Contributions to blind and semi-blind estimation and performance analysis.

Abla Kammoun

▶ **To cite this version:**

Abla Kammoun. Contributions to blind and semi-blind estimation and performance analysis.. Signal and Image processing. Ecole nationale supérieure des telecommunications - ENST, 2010. English. NNT : . tel-00560663v1

HAL Id: tel-00560663

https://theses.hal.science/tel-00560663v1

Submitted on 28 Jan 2011 (v1), last revised 29 Jun 2011 (v2)

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

TÉLÉCOM PARIS TECH
ÉCOLE DOCTORALE INFORMATIQUE TÉLÉCOMMUNICATION ET ÉLECTRONIQUE

THÈSE

Spécialité : Traitement de signal et des images

par

Abla Kammoun

Contributions à l'estimation aveugle et semi-aveugle et analyse de performances

---

Thèse soutenue le 16 Avril 2010

JURY

Rapporteurs	Prof. Jean-Pierre Delmas	TELECOM & Management Sud Paris (ex INT)
	Prof. Mohamed-Slim Alouini	KAUST University
Examinateurs	Prof. Philippe Loubaton	Université de Marne la vallée
	Prof. David Gesbert	Institut Eurecom
	Dr. Walid Hachem	Chercheur CNRS (Télécom Paris Tech)
	Dr. Jamal Najim	Chercheur CNRS (Télécom Paris Tech)
Directeurs de thèse	Dr. Karim Abed-Meraim	Télécom Paris-Tech
	Dr. Sofiène Affès	INRS-EMT Canada

Remerciements

Cette thèse n'aurait jamais pu voir le jour, sans l'aide et le soutien de plusieurs personnes. Au terme de la rédaction de ce manuscrit, il m'est particulièrement agréable d'exprimer ma gratitude à tous ceux qui m'ont aidé de près ou de loin à la réalisation de ce travail.

Je tiens à remercier en tout premier lieu les rapporteurs de cette thèse Monsieur Jean-Pierre Delmas et Monsieur Mohamed Slim Alouini pour l'intérêt qu'ils ont porté à mon travail. Je remercie également Monsieur Philippe Loubaton et aussi les autres membres du Jury de m'avoir honoré en acceptant d'examiner ma thèse.

Durant cette thèse, j'ai eu la grande chance de collaborer avec Jamal Najim et Walid Hachem. Jamais, je n'aurais pu enrichir mes connaissances en probabilité et statistiques si je n'avais pas bénéficié de leur aide précieux. L'acquisition de ces outils m'ont permis non seulement d'approfondir mon savoir dans le domaine mais aussi de pouvoir affronter avec plus de confiance, tout autre sujet de mathématiques. Pour tout cela, je tiens à les remercier très vivement.

Je veux remercier aussi vivement mon directeur de thèse, Karim-Abed Meraim pour avoir toujours été à l'écoute malgré son emploi de temps chargé, et aussi pour la patience infinie qu'il a su éprouver à mon égard durant les périodes les plus critiques de ma thèse. Outre ses qualités scientifiques indéniables, j'ai beaucoup apprécié son humilité, sa générosité et sa grande bienveillance.

Mes remerciements s'adressent également à Sofiène Affes d'avoir facilité mon séjour à l'INRS et de nous avoir toujours donné à temps des remarques avisés sur les articles soumis.

Enfin, je ne pourrais jamais oublier le soutien constant de mon frère, mes soeurs et mes parents. Les mots ne sauront jamais exprimer l'ampleur de la reconnaissance et de l'amour que j'ai à leur égard. Qu'ils trouvent dans ce travail le fruit de leurs sacrifices, soutiens et encouragements.

Table of contents

List of Figures

Notations and Abbreviations

Abbreviations

a.s. almost surely

AWGN Additive White Gaussian Noise

BER Bit Error Rate

BPSK Binary Phase Shift Keying Modulation

BTS Base station

cdf Cumulative Distribution Function

CDMA Code Division Multiple Access

CST Conventional superimposed training

dB Decibels

DDST Data-Dependent Superimposed Training

DFT Discrete Fourier Transform

DML Deterministic Maximum Likelihood

EM Expectation Maximization algorithm

FA-ML Maximum Likelihood with Finite Alphabet Constraints

GML Gaussian Maximum Likelihood

GSM Global System for Mobile communications

i.i.d. Independent and Identically Distributed

IDFT Inverse Discrete Fourier Transform

LMMSE Linear minimum mean square error

LMS Least Mean Squares Algorithm

LS Least Square

MGF Moment Generating Function

MIMO Multiple Input Multiple Output

MMSE Minimum Mean Square Error

MSE Mean Square Error

OFDM Orthogonal Frequency Division Multiplexing

pdf Probability Density Function

QAM Quadrature Amplitude Modulation

QPSK Quadrature Phase Shift Keying Modulation

RLS Recursive Least Squares Algorithm

SIMO Single Input Multiple Output

SINR Signal-to-Interference plus Noise Ratio

SISO Single Input Single Output

SML Stochastic Maximum Likelihood

SNR Signal to Noise Ratio

ST Superimposed Training

TDMT Time-Division Multiplexed Training

Notations

$*$ Convolution product

$\Im()$ Imaginary part

$\jmath$ The complex number verifying $\jmath^2 = -1$

$()^*$ Complex conjugate

$()^{\mathrm{H}}$ Complex conjugate transpose

$()^{\mathrm{T}}$ Transpose

log Natural Logarithm

$\mathbb{E}$ Statistical expectation

det() Determinant operator for matrices

diag($\mathbf{A}$) A diagonal matrix whose elements correspond to the diagonal element of $\mathbf{A}$

diag($\mathbf{v}$) A diagonal matrix whose diagonal is composed of entries of $\mathbf{v}$

Tr Trace operator

$||\ .\ ||$ Euclidean norm for vectors or spectral norm for matrices

$\|\| \, . \, \|\|_1$	The ℓ_1 norm
$\|\| \, . \, \|\|_p$	The ℓ_p quasi-norm
$\otimes$	Kronecker product operator
$\Re()$	Real part
f'	First Derivative of Function f
f''	Second Derivative of Function f
$f^{(k)}$	k-th Derivative of Function f
$\mathbf{A}^{-1}$	The inverse of matrix $\mathbf{A}$
$\mathbf{A}^{\#}$	The pseudo inverse of matrix $\mathbf{A}$
$\mathbf{I}_N$	Identity matrix of size N

Résumé

L'analyse de performance des systèmes de communication numérique est essentielle à plusieurs égards. Elle permet à titre d'exemple, de trouver les éléments déterminants de leurs performances, et aussi de guider la conception des systèmes futurs. Trouver des formules explicites qui traduisent de façon fidèle le comportement de ces systèmes est l'un des sujets de recherche les plus étudiés depuis quelques années. C'est dans ce contexte que nous situons notre travail dans la première et la deuxième partie du présent rapport. Nous nous intéressons en premier lieu aux performances en terme de taux d'erreur binaire et de probabilité de coupure des systèmes utilisant des séquences d'apprentissage pour l'estimation du canal. En nous basant sur les résultats théoriques obtenus, nous proposons aussi des stratégies optimales pour l'allocation de puissance. En second lieu, nous nous intéressons aux performances des systèmes utilisant le filtre LMMSE dans un contexte multi-utilisateur. Nous montrons que le SINR converge vers une Gaussienne dont on calcule la moyenne et la variance. Pour estimer de façon précise le taux d'erreur binaire et la probabilité de coupure, nous utilisons la distribution Gamma généralisée dont nous ajustons les paramètres de telle façon à ce que ses trois premiers moments égalisent les trois premiers moments asymptotiques. Nous montrons par simulations que cette approche permet une bonne approximation du taux d'erreur binaire et de la probabilité de coupure même pour des systèmes à dimensions réduites.

D'un autre côté, avec des applications de plus en plus exigeantes en terme de débit, la tendance actuelle est de chercher de nouvelles méthodes pour réduire au maximum le nombre de symboles pilotes. A cet égard, les méthodes d'estimation aveugle ont été souvent avancées compte tenu de leur avantage majeur par rapport aux schémas classiques, à savoir la suppression de la séquence d'apprentissage. Cependant, l'implémentation pratique de ces méthodes se heurte à de nombreuses difficultés ayant trait à leur sensibilité à la surestimation de l'ordre et à leur complexité, souvent perçue comme élevée. D'autres méthodes connues sous le nom de méthodes semi-aveugles ont été proposées. Parmi elles, se trouvent des méthodes qui combinent de façon linéaire via un coefficient dit de régularisation, un critère aveugle à un critère moindre carré issu de la séquence d'apprentissage. Le choix du coefficient de régularisation est fait de telle sorte à optimiser l'erreur quadratique moyenne asymptotique, pour laquelle plusieurs chercheurs, confortés par les résultats des simulations supposent l'existence d'un seul minimum. Néanmoins, aucune preuve théorique de ce constat n'a été établi. C'est dans ce contexte que s'inscrivent nos contributions dans la dernière partie de notre rapport. S'agissant des méthodes aveugles, nous proposons une méthode d'estimation robuste à la surestimation de l'ordre, et dont l'implémentation pourrait s'appuyer sur des algorithmes adaptatifs de faible complexité. Nous menons aussi une étude théorique sur la robustesse des méthodes aveugles basées sur la norme ℓ_p. Plus explicitement, nous apportons des preuves théoriques qui témoignent de l'efficacité de l'utilisation de la norme ℓ_p pour robustifier les méthodes aveugles. Concernant les méthodes semi-aveugles utilisant le principe de régularisation, deux contributions ont été apportées. La première porte sur la preuve de la quasi-convexité de l'erreur quadratique moyenne asymptotique. Elle confirme la conjecture déjà énoncée sur l'existence d'un unique coefficient de régularisation optimal. En ce qui concerne la deuxième contribution, nous proposons de substituer le coefficient de régularisation par une matrice de régularisation. Contrairement à toute attente, l'optimisation de l'erreur quadratique moyenne est plus facile dans ce cas, et aboutit à

une forme explicite de la matrice de régularisation optimale.

Mots-clés: Analyse de performance, système MIMO, Système multi-utilisateur, Méthodes d'estimation aveugle, Norme ℓ_p, Méthodes d'estimation semi-aveugle

Introduction

General context and motivations

Over the last decades, wireless communication systems have changed dramatically. While the first generation systems were analog and tailored for voice communications, second-generation systems, namely GSM, employ digital transmission and enable both voice and low-data rate services. With the extensive growth of Internet, new applications involving multimedia services emerged with the third generation systems. To meet the new performance requirements, these latter employ advanced technologies.

One of the major challenges of advanced broadband technologies is to combat the experienced channel fading. While narrow-band channels were valid for modelling first generation systems, they became totally irrelevant for second and third generation systems. As a matter of fact, when operating over broad bandwidths, the channel delay spread becomes higher than the sampling period, and as such, the received signal is the superposition of all echo signals arising due to multi-path propagation. To combat fading and to further increase the data rate for future systems, the use of MIMO techniques has been pushed forward by many theoretical studies. The work of Telatar, showing the great potential of MIMO channels in increasing the system capacity is considered as one of the most important findings in this area. It has spawned an intensive effort to study the statistics of the Shannon mutual information, for more practical (correlated) MIMO channels including the so called Kronecker correlated channels, generally spatially correlated channels and general variance profile channels. Such analysis provides insights about the maximum number of bits that can be transmitted reliably over a given channel. Nevertheless, the data rates stipulated by the mutual information are not achievable in practice. The best that we know, is that there exists an optimal strategy for which the data rate will be maximum. Therefore, other performance indexes, like the BER or the outage probability might be more appropriate especially for sub-optimum strategies. Despite there practical relevance, the study of such performance indexes was limited to very specific situations where the statistics of the post processing SINR could be derived for finite system dimensions. The general case which allows for correlated channel models or imperfect channel estimation was either not investigated or worked out under non-realistic assumptions. This motivates our research work in the two first parts to develop new techniques allowing closed-form derivations for BER and outage probability metrics.

Apart from performance analysis issues, another attractive area of research regarding channel estimation and equalization strategies have spurred the interest of many researchers. To improve the system spectral efficiency, blind methods have been often pushed forward. Nevertheless, their use in practice was limited to specific applications where pilot symbols are unavailable. This is principally attributed to several practical considerations pertaining their high computational complexity as well as their sensitivity towards channel order overmodeling. That's why, a major effort is now devoted to improve the robustness of blind equalization methods.

On the other hand, training based techniques were thus preferred for their low complexity

and their good channel estimation quality. The price to be paid, is a reduced spectral efficiently owing to the transmission of known symbols during the training period. Beyond finding solutions for the practical difficulties experienced by blind techniques, the research trend is now to look forward combined blind and training based methods, referred to as semi-blind techniques, in order to enhance the channel estimation quality or to shorten as much as possible the training periods.

Contributions and organization We organize our report into three parts. The first part deals with the performance analysis of training based systems. After a brief overview on channel models, we investigate in the second chapter, the performance in terms of BER and outage probability of three training based schemes: the time division multiplexing scheme as well as two superimposed training schemes based on the simultaneous transmission of pilot and data symbols. When imperfect channel estimation is taken into account, the statistics of the post-processing SNR has no closed-form expressions for finite size dimensions, and thus the outage probability and the BER expressions cannot be derived, either. To deal with this difficulty, we will rather study the case where the system dimensions grow to infinity at the same pace. In this asymptotic regime, we will show that closed-form expressions for the BER and the outage probability exist, and are accurate even for usual system dimensions. In a second step, we determine the optimal power allocation between data and training that minimizes the BER and the outage probability metrics.

The second part addresses also issues regarding asymptotic performance analysis, but in the context of multi-user systems. After a brief overview on the random matrix theory tools in chapter 3, we study in chapter 4 the asymptotic behaviour of the SINR at the output of the LMMSE receiver. While perfect channel knowledge is now assumed, our approach encompasses general channel variance profiles , thereby making the derived results apply for a large variety of systems ranging from (receive) correlated MIMO systems to CDMA uplink and downlink based systems. Nevertheless, the obtained asymptotic results could not help to yield accurate approximations for the BER and outage probability. To get insights about these performance metrics, we propose in chapter 5 to use an approximation based on the Generalized Gamma distribution.

As for the third part, two main issues are raised. The first one aims at solving the sensitivity of the blind methods to channel order overmodeling. More explicitly, After presenting a brief overview on blind system identification methods in chapter 6, we develop in chapter 7, nonzero delay blind equalization method that combines high performance with robustness to channel order overmodeling. In chapter 8, we focus on blind subspace based methods using ℓ_p quasi norms. We provide for the first time well founded arguments that supports the use of ℓ_p quasi norms to improve the robustness of subspace based methods.

The second issue deals with regularized semi-blind channel estimation. In these techniques, the regularizing scalar plays a key role in the MSE performance. It is set numerically so as to minimize the asymptotic MSE, the uniqueness of its minimum being only conjectured but never proved. In our work, we establish for the first time in chapter 9 the uniqueness of the asymptotic MSE in regularized channel estimation techniques. In a second step, we present in chapter 10 a new method based on matrix regularization. We prove that interestingly not only better performance is obtained but also the optimal regularizing matrix has closed-form expression.

Publications

Our work has resulted in the following publications:

Journal Articles

1. A. Kammoun, M. Kharouf, W. Hachem and J. Najim. A Central Limit Theorem for the SINR at the LMMSE estimator output for large dimensional signals *IEEE Inf. Theory*, Vol. 55 (11), nov. 2009

2. A. Kammoun, M. Kharouf, W. Hachem and J. Najim. "BER and Outage probability approximations for LMMSE detectors on correlated MIMO channels". *IEEE Inf. Theory*, Vol. 55 (10), oct. 2009 .

3. A. Kammoun, K. Abed-Meraim and S. Affès. "Quasi Convexity of the Asymptotic Channel MSE in Regularized Semi-blind Channel Estimation", *submitted to IEEE Inf. Theory*

4. A. Kammoun, K. Abed-Meraim and S. Affès. "Performance Analysis and Optimal Power Allocation for Linear Receivers Based on Superimposed Training", *to be prepared for IEEE Inf. Theory*

Proceedings

1. A. Kammoun, K. Abed-Meraim and S. Affès. "Performance Of Linear Receivers Based On Superimposed Training" *IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Helsinki (Finlande), July 2007.*

2. A. Kammoun, M. Kharouf, W. Hachem and J. Najim, "Fluctuations of the SNR at the Wiener Filter Output for Large Dimensional Signals" *IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Recife (Brazil), July 2008.*

3. A. Kammoun, M. Kharouf, W. Hachem and J. Najim "Outage Probability Approximation for the Wiener Filter SINR in MIMO Systems " *IEEE Workshop on Signal Processing Advances in Wireless Communications (SPAWC), Recife (Brazil), July SPAWC 2008.*

4. A. Kammoun, K. Abed-Meraim and S. Affès "An Efficient Regularized Semi Blind Estimator" *IEEE International Conference on Communications (ICC), Dresden (Germany) June 2009.*

5. A. Kammoun, K. Abed-Meraim and S. Affès "Superimposed or Time-multiplexed Training: A Performance Comparison" *17th European Signal Processing Conference (Eusipco), Glasgow Scotland August 2009* .

6. A. Kammoun, K. Abed-Meraim and S. Affès " Regularized Semi-Blind Estimator Over MIMO-OFDM Systems" *IEEE International Symposium on Signal Processing and Information Technology* Ajman (UAE) December 2009

7. A. Kammoun, K. Abed-Meraim and S. Affès "Blind Nonzero Delay MMSE Equalizer for SIMO FIR Systems " *Submitted to SPAWC 2010*

8. A. Kammoun, K. Abed-Meraim and S. Affès "Robustness of Blind Subspace Techniques using ℓ_p Quasi-norms" *Submitted to SPAWC 2010*

Part I

Performance analysis for MIMO systems with imperfect channel estimation

Introduction

At the start of 21st century, wireless mobile communication systems have witnessed an important growth that leaded to higher data rate and better spectral efficiency. This growth is particularly attributed to the emergence of MIMO systems, allowing the Shannon mutual information to increase linearly with the minimum number of transmit and receive antennas, [98]. However, the good performance of MIMO systems requires a priori knowledge of the channel at the receiver. In many practical systems, the receiver estimates the channel by time division multiplexing training symbols with the data. This option is commonly used in current digital communication systems for its low complexity and its good performance in terms of channel estimation quality. However, it might be not the optimal option in terms of bandwidth utilization since during the training period, no data can be transmitted. An other alternative is to transmit data and training simultaneously. This option is referred to as superimposed training, and in contrast to the time division multiplexed training technique, is efficient in terms of bandwidth utilization. However, since data act as an additive source of noise, channel estimation quality is severely affected. To deal with this problem, Ghogho *et al* proposed in [40] to distort the transmitted data symbols so that they become orthogonal to the pilot symbols. This scheme is referred to as data-dependent superimposed training (DDST) and was shown by simulations to outperform the conventional scheme.

In this part, we propose to study the aforementioned training design schemes in terms of BER and outage probability, while using linear receivers. Previous works have dealt with performance analysis issues by either assuming perfect channel knowledge or making some unrealistic assumptions about the decorrelation of the noise and the channel estimation error. Without these assumptions, performance analysis for finite size systems, is not possible.

To circumvent this difficulty, we carry out performance analysis in the asymptotic regime. More precisely, we consider the case where the sample size N, the number of transmitting antennas K and the number of receiving antennas M go to infinity, but with constant ratios, i.e. $\frac{K}{N} \to c_1$ and $\frac{M}{K} \to c_2$. For the time-division multiplexed training (TDMT) scheme, we assume on top of that, that the training and data frame sizes denoted by N_1 and N_2 converge to infinity with constant ratio, i.e. $\frac{N_2}{N_1} \to r$. Working with the asymptotic regime, allows us to simplify the derivations and at the same time, we observe that the obtained results apply as well to usual sample and antenna-array sizes.

We divide this part into two chapters. The first one, looks over previous works regarding training based schemes, while the second one derives under the asymptotic regime closed-form expressions for the bit error rate and outage probability for the three aforementioned training based schemes. Optimal power allocation that optimize the bit error rate and outage probability indexes are also provided.

Chapter 1

Overview on training based techniques

In this chapter, we briefly review the main characteristics of fading channels encountered in wireless communication systems. To estimate these channels at the receiver side, several training strategies have been proposed. They are briefly reviewed in this chapter, as well as, the performance indexes that are used to compare between them.

1.1 Fading Channel

In a wireless communication system, the received signal may experience fluctuations that affect its envelope and its phase. These fluctuations, referred to as fading, are essentially caused by two physical phenomena. The first one is related to the mobility of the transmitter, resulting in a shift in the frequency of the transmitted signal. This phenomenon is known as the *Doppler spread*. The second phenomenon is related to the multiple reflections experienced by the transmitted wave. Indeed, signals may be reflected by various scatterers like buildings, trees, objects, etc..., in such a way that the transmitted wave travel to the receiver via different paths. The maximum delay of the significant multipath component is referred to as the *maximum delay spread.* All these fluctuations induce frequency and time variations to the fading channel. To characterize these variations, we often define the following parameters:

- The coherence time which measures the period of time over which the fading is essentially invariant (the time duration over which the time correlation function taken at the same frequency is above a certain threshold). The coherence time T_c is often related to the channel Doppler spread f_d by:
$$T_c \simeq \frac{1}{f_d}$$
If the signal time duration T_c is smaller than the channel's coherence time T_c, the fading is said to be slow, otherwise it is considered to be fast.

- The coherence bandwidth which represents the frequency bandwidth over which the fading is essentially invariant (the bandwidth over which the frequency correlation function taken at the same time is above a certain threshold). The coherence bandwidth B_c is often related to the maximum delay spread $\tau_{\max}$ by:
$$B_c \simeq \frac{1}{\tau_{\max}}$$

The coherence time and the coherence bandwidth characterizes the nature of fading.

- Slow/Fast fading channel: if the symbol time duration T_s is smaller than the channel's coherence time, the channel fading is said to be slow, otherwise, it is considered to be fast. For slow fading channels, many successive symbols undergo the same particular fade level, thus leading to burst errors. Nevertheless, fewer training symbols per unit time are required to estimate the channel at the receiver side.
- Flat/frequency selective fading channel: If the transmitted signal bandwidth is much smaller than the channel's coherence bandwidth B_c, the channel fading is said to be flat otherwise it is considered to be frequency selective.

1.2 Channel Model

1.2.1 Single Input Single Ouput channel (SISO)

During the coherence time period, the channel between two antennae is often modeled as a linear filter characterized by the following finite length impulse response:

$$h(t) = \sum_{l=1}^{L} h_l \delta(t - \tau_l) \tag{1.1}$$

where h_l and τ_l correspond to the random channel attenuations and delays, respectively. For flat fading channel, (1.1) becomes:

$$h(t) = h_0 \delta(t)$$

Consequently, the received signal y is given by:

$$y(t) = h * s(t) = h_0 s(t) + v(t)$$

where $s(t)$ denotes the transmitted symbol at time t, and $v(t)$ is the additive Gaussian noise.

