis free-space equation alone. While difficult to calculate exactly, the actual path loss is often modeled as a function of the distance between the transmitter and the receiver raised to the *path loss exponent*. In this case, the received power is now represented as

$$P_r = P_t \left(\frac{\lambda}{4\pi d} \right)^{\gamma} G_t G_r, \tag{2.11}$$

with the path loss exponent $\gamma > 2$. It should also be pointed out that the aforementioned three major propagation mechanisms always come together instead of individually. The combined effect of the three propagation mechanisms will make the signal received behave like a random process.

In practice, the obstacles can be very different from one location to the other. The actual path loss measurements may thus vary greatly from the average. Experimental results

suggest that the path loss measured at any location is random and distributed log-normally around a mean path loss value with a given standard deviation. The random variation is often referred to as log-normal shadowing [72]. Typical values for the path loss exponent and the log-normal shadowing standard deviation have been tabulated for different environments, based on experimental measurements; the path loss exponent typically varies between 2 to 6 [1].

Small scale model

The small scale model, also referred to as small-scale fading, is used to describe the rapid fluctuations in received signal strength over a short time interval or travel distance, so that the large scale effects may be ignored. This phenomenon is caused by two or more images of the transmitted signal which arrive at the receiver through different paths with slightly different time delays, amplitudes, and phases. These waves, called multipath waves, give a resultant signal which can vary significantly in amplitude and phase, depending on the distribution of the intensity and relative propagation time of the waves, as well as the bandwidth of the transmitted signal [3].

In urban areas, giving the height of the mobile antennas is normally well below the height of surrounding structures, significant fading takes place since there is no line-of-sight propagation to the base station. Nevertheless, even when a line-of-sight exists such as in rural areas, multipath still occurs due to reflections from the ground and surrounding structures. The multipath structure, combined with the motion of the receiver, transmitter or surrounding objects in the radio channel, gives rise to variations in received signal amplitude as a function of time. These variations are usually described using standard statistical models, whose validity is proven through numerical experimental measurement campaigns over years. In the following, we introduce the most prominent properties induced

by the small scale models.

Delay spread

The multipath effect lengthens the time required for the signal to reach the receiver, and thus causes serious problems to the signal detection process at the receiver side. To depict the multipath effect, an idealized and classical model is the so-called double negative exponential model, in which the delay separation between paths increases exponentially with path delay, and the path amplitudes also fall off exponentially with delay [73].

In practice, the path delays as well as path amplitudes may show a considerable variability from the classical model. Nevertheless, we can examine the multipath effect by measuring the quantitative properties of a given multipath intensity profile or spectrum $\psi_{\mathrm{De}}(\tau)$, i.e., the average power of the channel output as a function of delay τ. Firstly, the *maximum excess delay* is defined to be the time delay during which multipath energy falls from the maximum to a level xdB below. That is, the maximum excess delay is defined as $\tau_x - \tau_0$, where τ_0 is the propagation delay of the first arrived and τ_x is the maximum delay at which a multipath component is within xdB of the strongest arriving multipath signal. The value of τ_x is sometimes called the excess delay spread, but in all cases it must be specified with a threshold that relates the multipath noise floor to the maximum received multipath component. Secondly, the average delay spread $\bar{\tau}$ is given by

$$\bar{\tau} = \frac{\int_0^{\tau_{max}} \tau \psi_{\mathrm{De}}(\tau) d\tau}{\int_0^{\tau_{max}} \psi_{\mathrm{De}}(\tau) d\tau}, \tag{2.12}$$

where τ_{max} is the maximum path delay, i.e. beyond which it is reasonable to assume $\psi_{\mathrm{De}} = 0$. Finally, the root mean square (RMS) delay spread of the channel, τ_{rms}, is defined

as

$$\tau_{rms} = \sqrt{\frac{\int_0^{\tau_{max}} (\tau - \bar{\tau})^2 \psi_{\mathrm{De}}(\tau) d\tau}{\int_0^{\tau_{max}} \psi_{\mathrm{De}}(\tau) d\tau}}. \tag{2.13}$$

Note that these delays are usually measured relative to the first detectable signal arriving at the receiver, i.e. assuming $\tau_0 = 0$. In addition, they are defined from a single multipath intensity profile, which is the temporal or spatial average of impulse response measurements collected and averaged over a local area. In practice, to determine the statistical range of multipath channel parameters for a mobile communication system, many measurements are conducted in various local areas over a large scale area, e.g., three multipath channels, in terms of low, medium, and high delay spread case, respectively, are defined by IMT-2000 for each environment [72].

Coherence bandwidth

The coherence bandwidth is another important measure of the mobile wireless channels, and is related to the multipath structure of the channel. It is a statistical measure of the range of frequencies over which the channel frequency response can be considered nearly constant (i.e. "flat") [2]. It can be practically defined as the maximum frequency difference for which two signals transmitted through the mobile wireless channels remain strongly correlated.

If the coherence bandwidth B_c is defined as the bandwidth over which the frequency correlation function is above 0.9, then the coherence bandwidth is approximately $B_c \approx 1/(50\tau_{rms})$. If the definition is relaxed so that the frequency correlation function is above 0.5, then the coherence bandwidth is approximately $B_c \approx 1/(5\tau_{rms})$ [1]. It is important to note that the relationship between coherence bandwidth and RMS delay spread remains empirical in nature, as B_c can be defined in different ways.

As the typical value of RMS delay spread is less than $0.05\mu s$ in flat rural environments, $0.2\mu s$ in urban areas, and $2 \sim 3\mu s$ in hilly terrains, respectively, the coherence bandwidth varies from several MHz to a few hundred kHz, depending on the terrain; nonetheless, for a particular multipath channel, its coherence bandwidth according to a given definition is always fixed.

Frequency flat and selective fading

A mobile wireless channel is said to be *frequency flat* fading, if the signal bandwidth is comparable to or smaller than the coherence bandwidth of the channel. On the other hand, if the signal bandwidth is larger than the coherence bandwidth, the channel is said to be *frequency selective* fading.

In frequency flat fading, the mobile wireless channel has a constant gain and linear phase response over a bandwidth which is greater than the signal bandwidth. Therefore, the spectral characteristics of the transmitted signal are preserved at the receiver, and so the flat fading channels are sometimes referred to as *narrowband* channels. However, the strength of the received signal may change over time, due to the fluctuation in the gain of the channel caused by the multipath effects. Seen from the time domain, flat fading occurs when the symbol duration of the transmitted signal is much longer than the delay spread of the flat fading channel and the multipath signal replicas are thus "unresolvable". Hence, the intersymbol interference (ISI) will not happen since the delayed replicas of the current symbol will not overlap with the next symbol. Nevertheless, it should be noted that a flat fading channel can also experience deep fades from time to time due to destructive superpositions of multipaths, and thus it may require 20 to 30dB more transmitting power to compensate the losses due to the deep fades. Therefore, additional processing techniques are necessary to counteract such negative effects.

Different from flat fading, frequency selective fading will cause different attenuations in the received signal at different frequencies, and so the propagation channels are sometimes referred to as *wideband* channels in this case. In the time-domain, the channel impulse response has a multipath delay spread which is greater than the symbol duration of the transmitted signal. Hence, the received signal includes multiple "resolvable" versions of the transmitted symbol waveform that are attenuated and delayed in time whose net effect is to induce ISI at the receiver side.

Doppler spread

Doppler spread is a measure of the spectral broadening caused by the mobility in the channel, and is defined as the range of frequencies over which the received Doppler spectrum is essentially nonzero [2]. When only a pure sinusoidal tone of frequency f_c is transmitted, the received signal spectrum will have components in the range $f_c - f_d$ to $f_c + f_d$, where f_d is the *Doppler shift*, defined as $f_d = (v/\lambda)\cos\theta$, with v, λ, and θ denoting the relative mobile speed, the wavelength of the carrier, and the angle between the mobile moving direction and the LOS from the transmitter to the receiver, respectively. In this simplified scenario, the Doppler spread can be simply expressed by $2f_d$. Hence, it is clear that the Doppler spread depends on the relative speed of the mobile v and the angle θ.

If the distribution of the Doppler spectrum is available, then this knowledge can be taken into consideration when measuring the Doppler spread. Let the Doppler power spectrum be denoted as $\psi_{\mathrm{Do}}(\nu)$, where ν represents the Doppler frequency shift. Thus, the Doppler spread can be computed by the RMS bandwidth of $\psi_{\mathrm{Do}}(\nu)$, defined as

$$\nu_{rms} = \sqrt{\frac{\int_{\mathcal{F}} (\nu - \bar{\nu})^2 \psi_{\mathrm{Do}}(\nu) d\nu}{\int_{\mathcal{F}} \psi_{\mathrm{Do}}(\nu) d\nu}}, \tag{2.14}$$

where $\bar{\nu}$ is the average of the Doppler frequency given by

$$\bar{\nu} = \frac{\int_{\mathcal{F}} \nu \psi_{\mathrm{Do}}(\nu) d\nu}{\int_{\mathcal{F}} \psi_{\mathrm{Do}}(\nu) d\nu}, \tag{2.15}$$

and $\mathcal{F}$ represents the interval in which the spectrum is nonzero.

Concerning the distribution of the Doppler power spectrum $\phi_{\mathrm{Do}}(\nu)$, if one assumes an idealized, uniformly distributed scattering around a terminal with vertical electromagnetic field for the receive and the transmit antennas, then the Doppler spectrum has the classical U-shaped form and is approximated by the Clarke's model [18, 74]. However, in reality, the Doppler spectrum can show considerable variations from this idealized model.

Coherence time

Coherence time is the time domain dual of the Doppler spread and is used to characterize the time-varying nature of the mobile wireless channel. It is a statistical measure of the time duration over which the channel can be considered unchanged. In other words, two signal samples separated by an interval longer than the coherence time can be considered independent to each other. When this occurs, the channel is considered to change significantly during the transmission of the signal, thus introducing a form of distortion in the received signal.

Similar to the relationship between the coherence bandwidth and the delay spread, a unique, standard relationship between the coherence time and Doppler spread does not appear to exist. Nevertheless, the coherence time T_c is typically related to the direct inverse of the Doppler spread and can be approximated as $T_c \approx 1/(2f_m)$, where f_m is the *maximum Doppler shift* given by $f_m = v/\lambda$. If the coherence time is defined as the time over which the time correlation function is above 0.5, then the coherence time is approximately

$T_c \approx 9/(16\pi f_m)$ [75]. A popular rule of thumb is to define the coherence time as the geometric mean of $1/(2f_m)$ and $9/(16\pi f_m)$, i.e. $T_c \approx 0.423/f_m$ [1].

Slow and fast fading

Depending on how rapidly the transmitted signal changes as compared to the rate of change of a mobile wireless channel, a channel may be categorized as either *slow* or *fast* fading [18]. Specifically, the channel is slow fading if the symbol duration is smaller than the coherence time; otherwise, the channel is fast fading. It should be clear that the relative speed of the mobile (or that of objects in the channel) and the symbol duration determine whether a signal undergoes slow or fast fading.

In a slow fading situation, the channel may be assumed to be static over one or several symbol durations, which is also called the *block fading* channel. In the frequency domain, this implies that the Doppler spread of the channel is less than the bandwidth of the baseband signal. In this case, a particular deep fade will affect consecutive symbols, leading to the so-called burst errors.

In a fast fading channel, on the other hand, the channel impulse response changes rapidly within the symbol duration. This causes frequency dispersion due to the Doppler spread, which may lead to severe signal distortion. Therefore, we can conclude that the fading speed is of great importance to determine the suitable estimation and detection strategies in many communication applications.

2.2.2 Delay profile and impulse response

The small scale fading of a mobile wireless channel can be fully characterized by the time-varying impulse response of the channel, where the time variation is due to receiver or transmitter motion in space. Once the impulse response of a particular mobile wireless

channel is available, it can be used to simulate or analyze any type of radio transmission systems through this channel.

In the following, we show that a mobile wireless channel can be modeled as a linear filter with a time-varying impulse response; the specific modeling of the impulse response is discussed in the next section for various channel generation models of interest. Without loss of generality, we consider that the time variation of the channel is due strictly to receiver motion in space. Therefore, at a certain position d of the receiver, the channel can be modeled as a linear time invariant *passband* system, and is expressed by $h_{pb}(d,t)$; this is due to the fact that the received multipath waves are arrived at the receiver with different propagation delays, amplitudes, and phases as its spatial location is changed. Hence, if $x(t)$ represents the transmitted signal, the received signal $y(d,t)$ at the position d can be expressed by

$$y(d,t) = x(t) * h_{pb}(d,t) = \int_{-\infty}^{\infty} x(\lambda) h_{pb}(d, t-\lambda) d\lambda, \tag{2.16}$$

where $*$ denotes the convolution operation. Since the position of a moving receiver is also function of time t, i.e. $d \equiv d(t)$ and the receiver moves along at a constant velocity over a short time (or distance) interval, we can arrive at [1, p.144]

$$y(d(t),t) \equiv y(t) = \int_{-\infty}^{\infty} x(t-\tau) h_{pb}(\tau, t) d\tau = x(t) * h_{pb}(\tau, t), \tag{2.17}$$

where $h_{pb}(\tau, t)$ represents the impulse response of the *passband* time-varying multipath channel, with the variable t representing the time variation due to motion, and τ representing the multipath delay for a fixed value of t.[1]

For passband wireless transmission, it is convenient to represent the signals of interest

[1] The impulse response of a linear time-varying channel $h_{pb}(\tau, t)$ is the channel output at t in response to an impulse applied to the channel at $t - \tau$ [76].

in terms of a low-pass equivalent representation [77]. Let $c(t)$ and $r(t)$ denote the complex envelope of $x(t)$ and $y(t)$, respectively, defined as

$$x(t) = \Re\{c(t)\exp(j2\pi f_c t)\}, \tag{2.18}$$

$$y(t) = \Re\{r(t)\exp(j2\pi f_c t)\}, \tag{2.19}$$

where f_c denotes the carrier frequency. Assuming the multipath channel is bandlimited, we can then rewrite (2.17) into its complex envelope representation [1, 3], i.e.,

$$r(t) = \frac{1}{2}\int_{-\infty}^{\infty} c(t-\tau)h(\tau,t)d\tau = \frac{1}{2}\,c(t) * h(\tau,t), \tag{2.20}$$

in which $h(\tau,t)$ represents the *complex baseband impulse response*, corresponding to the passband channel impulse response $h_{pb}(\tau,t)$. By doing so, we can remove the high frequency variations caused by the carrier, thus facilitating numerical experiments and signal analysis.

For computational and analytical purposes, we can discretize the multipath delay axis τ of the impulse response into the so-called *excess delay bins*, each with equal time delay segments of width $\Delta\tau$. In practice, we usually set the time delay of the first arriving multipath component to zero, i.e. $\tau_0 = 0$ by neglecting the propagation delay between the transmitter and receiver. The time delay of the ith bin is then specified as $\tau_i = i\Delta\tau$, $i = 0, 1, \cdots, L-1$, where L represents the total number of delay bins.

Since the received signal in a multipath channel consists of a series of attenuated, time delayed, phase shifted versions of the transmitted signal, the complex baseband impulse response can be expressed as

$$h(\tau,t) = \sum_{i=0}^{L-1} a_i(\tau,t)\exp[j(2\pi f_c\tau_i + \phi_i(\tau,t))]\delta(\tau-\tau_i), \tag{2.21}$$

where $a_i(\tau, t)$ and $(2\pi f_c \tau_i + \phi_i(\tau, t))$ are the real amplitudes and phase shifts, respectively, of the ith multipath component at time t. If we focus on a short time (or distance) interval, i.e., employing a processing window that is smaller than the channel coherence time, then the channel impulse response can be assumed to be time-invariant, and thus it may be further simplified as

$$h(\tau) = \sum_{i=0}^{L-1} a_i \exp(-j\theta_i)\delta(\tau - \tau_i). \tag{2.22}$$

In the following, we briefly review some well-known channel models which are actually developed on the basis of (2.21) and (2.22) to generate channel coefficients.

2.3 Overview of recent wireless channel simulation models

Modeling the radio channels has been one of the most challenging tasks of mobile radio system design. To exactly describe a practical channel would be very difficult due to the large number of variables involved. Hence, it is typical to resort to a statistical approach, based on measurements made specifically for an intended transmission scenario or spectrum allocation. In the following, we briefly review several wireless channel models that are largely used in simulations and analyses.

2.3.1 3GPP spatial channel model

The 3GPP spatial channel model (SCM), assuming scatterers are separated into $N = 6$ clusters, each with $M = 20$ scatterers, is illustrated in Fig. 2.3 [78, 79]. In this model, we assume that there are N_T transmit antennas at the base station (BS), and N_R receive antennas at the mobile station (MS). The line of sight (LOS) direction is denoted by θ_{BS}, referring to the bore-sight of the BS antenna array. The angle between the bore-sight of the MS antenna array and the LOS is represented by θ_{MS}. The velocity of the MS is assumed

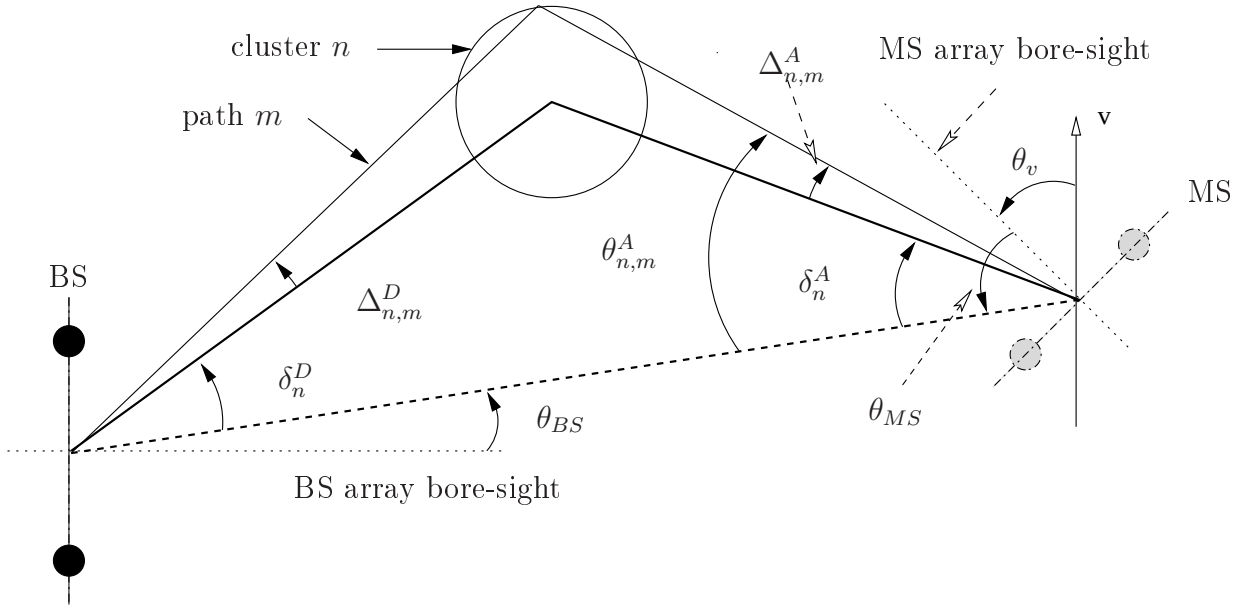


Fig. 2.3 3GPP multiple-input multiple-output spatial channel model.

to be $\vec{v}$ with direction θ_v.

In this model, every scatterer corresponds to one path, and therefore paths associated with the same cluster are assumed to have the same average power and path delay. Considering the downlink transmission scenario, the mean angle of departure (AoD) and mean angle of arrival (AoA) of paths in cluster n are denoted by δ^D_n and δ^A_n, respectively; then the AoD and AoA of the mth path in cluster n are respectively defined by

$$\theta^D_{n,m} \stackrel{\text{def}}{=} \theta_{BS} + \delta^D_n + \Delta^D_{n,m}, \tag{2.23}$$

$$\theta^A_{n,m} \stackrel{\text{def}}{=} \theta_{MS} + \delta^A_n + \Delta^A_{n,m}, \tag{2.24}$$

where $\Delta^D_{n,m}$ and $\Delta^A_{n,m}$ are the angle offsets with respect to the mean AoD and mean AoA, respectively.

Finally, we can express the channel impulse response associated with the qth transmit

antenna and the pth receive antenna as

$$h_{p,q}(\tau, t) = \sum_{n=1}^{N} h_{p,q,n}(t)\delta(\tau - \tau_n(t)), \tag{2.25}$$

where $\tau_n(t)$ denotes the delay spread associated with the nth cluster at some physical time t, and $h_{p,q,n}(t)$ represents the complex channel gain at time t between the pth receive and the qth transmit antenna, associated with the nth multipath, and is defined as

$$\begin{aligned} h_{p,q,n}(t) &= \sqrt{\frac{P_n}{M}} \sum_{m=1}^{M} \sqrt{G_{BS}(\theta^D_{n,m})} \sqrt{G_{MS}(\theta^A_{n,m})} \exp(jkd_q \sin(\theta^D_{n,m})) \\ &\quad \exp(j[kd_p \sin(\theta^A_{n,m}) + \Phi_{n,m}]) \exp(jk\|\vec{v}\| \cos(\theta^A_{n,m} - \theta_v)t), \end{aligned} \tag{2.26}$$

and parameters of the complex channel gain are defined in Table 2.1.

Table 2.1 Parameters of 3GPP-SCM complex channel gain.

P_n	power of the nth path
j	square root of -1
k	wave number $2\pi/\lambda$ (λ is the carrier wavelength in meters)
d_q	distance in meters from BS antenna element q to the reference ($q = 1$) antenna
d_p	distance in meters from MS antenna element p to the reference ($p = 1$) antenna
$\Phi_{n,m}$	phase of the mth subpath of the nth path
$\|\vec{v}\|$	magnitude of the MS velocity vector
θ_v	angle of the MS velocity vector
$G_{BS}(\theta^D_{n,m})$	BS effective antenna array gain
$G_{MS}(\theta^A_{n,m})$	MS effective antenna array gain

2.3.2 COST-207 models

The COST-207 channel models were mainly developed for macrocellular applications. Specifically, COST-207 introduces tap delay line models to represent the propagation channel within the framework of GSM developments [80], i.e., for typical channel characteristics of transmit bandwidths from 10 to 20 MHz, centered around 900MHz.

Based on large amounts of measured data, the typical average power delay profiles of these models were defined for each kind of mobile environment. These can be generally expressed by

$$h(\tau) = \sum_{i=0}^{N-1} a_i \, \delta(\tau - \tau_i), \tag{2.27}$$

where a_i and τ_i represent the complex gain and delay associated with the ith discrete multipath component, respectively, and N denotes the number of multipath components.

Some notable 6-ray multipath delay profiles are shown in Table 2.2, including the typical urban (TU), bad urban (BU), and hilly terrain (HT) scenarios [80]. Due to the simplicity with which it describes a wireless transmission environment, the concept of the COST-207 models has been adopted as a basis for evaluating many other mobile wireless systems, including the IMT2000, UMTS, and mobile DVB-T reception, etc.

2.3.3 Simplified tapped delay line models

The simplified tapped delay line (TDL) models, or transversal filters, are similar to the previous COST-207 models, except that the taps are now symbol-spaced. The transfer function of an mth order TDL model can be described as follows [81]:

$$H(z) = b_0 + b_1 z^{-1} + \cdots + b_m z^{-m}. \tag{2.28}$$

Table 2.2 COST-207 typical urban (TU), bad urban (BU), and hilly terrain (HT) 6-ray power delay profile.

Typical Urban		Bad Urban		Hilly Terrain	
Delay (μs)	Power profile	Delay (μs)	Power profile	Delay (μs)	Power profile
0.0	0.189	0.0	0.164	0.0	0.413
0.2	0.379	0.3	0.293	0.1	0.293
0.5	0.239	1.0	0.147	0.3	0.145
1.6	0.095	1.6	0.094	0.5	0.074
2.3	0.061	5.0	0.185	15.0	0.066
5.0	0.037	6.6	0.117	17.2	0.008

A TDL model includes many adjustable parameters, including choices of the number of taps as well as tap coefficients. The tap coefficients of the TDL models can be either correlated or uncorrelated; e.g., it is possible to incorporate spatial correlation in the TDL models for multiple-antenna transmission systems [82]. Since the correlated TDL models often lead to computational difficulties, the uncorrelated ones are more widely considered in numerical experiments.

Generally speaking, in the TDL models, tap coefficients can be characterized as ZMC-SCG [62] random variables. This can be justified by the central limit theorem on the basis that fading is caused by the superposition of a large number of independent scattered components. In the context of blind channel estimation, the TDL models are widely considered to generate channel coefficients for estimation.

Chapter 3

Survey of recent developments on MIMO-OFDM channel estimation

As we stated earlier in Section 1.1, modern technologies such as diversity techniques, beamforming, and spatial multiplexing can be applied to fully exploit the special structures of the MIMO-OFDM systems when accurate CSI is available. In this chapter, we provide further motivation to devise suitable channel estimation techniques to fully exploit the advantages brought about by MIMO-OFDM. We then present an overview of existing approaches for channel estimation in wideband MIMO-OFDM systems with special emphasis on subspace-based blind approaches. We conclude with a discussion of the limitation of current subspace-based blind estimators.

3.1 Motivating the need for channel estimation

Before justifying the significance of CSI in the above mentioned techniques for MIMO-OFDM, we digress slightly to overview several other detection techniques that do not require CSI at the receiver.

3.1.1 Coherent, noncoherent, and differential detections

In connection with the availability of CSI, there are mainly three types of communication paradigms for the design of communication signals and their detection, namely: coherent, noncoherent, and differential.

The first category, i.e. coherent signal detection, relies on the assumption that an accurate CSI is available at the receiver side. However, this assumption may not always be satisfied, particularly in mobile environments with relatively fast changing channel conditions [83–85]. Hence, the remaining categories, i.e., noncoherent and differential signal design and detection, which do not rely on this assumption, are suggested as alternatives in such scenarios.

The second category, i.e., noncoherent signal detection, has been successfully introduced for several MIMO applications. For example, an efficient and systematic construction of full diversity noncoherent space-frequency codes was presented in [83]. However, its performance is sensitive to the delay spread and the power delay profile. Another example is provided by the design and use of the so-called *training codes* in MIMO systems [86]. Although the CSI is not required for signal detection, one may argue that the code design "abuses" the terminology of *noncoherent*, by allowing part of its codeword to be known to the receiver before transmission for the purpose of estimating the wireless channels. Nonetheless, detection performance based on training codes is generally inferior to that of coherent detection.

The third category, i.e., differential signal detection, has been widely used in practical cellular mobile communication systems, such as IS-54, the 2nd generation standard digital cellular systems in North America. More recently, various differential techniques have been proposed for MIMO systems, including a subcarrier-reconstruction-based approach

[87], a differential space-time-frequency modulation [84], and a multiple-symbol differential detection [85]. But similar to the case of noncoherent detection, a generally worse performance than that of coherent detection is observed; moreover, a quasi-static channel (i.e. slowly-varying) is also required for a successful detection [84, 85, 87, 88].

Thus, with penalty in achievable detection performance, the noncoherent and differential detection techniques are specialized designs suitable for cases in which no accurate CSI is available, a situation often justified on the basis that accurate CSI is difficult to obtain in a relatively fast time-varying channel. However, this argument is questionable since to a great extent, the availability of accurate CSI is mostly a matter of such resources (e.g., computation, bandwidth) the system designer is willing to allocate to the task of channel estimation. In addition, as we have seen above, the noncoherent and differential techniques may be either too restrictive for certain power delay profiles, or impractical in the requirement of quasi-static fading channels. It is therefore questionable whether or not the noncoherent and differential techniques can replace the coherent ones in the aforementioned scenarios.

To summarize the above discussion, we can conclude that if an acceptable estimation performance can be achieved in the relatively fast time-varying channel with affordable cost in resource, then there is no reason to give up the coherent techniques. Besides, it is attractive to enjoy the various already well-designed and proven coherent detection schemes.

3.1.2 The need for channel estimation in coherent detection

The prospect of operating MIMO-OFDM systems close to the Shannon capacity, as enabled by the invention of various capacity-achieving techniques, relies heavily on the availability of advanced channel estimation technique [89].

It can be shown that when the CSI is known to *both* the transmitter and the receiver, the so-called *waterpouring* or *water-filling* algorithm can be employed so that the resulting capacity of the MIMO-OFDM channel is greater than (or equal to) that available when the channel is unknown to the transmitter [62]. However, channel knowledge at the transmitter is normally obtained through feedback from the receiver or based on the channel reciprocity principle in a duplex system. Both approaches may be problematic from a practical perspective. Feedback of channel information may consume excess amount of bandwidth, especially for rapidly time-varying channels. Furthermore, in this latter case, the channel information at the transmitter is likely to be outdated because of transmission and processing delays. Channel reciprocity has a very limited realm of applications: Reciprocity in time is only possible for quasi-static channels, while reciprocity in frequency is only applicable to narrow sub-bands within the coherence bandwidth of the wireless channel.

In this thesis, our interest is in practical broadband MIMO-OFDM transmission over time-varying channels. Therefore, in the following, we mainly focus on the scenarios in which the CSI is only available at the receiver and not at the transmitter.

Some of the most prominent examples of detection technique that justify the need of channel estimation in MIMO systems, include various recently proposed diversity and spatial multiplexing schemes, such as space-time coding [90–92], space-frequency (SF) coding [93–96], space-time-frequency (STF) coding [97–100], and BLAST systems [101, 102]. Generally speaking, diversity techniques provide the receiver with multiple independent *looks* or observations at the same transmitted signal, either in the space, time, frequency, polarization, or a combination of the above domains. As each observation represents a diversity branch, the probability that all branches experience a deep fade concurrently is reduced significantly; thus, diversity techniques improve the link reliability and therefore reduce the

bit error rate. On the contrary, spatial multiplexing techniques offer a linear increase in the signaling rate by exploiting the parallel transmission of different information streams from different antennas, without considering the link reliability or bit error rate.

To better explain the need for channel estimation in these techniques, consider the following simplified mathematical formulation. Let $x_q^m[k]$ denote a symbol transmitted at the qth transmit antenna over the kth subcarrier and the mth OFDM symbol time; this symbol may represent part of the codeword that results from the application of a certain diversity or spatial multiplexing technique. Without loss of generality, the codeword is assumed to be block-based, i.e., it spans over multiple OFDM symbols from $m = 1, 2, \cdots, N_F$ in the time domain, multiple transmit antennas from $q = 1, 2, \cdots, N_T$ in the spatial domain, and multiple subcarriers from $k = 0, 1, \cdots, N_C - 1$ in the frequency domain. Then on the basis of (2.7), we can represent the signal received at the pth receive antenna over the kth tone and the mth OFDM symbol as

$$y_p^m[k] = \sum_{q=1}^{N_T} h_{p,q}^m[k] x_q^m[k] + n_p^m[k], \quad p = 1, 2, \cdots, N_R. \tag{3.1}$$

Thus, the decision rule for maximum likelihood decoding to recover the transmitted codeword is equivalent to minimizing the metric

$$\sum_{k=0}^{N_C-1} \sum_{m=1}^{N_F} \sum_{p=1}^{N_R} |y_p^m[k] - \sum_{q=1}^{N_T} h_{p,q}^m[k] x_q^m[k]|^2 \tag{3.2}$$

over all possible codewords and deciding in favor of the codeword that minimizes the above sum. To this end, we can see that the CSI, i.e. the knowledge of the channel gains $h_{p,q}^m[k]$'s for all possible values of p, q, m, and k, is essential to implement the above decision rule.

Another implement example that justifies the requirement of channel estimation is the

use of various beamforming techniques [6, 16]. In receive beamforming, signals are coherently combined with appropriate steering vectors at the receiver so as to enhance the antenna array response in a preferred direction. The computation of the steering vectors is normally performed by applying the singular value decomposition on a matrix constructed from the CSI, and hence the latter is crucial to achieve the computation.

Having justified the importance of channel estimation in MIMO-OFDM systems, in the following, we review and survey recent advances on related channel estimation techniques. Specifically, we focus on the channel estimation and tracking of the MIMO wireless channels in a wideband OFDM context.

3.2 Overview of channel estimation for wideband MIMO-OFDM

The good performance of MIMO-OFDM depends on its ability to deal with the frequency and time selective nature of the wideband propagation channels [103], which in turn heavily relies on the availability of the corresponding CSI, as obtained by means of a channel estimation technique.

Generally speaking, channel estimation can be performed in two different ways, that is: *pilot-based* and *blind* estimation [89]. In pilot-based channel estimation, known signals or symbols are transmitted to assist the receiver in determining the required CSI. On the other hand, a blind channel estimation method determines the CSI without the aid of known symbols, and thus can achieve a higher bandwidth efficiency; however, it generally comes with a penalty of slower convergence speed, reduced estimation accuracy, and higher implementation complexity [104].

For the reasons stated above, one might erroneously conclude that there are limited uses for a blind estimation technique in practice, since the dynamic requirement associated

to the estimation of time-varying mobile channels further enhances the demands in terms of convergence speed and implementation complexity. However, in many emerging wireless standards, the transmission of overhead information such as these pilot symbols poses a major burden to the data transmission. In fact, the merits of the high bandwidth efficiency associated with the MIMO-OFDM systems can be greatly offset by overhead [105]. For example, it has been shown that to maximize the overall transmission rate in a BLAST-MIMO system, half of the available interval is used for training [102].

Hence, if a blind channel estimation technique can be devised to possess fast convergence speed with acceptable performance and a reasonable complexity, then it can replace pilot-based channel estimators, and be embraced by future wireless standards for the purpose of meeting the demand of high transmission rate. For completeness, both pilot-based and blind channel estimation techniques are further reviewed below.

3.2.1 Pilot-based channel estimation

Pilot-based channel estimation can be broadly categorized as employing either continuous-time sounding signals, i.e. pilot tones [89, 106–110], or a sequence of known symbols inserted among data symbols, i.e. pilot or training symbols [103, 105, 111–113].

The first category of technique, i.e. those based on pilot tones, employ a continuous-time waveform or tone to calibrate the multipath-induced effects in the receiver side [109]. For single-carrier systems, one generally needs to suppress the data signal power around the desired pilot frequencies [114]. In one of the earlier pilot tone techniques, the so-called tone calibrated technique (TCT), double sideband modulation is transmitted alongside the pilot [106]. In addition, to make room for the pilot in the presence of Doppler shift, the sidebands of lower frequencies relative to tone frequency must be shaped so as to have zero response in the neighborhood of the DC signal. Other techniques, such as transparent tone-

in-band (TTIB) [107], similarly *notch out* a spectral gap in the center of the data spectrum for the placement of the pilot. An alternative possibility is the dual-pilot tone calibration technique (DPTCT), in which a pair of pilots are symmetrically placed near the edges of the channel and outside the data spectrum [108]. For multi-carrier systems such as in OFDM, not only the placement of the pilot tones but also the pilot sequences must be optimized to obtain the minimal MSE of the least squares (LS) channel estimate. It has been shown that the optimal pilot tones should be equipowered, equispaced, and phase-shift orthogonal [89]. Although pilot-tone techniques lead to robust and simple receiver structures, a large fraction of the transmitted power is wasted in these tones, and thus there has been a trend away from these techniques in recent wireless standards.

In the second category of pilot-based channel estimation technique, i.e. those based on pilot symbols, the receiver first extracts the channel transfer function at those times and frequencies at which pilot symbols have been inserted. Then, the missing values of the transfer function between the positions of pilot symbols can be interpolated by means of filtering [89, 103, 115]. Note that performance of channel estimation is highly dependent on the distribution of the pilot symbols in relation to the coherence time and coherence bandwidth of the wireless channel. Thus, one needs to judiciously place the pilot symbols not only to gain better estimation performance but also to reduce the resources allocated to training. This is particularly important in a fast fading channel where pilots have to be inserted frequently in order to track channel variations [113].

Generally, to obtain acceptable channel estimation for SISO-OFDM systems, the first and the last subcarriers are modulated with pilot symbols. Then the other pilot symbols are inserted at every n_f subcarriers in the frequency direction and every n_t OFDM symbols in the time domain. The insertion periods n_f and n_t must satisfy the following conditions,

which result from the sampling theorem [111, 116]:

$$n_f < \frac{1}{\tau_{max}\Delta f}, \quad n_t < \frac{1}{2f_m T}, \tag{3.3}$$

where Δf denotes the subcarrier spacing, τ_{max} is the maximum path delay, T is the OFDM symbol duration, and f_m is the maximum Doppler shift. The above conditions can be reformulated into a more intuitive way, in terms of the coherence bandwidth and the coherence time, respectively, as given by

$$n_f < B_c/\Delta f, \quad n_t < T_c/T, \tag{3.4}$$

where $B_c \approx 1/\tau_{max}$ represents the coherence bandwidth and $T_c \approx 1/(2f_m)$ represents the coherence time. The bandwidth efficiency is reduced by a factor of approximately $1/(n_t n_f)$ due to the insertion of pilots as above. However, to achieve a reasonable noise reduction by filtering in the interpolation step, the density of the pilot symbols along both the time and frequency axes should be larger than twice the minimum density mentioned from (3.4), so that the channel transfer function is over-sampled. If multiple antennas are employed, then the situation is getting worse since additional pilot symbols are needed to estimate the channel transfer function between each combination of transmit and receive antenna.

More recently, significant efforts have been devoted to situations in which not all subcarriers are activated [117]. In these cases, the locations of the pilot symbols cannot be equally spaced due to the so-called virtual carriers, i.e. subcarriers that are set to zero with no any information being transmitted. Therefore, the optimal solution for the case of all subcarriers are activated cannot be applied directly. One possible solution was obtained by an iterative method to achieve minimum mean square error of the least squares estimate

[118]. However, in a time-invariant channel scenario (i.e., within one OFDM symbol period in this case), the minimum number of pilots with this approach is still nearly twice as large as the number of unknown channel coefficients. A more bandwidth efficient method was proposed in [113], such that the minimum number of pilots required is only half of the above method. However, no less than $N_T L$ pilots are still required for a time-invariant broadband wireless channel with channel order L.