1.2.2 Multiple Input Multiple Output Channel (MIMO)

For MIMO systems, the channel impulse response observed at the m-th receiving antenna is given by:

$$h_m(t) = \sum_{k=0}^{K} \sum_{l=0}^{L_k} h_{k,m} \delta(t - l)$$

where K denotes the number of transmitting antennae and L_k denotes the channel order corresponding to the kth transmitting antenna, (See fig. 1.1). The received signal observed by the m-th antenna is given by:

$$y_m(t) = \sum_{k=0}^{K} \sum_{l=0}^{L_k} h_{k,m} s_k(t - l) + v_m(t)$$

where $s_k(t)$ denotes the transmitted signal by the k-th antenna at time t and $v_m(t)$ is the noise experienced at the mth receiving antenna.

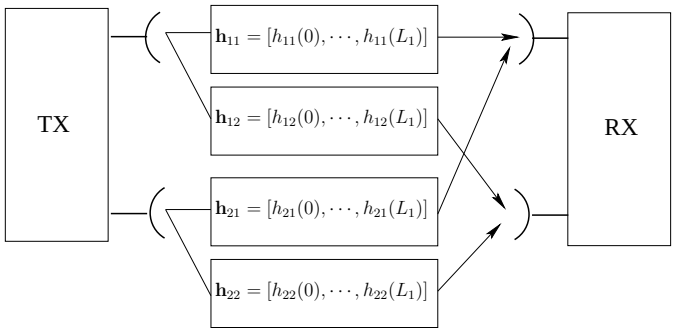


Figure 1.1: Block diagram of a 2×2 MIMO frequency selective fading channel

Hence, over the M receiving antennae, the received signal at time t is given by:

$$\mathbf{y}(t) = \begin{bmatrix} y_1(t) \\ \vdots \\ y_M(t) \end{bmatrix} = \sum_{k=1}^{K} \sum_{l=0}^{L_k} \mathbf{h}_k(l) s_k(t-l) + \mathbf{v}(t) \tag{1.2}$$

where $\mathbf{h}_k(l) = [h_{k,1}(l), \cdots, h_{k,M}(l)]^{\mathrm{T}}$ represents the channel response vector between the kth transmit antenna and the M receiving antenna. For sake of simplicity, let assume that all the channels have the same order L, i.e. $L_k = L$ for all k. Then (1.2) becomes:

$$\begin{aligned} \mathbf{y}(t) &= \sum_{k=1}^{K} [\mathbf{h}_k(0), \cdots, \mathbf{h}_k(L)] \begin{bmatrix} s_k(t) \\ \vdots \\ s_k(t-L) \end{bmatrix} + \mathbf{v}(t) \\ &= \sum_{k=1}^{K} \mathbf{H}_k \begin{bmatrix} s_k(t) \\ \vdots \\ s_k(t-L) \end{bmatrix} + \mathbf{v}(t) \end{aligned}$$

where

$$\mathbf{H}_k = [\mathbf{h}_k(0), \cdots, \mathbf{h}_k(L)] = \begin{bmatrix} h_{11}(0) & \cdots & h_{1k}(L) \\ \vdots & & \vdots \\ h_{Mk}(0) & \cdots & h_{Mk}(L) \end{bmatrix}.$$

Let assume that the channel is constant over N consecutive symbols which implies that the coherence time is greater than NT_s. Stacking N successive observations into vector $\mathbf{y} = [\mathbf{y}(n), \cdots, \mathbf{y}(n-N)]$, we get:

$$\mathbf{y} = \sum_{k=1}^{K} \boldsymbol{\mathcal{I}}_N(\mathbf{H}_k)\mathbf{s}_k + \mathbf{v} \tag{1.3}$$

where $s_k = [s_k(n), \cdots, s_k(n-N-L+1)]^{\mathrm{T}}$, $\mathbf{v} = [\mathbf{v}(n), \cdots, \mathbf{v}(n-N+1)]^{\mathrm{T}}$ and $\boldsymbol{\mathcal{I}}_N(\mathbf{H}_k)$ is the $MN \times (N+L)$ block Toeplitz matrix with first line block $[\mathbf{h}_k(0), \cdots, \mathbf{h}_k(L), \mathbf{0}, \cdots, \mathbf{0}]$ and first column block $\left[\mathbf{h}_k^{\mathrm{T}}(0) \;\; \mathbf{0}_{1\times M(N-1)}\right]^{\mathrm{T}}$. Denote by $\boldsymbol{\mathcal{I}}_N(\mathbf{H}) = [\boldsymbol{\mathcal{I}}_N(\mathbf{H}_1), \cdots, \boldsymbol{\mathcal{I}}_N(\mathbf{H}_K)]$ and $\mathbf{s} = [\mathbf{s}_1^{\mathrm{T}}, \cdots, \mathbf{s}_K^{\mathrm{T}}]^{\mathrm{T}}$, (1.3) becomes:

$$\mathbf{y} = \boldsymbol{\mathcal{I}}_N(\mathbf{H})\mathbf{s} + \mathbf{v}. \tag{1.4}$$

1.2.3 MIMO flat fading channel

For MIMO flat fading channels, only the first path is considered. The observed signal at the mth antenna is therefore given by:

$$\mathbf{y}_m(t) = \sum_{k=1}^{K} h_{mk} s_k(t) + v_m(t)$$

where h_{mk} represents the fading coefficient between the k-th transmitting antenna and the m-th receiving antenna, (See fig.1.2)

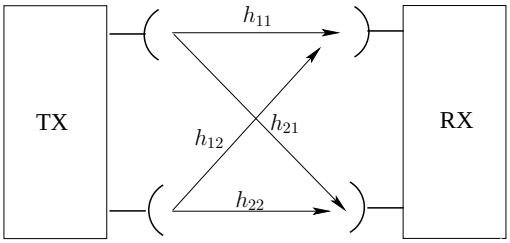


Figure 1.2: Block diagram of a 2×2 MIMO flat fading channel

The output of the M receiving antenna is therefore given by:

$$\mathbf{y}(t) = \mathbf{H}\mathbf{s}(t) + \mathbf{v}(t) \tag{1.5}$$

where $\mathbf{H}$ is the $M \times K$ channel gain matrix given by:

$$\mathbf{H} = \begin{bmatrix} h_{11} & \cdots & h_{1K} \\ \vdots & \ddots & \vdots \\ h_{M,1} & \cdots & h_{M,K} \end{bmatrix}$$

Assuming that the channel is constant over N successive time symbols, we get:

$$\mathbf{Y} = \mathbf{H}\mathbf{S} + \mathbf{V}$$

where $\mathbf{Y}$, $\mathbf{S}$ and $\mathbf{V}$ gather the N last observations of the signal vectors $\mathbf{y}(t)$, $\mathbf{s}(t)$ and $\mathbf{v}(t)$, i.e $\mathbf{Y} = [\mathbf{y}(n), \cdots, \mathbf{y}(n-N+1)]$, $\mathbf{S} = [\mathbf{s}(n), \cdots, \mathbf{s}(n-N+1)]$ and $\mathbf{V} = [\mathbf{v}(n), \cdots, \mathbf{v}(n-N+1)]$.

1.3 Selective fading vs flat fading channels

Unlike frequency selective fading channels, flat fading channels allow low-complexity detection since symbols can be retrieved easily by inverting the channel matrix. The data detection for frequency selective fading channels can be performed optimally using the viterbi algorithm with a complexity that is exponential with the number of paths.

In practice, the propagation environment might be rich with scatterers. This is the case for instance of urban and indoor environments. The channel is then highly frequency selective. However, it may be possible to get effective flat fading channels through the adding of some signal processing. This is, for instance, the case of the OFDM technology which transforms the multipath fading channel into parallel flat fading channels.

On the other hand, frequency selective channels offer better diversity, since each symbol is being transmitted through many paths. If one path undergoes a severe fading, we can expect that other paths will have enough energy to carry the transmitted symbol. For this reason, in practice, OFDM is used in conjunction with channel coding techniques, to overcome the incurred loss in multipath diversity [28].

The following table summarizes the advantages and drawbacks of frequency selective and flat fading channels:

Flat Fading channel	Frequency selective channel
✔Low complexity	✗ High complexity
✗Loss of multipath diversity	✔ Multipath diversity

Table 1.1: Comparison between frequency selective and flat fading channels

1.4 Training design techniques for flat fading channels

1.4.1 Time Divison multiplexed training for flat fading channels

Time-Division Multiplex Training (TDMT) is the most commonly used technique for channel estimation, [20]. Because of its simplicity, it has been used in many practical communication systems, e.g., in Global System for Mobile Communications (GSM) [89]. Although accurate channel estimation can be obtained with low-computational-complexity receivers, this technique results in a low bandwidth efficiency, especially when a large number of pilots is required.

For flat fading channels, the transmission for TDMT schemes has two phases (See fig.1.3):

1. **First phase :** In the first phase, each transmitting antenna sends N_1 pilot symbols. Let $\mathbf{Y}_1 = [\mathbf{y}_1, \cdots, \mathbf{y}_{N_1}]$ be the received data matrix within N time slots, then $\mathbf{Y}_1$ writes as:

$$\mathbf{Y}_1 = \mathbf{H}\mathbf{P}_t + \mathbf{V}_1$$

where $\mathbf{V}_1$ is the $M \times N_1$ Gaussian noise matrix whose entries are i.i.d. with zero mean and variance σ_v^2, and $\mathbf{P}_t$ is the $K \times N_1$ pilot matrix. It has been proved in [56] that the Mean Square Error (MSE) of the channel estimation is minimized subject to a fixed training power $\sigma_{P_t}^2$, when the pilot matrix satisfies:

$$\mathbf{P}_t\mathbf{P}_t^{\mathrm{H}} = N_1\sigma_{P_t}^2\mathbf{I}_K.$$

2. **Second Phase:** In the second phase, N_2 data symbols with power σ_s^2 are sent by each antenna so that the received signal $\mathbf{Y}_2 = [\mathbf{y}_{N_1+1}, \cdots, \mathbf{y}_N]$ writes as:

$$\mathbf{Y}_2 = \mathbf{H}\mathbf{S} + \mathbf{V}_2,$$

where $\mathbf{S}$ is the $K \times N_2$ data matrix with i.i.d. data symbols of power σ_s^2 and $\mathbf{V}_2$ is the $M \times N_2$ additive noise matrix.

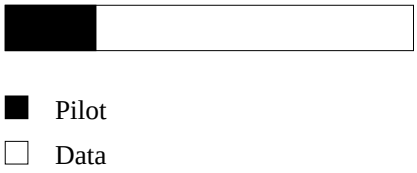


Figure 1.3: Time-multiplexed training scheme

Figure 1.4: Superimposed training scheme

1.4.2 Superimposed training (ST)

Recently, superimposed training (ST) has evolved as a new promising alternative to TDMT schemes due to its high bandwidth efficiency. Indeed, it consists in transmitting pilot and data symbols simultaneously at the same time and on the same frequency domains, fig. 1.4.

More explicitly, the received signal for flat fading channels could be expressed as :

$$\mathbf{Y} = \mathbf{HS} + \mathbf{V}$$

where $\mathbf{S} = \mathbf{W} + \mathbf{P}$, ($\mathbf{W}$ and $\mathbf{P}$ are respectively data and pilot symbols) Since during channel estimation, data symbols act as a source of noise, channel estimation is affected. To mitigate the cross-correlation between training and data symbols, the first proposed (ST) schemes use periodic pilot sequence that ensures cyclostationarity of the transmitted signals. This assumes a long data sequence, in order to ensure approximately a zero cross correlation between data and pilots. In the literature, the performance of the superimposed training in terms of maximum mutual information has been investigated. It has been shown that even after power allocation optimization, the ST scheme entails a gain over the TDMT scheme only in special scenarios (many receive antennas and/or short coherence time). In other scenarios, the ST scheme suffers from high channel estimation errors and its gain over the TDMT scheme is often lost, [13, 27].

1.4.3 Alternatives to Conventional superimposed training

Because of the low performance of the conventional ST scheme, many alternatives have been proposed. Among them, we cite principally:

1. Orthogonal affine precoding: Affine precoders were first proposed in [70] and investigated again in [68], [80] and [81]. They were principally proposed for frequency selective channels and can be classified into two classes:

   - The first class concerns block transmissions with cyclic prefix and was mainly studied by Giannakis *et al* in [81, 80] for SISO systems and then by Scaglione *et al* in [108, 110, 109] for MIMO systems. In this class, the transmitted data vector $\mathbf{w}$ is precoded

by a tall matrix **F**. This latter is devised so that the transmitted signal sequence **Fw** is orthogonal to the training sequence. The channel and symbol estimation are then decoupled.

- The second class concerns block transmissions without cyclic prefix. In this case, the orthogonality between signal and training sequences cannot be obtained by using the previously described precoding. Instead, a source data matrix **W** is first built from the source data vectors **w** and then post-multiplied by a fat (more columns than rows) precoder matrix **F** thus yielding **SF**. This technique was proposed for SISO systems in the work of [85] and has been recently generalized for MIMO systems in [102].

2. Data-dependent superimposed training (DDST): In Affine precoding schemes, the introduced redundancy serves essentially to ensure the orthogonality between data and training sequences. The increasing of the diversity gain can be also obtained but only when the redundancy is higher than a certain threshold. Motivated by this consideration, Ghogho *et al* [40] propose to use a square rank-deficient precoder matrix in order to improve the bandwidth efficiency utilization. Symbol recovery is still possible by virtue of the finite alphabet property of the data symbols. This technique was proposed for flat fading channels [40] and also for frequency selective channels [39, 41].

1.5 Performance indexes

The performance of digital communication systems can be measured by using many indexes. Among them, we can cite:

- Signal to Noise Ratio: is measured at the output of the receiver and as such is directly related to the data detection process.

- Outage probability: Outage occurs when the received power is less than the receiver's sensitivity. The probability of such event, is given by the probability that the output SNR falls below a certain threshold depending on the receiver's sensitivity.

- Bit error rate: Of the commonly known performance indexes, the bit error rate is the one that is most revealing about the system behaviour and the one most often used to evaluate the quality of service in communication systems. On the other hand, it is among the most difficult to compute.

1.6 Optimal Power allocation

The power allocation between training and data has a great impact on the system performance. On the one hand, to achieve high quality channel estimate, a high proportion of power needs to be spent for training transmission, which leaves little power to data. On the other hand, if too little power is given to training, the channel estimation will be poor, thus affecting the system performance. In the literature, many works propose to optimize the power allocation so as to minimize the mutual information [56, 107] or to maximize the SNR [81].

1.7 Conclusion

In this chapter, we have presented the most important characteristics of fading channels encountered in wireless communication systems. We have also reviewed the training design techniques that are used to combat fading channels. The definition and the significance of the most pertinent performance indexes are also provided. These performance indexes will be used in the next chapter to make a performance comparison between three training based schemes.

Chapter 2

Performance Analysis of Training based techniques

As mentioned in the previous chapter, the most known non-blind training based techniques can be classified into two categories: superimposed training (ST) and time-division multiplexed training (TDMT). While the TDMT based schemes use training periods where no data is transmitted, their ST counterparts, transmit data and training simultaneously. The price to be paid is a poor channel estimation quality for conventional ST schemes or a distortion on the data symbols for data-dependent superimposed training (DDST) based schemes.

2.1 Performance Analysis of Time-Division multiplexed training

Although the TDMT scheme was proposed a long time ago, many questions related to the bit error rate performance and the allocation of power between training and data remain unanswered. While issues related to the capacity and the mutual information have been widely investigated in previous works, performance indexes like the bit error rate and the outage probability have not been enough studied. The few works that have focused on this issue, have been based on unrealistic assumptions like the uncorrelation between the noise and the channel estimation error, [112, 2]. These assumptions make calculations feasible for fixed size dimensions, but are far away from being realistic. To make derivations possible while keeping realistic conditions, we will relax the assumption of finite size dimensions by allowing the system dimensions to grow to infinity at the same rate. We show that in this case, closed-form expressions for the BER and outage probability are possible and are accurate even for usual system dimensions.

2.1.1 Channel estimation

We recall that for the TDMT scheme, the receiver performs the channel estimation by using only the received symbols corresponding to the training period. Assume that N_1 symbols have been devoted to training, we have:

$$\mathbf{Y}_1 = \mathbf{H}\mathbf{P}_t + \mathbf{V}_1$$

where $\mathbf{Y}$ is the $M \times N_1$ received matrix, and $\mathbf{H}$ is the $M \times K$ channel matrix with i.i.d. centered entries with variance $\frac{1}{K}$. The entries of the additive Gaussian noise matrix $\mathbf{V}_1$ are also assumed to be i.i.d. with zero mean and variance σ_v^2. Assuming that the receiver estimates the channel

in the least square (LS) sense, the channel estimate is thus given by:

$$\begin{aligned}\widehat{\mathbf{H}}_t &= \mathbf{Y}_1\mathbf{P}_t^{\mathrm{H}}\left(\mathbf{P}_t\mathbf{P}_t^{\mathrm{H}}\right)^{-1}\\ &= \mathbf{H} + \mathbf{V}_1\mathbf{P}_t^{\mathrm{H}}\left(\mathbf{P}_t\mathbf{P}_t^{\mathrm{H}}\right)^{-1}\\ &= \mathbf{H} + \Delta\mathbf{H}_t\end{aligned}$$

where $\Delta\mathbf{H}_t = \mathbf{V}_1\mathbf{P}_t^{\mathrm{H}}\left(\mathbf{P}_t\mathbf{P}_t^{\mathrm{H}}\right)^{-1}$ and the index t refers to the TDMT scheme.

Thus, the mean square error writes as:

$$\mathrm{MSE}_t = M\sigma_v^2\mathrm{Tr}\left(\left(\mathbf{P}_t\mathbf{P}_t^{\mathrm{H}}\right)^{-1}\right). \tag{2.1}$$

As it has been already mentioned in chapter 1 section 1.4.1, the optimal training matrix that minimizes the MSE under a constant training energy $N_1\sigma_{P_t}^2$ should satisfy:

$$\mathbf{P}_t\mathbf{P}_t^{\mathrm{H}} = N_1\sigma_{P_t}^2\mathbf{I}_K,$$

where $\sigma_{P_t}^2$ denotes the amount of power devoted to the transmission of a pilot symbol. The optimal minimum value for the MSE_t is then given by:

$$\mathrm{MSE}_t = \frac{KM\sigma_v^2}{N\sigma_{P_t}^2}.$$

2.1.2 Data Detection

In the data transmission phase, the linear receiver uses the channel estimate in order to retrieve the transmitted data. Let $\mathbf{Y}_2$ denotes the $M \times N_2$ received matrix corresponding to data, then $\mathbf{Y}_2$ can be expressed as:

$$\mathbf{Y}_2 = \mathbf{HS} + \mathbf{V}_2$$

where $\mathbf{V}_2$ is independent of $\mathbf{V}_1$ but with entries being identically distributed as that of $\mathbf{V}_1$.

After channel inversion, the estimated data matrix is given by:

$$\widehat{\mathbf{S}} = \left(\widehat{\mathbf{H}}_t\right)^{\#}\mathbf{Y}_2. \tag{2.2}$$

where $\left(\widehat{\mathbf{H}}_t\right)^{\#}$ denotes the pseudo-inverse matrix of $\mathbf{H}$.

Assuming that the channel estimation error is small, the pseudo-inverse of the estimated matrix can be approximated by the linear part of the Taylor expansion as [69, Theorem 5, page 174]:

$$\left(\widehat{\mathbf{H}}_t\right)^{\#} = \mathbf{H}^{\#} - \mathbf{H}^{\#}\left(\Delta\mathbf{H}_t\right)\mathbf{H}^{\#} + \mathbf{H}^{\#}\left(\mathbf{H}^{\#}\right)^{\mathrm{H}}\Delta\mathbf{H}_t\left(\mathbf{I}_M - \mathbf{H}\mathbf{H}^{\#}\right) \tag{2.3}$$

Substituting $\mathbf{H}^{\#}$ by $\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{H}}$ in (2.3), we obtain:

$$\left(\widehat{\mathbf{H}}_t\right)^{\#} = \mathbf{H}^{\#} - \mathbf{H}^{\#}\left(\Delta\mathbf{H}_t\right)\mathbf{H}^{\#} + \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\Delta\mathbf{H}_t^{\mathrm{H}}\mathbf{\Pi} \tag{2.4}$$

where $\mathbf{\Pi} = \mathbf{I}_M - \mathbf{H}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\mathbf{H}^{\mathrm{H}}$ is the orthogonal projector on the null space of $\mathbf{H}$. Hence, the zero-forcing estimate of the transmitted matrix can be expressed as:

$$\widehat{\mathbf{S}} = \mathbf{S} - \mathbf{H}^{\#}\Delta\mathbf{H}_t\mathbf{S} + \left(\mathbf{H}^{\#} - \mathbf{H}^{\#}\Delta\mathbf{H}_t\mathbf{H}^{\#}\right)\mathbf{V}_2 + \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\left(\Delta\mathbf{H}_t\right)^{\mathrm{H}}\mathbf{\Pi}\mathbf{V}_2.$$

Consequently, the effective post-processing noise $\Delta\mathbf{S} = \widehat{\mathbf{S}} - \mathbf{S}$ could be written as:

$$\Delta\mathbf{S} = -\mathbf{H}^{\#}\Delta\mathbf{H}_t\mathbf{S} + \left(\mathbf{H}^{\#} - \mathbf{H}^{\#}\Delta\mathbf{H}_t\mathbf{H}^{\#} + \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\left(\Delta\mathbf{H}_t\right)^{\mathrm{H}}\mathbf{\Pi}\right)\mathbf{V}_2.$$

2.1.3 Performance Analysis for the TDMT scheme

For finite system dimensions, the performance analysis of the TDMT scheme is difficult. Instead, we will work under the asymptotic regime when N_1, N_2, M and K grow to infinity with a constant rate, $\frac{K}{N_1+N_2} \to c_1$ with $0 < c_1 < 1$, $\frac{M}{K} \to c_2$ and $\frac{N_2}{N_1} \to r$. Next, the notation $K \to +\infty$, will refer to this asymptotic regime. Let us now recall the main assumptions:

A1. *The channel matrix* $\mathbf{H}$ *is independent from the noise, it has i.i.d. complex Gaussian centered entries with variance* $\frac{1}{K}$.

A2. *Matrices* $\mathbf{V}_1$ *and* $\mathbf{V}_2$ *are independent. Their entries are i.i.d. complex Gaussian with zero mean and variance* σ_v^2.

A3. *The entries of the data matrix* $\mathbf{S}$ *are assumed to be i.i.d. They are also bounded and independent from* $\mathbf{H}$ *and the noise matrices* $\mathbf{V}_1$ *and* $\mathbf{V}_2$.

A4. *The training matrix* $\mathbf{P}_t$ *verifies:* $\mathbf{P}_t\mathbf{P}_t^{\mathrm{H}} = N_1\sigma_{P_t}^2\mathbf{I}_K$.