With respect to a receiver with ideal CSI, the BER curves typically degrade by no more than 1.5dB in SNR [111, 112], if a carefully designed pilot-based estimator is employed. In the following, we discuss the possibility of achieving similar or even better system performance by using a blind-based channel estimator, i.e. without employing specific pilot symbols.

3.2.2 Blind channel estimation

Without the assistance and the expense of pilot symbols, blind-based channel estimation presents a bandwidth efficient way to acquire the CSI needed for signal detection. Existing blind estimation methods can be broadly categorized as deterministic or statistical.

Deterministic methods, the first category of blind approaches, are in general favored when the input statistics are unknown, or there may not be sufficient time samples to obtain suitable estimate of the required statistics. To date, several interesting deterministic methods have been developed by, e.g., employing the maximum likelihood approach [32, 33], exploiting null guard intervals [34, 35], exploiting zeros of the channel impulse response [36], or by using fractional sampling as well as interpolation [37]. However, most of them are for SISO or single carrier exclusively, and hence are less applicable in the context of MIMO-OFDM systems.

In the second category, blind approaches can be further sub-divided into two classes

as follows: algorithms using higher order statistics (HOS) of the received signal, and algorithms using only second order statistics (SOS). In general, HOS-based algorithms are usually not considered for moderate to rapidly time-varying wireless channels, since a large number of received samples is needed before the HOS estimates are reliable. Instead, SOS-based algorithms can potentially achieve superior estimation performance for a given time averaging interval than approaches using higher order statistics (HOS) [27–31, 119].

Amid SOS-based blind approaches, subspace-based estimation is attractive since channel estimates can often be obtained in a simple form by optimizing a quadratic cost function [38]. Without employing any precoding at the transmitter, a noise subspace-based method is proposed for OFDM systems by utilizing the redundancy introduced by the cyclic prefix (CP) [39, 40], and it is further extended for MIMO-OFDM systems in [41]. Virtual carriers (VC) are subcarriers that are set to zero without any information being transmitted. The presence of VC provides another useful resource that can be used for channel estimation. Such a scheme is proposed for OFDM systems [42], and it is further extended for MIMO-OFDM systems in [43, 44].

On the basis of the above considerations, we conclude that to meet the demands of future wireless standards based on MIMO-OFDM, subspace-based blind channel estimation offers an attractive solution in terms of both estimation performance and the required number of time samples (i.e. OFDM symbols) to attain the performance.

3.3 Generalized subspace-based blind estimation

In the following, we present the concept of subspace-based blind estimation in mathematical term, and explore the problem of interests in this thesis, which results from current limitations on the use of these estimators. Then, we overview some examples of subspace-

based blind channel estimation methods that have been recently proposed in the signal processing and communications literature.

3.3.1 The mathematical basis of subspace-based blind estimation

Blind channel estimation, which is based on the exploitation of the structure in the channel and properties of the input, relies only on multiple observations of the channel output $\mathbf{y}$ to identify the unknown channel parameters [28]. Specifically, the subspace-based blind estimation problem, which uses the second order statistics of the observation, can be formulated as follows. Let $\mathbf{x} \in \mathcal{U} \subset \mathbb{C}^{d_s \times 1}$ be the input vector to a linear channel as shown in Fig. 3.1.[1] The output of the channel $\mathbf{s} \in \mathbb{C}^{d_c \times 1}$ can then be represented by

$$\mathbf{x} \mapsto \mathbf{s} = \mathcal{H}\mathbf{x}, \tag{3.5}$$

where matrix $\mathcal{H} \in \mathbb{C}^{d_c \times d_s}$ $(d_c > d_s)$ represents a linear transformation [120]. The *image* of $\mathcal{U}$, denoted by $\mathcal{V}$, is the set $\{\mathbf{s} = \mathcal{H}\mathbf{x} \text{ for some } \mathbf{x} \in \mathcal{U}\}$. Assuming that $\dim[\mathcal{U}] = d_s$ and matrix $\mathcal{H}$ is full rank, we have $\dim[\mathcal{V}] = d_s$. The observation $\mathbf{y} \in \mathcal{W} \subseteq \mathbb{C}^{d_c \times 1}$, which is the sum of channel output $\mathbf{s}$ and noise $\mathbf{n}$, can be written as $\mathbf{y} = \mathbf{s} + \mathbf{n}$.

Subspace methods rely on the assumptions that the observation space, i.e. $\mathcal{W}$, has $\dim[\mathcal{W}] = d_c > \dim[\mathcal{U}]$. Only a second order statistical characterization of the input vector $\mathbf{x}$ and noise vector $\mathbf{n}$ is generally involved. Without loss of generality, it is convenient to assume that $\mathbf{x}$ and $\mathbf{n}$ have zero mean, i.e., $E[\mathbf{x}] = \mathbf{0}$ and $E[\mathbf{n}] = \mathbf{0}$, where $E[\cdot]$ denotes statistical expectation. We denote by $\mathbf{R}_\mathbf{x} = E[\mathbf{x}\mathbf{x}^H]$ and $\mathbf{R}_\mathbf{n} = E[\mathbf{n}\mathbf{n}^H]$ the correlation matrix of $\mathbf{x}$ and $\mathbf{n}$, respectively.

From the structure of the linear signal model in Fig. 3.1, it follows that $\mathbf{R}_\mathbf{y} = E[\mathbf{y}\mathbf{y}^H]$,

[1] In practice, due to the finite alphabet property, the set $\mathcal{U}$ that encompasses all the possible input vectors $\mathbf{x}$ is a proper subset of $\mathbb{C}^{d_s \times 1}$.

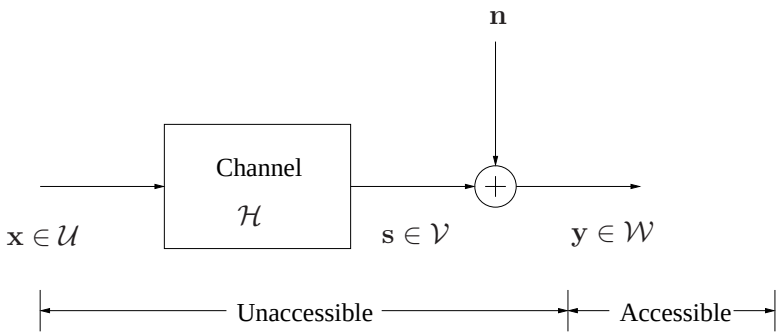


Fig. 3.1 A generic blind channel estimation problem.

the correlation matrix of the observation, can be expressed as

$$\mathbf{R}_{\mathbf{y}} = \mathcal{H}\mathbf{R}_{\mathbf{x}}\mathcal{H}^H + \mathbf{R}_{\mathbf{n}}. \tag{3.6}$$

Thus, provided that the noise correlation matrix is a sufficiently regular structure (e.g., $\mathbf{R}_{\mathbf{n}} = \sigma_n^2\,\mathbf{I}$), it appears that relevant information regarding the channel parameters can be obtained from the eigenvectors of $\mathbf{R}_{\mathbf{y}}$.

Consider a practical situation in which several observations (or realizations) of random vector $\mathbf{y}$ are available, and let $\mathbf{y}_{(j)}$ denote the jth observation at some physical time t_j. We can estimate $\mathbf{R}_{\mathbf{y}}$ as

$$\hat{\mathbf{R}}_{\mathbf{y}} = \frac{1}{T_{av}} \sum_{j=1}^{T_{av}} \mathbf{y}_{(j)}\mathbf{y}_{(j)}^H. \tag{3.7}$$

Under the stationary condition, $\hat{\mathbf{R}}_{\mathbf{y}}$ (and $\hat{\mathbf{R}}_{\mathbf{n}}$) converges to $\mathbf{R}_{\mathbf{y}} = E[\mathbf{y}\mathbf{y}^H]$ (and $\mathbf{R}_{\mathbf{n}}$) in the mean square sense as the time averaging interval $T_{av} \to \infty$ [40]. Note that we can write $\mathbf{R}_{\mathbf{y}} = \hat{\mathbf{R}}_{\mathbf{y}} + \Delta\mathbf{R}_{\mathbf{y}}$, where $\Delta\mathbf{R}_{\mathbf{y}}$ denotes the difference between the *true* and the sampled correlation matrix.

By applying an eigen-decomposition to $\mathbf{R}_{\mathbf{y}}$, we can obtain the *signal* and *noise subspaces* from the span of the eigenvectors which correspond to its d_s largest eigenvalues and $d_c - d_s$

smallest eigenvalues, respectively. Due to the inevitable noise perturbation and insufficient time averaging, minimizing the cost function $C(\mathcal{H}) = \mathrm{tr}\,[\mathcal{L}(\mathcal{H})^T\hat{\mathbf{\Psi}}\mathcal{L}^*(\mathcal{H})]$ subject to a quadratic constraint, i.e., $\mathcal{L}(\mathcal{H})\mathcal{L}^H(\mathcal{H}) = \mathbf{I}$, is usually employed to obtain the channel coefficients, where the matrix $\hat{\mathbf{\Psi}}$ is constructed from the eigenvectors of the perturbed noise subspace and the matrix $\mathcal{L}(\mathcal{H})$ is obtained by re-arranging the non-redundant elements of $\mathcal{H}$, depending on the dimensionality as well as the formulation of the matrix $\hat{\mathbf{\Psi}}$. The details of this step are application specific, i.e. the construction of $\mathcal{L}(\mathcal{H})$ and $\hat{\mathbf{\Psi}}$.

Let $\mathcal{H}_o$ be the *optimal* such that $C(\mathcal{H}_o) \leq C(\mathcal{H})$ under the chosen constraint. We have

$$C(\mathcal{H}_o) = \min_{\mathcal{L}(\mathcal{H})\mathcal{L}(\mathcal{H})^H = \mathbf{I}} \left\{ \mathrm{tr}\,[\mathcal{L}(\mathcal{H})^T\hat{\mathbf{\Psi}}\mathcal{L}^*(\mathcal{H})] \right\} = \sum_{j=1}^{r} \gamma_j(\hat{\mathbf{\Psi}}), \tag{3.8}$$

where $\gamma_j(\hat{\mathbf{\Psi}})$ represents the jth smallest eigenvalue of $\hat{\mathbf{\Psi}}$ and r depends on specific algorithms. To discuss the asymptotic behaviors of $C(\mathcal{H}_o)$ as $\|\Delta\mathbf{R}_\mathbf{y}\| \to 0$ and $\|\mathbf{R}_\mathbf{n}\| \to 0$, let $\gamma_j(\hat{\mathbf{\Psi}})$ and $\gamma_j(\mathbf{\Psi})$ be the jth smallest eigenvalue of $\hat{\mathbf{\Psi}} = \mathbf{\Psi} - \Delta\mathbf{\Psi}$ and $\mathbf{\Psi}$, respectively, where $\mathbf{\Psi}$ is constructed (in a similar way as $\hat{\mathbf{\Psi}}$) from the error free noise eigenvectors of $\mathbf{R}_\mathbf{y}$. In this case, $\mathbf{\Psi}$ is perturbed by the amount $\Delta\mathbf{\Psi}$ due to an insufficient time averaging interval. Since $\hat{\mathbf{\Psi}}$ and $\mathbf{\Psi}$ are Hermitian in the subspace-based problems, we have [121]

$$|\gamma_j(\hat{\mathbf{\Psi}}) - \gamma_j(\mathbf{\Psi})| \leq \|\Delta\mathbf{\Psi}\|_2. \tag{3.9}$$

First consider the case $\|\Delta\mathbf{R}_\mathbf{y}\| > 0$ and $\|\mathbf{R}_\mathbf{n}\| = 0$, which corresponds to a noise-free situation with finite time averaging; that is, $\mathbf{\Psi}$ is perturbed by the amount $\Delta\mathbf{\Psi}$ due to an insufficient time averaging interval. Here, since $\gamma_j(\mathbf{\Psi}) = 0$, $j = 1, 2, \cdots, r$, we have $C(\mathcal{H}_o) = \sum_{j=1}^{r} \gamma_j(\hat{\mathbf{\Psi}}) = \sum_{j=1}^{r} |\gamma_j(\hat{\mathbf{\Psi}}) - \gamma_j(\mathbf{\Psi})| \leq r\|\Delta\mathbf{\Psi}\|_2$. While the exact relationship between $\Delta\mathbf{\Psi}$ and $\Delta\mathbf{R}_\mathbf{y}$ depends on the perturbation of the matrix $\hat{\mathbf{\Psi}}$ in a given problem,

we can generally show for many cases of interest that $\|\Delta\mathbf{\Psi}\|_2 \leq \alpha\|\Delta\mathbf{R_y}\|_2$, where $\alpha > 0$ is a scalar. Therefore, $C(\mathcal{H}_o) \leq r\alpha\|\Delta\mathbf{R_y}\|_2$ and the estimation performance is determined by the closeness of the estimated correlation matrix to the true one.

Alternatively, consider the case $\|\Delta\mathbf{R_y}\| = 0$ and $\|\mathbf{R_n}\| > 0$, corresponding to a noisy situation but with infinite time averaging ($T_{av} \to \infty$); that is, $\mathbf{\Psi}$ is perturbed by the amount $\Delta\mathbf{\Psi}$ due to the noise. Hence, we can also arrive at $C(\mathcal{H}_o) \leq \alpha^{'}\|\mathbf{R_n}\|_2$, where $\alpha^{'} > 0$ is a scalar. In the specific case of $\mathbf{R_n} = \sigma_n^2\,\mathbf{I}$, we can arrive at $C(\mathcal{H}_o) \leq \alpha^{'}\|\hat{\mathbf{R}}_\mathbf{n}\|_2 = \alpha^{'}\sigma_n^2$. Therefore, in this case, when there is sufficient time averaging, the estimation performance is determined by the noise variance. In both cases above, we conclude that $C(\mathcal{H}_o) \to 0$ as $\|\Delta\mathbf{R_y}\| \to 0$ and $\|\mathbf{R_n}\| \to 0$.

In general, the performance of the channel estimator obtained from (3.8) largely depends on our ability to discriminate between the d_s largest eigenvalues of $\hat{\mathbf{R}}_\mathbf{y}$ and its $d_c - d_s$ smallest (noise) eigenvalues. Since wireless channels are in general non-stationary, we cannot choose an arbitrarily large T_{av} to estimate the correlation matrix. Nonetheless, this can be ensured by requiring that $T_{av} \geq d_c$ in a practical scenario, i.e., $\|\Delta\mathbf{R_y}\| > 0$ and $\|\mathbf{R_n}\| > 0$.

3.3.2 Overview of recent subspace-based blind approaches

To date, several interesting subspace-based blind channel estimation methods have been proposed. Concerning the linear transformation $\mathcal{H}$ in (3.5), it not only characterizes the input-output relationship of a subspace-based blind problem, but also plays an important role in the performance of the corresponding subspace-based blind channel estimator. Indeed, the number of rows in $\mathcal{H}$, i.e. d_c, is equal to the dimension of the correlation matrix $\hat{\mathbf{R}}_\mathbf{y}$, which is directly related to the minimum required time samples to estimate the correlation matrix. The linear transformation matrix $\mathcal{H}$ of some notable subspace-based blind channel estimators are briefly reviewed below.

a) SISO/SIMO systems [38]

This seminal work proposed a subspace-based blind estimation approach either by means of oversampling the received signal compared to the information data rate, or by using several sensors. Although several other approaches using SOS had been proposed beforehand, this approach is computationally more efficient, and can achieve a comparably or even better statistical performance.

In [38], two possible scenarios are considered as explained below.

(1) Oversampling a single sensor: This scenario considers a SISO system, with an underlying continuous-time channel impulse response $h(t)$ assumed to have finite support. The channel output vector is obtained by oversampling the continuous-time output at the rate M/T, where T is the symbol period and M is the oversampling factor.

Assuming a temporal observation window of length $N_F T$, the sampled output of the SISO channel can be arranged in a vector $\mathbf{s}$ such that (3.5) is satisfied with a suitable definition of $\mathcal{H}$. Specifically, define the polyphase impulse response by

$$h_l^i \stackrel{\text{def}}{=} h(t_0 - i\,(T/M) - lT); \;\; 0 \le i \le M-1, \;\; 0 \le l \le L-1, \tag{3.10}$$

where L denotes the channel order and t_0 is an appropriate reference time. The corresponding linear transformation of the blind estimation problem can thus be characterized as

$$\mathcal{H} = \left[\mathcal{H}^{(0)^T}\mathcal{H}^{(1)^T}\cdots\mathcal{H}^{(M-1)^T}\right]^T \in \mathbb{C}^{MN_F\times(N_F+L-1)}, \tag{3.11}$$

where $\mathcal{H}^{(i)}$ is defined as

$$\mathcal{H}^{(i)} \stackrel{\text{def}}{=} \begin{bmatrix} h^i_{L-1} & \cdots & h^i_0 & 0 & \cdots & \cdots & 0 \\ 0 & h^i_{L-1} & \cdots & h^i_0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h^i_{L-1} & \cdots & h^i_0 & 0 \\ 0 & \cdots & \cdots & 0 & h^i_{L-1} & \cdots & h^i_0 \end{bmatrix} \in \mathbb{C}^{N_F \times (N_F+L-1)}. \tag{3.12}$$

(2) Multiple sensors: This scenario considers a SIMO system where M now denotes the number of sensors or receiving antennas and $h^i(t)$, $i = 1, 2, \cdots, M$ are the continuous-time impulse responses of the propagation channel between the input and the ith sensor output. Then the corresponding linear transformation also follows (3.11)-(3.12), except that $h^i_l \stackrel{\text{def}}{=} h^i(t_0 - lT)$, $0 \le l \le L-1$, meaning that the ith sensor receives a signal that has traveled through the ith propagation channel.

b) OFDM systems with cyclic prefix [39]

It can be shown that cyclostationarity in the received signal allows the receiver to blindly identify the channel impulse response using only second order statistics [29]. In particular, by exploiting the cyclostationarity embedded at the transmitter due to the insertion of a cyclic prefix in each symbol, a subspace-based blind channel identification was proposed for SISO-OFDM systems in [39].

In the notation of [39], the discrete-time impulse response of the SISO channel is denoted as h_l, $0 \le l \le L-1$. The output vector $\mathbf{s}$ consists of the sequence of time samples from N_F consecutive OFDM symbols, with each OFDM symbol of size $N_C + N_{cp}$, where N_C is the IFFT size and N_{cp} is the length of cyclic prefix (See Section 2.1 for details). Then the

corresponding linear transformation of the blind estimation problem can be written as

$$\mathcal{H} = \begin{bmatrix} h_{L-1} & \cdots & h_0 & 0 & \cdots & \cdots & 0 \\ 0 & h_{L-1} & \cdots & h_0 & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & h_{L-1} & \cdots & h_0 & 0 \\ 0 & \cdots & \cdots & 0 & h_{L-1} & \cdots & h_0 \end{bmatrix}, \tag{3.13}$$

which is also a Toeplitz matrix of size $(N_F(N_C + N_{cp}) - L + 1) \times N_F(N_C + N_{cp})$.

c) OFDM systems without cyclic prefix [122]

Motivated by the multichannel signal model in single-carrier systems [38], another notable blind approach for channel estimation was proposed in the context of OFDM systems [122]. The method distinguishes itself from [39] mainly by eliminating the use of the CP for channel estimation. Two possible scenarios are considered as follows:

(1) Oversampling a single sensor: In this case, the continuous-time channel $h(t)$ is assumed to be of finite support. If M and L denote the oversampling factor in the OFDM system and the channel order, respectively, then for some reference time t_0, the discrete-time impulse response of interests are given as in [39] by

$$h_l^i \stackrel{\text{def}}{=} h(t_0 - i\,(T/M) - lT); \;\; 0 \le i \le M-1, \;\; 0 \le l \le L-1, \tag{3.14}$$

where T represents the duration of the $N_C + N_{cp}$ individual time samples comprising each OFDM symbol. The corresponding linear transformation of this blind estimation problem

can thus be characterized as

$$\mathcal{H} = \left[\mathcal{H}^{(0)^T}\mathcal{H}^{(1)^T}\cdots\mathcal{H}^{(M-1)^T}\right]^T \in \mathbb{C}^{M(N_C-L+1)\times N_C}, \tag{3.15}$$

where $\mathcal{H}^{(i)}$ is defined in (3.12) but now has dimension $(N_C - L + 1) \times N_C$. Note that in this approach only the non-ISI corrupted time samples are used in the construction of $\mathbf{s}$, which explains the different size of $\mathcal{H}^{(i)}$.

(2) Multiple sensors: If M now denotes the number of sensors in a SIMO-OFDM system, then by collecting only non-ISI corrupted OFDM time samples at each sensor and stacking them, we can construct another multichannel signal model. The corresponding linear transformation also follows the above case with $h_l^i \overset{\text{def}}{=} h^i(t_0 - lT)$, $0 \le l \le L-1$.

d) MIMO-OFDM systems [44]

With the growing popularity of MIMO-OFDM systems, an eminent subspace-based approach was proposed in [44]. Two possible scenarios are addressed as follows:

(1) MIMO-OFDM systems with $N_R \ge N_T$: In this case, let the temporal window of observations be N_F and $h_{p,q}(t)$ denotes the continuous-time channel impulse response between the qth transmit antenna and the pth receive antenna. The corresponding linear transformation of the blind estimation problem can be formulated as

$$\mathcal{H} = \begin{bmatrix} \mathbf{h}(L-1) & \cdots & \mathbf{h}(0) & 0 & \cdots & \cdots & 0 \\ 0 & \mathbf{h}(L-1) & \cdots & \mathbf{h}(0) & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & 0 & \mathbf{h}(L-1) & \cdots & \mathbf{h}(0) & 0 \\ 0 & \cdots & \cdots & 0 & \mathbf{h}(L-1) & \cdots & \mathbf{h}(0) \end{bmatrix}, \tag{3.16}$$

which is a block-Toeplitz matrix of size $(N_F Q - L + 1)N_R \times N_F Q N_T$, with $Q \stackrel{\text{def}}{=} N_C + N_{cp}$,

$$\mathbf{h}(l) \stackrel{\text{def}}{=} \begin{bmatrix} h_{11}(l) & h_{12}(l) & \cdots & h_{1N_T}(l) \\ h_{21}(l) & h_{22}(l) & \cdots & h_{2N_T}(l) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_R1}(l) & h_{N_R2}(l) & \cdots & h_{N_RN_T}(l) \end{bmatrix} \in \mathbb{C}^{N_R \times N_T}, \tag{3.17}$$

and $h_{p,q}(l) \stackrel{\text{def}}{=} h_{p,q}(t_0 - lT)$, $0 \le l \le L-1$, where T denotes the OFDM sampling period.

(2) MIMO-OFDM systems with $N_T > N_R$

In this case, an oversampling factor M is applied at the receiver with $M \ge \lceil N_T/N_R \rceil$. With $h_{p,q}^{(i)}(l) \stackrel{\text{def}}{=} h_{p,q}(t_0 - i\,(T/M) - lT)$, $0 \le i \le M-1$, $0 \le l \le L-1$, the lth lag of the oversampled-MIMO channel is represented as

$$\tilde{\mathbf{h}}(l) = \begin{bmatrix} h_{1,1}^{(0)}(l) & h_{1,2}^{(0)}(l) & \cdots & h_{1,N_T}^{(0)}(l) \\ \vdots & \vdots & \ddots & \vdots \\ h_{1,1}^{(M-1)}(l) & h_{1,2}^{(M-1)}(l) & \cdots & h_{1,N_T}^{(M-1)}(l) \\ h_{2,1}^{(0)}(l) & h_{2,2}^{(0)}(l) & \cdots & h_{2,N_T}^{(0)}(l) \\ \vdots & \vdots & \ddots & \vdots \\ h_{2,1}^{(M-1)}(l) & h_{2,2}^{(M-1)}(l) & \cdots & h_{2,N_T}^{(M-1)}(l) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_R,1}^{(0)}(l) & h_{N_R,2}^{(0)}(l) & \cdots & h_{N_R,N_T}^{(0)}(l) \\ \vdots & \vdots & \ddots & \vdots \\ h_{N_R,1}^{(M-1)}(l) & h_{N_R,2}^{(M-1)}(l) & \cdots & h_{N_R,N_T}^{(M-1)}(l) \end{bmatrix}. \tag{3.18}$$

The corresponding linear transformation matrix is then characterized as

$$\mathcal{H} = \begin{bmatrix} \tilde{\mathbf{h}}(L-1) & \cdots & \tilde{\mathbf{h}}(0) & 0 & \cdots & \cdots & 0 \\ 0 & \tilde{\mathbf{h}}(L-1) & \cdots & \tilde{\mathbf{h}}(0) & 0 & \cdots & 0 \\ \vdots & & & & & & \vdots \\ 0 & \cdots & 0 & \tilde{\mathbf{h}}(L-1) & \cdots & \tilde{\mathbf{h}}(0) & 0 \\ 0 & \cdots & \cdots & 0 & \tilde{\mathbf{h}}(L-1) & \cdots & \tilde{\mathbf{h}}(0) \end{bmatrix}, \tag{3.19}$$

which is of size $M(N_F Q - L + 1)N_R \times N_F Q N_T$. Here, oversampling is introduced to ensure that $\mathcal{H}$ is a tall matrix (i.e. full column rank), a condition needed in the application of a subspace approach.

3.3.3 Summary of the notable subspace-based blind approaches

For future reference, we briefly summarize the above representative subspace-based estimators [38, 39, 44, 122] in Table 3.1, in terms of the dimension of the signal subspace d_s, the dimension of the noise subspace $d_c - d_s$, the size of the linear transformation matrix $\mathcal{H}$, and the dimension of the corresponding correlation matrix $\hat{\mathbf{R}}_{\mathbf{y}}$, respectively.[2]

Table 3.1 Summary of dimensionality for some notable subspace-based blind channel estimators.

Estimator	d_s	$d_c - d_s$	size of $\mathcal{H}$	dimension of $\hat{\mathbf{R}}_{\mathbf{y}}$
SISO (oversampling) [38]	$N_F + L - 1$	$MN_F - N_F - L + 1$	$MN_F \times (N_F + L - 1)$	MN_F
SIMO (multiple sensors) [38]	$N_F + L - 1$	$MN_F - N_F - L + 1$	$MN_F \times (N_F + L - 1)$	MN_F
SISO-OFDM (with CP) [39]	$N_F N_C$	$N_F N_{cp} - L + 1$	$(N_F(N_C + N_{cp}) - L + 1) \times N_F(N_C + N_{cp})$	$N_F(N_C + N_{cp}) - L + 1$
SISO-OFDM (without CP) [122]	N_C	$M(N_C - L + 1) - N_C$	$M(N_C - L + 1) \times N_C$	$M(N_C - L + 1)$
SIMO-OFDM (without CP) [122]	N_C	$N_R(N_C - L + 1) - N_C$	$N_R(N_C - L + 1) \times N_C$	$N_R(N_C - L + 1)$
MIMO-OFDM ($N_R \geq N_T$) [44]	$N_F D N_T$	$(N_F Q - L + 1)N_R - N_F D N_T$	$(N_F Q - L + 1)N_R \times N_F Q N_T$	$(N_F Q - L + 1)N_R$
MIMO-OFDM ($N_R < N_T$) [44]	$N_F D N_T$	$M(N_F Q - L + 1)N_R - N_F D N_T$	$M(N_F Q - L + 1)N_R \times N_F Q N_T$	$M(N_F Q - L + 1)N_R$

[2] The parameter D in [44] denotes the number of useful subcarriers in the OFDM systems.

3.4 Limitations of current subspace-based blind estimators

In practice, to obtain the eigenvectors corresponding to the noise subspace in a subspace-based problem, the correlation matrix $\mathbf{R}_{\mathbf{y}}$ must be estimated through time averaging over multiple received samples $\mathbf{y}_{(j)}$, $j = 1, 2, \cdots, T_{av}$. To this end, the unknown channel must remain time-invariant throughout the averaging process, which may pose a serious problem in practical applications.

Clearly, to obtain a sampled correlation matrix with full rank, the number of time samples required must be no less than the dimension of the correlation matrix. Similar considerations can be found in the literature, specifically in the context of minimum variance beamformers [123], large dimensional random matrices [124], and the persistence of excitation assumption [25, 40]. Hence, when we consider the time invariance requirement of a practical MIMO-OFDM system with a large number of OFDM subcarriers, e.g., $N_C = 128$ or more, the traditional subspace-based methods require extremely large number of time samples for obtaining a good time-averaged correlation matrix, making them impractical. Indeed, as can simply observe from the last column of Table 3.1, the dimension of the correlation matrix in any one of the method in [39, 44, 122] is directly proportional to N_C, i.e. the number of OFDM subcarriers.

More recently, variants of the statistics-based blind channel estimation methods have been proposed, e.g., by inserting zero-padding instead of CP for each OFDM block [45], or by introducing the so-called repetition index [46] and re-modulation [47] on the received signal. However, the number of time samples required with these methods is still implicitly proportional to the size of the IFFT, i.e. N_C. Regarding deterministic approaches, we note that they still need to accumulate time samples in order to obtain channel estimates algebraically, and their performance in noise improves as the number of time samples

increases.

Therefore, as the time averaging interval T_{av} is the major limitation on the quality of the sampled correlation matrix, any subspace-based blind algorithm which uses a smaller dimension of the correlation matrix may achieve a better estimation performance for a given T_{av}. Equivalently, aiming at any reasonable estimation performance, a subspace-based algorithm which uses a correlation matrix of smaller dimension generally requires a shorter time average interval. In Chapter 4, to meet the demands of future wireless standards, we propose a novel subspace-based algorithm for blind channel estimation in MIMO-OFDM systems, in which the required dimension of correlation matrix can be significantly less than those of the previously reported algorithms (that is, for the same system setup and parameter values). This will be achieved mainly by exploiting the frequency correlation among adjacent subcarriers in the MIMO-OFDM systems.

Chapter 4

Subspace-based blind channel estimation with reduced time averaging for MIMO-OFDM

In this chapter, to relax the time invariance requirement in practical MIMO-OFDM systems, we propose a novel subspace-based blind channel estimation algorithm with reduced time averaging. This is achieved by exploiting the frequency correlation among adjacent subcarriers in OFDM transmissions through subcarrier grouping [60], for which some supportive field measurements can also be found in [61]. The resulting gain in performance comes at the cost of an ambiguity matrix with larger dimensions; however, this dimension can be easily reduced to the normal one when precoding is applied [125] (see also Section 6.3) or when the ratio of coherence bandwidth to the channel bandwidth is large.

In Chapter 7, through simulations over 3GPP-SCM wireless channels, the proposed approach is shown to outperform the approach from [47]. In particular, it can achieve a normalized mean square error (NMSE) of 10^{-4} on the channel estimates within only 50 time samples (when the SNR=15dB), which is also very competitive over the deterministic approaches designed exclusively for SISO and single-carrier transmissions.

In a nutshell, the contribution of this work is not only to show that the proposed blind approach can work with a small number of time samples but that it may come with improved performance and robustness over existing statistical and deterministic methods.

The chapter is organized as follows. Section 4.1 is devoted to the problem formulation, including a description of the system under consideration. We introduce the proposed subspace-based blind algorithm in Section 4.2. Identifiability conditions as well as performance analysis of the proposed algorithm are given in Section 4.3.

4.1 Problem formulation

We consider a MIMO-OFDM system with N_T transmit and N_R receive antennas, employing N_C subcarriers, as depicted in Fig. 2.2. Let the mth OFDM symbol transmitted over the kth subcarrier be denoted as $\mathbf{x}^m[k] \stackrel{\text{def}}{=} \left[\, x_1^m[k] \;\; x_2^m[k] \;\; \cdots \;\; x_{N_T}^m[k] \,\right]^T$, where $x_q^m[k]$ is the symbol transmitted at the qth transmit antenna. Then the mth OFDM symbol transmitted over the N_C subcarriers can be written in a concatenated form as $\mathbf{x}^m \stackrel{\text{def}}{=} \left[\mathbf{x}^m[0]^T \;\; \mathbf{x}^m[1]^T \;\; \cdots \;\; \mathbf{x}^m[N_C-1]^T\right]^T$. The *input vector*, which represents an OFDM block in our system setup, is assumed to consist of N_F OFDM symbols and thus can be written as $\mathbf{x} = \left[\mathbf{x}^{1^T} \;\; \mathbf{x}^{2^T} \;\; \cdots \;\; \mathbf{x}^{N_F{}^T}\right]^T$.

At the receiver, let the mth received OFDM symbol over the kth subcarrier be denoted as $\mathbf{y}^m[k] \stackrel{\text{def}}{=} \left[\, y_1^m[k] \;\; y_2^m[k] \;\; \cdots \;\; y_{N_R}^m[k] \,\right]^T$, where $y_p^m[k]$ is the symbol received at the pth receive antenna. Then the mth OFDM symbol received over N_C subcarriers can be written as $\mathbf{y}^m \stackrel{\text{def}}{=} \left[\mathbf{y}^m[0]^T \;\; \mathbf{y}^m[1]^T \;\; \cdots \;\; \mathbf{y}^m[N_C-1]^T\right]^T$, and $\mathbf{y} = \left[\mathbf{y}^{1^T} \;\; \mathbf{y}^{2^T} \;\; \cdots \;\; \mathbf{y}^{N_F{}^T}\right]^T$ represents the *observation*. In addition, let $\mathbf{n}$ represent the additive *noise* structured in a similar way.

In the following, we assume that: (1) the length of the CP appended to each OFDM symbol is longer than the maximum excess delay of the channel; (2) the channel is time-

invariant at least over each OFDM block; and (3) the average power of the transmit symbol alphabet is normalized: $E\left[|x_q^m[k]|^2\right] = 1$. Under these assumptions, the input-output relation of the MIMO-OFDM system may be expressed compactly by

$$\mathbf{y} = \mathcal{H}\,\mathbf{x} + \mathbf{n}, \tag{4.1}$$

where the matrix $\mathcal{H}$ is defined as

$$\mathcal{H} \stackrel{\text{def}}{=} \mathbf{I}_{N_F} \otimes \text{diag}(\mathbf{H}[0] \cdots \mathbf{H}[N_C - 1]) \tag{4.2}$$

with size $(N_R N_C N_F) \times (N_T N_C N_F)$ and the definition of diagonal blocks is given in (2.9).

In this chapter, our interest lies in the blind, SOS-based estimation of the channel coefficients, i.e. $\{h_{p,q}[k]\}$, directly from multiple observations of random vectors $\mathbf{y}$'s.

4.2 Subspace-based blind estimation

In the case of SOS-based blind approaches, the main concern is to estimate the correlation matrix $\mathbf{R}_\mathbf{y} = E[\mathbf{y}\mathbf{y}^H]$ while meeting a certain level of confidence, and over a time averaging interval as short as possible. We first briefly comment on the time averaging requirement in the *traditional* approaches: For subspace-based algorithms that apply channel estimation in the time domain and assuming $N_R \geq N_T$ (see e.g., [41, 44]), the channel matrix is block Toeplitz and can be written as $\mathcal{H}_{td} = \sum_{l=0}^{L-1} \mathbf{B}^l \otimes \mathbf{H}(l)$, where L represents the channel order, $\mathbf{B}$ is an $[(N_C + N_{cp})N_F - L + 1] \times (N_C + N_{cp})N_F$ *backward shift* matrix [121] with N_{cp} denoting the length of the cyclic prefix, and $\mathbf{H}(l) \stackrel{\text{def}}{=} (1/N_C) \sum_{k=0}^{N_C-1} \mathbf{H}[k] \exp(j2\pi kl/N_C)$, i.e. the MIMO channel impulse response of the lth tap.

Then, the dimension of the correlation matrix of the observations is $[(N_C + N_{cp})N_F -$

$L+1]N_R$, which can be approximated by $N_C N_F N_R$ if $N_C \gg N_{cp}$ and $N_C N_F \gg L$. On the basis of these considerations, we can conclude that in the context of MIMO-OFDM, we need to choose the number of time samples $T_{av} \geq N_C N_F N_R$ to achieve acceptable performance for these time-domain approaches. As N_C is normally chosen between 128 and 2048 so as to alleviate the adverse effects from the frequency-selective channels, we can see that these algorithms require an extremely large T_{av} for obtaining an acceptable time-averaged correlation matrix. Below, we develop an improved procedure that exploits correlation over the frequency domain to relax such a requirement.

4.2.1 Proposed approach

In the context of MIMO-OFDM, although a pilot-based subspace method in the frequency domain was proposed in [126], a blind one constructed directly from (4.1) has seldom been considered, mainly because there are a large number $N_T N_R N_C \geq N_T N_R L$ of unknowns to be estimated (recall that L represents the channel order). Nevertheless, the number of unknowns can be reduced by exploiting the frequency correlation among adjacent OFDM subcarriers with some loss in the estimation performance. In return, the dimension of the correlation matrix and hence the number of time samples required for time averaging can be reduced significantly. The details are given below for the case $N_R > N_T$; however, if oversampling is used at the receiver, the case $N_R \leq N_T$ is also possible.

Let the frequency span of P adjacent subcarriers reside inside the coherence bandwidth of the wireless channel, defined here as the range of frequencies over which the frequency response matrix of the MIMO channel does not change appreciably [127]. Let $\Omega \stackrel{\text{def}}{=} \{0, 1, \cdots, N_C - 1\}$, i.e. the index set of the N_C subcarriers, be partitioned into P disjoint subsets (assuming $\zeta \stackrel{\text{def}}{=} N_C/P \in \mathbb{Z}^+$) with each subset denoted as $\Omega_p \stackrel{\text{def}}{=} \{\omega_{p,1}, \omega_{p,2}, \cdots, \omega_{p,\zeta}\}$, where $\omega_{p,i} \stackrel{\text{def}}{=} p - 1 + (i-1)P$, $i = 1, 2, \cdots, \zeta$ for $p = 1, 2, \cdots, P$

(see Fig 4.1 for details). Note that $\Omega_1 \cup \Omega_2 \cup \cdots \cup \Omega_P = \Omega$, and $\Omega_i \cap \Omega_j = \emptyset$, where $\emptyset$ denotes the empty set.

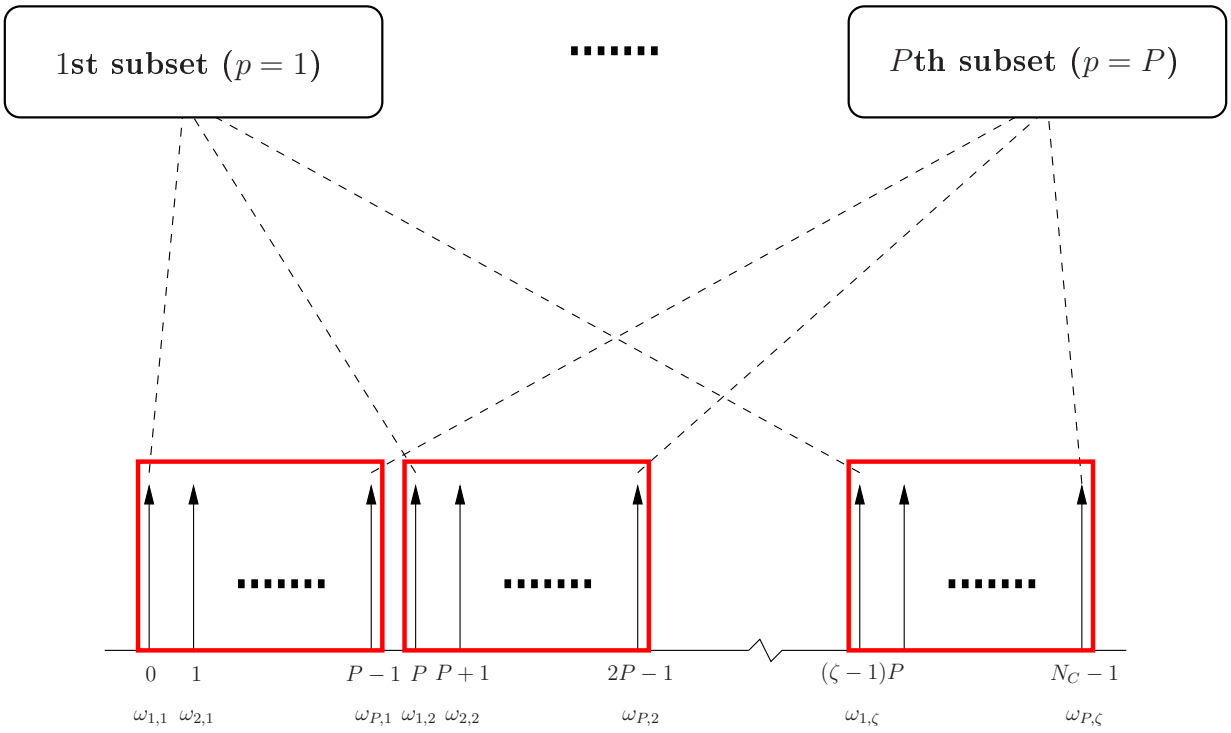


Fig. 4.1 A schematic of the partitioning of the subcarrier index set $\Omega = \{0, 1, \cdots, N_C - 1\}$ into P disjoint subsets, i.e. $\Omega_p = \{\omega_{p,1}, \omega_{p,2}, \cdots, \omega_{p,\zeta}\}$, $p = 1, 2, \cdots, P$.

Define $\mathbf{x}_p = [\mathbf{x}_p^{1^T} \mathbf{x}_p^{2^T} \cdots \mathbf{x}_p^{{N_F}^T}]^T$, $\mathbf{y}_p = [\mathbf{y}_p^{1^T} \mathbf{y}_p^{2^T} \cdots \mathbf{y}_p^{{N_F}^T}]^T$, $\mathbf{n}_p = [\mathbf{n}_p^{1^T} \mathbf{n}_p^{2^T} \cdots \mathbf{n}_p^{{N_F}^T}]^T$, where

$$\mathbf{x}_p^m \stackrel{\text{def}}{=} \{\, \mathbf{x}^m[k] \mid k \in \Omega_p \,\} = [\, \mathbf{x}^m[\omega_{p,1}]^T \; \mathbf{x}^m[\omega_{p,2}]^T \; \cdots \; \mathbf{x}^m[\omega_{p,\zeta}]^T \,]^T \tag{4.3}$$

and $\mathbf{y}_p^m$ and $\mathbf{n}_p^m$ are defined in a similar way. These vectors are obtained from the complete input vector $\mathbf{x}$, observation vector $\mathbf{y}$, and noise vector $\mathbf{n}$ above by retaining only the frequency component of the pth subset Ω_p. Then (4.1) can be re-written for the pth subset

as

$$\mathbf{y}_p = \mathcal{H}_p \, \mathbf{x}_p + \mathbf{n}_p, \quad p = 1, 2, \cdots, P, \tag{4.4}$$

where

$$\mathcal{H}_p \stackrel{\text{def}}{=} \mathbf{I}_{N_F} \otimes \text{diag}(\mathbf{H}[\omega_{p,1}] \cdots \mathbf{H}[\omega_{p,\zeta}]) \tag{4.5}$$

is assumed to be of full rank with size $(N_R N_F \zeta) \times (N_T N_F \zeta)$. The identification of $\mathcal{H}_p$ can then be achieved based on $\mathbf{R}_{\mathbf{y}_p} = E[\mathbf{y}_p \mathbf{y}_p^H]$, which can be re-written as

$$\mathbf{R}_{\mathbf{y}_p} = \mathcal{H}_p \mathbf{R}_{\mathbf{x}_p} \mathcal{H}_p^H + \mathbf{R}_{\mathbf{n}_p}, \tag{4.6}$$

where $\mathbf{R}_{\mathbf{x}_p} \stackrel{\text{def}}{=} E[\mathbf{x}_p \mathbf{x}_p^H]$ is assumed to be of full rank, and $\mathbf{R}_{\mathbf{n}_p} \stackrel{\text{def}}{=} E[\mathbf{n}_p \mathbf{n}_p^H] = \sigma_n^2 \mathbf{I}$. Since the P adjacent subcarriers are assumed to reside inside the coherence bandwidth, the sub-channel matrices $\mathcal{H}_p$, $p = 1, 2, \cdots, P$ can be *approximated*[1] by denoting $\bar{\mathcal{H}} \stackrel{\text{def}}{=} \mathcal{H}_1 = \mathcal{H}_2 = \cdots = \mathcal{H}_P$.

An estimate of the correlation matrix in (4.6) can be obtained as

$$\hat{\mathbf{R}}_{\bar{\mathbf{y}}} = \frac{1}{P T_{av}} \sum_{j=1}^{T_{av}} \sum_{p=1}^{P} \mathbf{y}_{p\,(j)} \mathbf{y}_{p\,(j)}^H, \tag{4.7}$$

where $\mathbf{y}_{p\,(j)} \in \mathbb{C}^{(N_R N_F \zeta) \times 1}$ denotes the jth observation of $\mathbf{y}_p$ at some physical time t_j. Therefore, the number of the time samples T_{av} required can be significantly reduced since the dimension of the correlation matrix is reduced by a factor of P, and an averaging over P subsets is applied at each time epoch, which is equivalent to the frequency averaging.

By applying the eigenvalue decomposition (EVD) to $\mathbf{R}_{\mathbf{y}_p}$, we can express (4.6) as

[1] In practice, there are always small variations of the sub-channel matrices over the assumed coherence bandwidth. The effects of such small variations on the estimation performance are considered and analyzed in Section 4.3.

$\mathbf{R}_{\mathbf{y}_p} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, where $\mathbf{U}$ is a matrix whose columns are the orthonormal eigenvectors of $\mathbf{R}_{\mathbf{y}_p}$, and which can be partitioned as $\mathbf{U} = [\,\mathbf{U}_s \,|\, \mathbf{U}_n\,] = [\,\mathbf{u}_1 \cdots \mathbf{u}_{d_s} \,|\, \mathbf{u}_{d_s+1} \cdots \mathbf{u}_{d_s+d_n}\,]$. The signal subspace can thus be denoted as $\mathfrak{R}(\mathbf{U}_s)$, while its orthogonal complement, the noise subspace, can be denoted as $\mathfrak{R}(\mathbf{U}_n)$, with $d_s \stackrel{\text{def}}{=} \text{rank}\,(\bar{\mathcal{H}}) = N_T N_F \zeta$ and $d_n \stackrel{\text{def}}{=} (N_R - N_T)N_F\zeta$. $\mathbf{\Lambda}$ is a diagonal matrix consisting of the corresponding eigenvalues of $\mathbf{R}_{\mathbf{y}_p}$, and is denoted as $\mathbf{\Lambda} = \text{diag}\,(\lambda_1, \lambda_2, \cdots, \lambda_{d_s+d_n})$ with $\lambda_{\max} = \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_{d_s} > \lambda_{d_s+1} = \cdots = \lambda_{d_s+d_n} \stackrel{\text{def}}{=} \lambda_{\min} \geq 0$. Under the assumption of white noise with non-zero variance, i.e., $\lambda_{\min} = \sigma_n^2 > 0$. Since $\bar{\mathcal{H}}$ and $\mathbf{U}_s$ share the same range space and are orthogonal to the range space of $\mathbf{U}_n$, we can arrive at the following orthogonality relationship

$$\mathbf{u}_j^H \bar{\mathcal{H}} = \mathbf{0}, \quad j = d_s + 1, \cdots, d_s + d_n. \tag{4.8}$$

Although $\bar{\mathcal{H}}$ can be solved from the set of homogeneous linear equations, due to the use of a finite time averaging interval, only an estimate of the noise subspace $\mathbf{U}_n$ is available in practice. In this case, by denoting $\hat{\mathbf{u}}_j$ as the perturbed version of $\mathbf{u}_j$, obtained by applying the EVD to the sample correlation matrix $\hat{\mathbf{R}}_{\bar{\mathbf{y}}}$ in (4.7), we may obtain the channel estimate by minimizing a quadratic cost function given by, e.g.,[2]

$$C(\bar{\mathcal{H}}) = \sum_{j=d_s+1}^{d_s+d_n} \|\hat{\mathbf{u}}_j^H \bar{\mathcal{H}}\|_2^2. \tag{4.9}$$

The trivial solution $\bar{\mathcal{H}} = \mathbf{0}$ can be avoided by introducing a suitable constraint as discussed below.

At this point, it is convenient to reformulate the quadratic cost function (4.9) in a form that is more convenient for its optimization over the unconstrained parameters in matrix $\bar{\mathcal{H}}$.

[2]Ideally, we should measure how close is $\hat{\mathbf{u}}_j^H \bar{\mathcal{H}}$ to the all zero vector $\mathbf{0}_{1\times N_T N_F \zeta}$. To this end, we are not restricted to the cost function in (4.9).

Let us first partition $\hat{\mathbf{u}}_j$ into N_F segments of equal dimension as $\hat{\mathbf{u}}_j = \left[\hat{\mathbf{u}}_{j,1}^T \hat{\mathbf{u}}_{j,2}^T \cdots \hat{\mathbf{u}}_{j,N_F}^T\right]^T$, and then define matrix $\hat{\mathbf{V}}_j \stackrel{\text{def}}{=} [\hat{\mathbf{u}}_{j,1} \hat{\mathbf{u}}_{j,2} \cdots \hat{\mathbf{u}}_{j,N_F}]$, where $\hat{\mathbf{u}}_{j,i} \in \mathbb{C}^{N_R\zeta \times 1}$ for $i = 1, 2, \cdots, N_F$. For a fixed but arbitrarily selected integer $\rho \in [1, P]$, let us define

$$\bar{\mathcal{H}}' \stackrel{\text{def}}{=} \bar{\mathcal{H}}'_\rho = \begin{bmatrix} \mathbf{H}[\omega_{\rho,1}] \\ \vdots \\ \mathbf{H}[\omega_{\rho,\zeta}] \end{bmatrix} = [\mathbf{h}_1^\rho \ \mathbf{h}_2^\rho \ \cdots \ \mathbf{h}_{N_T}^\rho], \tag{4.10}$$

where each column $\mathbf{h}_q^\rho$ is obtained by concatenating the corresponding columns of $\mathbf{H}[\omega_{\rho,k}]$, $k = 1, \cdots, \zeta$, that is,

$$\mathbf{h}_q^\rho \stackrel{\text{def}}{=} [h_{1,q}[\omega_{\rho,1}] \cdots \ h_{N_R,q}[\omega_{\rho,1}] \ \ \cdots \ \ h_{1,q}[\omega_{\rho,\zeta}] \cdots \ h_{N_R,q}[\omega_{\rho,\zeta}]]^T \tag{4.11}$$

for $q = 1, 2, \cdots, N_T$. Then the condition $C(\bar{\mathcal{H}}) = 0$ in (4.9) implies that $C'(\bar{\mathcal{H}}') = 0$, in which the latter is defined as (see Appendix A for details)

$$C'(\bar{\mathcal{H}}') = \text{tr}\left(\bar{\mathcal{H}}'^T \hat{\mathbf{\Psi}} \bar{\mathcal{H}}'^*\right), \tag{4.12}$$

and the matrix $\hat{\mathbf{\Psi}}$ is defined as

$$\hat{\mathbf{\Psi}} \stackrel{\text{def}}{=} \sum_{j=d_s+1}^{d_s+d_n} \hat{\mathbf{V}}_j^* \hat{\mathbf{V}}_j^T \in \mathbb{C}^{(N_R\zeta)\times(N_R\zeta)}. \tag{4.13}$$

We now proceed to obtain the channel estimate by minimizing the quadratic cost function in (4.12) as follows. Let the eigenvalues of $\hat{\mathbf{\Psi}}$ be ordered as $\gamma_{\min} = \gamma_1(\hat{\mathbf{\Psi}}) \leq \gamma_2(\hat{\mathbf{\Psi}}) \leq \cdots \leq \gamma_{(N_R\zeta)}(\hat{\mathbf{\Psi}}) = \gamma_{\max}$. Then from the Rayleigh-Ritz theorem [121], we know that for all

$\mathbf{Q} \in \mathbb{C}^{(N_R\zeta)\times r}$,

$$\gamma_1(\hat{\mathbf{\Psi}}) + \cdots + \gamma_r(\hat{\mathbf{\Psi}}) = \min_{\mathbf{Q}^H\mathbf{Q}=\mathbf{I}} \text{tr}\,(\mathbf{Q}^H\hat{\mathbf{\Psi}}\mathbf{Q}), \tag{4.14}$$

where r is a given integer with $1 \leq r \leq N_R\zeta$. The optimal solution $\hat{\mathbf{Q}}_o \in \mathbb{C}^{(N_R\zeta)\times r}$ is a matrix whose columns are chosen to be orthonormal eigenvectors corresponding to the r smallest eigenvalues of $\hat{\mathbf{\Psi}}$. Therefore we can carry on the minimization on the RHS of (4.14) to find $\hat{\mathbf{Q}}_o$ and obtain the desired solution of (4.12) by

$$\bar{\mathcal{H}}_o' = \hat{\mathbf{Q}}_o^*\mathbf{A}, \tag{4.15}$$

where $\mathbf{A} \in \mathbb{C}^{r\times N_T}$ can be seen as an *ambiguity matrix*. To ensure that enough basis functions are available for the adequate representation of the unknown channel matrix, r should be chosen so that $r = \dim[\Re(\hat{\mathbf{Q}}_o^*)] \geq \dim[\Re(\bar{\mathcal{H}}_o)]/N_F = N_T\zeta$; in our case, we simply choose $r = N_T\zeta$.

Table 4.1 Computational complexity of the proposed algorithm.

Main Step	Complexity (flops)
1. Compute $\hat{\mathbf{R}}_{\bar{\mathbf{y}}}$.	$\frac{3}{2}(PT_{av})(N_RN_F\zeta)^2$
2. Given $d_s = N_TN_F\zeta$ and $d_n = (N_R - N_T)N_F\zeta$, find eigenvectors $\hat{\mathbf{u}}_j$, $j = d_s+1, \cdots, d_s+d_n$, which correspond to the d_n smallest eigenvalues of $\hat{\mathbf{R}}_{\bar{\mathbf{y}}}$.	$\mathcal{O}((N_RN_F\zeta)^3)$
3. Partition $\hat{\mathbf{u}}_j = \left[\hat{\mathbf{u}}_{j,1}^T\hat{\mathbf{u}}_{j,2}^T\cdots\hat{\mathbf{u}}_{j,N_F}^T\right]^T$ and form the matrices $\hat{\mathbf{V}}_j$=$[\hat{\mathbf{u}}_{j,1}\hat{\mathbf{u}}_{j,2}\cdots\hat{\mathbf{u}}_{j,N_F}]$ from $\hat{\mathbf{u}}_j$, $j = d_s+1, \cdots, d_s+d_n$.	0
4. Form the matrix $\hat{\mathbf{\Psi}}$ from the $\hat{\mathbf{V}}_j$'s.	$d_n(N_F+1)(N_R\zeta)^2$
5. Find $\hat{\mathbf{Q}}_o$, whose columns are the eigenvectors which correspond to the $N_T\zeta$ smallest eigenvalues of $\hat{\mathbf{\Psi}}$.	$\mathcal{O}((N_R\zeta)^3)$
6. Obtain channel estimate $\bar{\mathcal{H}}_o'$.	$2(N_T\zeta)(N_R\zeta)N_T$

4.2.2 Further comments on the proposed approach

The main processing steps for the proposed algorithm are summarized in Table 4.1, along with their computational complexities in terms of the number of required (complex) flops. To meet the minimum requirement of time averaging in connection with (4.7), i.e., to avoid rank deficiency, we need $PT_{av} \geq N_R N_F \zeta$, or equivalently, $T_{av} \geq N_R N_F N_C / P^2$. Therefore, the reduction in the averaging time T_{av} is proportional *to the square* of the number of subsets, P. Assuming $PT_{av} = N_R N_F \zeta$ is chosen (i.e. the dimension of the correlation matrix), the total computational complexity of the proposed algorithm is given as $\mathcal{O}((N_R N_F \zeta)^3)$, including the two EVD operations. Although the steps of matrix computations are similar to those found in the traditional approaches, the complexity of the first EVD operation is in general much lower. A reduction by $P^3 \approx 10^{4.5}$ flops in the EVD operation can be expected for a typical value of $P = 32$.

The ambiguity matrix $\mathbf{A}$, inherent in all subspace-based blind channel estimation methods due to the second-order problem formulation, is required in order to obtain a final channel estimate in Step 6 (see also (4.15) for details). While the estimation of the ambiguity matrix for subspace methods is a general problem on its own that falls outside the scope of this thesis, several approaches are available in practice to implement this step, including the use of higher order statistics or the insertion of a limited number of pilot symbols (resulting in the so-called semi-blind approach). Concerning the size of the ambiguity matrix in the current problem, a simple precoding technique is presented in Chapter 6 to reduce it to $N_T \times N_T$, which is the size of the ambiguity in the corresponding time-domain subspace problems.

4.3 Performance analysis on the proposed subspace-based estimator

In this section, we analyze the performance of the newly proposed algorithm in Section 4.2. We first investigate its identifiability conditions, and then derive expressions for its asymptotic performance and associated Cramer-Rao bound. In Chapter 7, these analysis results will be used in conjunction with Monte Carlo numerical simulations to show that the proposed algorithm indeed achieves a better estimation accuracy than previous reported algorithms within a reasonable time averaging interval.

4.3.1 Identifiability conditions

In subspace-based blind estimation problems, the study of identifiability conditions aims to explore the structure of the channel and properties of the input such that we can uniquely determine the channel coefficients up to a certain degree of ambiguity [128]. The study is normally conducted explicitly from the viewpoint of the signal-noise orthogonality relationship, assuming that the time averaging interval is sufficient long to neglect estimation error of the sample correlation matrix; that is: assuming $\hat{\mathbf{R}}_{\bar{\mathbf{y}}} = \mathbf{R}_{\mathbf{y}_p}$. As a result, we can determine not only the type of channels that a specific blind algorithm can identify, but also the way that system parameters should be chosen.

In the following, assuming the dimensions of the signal and noise subspaces are known, we investigate to what extent the channel is still *identifiable*, i.e. channel coefficients can be uniquely determined up to a certain degree of ambiguity. First, let us express the orthogonality relationship in (4.8) in terms of the noise subspace eigenvector matrix $\mathbf{U}_n$ by

$$\mathbf{U}_n^H \bar{\mathcal{H}} = \mathbf{0}, \tag{4.16}$$

which is a homogeneous linear system.

Since $\bar{\mathcal{H}}$ is assumed to be of full rank, i.e. $\text{rank}(\bar{\mathcal{H}}) = N_T N_F \zeta = d_s$, (4.16) also implies $\mathfrak{R}(\bar{\mathcal{H}}) = \mathfrak{N}(\mathbf{U}_n^H)$.

The general solution of the system (4.16) is given as

$$\begin{aligned} \bar{\mathcal{H}} &= \left(\mathbf{I} - \left(\mathbf{U}_n^H\right)^{\dagger}\left(\mathbf{U}_n^H\right)\right)\mathbf{Y} \\ &= \left(\mathbf{I} - \mathbf{U}_n\mathbf{U}_n^H\right)\mathbf{Y} \\ &= \mathbf{P}_{\mathfrak{R}(\mathbf{U}_n)^{\perp}}\mathbf{Y} \end{aligned} \tag{4.17}$$

for some arbitrary matrix $\mathbf{Y}$ with $\mathbf{P}_{\mathfrak{R}(\mathbf{U}_n)^{\perp}} \stackrel{\text{def}}{=} \mathbf{I} - \mathbf{U}_n\mathbf{U}_n^H$ [129, p.140], where the second equality follows from the fact that $\mathbf{U}_n$ is a full column rank matrix with orthonormal columns. Therefore, we can always avoid the trivial solution $\bar{\mathcal{H}} = \mathbf{0}$ since $\mathbf{U}_n\mathbf{U}_n^H \neq \mathbf{I}$. Let $\bar{\mathcal{H}}' = \mathbf{Q}_o^*\mathbf{A}$ denote a solution to (4.14) constructed from the exact $\mathbf{Q}_o$ with corresponding $\bar{\mathcal{H}}$.[3] In addition, let $\bar{\mathcal{H}}_1' = \bar{\mathcal{H}}'\mathbf{B} = (\mathbf{Q}_o^*\mathbf{A})\mathbf{B} = \mathbf{Q}_o^*\mathbf{A}_1$ where $\mathbf{B}$ is a square matrix of dimension N_T. Then, it can be verified that the corresponding $\bar{\mathcal{H}}_1 = \bar{\mathcal{H}}\,(\mathbf{I}_\zeta \otimes \mathbf{B})$, which is also of the form $\mathbf{P}_{\mathfrak{R}(\mathbf{U}_n)^{\perp}}\mathbf{Y}_1$, with $\mathbf{Y}_1 = \mathbf{Y}\,(\mathbf{I}_\zeta \otimes \mathbf{B})$. This shows that $\bar{\mathcal{H}}'$ is uniquely determined up to the ambiguity matrix $\mathbf{A}$ ($\stackrel{\text{def}}{=} \mathbf{A}_1\mathbf{B}^{-1}$).

In the following, let us determine the sufficient condition for the channel to be identifiable. Recall that the dimension of the solution space of a homogeneous linear system $\mathbf{A}_{m\times n}\mathbf{X}_{n\times p} = \mathbf{0}_{m\times p}$ equals $p\dim[\mathfrak{N}(A)]$; the dimension of the solution space for $\bar{\mathcal{H}}$ in (4.16) is then given as $d_s\dim[\mathfrak{N}(\mathbf{U}_n^H)]$. Since $\bar{\mathcal{H}}$ is a block diagonal matrix with $d_s N_R$ nonzeros entries, a sufficient (but not necessary) condition for the channel to be identifiable can thus be written as $d_s\dim\left[\mathfrak{N}(\mathbf{U}_n^H)\right] \geq d_s N_R$, or simply $d_s \geq N_R$. We can rearrange the above inequality and arrive at $N_F\zeta \geq (N_R/N_T)$.

[3] As per the formating transformation $\bar{\mathcal{H}} \leftrightarrow \bar{\mathcal{H}}'$ defined in Section 4.2.

Therefore, $N_F = 1$ is possible for the proposed algorithm as long as $\zeta \geq (N_R/N_T)$, meaning that the aforementioned algorithm and identifiability conditions can be derived for the case of single-symbol processing, i.e., one OFDM input symbol at a time. However, our results can easily be extended to the general cases where the OFDM system operates on blocks of multiple OFDM symbols at a time, as is necessary with many traditional approaches [41, 44].

4.3.2 Perturbation analysis

By denoting $\mathcal{H}_p = \bar{\mathcal{H}} + \Delta\mathcal{H}_p$, the correlation matrix $\mathbf{R}_{\mathbf{y}_p}$ in (4.6) can be re-written as

$$\mathbf{R}_{\mathbf{y}_p} = \bar{\mathcal{H}}\,\mathbf{R}_{\mathbf{x}_p}\bar{\mathcal{H}}^H + \mathbf{R}_{\Delta\mathcal{H}_p} + \mathbf{R}_{\mathbf{n}_p}, \tag{4.18}$$

where we define

$$\mathbf{R}_{\Delta\mathcal{H}_p} = \Delta\mathcal{H}_p\,\mathbf{R}_{\mathbf{x}_p}\bar{\mathcal{H}}^H + \bar{\mathcal{H}}\,\mathbf{R}_{\mathbf{x}_p}\Delta\mathcal{H}_p^H + \Delta\mathcal{H}_p\,\mathbf{R}_{\mathbf{x}_p}\Delta\mathcal{H}_p^H. \tag{4.19}$$

In Section 4.2, an identification of $\bar{\mathcal{H}}$ was obtained by assuming $\|\Delta\mathcal{H}_p\| \to 0$, $\forall p$. In the following, the asymptotic performance of the proposed algorithm under high SNR and sufficiently large time averaging interval is studied.

Under the condition of stationarity and ergodicity, the estimate of the correlation matrix in (4.7) converges to:

$$\mathbf{R}_{\bar{\mathbf{y}}} = \underbrace{\bar{\mathcal{H}}\,\mathbf{R}_{\bar{\mathbf{x}}}\bar{\mathcal{H}}^H}_{\stackrel{\text{def}}{=}\mathbf{R}_{\bar{\mathcal{H}}}} + \mathbf{R}_{\Delta\mathcal{H}} + \mathbf{R}_{\bar{\mathbf{n}}} \;\text{ as }\; T_{av} \to \infty, \tag{4.20}$$

where $\mathbf{R}_{\bar{\mathbf{x}}} \stackrel{\text{def}}{=} (1/P)\sum_{p=1}^{P}\mathbf{R}_{\mathbf{x}_p}$, $\mathbf{R}_{\Delta\mathcal{H}} \stackrel{\text{def}}{=} (1/P)\sum_{p=1}^{P}\mathbf{R}_{\Delta\mathcal{H}_p}$ and $\mathbf{R}_{\bar{\mathbf{n}}} \stackrel{\text{def}}{=} (1/P)\sum_{p=1}^{P}\mathbf{R}_{\mathbf{n}_p}$.

Since $\mathbf{R}_{\bar{\mathbf{y}}} = \mathbf{R}_{\bar{\mathcal{H}}} + \mathbf{R}_{\Delta\mathcal{H}} + \mathbf{R}_{\bar{\mathbf{n}}}$ can be seen as a perturbed data matrix with $\mathbf{R}'_{\mathbf{n}} \stackrel{\text{def}}{=} \mathbf{R}_{\Delta\mathcal{H}} + \mathbf{R}_{\bar{\mathbf{n}}}$ being the perturbation source, we have the first order perturbation of the noise subspace [122, 130] denoted by

$$\Delta\mathbf{U}_{n,\bar{\mathcal{H}}} = -\mathbf{U}_{s,\bar{\mathcal{H}}}\Sigma^{-1}_{s,\bar{\mathcal{H}}}\mathbf{U}^{H}_{s,\bar{\mathcal{H}}}(\mathbf{R}'_{\mathbf{n}})^{H}\mathbf{U}_{n,\bar{\mathcal{H}}}, \tag{4.21}$$

assuming the EVD of $\mathbf{R}_{\bar{\mathcal{H}}}$ is written as

$$\mathbf{R}_{\bar{\mathcal{H}}} = [\ \mathbf{U}_{s,\bar{\mathcal{H}}} \mid \mathbf{U}_{n,\bar{\mathcal{H}}}\ ] \begin{bmatrix} \mathbf{\Sigma}_{\mathbf{s},\bar{\mathcal{H}}} & \\ & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U}^{H}_{s,\bar{\mathcal{H}}} \\ \mathbf{U}^{H}_{n,\bar{\mathcal{H}}} \end{bmatrix}. \tag{4.22}$$

Furthermore, by partitioning the jth column of $\mathbf{U}_{n,\bar{\mathcal{H}}}$ and $\Delta\mathbf{U}_{n,\bar{\mathcal{H}}}$ into N_F segments of equal dimension, we define new matrices $\mathbf{V}_j \stackrel{\text{def}}{=} [\mathbf{U}^{j,1}_{n,\bar{\mathcal{H}}}\mathbf{U}^{j,2}_{n,\bar{\mathcal{H}}} \cdots \mathbf{U}^{j,N_F}_{n,\bar{\mathcal{H}}}]$ and $\Delta\mathbf{V}_j \stackrel{\text{def}}{=} [\Delta\mathbf{U}^{j,1}_{n,\bar{\mathcal{H}}}\Delta\mathbf{U}^{j,2}_{n,\bar{\mathcal{H}}} \cdots \Delta\mathbf{U}^{j,N_F}_{n,\bar{\mathcal{H}}}]$, where $\mathbf{U}^{j,i}_{n,\bar{\mathcal{H}}}$ and $\Delta\mathbf{U}^{j,i}_{n,\bar{\mathcal{H}}}$ denote the ith segment of the jth column of $\mathbf{U}_{n,\bar{\mathcal{H}}}$ and $\Delta\mathbf{U}_{n,\bar{\mathcal{H}}}$, respectively. Similar to (4.13), $\mathbf{\Psi} \stackrel{\text{def}}{=} \sum_{j=1}^{d_n} \mathbf{V}^{*}_{j}\mathbf{V}^{T}_{j}$ is then constructed for the Hermitian matrix $\mathbf{Q}_o$ (note that $\mathbf{Q}_o$ is the un-perturbed version of $\hat{\mathbf{Q}}_o$). The EVD of $\mathbf{\Psi}$ can be written as

$$\mathbf{\Psi} = [\ \mathbf{U}_{s,\Psi} \mid \mathbf{U}_{n,\Psi}\ ] \begin{bmatrix} \mathbf{\Sigma}_{\mathbf{s},\mathbf{\Psi}} & \\ & \mathbf{0} \end{bmatrix} \begin{bmatrix} \mathbf{U}^{H}_{s,\Psi} \\ \mathbf{U}^{H}_{n,\Psi} \end{bmatrix}, \tag{4.23}$$

where we identify $\mathbf{Q}_o = \mathbf{U}_{n,\Psi}$ and hence $\hat{\mathbf{Q}}_o = \mathbf{Q}_o + \Delta\mathbf{Q}_o$ with

$$\Delta\mathbf{Q}_o = -\mathbf{U}_{s,\Psi}\Sigma^{-1}_{s,\Psi}\mathbf{U}^{H}_{s,\Psi}(\Delta\mathbf{\Psi})^{H}\mathbf{Q}_o, \tag{4.24}$$

where $\Delta\mathbf{\Psi} \stackrel{\text{def}}{=} \sum_{j=1}^{d_n} \mathbf{V}^{*}_{j}\Delta\mathbf{V}^{T}_{j} + \Delta\mathbf{V}^{*}_{j}\mathbf{V}^{T}_{j} + \Delta\mathbf{V}^{*}_{j}\Delta\mathbf{V}^{T}_{j}$. By assuming that $\bar{\mathcal{H}}' = [\mathbf{h}^{\rho}_{1}\ \mathbf{h}^{\rho}_{2}\ \cdots\ \mathbf{h}^{\rho}_{N_T}]$ for known, but arbitrary value of $\rho \in [1, P]$, we can obtain the ambiguity matrix from $\mathbf{A} = (\hat{\mathbf{Q}}^{*}_{o})^{\dagger}\bar{\mathcal{H}}'$. Therefore, the asymptotic channel estimate and asymptotic estimation

error can be written as[4]

$$\bar{\mathcal{H}}'_{o,a} = (\hat{\mathbf{Q}}^*_o)(\hat{\mathbf{Q}}^*_o)^\dagger \bar{\mathcal{H}}', \tag{4.25}$$

$$\Delta\bar{\mathcal{H}}'_{o,a} = \left[\mathbf{I} - (\hat{\mathbf{Q}}^*_o)(\hat{\mathbf{Q}}^*_o)^\dagger\right] \bar{\mathcal{H}}'. \tag{4.26}$$

Note that since $(\hat{\mathbf{Q}}^*_o)(\hat{\mathbf{Q}}^*_o)^\dagger = \mathbf{P}_{\Re(\hat{\mathbf{Q}}^*_o)}$ and $\mathbf{I}-(\hat{\mathbf{Q}}^*_o)(\hat{\mathbf{Q}}^*_o)^\dagger = \mathbf{P}_{\Re(\hat{\mathbf{Q}}^*_o)^\perp}$ [131, 132], each column of the estimated channel matrix $\bar{\mathcal{H}}'_{o,a}$ can be seen as the projection of the corresponding column of $\bar{\mathcal{H}}'$ on the range of $\hat{\mathbf{Q}}^*_o$. Hence, as long as $\Re(\bar{\mathcal{H}}') \nsubseteq \Re(\hat{\mathbf{Q}}^*_o)$, we have $\bar{\mathcal{H}}'_{o,a} = \mathbf{P}_{\Re(\hat{\mathbf{Q}}^*_o)}\bar{\mathcal{H}}' \neq \bar{\mathcal{H}}'$ and $\Delta\bar{\mathcal{H}}'_{o,a} = \mathbf{P}_{\Re(\hat{\mathbf{Q}}^*_o)^\perp}\bar{\mathcal{H}}' \neq \mathbf{0}$.

To quantify the estimation performance, let us define the ν-distance $\delta_\nu(\mathbf{\Gamma}, \mathbf{\Sigma})$ between two matrices $\mathbf{\Gamma}$ and $\mathbf{\Sigma}$ of the same dimension in a linear space by

$$\delta_\nu(\mathbf{\Gamma}, \mathbf{\Sigma}) = \|\mathrm{vec}\,(\mathbf{\Gamma} - \mathbf{\Sigma})\,\|_\nu \tag{4.27}$$

for any real number $\nu \geq 1$. The normalized mean square error (NMSE) and channel average bias (CAB) of the channel estimate can then be conveniently expressed by

$$\mathrm{NMSE} = E\left[\sum_{p=1}^{P} \delta_2^2(\bar{\mathcal{H}}'_o, \bar{\mathcal{H}}'_p)\right] / E\left[\sum_{p=1}^{P} \|\bar{\mathcal{H}}'_p\|_F^2,\right] \tag{4.28}$$

$$\mathrm{CAB} = \frac{1}{N_T N_R N_C} E\left[\sum_{p=1}^{P} \delta_1(\bar{\mathcal{H}}'_o, \bar{\mathcal{H}}'_p)\right], \tag{4.29}$$

where $\bar{\mathcal{H}}'_o$ is given either by (4.15) for evaluating the channel estimate or by (4.25) for the asymptotic channel estimate, and $\bar{\mathcal{H}}'_p$ is defined in Section 4.2.1.

[4]Here asymptotic error refers to a situation where SNR is large and time averaging interval is sufficiently long.

4.3.3 Cramer-Rao bound

To determine whether the proposed algorithm is *efficient* or not specifically in the high SNR regime, we evaluate the unbiased Cramer-Rao bound (CRB) for the problem under consideration. Let $\theta = [\sigma_n^2\,,\, \mathrm{vec}(\bar{\mathcal{H}}')^T]^T$ be the column vector of nonrandom parameters to be estimated from the observations. The Fisher information matrix (FIM) for the complex valued parameter vector θ can be obtained by [133]

$$\mathcal{J}_{\theta\theta} = E_{\mathbf{y}_p|\,\theta}\left(\frac{\partial \ln f(\mathbf{y}_p|\,\theta)}{\partial \theta}\right)\left(\frac{\partial \ln f(\mathbf{y}_p|\,\theta)}{\partial \theta}\right)^H, \tag{4.30}$$

where the observation can be characterized by the complex probability density function of $\mathbf{y}_p$. On the basis of central limit theorem, the latter is assumed to be of ZMCSCG distribution , i.e. $\mathcal{N}(\mathbf{0}, \mathbf{R}_{\mathbf{y}_p})$, where the covariance matrix is defined as $\mathbf{R}_{\mathbf{y}_p} = \sigma_s^2\bar{\mathcal{H}}\bar{\mathcal{H}}^H + \sigma_n^2\mathbf{I}$.