2.1.3.1 Bit error rate performance

In order to evaluate the bit error rate performance, we need to analyse the asymptotic behaviour of the post-processing noise observed at each entry of matrix $\Delta\mathbf{S}$. Using the 'characteristic function' approach, we can prove that conditioned on the channel, the noise behaves asymptotically like a Gaussian random variable. This result is stated in the following theorem but its proof is postponed in appendix A.1.

Theorem 2.1. *Under assumptions* **A1**, **A2**, **A3** *and* **A4** *and conditioned on the channel, the post-processing noise experienced by the i-th antenna at each time k,* $\Delta\mathbf{S}(i,k)$, *for the TDMT scheme behaves in the asymptotic regime as a Gaussian random variable:*

$$\mathbb{E}\left[e^{\jmath\Re(z^*\Delta\mathbf{S}(i,k))}\right] - e^{-\frac{\sigma_s^2\delta_t\left[\left(\mathbf{H}^H\mathbf{H}\right)^{-1}\right]_{i,i}|z|^2}{4}} \xrightarrow[K\to+\infty]{} 0$$

where

$$\delta_t = c_1(1+r)\frac{\sigma_v^2}{\sigma_P^2} + \frac{\sigma_v^2}{\sigma_s^2} + \frac{c_1(1+r)(c_2+1)\sigma_v^4}{\sigma_s^2\sigma_P^2(c_2-1)}.$$

The gaussianity of the post-processing noise being verified in the asymptotic case, we can derive the bit error rate for QPSK constellation and Gray encoding as [87]:

$$\mathrm{BER} = \mathbb{E}Q(\sqrt{x}) \tag{2.5}$$

where the expectation is taken with respect to the probability density function of the post processing SNR at the i-th branch γ_t defined as:

$$\gamma_t = \frac{1}{\delta_t\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}}.$$

From [46] and [113], we know that $\frac{1}{\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}}$ is a weighted chi-square distributed random variable with $2(M-K+1)$ degrees of freedom, whose density function is given by:

$$f(x) = \frac{K^{M-K+1}x^{M-K}e^{-Kx}}{(M-K)!}\mathbf{1}_{[0,+\infty[},$$

where $\mathbf{1}_{[0,+\infty[}$ is the indicator function corresponding to the interval $[0,+\infty[$. Hence, the probability density function of γ_t is given by:

$$f_{\gamma_t}(x) = \frac{(K\delta_t)^{M-K+1}x^{M-K}\exp(-K\delta_t x)}{(M-K)!}\mathbf{1}_{[0,+\infty[} \tag{2.6}$$

Plugging (2.6) into (2.5), we get:

$$\text{BER} = \frac{(K\delta_t)^{M-K+2}}{(M-K)!}\int_0^{+\infty} x^{M-K}\exp(-K\delta_t x)Q(\sqrt{x})dx \tag{2.7}$$

To compute (2.7), we use the following integral function:

$$J(m,a,b) = \frac{a^m}{\Gamma(m)}\int_0^{+\infty}\exp(-ax)x^{m-1}Q(\sqrt{bx})dx. \tag{2.8}$$

The BER is therefore equal to:

$$\text{BER} = J(M-K+1, K\delta_t, 1). \tag{2.9}$$

The integral in (2.8) has been shown to have, for $c>0$ the following closed-form expression, [99]:

$$J(m,a,b) = \frac{\sqrt{c/\pi}\Gamma(m+\frac{1}{2})}{2(1+c)^{m+\frac{1}{2}}\Gamma(m+1)}{}_2F_1(1, m+\frac{1}{2}; m+1; \frac{1}{1+c}), \quad c = \frac{b}{2a}$$

where ${}_2F_1(p,q;n,z)$ is the Gauss hyper-geometric function [49]. If $c=0$ equivalently $b=0$, it is easy to note that $J(m,a,0)$ is equal to $\frac{1}{2}$. When m is restricted to positive integer values, the above equation can be further simplified to [115]:

$$J(m,a,b) = \frac{1}{2}\left[1-\mu\sum_{k=0}^{m-1}\binom{2k}{k}\left(\frac{1-\mu^2}{4}\right)^k\right] \tag{2.10}$$

where $\mu = \sqrt{\frac{c}{1+c}}$. Plugging (2.10) into (2.9), we get:

$$\text{BER} = \frac{1}{2}\left[1-\mu_t\sum_{k=0}^{M-K}\binom{2k}{k}\left(\frac{1-\mu_t^2}{4}\right)^k\right] \tag{2.11}$$

where $\mu_t = \sqrt{\frac{1}{2K\delta_t+1}}$.

2.1.3.2 Outage Probability

Outage probability is the probability that the post-processing SNR γ_t falls below a certain threshold. It is therefore given by:

$$\mathbb{P}\left[\gamma_t \le \gamma_s\right] = \int_0^{\gamma}\frac{(K\delta_t)^{M-K+1}x^{M-K}\exp(-K\delta_t x)}{(M-K)!}dx \tag{2.12}$$

$$= \int_0^{K\delta_t\gamma_s}\frac{x^{M-K}}{(M-K)!}\exp(-u)du \tag{2.13}$$

$$= \frac{1}{(M-K)!}\gamma\left(M-K+1, K\delta_t\gamma_s\right) \tag{2.14}$$

where $\gamma(a,x)$ is the lower incomplete Gamma function defined as, [49]:

$$\gamma(a,x) = \int_0^x t^{a-1}\exp(-t)dt.$$

2.1.3.3 Optimal Power allocation

Referring to the expressions of outage probability and BER, we can easily see that the optimal amount of power allocated to data and pilot for the TDMT scheme is the one that minimizes δ_t. Let $\tilde{c}_1 = (1+r)c_1$, then minimizing δ_t with respect to σ_s^2 and $\sigma_{P_t}^2$ under the constraint that $N_1\sigma_{P_t}^2 + N_2\sigma_s^2 = (N_1+N_2)\sigma_T^2$ (σ_T^2 being the mean energy per symbol), we get:

$$\sigma_s^2 = \frac{(1+r)\sigma_T^2\sqrt{r\left((1+r)\sigma_T^2 + \frac{\tilde{c}_1\sigma_v^2(c_2+1)}{c_2-1}\right)}}{r\left(\sqrt{r\left((1+r)\sigma_T^2 + \frac{\tilde{c}_1\sigma_v^2(c_2+1)}{c_2-1}\right)} + \sqrt{\tilde{c}_1(\left((1+r)\sigma_T^2 + \frac{r\sigma_v^2(c_2+1)}{c_2-1}\right)}\right)}, \tag{2.15}$$

$$\sigma_{P_t}^2 = \frac{r(1+r)\sigma_T^2\sqrt{\tilde{c}_1\left((1+r)\sigma_T^2 + \frac{r\sigma_v^2(c_2+1)}{c_2-1}\right)}}{r\left(\sqrt{r\left((1+r)\sigma_T^2 + \frac{\tilde{c}_1\sigma_v^2(c_2+1)}{c_2-1}\right)} + \sqrt{\tilde{c}_1\left((1+r)\sigma_T^2 + \frac{r\sigma_v^2(c_2+1)}{c_2-1}\right)}\right)}. \tag{2.16}$$

2.2 Performance analysis for superimposed training based schemes

An other alternative to TDMT based schemes, is to send the training and data sequences at the same time. Since data is transmitted all the time, these schemes allow efficient bandwidth utilization, but may suffer from the interference caused by the training sequence. In this line, our work aims to assess the performance of ST based schemes and to compare it with that of TDMT based schemes. We particularly focus on the performance of conventional and data-dependent superimposed training based schemes.

2.2.1 System model for superimposed training based schemes

2.2.1.1 Conventional superimposed training (CST)

In the first CST based schemes, the pilot matrix is added to data without any precoding. The transmitted matrix is then given by:

$$\mathbf{S} = \mathbf{W} + \mathbf{P}$$

where the entries of $\mathbf{W}$ are assumed to be i.i.d. with variance $\sigma_{w_c}^2$ and independent from the training matrix $\mathbf{P}$. The total power per transmitted symbol σ_T^2 is given by:

$$\sigma_T^2 = \sigma_{w_c}^2 + \sigma_{P_c}^2$$

where $\sigma_{P_c}^2$ is the power of each entry of matrix $\mathbf{P}$. (Note that the subscript 'c' stands for 'conventional'.)

2.2.1.2 Data-dependent superimposed training DDST:

In [40], M. Ghogho *et al* propose to introduce a linear distortion to the data matrix at the transmitter so as to ensure the orthogonality between data and training sequences. It is suggested to add a perturbation matrix to the data matrix that is given by:

$$\mathbf{E} = -\mathbf{W}\mathbf{J}$$

where $\mathbf{J} = c_1\mathbf{1}_{\frac{1}{c_1}} \otimes \mathbf{I}_K$, (We assume that $\frac{1}{c_1} = \frac{N}{K}$ is integer valued).

The received signal at each block is therefore given by:

$$\mathbf{Y} = \mathbf{HW}\left(\mathbf{I}_N - \mathbf{J}\right) + \mathbf{HP} + \mathbf{V}$$

Hence, the total transmitted power is split between training and data as follows:

$$\sigma_T^2 = \underbrace{\sigma_{w_d}^2}_{\text{data power}} - \underbrace{c_1\sigma_{w_d}^2}_{\text{distortion power}} + \underbrace{\sigma_{P_d}^2}_{\text{training power}}, \tag{2.17}$$

where $\sigma_{w_d}^2$ and $\sigma_{P_d}^2$ denote the data and training powers, respectively, for the DDST scheme. (Note that subscript 'd' stands for 'data-dependent'.)

In the DDST scheme, The design of the training matrix should meet the following conditions:

$$\begin{cases} \mathbf{P}\mathbf{P}^{\mathrm{H}} = N\sigma_{P_d}^2\mathbf{I}_K \\ \mathbf{P}^{\mathrm{H}}\mathbf{P} = N\sigma_{P_d}^2\mathbf{J}. \end{cases}$$

It is easy to verify that under these conditions, the training matrix $\mathbf{P}$ is orthogonal to the matrix $\mathbf{I}_N - \mathbf{J}$ [that is $(\mathbf{I}_N - \mathbf{J})\mathbf{P}^{\mathrm{H}} = \mathbf{0}$]. Therefore, when multiplying the received signal on the right side by $\mathbf{P}^{\mathrm{H}}$ (this matrix multiplication is considered in the LS channel estimation as shown next), the obtained result is independent from the unknown data. A possible training matrix that fulfills the aforementioned requirements is given in [40] as:

$$\mathbf{P}(k,n) = \sqrt{\sigma_{P_d}^2}\exp\left(\jmath 2\pi kn/K\right).$$

2.2.2 Channel estimation

We derive in this section the MSE values of the LS channel estimates for the conventional and data-dependent ST schemes.

2.2.2.1 Channel estimation for the CST based scheme

The receiver estimates the channel by treating $\mathbf{HW}$ as an additive noise term. Hence the LS channel matrix estimate is given by :

$$\widehat{\mathbf{H}}_c \;=\; \mathbf{Y}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} = \mathbf{H} + \mathbf{HW}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{\mathbf{-1}} + \mathbf{V}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}.$$

Let $\Delta\mathbf{H}_c = \widehat{\mathbf{H}}_c - \mathbf{H}$ denote the channel estimation error matrix. Thus, the MSE is given by:

$$\mathrm{MSE}_c \;=\; \mathrm{Tr}\left(\mathbb{E}\left[\Delta\mathbf{H}_c\Delta\mathbf{H}_c^{\mathrm{H}}\right]\right) = \mathrm{Tr}\left(\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\right)\left(\sigma_{w_c}^2\mathrm{Tr}(\mathbf{H}\mathbf{H}^{\mathrm{H}}) + M\sigma_v^2\right). \tag{2.18}$$

Similarly to the TDMT case, the training matrix which minimizes the MSE subject to a fixed training power must verify:

$$\mathbf{P}\mathbf{P}^{\mathrm{H}} = N\sigma_P^2\mathbf{I}_K.$$

Thus, the expression for the MSE becomes:

$$\mathrm{MSE}_c = \frac{K}{N\sigma_{P_c}^2}\left(\sigma_{w_c}^2\mathrm{Tr}(\mathbf{H}\mathbf{H}^{\mathrm{H}}) + M\sigma_v^2\right).$$

Note that the estimation errors always exist even if the additive noise is not present. This is due to the presence of the unknown data that acts like an extra source of noise during the channel estimation step.

2.2.2.2 Channel estimation for the DDST scheme

The channel LS estimate is obtained by multiplying $\mathbf{Y}$ by $\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}$, thus giving:

$$\widehat{\mathbf{H}}_d \;=\; \mathbf{Y}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} = \mathbf{H} + \mathbf{V}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} = \mathbf{H} + \Delta\mathbf{H}_d,$$

where $\Delta\mathbf{H}_d = \mathbf{V}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}$ denotes the channel estimation error matrix for the DDST scheme. Thus, the MSE has the following expression:

$$\mathrm{MSE}_d \;=\; M\sigma_v^2 \mathrm{Tr}\left(\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\right) = \frac{KM\sigma_v^2}{N\sigma_{P_d}^2}. \tag{2.19}$$

Note that for an equal data and total power allocation, i.e $\sigma_{w_c}^2 = \sigma_{w_d}^2$ and $\sigma_{w_c}^2 + \sigma_{P_c}^2 = \sigma_{w_d}^2 - c_1\sigma_{w_d}^2 + \sigma_{P_d}^2$, the channel estimation error corresponding to the data dependent superimposed training scheme is lower than that of the conventional scheme, i.e $\mathrm{MSE}_d < \mathrm{MSE}_c$.

2.2.3 Data detection

In the following, we determine for both schemes the post-processing noise at the output of the linear zero-forcing detector, and establish theorems about its asymptotic behaviour. The proof of the theorems will need extra assumptions that we shall first mention:

A5. *The entries of* $\mathbf{W}$ *are zero-mean circular i.i.d. complex bounded random variables uncorrelated from the noise* $\mathbf{V}$ *with variance* $\sigma_{w_c}^2$ *for the CST scheme (resp.* $\sigma_{w_d}^2$ *for the DDST scheme).*

A6. *The entries of* $\mathbf{V}$ *are zero-mean circular i.i.d. complex random Gaussian variables with power* σ_v^2.

A7. *The entries of the training matrix* $\mathbf{P}$ *verifies:*

$$\mathbf{P}(k,n) = \sigma_P^2 \exp\left(2\jmath\pi kn/K\right), k \in \{1,\cdots,K\} \quad \text{and} \quad n \in \{1,\cdots,N\}$$

where $\sigma_P^2 = \sigma_{P_c}^2$ *when the CST scheme is considered. (resp.* $\sigma_P^2 = \sigma_{P_d}^2$ *when the DDST scheme is considered.)*

2.2.3.1 Data detection for the CST scheme

For the conventional superimposed training, the zero-forcing estimate of the transmitted data matrix writes as:

$$\widehat{\mathbf{W}}_c \;=\; \left(\widehat{\mathbf{H}}_c\right)^{\#}\left(\mathbf{Y} - \widehat{\mathbf{H}}_c\mathbf{P}\right) = \left(\widehat{\mathbf{H}}_c\right)^{\#}\mathbf{H}\left(\mathbf{W}+\mathbf{P}\right) + \left(\widehat{\mathbf{H}}_c\right)^{\#}\mathbf{V} - \mathbf{P}. \tag{2.20}$$

Assuming that the channel estimation error is small, the pseudo-inverse of the estimated matrix can be approximated by the linear part of the Taylor expansion as:

$$\left(\widehat{\mathbf{H}}_c\right)^{\#} = \mathbf{H}^{\#} - \mathbf{H}^{\#}\left(\Delta\mathbf{H}_c\right)\mathbf{H}^{\#} + \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\Delta\mathbf{H}_c^{\mathrm{H}}\mathbf{\Pi}. \tag{2.21}$$

Substituting (2.21) into (2.20), the zero-forcing estimate of the transmitted matrix can be further expressed as:

$$\widehat{\mathbf{W}}_c = \left(\mathbf{I}_K - \mathbf{H}^{\#}\Delta\mathbf{H}_c\right)\left(\mathbf{W}+\mathbf{P}\right) + \left(\mathbf{I}_K - \mathbf{H}^{\#}\Delta\mathbf{H}_c\right)\mathbf{H}^{\#}\mathbf{V} + \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\mathbf{\Delta H_c^{\mathrm{H}}\Pi V} - \mathbf{P}.$$

Hence, the effective post-processing noise $\Delta\mathbf{W}_c = \widehat{\mathbf{W}}_c - \mathbf{W}$ could be written as:

$$\Delta\mathbf{W}_c \;=\; -\mathbf{H}^{\#}\Delta\mathbf{H}_c\left(\mathbf{W}+\mathbf{P}\right) + \left(\mathbf{I}_K - \mathbf{H}^{\#}\Delta\mathbf{H}_c\right)\mathbf{H}^{\#}\mathbf{V} + \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-\mathbf{1}}\boldsymbol{\Delta}\mathbf{H}_{\mathbf{c}}^{\mathrm{H}}\boldsymbol{\Pi}\mathbf{V},$$

The asymptotic behavior of each entry of the noise matrix is given by the following theorem:

Theorem 2.2. *Under assumptions* **A1**, **A6**, **A5** *and* **A7**, *the post-processing noise experienced by the i-th antenna at each time k,* $[\Delta\mathbf{W}_c]_{i,k}$,*behaves asymptotically as a Gaussian mixture random variable, i.e.,*

$$\mathbb{E}\left[\exp\left(\jmath\Re s^*\left[\Delta\mathbf{W}_c\right]_{i,k}\right)\right] - \sum_{i=1}^{\mathcal{Q}} p_i \exp\left(\jmath\Re\left(s^*(\alpha_i+\theta)\right)\right)\exp\left(-\frac{|s^2|\sigma_{w_c}^2\left(\delta_c\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}+\tau_c\right)}{4}\right) \xrightarrow[K\to\infty]{} 0 \tag{2.22}$$

where :

$$\delta_c = (1-c_1)\frac{\sigma_v^2}{\sigma_{w_c}^2} + \frac{c_1\sigma_v^2}{\sigma_{P_c}^2} + \frac{c_1\sigma_v^4(c_2+1)}{(c_2-1)\sigma_{P_c}^2\sigma_{w_c}^2}$$

$$\tau_c = \frac{c_1\sigma_v^2}{\sigma_P^2(c_2-1)} + \frac{c_1\sigma_{w_c}^2}{\sigma_P^2}$$

$$\theta = c_1(c_2-1)\frac{\sigma_v^2\widetilde{\mathbf{h}}_i\mathbf{p}_j}{\sigma_P^2}$$

and $\mathcal{Q}$ *is the cardinal of the set of all possible values of* $\overline{W}_i = c_1\sum_{l=1}^{\frac{1}{c_1}}[\mathbf{W}]_{i,l}$, *and* p_i *is the probability that* $\overline{W}_i$ *takes the value* α_i.

We can also prove that conditioning on the fact that $[\mathbf{W}]_{i,k} = (\epsilon_1+\jmath\epsilon_2)\sqrt{\frac{\sigma_{w_c}^2}{2}}$ *where* $\epsilon_1 = \pm 1$ *and* $\epsilon_2 = \pm 1$ *the post-processing noise satisfies:*

$$\mathbb{E}\left[\exp\left(\jmath\Re\left(s^*\left[\Delta\mathbf{W}_c\right]_{i,k}\right)\right) | \left[\mathbf{W}_c\right]_{i,k} = (\epsilon_1+\jmath\epsilon_2)\sqrt{\frac{\sigma_{w_c}^2}{2}}\right] - \sum_{i=1}^{\mathcal{Q}'} p_i' \exp\left(\jmath\Re\left(s^*(-c_1(\epsilon_1+\jmath\epsilon_2)\sqrt{\frac{\sigma_{w_c}^2}{2}}+\alpha_i'+\theta)\right)\right)$$
$$\times\exp\left(-\frac{|s^2|\sigma_{w_c}^2\left(\delta_c\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}+\tau_c\right)}{4}\right) \xrightarrow[K\to\infty]{} 0 \tag{2.23}$$

where $\mathcal{Q}'$ *is the cardinal of the set of all possible values of* $\overline{\overline{W}}_i = c_1\sum_{l=1}^{\frac{1}{c_1}-1}[\mathbf{W}]_{i,l}$ *and* p_i' *is the probability that* $\overline{\overline{W}}_i$ *takes the value* α_i'.

Proof. See Appendix A.2. □

Remark 2.1. *The presence of the term* θ *makes the derivation of the BER difficult. By simulations, we have seen that this term, has a little impact and thus in the following, we will assume that the distribution of the post-processing noise behaves satisfies:*

$$\mathbb{E}\left[\exp\left(\jmath\Re\left(s^*\left[\Delta\mathbf{W}_c\right]_{i,k}\right)\right) | \left[\mathbf{W}_c\right]_{i,k} = (\epsilon_1+\jmath\epsilon_2)\sqrt{\frac{\sigma_{w_c}^2}{2}}\right] - \sum_{i=1}^{\mathcal{Q}'} p_i' \exp\left(\jmath\Re\left(s^*(-c_1(\epsilon_1+\jmath\epsilon_2)\sqrt{\frac{\sigma_{w_c}^2}{2}}+\alpha_i')\right)\right)$$
$$\times\exp\left(-\frac{|s^2|\sigma_{w_c}^2\left(\delta_c\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}+\tau_c\right)}{4}\right) \simeq 0 \tag{2.24}$$

2.2.3.2 Data detection for the DDST scheme

For the DDST scheme, we consider a zero-forcing receiver which, prior to inverting the channel matrix, cancels the contribution of the training symbols by right-multiplying $\mathbf{Y}$ by $(\mathbf{I}-\mathbf{J})$. Thus, the zero-forcing estimate is given by:

$$\begin{aligned}
\widehat{\mathbf{W}}_d &= \left(\widehat{\mathbf{H}}_d\right)^{\#} \mathbf{Y}\left(\mathbf{I}-\mathbf{J}\right) \\
&= \left(\mathbf{H}^{\#} - \mathbf{H}^{\#}\Delta\mathbf{H}_d\mathbf{H}^{\#} + \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\Delta\mathbf{H}_d^{\mathrm{H}}\mathbf{\Pi}\right)\mathbf{H}\mathbf{W}\left(\mathbf{I}-\mathbf{J}\right) \\
&+ \left(\mathbf{H}^{\#} - \mathbf{H}^{\#}\Delta\mathbf{H}_d\mathbf{H}^{\#} + \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\Delta\mathbf{H}_d^{\mathrm{H}}\mathbf{\Pi}\right)\mathbf{V}\left(\mathbf{I}-\mathbf{J}\right) \\
&= \left(\mathbf{I} - \mathbf{H}^{\#}\Delta\mathbf{H}_d\right)\mathbf{W}\left(\mathbf{I}-\mathbf{J}\right) + \left(\mathbf{H}^{\#} - \mathbf{H}^{\#}\Delta\mathbf{H}_d\mathbf{H}^{\#}\right)\mathbf{V}\left(\mathbf{I}-\mathbf{J}\right) \\
&+ \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\Delta\mathbf{H}_d^{\mathrm{H}}\mathbf{\Pi}\mathbf{V}\left(\mathbf{I}-\mathbf{J}\right) \\
&= \mathbf{W}\left(\mathbf{I}-\mathbf{J}\right) - \mathbf{H}^{\#}\Delta\mathbf{H}_d\mathbf{W}\left(\mathbf{I}-\mathbf{J}\right) + \left(\mathbf{H}^{\#} - \mathbf{H}^{\#}\Delta\mathbf{H}_d\mathbf{H}^{\#}\right)\mathbf{V}\left(\mathbf{I}-\mathbf{J}\right) + \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\Delta\mathbf{H}_d^{\mathrm{H}}\mathbf{\Pi}\mathbf{V}(\mathbf{I}-\mathbf{J}) \\
&= \mathbf{W} + \left(-\mathbf{W}\mathbf{J} - \mathbf{H}^{\#}\Delta\mathbf{H}_d\mathbf{W}\left(\mathbf{I}-\mathbf{J}\right) + \left(\mathbf{H}^{\#} - \mathbf{H}^{\#}\Delta\mathbf{H}_d\mathbf{H}^{\#}\right)\mathbf{V}\left(\mathbf{I}-\mathbf{J}\right)\right) \qquad (2.25) \\
&+ \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\Delta\mathbf{H}_d^{\mathrm{H}}\mathbf{\Pi}\mathbf{V}\left(\mathbf{I}-\mathbf{J}\right) \qquad (2.26)
\end{aligned}$$

Hence,

$$\Delta\mathbf{W}_d = -\mathbf{W}\mathbf{J} - \mathbf{H}^{\#}\Delta\mathbf{H}_d\mathbf{W}\left(\mathbf{I}-\mathbf{J}\right) + \left(\mathbf{H}^{\#} - \mathbf{H}^{\#}\Delta\mathbf{H}_d\mathbf{H}^{\#}\right)\mathbf{V}\left(\mathbf{I}-\mathbf{J}\right) + \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\Delta\mathbf{H}_d^{\mathrm{H}}\mathbf{\Pi}\mathbf{V}\left(\mathbf{I}-\mathbf{J}\right) \qquad (2.27)$$

The asymptotic distribution of $\left[\Delta\mathbf{W}_d\right]_{i,k}$ is given by the following lemma:

Theorem 2.3. *Under assumptions* **A1**, **A6**, **A5** *and* **A7** *the post-processing noise experienced by the i-th antenna at each time k behaves asymptotically as a Gaussian mixture random variable, i.e,*

$$\mathbb{E}\left[\exp\left(\jmath\Re\left(s^*\left[\Delta\mathbf{W}_d\right]_{i,k}\right)\right)\right] - \sum_{i=1}^{\mathcal{Q}} p_i \exp\left(\jmath\Re\left(s^*\alpha_i\right)\right)\exp\left(-\frac{|s|^2\delta_d\sigma_{w_d}^2\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}}{4}\right) \xrightarrow[K\to\infty]{} 0 \qquad (2.28)$$

where :

$$\delta_d = \left(1-c_1\right)\left(\frac{c_1\sigma_v^2}{\sigma_{P_d}^2} + \frac{\sigma_v^2}{\sigma_{w_d}^2} + \frac{c_1\sigma_v^4(c_2+1)}{(c_2-1)\sigma_{P_d}^2\sigma_{w_d}^2}\right) \qquad (2.29)$$

and $\mathcal{Q}$ is the cardinal of the set of all possible values of $\left[\overline{\mathbf{W}}\right]_{i,k} = c_1\sum_{k=1}^{\frac{1}{c_1}}\left[\mathbf{W}\right]_{i,k}$, and p_i is the probability that $\left[\overline{\mathbf{W}}\right]_{i,k}$ takes the value α_i.

We can also prove that conditioning on the fact that $\left[\mathbf{W}\right]_{i,k} = \epsilon_1\sqrt{\frac{\sigma_{w_d}^2}{2}} + \jmath\epsilon_2\sqrt{\frac{\sigma_{w_d}^2}{2}}$ where $\epsilon_1 = \pm 1$ and $\epsilon_2 = \pm 1$ the post-processing noise satisfies:

$$\mathbb{E}\left[\exp\left(\jmath\Re\left(s^*\left[\Delta\mathbf{W}_d\right]_{i,k}\right)\right) | \left[\mathbf{W}\right]_{i,k} = (\epsilon_1+\jmath\epsilon_2)\sqrt{\frac{\sigma_w^2}{2}}\right] - \sum_{i=1}^{\mathcal{Q}'} p_i' \exp\left(\jmath\Re\left(s^*\left(-c_1\left(\epsilon_1+\jmath\epsilon_2\right)\sqrt{\frac{\sigma_{w_d}^2}{2}} + \alpha_i'\right)\right)\right) \qquad (2.30)$$

$$\times \exp\left(-\frac{|s|^2\delta_d\sigma_{w_d}^2\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}}{4}\right) \xrightarrow[K\to\infty]{} 0 \qquad (2.31)$$

where $\mathcal{Q}^{'}$ *is the cardinal of the set of all possible values* $\overline{\overline{W}}_i = c_1 \sum_{l=1}^{\frac{1}{c_1}-1} [\mathbf{W}]_{i,l}$, *and* $p_i^{'}$ *is the probability that* $\overline{\overline{W}}_i$ *takes the value* $\alpha_i^{'}$.

Proof. See Appendix A.3. □

2.2.4 Bit error rate performance

The assumption of the gaussianity of the post processing noise has been always assumed. For time division multiplexed training, this assumption is well-founded, since the post-processing noise, converges to a Gaussian distribution in the asymptotic regime, (See theorem 2.1). Moreover, for conventional cases, which do not take into account the imperfect channel estimation, the asymptotic gaussianity of the post-processing noise holds in case of additive Gaussian noise.

In the superimposed training case, the distortion caused by the presence of data symbols affects the distribution of the post-processing noise which becomes asymptotically Gaussian mixture distributed. To assess the system performance in this particular case, we will start from the elementary definition of the bit error rate. Let $\Delta\mathbf{W}_{i,k}$ denotes the post processing noise experienced at the i-th antenna at time k ($\Delta\mathbf{W}_{i,k} = [\Delta\mathbf{W}_c]_{i,k}$ for the CST scheme and $\Delta\mathbf{W}_{i,k} = [\Delta\mathbf{W}_d]_{i,k}$ for the DDST scheme).

As it has been previously shown, $\Delta\mathbf{W}_{i,k}$ behaves as a Gaussian mixture random variable. Let σ_∞^2 be the asymptotic variance of $\Delta\mathbf{W}_{i,k}$, i.e, $\sigma_\infty^2 = \sigma_{w_d}^2 \delta_d \left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}$ for the DDST scheme and $\sigma_\infty^2 = \sigma_{w_c}^2 \left(\tau_c + \delta_c \left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}\right)$ for the CST scheme.

Using the symmetry of the transmitted data, the BER expression at the ith branch, under QPSK constellation and for a given channel realization is given by:

$$\begin{aligned}\mathrm{BER}_i &= \frac{1}{2}\mathbb{P}\left[\Re\left(\widehat{\mathbf{W}}_{i,k}\right) > 0 | \Re\left(\mathbf{W}_{i,k}\right) = -\sqrt{\frac{\sigma_w^2}{2}}\right] + \frac{1}{2}\mathbb{P}\left[\Re\left(\widehat{\mathbf{W}}_{i,k}\right) < 0 | \Re\left(\mathbf{W}_{i,k}\right) = \sqrt{\frac{\sigma_w^2}{2}}\right] \\ &= \frac{1}{2}\mathbb{P}\left[\Re\left(\Delta\mathbf{W}_{i,k}\right) > \sqrt{\frac{\sigma_w^2}{2}} | \Re\left(\mathbf{W}_{i,k}\right) = -\sqrt{\frac{\sigma_w^2}{2}}\right] + \frac{1}{2}\mathbb{P}\left[\Re\left(\Delta\mathbf{W}_{i,k}\right) < -\sqrt{\frac{\sigma_w^2}{2}} | \Re\left(\mathbf{W}_{i,k}\right) = \sqrt{\frac{\sigma_w^2}{2}}\right]\end{aligned}$$

In the asymptotic regime, $\left[\Re\left(\Delta\mathbf{W}_{i,k}\right) | \Re\left[\mathbf{W}\right]_{i,k} = \epsilon\sqrt{\frac{\sigma_w^2}{2}}\right]$ converges to a mixed Gaussian distribution with the probability density function:

$$f(x) = \frac{1}{\sqrt{\pi\sigma_\infty^2}} \sum_{s=1}^{\sqrt{\mathcal{Q}'}} p_s \exp\left(-\frac{(x + c_1\epsilon\sqrt{\frac{\sigma_w^2}{2}} - \Re(\alpha_s))^2}{\sigma_\infty^2}\right)$$

Hence, conditioned on the channel, the asymptotic bit error rate can be approximated by:

$$\begin{aligned}\mathrm{BER}_{i,\infty} = &\frac{1}{2}\frac{1}{\sqrt{\pi\sigma_\infty^2}} \int_{\sqrt{\frac{\sigma_w^2}{2}}}^{+\infty} \sum_{s=1}^{\sqrt{\mathcal{Q}'}} p_s^{'} \exp\left(-\frac{(x - c_1\sqrt{\frac{\sigma_w^2}{2}} - \Re(\alpha_s))^2}{\sigma_\infty^2}\right) dx \\ &+ \frac{1}{2}\frac{1}{\sqrt{\pi\sigma_\infty^2}} \int_{-\infty}^{-\sqrt{\frac{\sigma_w^2}{2}}} \sum_{s=1}^{\sqrt{\mathcal{Q}'}} p_s^{'} \exp\left(-\frac{(x + c_1\sqrt{\frac{\sigma_w^2}{2}} - \Re(\alpha_s))^2}{\sigma_\infty^2}\right) dx\end{aligned}$$

Finally, the proposed approximation of the BER can be obtained by averaging with respect to the channel realization $\mathbf{H}$, thus giving:

$$\mathrm{BER}_\infty = \mathbb{E}\frac{1}{2}\sum_{s=1}^{\sqrt{\mathcal{Q}'}} p_s' Q\left(\sqrt{\frac{\sigma_w^2}{\sigma_\infty^2}}(1-c_1) - \frac{\Re(\alpha_s)}{\sqrt{\frac{\sigma_\infty^2}{2}}}\right) + \frac{1}{2}\sum_{s=1}^{\sqrt{\mathcal{Q}'}} p_s' Q\left(\sqrt{\frac{\sigma_w^2}{\sigma_\infty^2}}(1-c_1) + \frac{\Re(\alpha_s)}{\sqrt{\frac{\sigma_\infty^2}{2}}}\right)$$

For QPSK constellations, it can be shown that $\sqrt{\mathcal{Q}'} = \frac{1}{c_1}$, where $\frac{1}{c_1} = \frac{N}{K}$ is assumed to be integer. Moreover, the set $\mathcal{S}$ of the values taken by $\Re(\alpha_s)$ can be given by:

$$\mathcal{S} = \left\{ \Re(\alpha_s) = c_1\sqrt{\frac{\sigma_w^2}{2}}(\frac{1}{c_1} - 2s - 1), s \in \left\{0, \cdots, \frac{1}{c_1} - 1\right\}\right\}.$$

with probability $p_s = \frac{\binom{\frac{1}{c_1}-1}{s}}{2^{\frac{1}{c_1}-1}}$.

Let $\gamma = \frac{\sigma_w^2}{\sigma_\infty}$ ($\gamma = \frac{1}{\tau_c + \delta_c\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}}$ for the CST scheme and $\gamma = \frac{1}{\delta_d(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}}$ for the DDST scheme) then, the BER expression becomes:

$$\mathrm{BER}_\infty = \mathbb{E}\sum_{s=0}^{\frac{1}{c_1}-1} \frac{\binom{\frac{1}{c_1-1}}{s}}{2^{\frac{1}{c_1}-1}} Q(2sc_1\sqrt{\gamma}) \tag{2.32}$$

where the expectation is taken over the distribution of γ. Expression (2.32) will be used to evaluate and compare the performance in terms of BER of the superimposed training based schemes. For the CST scheme, giving a closed form expression for (2.32) is not easy, we will provide instead a numerical estimation based on Taylor expansion of the error function.

2.2.4.1 BER for the DDST scheme

For the DDST scheme, the expectation in (2.32) is taken over the distribution of $\gamma = \gamma_d = \frac{1}{\delta_d\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}}$ given by:

$$f_{\gamma_d}(x) = \frac{(K\delta_d)^{M-K+1}x^{M-K}}{(M-K)!}\exp\left(-K\delta_d x\right)$$

The computation of the Bit error rate can be treated similarly to the TDMT scheme, thus leading to:

$$\mathrm{BER}_\infty = \frac{1}{2^{\frac{1}{c_1}-1}}\sum_{s=0}^{\frac{1}{c_1}-1}\binom{\frac{1}{c_1}-1}{s} J(M-K+1, K\delta_d, 4s^2c_1^2)$$

Using (2.10), it can be possible to derive the following closed-form expression:

$$\mathrm{BER}_\infty = \frac{1}{2^{\frac{1}{c_1}}}\sum_{s=0}^{\frac{1}{c_1}-1}\binom{\frac{1}{c_1}-1}{s}\left[1 - \mu_s\sum_{k=0}^{M-K}\binom{2k}{k}\left(\frac{1-\mu_s^2}{4}\right)^k\right] \tag{2.33}$$

where $\mu_s = \sqrt{\frac{2s^2c_1^2}{K\delta_d + 2s^2c_1^2}}$.

2.2.4.2 BER for the CST scheme

Unlike the DDST scheme, the CST scheme does not seem to have a straightforward closed-form expression for the BER. Indeed, as compared to the DDST and TDMT based schemes, the variable over which the expectation is taken is not proportional to $\frac{1}{\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}}$.

For the CST scheme, the expectation in (2.32) is taken over the distribution of $\gamma_c = \frac{1}{\tau_c+\delta_c\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}}$, given by:

$$f_{\gamma_c}(x) = \frac{(K\delta_c)^{M-K+1}x^{M-K}}{(M-K)!(1-\tau_c x)^{M-K+2}} \exp\left(-\frac{K\delta_c x}{1-\tau_c x}\right) \mathbf{1}_{\left\{0\leq x\leq \frac{1}{\tau_c}\right\}} \tag{2.34}$$

Whenever straightforward closed-form expressions for the BER do not exist, numerical methods have often been a good alternative to approximate the BER. A class of these methods is based on approximating the error function $Q(x)$ by a function series, [6, 78, 36]. In our case, we opt for using a Taylor expansion for $Q(x)$ given by:

$$Q(x) = Q(x_0) + \sum_{n=1}^{\infty} \frac{d^n Q}{dx^n}_{|x_0} \frac{(x-x_0)^n}{n!} \tag{2.35}$$

where x_0 is the point around which the Taylor expansion is made. Our choice is accounted for by the fact that, in the considered case, all moments of $\sqrt{\gamma_c}$ can be computed in closed-forms. Actually, let r be a positive real, then the rth moment of γ_c is given by [49, page 348]:

$$\mathbb{E}\left[\gamma_c^r\right] = \frac{\left(\frac{K\delta_c}{\tau_c}\right)^{\frac{M-K}{2}}}{(M-K)!\tau_c^r} \exp\left(\frac{K\delta_c}{2\tau_c}\right) \Gamma(M-K+r+1) W_{\frac{-M+K-2r}{2},\frac{-1-M+K}{2}}\left(\frac{K\delta_c}{\tau_c}\right) \tag{2.36}$$

where $W_{\lambda,\mu}(z)$ denotes the Whittaker function defined in [49]. To approximate the bit error rate, we will substitute $Q(x)$ in (2.32) by its Taylor expansion given by (2.35), and permute the infinite sum and the expectation operator, thus yielding:

$$\mathrm{BER}_\infty = \sum_{s=0}^{\frac{1}{c_1}-1} \frac{\binom{\frac{1}{c_1}-1}{s}}{2^{\frac{1}{c_1}-1}} \left[Q(2sc_1\mathbb{E}\sqrt{\gamma_c}) + \sum_{n=1}^{\infty} \frac{d^n Q}{dx^n}_{|2sc_1\mathbb{E}\sqrt{\gamma_c}} \frac{(2sc_1)^n \mathbb{E}(\sqrt{\gamma_c} - \mathbb{E}\sqrt{\gamma_c})^n}{n!} \right] \tag{2.37}$$

The next step is remove the infinite sum in (2.37) by truncating at a given order J depending on the desired accuracy, thus leading to:

$$\mathrm{BER}_\infty \simeq \sum_{s=0}^{\frac{1}{c_1}} \frac{\binom{\frac{1}{c_1}}{k}}{2^{\frac{1}{c_1}}} \left[Q(2sc_1\mathbb{E}\sqrt{\gamma_c}) + \sum_{n=1}^{J} \frac{d^n Q}{dx^n}_{|2sc_1\mathbb{E}\sqrt{\gamma_c}} \frac{(2sc_1)^n \mathbb{E}(\sqrt{\gamma_c} - \mathbb{E}\sqrt{\gamma_c})^n}{n!} \right] \tag{2.38}$$

All the deterministic quantities in (2.38) can be numerically computed. Hereafter, we show how this can be done:

- The nth derivative of $Q(x)$ can be expressed as:

$$\frac{d^n Q}{dx^n} = \frac{(-1)^n \exp(-x^2/2)}{\sqrt{2\pi}} H_{n-1}(x)$$

  where $H_n(x)$ denotes the Hermite Polynomial with order n given by:

$$H_n(x) = (-1)^n \exp(x^2/2) \frac{d^n}{dx^n} \exp(-x^2/2)$$

- The central moments of $\sqrt{\gamma_c}$ can be evaluated by using the non-central moments of γ. In fact, the expansion of the expression of the central moment of $\sqrt{\gamma_c}$ is given by :

$$\mathbb{E}\left[(\sqrt{\gamma_c}-\mathbb{E}\sqrt{\gamma_c})^n\right]=\sum_{m=0}^{n}\binom{n}{m}\mathbb{E}\gamma_c^{\frac{m}{2}}(-1)^{n-m}\left(\mathbb{E}\sqrt{\gamma_c}\right)^{n-m} \tag{2.39}$$

  and thus can be computed using (2.36)

2.2.4.3 BER Performance comparison

Though a closed-form expression for the BER does not exist for the CST scheme, we can still make performance comparison between both schemes, based on the distribution of the post-processing noise.

Actually, one can observe that for equal data power allocation, σ_w^2 satisfying :

$$\sigma_w^2+\sigma_{P_c}^2=\sigma_w^2(1-c_1)+\sigma_{P_d}^2=\sigma_T^2,$$

we have:

$$\delta_c(\sigma_w^2)\geq\delta_d(\sigma_w^2)$$

and hence:

$$\frac{1}{\tau_c+\delta_c\left[(\mathbf{H}^{\text{H}}\mathbf{H})^{-1}\right]_{i,i}}\leq\frac{1}{\delta_d\left[(\mathbf{H}^{\text{H}}\mathbf{H})^{-1}\right]_{i,i}}\leq\max_{0\leq\sigma_w^2\leq\sigma_T^2}\frac{1}{\delta_d\left[(\mathbf{H}^{\text{H}}\mathbf{H})^{-1}\right]_{i,i}}$$

This result proves that, for any data power allocation, the DDST scheme outperforms the conventional one in terms of BER. As a consequence, for an equal total power allocation, the DDST scheme with optimal power allocation outperforms the conventional one in terms of BER.

2.2.5 Outage probability performance

At the output of the zero-forcing receiver, we define the SNR as the ratio between the power signal σ_w^2 and the power of the post-processing noise. Using straightforward but tedious calculations, it is easy to show that the entries of the post-processing noise matrix for the CST and DDST schemes satisfy:

$$\mathbb{E}\left[|\left[\Delta\mathbf{W}_c\right]_{i,k}|^2\right]-\sigma_w^2\left(c_1+\tau_c+\delta_c\left(\mathbf{H}^{\text{H}}\mathbf{H}\right)^{-1}\right)\simeq 0 \tag{2.40}$$

$$\mathbb{E}\left[|\left[\Delta\mathbf{W}_d\right]_{i,k}|^2\right]-\sigma_w^2\left(c_1+\delta_d\left(\mathbf{H}^{\text{H}}\mathbf{H}\right)^{-1}\right)\xrightarrow[K\to+\infty]{}\simeq 0 \tag{2.41}$$

Let $\tilde{\gamma}_c$ and $\tilde{\gamma}_d$ denote the post-processing SNR for the CST and DDST schemes. Then, using (2.40) and (2.41), we can easily check that:

$$\widetilde{\gamma}_c-\frac{1}{c_1+\tau_c+\delta_c\left(\mathbf{H}^{\text{H}}\mathbf{H}\right)^{-1}}\simeq 0 \text{ almost surely}$$

$$\widetilde{\gamma}_d-\frac{1}{c_1+\delta_d\left(\mathbf{H}^{\text{H}}\mathbf{H}\right)^{-1}}\simeq 0 \text{ almost surely}$$

The pdf of the asymptotic equivalents of $\widetilde{\gamma}_c$ and $\widetilde{\gamma}_d$ has the same shape as that of γ_c given by (2.34). The approximation for the outage probability can be then obtained by integrating (2.34), thus giving:

$$\mathbb{P}\left[\widetilde{\gamma} \leq x\right] \simeq \begin{cases} \dfrac{1}{(M-K)!}\gamma(M-K+1, \dfrac{K\delta x}{1-\widetilde{\tau}x}) \quad \text{if} \quad 0 \leq x \leq \dfrac{1}{\widetilde{\tau}} & (2.42) \\ 1 \quad \text{if} \quad x \geq \dfrac{1}{\widetilde{\tau}} & (2.43) \end{cases}$$

where $\delta = \delta_c$, $\widetilde{\tau} = c_1 + \tau_c$ (resp. $\delta = \delta_d$, $\widetilde{\tau} = c_1$) refer to the CST scheme (resp. to the DDST scheme).

2.2.6 Optimal power allocation

So far, we have provided approximations of the BER and outage probabilities for CST and DDST schemes. As it has been previously shown, these approximations, depend on the power allocated to data and training, in addition to other parameters. While the system has no control over the noise power or the number of transmitting and receiving antennas, it still can optimize the power allocation in such a way to minimize the performance indexes. Next, we provide, for CST and DDST schemes, the optimal data and training power amounts that minimize the BER and outage probabilities, under the constraint of a constant total power.