We can estimate the real vector given by $\theta_R = [\Re(\theta^T), \Im(\theta^T)]^T$, where $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary part of a complex vector, respectively. The real FIM $\mathcal{J}_{\theta_R\theta_R}$ can thus be determined from $\mathcal{J}_{\theta\theta}$ by

$$\mathcal{J}_{\theta_R\theta_R} = \mathbf{M}\begin{bmatrix} \mathcal{J}_{\theta\theta} & \mathcal{J}_{\theta\theta^*} \\ \mathcal{J}^*_{\theta\theta^*} & \mathcal{J}^*_{\theta\theta} \end{bmatrix}\mathbf{M}^H, \tag{4.31}$$

where

$$\mathbf{M} = \frac{1}{2}\begin{bmatrix} \mathbf{I} & \mathbf{I} \\ -i\mathbf{I} & i\mathbf{I} \end{bmatrix},\quad \theta_R = \begin{bmatrix} \Re(\theta) \\ \Im(\theta) \end{bmatrix} = \mathbf{M}\begin{bmatrix} \theta \\ \theta^* \end{bmatrix}. \tag{4.32}$$

For the above complex Gaussian model, we have [134]

$$[\mathcal{J}_{\theta\theta}]_{i,j} = \mathrm{tr}\left\{\mathbf{R}_{\mathbf{y}_p}^{-1}\left(\frac{\partial \mathbf{R}_{\mathbf{y}_p}}{\partial \theta_i^*}\right)\mathbf{R}_{\mathbf{y}_p}^{-1}\left(\frac{\partial \mathbf{R}_{\mathbf{y}_p}}{\partial \theta_j^*}\right)^H\right\}, \tag{4.33}$$

$$[\mathcal{J}_{\theta\theta^*}]_{i,j} = \mathrm{tr}\left\{\mathbf{R}_{\mathbf{y}_p}^{-1}\left(\frac{\partial \mathbf{R}_{\mathbf{y}_p}}{\partial \theta_i^*}\right)\mathbf{R}_{\mathbf{y}_p}^{-1}\left(\frac{\partial \mathbf{R}_{\mathbf{y}_p}}{\partial \theta_j^*}\right)\right\}. \tag{4.34}$$

Note that we still use $\mathbf{R}_{\mathbf{y}_p}$ instead of $\hat{\mathbf{R}}_{\bar{\mathbf{y}}}$ in (4.33)-(4.34), because perturbations among adjacent channel coefficients inside the coherence bandwidth may exist. However, in ideal cases when these perturbations can be neglected, we can replace $\mathbf{R}_{\mathbf{y}_p}$ by $\hat{\mathbf{R}}_{\bar{\mathbf{y}}}$ with sufficient time averaging and obtain the minimum constrained CRB [134] from PT_{av} independent and identically distributed (i.i.d.) observations as

$$\mathrm{Var}(\hat{\theta}_{R_i}) \geq [(PT_{av}\mathcal{J}_{\theta_R\theta_R})^\dagger]_{i,i}, \tag{4.35}$$

where $\hat{\theta}_{R_i}$ denotes the ith entry of $\hat{\theta}_R$. Combining the CRB of the real and imaginary parts of the ith entry of the complex parameter vector θ, we can arrive at

$$\mathrm{Var}(\hat{\theta}_i) \geq \frac{1}{PT_{av}}\left\{[\mathcal{J}_{\theta_R\theta_R})^\dagger]_{i,i} + [\mathcal{J}_{\theta_R\theta_R})^\dagger]_{i+N_b,i+N_b}\right\}, \tag{4.36}$$

$i = 1, 2, \cdots, N_b$, where $N_b = N_T N_R \zeta + 1$ denotes the length of the nonrandom vector θ. Recall that $\theta_1 = \sigma_n^2$ and θ_i, $i = 2, \cdots, N_b$ represent the channel coefficients of interests. Hence, for any channel coefficient of interests, we can apply the following lower bound

$$\mathrm{Var}(\hat{h}_{p,q}[k]) \geq \min_i \frac{1}{PT_{av}}\left\{[\mathcal{J}_{\theta_R\theta_R})^\dagger]_{i,i} + [\mathcal{J}_{\theta_R\theta_R})^\dagger]_{i+N_b,i+N_b}\right\}, \tag{4.37}$$

for $i = 2, \cdots, N_b$.

Since above expression (4.37) gives a lower bound on the variance of deterministic unknowns, it can be used to determine the lower bound on the variance of the channel coefficients if the channel is fixed. However, in order to determine the lower bound on the variance of the channel coefficients when the latter as drawn from a given probability distribution, we can run a series of experiments and determine the lower bound by

$$\mathrm{Var}(\hat{h}_{p,q}[k]) = \lim_{j\to\infty} \inf \ \mathrm{Var}(\hat{h}^{(j)}_{p,q}[k]), \tag{4.38}$$

where $\mathrm{Var}(\hat{h}^{(j)}_{p,q}[k])$ represents the lower bound evaluated for (i.e. conditioned on) the jth channel realization. In Chapter 7, the above formulas for the CRB will be evaluated numerically and used as a benchmark in the performance evaluation of the proposed estimators via Monte Carlo simulations.

Chapter 5

Subspace tracking based on orthogonal iteration: convergence behavior

As a first step toward the development of adaptive subspace-based channel estimator for MIMO-OFDM systems, this chapter studies convergence properties of subspace trackers using orthogonal iteration. In the context of blind estimation of a time-varying channel, orthogonal iteration and its variants have been widely considered for tracking the channel parameters by updating the EVD of an exponentially weighted correlation matrix. It is well known that when orthogonal iteration is applied to a fixed matrix, it converges exponentially to the EVD (or dominant subspace) of the matrix with arbitrary non-zero initial conditions. However, orthogonal-iteration-based subspace trackers can only inherit these merits when the channels considered undergo extremely slow time-variations. In this chapter, we extend the traditional (i.e. fixed subspace) convergence analysis of the orthogonal iteration to include non-stationary situations as well. We use the results to investigate the convergence behavior of subspace trackers based on orthogonal iteration under slow, moderate and rapid time-variations of the underlying subspace. In the latter

case, we expose a fundamental limitation of the orthogonal iteration, i.e. practical limit on subspace variations to ensure effective tracking.[1]

5.1 Motivations of using subspace tracking

Subspace decomposition has proved to be an important tool in various signal processing applications. To this end, a straightforward approach is simply to employ either an eigenvalue decomposition (EVD) or singular value decomposition (SVD). These approaches, which belong to the family of direct or block processing techniques, are characterized as computationally demanding procedures and unsuitable for online processing due to their lack of repetitive structure [55]. Furthermore, they are often implemented in a batch mode, using an estimated correlation matrix obtained by collecting time samples over a sufficiently long observation interval. Therefore, these approaches, which rely on the assumption of statistical stationarity of the data, cannot be used in situations where the characteristics of the received signals change with time [135]. Computationally efficient and sequential algorithms that produce an exact or approximate EVD or SVD at each time step are generally favored in signal processing applications.

Thus, a considerable effort has gone into the development of sequential adaptive algorithms, also known as *subspace trackers*. To date, several signal-subspace trackers have been proposed for non-stationary environments. Instead of recomputing the EVD or SVD from scratch with every update, these algorithms attempt to recursively update these decompositions so as to minimize the amount of computations involved (see e.g. [135–138] and references therein). While there are many more signal-subspace trackers than noise-

[1] While most fast subspace trackers with low complexity assume a rank one update, they are not applicable to the adaptive subspace-based channel estimator that will be proposed in the next chapter. We thus focus on the orthogonal iteration that is more general and suitable for our design.

subspace trackers in the literature [139], nonetheless, in the applications of blind channel estimation, we can transform the noise-subspace problems into signal-subspace ones without much effort [38].

Orthogonal iteration is a simple technique that can be used to compute higher-dimensional invariant subspaces [140]. It is shown to have a global and exponential convergence property under a mild assumption on the distribution of eigenvalues, with arbitrary initial conditions [141]. In addition, it is suitable for real-time processing because it is well structured [136]. Therefore, orthogonal iteration and its variants have been considered for blind adaptive channel estimation to a great extent. Existing subspace tracking algorithms can be broadly categorized as whether or not they are based on orthogonal iteration. For the orthogonal-iteration-based subspace trackers, their variants include the low rank adaptive filter (LORAF) [136], the orthogonal projection approximation and subspace tracking (OPAST) [142], the Oja's method, and the novel information criterion (NIC) [143]. Recently, improvements on these existing approaches can also be found in [139, 144, 145].

In this chapter, given that numerous subspace trackers in the literature are fundamentally derived from the concept of orthogonal iteration, we first investigate the convergence properties when orthogonal iteration is applied in non-stationary scenarios. Specifically, we are interested in the distance between the *true* and the orthogonal-iterated subspaces. Then we study a fundamental limitation on the application of orthogonal-iterated subspace trackers in time-varying scenarios. Our results will be useful for better understanding the behavior of subspace trackers based on orthogonal iteration when applied to estimate time-varying MIMO channels in the next chapter.

5.2 Orthogonal iteration and its applications

Given a tall, column orthonormal matrix $\mathbf{Q}_0 \in \mathbb{C}^{N\times r}$, the so-called method of *orthogonal iteration* generates a sequence of matrices $\mathbf{Q}_m$, whose column span is assumed to approximate the span of the $r-$ dimensional dominant subspace of the matrix $\mathbf{W} \in \mathbb{C}^{N\times N}$, according to the following recurrence:

$$\begin{aligned} \mathbf{A}_m &= \mathbf{W}\mathbf{Q}_{m-1}, \quad m = 1, 2, \cdots \\ \mathbf{Q}_m\mathbf{R}_m &= \mathbf{A}_m, \end{aligned} \tag{5.1}$$

where $\mathbf{Q}_m$ and $\mathbf{R}_m$ denote the QR decomposition of the matrix $\mathbf{A}_m$ at the mth iteration. If $\mathbf{W}$ does not change over time, one can show that the subspace $\mathfrak{R}(\mathbf{Q}_m)$ converges to $\mathfrak{D}_r(\mathbf{W})$ at a rate proportional to $|\lambda_{r+1}(\mathbf{W})/\lambda_r(\mathbf{W})|^m$ [140]. Therefore, the usefulness of the method depends on this ratio, since it determines the rate of convergence. Note that when $r = 1$, (5.1) is just the well-known *power* method [146].

In several applications of interest in signal processing and communications, however, the assumption on the stationarity of $\mathbf{W}$ is usually not valid. Instead, a time-varying sequence $\{\mathbf{W}_m\}_{m=1}^{\infty}$ is often used, which is updated recursively as in e.g.:

$$\mathbf{W}_m = \alpha\mathbf{W}_{m-1} + (1-\alpha)\mathbf{z}_m\mathbf{z}_m^H, \tag{5.2}$$

where m now represents the discrete-time index, $\alpha \in [0, 1]$ represents the forgetting factor (typically close to 1), and $\mathbf{z}_m \in \mathbb{C}^{N\times 1}$ denotes an observation vector at time m, often modeled as an i.i.d. sequence of random vectors. In this case, we may sequentially track the $r-$ dimensional dominant subspace of the time-varying sequence $\{\mathbf{W}_m\}_{m=1}^{\infty}$ simply by replacing the stationary matrix $\mathbf{W}$ in (5.1) with $\mathbf{W}_m$ [147].

5.3 Convergence analysis

In order to motivate the method and to derive its convergence properties in non-stationary scenarios, we follow the analysis as well as the notation for the stationary case given in [140], and generalize the orthogonal iteration as follows.

To begin, let us consider k iterations of the recurrence in (5.1) and use induction to express it by

$$\underbrace{\mathbf{W}_k\mathbf{W}_{k-1}\cdots\mathbf{W}_1}_{\stackrel{\text{def}}{=}\bar{\mathbf{W}}_k}\mathbf{Q}_0 = \mathbf{Q}_k\mathbf{R}_k\mathbf{R}_{k-1}\cdots\mathbf{R}_1, \tag{5.3}$$

where $\mathbf{W}_1,\cdots,\mathbf{W}_k$ represent matrices of interest over the first k time iterations, respectively. Assume that

$$\bar{\mathbf{U}}_k^H\bar{\mathbf{W}}_k\bar{\mathbf{U}}_k = \bar{\mathbf{\Lambda}}_k = \mathrm{diag}(\bar{\lambda}_{i,k}) \tag{5.4}$$

is an EVD of $\bar{\mathbf{W}}_k$ with $\bar{\lambda}_{1,k} \geq \bar{\lambda}_{2,k} \geq \cdots \geq \bar{\lambda}_{N,k} \geq 0$ and $\bar{\mathbf{U}}_k^H\bar{\mathbf{U}}_k = \mathbf{I}$. Partition $\bar{\mathbf{U}}_k$ and $\bar{\mathbf{\Lambda}}_k$ as follows:

$$\bar{\mathbf{U}}_k = \left[\bar{\mathbf{U}}_{1,k}\ \bar{\mathbf{U}}_{2,k}\right],\ \ \bar{\mathbf{\Lambda}}_k = \begin{bmatrix} \bar{\mathbf{\Lambda}}_{1,k} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{\Lambda}}_{2,k} \end{bmatrix}, \tag{5.5}$$

where $\bar{\mathbf{U}}_{1,k} \in \mathbb{C}^{N\times r}$, $\bar{\mathbf{U}}_{2,k} \in \mathbb{C}^{N\times(N-r)}$, $\bar{\mathbf{\Lambda}}_{1,k} \in \mathbb{C}^{r\times r}$, and $\bar{\mathbf{\Lambda}}_{2,k} \in \mathbb{C}^{(N-r)\times(N-r)}$. Then we can arrive at

$$\begin{bmatrix} \bar{\mathbf{\Lambda}}_{1,k} & \mathbf{0} \\ \mathbf{0} & \bar{\mathbf{\Lambda}}_{2,k} \end{bmatrix}\begin{bmatrix} \bar{\mathbf{U}}_{1,k}^H\mathbf{Q}_0 \\ \bar{\mathbf{U}}_{2,k}^H\mathbf{Q}_0 \end{bmatrix} = \begin{bmatrix} \bar{\mathbf{U}}_{1,k}^H\mathbf{Q}_k \\ \bar{\mathbf{U}}_{2,k}^H\mathbf{Q}_k \end{bmatrix}(\mathbf{R}_k\mathbf{R}_{k-1}\cdots\mathbf{R}_1). \tag{5.6}$$

If we let

$$\bar{\mathbf{U}}_k^H \mathbf{Q}_l = \begin{bmatrix} \bar{\mathbf{U}}_{1,k}^H \mathbf{Q}_l \\ \bar{\mathbf{U}}_{2,k}^H \mathbf{Q}_l \end{bmatrix} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{V}_l \\ \mathbf{Y}_l \end{bmatrix}, \quad l = 0, 1, \cdots, k, \tag{5.7}$$

then

$$\mathbf{Y}_k = \bar{\mathbf{\Lambda}}_{2,k} \mathbf{Y}_0 \mathbf{V}_0^{-1} \bar{\mathbf{\Lambda}}_{1,k}^{-1} \mathbf{V}_k \tag{5.8}$$

can be obtained by using (5.6) and (5.7). We can define the distance between the two subspaces $\mathfrak{D}_r(\bar{\mathbf{W}}_k)$ and $\mathfrak{R}(\mathbf{Q}_k)$ according to [140]

$$\text{dist}\left(\mathfrak{D}_r(\bar{\mathbf{W}}_k), \mathfrak{R}(\mathbf{Q}_k)\right) = \|\bar{\mathbf{U}}_{2,k}^H \mathbf{Q}_k\|_2 = \|\mathbf{Y}_k\|_2. \tag{5.9}$$

By invoking (5.8), we can obtain

$$\|\mathbf{Y}_k\|_2 \leq \|\bar{\mathbf{\Lambda}}_{2,k}\|_2 \|\mathbf{Y}_0\|_2 \|\mathbf{V}_0^{-1}\|_2 \|\bar{\mathbf{\Lambda}}_{1,k}^{-1}\|_2 \|\mathbf{V}_k\|_2. \tag{5.10}$$

Let $\bar{\theta}_k \in [0, \pi/2]$ be defined to provide another measure of the closeness of the two subspaces $\mathfrak{D}_r(\bar{\mathbf{W}}_k)$ and $\mathfrak{R}(\mathbf{Q}_0)$, according to

$$\cos\left(\bar{\theta}_k\right) = \min_{\mathbf{u} \in \mathfrak{D}_r(\bar{\mathbf{W}}_k),\, \mathbf{v} \in \mathfrak{R}(\mathbf{Q}_0)} \frac{|\mathbf{u}^H \mathbf{v}|}{\|\mathbf{u}\|_2 \|\mathbf{v}\|_2}. \tag{5.11}$$

Then $\cos\left(\bar{\theta}_k\right) = \sigma_r(\bar{\mathbf{U}}_{1,k}^H \mathbf{Q}_0) = \sigma_r(\mathbf{V}_0)$ and $\|\mathbf{Y}_0\|_2 = \sin(\bar{\theta}_k)$, where $\sigma_r(\mathbf{V}_0)$ denotes the rth largest singular value of the matrix $\mathbf{V}_0$ [140]. Combining with (5.9)-(5.11), we can finally arrive at

$$\text{dist}(\mathfrak{D}_r(\bar{\mathbf{W}}_k), \mathfrak{R}(\mathbf{Q}_k)) \leq \tan(\bar{\theta}_k) \left(\frac{\bar{\lambda}_{r+1,k}}{\bar{\lambda}_{r,k}}\right). \tag{5.12}$$

In the following, we categorize the non-stationary scenarios into three main cases and show how the result in (5.12) can be used to study the convergence properties in each case.

Case 1: Very small time variations

In general, we can express $\mathbf{W}_{k-i} = \mathbf{W}_k + \Delta\mathbf{W}_{k,i}$ for $i = 1, 2, \cdots, k-1$. Therefore, $\bar{\mathbf{W}}_k$ in (5.3) can be re-written as

$$\begin{aligned} \bar{\mathbf{W}}_k &= \mathbf{W}_k\mathbf{W}_{k-1}\cdots\mathbf{W}_1 \\ &= \mathbf{W}_k(\mathbf{W}_k + \Delta\mathbf{W}_{k,1})\cdots(\mathbf{W}_k + \Delta\mathbf{W}_{k,k-1}) \\ &= (\mathbf{W}_k)^k + \Delta\bar{\mathbf{W}}_k, \end{aligned} \tag{5.13}$$

where $\Delta\bar{\mathbf{W}}_k = \bar{\mathbf{W}}_k - (\mathbf{W}_k)^k$ comprises products of $\mathbf{W}_k$ and $\Delta\mathbf{W}_{k,i}$ $i = 1, 2, \cdots, k-1$. Let us further assume that

$$\mathbf{U}_k^H\mathbf{W}_k\mathbf{U}_k = \mathbf{\Lambda}_k = \mathrm{diag}(\lambda_{i,k}) \tag{5.14}$$

with $\lambda_{1,k} \geq \lambda_{2,k} \geq \cdots \geq \lambda_{N,k} \geq 0$. If $\Delta\bar{\mathbf{W}}_k \to \mathbf{0}$, then $\bar{\mathbf{W}}_k$ can be approximated by $(\mathbf{W}_k)^k$ alone. Hence, we can rewrite (5.12) as

$$\mathrm{dist}(\mathfrak{D}_r(\mathbf{W}_k), \mathfrak{R}(\mathbf{Q}_k)) \leq \tan(\bar{\theta}_k)\left(\frac{\lambda_{r+1,k}}{\lambda_{r,k}}\right)^k. \tag{5.15}$$

We may interpret the above result as follows: Given *very* small variations of $\mathbf{W}_i$ for $i = 1, 2, \cdots, k$, the distance between $\mathfrak{D}_r(\mathbf{W}_k)$ and $\mathfrak{R}(\mathbf{Q}_k)$ converges to zero with a rate equal to $(\frac{\lambda_{r+1,k}}{\lambda_{r,k}})^k$ (assuming $\lambda_{r,k} > \lambda_{r+1,k}$), which is the well-known property of the subspace-tracking algorithms using orthogonal iteration.

Case 2: Moderate time variations

For moderate variations of $\mathbf{W}_i$ over $i = 1, 2, \cdots, k$, however, the above property generally does not hold anymore. We first notice from (5.12) that the orthogonal iteration attempts

to converge to $\mathfrak{D}_r(\bar{\mathbf{W}}_k)$, which may be largely different from $\mathfrak{D}_r(\mathbf{W}_k)$. Apart from this, we also wonder how the convergence rate is affected by the time variation in $\mathbf{W}_k$. To answer this question, it is of interest to view the effect $\Delta\bar{\mathbf{W}}_k$ in (5.13) as a perturbation to the matrix $(\mathbf{W}_k)^k$. Therefore, the corresponding perturbation in the eigenvalues of $(\mathbf{W}_k)^k$ can be described according to

$$\left|\lambda_i(\bar{\mathbf{W}}_k) - \lambda_i((\mathbf{W}_k)^k)\right| \leq \|\Delta\bar{\mathbf{W}}_k\|_2, \tag{5.16}$$

assuming $(\mathbf{W}_k)^k$ is normal, which is the case in applications of general interests. Fig. 5.1(a) illustrates the possible location of the perturbed eigenvalues $\lambda_{r+1}(\bar{\mathbf{W}}_k)$ and $\lambda_r(\bar{\mathbf{W}}_k)$, which are bounded by a circle of radius $\|\Delta\bar{\mathbf{W}}_k\|_2$, with centers located at $\lambda_{r+1}((\mathbf{W}_k)^k)$ and $\lambda_r((\mathbf{W}_k)^k)$, respectively. Therefore, the ratio $(\frac{\bar{\lambda}_{r+1,k}}{\bar{\lambda}_{r,k}})$ in (5.12) governing the convergence rate is now bounded by

$$\frac{\lambda_{r+1}((\mathbf{W}_k)^k) - \delta}{\lambda_r((\mathbf{W}_k)^k) + \delta} \leq \left(\frac{\bar{\lambda}_{r+1,k}}{\bar{\lambda}_{r,k}}\right) \leq \frac{\lambda_{r+1}((\mathbf{W}_k)^k) + \delta}{\lambda_r((\mathbf{W}_k)^k) - \delta}, \tag{5.17}$$

where $\delta \stackrel{\text{def}}{=} \|\Delta\bar{\mathbf{W}}_k\|_2 \geq 0$. This implies that the convergence rate may be slightly increased or decreased, depending on the specific nature of the perturbation source $\Delta\bar{\mathbf{W}}_k$.

Case 3: Large time variations

For large variations of $\mathbf{W}_i$ over $i = 1, 2, \cdots, k$, we can generally assume that $\mathfrak{D}_r(\bar{\mathbf{W}}_k)$ can be significantly different from $\mathfrak{D}_r(\mathbf{W}_k)$, potentially making the subspace tracking ineffectual. It is therefore natural to ask, to what extent can we still *track* the subspace by orthogonal iteration, given the matrices $\mathbf{W}_i$ are rapidly changing. In other words, we seek to know what is the maximum allowable time-variation of $\mathbf{W}_i$.

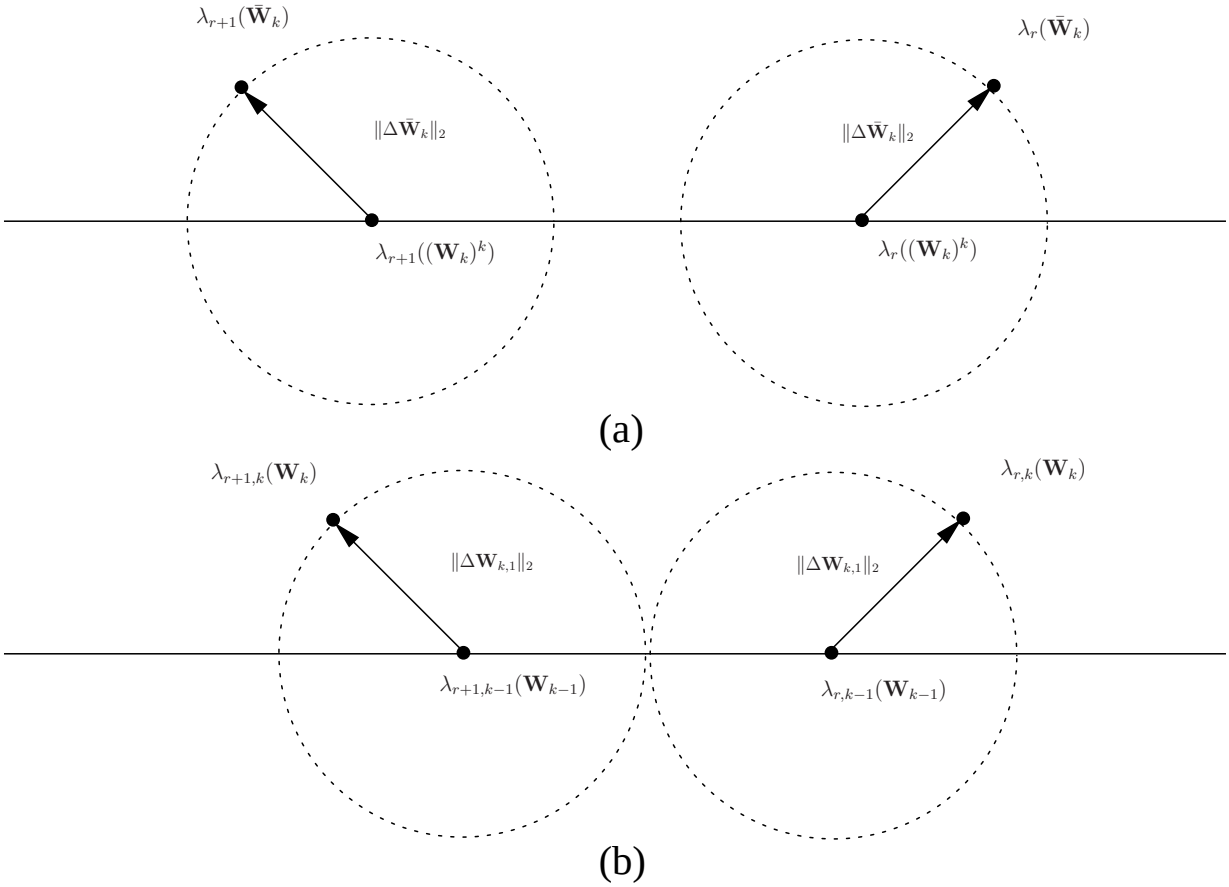


Fig. 5.1 (a)Perturbation of the eigenvalues $\lambda_{r+1}((\mathbf{W}_k)^k)$ and $\lambda_r((\mathbf{W}_k)^k)$ due to $\|\Delta\bar{\mathbf{W}}_k\|_2$. (b)Perturbation of the eigenvalues $\lambda_{r+1,k-1}(\mathbf{W}_{k-1})$ and $\lambda_{r,k-1}(\mathbf{W}_k)$ due to $\|\Delta\mathbf{W}_{k,1}\|_2$.

One possible approach is to restrain the variation from $\mathbf{W}_{k-1}$ to $\mathbf{W}_k$ to at most half the distance between the rth and $(r+1)$th eigenvalues of $\mathbf{W}_{k-1}$, i.e.,

$$\|\Delta\mathbf{W}_{k,1}\|_2 < \frac{1}{2}\,|\lambda_{r,k-1}(\mathbf{W}_{k-1}) - \lambda_{r+1,k-1}(\mathbf{W}_{k-1})|. \tag{5.18}$$

Focusing on the kth iteration alone, i.e. $\mathbf{W}_k\mathbf{Q}_{k-1} = \mathbf{Q}_k\mathbf{R}_k$, we can re-state the problem from the viewpoint of initial condition $\mathbf{Q}_{k-1}$ with one-step iteration. On the basis of earlier discussions, we know that

$$\mathrm{dist}(\mathfrak{D}_r(\mathbf{W}_k), \mathfrak{R}(\mathbf{Q}_k)) \leq \tan(\theta_{k-1}) \left(\frac{\lambda_{r+1,k}(\mathbf{W}_k)}{\lambda_{r,k}(\mathbf{W}_k)}\right),$$

where $\theta_{k-1} \in [0, \pi/2]$ is defined according to

$$\cos(\theta_{k-1}) = \min_{\mathbf{u}\in\mathfrak{D}_r(\mathbf{W}_k),\, \mathbf{v}\in\mathfrak{R}(\mathbf{Q}_{k-1})} \frac{|\mathbf{u}^H\mathbf{v}|}{\|\mathbf{u}\|_2\|\mathbf{v}\|_2}. \tag{5.19}$$

We can then clearly see that the distance between the $(r+1)$th and the rth eigenvalue of the matrix $\mathbf{W}_k$ should be maximized in order to minimizing the ratio $(\frac{\lambda_{r+1,k}(\mathbf{W}_k)}{\lambda_{r,k}(\mathbf{W}_k)})$, implying that the boundaries of the perturbed eigenvalue as illustrated in Fig. 5.1(b) should not be touching each other.

Summary

Our main observations regarding the convergence of the orthogonal iteration are summarized below:

1. *Very small time variations*: The orthogonal iteration converges toward to $\mathfrak{D}_r(\bar{\mathbf{W}}_k)$ at the rate given by (5.15).

2. *Moderate time variations*: For moderate time variations, the convergence rate may be increased or decreased according to (5.17).

3. *Large time variations*: To ensure effective tracking, the rate of change in the underlying subspace should not exceed the fundamental limit provided by (5.18).

5.4 Numerical experiments of orthogonal iteration

In order to support the above claims, we provide numerical results as follows. We start by constructing a fixed Hermitian matrix $\mathbf{W} \in \mathbb{C}^{16\times 16}$. We first show that when a noisy sample $\mathbf{W}'$ is used instead of $\mathbf{W}$ in the orthogonal iteration, the algorithm may converge to a subspace that is very different from $\mathfrak{D}_r(\mathbf{W})$. In this experiment, $\mathbf{W}'$ remains constant but is modeled as $\mathbf{W}+\Delta\mathbf{W}$ where each entry of $\Delta\mathbf{W}$ is a realization of an i.i.d. Gaussian r.v. with zero mean and variance σ^2. The experimental results for $r = 2$ and $\sigma^2 = 0, 10^{-3}, 10^{-2}, 10^{-1}$ are shown in Fig. 5.2, where the distance between $\mathfrak{D}_r(\mathbf{W})$ and $\mathfrak{R}(\mathbf{Q}_k)$, i.e.,

$$d \stackrel{\text{def}}{=} \text{dist}(\mathfrak{D}_r(\mathbf{W}), \mathfrak{R}(\mathbf{Q}_k)), \tag{5.20}$$

is plotted versus the iteration index k (each curve is averaged over 200 independent runs). As we can observe in the steady-state, the distance between $\mathfrak{D}_r(\mathbf{W})$ and $\mathfrak{R}(\mathbf{Q}_k)$ increases as σ^2 is increased, simply because the orthogonal iteration converges to $\mathfrak{D}_r(\mathbf{W}')$ instead of $\mathfrak{D}_r(\mathbf{W})$. In the context of subspace tracking a wireless channel, this situation occurs when an *estimated* correlation matrix is actually employed for the algorithm. The estimation errors can be due mainly to: insufficient time samples for the correlation matrix averaging, fast time-varying nature of the wireless channel, improper choices of parameters for the exponential or rectangular windowing, or even a combination of the above. In such cases,

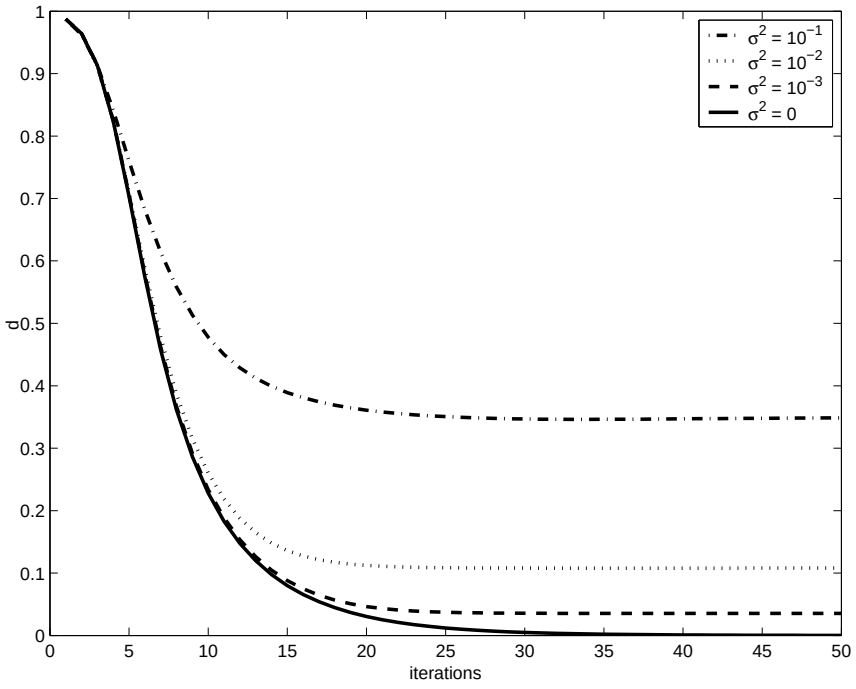


Fig. 5.2 $\mathrm{dist}(\mathfrak{D}_r(\mathbf{W}), \mathfrak{R}(\mathbf{Q}_k))$ versus the number of iterations for various σ^2's.

we are inevitably falling into the above situation.

Next, we want to show that the convergence rate in Case 2 might be slightly increased or decreased in accordance with (5.17). To this end, we consider a fixed $\mathbf{W}$ and several realizations of $\Delta\mathbf{W}$ with $\sigma^2 = 10^{-6}$ for $\mathbf{W}' = \mathbf{W} + \Delta\mathbf{W}$. In each case, we compare the convergence curves (which are now controlled by the rate factor $\lambda_{r+1}(\mathbf{W}')/\lambda_r(\mathbf{W}')$) with that of the unperturbed case (i.e., $\sigma^2 = 0$). With such a perturbation on the matrix $\mathbf{W}$, we can observe the convergence rates ranging from below to above that of the ideal case, as clearly shown in Fig. 5.3. Note that we also show the curve ρ^k along with the upper and lower bounds ρ_u^k and ρ_l^k, respectively, in the logarithmic scale for reference, where

$$\rho \stackrel{\text{def}}{=} \frac{\lambda_{r+1}(\mathbf{W})}{\lambda_r(\mathbf{W})}. \tag{5.21}$$

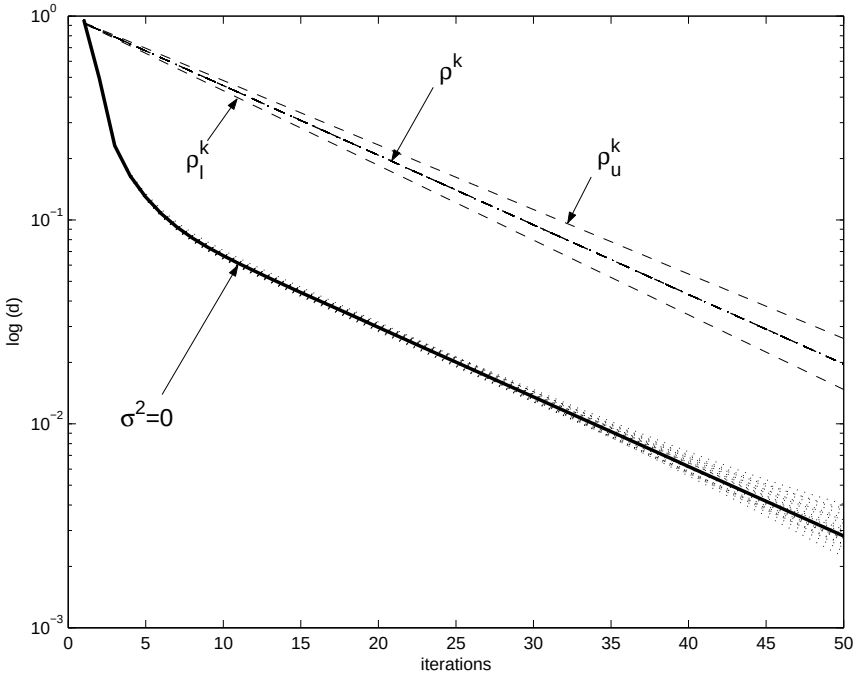


Fig. 5.3 $\log(\mathrm{dist}(\mathfrak{D}_r(\mathbf{W}_k'), \mathfrak{R}(\mathbf{Q}_k)))$ versus the number of iterations for $\sigma^2 = 0$ (solid line) and 10^{-6} (dash-dot lines).

and ρ_l, ρ_u are obtained as in (5.17), i.e.

$$\rho_l \stackrel{\text{def}}{=} \frac{\lambda_{r+1}(\mathbf{W}) - \delta}{\lambda_r(\mathbf{W}) + \delta}, \quad \rho_u \stackrel{\text{def}}{=} \frac{\lambda_{r+1}(\mathbf{W}) + \delta}{\lambda_r(\mathbf{W}) - \delta}, \tag{5.22}$$

with δ being the maximal 2-norm of the realizations of $\Delta\mathbf{W}$.