2.2.6.1 Optimum allocation for minimizing the BER

2.2.6.1.1 Optimization of the power allocation for the DDST scheme For the DDST scheme, we can deduce from (2.32) that maximizing γ_d leads to minimize the BER. To maximize γ_d, we need to optimize δ_d as a function of σ_w^2 and under the constraint that $\sigma_{P_d}^2 + (1-c_1)\sigma_{w_d}^2 = \sigma_T^2$. After straightforward calculations, we can find that the optimal values for $\sigma_{w_d}^2$ and $\sigma_{P_d}^2$ are given by:

$$\sigma_{w_d}^2 = \frac{\sqrt{(1-c_1)\left(\sigma_T^2 + \frac{c_1(c_2+1)\sigma_v^2}{c_2-1}\right)}\sigma_T^2}{(1-c_1)\left(\sqrt{(1-c_1)\left(\sigma_T^2 + \frac{c_1(c_2+1)\sigma_v^2}{c_2-1}\right)} + \sqrt{c_1\sigma_T^2 + \frac{c_1(c_2+1)(1-c_1)\sigma_v^2}{c_2-1}}\right)}, \tag{2.44}$$

$$\sigma_{P_d}^2 = \frac{\sqrt{c_1\sigma_T^2 + \frac{c_1(c_2+1)(1-c_1)\sigma_v^2}{c_2-1}}\sigma_T^2}{\sqrt{(1-c_1)\left(\sigma_T^2 + \frac{c_1(c_2+1)\sigma_v^2}{c_2-1}\right)} + \sqrt{c_1\sigma_T^2 + \frac{c_1(c_2+1)(1-c_1)\sigma_v^2}{c_2-1}}}. \tag{2.45}$$

2.2.6.1.2 Optimization of the power allocation for the CST scheme In the CST scheme, the values of $\sigma_{w_c}^2$ and $\sigma_{P_c}^2$ that minimize δ_c may not be optimal in the sense that they do not necessarily minimize the BER, given that τ_c depends on the data power $\sigma_{w_c}^2$. However, since in practice the data block size N is much higher than the number of transmitting antennae, the effect of τ_c could be neglected and as such a suboptimal power allocation would be the one that minimizes δ_c. The corresponding values of $\sigma_{w,c}^2$ and $\sigma_{P,c}^2$ are given by:

$$\sigma_{w,c}^2 = \frac{\sqrt{(1-c_1)\,\sigma_T^2 + \frac{c_1(c_2+1)\sigma_v^2}{c_2-1}}\sigma_T^2}{\sqrt{c_1\sigma_T^2 + \frac{c_1(c_2+1)\sigma_v^2}{c_2-1}} + \sqrt{(1-c_1)\,\sigma_T^2 + \frac{c_1(c_2+1)\sigma_v^2}{c_2-1}}}, \tag{2.46}$$

$$\sigma_{P,c}^2 = \frac{\sqrt{c_1\sigma_T^2 + \frac{c_1(c_2+1)\sigma_v^2}{c_2-1}}\sigma_T^2}{\sqrt{(1-c_1)\,\sigma_T^2 + \frac{c_1(c_2+1)\sigma_v^2}{c_2-1}} + \sqrt{c_1\sigma_T^2 + \frac{c_1(c_2+1)\sigma_v^2}{c_2-1}}} \tag{2.47}$$

2.3 Simulation results

Despite being valid only for the asymptotic regime, our results are found to yield a good accuracy even for very small system dimensions. In this section, we present simulation results that compares between all the aforementioned described training based schemes. More particularly, we carry out simulations that compare the DDST scheme with its TDMT and CST counterparts.

2.3.1 Performance comparison between DDST and TDMT based schemes

In this section, except when mentioning, we consider a 2×4 MIMO system ($K = 2$, $M = 4$) with a data block size $N = 32$.

2.3.1.1 Bit error rate performance

Fig. 2.1 plots the empirical and theoretical BER under QPSK constellation for $N = 32$, $K = 2$ and $M = 4$ for the TDMT and DDST based schemes. All comparisons are conducted in the context when both schemes have the same total energy. The number of training symbols is set to $N_1 = 2$ ($N_2 = 30$) for the TDMT scheme. For low SNR values (SNR below 6 dB), both

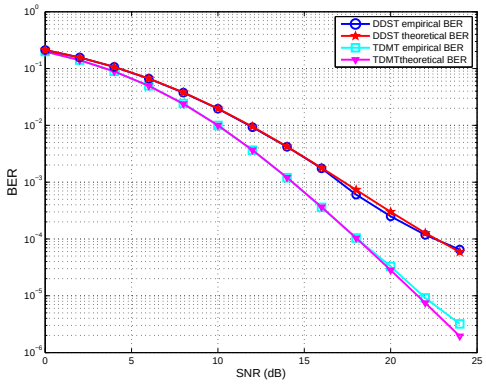


Figure 2.1: Theoretical and empirical BER for the TDMT and DDST based schemes.

schemes achieve approximatively the same BER performance, and therefore, the DDST scheme outperforms its TDMT counterpart in terms of data rate, since it has a better bandwidth efficiency. For high SNR values, the noise caused by the data distortion is higher than the additive Gaussian noise, thus affecting the performance of the DDST scheme.

2.3.1.2 Applications

To compare the efficiency of the TDMT and DDST schemes, we consider applications in which the BER should be below a certain threshold, say 10^{-2}. This may be the case for instance of circuit-switched voice applications. Note that for non-coded systems, a target BER of 10^{-2} is commonly used.

2.3.1.2.1 Application 1 In this scenario, we set the SNR $\triangleq \frac{\sigma_T^2}{\sigma_v^2}$ to 15 dB. We then vary the ratio $c_1 = \frac{K}{N}$ from 0.01 to 0.5. Since we consider $K = 2$ and $M = 4$, $N = K/c_1$ varies also with c_1. For each value of c_1 we compute the BER by using (2.11) and (2.33). Fig. 2.2 illustrates the obtained results. We note that the DDST scheme may be interesting for low values of c_1

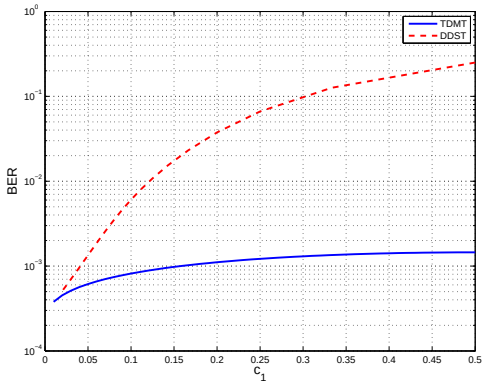


Figure 2.2: BER with respect to c_1 when $K = 2$, $M = 4$ and SNR=15 dB

(say below 0.125), i.e., for long enough frames. For small frames (high distortion ratio c_1), the distortion of the data becomes too high thus reducing the interest of the DDST scheme.

2.3.1.2.2 Application 2 In this experiment, we also consider a scenario where the BER should be below 10^{-2}. Using (2.15), (2.16) and (2.11), we determine the minimum number of required training symbols to meet the BER lower bound requirement. We then, plot the corresponding ratio $r = \frac{N_2}{N_1}$ with respect to the SNR. We note that if the SNR is below 2 dB, the BER requirement could not be achieved. This is to be compared with the DDST scheme where the SNR should be set at least to 10.5 dB so as to meet the BER lower bound requirement as it can be shown in fig. 2.3. Moreover, for a SNR more than 8.5 dB, the minimum number of pilot symbols for channel identification (equal to K) is sufficient to meet the BER requirement.

2.3.1.3 Outage probability performance

In this experiment, we set the SNR to 10dB. Fig. 2.4 shows the outage probability for respectively short ($N = 32$) and long frame ($N = 128$). In the legend 'TDMT' and 'DDST' refer to the outage probability values obtained from (2.14) for the TDMT scheme and (2.42) and (2.43) for the DDST scheme, whereas 'TDMT empirical' and 'DDST empirical' are the empirical cumulative distribution functions of γ_t and γ_d. We note that for high distortion ($N = 32$), the gain of the TDMT scheme over the DDST scheme is much larger that the observed one for small distortion ($N = 128$).

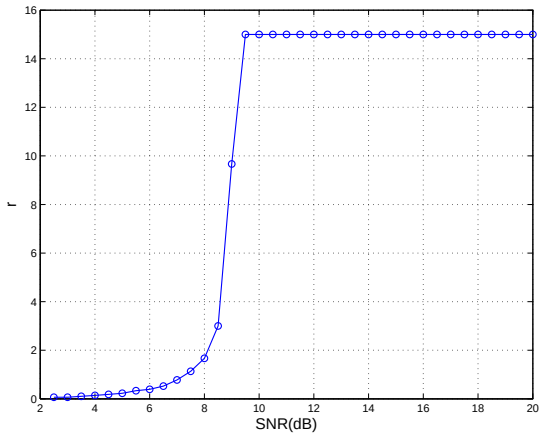
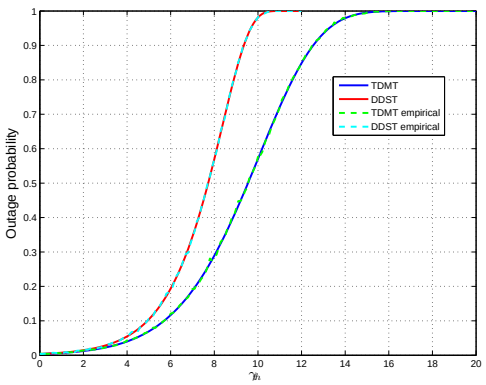
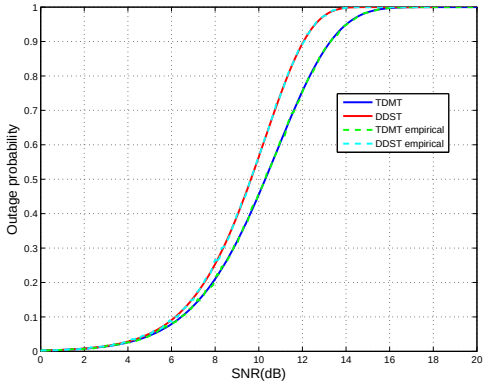


Figure 2.3: Required r versus SNR for BER $\leq 10^{-2}$.



(a) Short frame ($N = 32$)



(b) Long frame ($N = 128$)

Figure 2.4: Outage probability for TDMT and DDST based schemes for short and long frames

2.3.2 Performance comparison between DDST and CST based schemes

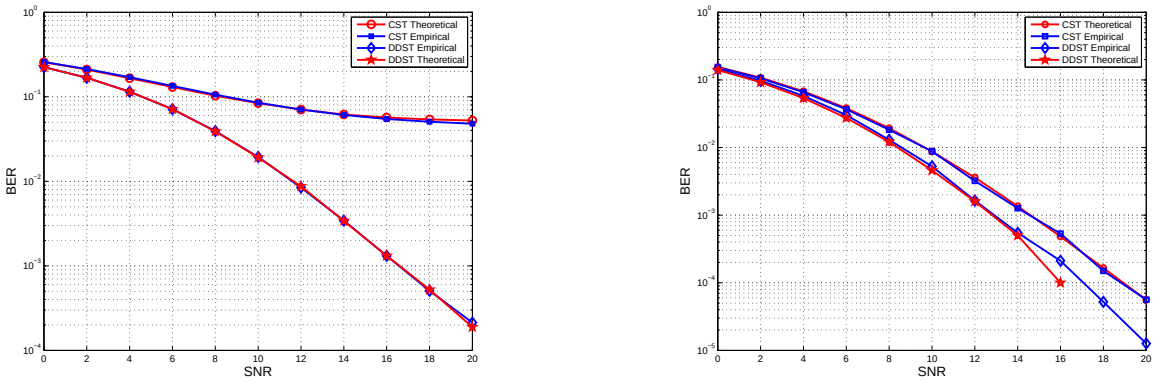
The objective of this section is to validate the theoretical performance analysis for the CST scheme and to compare it with that of the DDST scheme. In all our simulations, we set $\sigma_T^2 = 1$ and consider two simulation contexts corresponding respectively to the cases for long and short transmitted frames:

- In the first context, we consider a MIMO system using $K = 2$ transmit antennas and $M = 4$ receive antennas. The data block size N is set to $N = 1024$.
- In the second context, we consider a MIMO system using $K = 4$ transmit antennas and $M = 8$ received antennas. The data block size is set to $N = 64$.

Except when mentioned, the power allocation for the CST and DDST schemes are set according to (2.44),(2.45) and (2.46),(2.47).

2.3.2.1 Bit error rate performance

Fig. 2.5 displays the empirical and theoretical BER for the CST and DDST schemes for both contexts. We note that the gain in performance achieved by the DDST scheme is more significant in the case of short frames than that in the case of long frames.



(a) Short frame ($N = 64$)

(b) Long frame ($N = 1024$)

Figure 2.5: BER for the CST and DDST schemes for short and long frames

2.4 Conclusion

In this chapter, we have carried out theoretical studies on BER and outage probability for three training based schemes, namely, the basic time division multiplexed training scheme, and two superimposed training based schemes. To make derivations possible, the asymptotic regime where all the system dimensions grow to infinity with a constant pace has been considered. For each scheme, we have derived closed-form approximations for the BER and outage probability. We have also determined optimal power allocations of power between data and training in order to minimize the BER or the outage probability.

Appendices

Appendix A

Proof of results in chapter 2

This appendix provides the proofs of the derived results in chapter 2.

A.1 Proof of theorem 2.1

In the sequel, we propose to determine the asymptotic distribution of the post-processing noise of each entry of the matrix $\Delta\mathbf{S}$. Actually the (i,j) entry of $\Delta\mathbf{S}$ is given by:

$$(\Delta\mathbf{S})_{i,j} = -\mathbf{h}_i^{\#}\Delta\mathbf{H}_t\mathbf{s}_j + \mathbf{h}_i^{\#}\left(\mathbf{I}_K - \Delta\mathbf{H}_t\mathbf{H}^{\#}\right)\mathbf{v}_{2,j} + \widetilde{\mathbf{h}}_i(\Delta\mathbf{H}_t)^{\mathrm{H}}\mathbf{\Pi}\mathbf{v}_{2,j}$$

where $\mathbf{h}_i^{\#}$ and $\widetilde{\mathbf{h}}_i$ denote respectively the ith row of $\mathbf{H}^{\#}$ and $(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}$, and $\mathbf{s}_j$ and $\mathbf{v}_{2,j}$ denote jth columns of $\mathbf{S}$ and $\mathbf{V}_2$, respectively. Conditioned on $\mathbf{H}$, $\mathbf{V}_1$ and $\mathbf{S}$, $(\Delta\mathbf{S})_{i,j}$ is a Gaussian random variable with mean equal to $-\mathbf{h}_i^{\#}\Delta\mathbf{H}_t\mathbf{s}_j$ and variance

$$\sigma^2_{s,K} = \sigma_v^2\left(\mathbf{h}_i^{\#} - \mathbf{h}_i^{\#}\Delta\mathbf{H}_t\mathbf{H}^{\#} + \widetilde{\mathbf{h}}_i(\Delta\mathbf{H}_t)^{\mathrm{H}}\mathbf{\Pi}\right)\left(\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}} - \left(\mathbf{H}^{\#}\right)^{\mathrm{H}}\Delta\mathbf{H}_t^{\mathrm{H}}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}} + \mathbf{\Pi}\Delta\mathbf{H}_t\left(\widetilde{\mathbf{h}}_i\right)^{\mathrm{H}}\right)$$

Since our proof will be based on the 'characteristic function' approach, we shall first recall the expression of the characteristic function for complex random variables:

Theorem A.1. *Let X_n be a complex Gaussian random variable with mean $m_{X,n}$ and variance $\sigma^2_{X,n}$, such that $\mathbb{E}(X_n - m_{X,n})^2 = 0$. Then, X_n can be seen as a two-dimensional random variable corresponding to its real and imaginary parts. The characteristic function of X_n is therefore given by:*

$$\mathbb{E}\left[\exp\left(\jmath\Re(z^*X_n)\right)\right] = \exp\left(\jmath\Re\left(z^*m_{X,n}\right)\right)\exp\left(-\frac{1}{4}|z|^2\sigma^2_{X,n}\right).$$

Applying TheoremA.1, the conditional characteristic function of $(\Delta\mathbf{S})_{i,j}$ can be written as:

$$\mathbb{E}\left[\exp\left(\jmath\Re\left(z^*\left(\Delta\mathbf{S}\right)_{i,j}\right)\right)|\mathbf{V}_1,\mathbf{H},\mathbf{S}\right] = \exp\left(-\jmath\Re\left(z^*\mathbf{h}_i^{\#}\Delta\mathbf{H}_t\mathbf{s}_j\right)\right)\exp\left(-\frac{1}{4}|z|^2\sigma^2_{s,K}\right). \quad \text{(A.1)}$$

To remove the condition expectation on $\mathbf{V}_1$ and $\mathbf{S}$, one should prove that $\sigma^2_{s,K}$ converges almost surely to a deterministic quantity. Actually, $\sigma^2_{s,K}$ can be expanded as follows:

$$\sigma^2_{s,K} = \sigma_v^2\mathbf{h}_i^{\#}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}} + \sigma_v^2\mathbf{h}_i^{\#}\Delta\mathbf{H}_t\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\left(\Delta\mathbf{H}_t\right)^{\mathrm{H}}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}} - 2\sigma_v^2\Re\left(\mathbf{h}_i^{\#}\Delta\mathbf{H}_t\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\right)$$
$$+ \sigma_v^2\widetilde{\mathbf{h}}_i\Delta\mathbf{H}_t^{\mathrm{H}}\mathbf{\Pi}\Delta\mathbf{H}_t\left(\widetilde{\mathbf{h}}_i\right)^{\mathrm{H}}.$$

Let

$$A_{\sigma,K} = \sigma_v^2 \mathbf{h}_i^{\#} \Delta\mathbf{H}_t \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1} \left(\Delta\mathbf{H}_t\right)^{\mathrm{H}} \left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}$$
$$B_{\sigma,K} = \sigma_v^2 \widetilde{\mathbf{h}}_i \Delta\mathbf{H}_t^{\mathrm{H}} \mathbf{\Pi} \Delta\mathbf{H}_t \left(\widetilde{\mathbf{h}}_i\right)^{\mathrm{H}}$$
$$\epsilon_{\sigma,K} = \mathbf{h}_i^{\#} \Delta\mathbf{H}_t \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1} \left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}.$$

The limiting behaviour of $A_{\sigma,K}$ can be derived by using the following known results describing the asymptotic behaviour of an important class of quadratic forms:

Lemma A.1. *[3, Lemma 2.7] Let $\mathbf{x} = [X_1, \cdots, X_N]^{\mathrm{T}}$ be a $N \times 1$ vector where the X_n are centered i.i.d. complex random variables with unit variance. Let $\mathbf{A}$ be a deterministic $N \times N$ complex matrix. Then, for any $p \geq 2$ there exists a constant C_p depending on p only such that:*

$$\mathbb{E}\left|\frac{1}{N}\mathbf{x}^{\mathrm{H}}\mathbf{A}\mathbf{x} - \frac{1}{N}\mathrm{Tr}(\mathbf{A})\right|^p \leq \frac{C_p}{N^p}\left(\left(\mathbb{E}|X_1|^4 \mathrm{Tr}\left(\mathbf{A}\mathbf{A}^{\mathrm{H}}\right)\right)^{p/2} + \mathbb{E}|X_1|^{2p}\mathrm{Tr}\left(\left(\mathbf{A}\mathbf{A}^{\mathrm{H}}\right)^{p/2}\right)\right) \tag{A.2}$$

Noticing that $\mathrm{Tr}\left(\mathbf{A}\mathbf{A}^{\mathrm{H}}\right) \leq N\|\mathbf{A}\|^2$ and that $\mathrm{Tr}\left(\left(\mathbf{A}\mathbf{A}^{\mathrm{H}}\right)^{p/2}\right) \leq N\|\mathbf{A}\|^p$, we obtain the simpler inequality:

$$\mathbb{E}\left|\frac{1}{N}\mathbf{x}^{\mathrm{H}}\mathbf{A}\mathbf{x} - \frac{1}{N}\mathrm{Tr}(\mathbf{A})\right|^p \leq \frac{C_p}{N^{p/2}}\|\mathbf{A}\|^p\left(\left(\mathbb{E}|X_1|^2\right)^{p/2} + \mathbb{E}|X_1|^{2p}\right) \tag{A.3}$$

Hence, if $\mathbf{A}$ and $\mathbf{x}$ have respectively finite spectral norm and finite eigth moment, we can conclude, using Borel-Cantelli lemma, about the almost convergence of the quadratic form $\frac{1}{N}x^{\mathrm{H}}\mathbf{A}\mathbf{x}$, thus yielding the following corollary:

Corollary A.1. *Let $\mathbf{x} = [x_1, \cdots, x_N]^{\mathrm{T}}$ be a $N \times 1$ vector where the entries x_i are centered i.i.d. complex random variables with unit variance and finite eight order. Let $\mathbf{A}$ be a determinsitic $N \times N$ complex matrix with bounded spectral norm. Then,*

$$\frac{1}{N}\mathbf{x}^{\mathrm{H}}\mathbf{A}\mathbf{x} - \frac{1}{N}\mathrm{Tr}(\mathbf{A}) \longrightarrow 0 \qquad \text{almost surely.}$$

By corollary A.1, the asymptotic behavior of $\mathbf{A}_{\sigma,K}$ is then given by:

$$A_{\sigma,K} - \frac{\sigma_v^2\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}}{N_1\sigma_P^2}\mathrm{Tr}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1} \longrightarrow 0 \qquad \text{almost surely.}$$

Since $\frac{1}{K}\mathrm{Tr}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}$ converges asymptotically to $\frac{1}{c_2-1}$ as the dimensions go to infinity [105], we get:

$$A_{\sigma,K} - \frac{c_1(1+r)\sigma_v^4}{(c_2-1)\sigma_P^2}\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i} \longrightarrow 0.$$

Note that TheoremA.1 can be applied since the smallest eigenvalue of the Wishart matrix $\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)$ are almost surely uniformely bounded away from zero by $(1-\sqrt{c_2})^2 > 0$, [92].

Before determining the limiting behavior of $B_{\sigma,K}$, we shall need the following lemma:

Lemma A.2. *Let* $\mathbf{Y} = \left(\frac{1}{\sqrt{K}} y_{i,j}\right)_{i=1,j=1}^{M,K}$ *be a* $M \times K$ *with Gaussian i.i.d entries. Then, in the asymptotic regime given by:*

$$M, K \to \infty \text{ such that } \frac{M}{K} \to c_2 > 1$$

we have:

$$\left[(\mathbf{Y}^{\mathrm{H}}\mathbf{Y})^{-2}\right]_{i,i} - \frac{c_2}{c_2 - 1}\left(\left[(\mathbf{Y}^{\mathrm{H}}\mathbf{Y})^{-1}\right]_{i,i}\right)^2 \to 0$$

Proof. Without loss of generality, we can restrict our proof to the case where $i = 1$. Let $\mathbf{y}_1, \cdots, \mathbf{y}_K$ denote the columns of $\mathbf{Y}$. Matrix $\mathbf{Y}^{\mathrm{H}}\mathbf{Y}$ is then given by:

$$\mathbf{Y}^{\mathrm{H}}\mathbf{Y} = \begin{bmatrix} \mathbf{y}_1^{\mathrm{H}}\mathbf{y}_1 & \mathbf{y}_1^{\mathrm{H}}\mathbf{y}_2 & \cdots & \mathbf{y}_1^{\mathrm{H}}\mathbf{y}_K \\ \vdots & & & \vdots \\ \mathbf{y}_K^{\mathrm{H}}\mathbf{y}_1 & \mathbf{y}_K^{\mathrm{H}}\mathbf{y}_2 & \cdots & \mathbf{y}_K^{\mathrm{H}}\mathbf{y}_K \end{bmatrix}$$

Let $\mathbf{v}_y = \left[\left[(\mathbf{Y}^{\mathrm{H}}\mathbf{Y})^{-1}\right]_{1,2}, \cdots, \left[(\mathbf{Y}^{\mathrm{H}}\mathbf{Y})^{-1}\right]_{1,K}\right]$. Then, using the formula of the inverse of block matrices, we get:

$$\mathbf{v}_y = -\left[(\mathbf{Y}^{\mathrm{H}}\mathbf{Y})^{-1}\right]_{1,1} \mathbf{y}_1^{\mathrm{H}}\widetilde{\mathbf{Y}}\left(\widetilde{\mathbf{Y}}^{\mathrm{H}}\widetilde{\mathbf{Y}}\right)^{-1}$$

where $\widetilde{\mathbf{Y}} = [\mathbf{y}_2, \cdots, \mathbf{y}_K]$.

On the other hand,

$$\begin{aligned}\left[(\mathbf{Y}^{\mathrm{H}}\mathbf{Y})^{-2}\right]_{1,1} &= \left(\left[(\mathbf{Y}^{\mathrm{H}}\mathbf{Y})^{-1}\right]_{1,1}\right)^2 + \mathbf{v}_y\mathbf{v}_y^{\mathrm{H}} \\ &= \left(\left[(\mathbf{Y}^{\mathrm{H}}\mathbf{Y})^{-1}\right]_{1,1}\right)^2 \left(1 + \mathbf{y}_1^{\mathrm{H}}\widetilde{\mathbf{Y}}\left(\widetilde{\mathbf{Y}}^{\mathrm{H}}\widetilde{\mathbf{Y}}\right)^{-2}\widetilde{\mathbf{Y}}^{\mathrm{H}}\mathbf{y}_1\right)\end{aligned}$$

Using corollary A.1, we have:

$$\mathbf{y}_1^{\mathrm{H}}\widetilde{\mathbf{Y}}\left(\widetilde{\mathbf{Y}}^{\mathrm{H}}\widetilde{\mathbf{Y}}\right)^{-2}\widetilde{\mathbf{Y}}^{\mathrm{H}}\mathbf{y}_1 - \frac{1}{K}\mathrm{Tr}\left(\widetilde{\mathbf{Y}}^{\mathrm{H}}\widetilde{\mathbf{Y}}\right)^{-1} \to 0 \text{ almost surely.}$$

Since $\frac{1}{K}\mathrm{Tr}\left(\widetilde{\mathbf{Y}}^{\mathrm{H}}\widetilde{\mathbf{Y}}\right)^{-1}$ tends to $\frac{1}{c_2 - 1}$ almost surely, we get the desired result. □

We are now in position to deal with the term $B_{\sigma,K}$. Using corollary A.1, we get:

$$B_{\sigma,K} - \frac{\sigma_v^4(M-K)}{N_1\sigma_P^2}\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-2}\right]_{i,i} \to 0 \text{ almost surely}$$

Hence,

$$B_{\sigma,K} - \frac{\sigma_v^4 c_1(c_2-1)(1+r)}{\sigma_P^2}\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-2}\right]_{i,i} \to 0 \text{ almost surely}$$

Using lemma A.2, we get that:

$$B_{\sigma,K} - \frac{\sigma_v^4 c_1 c_2(1+r)}{\sigma_P^2}\left(\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}\right)^2 \to 0 \text{ almost surely}$$

It can be shown that $\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}$ converge almost surely to $\frac{1}{c_2-1}$, (its inverse is the mean of independent random variables [113]), then:

$$B_{\sigma,K} - \frac{\sigma_v^4 c_1 c_2(1+r)}{\sigma_P^2(c_2-1)}\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i} \to 0 \text{ almost surely}$$

To prove the almost sure convergence to zero of $\epsilon_{\sigma,K}$, we will be based on the following result, about the asymptotic behaviour of weighted averages:

Theorem A.2. *Almost sure convergence of weighted averages [5] Let* $\mathbf{a} = [a_1, \cdots, a_N]^{\mathrm{T}}$ *be a sequence of* $N \times 1$ *deterministic real vectors with* $\sup_N \frac{1}{N}\mathbf{a}_N^{\mathrm{T}}\mathbf{a}_N < +\infty$. *Let* $\mathbf{x}_N = [x_1, \cdots, x_N]$ *be a* $N \times 1$ *real random vector with i.i.d. entries, such that* $\mathbb{E}x_1 = 0$ *and* $\mathbb{E}|x_1| < +\infty$. *Therefore,* $\frac{1}{N}\mathbf{a}_N^{\mathrm{T}}\mathbf{x}_N$ *converges almost surely to zero as* N *tends to infinity.*

This theorem was proved in [5] for real variables. Since we are interested in the asymptotic convergence of the real part of $\epsilon_{\sigma,K}$, it can be possible to transpose our problem into the real case. Indeed, let $\mathbf{x} = \mathbf{V}_1^{\mathrm{H}}\mathbf{h}_i^{\#}$ and $\mathbf{a} = \mathbf{P}_t^{\mathrm{H}}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\mathbf{h}_i^{\#}$, then $\Re\left(\epsilon_{\sigma,K}\right)$ is given by:

$$\Re\left(\epsilon_{\sigma,K}\right) = \frac{1}{N_1\sigma_P^2}\Re(\mathbf{x}^{\mathrm{H}}\mathbf{a})$$

Let $\mathbf{a}_r, \mathbf{x}_r$ (resp. $\mathbf{a}_i, \mathbf{x}_i$) denote respectively the real parts (resp. imaginary parts) of $\mathbf{a}$ and $\mathbf{x}$, then

$$\Re\left(\epsilon_{\sigma,K}\right) = \frac{1}{N_1\sigma_P^2}\mathbf{a}_r^{\mathrm{T}}\mathbf{x}_r - \mathbf{a}_i^{\mathrm{T}}\mathbf{x}_i$$

Referring to theorem A.2, the convergence to zero of $\Re\left(\epsilon_{\sigma,K}\right)$ is ensured if $\frac{1}{2N_1}\left(\mathbf{a}_r^{\mathrm{T}}\mathbf{a}_r + \mathbf{a}_i^{\mathrm{T}}\mathbf{a}_i\right) = \frac{1}{2N_1}\|\mathbf{a}\|_2^2$is finite. This is almost surely true, since:

$$\begin{aligned}\frac{1}{N_1\sigma_P^2}\|\mathbf{a}\|_2^2 &= \frac{1}{N_1\sigma_P^2}\mathrm{Tr}\left(\mathbf{P}_t^{\mathrm{H}}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\mathbf{h}_i^{\#}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\mathbf{h}_i^{\#}\right)\\ &= \mathbf{h}_i^{\#}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-2}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}} < \|\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-2}\|_2\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}\end{aligned}$$

This leads to

$$\sigma_{s,K}^2 - \tilde{\sigma}_{s,K}^2 \longrightarrow 0 \qquad \text{almost surely.}$$

where $\tilde{\sigma}_{s,K}^2$ is given by:

$$\tilde{\sigma}_{s,K}^2 = \sigma_v^2\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i} + \frac{c_1(c_2+1)(1+r)\sigma_v^4}{(c_2-1)\sigma_P^2}\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}.$$

Substituting $\sigma_{s,K}^2$ by its asymptotic equivalent in (A.1), we get:

$$\mathbb{E}\left[\exp\left(\jmath\Re\left(z^*\left(\Delta\mathbf{S}\right)_{i,j}\right)\right)|\mathbf{H},\mathbf{S}\right] - \mathbb{E}\left[\exp\left(-\jmath\Re\left(z^*\mathbf{h}_i^{\#}\Delta\mathbf{H}_t\mathbf{s}_j\right)\right)|\mathbf{S},\mathbf{H}\right]\exp\left(-\frac{1}{4}|z|^2\tilde{\sigma}_{s,N}^2\right) \longrightarrow 0 \text{ almost surely.}$$

Also conditioning on $\mathbf{S}$ and $\mathbf{H}$, $\mathbf{h}_i^{\#}\Delta\mathbf{H}_t\mathbf{s}_j$ is a Gaussian random variable with zero mean and variance

$$\sigma_{m,K}^2 = \frac{\sigma_v^2}{N_1\sigma_P^2}\mathbf{h}_i^{\#}{\mathbf{s}_j}^{\mathrm{H}}\left(\mathbf{P}_t\mathbf{P}_t^{\mathrm{H}}\right)^{-1}\mathbf{s}_j\left(\mathbf{h}_i\right)^{\#}.$$

Since $\frac{1}{K}\mathbf{s}_j{}^{\mathrm{H}}\mathbf{s}_j \longrightarrow \sigma_s^2$ almost surely, we get that $\sigma_{m,K}^2$ converges almost surely to $\tilde{\sigma}_{m,K}^2$ where

$$\tilde{\sigma}_{m,K}^2 = \frac{c_1(1+r)\sigma_v^2\sigma_s^2}{\sigma_P^2}\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i},$$

Using the fact that the characteristic function of $\mathbf{h}_i^{\#}\Delta\mathbf{H}_t\mathbf{s}_j$ is

$$\mathbb{E}\left[\exp\left(-\jmath\Re\left(z^*\mathbf{h}_i^{\#}\Delta\mathbf{H}_t\mathbf{s}_j\right)\right)|\mathbf{S},\mathbf{H}\right] = \exp\left(-\frac{1}{4}|z|^2\sigma_{m,K}^2\right),$$

we obtain that conditionally on the channel:

$$\mathbb{E}\left[\exp\left(\jmath\Re\left(z^*\left(\Delta\mathbf{S}\right)_{i,j}\right)\right)\right] - \exp\left(-\frac{1}{4}|z|^2\left(\tilde{\sigma}_{m,K}^2+\tilde{\sigma}_{s,K}^2\right)\right) \longrightarrow 0 \text{ almost surely.}$$

We end up the proof by noticing that $\tilde{\sigma}_{m,K}^2+\tilde{\sigma}_{s,K}^2 = \sigma_s^2\delta_t\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}$.

A.2 Proof of theorem 2.2

For the conventional scheme, the effective post-processing noise writes as:

$$\Delta\mathbf{W}_c = -\mathbf{H}^{\#}\Delta\mathbf{H}_c\left(\mathbf{W}+\mathbf{P}\right) + \left(\mathbf{I}_K - \mathbf{H}^{\#}\Delta\mathbf{H}_c\right)\mathbf{H}^{\#}\mathbf{V} + \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\Delta\mathbf{H}_c^{\mathrm{H}}\mathbf{\Pi}\mathbf{V}$$

Substituting $\Delta\mathbf{H}_c$ by its value, we get:

$$\begin{aligned}\Delta\mathbf{W}_c &= -\mathbf{W}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{W} - \mathbf{W}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P} - \mathbf{H}^{\#}\mathbf{V}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{W} - \mathbf{H}^{\#}\mathbf{V}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}\\ &\quad +\mathbf{H}^{\#}\mathbf{V} - \mathbf{W}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{H}^{\#}\mathbf{V} - \mathbf{H}^{\#}\mathbf{V}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{H}^{\#}\mathbf{V} + \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}\mathbf{V}^{\mathrm{H}}\mathbf{\Pi}\mathbf{V}\end{aligned}$$

Denote by $\mathbf{w}_i$, $\tilde{\mathbf{w}}_i$ the i th column and row of $\mathbf{W}$. Also, denote by $\mathbf{p}_j$, $\mathbf{v}_j$, $\mathbf{h}_i^{\#}$ and $\widetilde{\mathbf{h}}_i$ the j th column of $\mathbf{P}$ and $\mathbf{V}$ and the i th rows of $\mathbf{H}^{\#}$ and $\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}$, respectively. Then, $\left[\Delta\mathbf{W}_c\right]_{i,j}$ can be written as:

$$\begin{aligned}\left[\Delta\mathbf{W}_c\right]_{i,j} &= -\tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{w}_j - \tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{p}_j - \mathbf{h}_i^{\#}\mathbf{V}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{w}_j - \mathbf{h}_i^{\#}\mathbf{V}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{p}_j\\ &\quad +\mathbf{h}_i^{\#}\mathbf{v}_j - \tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{H}^{\#}\mathbf{v}_j - \mathbf{h}_i^{\#}\mathbf{V}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{H}^{\#}\mathbf{v}_j + \widetilde{\mathbf{h}}_i\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}\mathbf{V}^{\mathrm{H}}\mathbf{\Pi}\mathbf{v}_j\end{aligned}$$

Hereafter, we will derive the asymptotic distribution for the random variable $\left[\Delta\mathbf{W}_c\right]_{i,j}$. Similarly to the TDMT case, we will be based on the 'characteristic function approach'. We refer the reader to theorem A.1 in A.1 for the expression of the characteristic function.

Conditioning on $\mathbf{v}_j$, $\tilde{\mathbf{w}}_i$ and $\mathbf{w}_j$, $\left[\Delta\mathbf{W}_c\right]_{i,j}$ is Gaussian distributed, since it involves a linear combination of independent Gaussian random variables. Hence, the conditional characteristic function of $\left[\Delta\mathbf{W}_c\right]_{i,j}$ writes as:

$$\mathbb{E}\left(\exp\left(\jmath\Re\left(z^*\left[\Delta\mathbf{W}_c\right]_{i,j}\right)\right)|\mathbf{v}_j,\tilde{\mathbf{w}}_i,\mathbf{w}_j\right) = \exp\left(\jmath s m_N - |z|^2\sigma_N^2/2\right)$$

where

$$\begin{aligned}m_N &= -\tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{w}_j - \tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{p}_j - \mathbf{h}_i^{\#}\mathbf{v}_j\mathbf{p}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{w}_j - \mathbf{h}_i^{\#}\mathbf{v}_j\mathbf{p}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{p}_j\\ &\quad +\mathbf{h}_i^{\#}\mathbf{v}_j - \tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{H}^{\#}\mathbf{v}_j - \mathbf{h}_i^{\#}\mathbf{v}_j\mathbf{p}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{H}^{\#}\mathbf{v}_j + \widetilde{\mathbf{h}}_i\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{p}_j\mathbf{v}_j^{\mathrm{H}}\mathbf{\Pi}\mathbf{v}_j\\ \sigma_N^2 &= \mathbb{E}\left[\left|\mathbf{h}_i^{\#}\mathbf{V}_j\mathbf{P}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{w}_j + \mathbf{h}_i^{\#}\mathbf{V}_j\mathbf{P}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{p}_j + \mathbf{h}_i^{\#}\mathbf{V}_j\mathbf{P}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{H}^{\#}\mathbf{v}_j\right.\right.\\ &\quad \left.\left.+\widetilde{\mathbf{h}}_i\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}_j\mathbf{V}_j^{\mathrm{H}}\mathbf{\Pi}\mathbf{v}_j\right|^2|\mathbf{v}_j,\tilde{\mathbf{w}}_i,\mathbf{w}_j\right],\end{aligned}$$

where $\mathbf{P}_j$ and $\mathbf{V}_j$ being the $\mathbf{P}$ and $\mathbf{V}$ matrices with the j th column removed.

$$\sigma_N^2 = X_N + A + Y_N + Z_N + 2\Re(\epsilon_{N,1}) + 2\Re(\epsilon_{N,2}) + 2\Re(\epsilon_{N,3}),$$

where:

$$\begin{aligned}
X_N &= \mathbb{E}\left[\mathbf{h}_i^{\#}\mathbf{V}_j\mathbf{P}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{w}_j\mathbf{w}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}_j\mathbf{V}_j^{\mathrm{H}}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}|\mathbf{v}_j,\tilde{\mathbf{w}}_i,\mathbf{w}_j\right]\\
A &= \mathbb{E}\left[\mathbf{h}_i^{\#}\mathbf{V}_j\mathbf{P}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{p}_j\mathbf{p}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}_j\mathbf{V}_j^{\mathrm{H}}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\right]\\
Y_N &= \mathbb{E}\left[\mathbf{h}_i^{\#}\mathbf{V}_j\mathbf{P}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{H}^{\#}\mathbf{v}_j\mathbf{v}_j^{\mathrm{H}}\left(\mathbf{H}^{\#}\right)^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}_j\mathbf{V}_j^{\mathrm{H}}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}|\mathbf{v}_j,\tilde{\mathbf{w}}_i,\mathbf{w}_j\right]\\
Z_N &= \mathbb{E}\left[\widetilde{\mathbf{h}}_i\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}_j\mathbf{V}_j^{\mathrm{H}}\mathbf{\Pi}\mathbf{v}_j\mathbf{v}_j^{\mathrm{H}}\mathbf{\Pi}\mathbf{V}_j\mathbf{P}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\widetilde{\mathbf{h}}_i^{\mathrm{H}}|\mathbf{v}_j,\tilde{\mathbf{w}}_i,\mathbf{w}_j\right]\\
\epsilon_{N,1} &= \mathbb{E}\left[\mathbf{h}_i^{\#}\mathbf{V}_j\mathbf{P}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{w}_j\mathbf{p}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}_j\mathbf{V}_j^{\mathrm{H}}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}|\mathbf{v}_j,\tilde{\mathbf{w}}_i,\mathbf{w}_j\right]\\
\epsilon_{N,2} &= \mathbb{E}\left[\mathbf{h}_i^{\#}\mathbf{V}_j\mathbf{P}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{H}^{\#}\mathbf{v}_j\mathbf{w}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}_j\mathbf{V}_j^{\mathrm{H}}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}|\mathbf{v}_j,\tilde{\mathbf{w}}_i,\mathbf{w}_j\right]\\
\epsilon_{N,3} &= \mathbb{E}\left[\mathbf{h}_i^{\#}\mathbf{V}_j\mathbf{P}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{p}_j\mathbf{v}_j^{\mathrm{H}}\left(\mathbf{H}^{\#}\right)^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}_j\mathbf{V}_j^{\mathrm{H}}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}|\mathbf{v}_j,\tilde{\mathbf{w}}_i,\mathbf{w}_j\right]
\end{aligned}$$

A is a determinsitic quantity which can be computed easily as follows:

$$\begin{aligned}
A &= \sigma_v^2\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}\mathrm{Tr}\left(\mathbf{P}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{p}_j\mathbf{p}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}_j\right)\\
&= \sigma_v^2\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}\mathbf{p}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{p}_j\left(1-\mathbf{p}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{p}_j\right)
\end{aligned}$$

One can prove easily that for our particular choice of the matrix $\mathbf{P}$, we have $\mathbf{p}_j^{\mathrm{H}}\mathbf{p}_j = K\sigma_P^2$. Therefore,

$$A = \sigma_v^2\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}c_1\left(1-c_1\right) \tag{A.4}$$

The remaining task now is to determine the limiting behavior of the introduced random quantities. For that, we will sequentially determine almost sure limits for the random variables X_N, Y_N and Z_N and prove the almost sure convergence to zero of $\epsilon_{N,1}$, $\epsilon_{N,2}$ and $\epsilon_{N,3}$.

1. Limiting behavior of X_N. Computing the expectation with respect to the probability density function of $\mathbf{V}_j$, we can prove that:

$$X_N = \sigma_v^2\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}\mathbf{w}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}_j\mathbf{P}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{w}_j$$

   X_N is therefore a quadratic form expression with respect to $\mathbf{w}_j$. Hence, the limiting behavior of X_N can be derived by using the result about the almost sure convergence of quadratic forms that has been stated in A.1. Indeed, using corollary A.1, it is easy to see that X_N satisfies:

$$X_N - \frac{c_1\sigma_v^2\sigma_{w_c}^2}{\sigma_P^2}\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}\xrightarrow[K\to\infty]{}0 \tag{A.5}$$

2. Limiting behavior of Y_N: By the same way, we compute first the expectation with respect to $\mathbf{V}_j$, thus leading to:

$$Y_N = \sigma_v^2\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}\mathbf{v}_j^{\mathrm{H}}\left(\mathbf{H}^{\#}\right)^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}_j\mathbf{P}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{H}^{\#}\mathbf{v}_j.$$

Using corollary A.1, we get:

$$Y_N - \frac{\sigma_v^4}{N\sigma_P^2}\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}\mathrm{Tr}\,(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1} + \frac{\sigma_v^4}{N^2\sigma_P^4}\mathbf{p}_j^{\mathrm{H}}(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\mathbf{p}_j \xrightarrow[K\to\infty]{} 0 \text{ almost surely}$$

Since the spectral norm of $(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}$ is bounded almost surely [3], the last term in the above equation could be shown to converge to zero almost surely. Hence, we get:

$$Y_N - \frac{\sigma_v^4}{N\sigma_P^2}\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}\mathrm{Tr}\,(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1} \xrightarrow[K\to\infty]{} 0 \text{ almost surely}$$

Knowing that $\frac{1}{K}\mathrm{Tr}\,(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}$ converges to $\frac{1}{c_2-1}$ almost surely, we obtain:

$$Y_N - \frac{c_1\sigma_v^4}{(c_2-1)\sigma_P^2}\left[(\mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}\right]_{i,i} \xrightarrow[K\to\infty]{} 0 \text{ almost surely} \tag{A.6}$$

3. Limiting behaviour of Z_N:

   Computing the expectation with respect to the probability distribution of $\mathbf{V}_j$, we get:

$$\begin{aligned}
Z_N &= \mathbb{E}\left[\widetilde{\mathbf{h}}_i(\mathbf{P}\mathbf{P}^{\mathrm{H}})^{-1}\mathbf{P}_j\mathbf{V}_j^{\mathrm{H}}\mathbf{\Pi}\mathbf{v}_j\mathbf{v}_j^{\mathrm{H}}\mathbf{\Pi}\mathbf{V}_j\mathbf{P}_j^{\mathrm{H}}(\mathbf{P}\mathbf{P}^{\mathrm{H}})^{-1}\widetilde{\mathbf{h}}_i^{\mathrm{H}}\right] \\
&= \sigma_v^2\mathbf{v}_j^{\mathrm{H}}\mathbf{\Pi}\mathbf{v}_j\widetilde{\mathbf{h}}_i(\mathbf{P}\mathbf{P}^{\mathrm{H}})^{-1}\mathbf{P}_j\mathbf{P}_j^{\mathrm{H}}(\mathbf{P}\mathbf{P}^{\mathrm{H}})^{-1}\widetilde{\mathbf{h}}_i^{\mathrm{H}} \\
&= \frac{\sigma_v^2\mathbf{v}_j^{\mathrm{H}}\mathbf{\Pi}\mathbf{v}_j}{N\sigma_P^2}\left(\widetilde{\mathbf{h}}_i\widetilde{\mathbf{h}}_i^{\mathrm{H}} - \frac{1}{N\sigma_P^2}\widetilde{\mathbf{h}}_i\mathbf{p}_j\mathbf{p}_j^{\mathrm{H}}\widetilde{\mathbf{h}}_i^{\mathrm{H}}\right) \\
&= \frac{\sigma_v^2\mathbf{v}_j^{\mathrm{H}}\mathbf{\Pi}\mathbf{v}_j}{N\sigma_P^2}\left(\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-2}\right]_{i,i} - \frac{1}{N\sigma_P^2}\widetilde{\mathbf{h}}_i\mathbf{p}_j\mathbf{p}_j^{\mathrm{H}}\widetilde{\mathbf{h}}_i^{\mathrm{H}}\right)
\end{aligned}$$

   Using corollary A.1, we get that $\frac{\mathbf{v}_j^{\mathrm{H}}\mathbf{\Pi}\mathbf{v}_j}{N\sigma_P^2} - \frac{(M-K)\sigma_v^2}{N\sigma_P^2} \to 0$ almost surely. Also, from A.2, we get that: $\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-2}\right]_{i,i} - \frac{c_2}{(c_2-1)^2}\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i} \to 0$ almost surely. Also using the same technique as in lemma A.2, we can prove that: $\frac{1}{N\sigma_P^2}\widetilde{\mathbf{h}}_i\mathbf{p}_j\mathbf{p}_j^{\mathrm{H}}\widetilde{\mathbf{h}}_i^{\mathrm{H}} \to 0$ almost surely. This implies that:

$$Z_N - \frac{c_1c_2\sigma_v^4}{(c_2-1)\sigma_P^2}\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i} \to 0 \text{ almost surely.} \tag{A.7}$$

4. Almost sure convergence to zero of $\Re(\epsilon_{N,1})$, $\Re(\epsilon_{N,2})$ and $\Re(\epsilon_{N,3})$. Computing the expectation with respect to $\mathbf{V}_j$, we get:

$$\begin{aligned}
\epsilon_{N,1} &= \frac{\sigma_v^2(1-c_1)}{N\sigma_P^2}\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}\mathbf{p}_j^{\mathrm{H}}\mathbf{w}_j \\
\epsilon_{N,2} &= \frac{\sigma_v^2}{N\sigma_P^2}\left[\mathbf{w}_j^{\mathrm{H}}\mathbf{H}^{\#}\mathbf{v}_j - \frac{1}{N\sigma_P^2}\mathbf{w}_j^{\mathrm{H}}\mathbf{p}_j\mathbf{p}_j^{\mathrm{H}}\mathbf{H}^{\#}\mathbf{v}_j\right] \\
\epsilon_{N,3} &= \frac{(1-c_1)\sigma_v^2}{N\sigma_P^2}\mathbf{v}_j^{\mathrm{H}}\left(\mathbf{H}^{\#}\right)^{\mathrm{H}}\mathbf{p}_j
\end{aligned}$$

   To prove the almost sure convergence to zero of all these terms, we will be based, in the same way as in the TDMT case, on theorem A.2 about the asymptotic behaviour of

weighted averages: We will prove the almost sure convergence of $\Re(\epsilon_{N,3})$. For $\Re(\epsilon_{N,1})$ and $\Re(\epsilon_{N,2})$, the proof is quite similar.