Finally, let us verify (5.18) by introducing another fixed Hermitian matrix $\Delta\mathbf{W}$. In this experiment, $\mathbf{W}' = \mathbf{W} + \beta\,\Delta\mathbf{W}$, where $\beta \in \mathbb{R}^+$ and $\|\Delta\mathbf{W}\|_2 = \|\mathbf{W}\|_2$. Given $\lambda_r(\mathbf{W}) = 7.16$, $\lambda_{r+1}(\mathbf{W}) = 6.61$ and $\|\Delta\mathbf{W}\|_2 = 9.67$, we need to have

$$\beta \leq \frac{|\lambda_r(\mathbf{W}) - \lambda_{r+1}(\mathbf{W})|}{2\|\Delta\mathbf{W}\|_2} \approx 0.028. \tag{5.23}$$

to restrain the variations. We consider a sudden change from $\mathbf{W}$ to $\mathbf{W}'$ at the 50th iteration

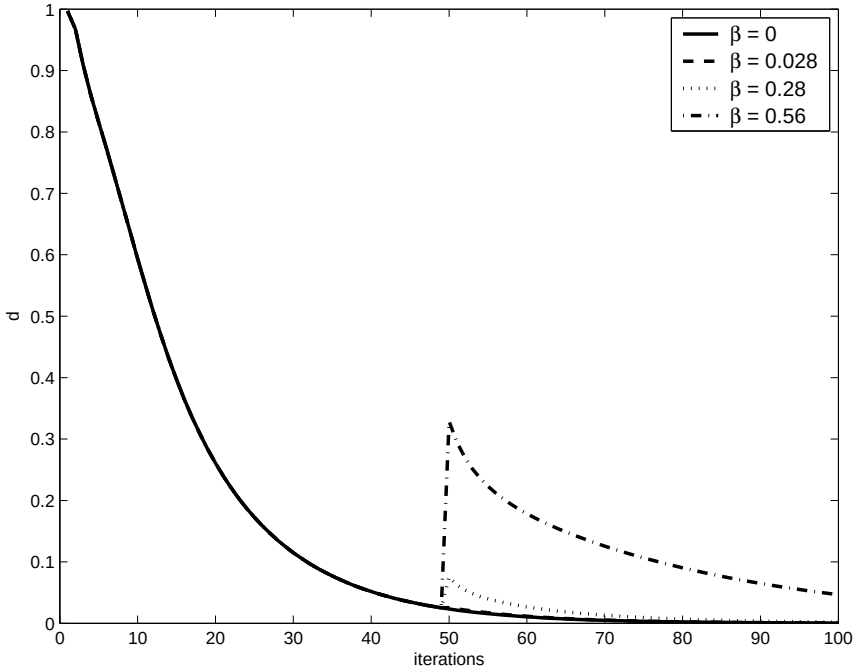


Fig. 5.4 A sudden change of $\mathbf{W}$ to $\mathbf{W}'$ at the 50th iteration.

for various β's. As can be observed in Fig. 5.4, we can see that for cases with $\beta > 0.028$, this sudden change substantially enlarges the distance between $\mathfrak{D}_r(\mathbf{W}')$ and $\mathfrak{R}(\mathbf{Q}_k)$.

For practical concerns, we also consider the popular time-varying model as mentioned in (5.2), with $\mathbf{z}_m \in \mathbb{C}^{16\times 1}$ given by

$$\mathbf{z}_m = \mathbf{H}\mathbf{x}_m + \mathbf{n}_m, \tag{5.24}$$

where $\mathbf{H} \in \mathbb{C}^{16\times 16}$ is a fixed channel matrix with $\text{rank}(\mathbf{H}) = 2$, $\mathbf{x}_m$ is an i.i.d. random vector from a QAM constellation, i.e., with entries randomly selected from $(1/\sqrt{2})(\pm 1 \pm j)$ with equal probability, and $\mathbf{n}_m$ is an i.i.d. Gaussian random vector with zero mean and variance σ_n^2. We choose $\mathbf{W}_0 = (1/500)\sum_{j=1}^{500} \mathbf{z}_j \mathbf{z}_j^H$ as our initial condition. Fig. 5.5 shows the probability $p \stackrel{\text{def}}{=} \text{prob}\left(|\lambda_{r,k-1}(\mathbf{W}_{k-1}) - \lambda_{r+1,k-1}(\mathbf{W}_{k-1})| \le 2\|\Delta\mathbf{W}_{k,1}\|_2\right)$ versus the forgetting

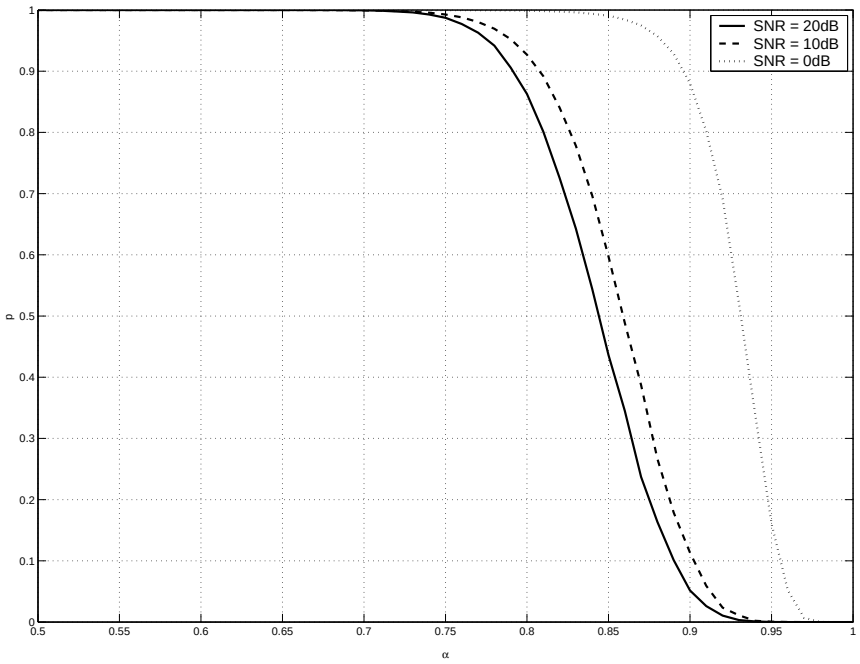


Fig. 5.5 $\text{prob}\left(|\lambda_{r,k-1}(\mathbf{W}_{k-1}) - \lambda_{r+1,k-1}(\mathbf{W}_{k-1})| \leq 2\|\Delta\mathbf{W}_k\|_2\right)$ versus the forgetting factor α in the time-varying model.

factor α for different SNR's, where $\text{SNR} \stackrel{\text{def}}{=} 10\log_{10}(1/2\sigma_n^2)$. For the case SNR = 0dB, we can observe that the probability $p = 0.5$, 0.15, and 0 when $\alpha = 0.93$, 0.95, and 0.98, respectively. Indeed, when α is close to one, the rate of change in $\mathbf{W}_m$ in (5.2) is very small, and therefore the probability of $\|\Delta\mathbf{W}_{k,1}\|_2$ exceeding the limit (5.18) is close to 0. As we increase the SNR, smaller value of α can be used at a given p level, that is, the tracking ability is improved.

To see how the forgetting factor α affects the tracking process, Fig. 5.6 - 5.8 present some realizations of d versus the number of iterations, when the above mentioned values α's are considered. From these figures, we can conclude that orthogonal iteration can only achieve satisfactory performance when the probability p is small. Hence, (5.18) can serve as a fundamental limitation to determine whether or not orthogonal iteration can be applied

in a rapidly time-varying scenario.

5.5 Summary of the convergence analysis

In this chapter, we extended the convergence analysis of orthogonal iteration from stationary cases to non-stationary ones. In particular, we investigated certain properties of orthogonal iteration when it is applied to subspace tracking of practical wireless time-varying channels. In the context of blind subspace tracking problems, we can conclude that the performance of blind channel estimation using orthogonal iteration is mainly determined by whether we can obtain a good estimate of the *time-varying* correlation matrix. In the case of moderate time variations, we showed that the rate of convergence may be increased or decreased, depending on the nature of the perturbation source. We also discussed a fundamental limitation on the use of orthogonal iteration over rapidly time-varying wireless channels. In the following, on the basis of the conclusions arrived, we present a blind subspace tracking algorithm suitable for time-varying MIMO wireless channels.

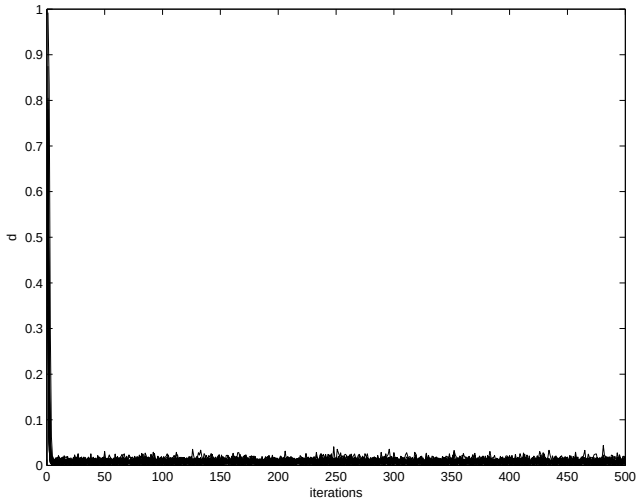


Fig. 5.6 $\mathrm{dist}(\mathfrak{D}_r(\mathbf{W}_k), \mathfrak{R}(\mathbf{Q}_k))$ versus the number of iterations in the time-varying model with $\alpha = 0.98$ when SNR = 0dB.

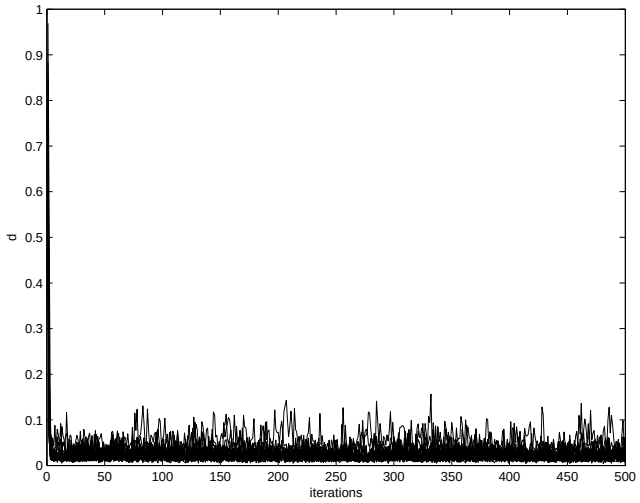


Fig. 5.7 $\mathrm{dist}(\mathfrak{D}_r(\mathbf{W}_k), \mathfrak{R}(\mathbf{Q}_k))$ versus the number of iterations in the time-varying model with $\alpha = 0.95$ when SNR = 0dB.

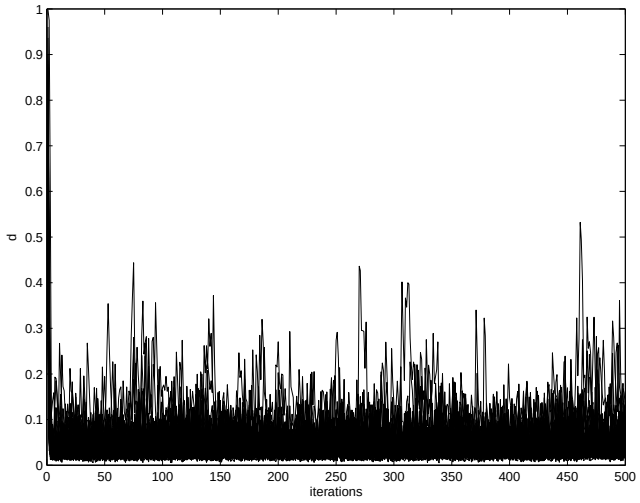


Fig. 5.8 $\mathrm{dist}(\mathfrak{D}_r(\mathbf{W}_k), \mathfrak{R}(\mathbf{Q}_k))$ versus the number of iterations in the time-varying model with $\alpha = 0.93$ when SNR = 0dB.

Chapter 6

Blind recursive subspace-based identification of time-varying wideband MIMO channels

In this chapter, we present a blind recursive algorithm for tracking rapidly time-varying wireless channels in precoded MIMO-OFDM systems. Subspace-based tracking is normally considered for slowly time-varying channels only. Thanks to the frequency correlation of the wireless channels, the proposed scheme is able to collect data not only from the time but also from the frequency domain to speed up the update of the required second order statistics. After each such update, the subspace information is recomputed using the orthogonal iteration, and then a new channel estimate is obtained. We also investigate the choice of precoder, in terms of the trade-off between the symbol recovery capability and the channel estimation performance, and demonstrate the convergence properties of our approach. In Chapter 7, the proposed algorithm will be evaluated in a 3GPP-SCM Suburban Macro scenario, in which a mobile station is allowed to move at a speed up to 100km/h. Then it will be shown that the NMSE of the channel estimates can converge to a very low level within less than 5 OFDM symbols.

To avoid the need of training sequence in the estimation of rapidly TV channels, the subspace-based blind estimation exploiting frequency correlation in MIMO-OFDM system over TI channels, that we developed in Chapter 4 (see also [148, 149]), is extended to the case of TV scenarios. While this approach in the TI case requires a larger dimension of the ambiguity matrix and a high-complexity singular value decomposition, these limitations are overcome in the TV case by using a precoder at the transmitter side, and a computationally efficient orthogonal iteration for subspace tracking at the receiver side, respectively. The resulting approach can track a fast time-varying MIMO channel in which the wireless channels may be changing at *each* OFDM symbol time. In addition, it offers the flexibility in choosing the number of transmit as well as receive antennas used (i.e. $N_T \geq N_R$ is also possible), with bandwidth efficiency approximately given by $N_T \log_2 |\mathcal{A}|$ bps/Hz, where N_T denotes the number of transmit antennas and $|\mathcal{A}|$ denotes the size of symbol alphabet used.

For a 256-point IFFT, the proposed algorithm will be evaluated in Chapter 7. Our simulation results will show that the NMSE can converge to a very low level within 5 OFDM symbols even when the maximum Doppler shift is about 230Hz, which outperforms [46, 55] in terms of estimation performance.

6.1 Problem formulation

In this section, we introduce the precoded MIMO-OFDM system model under consideration and formulate the problem of interest, i.e., the blind subspace-based estimation and tracking of the TV-MIMO channels.

6.1.1 Precoded MIMO-OFDM system with subcarrier grouping

Conventional subspace-based blind estimators are in general not favored when a fast time-varying channel is considered, since there may not be sufficient data samples to obtain the required statistics. The situation becomes even worse in the context of a MIMO-OFDM system, where a large dimension of the correlation matrix (up to thousands) is normally required. In Chapter 4, we have shown that the TI requirement in subspace-based blind channel estimation for MIMO-OFDM systems can be significantly relaxed, by making use of the subcarrier grouping to exploit the frequency correlation among adjacent subcarriers. However, this reduction in time averaging period comes at the price of a higher dimension of the ambiguity matrix. In order to overcome this problem, we consider here a precoded MIMO-OFDM system as described below.

The system under consideration employs N_C subcarriers, N_T transmit and N_R receive antennas, as per the block diagram shown in Fig. 6.1. To exploit the frequency correlation through the concept of subcarrier grouping, we assume that the frequency span of P adjacent subcarriers reside inside the coherence bandwidth of the wireless channel, defined here as the range of frequencies over which the frequency response matrix of the MIMO channel does not change appreciably [127]. As in Chapter 4, let $\Omega \stackrel{\text{def}}{=} \{0, 1, \cdots, N_C - 1\}$, i.e. the index set of the N_C subcarriers. We partition Ω into P disjoint subsets (assuming $N_C/P = \zeta \in \mathbb{Z}^+$) with the pth subset denoted as $\Omega_p \stackrel{\text{def}}{=} \{\omega_{p,1}, \omega_{p,2}, \cdots, \omega_{p,\zeta}\}$, where $\omega_{p,i} \stackrel{\text{def}}{=} p - 1 + (i-1)P$, $i = 1, 2, \cdots, \zeta$ for $p = 1, 2, \cdots, P$ (see Fig. 4.1). Let $\mathbf{x}_p^m \stackrel{\text{def}}{=} [\mathbf{x}_{1,p}^{m\ T}\ \mathbf{x}_{2,p}^{m\ T} \cdots \mathbf{x}_{N_T,p}^{m\ T}]^T$, where

$$\mathbf{x}_{j,p}^m \stackrel{\text{def}}{=} [\, x_j^m[\omega_{p,1}]\ x_j^m[\omega_{p,2}]\ \cdots\ x_j^m[\omega_{p,\zeta}]\,]^T, \tag{6.1}$$

with $x_j^m[k]$ denoting the signal transmitted at the kth subcarrier, the jth transmit an-

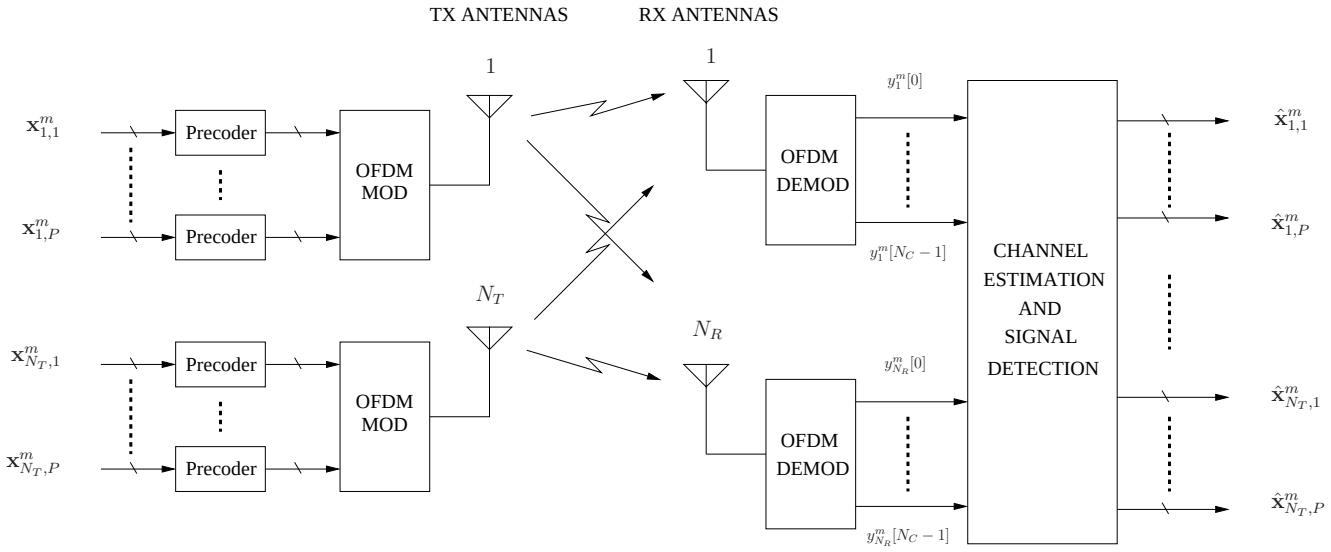


Fig. 6.1 The precoded MIMO-OFDM system model.

tenna and the mth OFDM symbol. In addition, let $\mathbf{y}_p^m \stackrel{\text{def}}{=} \left[\mathbf{y}_{1,p}^{m\,T}\ \mathbf{y}_{2,p}^{m\,T}\ \cdots\ \mathbf{y}_{N_R,p}^{m\,T}\right]^T$ and $\mathbf{n}_p^m \stackrel{\text{def}}{=} \left[\mathbf{n}_{1,p}^{m\,T}\ \mathbf{n}_{2,p}^{m\,T}\ \cdots\ \mathbf{n}_{N_R,p}^{m\,T}\right]^T$, where $\mathbf{y}_{i,p}^m \stackrel{\text{def}}{=} [y_i^m[\omega_{p,1}]\ \ y_i^m[\omega_{p,2}]\ \ \cdots\ \ y_i^m[\omega_{p,\zeta}]]^T$, $\mathbf{n}_{i,p}^m \stackrel{\text{def}}{=} [n_i^m[\omega_{p,1}]\ \ n_i^m[\omega_{p,2}]\ \ \cdots\ \ n_i^m[\omega_{p,\zeta}]]^T$, with $y_i^m[k]$ and $n_i^m[k]$ denoting the signal and noise received at the kth subcarrier, ith received antenna and the mth OFDM symbol, respectively. In the following, we assume that: (1) the length of the cyclic prefix (CP) appended to each OFDM symbol is longer than the maximum excess delay of the channel; (2) the average power of the transmit symbol alphabet is normalized to unity: $E\left[|x_q^m[k]|^2\right] = 1$.

Suppose that each input vector $\mathbf{x}_{j,p}^m$ in (6.1) is precoded by the matrix $\mathbf{\Psi} \in \mathbb{C}^{\zeta\times\zeta}$ (the choice of the non-redundant precoder matrix $\mathbf{\Psi}$ is considered in Section 6.4). Then the

input-output relationship for the pth frequency subset can be written as

$$\mathbf{y}_p^m = \mathcal{H}_p^m \left(\mathbf{I}_{N_T} \otimes \mathbf{\Psi}\right) \mathbf{x}_p^m + \mathbf{n}_p^m, \tag{6.2}$$

with the channel matrix $\mathcal{H}_p^m$ defined as

$$\mathcal{H}_p^m = \begin{bmatrix} \mathcal{H}_{1,1,p}^m & \mathcal{H}_{1,2,p}^m & \cdots & \mathcal{H}_{1,N_T,p}^m \\ \mathcal{H}_{2,1,p}^m & \mathcal{H}_{2,2,p}^m & \cdots & \mathcal{H}_{2,N_T,p}^m \\ \vdots & \vdots & \ddots & \vdots \\ \mathcal{H}_{N_R,1,p}^m & \mathcal{H}_{N_R,2,p}^m & \cdots & \mathcal{H}_{N_R,N_T,p}^m \end{bmatrix}. \tag{6.3}$$

Note that $\mathcal{H}_{i,j,p}^m = \mathrm{diag}(\mathbf{h}_{i,j,p}^m)$, where $\mathbf{h}_{i,j,p}^m \overset{\text{def}}{=} [h_{i,j}^m[\omega_{p,1}], h_{i,j}^m[\omega_{p,2}], \cdots, h_{i,j}^m[\omega_{p,\zeta}]]^T$, with $h_{i,j}^m[k]$ representing the equivalent frequency response between the ith receive and the jth transmit antenna, over the kth subcarrier and the mth OFDM symbol. In this chapter, since the frequency span of the P adjacent subcarriers have been assumed to reside inside the coherence bandwidth, we shall assume that the variations in the channel matrices $\mathcal{H}_p^m$ across these P subcarriers are negligible. Accordingly, we can define a new representative channel matrix $\mathcal{H}^m \overset{\text{def}}{=} \mathcal{H}_1^m \cong \mathcal{H}_2^m \cong \cdots \cong \mathcal{H}_P^m$, and drop the index p for all channel related quantities, including $\mathcal{H}_{i,j,p}^m$ and $\mathbf{h}_{i,j,p}^m$.

6.1.2 Problem statement

As the future generation of wireless systems aim at providing high-capacity transmission for high-mobility users, there is a strong need to further push the capability of adaptive channel tracking for wideband TV-MIMO channels, without extensively using pilot signals.

In this chapter, our first interest lies in the blind estimation/tracking of rapidly TV-MIMO channels with normalized Doppler frequencies that may reach significant values,

e.g. up to 2% or more [150], and for which the channel matrix $\mathcal{H}^m$ is allowed to change at each OFDM symbol time. To this end, we seek to develop a subspace-based blind channel estimator $\hat{\mathcal{H}}^m$, which is a function of the observed data up to the current symbol time m, i.e., $\{\mathbf{y}_p^l\}$, $p = 1, 2, \cdots, P$; $l = 1, 2, \cdots, m$, and which can be recursively updated in a computationally efficient manner. In other words, our objective is to propose an efficient updating algorithm $\phi(\cdot)$ in which the channel estimate at the mth symbol time can be represented as $\hat{\mathcal{H}}^m = \phi(\hat{\mathcal{H}}^{m-1}, \{\mathbf{y}_p^m\}_{p=1}^P)$. In addition, as the precoder is placed at the transmitter side without having any feedback of the channel knowledge from the receiver side, our second interest lies in determining the optimal precoder coefficients to further enhance the estimation performance.

6.2 Precoded subspace-based approach

In this section, we first introduce a block-based subspace channel estimation approach that exploits the frequency correlation among the adjacent subcarriers in the *precoded* MIMO-OFDM system. This approach is then extended to recursive subspace-based identification for TV channels in Section 6.3.

6.2.1 Subspace-based identification

For simplicity in notation, let us temporarily drop the time-index m of all the channel related coefficients. On the basis of (6.2) and under the assumption that channel variations over P adjacent subcarriers are negligible, the correlation matrix $\mathbf{R}_\mathbf{y} \stackrel{\text{def}}{=} E[\mathbf{y}_1\mathbf{y}_1^H] = \cdots = E[\mathbf{y}_P\mathbf{y}_P^H]$ can be written as

$$\mathbf{R}_\mathbf{y} = \mathcal{H}\left(\mathbf{I}_{N_T} \otimes \mathbf{\Psi}\mathbf{\Psi}^H\right)\mathcal{H}^H + \sigma_n^2\mathbf{I}_{\zeta N_R}, \tag{6.4}$$

where we have assumed that $\mathbf{n}_p$ and $\mathbf{x}_p$ are uncorrelated, $E[\mathbf{n}_p\mathbf{n}_p^H] = \sigma_n^2\mathbf{I}_{\zeta N_R}$, and $E[\mathbf{x}_p\mathbf{x}_p^H] = \mathbf{I}_{\zeta N_T}$. Similar to [151], by partitioning $\mathbf{R}_\mathbf{y}$ into sub-matrices of size $\zeta \times \zeta$, we may express its (u, v)th sub-matrix as

$$\mathbf{R}_{\mathbf{y},uv} = \sum_{j=1}^{N_T} \mathcal{H}_{u,j}\mathbf{\Psi}\mathbf{\Psi}^H\mathcal{H}_{v,j}^H + \delta_{uv}\sigma_n^2\mathbf{I}_\zeta \tag{6.5}$$

$$= \left(\sum_{j=1}^{N_T} \mathbf{h}_{u,j}\mathbf{h}_{v,j}^H\right) \odot \left(\mathbf{\Psi}\mathbf{\Psi}^H\right) + \delta_{uv}\sigma_n^2\mathbf{I}_\zeta, \quad u, v \in \{1, 2, \cdots, N_R\} \tag{6.6}$$

where $\delta_{uv} = 1$ if $u = v$ and zero otherwise. Let

$$\mathbf{W}_{uv} \stackrel{\text{def}}{=} \left[\mathbf{R}_{\mathbf{y},uv} - \delta_{uv}\sigma_n^2\mathbf{I}_\zeta\right] \oslash \mathbf{\Psi}\mathbf{\Psi}^H \tag{6.7}$$

be the (u, v)th sub-matrix of a new matrix $\mathbf{W}$. Then from (6.6) we can arrive at

$$\mathbf{W} = \mathbf{H}\mathbf{H}^H, \tag{6.8}$$

where

$$\mathbf{H} \stackrel{\text{def}}{=} \begin{bmatrix} \mathbf{h}_{1,1} & \mathbf{h}_{1,2} & \cdots & \mathbf{h}_{1,N_T} \\ \mathbf{h}_{2,1} & \mathbf{h}_{2,2} & \cdots & \mathbf{h}_{2,N_T} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{h}_{N_R,1} & \mathbf{h}_{N_R,2} & \cdots & \mathbf{h}_{N_R,N_T} \end{bmatrix}. \tag{6.9}$$

Assuming $\mathbf{H} \in \mathbb{C}^{(\zeta N_R)\times N_T}$ has full column rank, we can thus express the channel coefficients by means of

$$\mathbf{H} = \mathbf{Q}\mathbf{A}, \tag{6.10}$$

where the columns of the matrix $\mathbf{Q} \in \mathbb{C}^{(\zeta N_R)\times N_T}$ are obtained from the eigenvectors of the matrix $\mathbf{W}$ corresponding to the N_T largest eigenvalues, and $\mathbf{A} \in \mathbb{C}^{N_T \times N_T}$ is an ambiguity matrix.[1] As it has been shown in [151] for a similarly structured subspace problem, the matrix of interest $\mathbf{H}$ is identifiable as long as it is is a tall matrix, i.e. $\zeta N_R > N_T$. Therefore, this approach indeed offers the flexibility in choosing the number of transmit and receive antennas since $\zeta > 1$ is normally fulfilled, meaning that $N_T \geq N_R$ is also applicable.

6.2.2 Blind estimation algorithm

In practice, the channel estimate $\hat{\mathbf{H}}$ can be obtained from $\hat{\mathbf{H}} = \hat{\mathbf{Q}}\mathbf{A}$, where $\hat{\mathbf{Q}}$ denotes the matrix whose columns are the eigenvectors corresponding to the N_T largest eigenvalues of an estimated matrix $\hat{\mathbf{W}}$, with its (u, v)th sub-matrix denoted $\hat{\mathbf{W}}_{uv}$. The latter can be obtained as

$$\hat{\mathbf{W}}_{uv} = \left[\hat{\mathbf{R}}_{\mathbf{y},uv} - \delta_{uv}\hat{\sigma}_n^2\mathbf{I}_\zeta\right] \oslash \mathbf{\Psi}\mathbf{\Psi}^H, \tag{6.11}$$

where $\hat{\mathbf{R}}_{\mathbf{y},uv}$ denotes the (u, v)th sub-matrix of the sampled correlation matrix $\hat{\mathbf{R}}_{\mathbf{y}}$, and $\hat{\sigma}_n^2$ is an estimate of the noise variance. Therefore, the accuracy of the channel estimate largely depends on the estimation performance of the correlation matrix and the noise variance, i.e. $\hat{\mathbf{R}}_{\mathbf{y}}$ and $\hat{\sigma}_n^2$, respectively.

In general, to achieve satisfactory performance in the channel estimation step, i.e., $\hat{\mathbf{H}} = \hat{\mathbf{Q}}\mathbf{A}$, the time averaging period T_{av} for the above estimation of the correlation matrix $\hat{\mathbf{R}}_{\mathbf{y}}$ must be larger than ζN_R, i.e. $T_{av} \geq \zeta N_R$, since $\hat{\mathbf{R}}_{\mathbf{y}}$ is of size $\zeta N_R \times \zeta N_R$ [149, 152]. However, by exploiting the concept of frequency averaging within the coherence bandwidth, the required T_{av} can be effectively reduced by a factor P^2, i.e. $T_{av} \geq \zeta N_R/P^2$. Specifically,

[1] While the estimation of the ambiguity matrix for subspace methods is a general problem on its own that goes beyond the scope of this thesis, several approaches are available in practice to implement this step, including the use of higher order statistics or the insertion of a limited number of pilot symbols (resulting in the so-called semi-blind approach).

the estimate of the correlation matrix $\mathbf{R}_{\mathbf{y}}$ can be obtained by

$$\hat{\mathbf{R}}_{\mathbf{y}} = \frac{1}{PT_{av}} \sum_{n=1}^{T_{av}} \sum_{p=1}^{P} \mathbf{y}_p^n \mathbf{y}_p^{nH}. \tag{6.12}$$

The merits of this approach in practical MIMO-OFDM systems are demonstrated in Chapter 7 (see also [149]).

In Algorithm 1, we briefly summarize the above *precoded* subspace-based estimation, which will serve a basis in the derivation of the proposed channel tracking algorithm. Note that without employing the precoders $\mathbf{\Psi}$'s at the transmitter side, the dimension of the ambiguity matrix in [148] is $\zeta N_T \times N_T$ with $\zeta > 1$. While here, the use of the precoder matrix $\mathbf{\Psi}$ makes it possible to reduce this dimension to $N_T \times N_T$. The design of $\mathbf{\Psi}$ is addressed in Chapter 6.4.

Algorithm 1 Blind block-based subspace estimation of TI-MIMO channels

Step 1: Use the observed data to compute the estimates of $\hat{\mathbf{R}}_{\mathbf{y}}$ and $\hat{\sigma}_n^2$, and then construct $\hat{\mathbf{W}}$ by using (6.11).

Step 2: Form the matrix $\hat{\mathbf{Q}}$, whose columns are the eigenvectors which corresponds to the N_T largest eigenvalues of $\hat{\mathbf{W}}$.

Step 3: Obtain the channel estimate $\hat{\mathbf{H}} = \hat{\mathbf{Q}}\mathbf{A}$, where $\mathbf{A}$ is an ambiguity matrix.

6.3 Channel tracking

We now consider a fast time-varying scenario in which the MIMO wireless channel could be changing at each OFDM symbol time. Accordingly, we shall reintroduce the time-index m for all the channel related quantities, including those associated to the above block-based subspace estimation. Clearly, the matrix $\mathbf{Q}^m$, i.e., $\mathbf{Q}$ in (6.10) at the mth symbol time,

needs to be updated as new data samples become available to properly reflect changes in the unknown channel. Instead of applying an EVD on $\mathbf{W}^m$ at each time step, we can recursively update the EVD via an efficient subspace tracking algorithm so as to minimize the amount of computations involved. We notice that most fast subspace trackers with low complexity assume a rank one update [137, 138], and hence are not applicable here. On the basis of (6.10), we propose a new algorithm that combines the well-known orthogonal iteration with a joint time-frequency averaging to track the above mentioned time-varying channel, without incurring EVD operations repeatedly.

6.3.1 Recursive approach based on orthogonal iteration

Orthogonal iteration and its variants have been considered for blind adaptive estimation to a great extent (see [55, 145] and references therein). To track the fast time-varying channel $\mathcal{H}^m$ with low complexity, we also consider to recursively update $\mathbf{Q}^m$ by employing orthogonal iteration, which is known to converge exponentially with arbitrary initial conditions [140], and suitable for real-time processing because it is well structured. However, the main challenge still lies in whether we can estimate the required second order statistics within a sufficiently short processing window.

Since the frequency response matrices of the MIMO channel can be related by $\mathcal{H}^m \cong \mathcal{H}_p^m$, an estimate of the *time-varying* correlation matrix at the mth OFDM symbol time can be obtained by combining traditional window-based time averaging with frequency averaging over the P frequency subsets Ω_p, for $p \in \{1, \cdots, P\}$. This results into

$$\hat{\mathbf{R}}_{\mathbf{y}}^m = \sum_{n=m-l+1}^{m} \sum_{p=1}^{P} \beta^{m-n} \mathbf{y}_p^n \mathbf{y}_p^{nH} \tag{6.13}$$

$$= \beta \hat{\mathbf{R}}_{\mathbf{y}}^{m-1} + \sum_{p=1}^{P} \mathbf{y}_p^m \mathbf{y}_p^{mH} - \sum_{p=1}^{P} \beta^l \mathbf{y}_p^{m-l} \mathbf{y}_p^{m-l^H}, \tag{6.14}$$

where $l \in \mathbb{N}$ and $0 \leq \beta \leq 1$ denotes the window length and the forgetting factor, respectively. Considering a scenario in which no windowing is applied, i.e. $l = 1$, we can still collect $\{\mathbf{y}_p^m\}$, $p = 1, 2, \cdots, P$ at the mth OFDM symbol time, without referring to the OFDM symbols of the previous time instances, i.e. $\mathbf{y}_p^n$ for $n < m$. Hence, we can conclude that it is possible to track the fast time-varying channel provided $P > \zeta N_R$. In practice, this condition is not stringent, e.g. the choices $(P, \zeta N_R) = (32, 24)$ and $(64, 12)$, both fulfilling $P > \zeta N_R$, were reported in [149], where both the WiMAX specification and the 3GPP Spatial Channel Model (SCM) are considered. Of course, the window length l can be increased if the condition is not met or if it is desired to obtain better smoothing of the channel estimate. The choice of the parameter β and l is further discussed along with the presentation of our simulation results in Chapter 7.

Let $\hat{\mathbf{W}}^m$ be an estimate of the matrix $\mathbf{W}^m$, with its (u, v)th sub-matrix given as

$$\hat{\mathbf{W}}_{uv}^m \stackrel{\text{def}}{=} \left[\hat{\mathbf{R}}_{\mathbf{y},uv}^m - \delta_{uv}\hat{\sigma}_n^2\mathbf{I}_\zeta\right] \oslash \mathbf{\Psi}\mathbf{\Psi}^H. \tag{6.15}$$

In this work, we propose to recursively update the principal eigenvectors of $\hat{\mathbf{W}}^m$ using orthogonal iteration and use them to estimate the unknown channel matrix $\hat{\mathbf{H}}^m$ by

$$\hat{\mathbf{H}}^m = \hat{\mathbf{Q}}_{n_d}^m \mathbf{A}^m. \tag{6.16}$$

In (6.16), the columns of $\hat{\mathbf{Q}}_{n_d}^m$ are the *approximate* principal eigenvectors of $\hat{\mathbf{W}}^m$ resulting from the application of the n_dth orthogonal iteration at the mth OFDM symbol time, and $\mathbf{A}^m$ represents the corresponding ambiguity matrix. The details of iteration process are summarized as follows.

Given a tall, column orthonormal matrix $\hat{\mathbf{Q}}_0^m \in \mathbb{C}^{\zeta N_R \times N_T}$ at the mth OFDM symbol

time, the method of *orthogonal iteration* generates a sequence of matrices $\hat{\mathbf{Q}}_{\mu}^{m}$, whose column span is assumed to approximate the span of the N_T-dimensional dominant subspace of the matrix $\hat{\mathbf{W}}^{m} \in \mathbb{C}^{\zeta N_R \times \zeta N_R}$, according to the following recurrence:

$$\hat{\mathbf{Z}}_{\mu}^{m} = \hat{\mathbf{W}}^{m} \hat{\mathbf{Q}}_{\mu-1}^{m}, \quad \mu = 1, 2, \cdots, n_d, \tag{6.17}$$

$$\hat{\mathbf{Q}}_{\mu}^{m} \hat{\mathbf{R}}_{\mu}^{m} = \hat{\mathbf{Z}}_{\mu}^{m} \quad \text{(QR decomposition)}. \tag{6.18}$$

Note that in practice we choose $\hat{\mathbf{Q}}_{0}^{m} = \hat{\mathbf{Q}}_{n_d}^{m-1}$ except when $m = 0$ (the initial condition). On the basis of orthogonal iteration, we also notice that estimation of the dominant subspace of a slowly time-varying correlation matrix was considered in [147]; here we extend the use of orthogonal iteration by allowing $n_d \geq 1$ for tracking a time-varying MIMO channel.

6.3.2 Convergence properties

To motivate the use of the proposed recursive method in a fast time-varying wireless channel, we investigate its convergence properties as follows. Let us first assume that

$$\mathbf{U}^{mH} \hat{\mathbf{W}}^{m} \mathbf{U}^{m} = \mathbf{\Lambda}^{m} = \text{diag}(\lambda_i^m) \tag{6.19}$$

is an EVD of $\hat{\mathbf{W}}^{m}$ with $\lambda_1^m \geq \lambda_2^m \geq \cdots \geq \lambda_{\zeta N_R}^m \geq 0$ and $\mathbf{U}^{mH}\mathbf{U}^{m} = \mathbf{U}^{m}\mathbf{U}^{mH} = \mathbf{I}$. If $\mathbf{U}^{m}$ is partitioned as $\mathbf{U}^{m} = [\mathbf{U}_1^m \ \mathbf{U}_2^m]$, where $\mathbf{U}_1^m \in \mathbb{C}^{\zeta N_R \times N_T}$ and $\mathbf{U}_2^m \in \mathbb{C}^{\zeta N_R \times (\zeta N_R - N_T)}$, we can define the distance between the two subspaces $\mathfrak{D}_{N_T}(\hat{\mathbf{W}}^{m})$ and $\mathfrak{R}(\hat{\mathbf{Q}}_{\mu}^{m})$ according to [140] by

$$\text{dist}\left(\mathfrak{D}_{N_T}(\hat{\mathbf{W}}^{m}), \mathfrak{R}(\hat{\mathbf{Q}}_{\mu}^{m})\right) = \|(\mathbf{U}_2^m)^{H} \hat{\mathbf{Q}}_{\mu}^{m}\|_2. \tag{6.20}$$

Let the angle $\theta^m \in [0, \pi/2]$ be defined to provide a measure of the closeness of the two subspaces $\mathfrak{D}_{N_T}(\hat{\mathbf{W}}^m)$ and $\mathfrak{R}(\hat{\mathbf{Q}}_0^m)$ by means of (see also (5.11) for details)

$$\cos(\theta^m) \stackrel{\text{def}}{=} \min_{\mathbf{u}\in\mathfrak{D}_{N_T}(\hat{\mathbf{W}}^m),\, \mathbf{v}\in\mathfrak{R}(\hat{\mathbf{Q}}_0^m)} \frac{|\mathbf{u}^H\mathbf{v}|}{\|\mathbf{u}\|_2\|\mathbf{v}\|_2}, \tag{6.21}$$

where $\hat{\mathbf{Q}}_0^m$ represents the initial condition at the mth OFDM symbol time. Then from preceding as in Chapter 5, we can similarly arrive at

$$\text{dist}(\mathfrak{D}_{N_T}(\hat{\mathbf{W}}^m), \mathfrak{R}(\hat{\mathbf{Q}}_\mu^m)) \leq \tan(\theta^m) \left(\frac{\lambda_{N_T+1}^m}{\lambda_{N_T}^m}\right)^\mu, \quad \mu = 1, 2, \cdots, n_d. \tag{6.22}$$

According to (6.22), as long as the ratio $(\lambda_{N_T+1}^m/\lambda_{N_T}^m) < 1$, the iterated subspace $\mathfrak{R}(\hat{\mathbf{Q}}_\mu^m)$ converges to $\mathfrak{D}_{N_T}(\hat{\mathbf{W}}^m)$ exponentially with an arbitrary initial condition $\hat{\mathbf{Q}}_0^m = \hat{\mathbf{Q}}_{n_d}^{m-1}$ (except when $m = 0$, i.e. the initial condition); this crucial factor indeed allows the tracking of a fast time-varying channel. The convergence behavior of the orthogonal iteration scheme (6.17)-(6.18) as a function of μ is well predicted by (6.22) in the current application. To illustrate this point, Fig. 6.2 shows a plot of the subspace distance (6.20) as a function of μ when the orthogonal iteration (6.17)-(6.18) is used to approximate the 2-dimensional dominant subspace of a particular matrix $\hat{\mathbf{W}}_m \in \mathbb{C}^{12\times 12}$, appearing at a given symbol time m in one of our simulations; we also show a plot of $(\lambda_3(\hat{\mathbf{W}}^m)/\lambda_2(\hat{\mathbf{W}}^m))^\mu$ for reference.

Although the estimation performance can be improved by increasing n_d in a general sense, the iterated subspace actually converges to $\mathfrak{D}_{N_T}(\hat{\mathbf{W}}^m)$ instead of $\mathfrak{D}_{N_T}(\mathbf{W}^m)$. Therefore, the performance largely depends on whether or not we can obtain a good estimate of $\mathbf{W}^m$ at each OFDM symbol time. Thanks to the use of additional frequency domain samples as shown in (6.13)-(6.14), we can meet the requirement of the minimum number of data samples (i.e., the dimension of the correlation matrix) even in a fast time-varying

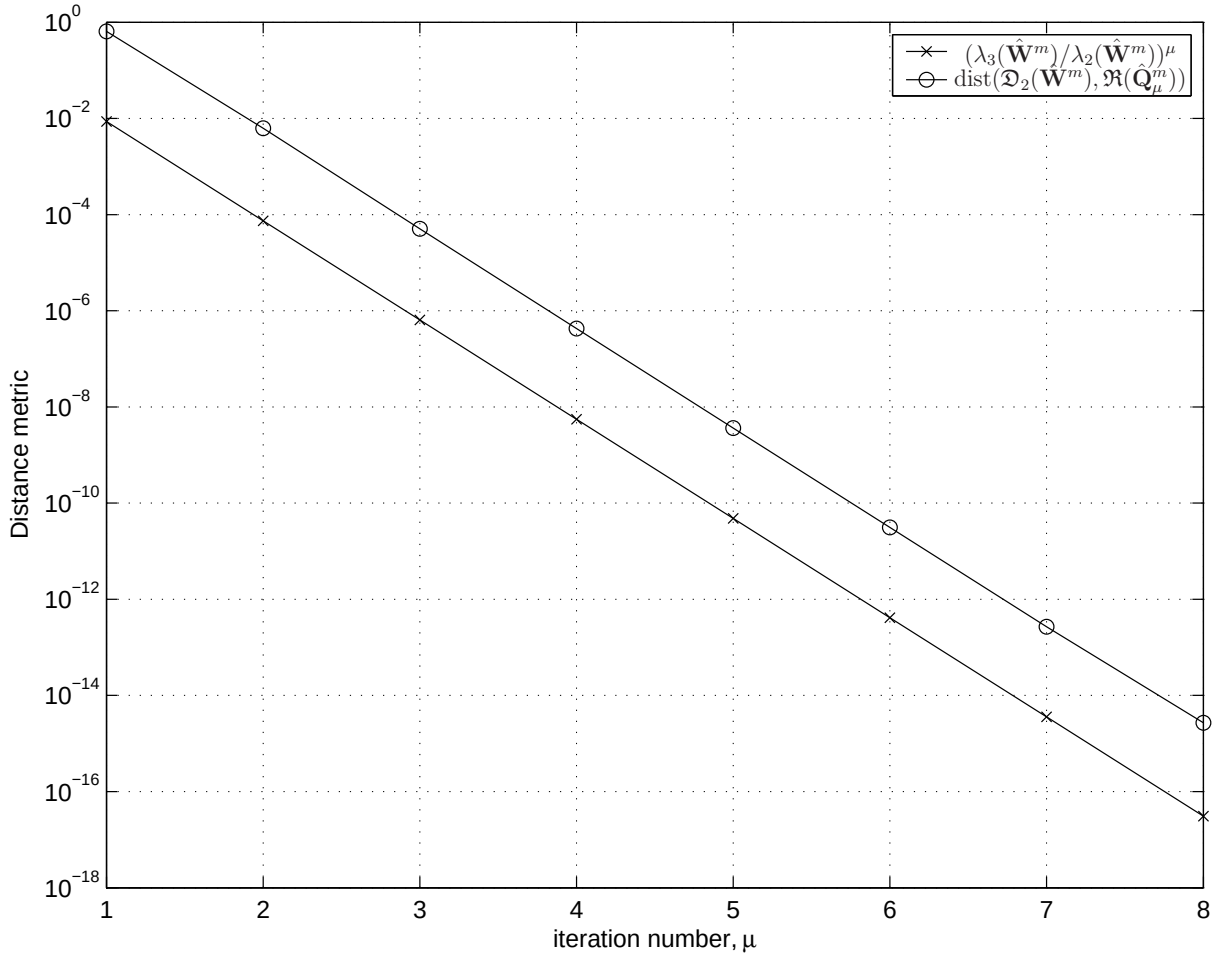


Fig. 6.2 A demonstration on the rate of convergence in subspace estimation by using orthogonal iteration.

wireless channel, as will be demonstrated in Chapter 7. Otherwise, the orthogonal iteration may converge exponentially to a wrong subspace.

6.3.3 Proposed blind recursive estimation algorithm

We briefly summarize the precoded subspace-based tracking algorithm in Algorithm 2. In practice, to acquire the ambiguity matrix $\mathbf{A}^m$ needed at the mth OFDM symbol time, we can employ another estimation process including the use of higher order statistics or pilot symbols (resulting in the so-called semi-blind approach). Note that for moderate choices of the number of transmit and receive antennas, the computational complexity (flops) of the proposed tracking algorithm for each iteration is $\mathcal{O}(\zeta^4)$. This figure is generally smaller than that of the blind adaptive channel estimators using either the least mean squares or the recursive least squares adaption, developed for SISO-OFDM [55].

6.4 Precoder design

To simplify the notation for the following discussions, let us define $\mathbf{\Gamma} \stackrel{\text{def}}{=} \mathbf{\Psi}\mathbf{\Psi}^H$ and let $\psi_{i,j}$ and $\gamma_{i,j}$ denote the (i,j)th entry of matrices $\mathbf{\Psi}$ and $\mathbf{\Gamma}$, respectively. On the basis of (6.15), the choice of $\mathbf{\Psi}$ does not appear to be restricted, except for the trivial constraint that the entries of $\mathbf{\Gamma}$ cannot be zeros, i.e., $\gamma_{i,j} \neq 0$, $\forall i,j$. However, we can choose the precoder matrix judiciously to simplify the channel estimator and optimize its performance.

First, we note that if the diagonal entries of $\mathbf{\Gamma}$ are identical, i.e., $\gamma_{i,i}$ is a constant for $i = 1, 2, \cdots, \zeta$, then the additional estimation of the noise variance in (6.15) can be avoided. To be more specific, let us define a new matrix $\hat{\mathbf{T}}^m$ with its (u,v)th sub-matrix given as

$$\hat{\mathbf{T}}^m_{uv} \stackrel{\text{def}}{=} \hat{\mathbf{R}}^m_{\mathbf{y},uv} \oslash \mathbf{\Gamma}, \quad u,v \in \{1,2,\cdots,N_R\}. \tag{6.23}$$

Algorithm 2 Blind recursive subspace-based identification of TV-MIMO channels

Initialization: $\hat{\mathbf{Q}}_{n_d}^{l-1} = \mathbf{I}(:, 1:N_T)$, $\hat{\mathbf{R}}_{\mathbf{y}}^{0} = \mathbf{0}$, $\tilde{\mathbf{R}}_{\mathbf{y}}^{0} = \mathbf{0}$

for $m = 1, 2, \cdots$ **do**
  Input vector: $\mathbf{y}_1^m, \cdots, \mathbf{y}_P^m$
  $\tilde{\mathbf{R}}_{\mathbf{y}}^{m} = \sum_{p=1}^{P} \mathbf{y}_p^m \mathbf{y}_p^{mH}$
  if $m < l$ **then**
    $\hat{\mathbf{R}}_{\mathbf{y}}^{m} = \beta \hat{\mathbf{R}}_{\mathbf{y}}^{m-1} + \tilde{\mathbf{R}}_{\mathbf{y}}^{m}$
  else
    $\hat{\mathbf{R}}_{\mathbf{y}}^{m} = \beta \hat{\mathbf{R}}_{\mathbf{y}}^{m-1} + \tilde{\mathbf{R}}_{\mathbf{y}}^{m} - \beta^l \tilde{\mathbf{R}}_{\mathbf{y}}^{m-l}$
    $\hat{\mathbf{W}}_{uv}^{m} = \left[\hat{\mathbf{R}}_{\mathbf{y},uv}^{m} - \delta_{uv}\hat{\sigma}_n^2 \mathbf{I}\right] \oslash \mathbf{\Psi}\mathbf{\Psi}^H$
    $\hat{\mathbf{Q}}_0^m = \hat{\mathbf{Q}}_{n_d}^{m-1}$
    for $\mu = 1, 2, \cdots, n_d$ **do**
      $\hat{\mathbf{Z}}_{\mu}^{m} = \hat{\mathbf{W}}^m \hat{\mathbf{Q}}_{\mu-1}^{m}$
      $\hat{\mathbf{Q}}_{\mu}^{m} \hat{\mathbf{R}}_{\mu}^{m} = \hat{\mathbf{Z}}_{\mu}^{m}$ (QR factorization on $\hat{\mathbf{Z}}_{\mu}^{m}$)
    end for
    $\hat{\mathbf{H}}^m = \hat{\mathbf{Q}}_{n_d}^{m} \mathbf{A}^m$
  end if
end for

Then we can arrive at $\hat{\mathbf{T}}^m = \hat{\mathbf{W}}^m + \rho \mathbf{I}$ (for some $\rho \in \mathbb{R}$). Since $\hat{\mathbf{T}}^m$ has the same invariant subspaces as $\hat{\mathbf{W}}^m$, we can simply apply $\hat{\mathbf{T}}^m$ instead of $\hat{\mathbf{W}}^m$ in the Algorithm 2 to eliminate the noise variance estimation.

Second, letting $\Delta\mathbf{R}_{\mathbf{y}}^{m}$ denote the difference between the estimated and the *true* correlation matrix, i.e. $\hat{\mathbf{R}}_{\mathbf{y}}^{m} = \mathbf{R}_{\mathbf{y}}^{m} + \Delta\mathbf{R}_{\mathbf{y}}^{m}$, we may express $\hat{\mathbf{T}}_{uv}^{m}$ as follows

$$\hat{\mathbf{T}}_{uv}^{m} = \underbrace{\mathbf{R}_{\mathbf{y},uv}^{m} \oslash \mathbf{\Gamma}}_{\stackrel{\text{def}}{=}\mathbf{T}_{uv}^{m}} + \underbrace{\Delta\mathbf{R}_{\mathbf{y},uv}^{m} \oslash \mathbf{\Gamma}}_{\stackrel{\text{def}}{=}\Delta\mathbf{T}_{uv}^{m}}. \tag{6.24}$$

Then it becomes clear that the choice of the precoder should focus on eliminating the error term $\Delta\mathbf{T}_{uv}^{m}$ in (6.24).

The matrix $\Delta\mathbf{R}_{\mathbf{y},uv}^{m}$ in (6.24) has a random nature resulting from the effects of the

time-varying channels, additive noise, and insufficient number of data samples. Let

$$J(\mathbf{\Psi}) \stackrel{\text{def}}{=} \sum_{u,v} E\|\Delta\mathbf{T}_{uv}^{m}\|_F^2 = \sum_{u,v} E\|\Delta\mathbf{R}_{\mathbf{y},uv}^{m} \oslash \mathbf{\Psi}\mathbf{\Psi}^H\|_F^2. \tag{6.25}$$

Given that $\oslash$ is an element-wise division, the minimization of $J(\mathbf{\Psi})$ is equivalent to maximizing (in a weighted sense) the entries of $\mathbf{\Gamma} = \mathbf{\Psi}\mathbf{\Psi}^H$. Nevertheless, the choice of a precoder is subject to a fixed transmit power and thus cannot be arbitrarily large; the entries of $\mathbf{\Gamma}$ should therefore be maximized based on the statistics of $\Delta\mathbf{R}_{\mathbf{y}}^{m}$ *and* subject to a normalization factor.

In summary, we suggest that the choice of the precoder be optimized via the objective function

$$\min_{\mathbf{\Psi}} J(\mathbf{\Psi}), \tag{6.26}$$

subject to the following constraints:

(C1) To guarantee that the element-wise division in (6.15) and (6.23) is feasible, the precoder must fulfill the trivial condition: $\gamma_{i,j} \neq 0$, $\forall i, j$.

(C2) To normalize the average transmit power, we require that $\sum_j |\psi_{i,j}|^2 = 1$, $\forall\, i$. Note that this constraint also implies that $\gamma_{i,i} = 1$, $\forall\, i$, meaning that the diagonal entries of $\mathbf{\Gamma}$ are identical, and hence there is no need for noise variance estimation.

In the absence of a more specific model, we consider a worst case situation and assume that the entries of $\Delta\mathbf{R}_{\mathbf{y}}^{m}$ are i.i.d. random variables with zero mean and equal variance; this choice is further supported by our numerical observations. Based on this assumption, the objective function in (6.26) becomes a standard optimization problem and can be solved by using Lagrange multiplier. Accordingly, the optimal precoder assuming $\gamma_{i,j} \in \mathbb{R}^+$ is obtained as $\mathbf{\Psi}_0 = (1/\sqrt{\zeta})\mathbf{1}_{\zeta\times\zeta}$, that is $\mathbf{\Gamma} = \mathbf{1}_{\zeta\times\zeta}$, where $\mathbf{1}_{\zeta\times\zeta}$ denotes a $\zeta \times \zeta$ matrix of all

ones. Note that this result coincides with the optimal choice of the precoder in terms of estimation performance, obtained from numerical results in [125].

Nevertheless, the above precoder $\mathbf{\Psi}_0$ has rank 1 (condition number $= \infty$), and thus is not a good choice from the perspective of symbol recovery. To make $\mathbf{\Psi}_0$ non-singular while keeping the estimation performance close to the optimum, we can perturb the entries of $\mathbf{\Psi}_0$ in the following manner: $\mathbf{\Psi}_0 \rightarrow \mathbf{\Psi}$, where the diagonal entries of $\mathbf{\Psi}$ now slightly exceed the off-diagonal ones. This approach is motivated by the following fact: a loose bound on $\mathrm{rank}(\mathbf{\Psi})$ is provided by [153, Section 4.10] as

$$\mathrm{rank}(\mathbf{\Psi}) \geq \sum_{i=1}^{\zeta} |\psi_{i,i}|/b_i, \tag{6.27}$$

where $b_i \stackrel{\text{def}}{=} \sum_{j=1}^{\zeta} |\psi_{i,j}|$. This implies that given the constraint (C2), we can increase $\mathrm{rank}(\mathbf{\Psi})$ from 1 by boosting the ratios $|\psi_{i,i}|/b_i$. Here we propose to use a simple Toeplitz matrix to accomplish this goal. That is, we define

$$\mathbf{\Psi} = \mathbf{\Psi}(\nu) \stackrel{\text{def}}{=} \frac{1}{\sqrt{1+(\zeta-1)\nu^2}} \begin{bmatrix} 1 & \nu & \cdots & \nu \\ \nu & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \nu \\ \nu & \cdots & \nu & 1 \end{bmatrix}_{\zeta \times \zeta}, \tag{6.28}$$

where ν $(0 < \nu \leq 1)$ can be seen as the common perturbed value of the off-diagonal entries of $\mathbf{\Psi}_0$. The condition number of $\mathbf{\Psi}(\nu)$ is given by $\kappa = (1+(\zeta-1)\nu)/(1-\nu)$ [154]. We can now impose some constraint on the condition number, e.g., $\kappa \leq \kappa^*$ for some practical but finite κ^* and relate the choice of ν to κ^* as $\nu \leq (\kappa^*-1)/(\zeta-1+\kappa^*)$.

Intuitively, there exists an optimal trade-off in terms of ν between the symbol recovery

($\nu < 1$) and channel estimation performance ($\nu = 1$) for a given SNR. However, the analysis for determining an optimal value of ν for this combined objective appears difficult. In the following chapter, we shall approach this problem from an experimental perspective using simulations. Note that a similar structure was employed for a *block-based* channel estimation scenario in [151] (i.e. quasi-stationary over several OFDM symbols). Our analysis explicitly shows that the optimal estimation performance is achieved when $\nu = 1$ and we provide insight into the trade-off among various choices of ν's.

Chapter 7

Numerical experiments and results

In this chapter, we present and discuss the results of numerical experiments aimed at evaluating the performance of the proposed algorithms developed earlier. Both time-invariant and time-variant wireless channels are considered in the evaluations.

7.1 Time-invariant scenarios

7.1.1 Methodology

Numerical evaluations of the performance of the proposed algorithm over block fading channels are presented in this section. A comparison of the proposed and referenced algorithms over a simplified TDL model will be given in the first place. Having demonstrated that better performance can readily be achieved by the proposed scheme under such a condition, we will then explore the performance of the proposed algorithm over practical scenarios where both WiMAX specification [70] and 3GPP Spatial Channel Model (SCM) [79] are considered. We refer the reader to Section 2.3 for further details on these channel models. For each 3GPP-SCM scenario, after identifying the best value of the parameter P (i.e. the number of frequency subsets for averaging) in the proposed algorithm presented in Ta-

ble 4.1, we will compare their NMSEs with the corresponding CRBs. In our experiments, the NMSE of the channel estimate is defined by

$$\text{NMSE} = \frac{1}{N_T N_R N_C} \sum_{i,j,k} \frac{E[|h_{i,j}[k] - \hat{h}_{i,j}[k]|^2]}{E[|h_{i,j}[k]|^2]}, \tag{7.1}$$

where we consider $\hat{h}_{i,j}[\omega_{1,i}] = \cdots = \hat{h}_{i,j}[\omega_{P,i}]$, $i = 1, 2, \cdots, \zeta$ (see Fig. 4.1 for more details). Note that the asymptotic performance of one of the most recent algorithms [47] will also be given over these practical scenarios for comparisons.

Throughout this section, we consider a MIMO-OFDM system with 2 transmit ($N_T = 2$) and 3 receive antennas ($N_R = 3$). The number of subcarriers used in the OFDM system is 256 ($N_C = 256$). For each time epoch, the incoming symbol streams are independent and identically distributed (i.i.d.) QPSK symbols. The SNR is defined as the ratio of the signal power to the noise power on a subcarrier basis. By referring to (3.1), we can also express it by $10 \log_{10}[E_s(\sigma_h^2/\sigma_n^2)]$, where σ_h^2 denotes the variance of the channel coefficients. All simulation results are obtained by averaging over 200 independent Monte Carlo runs except when evaluating the BER and CRBs, in which the latter are obtained by averaging over 10^7 independent Monte Carlo runs instead. In addition, the wireless channel is assumed to remain stationary over the time averaging intervals, and we employ $\mathbf{A} = (\hat{\mathbf{Q}}_o^*)^\dagger \bar{\mathcal{H}}'$ from (4.15) to obtain the ambiguity matrix, assuming $\bar{\mathcal{H}}'$ is known.

7.1.2 Comparison with referenced schemes

Numerical results of the proposed as well as the referenced subspace-based methods from [44], including the CP and VC approaches for MIMO-OFDM systems, are presented in this part. For each time epoch, the incoming QPSK symbols are chosen to span 2 OFDM symbols ($N_F = 2$) in order to fulfill the identifiability condition of the referenced schemes.

For the proposed method, we consider $P = 2$, 8, 32, and 64.

In order to save simulation time and overcome the issue of an accurate channel-order estimation [155] for the referenced schemes, we consider a simplified TDL model with 2 taps. Therefore, the excess delays are given as $\tau_n = (n-1)T$, where T denotes the OFDM sampling time interval. The tap coefficients are assumed to be i.i.d., ZMCSCG random variables with unit variance for both the real and imaginary parts. Under these conditions, there are 10 subcarriers residing inside the coherence bandwidth (P=10) if the latter is defined as the bandwidth over which the frequency correlation function is above 0.9, while there are 100 subcarriers residing inside the coherence bandwidth (P=100) if the definition is relaxed so that the frequency correlation function is above 0.5.[1]

Fig. 7.1 shows the NMSEs of the proposed and referenced methods as a function of the numbers of OFDM blocks, employed to obtain a sampled correlation matrix (each OFDM block is constituted of 2 OFDM symbols) when SNR = 20dB. As expected, the estimation performance generally improves when the number of the OFDM blocks is increased for time averaging. An exception to this is when the proposed method is with $P = 32$ and 64 as shown in Fig. 7.2, in which cases the frequency responses of the P sub-channels are not *flat* in any sense.

For each referenced method, we consider a dimension of the noise subspace equals to either 8 or 16.[2] When comparing a specific referenced method with different dimension of the noise subspace, either the CP or VC method with a larger dimension of the noise subspace outperforms the same method with a smaller dimension of the noise subspace.

[1] Under this scenario, the RMS delay spread τ_{rms} can be calculated as $T/2$ from (2.13). Since the OFDM subcarrier spacing $\Delta f = 1/(N_C T) = 1/(256T)$, we can arrive at $B_c \approx 1/(5\tau_{rms}) = 0.4/T \approx 100\Delta f$ when the frequency correlation function is above 0.5, and $B_c \approx 1/(50\tau_{rms}) = 0.04/T \approx 10\Delta f$ when the frequency correlation function is above 0.9, respectively.

[2] For the CP method, the dimension of the noise subspace is equal to the size of the cyclic prefix. For the VC method, the dimension of the noise subspace is equal to the sum of the size of the cyclic prefix and the size of nulls.

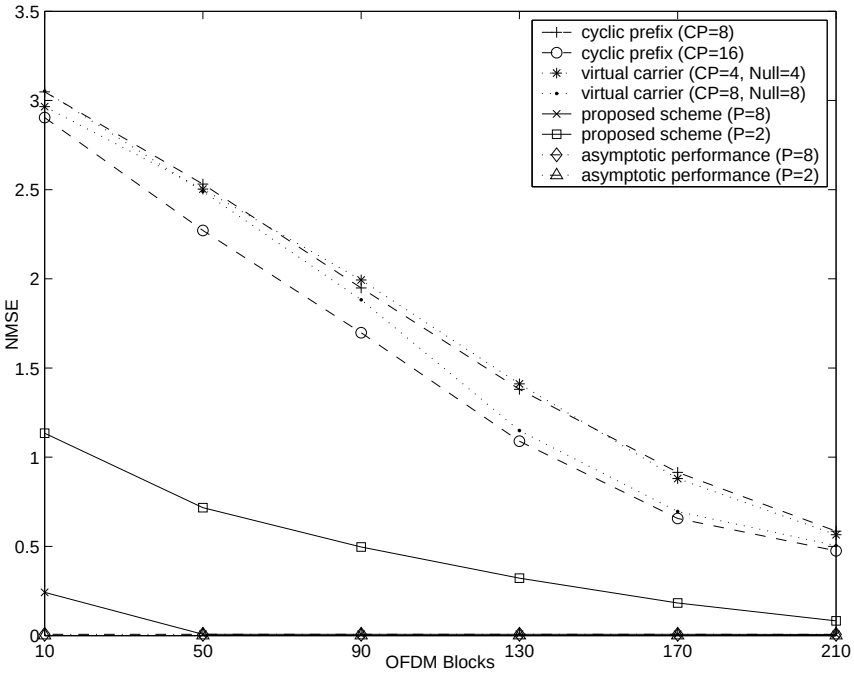


Fig. 7.1 NMSE versus number of OFDM blocks (SNR=20dB).

When comparing two referenced methods with the same dimension of the noise subspace, the CP method outperforms the VC method since the dimension of the CP's eigenvectors is larger, imposing additional constraints on the channel estimate [42, 43].

Within the given time averaging intervals, we note that the proposed method outperforms the referenced ones with any given dimension of the noise subspace. We can also observe from Fig. 7.1 and Fig. 7.2 that the number of the time samples required (i.e., the dimension of the correlation matrix) is reduced when P is increased. However, the estimation results also deteriorate since the proposed algorithm is based on the assumption that the channel coefficients of the adjacent P subcarriers are similar. On the contrary, the number of the time samples required is increased when P is decreased. Nevertheless, the estimation results also improve. To achieve the best tradeoff, we conclude that P should

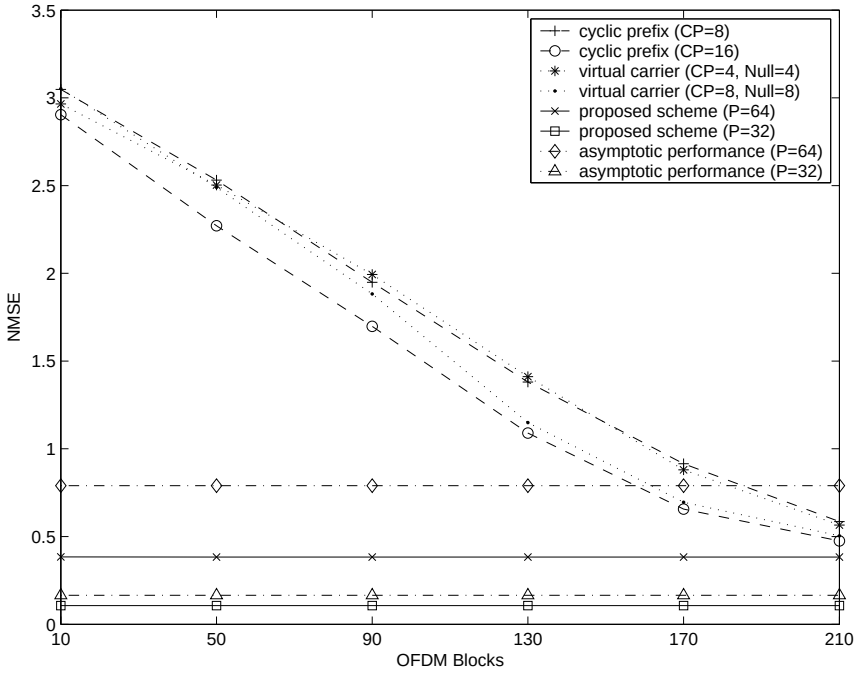


Fig. 7.2 NMSE versus number of OFDM blocks (SNR=20dB).

be chosen to restrict the frequency spans of the P adjacent subcarriers to reside inside the coherence bandwidth, for which the frequency correlation function is above 0.9, i.e. $P \approx 10$ in this case.

Fig. 7.1 and 7.2 also include asymptotic performance of the proposed scheme, as defined in (4.28). We do not evaluate the CRB here since the simplified TDL model is a less practical scenario. To evaluate the asymptotic bound for different values of P, an estimate of $\mathbf{R}_{\bar{\mathbf{y}}}$ in (4.20), with a sufficient time averaging and at a high SNR is employed to obtain a sufficiently good approximation to the *true* correlation matrix. For $P = 8$, we can see that the asymptotic performance closely matches the simulation result for $T_{av} \geq 50$. In order to reach the asymptotic performance for $P = 2$, the required number of OFDM blocks is increased to $T_{av} > 210$ OFDM blocks. Note that the accuracy of the bounds relies on

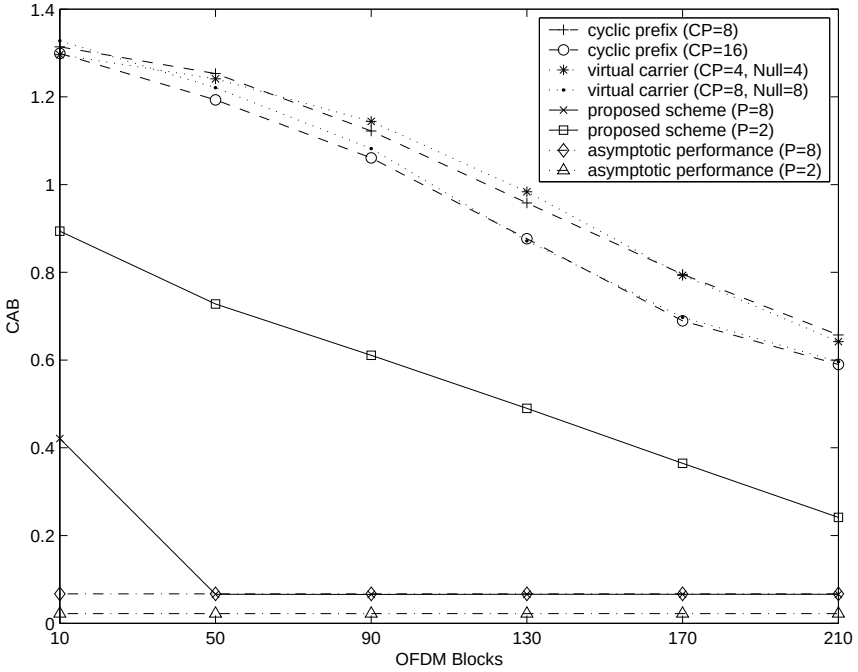


Fig. 7.3 CAB versus number of OFDM blocks (SNR=20dB).

the assumption that $\|\Delta\mathcal{H}_p\|$ is small. Otherwise, $\mathbf{R}_{\Delta\mathcal{H}}$ in (4.20) becomes significant and is equivalent to a low SNR condition. Therefore, asymptotic performance for $P = 32$ and 64 can only serve to indicate that whether or not the proposed scheme has met a certain level of confidence. Accordingly, we can conclude that the proposed algorithm reaches its asymptotic performance for $T_{av} \geq 10$ OFDM blocks when $P = 32$ and 64.

Fig. 7.3 and 7.4, which show the corresponding CABs as defined in (4.29), lead to the same conclusions as above. It should be noted that the error floor of the performance of the proposed algorithm is due to the variations across coherence bandwidth (see Section 4.3.2 for details). However, it can be eliminated by increasing the value of N_C (i.e., the size of FFT/IFFT) when the channel bandwidth is fixed.

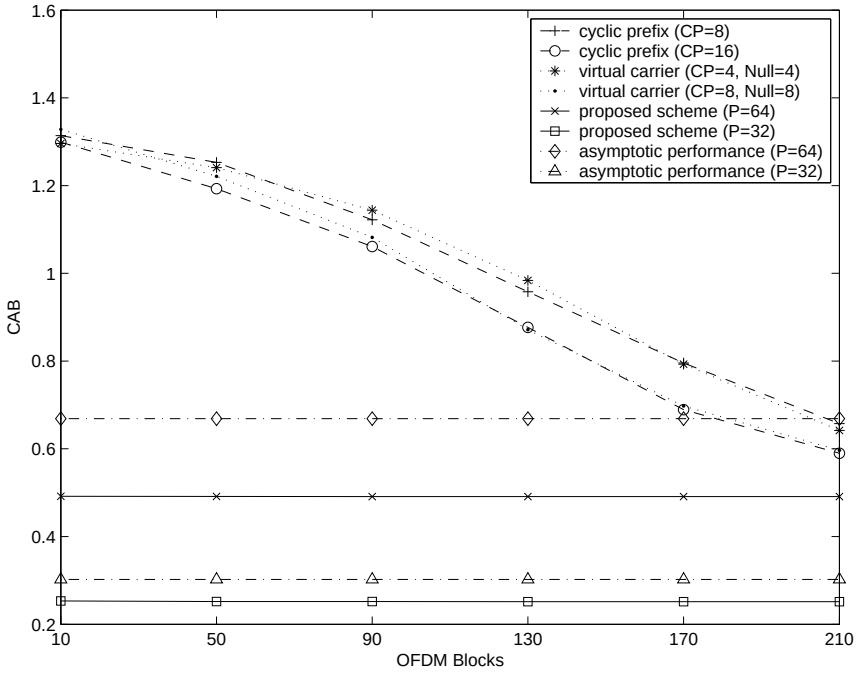


Fig. 7.4 CAB versus number of OFDM blocks (SNR=20dB).

7.1.3 Practical applications

We have shown that the proposed algorithm can achieve better performance than the referenced approaches within reasonable time averaging intervals. The superior performance relies on choosing the maximum allowable P for the purpose of minimizing the dimension of the correlation matrix without affecting the estimation performance. While the above simplified TDL model is useful for comparing various algorithms, it has limitations so that it is difficult to infer what happens in practical wideband situations. Therefore, in order to determine the maximum achievable P in practical scenarios, we consider to adapt part of the Mobile WiMAX OFDMA-PHY [70] for our OFDM system and to simulate it over the 3GPP-SCM [79].

In our OFDM system setup, the subcarrier spacing is chosen as 10.94kHz, given the

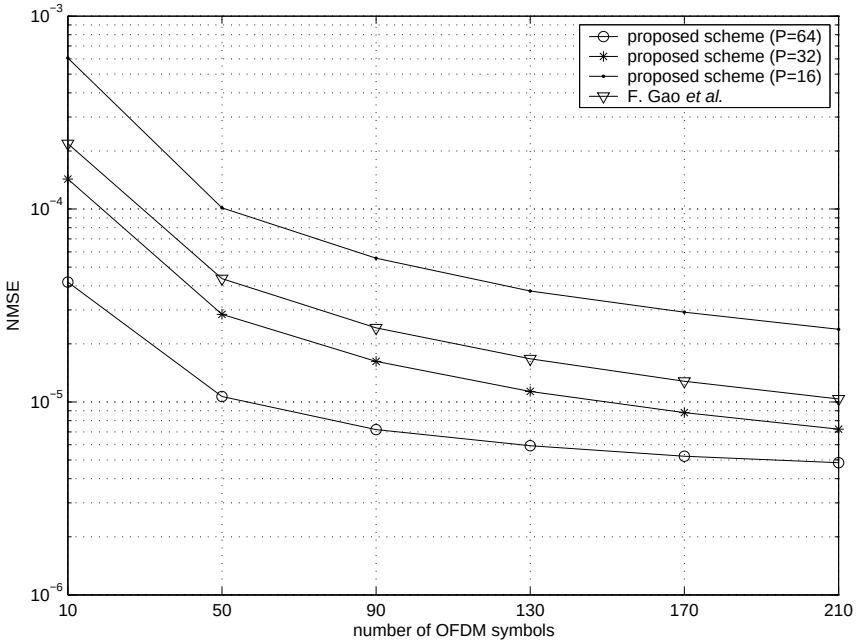


Fig. 7.5 NMSE versus number of OFDM blocks over 3GPP Urban Micro (SNR=20dB).

OFDM useful symbol duration is 91.4μs and the cyclic prefix length is 11.4μs [70]. Since we consider $N_C = 256$, the channel bandwidth is approximately 2.5MHz. For each time epoch, the incoming QPSK symbols are chosen only to span over 1 OFDM symbol ($N_F = 1$), and each Mobile WiMAX frame consists of 48 OFDM symbols. In the 3GPP-SCM setup, the carrier frequency is 2.5GHz. Base station antenna spacing is 10λ and MS antenna spacing is $\lambda/2$, where λ is the wavelength at the carrier frequency. The channel coefficients of each 3GPP-SCM scenario are generated according to the implementation in [79]. We also present the asymptotic performance of the approach given in [47] (as indicated by "F. Gao *et al.*" in the legends of the simulation figures), tailored into our system setup for comparisons. Note that accurate channel-order estimation is also assumed for this referenced algorithm.