Let $\mathbf{g}_j = \mathbf{H}^{\#}\mathbf{v}_j - \frac{1}{N\sigma_P^2}\mathbf{p}_j\mathbf{p}_j^{\mathrm{H}}\mathbf{H}^{\#}\mathbf{v}_j$,then $\epsilon_{N,3}$ can be given as:

$$\epsilon_{N,2} = \frac{\sigma_v^2}{N\sigma_P^2}\mathbf{w}_j^{\mathrm{H}}\mathbf{g}_j$$

Let $\mathbf{w}_{j,r}$, $\mathbf{g}_{j,r}$ and $\mathbf{w}_{j,i}$, $\mathbf{g}_{j,i}$ be respectively the real and imaginary parts of $\mathbf{w}_j$ and $\mathbf{g}_j$. Then

$$\Re\epsilon_{N,2} = \frac{\sigma_v^2}{N\sigma_P^2}\left(\mathbf{w}_r^{\mathrm{T}}\mathbf{g}_r - \mathbf{w}_i^{\mathrm{T}}\mathbf{g}_i\right).$$

According to theorem A.2, we need only to prove that $\frac{1}{2N}\left(\mathbf{g}_r^{\mathrm{T}}\mathbf{g}_r + \mathbf{g}_i^{\mathrm{T}}\mathbf{g}_i\right) = \frac{1}{2N}\|\mathbf{g}\|_2^2$ is bounded to conclude about the almost sure convergence to zero of $\Re(\epsilon_{N,2})$. This is almost surely true, since:

$$\frac{1}{2N}\|\mathbf{g}\|_2^2 \leq \frac{1}{N}\mathbf{v}_j^{\mathrm{H}}\left(\mathbf{H}^{\#}\right)^{\mathrm{H}}\mathbf{H}^{\#}\mathbf{v}_j + \frac{1}{N^2\sigma_P^2}\mathbf{v}_j^{\mathrm{H}}\left(\mathbf{H}^{\#}\right)^{\mathrm{H}}\mathbf{p}_j\mathbf{p}_j^{\mathrm{H}}\mathbf{p}_j\mathbf{p}_j^{\mathrm{H}}\mathbf{H}^{\#}\mathbf{v}_j \tag{A.8}$$

Using Corollary A.1, the right hand side in (A.8) is almost surely equivalent to $\frac{1}{N}\sigma_v^2\mathrm{Tr}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1} + \frac{c_1}{N}\mathbf{p}_j^{\mathrm{H}}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\mathbf{p}_j$. The spectral norm of $\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}$ is almost surely bounded, hence so is $\frac{1}{2N}\|\mathbf{g}\|_2^2$.

From (A.4), (A.5), (A.6) and (A.7), it ensues that:

$$\sigma_N^2 - \tilde{\sigma}_N^2 \xrightarrow[K\to\infty]{} 0 \text{ almost surely}$$

where

$$\tilde{\sigma}_N^2 = \left(c_1(1-c_1)\sigma_v^2 + \frac{c_1\sigma_v^2\sigma_{w_c}^2}{\sigma_P^2} + \frac{c_1(c_2+1)\sigma_v^4}{(c_2-1)\sigma_P^2}\right)\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}$$

Hence, almost surely, the following convergence holds:

$$\mathbb{E}\left[\exp\left(\jmath\Re\left(z^*\left[\Delta\mathbf{W}_c\right]_{i,j}\right)\right)|\mathbf{v}_j,\tilde{\mathbf{w}}_i,\mathbf{w}_j\right] - \mathbb{E}\left[\exp\left(\jmath\Re\left(z^*m_N\right)\right)|\mathbf{v}_j,\tilde{\mathbf{w}}_i,\mathbf{w}_j\right]\exp\left(-\frac{1}{4}||z|^2|\tilde{\sigma}_N^2\right) \xrightarrow[K\to\infty]{} 0 \tag{A.9}$$

This finishes the first part of the proof. It remains now to work out the characteristic function of the random variable m_N. We note that m_N can be written as the sum of $M_{N,1}$, $M_{N,2}$, $M_{N,3}$ and $M_{N,4}$ where:

$$\begin{aligned}
M_{N,1} &= -\tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{w}_j - \tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{p}_j\\
M_{N,2} &= -\mathbf{h}_i^{\#}\mathbf{v}_j\mathbf{p}_j{}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{w}_j - \mathbf{h}_i^{\#}\mathbf{v}_j\mathbf{p}_j{}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{H}^{\#}\mathbf{v}_j\\
M_{N,3} &= -\mathbf{h}_i^{\#}\mathbf{v}_j\mathbf{p}_j{}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{p}_j + \mathbf{h}_i^{\#}\mathbf{v}_j - \tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{H}^{\#}\mathbf{v}_j\\
M_{N,4} &= \widetilde{\mathbf{h}}_i\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{p}_j\mathbf{v}_j^{\mathrm{H}}\mathbf{\Pi}\mathbf{v}_j
\end{aligned}$$

Since $M_{N,1}$ does not depend on $\mathbf{v}_j$, we have:

$$\mathbb{E}\left[\exp(\jmath\Re\left(z^*m_N\right))|\tilde{\mathbf{w}}_i,\mathbf{w}_j\right] = \exp\left(\jmath\Re\left(z^*M_{N,1}\right)\right)\mathbb{E}\left[\exp\left(\jmath\Re\left(z^*M_{N,2}+z^*M_{N,3}\right)\right)|\tilde{\mathbf{w}}_i,\mathbf{w}_j\right] \tag{A.10}$$

To determine the characteristic function of m_N, we will next sequentially:

- Prove that $M_{N,2}$ converges to zero in probability,
- Find an asymptotic equivalent of the characteristic function of $M_{N,3} + M_{N,4}$,
- Determine an asymptotic equivalent of the characteristic function of $M_{N,1}$.

1. Convergence in probability of $M_{N,2}$: $M_{N,2}$ can be decomposed into the sum of two terms:

$$M_{N,2} = M_{N,2,1} + M_{N,2,2}$$

where

$$\begin{aligned} M_{N,2,1} &= -\mathbf{h}_i^{\#}\mathbf{v}_j\mathbf{p}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{w}_j \\ M_{N,2,2} &= -\mathbf{h}_i^{\#}\mathbf{v}_j\mathbf{p}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{H}^{\#}\mathbf{v}_j \end{aligned}$$

To prove the convergence of $M_{N,2}$ to zero, we need only to show that $M_{N,2,1}$ and $M_{N,2,2}$ converge to zero in probability. This can be verified, by showing that the second moment of $M_{N,2,1}$ and $M_{N,2,2}$ converge to zero in the asymptotic regime. Actually, we have:

$$\begin{aligned} \mathbb{E}\left[|M_{N,2,1}|^2\right] &= \mathbb{E}\left[\left|\mathbf{h}_i^{\#}\mathbf{v}_j\mathbf{p}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{w}_j\right|^2\right] \\ &= \sigma_v^2\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}\mathrm{Tr}\left(\mathbf{p}_j^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-2}\mathbf{p}_j\right) \\ &= \frac{K\sigma_v^2}{N^2\sigma_P^2}\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i} \\ &\xrightarrow[K\to\infty]{} 0 \end{aligned}$$

To compute the second moment of $M_{N,2,2}$, we shall need the following lemma

Lemma A.3. *Let* $\mathbf{x}$ *be a* $N\times 1$ *complex random vector with circular complex i.i.d entries. Let* $\mathbf{y}$ *be a* $N\times 1$ *deterministic vector, and* $\mathbf{A}$ *be a* $N\times N$ *deterministic matrix then:*

$$\mathbb{E}\left[\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}\mathbf{x}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{x}^*\right] = \kappa\sum_{i=1}^{N}|A_{i,i}|^2 + 2\left(\mathbb{E}|x_1|^2\right)^2\mathrm{Tr}(\mathbf{A}\mathbf{A}^{\mathrm{H}})$$

where $\kappa = \mathbb{E}|x_1|^4 - 2\left(\mathbb{E}|x_1|^2\right)^2$.

Proof.

$$\begin{aligned} \mathbb{E}\left[\mathbf{x}^{\mathrm{T}}\mathbf{A}\mathbf{x}\mathbf{x}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{x}^*\right] &= \mathbb{E}\sum_{i,j,k,l}x_iA_{i,j}x_jx_k^*A_{lk}^*x_l^* \\ &= \sum_{i=1}^{N}\mathbb{E}|x_1|^4|A_{i,i}|^2 + 2\left(\mathbb{E}|x_1|^2\right)^2\sum_{i,j,i\neq j}A_{i,j}A_{j,i}^* \\ &= \left(\mathbb{E}|x_1|^4 - 2\left(\mathbb{E}|x_1|^2\right)^2\right)\sum_{i=1}^{N}|A_{i,i}|^2 + 2\left(\mathbb{E}|x_1|^2\right)^2\sum_{i,j}A_{i,j}A_{j,i}^* \\ &= \kappa\sum_{i=1}^{N}|A_{i,i}|^2 + 2\left(\mathbb{E}|x_1|^2\right)^2\mathrm{Tr}\left(\mathbf{A}\mathbf{A}^{\mathrm{H}}\right) \end{aligned}$$

□

We are now in position to compute $\mathbb{E}|M_{N,2,2}|^2$. Since $\mathbf{v}$ has Gaussian entries, $\mathbb{E}|v_1|^4 - 2\left(\mathbb{E}|v_1|^2\right)^2 = 0$. Using lemma A.3, we get:

$$\begin{aligned}
\mathbb{E}|M_{N,2,2}|^2 &= 2\sigma_v^4 \mathrm{Tr}\left(\left(\mathbf{h}_i^{\#}\right)^{\mathrm{T}} \mathbf{p}_j^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{H}^{\#} \left(\mathbf{H}^{\#}\right)^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{p}_j \mathbf{h}_i^{\#}\right) \\
&= \frac{2\sigma_v^4}{N^2\sigma_P^2} \mathbf{p}_j^{\mathrm{H}} \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1} \mathbf{p}_j \mathrm{Tr}(\left(\mathbf{h}_i^{\#}\right)^{\mathrm{T}} \mathbf{h}_i) \\
&= \frac{2\sigma_v^4}{N^2\sigma_P^2} \mathbf{p}_j^{\mathrm{H}} \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1} \mathbf{p}_j \left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i} \\
&\leq \frac{2K\sigma_v^4}{N^2} \| \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1} \|_2^2 \\
&\xrightarrow[K\to\infty]{} 0
\end{aligned}$$

2. Asymptotic equivalent of the characteristic function of $M_{N,3} + M_{N,4}$:

   Conditioned on $\mathbf{w}_i$, $M_{N,3}$ is a Gaussian random variable with zero mean. Hence, the characteristic function of $M_{N,3}$ is given by:

$$\mathbb{E}\left[\exp\left(j\Re(z^* M_{N,3})\right) | \tilde{\mathbf{w}}_i\right] = \exp\left(-\frac{1}{4}|z|^2\sigma_M^2\right)$$

   where σ_M^2 is given by:

$$\begin{aligned}
\sigma_M^2 &= \mathbb{E}\left[\mathbf{h}_i^{\#}\mathbf{v}_j\mathbf{p}_j^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{p}_j\mathbf{p}_j^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{p}_j\mathbf{v}_j^{\mathrm{H}} \left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\right] + \mathbb{E}\left[\mathbf{h}_i^{\#}\mathbf{v}_j\mathbf{v}_j^{\mathrm{H}} \left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\right] \\
&+ \mathbb{E}\left[\tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{H}^{\#}\mathbf{v}_j\mathbf{v}_j^{\mathrm{H}} \left(\mathbf{H}^{\#}\right)^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{P}\tilde{\mathbf{w}}_i^{\mathrm{H}}\right] - 2\mathbb{E}\Re\left(\mathbf{h}_i^{\#}\mathbf{v}_j\mathbf{p}_j^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{p}_j\mathbf{v}_j^{\mathrm{H}} \left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\right) \\
&- 2\mathbb{E}\Re\left(\tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{H}^{\#}\mathbf{v}_j\mathbf{v}_j^{\mathrm{H}} \left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\right) + 2\mathbb{E}\Re\left(\tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{H}^{\#}\mathbf{v}_j\mathbf{p}_j^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{p}_j\mathbf{v}_j^{\mathrm{H}} \left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\right) \\
&= \sigma_v^2 \mathrm{Tr}\left(\mathbf{p}_j^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{p}_j\mathbf{p}_j^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{p}_j\mathbf{h}_i^{\#} \left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\right) + \sigma_v^2 \left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i} \\
&+ \sigma_v^2 \tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \tilde{\mathbf{w}}_i^{\mathrm{H}} - 2c_1\sigma_v^2 \left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i} - 2\sigma_v^2 \Re\left(\tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{H}^{\#} \left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\right) \\
&+ 2\sigma_v^2 c_1 \Re\left(\tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{H}^{\#} \left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\right) \\
&= \sigma_v^2(1-c_1)^2 \left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i} + \sigma_v^2 \tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \tilde{\mathbf{w}}_i^{\mathrm{H}}
\end{aligned}$$

Using corollaryA.1, we can prove that:

$$\frac{\sigma_v^2}{N^2\sigma_P^4} \tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}} \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1} \mathbf{P}\tilde{\mathbf{w}}_i^{\mathrm{H}} - \frac{\sigma_v^2\sigma_{w_c}^2}{N\sigma_P^2} \mathrm{Tr}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1} \xrightarrow[K\to\infty]{} 0 \text{ almost surely.}$$

Since $\frac{1}{K}\mathrm{Tr}\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1} \xrightarrow[K\to\infty]{} \frac{1}{c_2-1}$ almost surely, we get:

$$\frac{\sigma_v^2}{N^2\sigma_P^4} \tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}} \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1} \mathbf{P}\tilde{\mathbf{w}}_i^{\mathrm{H}} \xrightarrow[K\to\infty]{} \frac{c_1\sigma_v^2\sigma_{w_c}^2}{\sigma_P^2(c_2-1)}$$

Also using theorem A.2, we can prove by the same way as before that $\Re\left(\tilde{\mathbf{w}}_i\mathbf{P}^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{H}^{\#} \left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\right)$ converge almost surely to zero. Consequently;

$$\sigma_M^2 - \tilde{\sigma}_M^2 \xrightarrow[K\to\infty]{} 0 \text{ almost surely}$$

where

$$\tilde{\sigma}_M^2 = \sigma_v^2 (1-c_1)^2 \left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1} \right]_{i,i} + \frac{c_1 \sigma_v^2 \sigma_{w_c}^2}{\sigma_P^2 (c_2 - 1)}$$

An asymptotic equivalent for the characteristic function of $M_{N,3}$ can be therefore given by:

$$\mathbb{E}\left[\exp\left(\jmath \Re\left(z^* M_{N,3}\right)\right)\right] - \exp\left(-\frac{|z|^2}{4}\tilde{\sigma}_M^2\right) \xrightarrow[K\to\infty]{} 0 \text{ almost surely.}$$

Let us now deal with the term $M_{N,4}$. We decompose $M_{N,4}$ as

$$M_{N,4} = \frac{\widetilde{\mathbf{h}}_i \mathbf{p}_j}{\sqrt{N\sigma_P^2}} \left(\frac{\mathbf{v}_j \mathbf{\Pi} \mathbf{v}_j}{\sqrt{N\sigma_P^2}} - \frac{(M-K)\sigma_v^2}{\sqrt{N\sigma_P^2}} \right) + \frac{(M-K)\sigma_v^2 \widetilde{\mathbf{h}}_i \mathbf{p}_j}{N\sigma_P^2}.$$

One can see that the right hand side converges to zero in distribution, since $\frac{\widetilde{\mathbf{h}}_i \mathbf{p}_j}{\sqrt{N\sigma_P^2}}$ converges to zero in probability. Hence, we get:

$$M_{N,4} - c_1(c_2-1)\frac{\sigma_v^2 \widetilde{\mathbf{h}}_i \mathbf{p}_j}{\sigma_P^2} \to 0. \text{ in probability.}$$

Consequently, $M_{N,4} + M_{N,3}$ satisfies:

$$\mathbb{E}\left[\exp\left(\jmath\Re(z^*(M_{N,3}+M_{N,4}))\right)\right] - \exp(-\frac{|z|^2}{4}\widetilde{\sigma}_M^2) \exp\left(\jmath\Re\left(z^* c_1(c_2-1)\frac{\sigma_v^2 \widetilde{\mathbf{h}}_i \mathbf{p}_j}{\sigma_P^2}\right)\right) \to 0$$

Since $M_{N,2}$ converges to zero in probability, we can prove by using Slutsky theorem, that:

$$\mathbb{E}\left[\exp\left(\jmath\Re\left(z^* M_{N,3} + z^* M_{N,2}\right)\right)\right] - \exp\left(-\frac{|z|^2}{4}\tilde{\sigma}_M^2\right) \exp\left(\jmath\Re\left(z^*\theta\right)\right) \xrightarrow[K\to\infty]{} 0 \text{ almost surely.} \tag{A.11}$$

where $\theta = c_1(c_2-1)\frac{\sigma_v^2 \widetilde{\mathbf{h}}_i \mathbf{p}_j}{\sigma_P^2}$

3. Asymptotic equivalent of the characteristic function of $M_{N,1}$ To determine an asymptotic equivalent of the characteristic function of $M_{N,1}$, we first decompose $M_{N,1}$ as the sum of two terms, $M_{N,1,1}$ and $M_{N,1,2}$ where,

$$M_{N,1,1} = -\tilde{\mathbf{w}}_i \mathbf{P}^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{w}_j$$
$$M_{N,1,2} = -\tilde{\mathbf{w}}_i \mathbf{P}^{\mathrm{H}} \left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1} \mathbf{p}_j$$

Note that $\tilde{\mathbf{w}}_i$ and $\mathbf{w}_j$ are not independent since they share in common the variable $w_{i,j}$, but for sake of simplicity, we assume that they are independent, the final result remaining the same since the contribution of the term $w_{i,j}$ converges to zero almost surely.

Consequently, conditioned on $\tilde{\mathbf{w}}_i$, the characteristic function of $M_{N,1}$ writes as:

$$\mathbb{E}\left[\exp\left(\jmath\Re\left(z^* M_{N,1}\right)\right)|\tilde{\mathbf{w}}_i\right] = \mathbb{E}\left[\exp\left(\jmath\Re\left(z^* M_{N,1,1}\right)\right)|\tilde{\mathbf{w}}_i\right] \exp\left(\jmath\Re\left(z^* M_{N,1,2}\right)\right). \tag{A.12}$$

Using the CLT theorem for martingales, it can be shown that $M_{N,1,1}$ behaves asymptotically like a Gaussian random variable. More particularly, we can show that under some conditions, a bilinear form converges in distribution to a Gaussion random variable. For that, we will use the CLT theorem for martingales, which can be stated as follows.

Theorem A.3. *For each $n \in \mathbb{N}$, let $\{\xi_{n,j}, j = 0, \cdots, n\}$ be a martingale difference array with respect to a filtration $\{\mathcal{F}_{nj}\}$, for which,*

- $\sum_j \xi_{nj}^2 \to 1$ *in probability,*
- $\max_j |\xi_{nj}| \to 0$ *in probability,*
- $\sup_n \mathbb{E} \max_j \xi_{nj}^2 < \infty$

Then $\sum_j \xi_{nj}$ converges in distribution to $\mathcal{N}(0,1)$ as $n \to \infty$.