Fig. 7.5 - 7.7 show the NMSE versus number of OFDM blocks over the 3GPP Urban Micro, Urban Macro, and Suburban Macro models [79], respectively. Note that SNR =

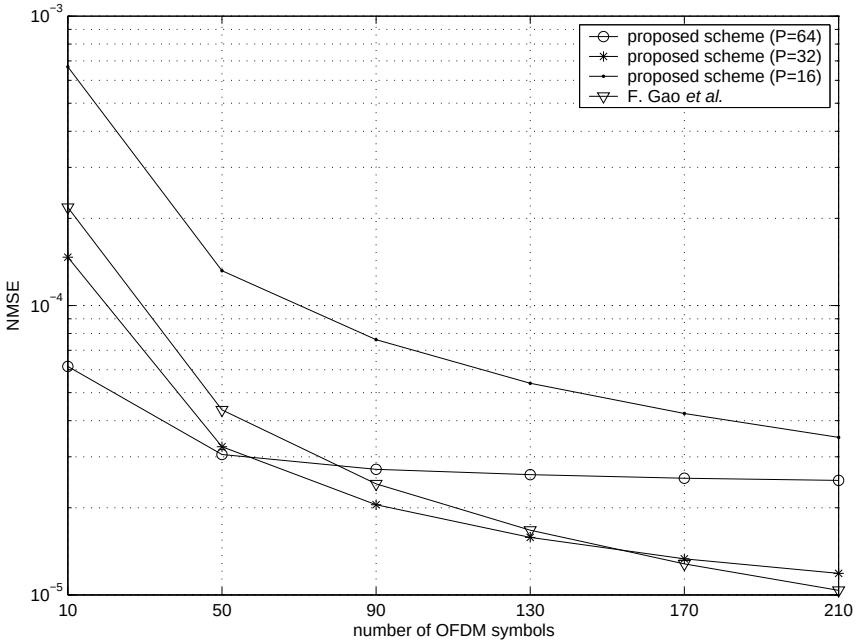


Fig. 7.6 NMSE versus number of OFDM blocks over 3GPP Urban Macro (SNR=20dB).

20dB is considered for all the scenarios. We observe that with a suitable choice of P, the proposed algorithm can reach NMSE $\leq 3 \times 10^{-5}$ in all the cases within 50 OFDM blocks (or approximately 1 Mobile WiMAX frame). We also present NMSE versus the choice of P of the proposed algorithm over the Urban Macro and Suburban Macro models in Fig. 7.8, and we can observe that the best choice of P for these models should fall between 32 and 64.

To determine the efficiency of the proposed algorithm, we evaluate the associated CRB over each 3GPP-SCM scenario. Specifically, we evaluate the CRBs as given in (4.38) by considering the adjacent P channel coefficients are the same (i.e., $\bar{\mathcal{H}}'_1 = \bar{\mathcal{H}}'_2 = \cdots = \bar{\mathcal{H}}'_P$), which constitutes the optimal condition of the proposed algorithm. Fig. 7.9(a) - 7.9(c) show the NMSE and the corresponding CRB over 3GPP Urban Micro, Urban Macro, and Suburban Macro models, respectively. From the results, we can observe that there is about

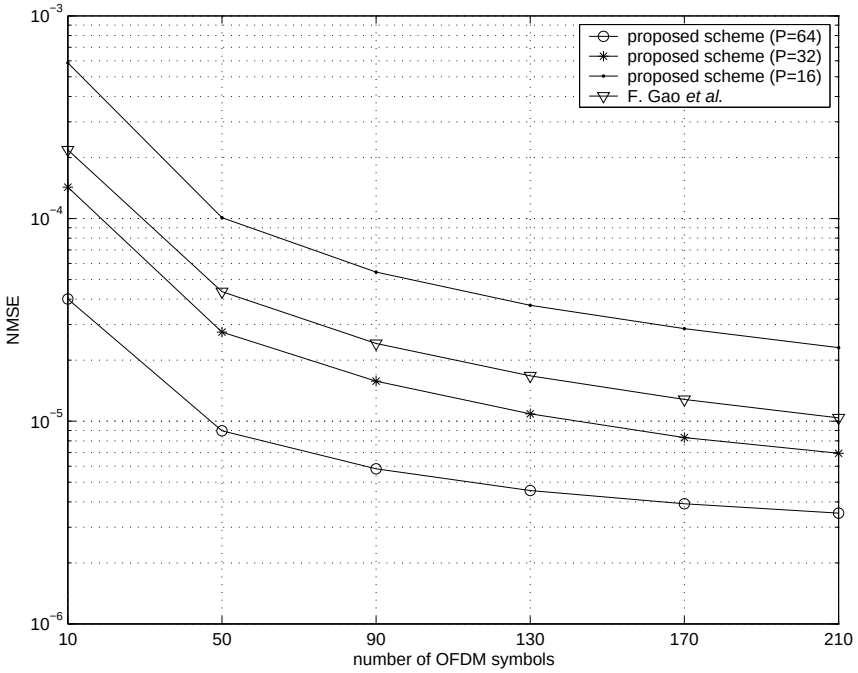


Fig. 7.7 NMSE versus number of OFDM blocks over 3GPP Suburban Macro (SNR=20dB).

3−6dB gap between the NMSE curve and the corresponding CRB bound over each scenario when SNR $\geq$ 20dB. We also observe in these experiments that the referenced algorithm [47] does not show advantages over our approach, since the dimension of its correlation matrix is given as $N_R(N_C + N_{cp}) > N_R N_C = 768$, as compared to $N_R N_C/P \leq 24$ in our approach. We stress again that these referenced curves in Fig. 7.9(a) - 7.9(c) represent the ideal, asymptotic performance of the algorithm proposed in [47]. In other words, the referenced algorithm could not reach the asymptotic performance at such a small number of time samples and low SNRs.

Finally, we present the BER curves of the proposed algorithm over Urban Macro and Suburban Macro scenario in Fig. 7.10(a) and 7.10(b), respectively. For SNR $\geq$ 15dB, we can conclude that the proposed algorithm employing only 50 time samples can reach the

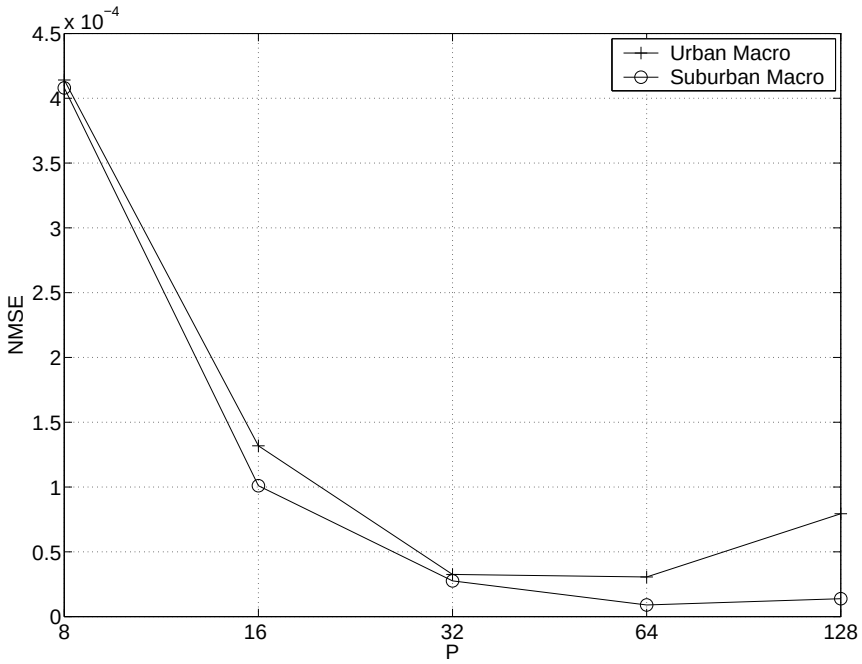
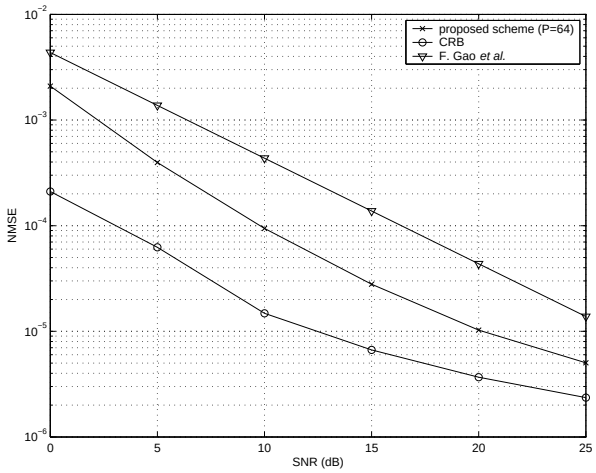


Fig. 7.8 NMSE versus P (when the number of OFDM symbols $T_{av} = 50$).

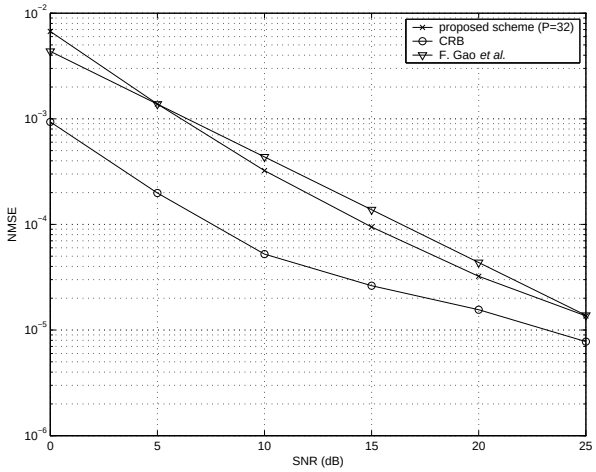
same performance as if perfect CSI is known at the receiver. However, more than 210 time samples are required for channel estimation if SNR $\leq$ 15dB.

7.2 Discussions of time-invariant scenarios

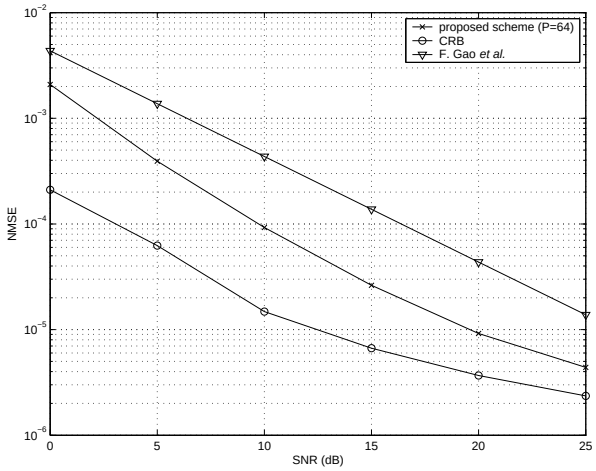
Our first contribution is in developing and analyzing a new scheme to overcome some fundamental limitation of the subspace-based blind approach when applied to MIMO-OFDM transmission over time-varying channels. Specifically, when considering the time invariance requirement of a practical MIMO-OFDM system with a large number of OFDM subcarriers, e.g., 128 or more, the traditional subspace-based methods require extremely large number of time samples for obtaining a good time-averaged correlation matrix, making them impractical. By exploiting the frequency correlation among adjacent subcarriers (i.e.,



(a) 3GPP Urban Micro

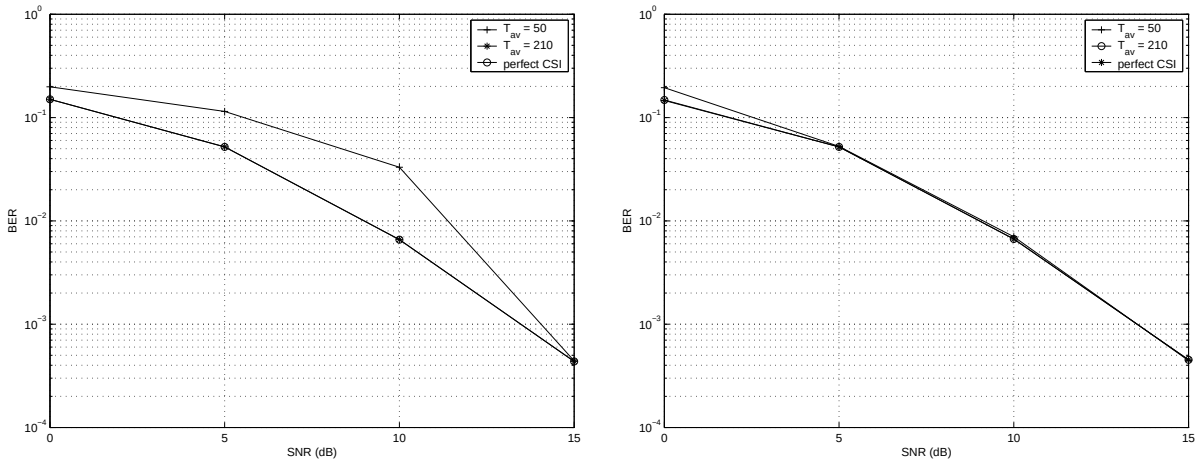


(b) 3GPP Urban Macro



(c) 3GPP Suburban Macro.

Fig. 7.9 Performance of the proposed scheme over various 3GPP-SCM scenarios ($T_{av} = 50$) as a function of SNR.



(a) 3GPP Urban Macro (b) 3GPP Suburban Macro

Fig. 7.10 BER of proposed scheme over 3GPP-SCM scenarios.

within the coherence bandwidth) through the concept of subcarrier grouping, we proposed a novel subspace-based estimation method which requires a significantly smaller number of time samples. The above simulation results showed that the proposed method could achieve a better estimation accuracy than existing benchmark approaches within a reasonable time averaging interval.

To explain why the proposed algorithm generally performs better in the Suburban Macro scenario than in the Urban Macro one, the cumulative distribution functions (CDFs) of the RMS delay spread (DS), i.e., $Pr\{\tau_{RMS} \leq \text{abscissa}\}$, of these scenarios are shown in Fig. 7.11. From this figure, we can clearly see that the corresponding RMS delay spread of the Suburban Macro scenario is much smaller than that of the Urban Macro one, meaning that the coherence bandwidth of the Suburban Macro scenario is much larger than that of the Urban Macro one. It thus implies that our approach is better suited for the 3GPP Suburban Macro case.

It should be noted that unlike the traditional approaches which require explicit channel-order information for estimating the channel matrix, the proposed algorithm requires only

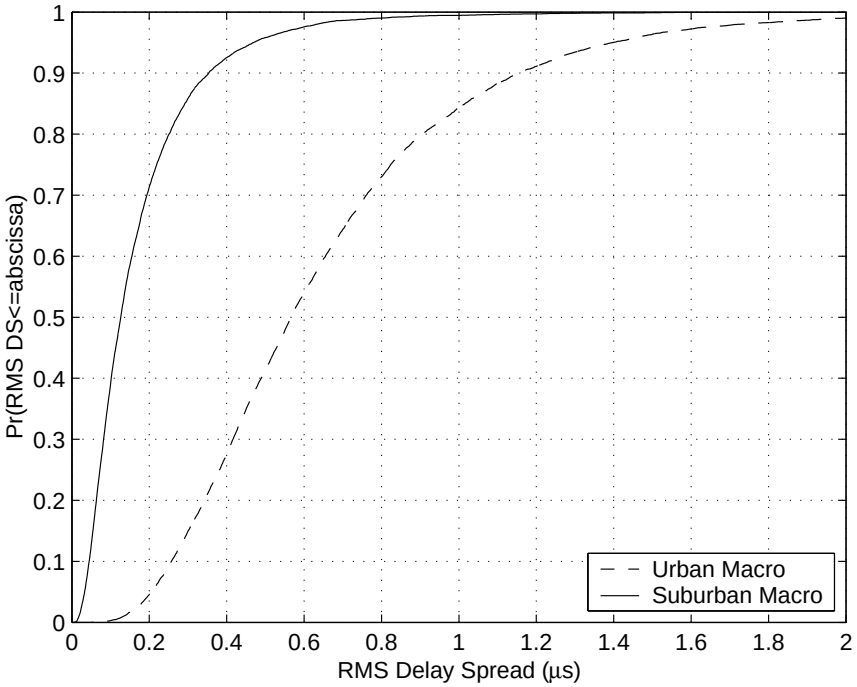


Fig. 7.11 CDF of the RMS delay spread.

an upper bound of the channel order to determine the range of the parameter P. Therefore, the proposed algorithm is less sensitive to the channel modeling errors.

7.3 Time-variant scenarios

The MIMO system under consideration consists of $N_T = 2$ transmit and $N_R = 3$ receive antennas. The number of subcarriers N_C used in the OFDM modulation is set to 256. For each time epoch, the incoming symbol streams are independent and identically distributed (i.i.d.) QPSK symbols. The OFDM useful symbol duration is 91.4μs and the cyclic prefix length is 11.4μs, resulting in a subcarrier spacing of 10.94kHz. Since we consider $N_C = 256$, the channel bandwidth is approximately 2.5MHz.

Modeling of the time-varying MIMO channel is also based on the 3GPP-SCM setup;

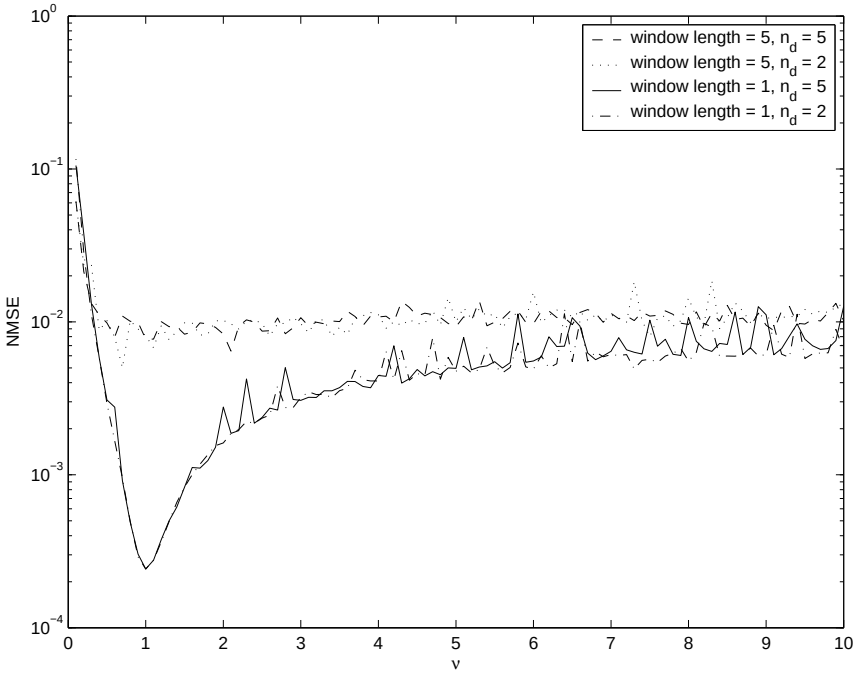


Fig. 7.12 NMSE versus precoder coefficient ν when MS speed is 100km/h ($E_b/N_0 = 14$dB).

we refer the reader to [79] for additional details. To evaluate our algorithm, we consider a Suburban Macro scenario with the carrier frequency $f_c = 2.5$GHz, where the mobile station (MS) is allowed to travel in a random direction at a constant speed of 100km/h. Hence, the maximum Doppler shift is 231.48Hz and the normalized Doppler frequency is 0.02. Experimentally, we have found that a suitable value of P in this given scenario is 64. To obtain the ambiguity matrix $\mathbf{A}^m$ needed at the mth OFDM symbol time, we employ $\mathbf{A}^m = (\hat{\mathbf{Q}}^m_{n_d})^\dagger \mathbf{H}^m_{\lceil P/2 \rceil}$ in the simulations. This type of approach is common in the literature on subspace-based blind channel identification.

Considering a rectangular window (i.e. $\beta = 1$) of length $l = 1$ and 5, we first investigate the choice of the precoder coefficient ν from the perspective of channel estimation performance. In our experiments, the NMSE for the mth channel estimate is defined here

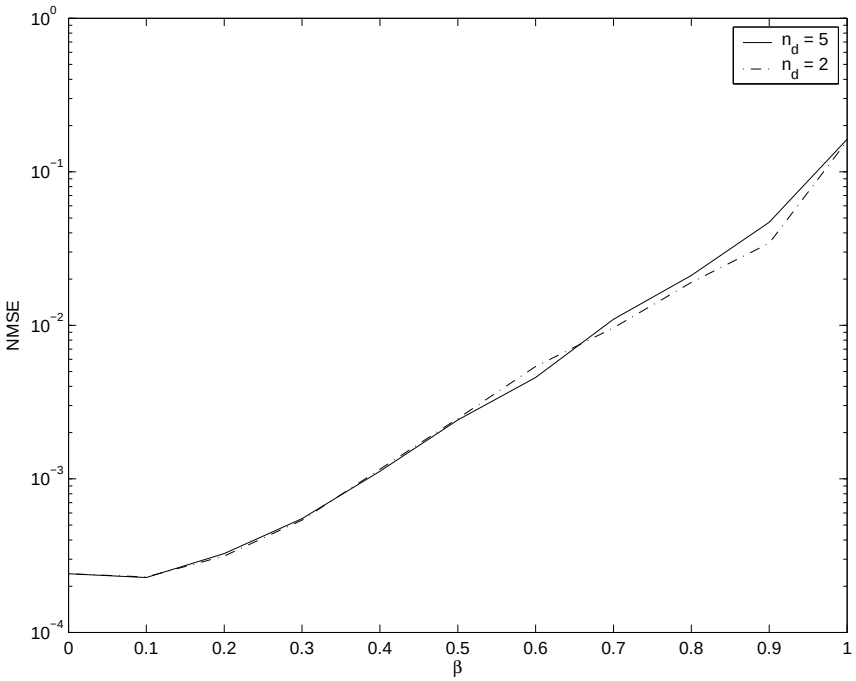


Fig. 7.13 NMSE versus forgetting factor β when MS speed is 100km/h ($\nu = 1$ and $E_b/N_0 = 14$dB).

as

$$\text{NMSE} = \sum_{i,j,k} E[|\hat{h}^m_{i,j}[k] - h^m_{i,j}[k]|^2] / \sum_{i,j,k} E[|h^m_{i,j}[k]|^2], \tag{7.2}$$

and the ensemble average is taken over 200 independent realizations of the random process. Fig. 7.12 shows the NMSE of the channel estimates versus ν when a hundred OFDM symbols are observed and the E_b/N_0 (i.e. SNR per bit) is 14dB, where E_b and N_0 denote the energy per bit, and the one-sided noise power spectral density, respectively. We are not surprised to see that choosing the window length $l = 1$ gives the best performance since the wireless channel is changing so rapidly in this case. In particular, we observe that the NMSE reaches its minimum, i.e. 2.5×10^{-4} when $\nu = 1$, which coincides with our analysis in Section 6.5.

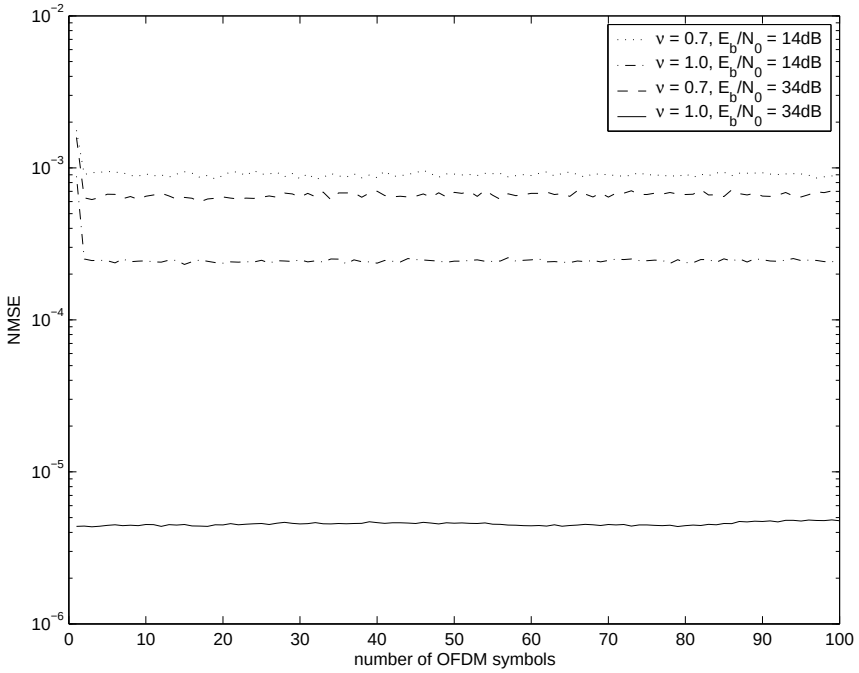


Fig. 7.14 NMSE versus number of OFDM symbols when MS speed is 100km/h ($n_d = 2$).

In Fig. 7.13, we investigate the choice of the forgetting factor β, when an exponential window of infinite length is considered under the same condition as above. We can see that the NMSE reaches a minimum (a value comparable to that of the rectangular window with $l = 1$) around NMSE $= 2 \times 10^{-4}$ when $\beta \in [0, 0.1]$, meaning that previous data samples are of little use for the estimation of the current channel statistics for this rapidly TV channel. Thus, employing an exponential window cannot gain additional estimation performance in this scenario. We also notice that there is no significant improvement in the estimation performance when n_d is increased from 2 to 5 in both figures. Hence, we simply assign $n_d = 2$ and employ a rectangular window of length $l = 1$ in the following.

Fig. 7.14 presents the NMSE of the channel estimates versus the number of OFDM symbols received when $\nu = 1$ and 0.7 at the $E_b/N_0 = 14$, 34dB. We can see that the

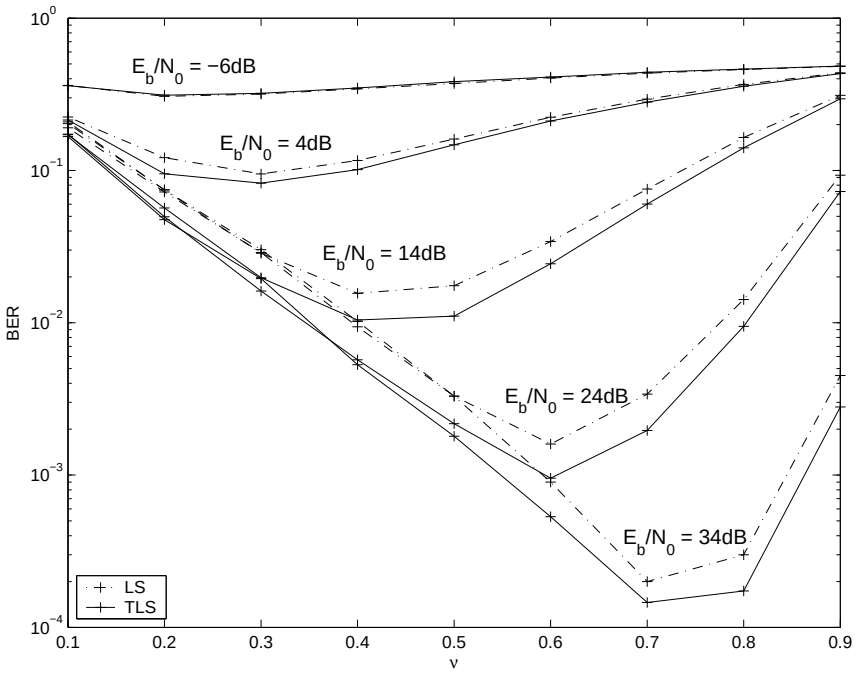


Fig. 7.15 BER versus precoder parameter ν when MS speed is 100km/h ($n_d = 2$).

proposed algorithm can track the fast time-varying channel in less than 5 OFDM symbols in all the cases, and maintain its performance over time despite the rapid variations in the channel coefficients. These results show in particular that adjusting the trade-off between estimation performance and symbol recovery in these cases (e.g., ν is decreased from 1 to a smaller value) will not affect the convergence rate. In addition, we observe that the estimation performance of the case $\nu = 1.0$ at $\mathrm{E_b/N_0}$ = 14dB outperforms that of the case $\nu = 0.7$ at E_b/N_0 = 34dB, implying that choosing a proper precoder coefficient is rather important.

In Fig. 7.15, we show the BER versus the precoder coefficient ν for various E_b/N_0's. We consider both the least squares (LS) and the total least squares (TLS) [120] estimation for symbol recovery. We first notice that for a given ν, a higher E_b/N_0 in general gives a

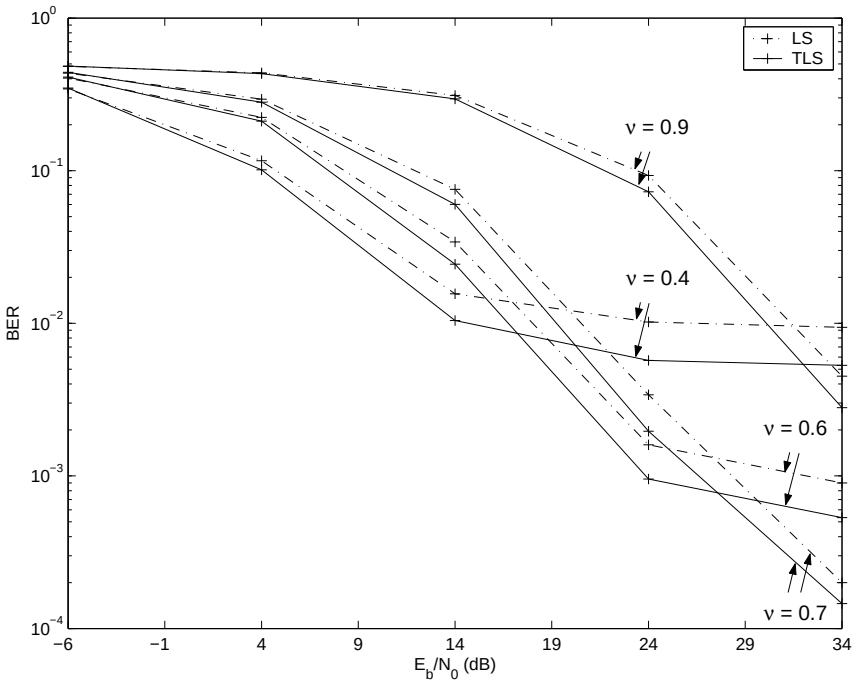


Fig. 7.16 BER versus E_b/N_0 when MS speed is 100km/h ($n_d = 2$).

better BER. We also observe that the higher the E_b/N_0, the larger the optimal choice of ν, and hence the lower the BER. This can be explained as follows: For a less noisy scenario, a shorter distance between any pair of the precoder outputs is allowed, and thus we can increase the value of ν to gain better estimation performance and so achieving a lower BER. Furthermore, we observe that for a given E_b/N_0, the TLS estimation (solid lines) generally outperforms the LS estimation (dash-dot lines) estimation, since the TLS estimation takes the channel estimation errors into account while performing the symbol recovery.

Fig. 7.16 demonstrates the BER versus the E_b/N_0 for various ν's, considering both the LS estimation (dash-dot lines) and the TLS estimation (solid lines) estimation for symbol recovery. We can see that when $\nu = 0.4$, the proposed algorithm performs the best in the low E_b/N_0 region due to its largest distance between any pair of the precoder outputs;

however, it performs the worst in the high E_b/N_0 region because the most inaccurate channel information is used. The case $\nu = 0.9$ yields the worst performance for almost all E_b/N_0 values due to its extremely short distance between any pair of the precoder outputs, even though the best estimation performance is achieved. A good choice of ν should fall between 0.6 to 0.7 when the E_b/N_0 is moderate to high, and a $1-2$dB gain can be achieved by using the TLS instead of the LS estimation at the $E_b/N_0 = 19$dB.

7.4 Discussions of time-variant scenarios

To estimate wideband time-varying channels with large Doppler shift, one typically resorts to pilot placements at consecutive OFDM symbol times over specific subcarriers, followed by different interpolation schemes; this is because blind channel estimation normally requires a long observation interval and tends to exhibit a slow convergence rate, making it difficult to apply on these channels. Our second main contribution is in developing a new scheme to blindly track a wideband time-varying wireless channel which may be changing at each OFDM symbol time, without using any preambles or training sequence. In particular, our approach offers the flexibility in choosing the number of transmit as well as receive antennas, and offers high bandwidth efficiency and low complexity. In a realistic mobile wireless channel environment in which the maximum Doppler shift is 231.48Hz, the numerical results showed that our approach can achieve a BER at the level of 10^{-2} when the $\mathrm{E_b/N_0} \geq 14$dB.

Chapter 8

Summary and conclusion

In this thesis, we analyzed existing subspace-based blind channel estimators and highlighted some of their limitations. Subsequently, we proposed novel estimators of this type for broadband MIMO-OFDM systems that exploit channel correlation over the coherence bandwidth, to improve the convergence speed while maintaining satisfactory estimation performance. In this chapter, we summarize the main results and ideas developed in the thesis, and then present some concluding remarks that may lead to some possible avenues for future research.

8.1 Summary of the work

Over many years, pilot-based channel estimation has been widely employed in various communication systems to acquire the CSI, as needed in several capacity-achieving techniques, such as (frequency-)space-time coding and spatial multiplexing. To take advantage of the blind channel estimation, i.e., acquiring the CSI without using pilots, it is essential to mitigate the slow convergence rate associated with blind channel estimators. Fortunately, several blind approaches using second order statistics have been proposed that can achieve

a reasonable convergence rate for a time-invariant channel. This includes the subspace methods, which are particularly attractive due to their good performance and moderate complexity.

In the subspace-based estimation approach, channel estimates can often be obtained in a simple form by optimizing a quadratic cost function. In Chapter 3, we explained that to guarantee an acceptable estimation performance for a time-varying channel, the dimension of the correlation matrix of the received signal cannot be too large. Nevertheless, most existing subspace-based approaches were employed in the time domain, including applications for OFDM-based systems. This design leads to the inherent property that the dimension of the correlation matrix is a multiple of the number of OFDM subcarriers. Thus, the aforementioned subspace-based approaches lead to extremely unfavorable conditions for these systems, since 256 to 2048 OFDM subcarriers are normally considered.

These observations motivated the development in Chapter 4, i.e., a new subspace-based blind channel estimator for MIMO-OFDM systems that can exploit the frequency correlation among adjacent channel coefficients within the coherence bandwidth, for the purpose of dramatically reducing the dimension of the correlation matrix. It was found that, given some rough knowledge of the RMS delay spread of the wireless channel, the proposed estimator also avoids the need of channel order estimation, and thus yield an added flexibility in terms of estimation performance and robustness. We further studied the identifiability condition of this new estimator along with its performance measures, including perturbation and Cramer-Rao bound analysis.

In Chapter 5, to reduce the computational complexity associated with eigenvalue decomposition in the proposed method, we also investigated the convergence properties (over non-stationary scenarios) of a simple yet powerful subspace tracking approach, namely, orthogonal iteration. Then in Chapter 6, by incorporating the subspace tracking approach

in our estimator, we developed a fully recursive algorithm to *update* the signal subspace needed in the estimate of time-varying channels. This approach is computationally efficient since it avoids re-computing the subspace from scratch at every time iteration. To further reduce the complexity associated with the ambiguity matrix in our proposed estimator, a precoder was also introduced at the transmitter.

To complete the study, the choice of the precoder was also derived and analyzed. We first showed that a precoder matrix whose entries are all 1's results in the best estimation performance; however, its condition number becomes infinity, and hence the transmitted symbols cannot be recovered at the receiver. We then showed mathematically that by properly perturbing the off-diagonal entries of the precoder matrix, we can decrease its condition number to a reasonable value.

Finally, in order to assess the performance of the proposed algorithms, we presented numerous computer simulations in Chapter 7. When we consider a block fading channel, it was shown that the proposed method requires a significantly shorter time-averaging period than the benchmark methods, particularly when the wireless channels are characterized by smaller RMS delay spread. We also confirmed that the proposed tracking algorithm indeed outperforms the benchmark algorithm, and is capable of properly tracking time-varying channels with the maximum Doppler shift up to 230Hz. Regarding the precoder design, we show empirically that the best trade-off between channel estimation and symbol detection performance is achieved when the ratio of its off-diagonal to diagonal entries is on the order of $0.6 - 0.7$.

8.2 Concluding remarks

The IEEE 802.11n PHY layer standard, which aims at providing an 11-fold increase in transmission speed over 802.11g, or a 55-fold increase in transmission speed over 802.11a, can be regarded as one of the prominent examples of wireless communication systems nowadays: Combining the MIMO-OFDM techniques, it uses training or pilot signals embedded in transmitted data streams to facilitate channel estimation and synchronization in the system. The presence of pilot signals implies that data throughput is decreased, e.g. at least 6% loss in capacity is expected in the IEEE 802.11n systems. Specifically, for a time-varying channel where mobile is expected to travel with a high speed, the throughput loss due to the periodic insertion of training or pilot signals is huge. Therefore, employing a fast-converging and reliable blind channel estimation in the design of future wireless system seems to be an attractive solution.

Nevertheless, blind channel estimation has not been employed or considered in *any* of these communication systems yet, including the latest proposals of wireless standards such as IEEE 802.11n, WiMAX, and 3GPP LTE. We may wonder, given that numerous works in blind channel estimation have demonstrated superior performance of their methods, for what reasons do people still not embrace these apparently capacity-saving techniques?

Similar question was raised by Z. Ding and Y. Li in the preface of [156] about a decade ago, and the reasons, according to the authors, may be attributed to: (1) the inadequate understanding about the problem itself and various proposed blind estimation schemes by many practicing engineers, (2) a reliable blind estimation algorithm is yet to be established that can guarantee speedy convergence, and with reasonable complexity. Specifically, the algorithm should not be very sensitive to parameters such as the estimated channel order.

A decade later, many more practicing engineers may have acquainted the problem itself

and several new blind estimation schemes have been proposed; however, in the context of MIMO-OFDM systems, a large number of different blind approaches that have been proposed to date still suffer from various shortcomings, including the slow convergence rate and sensitivity to the estimated channel order.

The work in this thesis led to many interesting developments in subspace-based blind channel estimation and tracking algorithms. Specifically, the proposed algorithms, which exploit correlation in the frequency channel coefficients within the coherence bandwidth, have shown significant improvement in the convergence speed, while maintaining satisfactory estimation performance. In addition, they do not require exact knowledge of the wireless channel order, and the requirement of their computations is lower than that of the benchmark approaches. We hope that the small step taken in the thesis toward a simple, reliable, and fast blind channel estimation algorithm, may eventually lead to widespread of blind channel estimation approaches in future wireless standards.

8.3 Future work

Several promising avenues for future research have emerged based on the work presented in this thesis. They are summarized briefly below:

1. In Chapter 4 and 7, estimating the ambiguity matrix is treated as a separate problem. It would be interesting to investigate how the estimation of ambiguity matrix can be effectively incorporated into the proposed algorithms. This can possibly be achieved with the help of an HOS approach, the use of a training sequence (resulting in the so-called semi-blind approach that combines the blind and the non-blind approaches), or other new signal processing approaches.

2. As space-time block coding is considered in IEEE 802.11n wireless networking stan-

dard, it would thus be of interest to further investigate the scenario in which a correlated input sequence, e.g. Alamouti's space-time block code, is applied at the transmitter side. We have assumed that the input sequence of the OFDM modulator is i.i.d. throughout this thesis. Therefore, a good starting point would be to consider an interleaver, for the purpose of scrambling the encoder output.

3. In Chapter 4, the asymptotic performance and the Cramer-Rao bound of the proposed algorithm are studied but the results remain in the form of somewhat general, which can only be evaluated through numerical computations. It would be relevant to generalize these bounds in terms of parameters such as noise variance, dimension of the eigenvectors, etc., in order to gain more insights on how the estimation performance is affected by these parameters. Such results would provide an even better understanding of the proposed subspace-based estimators.

4. It will be challenging to consider an even higher Doppler rate, in which the wireless channel may be changing within an OFDM symbol time. In this case, a subspace-based blind channel estimator will be in no doubt very attractive for future wireless standards if acceptable performance can be achieved.

5. It would be interesting to further extend the proposed blind channel estimation/tracking algorithms from single user scenarios to multi-user ones. In the multi-user case, it is expected that the proposed approaches may be affected by the interference from other users, resulting in worse estimation/tracking performance. Another interesting research avenue would be to explore the effects on the performance of the proposed algorithms when a degenerate channel condition is occurred.

Appendix A

Quadratic cost function

From (4.8), we can arrive at

$$\mathbf{u}_{j,i}^H \operatorname{diag}(\mathbf{H}[\omega_{\rho,1}] \cdots \mathbf{H}[\omega_{\rho,\zeta}]) = \mathbf{0}_{1\times N_T\zeta}, \quad i = 1, \cdots, N_F. \tag{A.1}$$

By further partitioning $\mathbf{u}_{j,i}$ into ζ segments of equal dimension, with the kth segment denoted as $\mathbf{u}_{j,i,k} \in \mathbb{C}^{N_R\times 1}$, $k = 1, 2, \cdots, \zeta$, we can obtain from the above that

$$\mathbf{u}_{j,i,k}^H \mathbf{H}[\omega_{\rho,k}] = \mathbf{0}_{1\times N_T}, \quad k = 1, \cdots, \zeta. \tag{A.2}$$

Therefore, we can arrive at

$$\mathbf{u}_{j,i}^H \begin{bmatrix} \mathbf{H}[\omega_{\rho,1}] \\ \vdots \\ \mathbf{H}[\omega_{\rho,\zeta}] \end{bmatrix} = \mathbf{u}_{j,i}^H \bar{\mathcal{H}}' = \mathbf{0}_{1\times N_T}. \tag{A.3}$$

In practice, only an estimate of the noise eigenvector $\mathbf{u}_j$ is available, denoted here as $\hat{\mathbf{u}}_j$. Based on (A.3), we can define a new cost function that is more convenient to perform the

optimization:

$$C'(\bar{\mathcal{H}}') \stackrel{\text{def}}{=} \sum_{j} \sum_{i=1}^{N_F} \|\hat{\mathbf{u}}_{j,i}^H \bar{\mathcal{H}}'\|_2^2. \tag{A.4}$$

Clearly, since $\mathbf{u}_j^H \bar{\mathcal{H}} = \mathbf{0}$ implies $\mathbf{u}_{j,i}^H \bar{\mathcal{H}}' = \mathbf{0}$ as shown above, it follows that the condition $C(\bar{\mathcal{H}}) = 0$ in (4.9) implies that $C'(\bar{\mathcal{H}}') = 0$. Finally,

$$\begin{aligned} C'(\bar{\mathcal{H}}') &= \sum_{j} \sum_{i=1}^{N_F} \|\hat{\mathbf{u}}_{j,i}^H \bar{\mathcal{H}}'\|_2^2 = \sum_{j} \sum_{i=1}^{N_F} \|\bar{\mathcal{H}}'^T \hat{\mathbf{u}}_{j,i}^*\|_2^2 \\ &= \sum_{j} \|\bar{\mathcal{H}}'^T \hat{\mathbf{V}}_j^*\|_F^2 = \operatorname{tr}\left(\bar{\mathcal{H}}'^T \hat{\mathbf{\Psi}} \bar{\mathcal{H}}'^*\right), \end{aligned}$$

which gives exactly (4.12). Note that by employing the new cost function in (A.4), we now measure how close is $\hat{\mathbf{u}}_{j,i}^H \bar{\mathcal{H}}'$ to $\mathbf{0}_{1\times N_T}$.

References

[1] T. S. Rappaport, *Wireless Communications: Principles and Practice.* Prentice Hall, 1996.

[2] V. K. Garg and J. E. Wilkes, *Wireless and Personal Communications Systems.* Prentice-Hall, 1996.

[3] G. D. Durgin, *Space-Time Wireless Channels.* Prentice Hall, 2003.

[4] Y. K. Kim and R. Prasad, *4G Roadmap and Emerging Communication Technologies.* Artech House universal personal communications series, 2006.

[5] A. Clapton, Ed., *Future Mobile Networks: 3G and Beyond.* The Institution of Electrical Engineers, 2001.

[6] L. Hanzo, J. Blogh, and S. Ni, *3G, HSPA and FDD versus TDD Networking: Smart Antennas and Adaptive Modulation (second edition).* John Wiley and Sons, 2008.

[7] H.-W. Chen and M. Guizani, *Next Generation Wireless Systems and Networks.* John Wiley and Sons, 2006.

[8] W. W. Lu, Ed., *Broadband Wireless Mobile: 3G and Beyond.* John Wiley and Sons, 2002.

[9] B. Furht and S. A. Ahson, Eds., *Long Term Evolution: 3GPP LTE Radio and Cellular Technology.* Auerbach Publications, 2009.

[10] F. Khan, *LTE for 4G Mobile Broadband: Air Interface Technologies and Performance.* Cambridge University Press, 2009.

[11] E. Dahlman, S. Parkvall, J. Sköld, and P. Beming, *3G Evolution: HSPA and LTE for Mobile Broadband.* Elsevier, 2007.

[12] 3rd Generation Partnership Project, "Technical Specification Group Radio Access Network; Requirements for Evolved UTRA (E-UTRA) and Evolved UTRAN (E-UTRAN)," 3GPP TR 25.913 Technical Report V8.0.0 (2008-12), 2008.

[13] S. Sesia, I. Toufik, and M. Baker, Eds., *LTE: The UMTS Long Term Evolution from Theory to Practice.* John Wiley and Sons, 2009.

[14] T.-D. Chiueh and P.-Y. Tsai, *OFDM Baseband Receiver Design for Wireless Communications.* John Wiley and Sons (Asia), 2007.

[15] 3rd Generation Partnership Project, "Technical Specification Group Radio Access Network; Evolved Universal Terrestrial Radio Access (E-UTRA); LTE Physical Layer - General Description," 3GPP TS 36.201 V8.3.0 (2009-03), 2009.

[16] R. Janaswamy, *Radiowave Propagation and Smart Antennas for Wireless Communications.* Kluwer Academic Publishers, 2001.

[17] G. L. Stüber, J. R. Barry, S. W. Mclaughlin, Y. Li, M. A. Ingram, and T. G. Pratt, "Broadband MIMO-OFDM wireless communications," *Proc. IEEE*, vol. 92, no. 2, pp. 271–294, Feb. 2004.

[18] G. Ferrari, G. Colavolpe, and R. Raheli, *Detection Algorithms for Wireless Communications with Applications to Wired and Storage Systems.* John Wiley and Sons, 2004.

[19] V. Kühn, *Wireless Communications over MIMO Channels: Applications to CDMA and Multiple Antenna Systems.* John Wiley and Sons, 2006.

[20] L. Hanzo, O. Alamri, M. El-Hajjar, and N. Wu, *Near-Capacity Multi-Functional MIMO Systems: Sphere-Packing, Iterative Detection and Cooperation.* John Wiley and Sons, 2009.

[21] B. Vucetic and J. Yuan, *Space-Time Coding.* John Wiley and Sons, 2003.

[22] Y. Li and Z. Ding, "Blind channel identification based on second order cyclostationary statistics," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process.*, vol. 4, Apr. 1993, pp. 81–84.

[23] Y. Hua, H. Yang, and W. Qiu, "Source correlation compensation for blind channel identification based on second-order statistics," *IEEE Signal Process. Lett.*, vol. 1, no. 8, pp. 119–120, Aug. 1994.

[24] L. Tong, G. Xu, B. Hassibi, and T. Kailath, "Blind channel identification based on second-order statistics: A frequency-domain approach," *IEEE Trans. Inf. Theory*, vol. 41, no. 1, pp. 329–334, Jan. 1995.

[25] G. Xu, H. Liu, L. Tong, and T. Kailath, "A least-squares approach to blind channel identification," *IEEE Trans. Signal Process.*, vol. 43, no. 12, pp. 2982–2993, Dec. 1995.

[26] G. B. Giannakis and C. Tepedelenlioglu, "Direct blind equalizers of multiple FIR channels: A deterministic approach," *IEEE Trans. Signal Process.*, vol. 47, no. 1, pp. 62–74, Jun. 1999.

[27] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: A time domain approach," *IEEE Trans. Inf. Theory*, vol. 40, no. 2, pp. 340–349, Mar. 1994.

[28] L. Tong and S. Perreau, "Multichannel blind identification: from subspace to maximum likelihood methods," *Proc. IEEE*, vol. 86, no. 10, pp. 1951–1968, Oct. 1998.

[29] R. W. Heath Jr. and G. B. Giannakis, "Exploiting input cyclostationarity for blind channel identification in OFDM systems," *IEEE Trans. Signal Process.*, vol. 47, no. 3, pp. 848–856, Mar. 1999.

[30] H. Gazzah, P. A. Regalia, J.-P. Delmas, and K. Abed-Meraim, "A blind multichannel identification algorithm robust to order overestimation," *IEEE Trans. Signal Process.*, vol. 50, no. 6, pp. 1449–1458, Jun. 2002.

[31] K. Rahbar, J. P. Reilly, and J. H. Manton, "Blind identification of MIMO FIR systems driven by quasistationary sources using second-order statistics: a frequency domain approach," *IEEE Trans. Signal Process.*, vol. 52, no. 2, pp. 406–417, Feb. 2004.

[32] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Trans. Signal Process.*, vol. 44, no. 3, pp. 661–672, Mar. 1996.

[33] F. Alberge, P. Duhamel, and M. Nikolova, "Adaptive solution for blind identification/equalization using deterministic maximum likelihood," *IEEE Trans. Signal Process.*, vol. 50, no. 4, pp. 923–936, Apr. 2002.

[34] S. Barbarossa, A. Scaglione, and G. B. Giannakis, "Performance analysis of a deterministic channel estimator for block transmission systems with null guard intervals," *IEEE Trans. Signal Process.*, vol. 50, no. 3, pp. 684–695, Mar. 2002.

[35] D. H. Pham and J. H. Manton, "A subspace algorithm for guard interval based channel identification and source recovery requiring just two received blocks," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process.*, vol. 4, Apr. 2003, pp. 317–320.

[36] H. Murakami, "Blind estimation of a fractionally sampled FIR channel for OFDM transmission using residue polynomials," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 225–234, Jan. 2006.

[37] ——, "Deterministic blind channel estimation for a block transmission system using fractional sampling and interpolation," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 4969–4978, Oct. 2007.

[38] E. Moulines, P. Duhamel, J.-F. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Trans. Signal Process.*, vol. 43, no. 2, pp. 516–525, Feb. 1995.

[39] X. Cai and A. N. Akansu, "A subspace method for blind channel identification in OFDM systems," in *Proc. IEEE Int. Conf. on Commun.*, vol. 2, Jun. 2000, pp. 929–933.

[40] B. Muquet, M. de Courville, and P. Duhamel, "Subspace-based blind and semi-blind channel estimation for OFDM systems," *IEEE Trans. Signal Process.*, vol. 50, no. 7, pp. 1699–1712, Jul. 2002.

[41] W. Bai, C. He, L.-G. Jiang, and H.-W. Zhu, "Blind channel estimation in MIMO-OFDM systems," in *Proc. IEEE Global Commun. Conf.*, vol. 1, Nov. 2002, pp. 317–321.

[42] C. Li and S. Roy, "Subspace-based blind channel estimation for OFDM by exploiting virtual carriers," *IEEE Trans. Wireless Commun.*, vol. 2, no. 1, pp. 141–150, Jan. 2003.

[43] C. Shin and E. J. Powers, "Blind channel estimation for MIMO-OFDM systems using virtual carriers," in *Proc. IEEE Global Commun. Conf.*, vol. 4, Nov. 2004, pp. 2465–2469.

[44] C. Shin, R. W. Heath, Jr., and E. J. Powers, "Blind channel estimation for MIMO-OFDM systems," *IEEE Trans. Veh. Technol.*, vol. 56, no. 2, pp. 670–685, Mar. 2007.

[45] Y. Zeng and T.-S. Ng, "A semi-blind channel estimation method for multiuser multi-antenna OFDM systems," *IEEE Trans. Signal Process.*, vol. 52, no. 5, pp. 1419–1429, May 2004.

[46] B. Su and P. P. Vaidyanathan, "Subspace-based blind channel identification for cyclic prefix systems using few received blocks," *IEEE Trans. Signal Process.*, vol. 55, no. 10, pp. 4979–4993, Oct. 2007.

[47] F. Gao, Y. Zeng, A. Nallanathan, and T.-S. Ng, "Robust subspace blind channel estimation for cyclic prefixed MIMO OFDM systems: algorithm, identifiability and performance analysis," *IEEE J. Sel. Areas Commun.*, vol. 26, no. 2, pp. 378–388, Feb. 2008.

[48] K.-D. Kammeyer, V. Kühn, and T. Petermann, "Blind and nonblind turbo estimation for fast fading GSM channels," *IEEE J. Sel. Areas Commun.*, vol. 19, no. 9, pp. 1718–1728, Sep. 2001.

[49] M. K. Tsatsanis and G. B. Giannakis, "Subspace methods for blind estimation of time-varying FIR channels," *IEEE Trans. Signal Process.*, vol. 45, no. 12, pp. 3084–3093, Dec. 1997.

[50] H. Liu and G. B. Giannakis, "Deterministic approaches for blind equalization of time-varying channels with antenna arrays," *IEEE Trans. Signal Process.*, vol. 46, no. 11, pp. 3003–3013, Nov. 1998.

[51] J. K. Tugnait and W. Luo, "Blind identification of time-varying channels using multistep linear predictors," *IEEE Trans. Signal Process.*, vol. 52, no. 6, pp. 1739–1749, Jun. 2004.

[52] B. Champagne, A. El-Keyi, and C.-C. Tu, "A subspace method for the blind identification of multiple time-varying FIR channels," in *Proc. IEEE Global Commun. Conf.*, Honolulu, Hawaii, Dec. 2009 (6 pages).

[53] C. Tepedelenlioğlu and G. B. Giannakis, "Transmitter redundancy for blind estimation and equalization of time- and frequency-selective channels," *IEEE Trans. Signal Process.*, vol. 48, no. 7, pp. 2029–2043, Jul. 2000.

[54] Y. Sung and L. Tong, "Tracking of fast-fading channels in long-code CDMA," *IEEE Trans. Signal Process.*, vol. 52, no. 3, pp. 786–795, Mar. 2004.

[55] X. G. Doukopoulos and G. V. Moustakides, "Blind adaptive channel estimation in OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 5, no. 7, pp. 1716–1725, Jul. 2006.

[56] A. K. Salkintzis and N. Passas, Eds., *Emerging Wireless Multimedia: Services and Technologies*. John Wiley and Sons, 2005.

[57] H.-G. Jeon and E. Serpedin, "A novel simplified channel tracking method for MIMO-OFDM systems with null sub-carriers," *Signal Process.*, vol. 88, no. 4, pp. 1002–1016, 2008.

[58] J. Gao and H. Liu, "Low-complexity MAP channel estimation for mobile MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 7, no. 3, pp. 774–780, Mar. 2008.

[59] H. M. Karkhanechi and B. C. Levy, "An efficient adaptive channel estimation algorithm for MIMO OFDM systems - study of Doppler," *J. Signal Process. Syst.*, vol. 56, pp. 261–271, 2009.

[60] A. F. Molisch, M. Z. Win, and J. H. Winters, "Space-time-frequency (STF) coding for MIMO-OFDM systems," *IEEE Commun. Lett.*, vol. 6, no. 9, pp. 370–372, Sep. 2002.

[61] W. Yamada, N. Kita, A. Ando, A. Sato, T. Takao, and D. Mori, "Polarization dependency on frequency correlation of eigenvector for MIMO-OFDM systems," in *Proc. IEEE Veh. Technol. Conf.*, vol. 1, Sep. 2004, pp. 115–119.

[62] A. Paulraj, R. Nabar, and D. Gore, *Introduction to Space-Time Wireless Communications*. Cambridge University Press, 2003.

[63] C. Berrou, A. Glavieux, and P. Thitimajshima, "Near Shannon limit error-correcting coding and decoding: Turbo codes (1)," in *Proc. IEEE Int. Conf. on Commun.*, vol. 2, May 1993, pp. 1064–1070.

[64] R. G. Gallager, "Low density parity check codes," *IRE Trans. Inf. Theory*, vol. 8, no. 21-28, Jan. 1962.

[65] G. Foschini and M. Gans, "On limits of wireless communications in a fading environment when using multiple antennas," *Wireless Pers. Commun.*, vol. 6, no. 3, pp. 311–335, Mar. 1998.

[66] I. Telatar, "Capacity of multi-antenna Gaussian channels," *Euporean Trans. Tel.*, vol. 10, no. 6, pp. 585–595, Nov./Dec. 1999.

[67] "ETS 300 401 (third edition): Radio Broadcast Systems: Digital Audio Broadcasting (DAB) to Mobile, Portable and Fixed Receivers," European Telecommunications Standards Institute, Tech. Rep., 2000.

[68] "Digital Video Broadcasting (DVB): Frame Structure, Channel Coding and Modulation for a Second Generation Terrestrial Television Broadcasting System (DVB-T2)," DVB Document, A122, 2009.

[69] "Functional Specification of Technological Enhancements for HIPERLAN/2 to Meet the Requirements of a Potential 4G System," IST-2000-28584 MIND, 2001.

[70] J. G. Andrews, A. Ghosh, and R. Muhamed, *Fundamentals of WiMAX: Understanding Broadband Wireless Networking*. Prentice Hall Communications Engineering and Emerging Technologies Series, 2007.

[71] M. W. Gast, *802.11 Wireless Networks: The Definitive Guide*. O'Reilly and Associates, 2002.

[72] V. K. Garg, *Wireless Network Evolution: 2G to 3G*. Prentice-Hall, 2002.

[73] W. R. Braun and U. Dersch, "A physical mobile radio channel model," *IEEE Trans. Veh. Technol.*, vol. 40, no. 2, pp. 472–482, May 1991.

[74] R. H. Clarke, "A statistical theory of mobile radio reception," *Bell Systems Technical Journal*, vol. 47, no. 6, pp. 957–1000, 1968.

[75] R. Steele, Ed., *Mobile Radio Communications.* IEEE Press, 1992.

[76] J. W. Mark and W. Zhuang, *Wireless Communications and Networking.* Prentice Hall, 2003.

[77] J. G. Proakis, *Digital Communications (Third Edition).* McGRAW-HILL, 1995.

[78] 3rd Generation Partnership Project, "Technical Specification Group Radio Access Network; Spatial Channel Model for Multiple-Input Multiple-Output Simulations (release 6)," TSG-RAN Meeting #19, Birminghan, UK, 11th-14th, 3GPP TR 25.996 Technical Report V1.0.0 (2003-03), 2003.

[79] ——, "Technical Specification Group Radio Access Network; Spatial Channel Model for Multiple-Input Multiple-Output Simulations (Release 6)," 3GPP TR 25.996 Technical Report V6.0.0 (2003-09), 2003.

[80] "Commission of the European Communities: COST 207 Digital Land Mobile Radio Communications: Final Report," Tech. Rep., 1989.

[81] R. J. Schilling and S. L. Harris, *Fundamentals of Digital Signal Processing Using MATLAB.* Thomson, 2005.

[82] A. Stéphenne and B. Champagne, "Effective multi-path vector channel simulator for antenna array systems," *IEEE Trans. Veh. Technol.*, vol. 49, no. 6, pp. 2370–2381, Nov. 2000.

[83] M. Borgmann and H. Bolcskei, "Noncoherent space-frequency coded MIMO-OFDM," *IEEE J. Sel. Areas Commun.*, vol. 23, no. 9, pp. 1799–1810, Sep. 2005.

[84] T. Himsoon, W. Su, and K. J. Ray Liu, "Single-block differential transmit scheme for broadband wireless MIMO-OFDM systems," *IEEE Trans. Signal Process.*, vol. 54, no. 9, pp. 3305–3314, Sep. 2006.

[85] V. Pauli, L. Lampe, and J. Huber, "Differential space-frequency modulation and fast 2-D multiple-symbol differential detection for MIMO-OFDM," *IEEE Trans. Veh. Technol.*, vol. 57, no. 1, pp. 297–310, Jan. 2008.

[86] P. Dayal, M. Brehler, and M. K. Varanasi, "Leveraging coherent space-time codes for noncoherent communication via training," *IEEE Trans. Inf. Theory*, vol. 50, no. 9, pp. 2058–2080, Sep. 2004.

[87] S. Liu and J.-W. Chong, "An efficient scheme to achieve differential unitary space-time modulation on MIMO-OFDM systems," *Electron. Telecommun. Research Inst. J.*, vol. 26, no. 6, pp. 565–574, Dec. 2004.

[88] K. Aksoy and Ü. Aygölü, "An efficient differential MIMO-OFDM scheme with coordinate interleaving," *EURASIP J. on Wireless Commun. and Networking*, vol. 2008, Article ID 734258, 8 pages, 2008.

[89] I. Barhumi, G. Leus, and M. Moonen, "Optimal training design for MIMO OFDM systems in mobile wireless channels," *IEEE Trans. Signal Process.*, vol. 51, no. 6, pp. 1615–1624, Jun. 2003.

[90] V. Tarokh, N. Seshadri, and A. R. Calderbank, "Space-time codes for high data rate wireless communication: performance criterion and code construction," *IEEE Trans. Inf. Theory*, vol. 44, no. 2, pp. 744–765, Mar. 1998.

[91] S. M. Alamouti, "A simple transmit diversity technique for wireless communications," *IEEE J. Sel. Areas Commun.*, vol. 16, no. 8, pp. 1451–1458, Oct. 1998.

[92] V. Tarokh, H. Jafarkhani, and A. R. Calderbank, "Space-time block codes from orthogonal design," *IEEE Trans. Inf. Theory*, vol. 45, no. 5, pp. 1456–1467, Jul. 1999.

[93] H. Bolcskei and A. J. Paulraj, "Space-frequency coded broadband OFDM systems," in *Proc. IEEE Wireless Commun. Networking Conf.*, vol. 1, Sep. 2000 (6 pages).

[94] H. Bolcskei, M. Borgmann, and A. J. Paulraj, "Impact of the propagation environment on the performance of space-frequency coded MIMO-OFDM," *IEEE J. Sel. Areas Commun.*, vol. 21, no. 3, pp. 427–439, Apr. 2003.

[95] W. Su, Z. Safar, M. Olfat, and K. J. R. Liu, "Obtaining full-diversity space-frequency codes from space-time codes via mapping," *IEEE Trans. Signal Process.*, vol. 51, no. 11, pp. 2905–2916, Nov. 2003.

[96] W. Su, Z. Safar, and K. J. R. Liu, "Full-rate full-diversity space-frequency codes with optimum coding advantage," *IEEE Trans. Inf. Theory*, vol. 51, no. 1, pp. 229–249, Jan. 2005.

[97] Z. Liu, Y. Xin, and G. B. Giannakis, "Space-time-frequency coded OFDM over frequency-selective fading channels," *IEEE Trans. Signal Process.*, vol. 50, no. 10, pp. 2465–2476, Oct. 2002.

[98] B. Lu, X. Wang, and K. R. Narayanan, "LDPC-based space-time coded OFDM systems over correlated fading channels: Performance analysis and receiver design," *IEEE Trans. Commun.*, vol. 50, no. 1, pp. 74–88, Jan. 2002.

[99] M. Fozunbal, S. W. McLaughlin, and R. W. Schafer, "On space-time-frequency coding over MIMO-OFDM systems," *IEEE Trans. Wireless Commun.*, vol. 4, no. 1, pp. 320–331, Jan. 2005.

[100] W. Su, Z. Safar, and K. J. R. Liu, "Towards maximum achievable diversity in space, time, and frequency: performance analysis and code design," *IEEE Trans. Wireless Commun.*, vol. 4, no. 4, pp. 1847–1857, Jul. 2005.

[101] G. D. Golden, C. J. Foschini, R. A. Valenzuela, and P. W. Wolniansky, "Detection algorithm and initial laboratory results using V-BLAST space-time communication architecture," *Elect. Lett.*, vol. 35, no. 1, pp. 14–16, 1999.

[102] T. Marzetta, "BLAST training: estimating channel characteristics for high capacity space-time wireless," in *Proc. 37th Annual Allerton Conf. Commun. Control Comput.*, vol. 37, Sep. 1999, pp. 958–966.

[103] E. G. Larsson, "Performance bounds for MIMO-OFDM channel estimation," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 1901–1916, May 2009.

[104] D. Angelosante, E. Biglieri, and M. Lops, "Sequential estimation of multipath MIMO-OFDM channels," *IEEE Trans. Signal Process.*, vol. 57, no. 8, pp. 3167–3181, Aug. 2009.

[105] G. Coluccia, E. Riegler, C. Mecklenbräuker, and G. Taricco, "Optimum MIMO-OFDM detection with pilot-aided channel state information," *IEEE J. Sel. Areas Signal Process.*, vol. 3, no. 6, pp. 1053–1065, Dec. 2009.

[106] F. Davarian, "High performance digital communications in mobile channels," in *Proc. IEEE Veh. Technol. Conf.*, May 1984, pp. 114–118.

[107] J. McGeehan and A. Bateman, "Phase-locked transparent tone-in-band (TTIB): a new spectrum configuration particularly suited to the transmission of data over SSB mobile radio networks," *IEEE Trans. Commun.*, vol. 32, no. 1, pp. 81–87, 1984.

[108] M. K. Simon, "Dual-pilot tone calibration technique," *IEEE Trans. Veh. Technol.*, vol. 35, no. 2, pp. 63–70, May 1986.

[109] F. Davarian, "Mobile digital communications via tone calibration," *IEEE Trans. Veh. Technol.*, vol. 36, no. 2, pp. 55–62, May 1987.

[110] J. K. Cavers, "The performance of phase locked transparent tone-in-band with symmetric phase detection," *IEEE Trans. Commun.*, vol. 39, no. 9, pp. 1389–1399, Sep. 1991.

[111] A. F. Molisch, Ed., *Wideband Wireless Digital Communications*. Prentice Hall, 2001.

[112] H. Bolcskei, D. Gesbert, C. B. Papadias, and A. J. van der Veen, Eds., *Space-Time Wireless Systems: from Array Processing to MIMO Communications.* Cambridge University Press, 2006.

[113] Q. Huang, M. Ghogho, and S. Freear, "Pilot design for MIMO OFDM systems with virtual carriers," *IEEE Trans. Signal Process.*, vol. 57, no. 5, pp. 2024–2029, May 2009.

[114] D. Y. Kim and K. Feher, "Power suppression at the Nyquist frequency for pilot-aided PAM and QAM systems," *IEEE Trans. Commun.*, vol. 37, no. 9, pp. 984–986, Sep. 1989.

[115] X. Dong, W.-S. Lu, and A. C. K. Soong, "Linear interpolation in pilot symbol assisted channel estimation for OFDM," vol. 6, no. 5, pp. 1910–1920, May 2007.

[116] R. V. Nee and R. Prasad, *OFDM for Wireless Multimedia Communications.* Artech House universal personal communications library, 2000.

[117] J.-Y. Son, J. Kim, and D.-S. Kwon, "The pilot design for channel estimation in a practical MIMO OFDM system," in *Proc. Int. Conf. Comput., Eng. and Inf.*, Apr. 2009, pp. 325–328.

[118] D. Hu, L. Yang, Y. Shi, and L. He, "Optimal pilot sequence design for channel estimation in MIMO OFDM systems," *IEEE Commun. Lett.*, vol. 10, no. 1, pp. 1–3, Jan. 2006.

[119] H. Bölcskei, R. W. Heath, Jr., and A. J. Paulraj, "Blind channel identification and equalization in OFDM-based multiantenna systems," *IEEE Trans. Signal Process.*, vol. 50, no. 1, pp. 96–109, Jan. 2002.

[120] J. M. Mendel, *Lessons in Estimation Theory for Signal Processing, Communications, and Contol.* Prentice Hall, 1995.

[121] R. A. Horn and C. R. Johnson, *Matrix Analysis.* Cambridge University Press, 1985.

[122] S. Roy and C. Li, "A subspace blind channel estimation method for OFDM systems without cyclic prefix," *IEEE Trans. Wireless Commun.*, vol. 1, no. 4, pp. 572–579, Oct. 2002.

[123] X. Mestre and M. A. Lagunas, "Finite sample size effect on minimum variance beamformers: Optimum diagonal loading factor for large arrays," *IEEE Trans. Signal Process.*, vol. 54, no. 1, pp. 69–82, Jan. 2006.

[124] J. W. Silverstein and P. L. Combettes, "Large dimensional random matrix theory for signal detection and estimation in array processing," in *Proc. Stat. Signal Array Process.*, Oct. 1992, pp. 276–279.

[125] C.-C. Tu and B. Champagne, "Subspace tracking of fast time-varying channels in precoded MIMO-OFDM systems," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process.*, Apr. 2009, pp. 2565–2568.

[126] M. Cicerone, O. Simeone, and U. Spagnolini, "Channel estimation for MIMO-OFDM systems by modal analysis/filtering," *IEEE Trans. Commun.*, vol. 54, no. 11, pp. 2062–2074, Nov. 2006.

[127] X.-D. Wang and H. V. Poor, *Wireless Communication Systems: Advanced Techniques for Signal Reception.* Prentice Hall, 2004.

[128] E. de Carvalho and D. T. M. Slock, "Blind and semi-blind FIR multichannel estimation: (global) identifiability conditions," *IEEE Trans. Signal Process.*, vol. 52, no. 4, pp. 1053–1064, Apr. 2004.

[129] D. A. Harville, *Matrix Algebra from a Statistician's Perspective.* Springer-Verlag New York, 1997.

[130] F. Li, H. Liu, and R. J. Vaccaro, "Performance analysis for DOA estimation algorithms: unification, simplification, and observations," *IEEE Trans. Aerosp. Electron. Syst.*, vol. 29, no. 4, pp. 1170–1184, Oct. 1993.

[131] A. Albert, *Regression and the Moore-Penrose Pseudoinverse.* Academic Press, 1972.

[132] A. J. Laub, *Matrix Analysis for Scientists and Engineers.* Soc. for Industrial and Applied Mathematics, 2005.

[133] H. V. Poor, *An Introduction to Signal Detection and Estimation.* Springer-Verlag, 1994.

[134] E. de Carvalho, J. Cioffi, and D. Slock, "Cramer-Rao bounds for blind multichannel estimation," in *Proc. IEEE Global Commun. Conf.*, vol. 2, Dec. 2000, pp. 1036–1040.

[135] B. Champagne, "Adaptive eigendecomposition of data covariance matrices based on first-order perturbations," *IEEE Trans. Signal Process.*, vol. 42, no. 10, pp. 2758–2770, Oct. 1994.

[136] P. Strobach, "Low-rank adaptive filters," *IEEE Trans. Signal Process.*, vol. 44, no. 12, pp. 2932–2947, Dec. 1996.

[137] B. Yang, "Projection approximation subspace tracking," *IEEE Trans. Signal Process.*, vol. 43, no. 1, pp. 95–107, Jan. 1995.

[138] B. Champagne and Q.-G. Liu, "Plane rotation-based EVD updating schemes for efficient subspace tracking," *IEEE Trans. Signal Process.*, vol. 46, no. 7, pp. 1886–1900, Jul. 1998.

[139] X. G. Doukopoulos and G. V. Moustakides, "Fast and stable subspace tracking," *IEEE Trans. Signal Process.*, vol. 56, no. 4, pp. 1452–1465, Apr. 2008.

[140] G. H. Golub and C. F. V. Loan, *Matrix Computation (3rd edition)*. The Johns Hopkins University Press, 1996.

[141] Y. Hua, Y. Xiang, T. Chen, K. Abed-Meraim, and Y. Miao, "A new look at the power method for fast subspace tracking," *Digital Signal Process.*, vol. 9, no. 4, pp. 297–314, 1999.

[142] K. Abed-Meraim, A. Chkeif, and Y. Hua, "Fast orthonormal PAST algorithm," *IEEE Signal Process. Lett.*, vol. 7, no. 3, pp. 60–62, Mar. 2000.

[143] Y. Miao and Y. Hua, "Fast subspace tracking and neural networking learning by a novel information criterion," *IEEE Trans. Signal Process.*, vol. 46, no. 7, pp. 1967–1979, Jul. 1998.

[144] R. Badeau, B. David, and G. Richard, "Fast approximated power iteration subspace tracking," *IEEE Trans. Signal Process.*, vol. 53, no. 8, pp. 2931–2941, Aug. 2005.

[145] D.-Z. Feng and W. X. Zheng, "Fast approximate inverse power iteration algorithm for adaptive total least-squares FIR filtering," *IEEE Trans. Signal Process.*, vol. 54, no. 10, pp. 4032–4039, Oct. 2006.

[146] D. S. Watkins, *Fundamentals of Matrix Computations*. John Wiley and Sons, 2002.

[147] N. L. Owsley, "Adaptive data orthogonalization," in *Proc. IEEE Int. Conf. on Acoust., Speech, Signal Process.*, vol. 3, Apr. 1978, pp. 109–112.

[148] C.-C. Tu and B. Champagne, "Subspace blind MIMO-OFDM channel estimation wih short averaging periods: performance analysis," in *Proc. IEEE Wireless Commun. Networking Conf.*, Mar. 2008, pp. 24–29.

[149] ——, "Subspace-based blind channel estimation for MIMO-OFDM systems with reduced time averaging," *IEEE Trans. Veh. Technol.*, vol. 59, no. 3, pp. 1539–1544, Mar. 2010.

[150] Y.-S. Choi, P. J. Voltz, and F. A. Cassara, "On channel estimation and detection for multicarrier signals in fast and selective Rayleigh fading channels," *IEEE Trans. Commun.*, vol. 49, no. 8, pp. 1375–1387, Aug. 2001.

[151] F. Gao and A. Nallanathan, "Blind channel estimation for MIMO OFDM systems via nonredundant linear precoding," *IEEE Trans. Signal Process.*, vol. 55, no. 2, pp. 784–789, Feb. 2007.

[152] X. Xu, Y. Jing, and X. Yu, "Subspace-base noise variance and SNR estimation for OFDM systems," in *Proc. IEEE Wireless Commun. Networking Conf.*, vol. 1, Mar. 2005, pp. 23–26.

[153] D. S. Bernstein, *Matrix Mathematics: Theory, Facts, and Formulas with Application to Linear Systems Theory.* Princeton University Press, 2005.

[154] R. A. Brualdi and J. J. Q. Massey, "Some applications on elementary linear algebra in combinatorics," *The College Math. J.*, vol. 24, no. 1, pp. 10–19, Jan. 1992.

[155] J. R. Treichler, I. Fijalkow, and C. R. Johnson, Jr., "Fractionally spaced equalizers. How long should they really be?" *IEEE Signal Process. Mag.*, vol. 13, no. 3, pp. 65–81, May 1996.

[156] Z. Ding and Y. Li, *Blind Equalization and Identification.* Marcel Dekker, 2001.

I want to dedicate this work to the French composer and pianist Erik Satie (1866-1925) for his masterpiece:

Gymnopédies

"The melodies of the pieces use deliberate, but mild, dissonances against the harmony, producing a piquant, melancholy effect that matches the performance instructions, which are to play each piece *slowly*, *dolorously* or *gravely*" - Wikipedia