Based on theorem A.3, we will prove the following result about the convergence in distribution of a bilinear form:

Theorem A.4. *Let $\mathbf{x}$ be a $N \times 1$ complex random vector with i.i.d. entries with zero mean and unit variance and with bounded entries. $\mathbf{A}$ be a sequence of deterministic $K \times N$ matrices with bounded spectral norm, and such that the maximum of the entries of $\mathbf{A}$ are of order $\mathcal{O}(\sqrt{N})$, i.e, there exists a constant C_m and C_M such that:*

$$\frac{C_m}{\sqrt{N}} \leq |\mathbf{A}_{i,j}| \leq \frac{C_M}{\sqrt{N}}.$$

$\mathbf{y}$ be a sequence of $K \times 1$ random vectors independent from $\mathbf{x}$, with independent and bounded entries. Then, in the asymptotic regime defined as $N \to \infty, K \to \infty$ and $\frac{K}{N} \to c_1$ we have:

$$\sqrt{\frac{1}{\mathrm{Tr}\left(\mathbf{A}^{\mathrm{H}}\mathbf{A}\right)}}\mathbf{x}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{y} \to \mathcal{N}(0,1) \quad \text{in distribution}$$

Proof. To prove theorem A.4, we need to compute, the characteristic function given by:

$$\Phi(z) = \mathbb{E}\left[\exp\left(\jmath\Re\left(\frac{z^*}{\sqrt{\mathrm{Tr}(\mathbf{A}^{\mathrm{H}}\mathbf{A})}}\mathbf{x}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{y}\right)\right)\right] \tag{A.13}$$

Theorem A.3 being valid for real random variables, we should transform our problem into the real case. Let z_1, z_2, $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{y}_1$, $\mathbf{y}_2$, $\mathbf{A}_1$ and $\mathbf{A}_2$ be real variables defined as:

$$\begin{aligned} z &= z_1 + \jmath z_2 \\ \mathbf{x} &= \mathbf{x}_1 + \jmath\mathbf{x}_2 \\ \mathbf{y} &= \mathbf{y}_1 + \jmath\mathbf{y}_2 \\ \mathbf{A}^{\mathrm{H}} &= \mathbf{A}_1 + \jmath\mathbf{A}_2 \end{aligned}$$

Then, after some straighforward calculations, we can show that:

$$\Re\left(z^*\mathbf{x}^{\mathrm{H}}\mathbf{A}^{\mathrm{H}}\mathbf{y}\right) = \mathbf{x}_r^{\mathrm{T}}\mathbf{A}_z\mathbf{y}_r \tag{A.14}$$

where,

$$\begin{aligned} \mathbf{x}_r &= \begin{bmatrix}\mathbf{x}_1^{\mathrm{T}} & \mathbf{x}_2^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \\ \mathbf{y}_r &= \begin{bmatrix}\mathbf{y}_1^{\mathrm{T}} & \mathbf{y}_2^{\mathrm{T}}\end{bmatrix}^{\mathrm{T}} \\ \mathbf{A}_z &= \begin{bmatrix} z_1\mathbf{A}_1 - z_2\mathbf{A}_2 & -z_1\mathbf{A}_2 - z_2\mathbf{A}_1 \\ -z_1\mathbf{A}_2 - z_2\mathbf{A}_1 & -z_1\mathbf{A}_1 + z_2\mathbf{A}_2 \end{bmatrix}. \end{aligned}$$

Besides, one can easily prove that:

$$\mathrm{Tr}\left(\mathbf{A}_z^{\mathrm{T}}\mathbf{A}_z\right) = 2(z_1^2+z_2)\left(\mathrm{Tr}(\mathbf{A}_1\mathbf{A}_1^{\mathrm{T}}) + \mathrm{Tr}\left(\mathbf{A}_2\mathbf{A}_2^{\mathrm{T}}\right)\right) = 2(z_1^2+z_2^2)\mathrm{Tr}\left(\mathbf{A}^{\mathrm{H}}\mathbf{A}\right) \tag{A.15}$$

Plugging (A.14) into (A.13), we get:

$$\Phi(z) = \mathbb{E}\left[\exp\left(\frac{\jmath\sqrt{2(z_1^2+z_2^2)}}{\sqrt{\mathrm{Tr}(\mathbf{A}_z\mathbf{A}_z^{\mathrm{T}})}}\left(\mathbf{x}_r^{\mathrm{T}}\mathbf{A}_z\mathbf{y}_r\right)\right)\right] \tag{A.16}$$

Let $\tilde{\mathbf{a}}_{i,z}$ denotes the ith row of $\mathbf{A}_z$. Then the random variable $\frac{2}{\mathrm{Tr}(\mathbf{A}_z\mathbf{A}_z^{\mathrm{T}})}\mathbf{x}_r^{\mathrm{T}}\mathbf{A}_z\mathbf{y}_r$ can be decomposed into the sum of $2N$ random variables as follows:

$$\frac{2}{\sqrt{\mathrm{Tr}(\mathbf{A}_z\mathbf{A}_z^{\mathrm{T}})}}\mathbf{x}_r^{\mathrm{T}}\mathbf{A}_z\mathbf{y}_r = \frac{2}{\sqrt{\mathrm{Tr}(\mathbf{A}_z\mathbf{A}_z^{\mathrm{T}})}}\sum_{i=1}^{2N} x_{r,i}\tilde{\mathbf{a}}_{i,z}\mathbf{y}_r$$

where $x_{r,i}$ denotes the ith element of vector $\mathbf{x}_r$.

Consider the increasing filtration $\mathcal{F}_i = \sigma\left(\mathbf{y}_r, x_{r,1}, \cdots, x_{r,i}\right), i \in \{1, \cdots, N\}$. Denote $\mathbb{E}_i$ the conditional expectation with respect to $\mathcal{F}_i$, that is, $\mathbb{E}_i = \mathbb{E}\left[.|\mathcal{F}_i\right]$. By convention, $\mathcal{F}_0$ is the trivial $\sigma-$field, i.e $\mathbb{E}_0 = \mathbb{E}$.

Let $\xi_i = \frac{2}{\sqrt{\mathrm{Tr}(\mathbf{A}_z\mathbf{A}_z^{\mathrm{T}})}}x_{r,i}\tilde{\mathbf{a}}_{i,z}\mathbf{y}_r$. Then, it is obvious that $(\xi_i)_{1\leq i\leq N}$ is adapted to the filtration $(\mathcal{F}_i)_{1\leq i\leq N}$. Moreover, since, $\mathbb{E}_{i-1}(\xi_i) = 0$, ξ_i is a difference martingale sequence.

To prove the CLT, we need to check the three conditions in theorem A.4: Let us prove that $\sum_{i=1}^{2N}\xi_i^2 \to 1$ in probability. We have:

$$\sum_{i=1}^{2N}\xi_i^2 = \frac{4}{\mathrm{Tr}(\mathbf{A}_z\mathbf{A}_z^{\mathrm{T}})}\mathbf{y}_r^{\mathrm{T}}\left(\sum_{i=1}^{2N}\tilde{\mathbf{a}}_{i,z}^{\mathrm{T}}\mathbf{a}_{i,z}x_{r,i}^2\right)\mathbf{y}_r$$

Since $x_{r,i}$ are bounded random variables, the spectral norm of $\sum_{i=1}^{2N}\tilde{\mathbf{a}}_{i,z}^{\mathrm{T}}\mathbf{a}_{i,z}x_{r,i}^2$ is bounded. Moreover, one can easily check that $\mathrm{Tr}(\mathbf{A}_z\mathbf{A}_z^{\mathrm{T}})$ is of order N. Therefore, using corollary A.1, we get:

$$\sum_{i=1}^{2N}\xi_i^2 - \frac{2}{\mathrm{Tr}(\mathbf{A}_z\mathbf{A}_z^{\mathrm{T}})}\left(\tilde{\mathbf{a}}_{i,z}\tilde{\mathbf{a}}_{i,z}^{\mathrm{T}}x_{r,i}^2\right) \xrightarrow[K\to\infty]{} 0 \text{ almost surely.}$$

Using theoremA.2, we can prove that:

$$\frac{2}{\mathrm{Tr}(\mathbf{A}_z\mathbf{A}_z^{\mathrm{T}})}\sum_{i=1}^{2N}\left(\tilde{\mathbf{a}}_{i,z}\tilde{\mathbf{a}}_{i,z}^{\mathrm{T}}(x_{r,i}^2 - \frac{1}{2})\right) \xrightarrow[K\to\infty]{} 0\text{almost surely.}$$

Since $\sum_{i=1}^{2N}\tilde{\mathbf{a}}_{i,z}\tilde{\mathbf{a}}_{i,z}^{\mathrm{T}} = \mathrm{Tr}(\mathbf{A}_z\mathbf{A}_z^{\mathrm{T}})$, we have:

$$\sum_{i=1}^{2N}\xi_i^2 - 1 \xrightarrow[K\to\infty]{} 0 \text{ almost surely.}$$

It remains now to show that $\max_j|\xi_{n,j}|$ converges to zero in probability. For that, we will make use of Hoeffding inequality which can be stated as follows:

Theorem A.5. *Hoeffding inequality Let $X_1, \cdots, X_n$ be n independent random variables with zero mean. Assume that the X_i are almost surely bounded, that there exists b and a such that*

$$\mathrm{P}(X_i \in [a, b]) = 1.$$

Then, for the sum of these variables

$$S = \sum_{i=1}^{n} X_i$$

we have the inequality:

$$\mathrm{P}\left(|S| \geq t\right) \leq 2\exp\left(-\frac{2t^2}{n(b-a)^2}\right)$$

Since the entries of $\tilde{\mathbf{a}}_{i,z}$ are of order $\frac{1}{\sqrt{N}}$ and that of $\mathbf{y}_r$ and $\mathbf{x}_r$ are bounded, $\frac{4}{\sqrt{\mathrm{Tr}(\mathbf{A}_z\mathbf{A}_z^{\mathrm{T}})}}\tilde{\mathbf{a}}_{i,j,s}\mathbf{y}_{j,r}$ is of order $\mathcal{O}(\frac{1}{N})$, $\tilde{\mathbf{a}}_{i,j,s}$ and $\mathbf{y}_{j,r}$ being the jth entries of $\tilde{\mathbf{a}}_{i,z}$ and $\mathbf{y}_r$. Therefore, there exists a constant C such that:

$$|\frac{4}{\sqrt{\mathrm{Tr}(\mathbf{A}_z\mathbf{A}_z^{\mathrm{T}})}}\tilde{\mathbf{a}}_{i,j,s}\mathbf{y}_{j,r}| \leq \frac{C}{N}$$

Applying Hoeffding inequality, we get that:

$$\mathbb{P}\left(|\xi_j| > t\right) \leq 2\exp\left(-\frac{2Nt^2}{8C^2}\right)$$

Using the union bound inequality, we deduce, that,

$$\mathbb{P}\left(\max_j |\xi_j| > t\right) \leq 4N\exp\left(-\frac{2Nt^2}{8C^2}\right) \xrightarrow[N\to\infty]{} 0$$

To conclude about the CLT, we need to verify that $\sup_n \mathbb{E}\max_j \xi_j^2 \leq \infty$. This is true, since, all the variables are bounded. This shows that, under the assumption of our theorem, we have

$$\frac{2}{\sqrt{\mathrm{Tr}(\mathbf{A}_z\mathbf{A}_z^{\mathrm{T}})}}\mathbf{x}_r^{\mathrm{T}}\mathbf{A}_z\mathbf{y}_r \xrightarrow[K\to\infty]{} \mathcal{N}(0,1).$$

The characteristic function (A.16) converges to:

$$\Phi(z) - \exp(-\frac{|z|^2}{4}) \xrightarrow[K\to\infty]{} 0$$

□

We are now in position to determine the asymptotic behavior of $M_{N,1,1}$. Applying theorem A.4, we can prove that:

$$\mathbb{E}\left[j\Re(z^* M_{N,1,1})\right] \xrightarrow[K\to\infty]{} \exp(-\frac{c_1\sigma_{w_c}^4}{4\sigma_P^2})$$

For our particular choice of the pilot matrix $\mathbf{P}$, we can show that $M_{N,1,2}$ involves the sum of $\frac{N}{K} = \frac{1}{c_1}$ random entries of $c_1\tilde{\mathbf{w}}_i$, therefore, in the asymptotic regime $M_{N,1}$ behaves in the asymptotic regime as a Gaussian mixture random variable, i.e:

$$\mathbb{E}\left[\exp\left(\jmath\Re\left(z^* M_{N,1}\right)\right)\right] - \sum_{i=1}^{\mathcal{Q}} p_i \exp(\jmath\Re(z^*\alpha_i)) \exp\left(-\frac{c_1|z|^2\sigma_{w_c}^4}{4\sigma_P^2}\right) \xrightarrow[K\to\infty]{} 0$$

where $\mathcal{Q}$ is the cardinal of the set of all possible values of $\overline{\mathbf{W}}_{i,k} = c_1 \sum_{k=1}^{\frac{1}{c_1}} \mathbf{W}_{i,k}$, and p_i is the probability that $\overline{\mathbf{W}}_{i,k}$ takes the value α_i.

Using (A.9),(A.10) and (A.11) , we conclude that:

$$\mathbb{E}\left[\exp\left(\jmath\Re\left(z^*\left[\Delta\mathbf{W}_c\right]_{i,j}\right)\right)\right] - \sum_{i=1}^{\mathcal{Q}} p_i \exp(\jmath\Re(z^*(\alpha_i+\theta)) \exp\left(-\frac{c_1|z|^2\sigma_{w_c}^4}{4\sigma_P^2}\right) \exp(-|z|^2\frac{\tilde{\sigma}_M^2+\tilde{\sigma}_N^2}{4}) \xrightarrow[K\to\infty]{} 0$$

Substituting $\tilde{\sigma}_N^2$ and $\tilde{\sigma}_M^2$ by their expression, we get after simplification:

$$\mathbb{E}\left[\exp\left(\jmath\Re\left(z^*\left[\Delta\mathbf{W}_c\right]_{i,j}\right)\right)\right] - \sum_{i=1}^{\mathcal{Q}} p_i \exp(\jmath\Re(z^*(\alpha_i+\theta))) \exp\left(-\frac{|z|^2\sigma_{w_c}^2\left(\tau_c + |z|^2\delta_c\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}\right)}{4}\right)$$

where

$$\delta_c = (1-c_1)\frac{\sigma_v^2}{\sigma_{w_c}^2} + \frac{c_1\sigma_v^2}{\sigma_P^2} + \frac{c_1\sigma_v^4(c_2+1)}{(c_2-1)\sigma_P^2\sigma_{w_c}^2}$$
$$\tau_c = \frac{c_1\sigma_v^2}{\sigma_P^2(c_2-1)} + \frac{c_1\sigma_{w_c}^2}{\sigma_P^2}.$$

A.3 Proof of theorem 2.3

For the DDST scheme, the post-processing noise matrix $\Delta\mathbf{W}_d$ is given by:

$$\begin{aligned}\Delta\mathbf{W}_d &= -\mathbf{W}\mathbf{J} - \mathbf{H}^{\#}\Delta\mathbf{H}_d\mathbf{W}\left(\mathbf{I}_N - \mathbf{J}\right) + \left(\mathbf{H}^{\#} - \mathbf{H}^{\#}\Delta\mathbf{H}_d\mathbf{H}^{\#}\right)\mathbf{V}\left(\mathbf{I}_N - \mathbf{J}\right)\\ &\quad + \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\Delta\mathbf{H}_d^{\mathrm{H}}\mathbf{\Pi}\mathbf{V}\left(\mathbf{I}_N - \mathbf{J}\right)\\ &= -\mathbf{W}\mathbf{J} - \mathbf{H}^{\#}\Delta\mathbf{H}_d\mathbf{W}\left(\mathbf{I}_N - \mathbf{J}\right) + \mathbf{H}^{\#}\mathbf{V}\left(\mathbf{I}_N - \mathbf{J}\right) - \mathbf{H}^{\#}\Delta\mathbf{H}_d\mathbf{H}^{\#}\mathbf{V}\left(\mathbf{I}_N - \mathbf{J}\right)\\ &\quad + \left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\Delta\mathbf{H}_d^{\mathrm{H}}\mathbf{\Pi}\mathbf{V}\left(\mathbf{I}_N - \mathbf{J}\right).\end{aligned}$$

Hence,

$$\begin{aligned}\left(\Delta\mathbf{W}_d\right)_{i,j} &= -\tilde{\mathbf{w}}_i\mathbf{J}_j - \mathbf{h}_i^{\#}\mathbf{V}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{W}\left(\mathbf{e}_j - \mathbf{J}_j\right) + \mathbf{h}_i^{\#}\mathbf{V}\left(\mathbf{e}_j - \mathbf{J}_j\right) - \mathbf{h}_i^{\#}\mathbf{V}\mathbf{P}^{\mathrm{H}}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{H}^{\#}\mathbf{V}\left(\mathbf{e}_j - \mathbf{J}_j\right)\\ &\quad + \widetilde{\mathbf{h}}_i\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}\mathbf{V}^{\mathrm{H}}\mathbf{\Pi}\mathbf{V}\left(\mathbf{e}_j - \mathbf{J}_j\right)\end{aligned}$$

where $\mathbf{e}_j$ and $\mathbf{J}_j$ denotes the jth columns of $\mathbf{I}_N$ and $\mathbf{J}$, respectively and $\tilde{\mathbf{w}}_i$ denotes the ith row of the matrix $\mathbf{W}$.

Let $\mathbf{v}_1 = \mathbf{V}\left(\mathbf{e}_j - \mathbf{J}_j\right)$, and $\mathbf{v}_2 = \mathrm{vec}(\mathbf{V}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}^{\mathrm{H}})$

The vector $[\mathbf{v}_1^{\mathrm{T}}, \mathbf{v}_2^{\mathrm{T}}]^{\mathrm{T}}$ is a Gaussian vector. Since $\mathbb{E}[\mathbf{v}_1\mathbf{v}_2^{\mathrm{H}}] = 0$, we conclude that $\mathbf{v}_1$ and $\mathbf{v}_2$ are independent. Then $\mathbf{v}_1$ and $\mathbf{V}_2 = \mathbf{V}\left(\mathbf{P}\mathbf{P}^{\mathrm{H}}\right)^{-1}\mathbf{P}^{\mathrm{H}}$ are also independent. Moreover, $\mathbb{E}[\mathbf{v}_1\mathbf{v}_1^{\mathrm{H}}] = \sigma_v^2\left(1-\frac{K}{N}\right)\mathbf{I}_N$.

Conditioning on $\mathbf{V}_2$, $\mathbf{H}$ and $\mathbf{W}$, $(\Delta\mathbf{W}_d)_{i,j}$ is a Gaussian random variable with mean equal to $-\tilde{\mathbf{w}}_i\mathbf{J}_j - \mathbf{h}_i^{\#}\mathbf{V}_2\mathbf{W}\left(\mathbf{e}_j - \mathbf{J}_j\right)$ and variance $\sigma^2_{w_d,N}$ equal to:

$$\begin{aligned}
\sigma^2_{w_d,N} &= \mathbb{E}\left[\left(\mathbf{h}_i^{\#} - \mathbf{h}_i^{\#}\mathbf{V}_2\mathbf{H}^{\#} + \widetilde{\mathbf{h}}_i\mathbf{V}_2^{\mathrm{H}}\mathbf{\Pi}\right)\mathbf{v}_1\mathbf{v}_1^{\mathrm{H}}\left(\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}} - \left(\mathbf{H}^{\#}\right)^{\mathrm{H}}\mathbf{V}_2^{\mathrm{H}}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}} + \mathbf{\Pi}\mathbf{V}_2\widetilde{\mathbf{h}}_i^{\mathrm{H}}\right)|\mathbf{V}_2\right] \\
&= \mathbb{E}\left[\mathbf{h}_i^{\#}\mathbf{v}_1\mathbf{v}_1^{\mathrm{H}}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\right] + \mathbb{E}\left[\mathbf{h}_i^{\#}\mathbf{V}_2\mathbf{H}^{\#}\mathbf{v}_1\mathbf{v}_1^{\mathrm{H}}\left(\mathbf{H}^{\#}\right)^{\mathrm{H}}\mathbf{V}_2^{\mathrm{H}}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\right] - 2\mathbb{E}\left[\Re\left(\mathbf{h}_i^{\#}\mathbf{V}_2\mathbf{H}^{\#}\mathbf{v}_1\mathbf{v}_1^{\mathrm{H}}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\right)\right] \\
&+ \sigma_v^2(1-\frac{K}{N})\widetilde{\mathbf{h}}_i\mathbf{V}_2^{\mathrm{H}}\mathbf{\Pi}\mathbf{V}_2(\widetilde{\mathbf{h}}_i)^{\mathrm{H}} \\
&= (1-\frac{K}{N})\sigma_v^2\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i} + \sigma_v^2(1-\frac{K}{N})\mathbf{h}_i^{\#}\mathbf{V}_2\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\mathbf{V}_2^{\mathrm{H}}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}} - 2(1-\frac{K}{N})\sigma_v^2\Re\left(\mathbf{h}_i^{\#}\mathbf{V}_2\mathbf{H}^{\#}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\right) \\
&+ \sigma_v^2(1-\frac{K}{N})\widetilde{\mathbf{h}}_i\mathbf{V}_2^{\mathrm{H}}\mathbf{\Pi}\mathbf{V}_2(\widetilde{\mathbf{h}}_i)^{\mathrm{H}}
\end{aligned}$$

Using the same techniques as before, it can be proved that:

$$(1-\frac{K}{N})\sigma_v^2\mathbf{h}_i^{\#}\mathbf{V}_2\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\mathbf{V}_2^{\mathrm{H}}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}} - \frac{c_1(1-c_1)\sigma_v^4}{(c_2-1)\sigma_P^2}\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i} \to 0 \text{ almost surely.}$$

and also that,

$$\Re\left(\mathbf{h}_i^{\#}\mathbf{V}_2\mathbf{H}^{\#}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}\right) \longrightarrow 0 \text{ almost surely.}$$

On the other hand, we have:

$$\sigma_v^2(1-c_1)\widetilde{\mathbf{h}}_i\mathbf{V}_2^{\mathrm{H}}\mathbf{\Pi}\mathbf{V}_2\left(\widetilde{\mathbf{h}}_i\right)^{\mathrm{H}} - \frac{c_1\sigma_v^4(1-c_1)(M-K)}{N\sigma_P^2}\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-2}\right]_{i,i} \to 0 \text{ almost surely.}$$

Since $\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-2}\right] - \frac{c_2}{c_2-1}\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]^2_{i,i} \to 0$ by lemma A.2, we get that:

$$\sigma_v^2(1-c_1)\widetilde{\mathbf{h}}_i\mathbf{V}_2^{\mathrm{H}}\mathbf{\Pi}\mathbf{V}_2\left(\widetilde{\mathbf{h}}_i\right)^{\mathrm{H}} - \frac{\sigma_v^4(1-c_1)c_1c_2}{(c_2-1)}\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i} \to 0.$$

Therefore,

$$\sigma^2_{w_d,N} - \tilde{\sigma}^2_{w_d,N} \longrightarrow 0 \text{ almost surely}$$

where,

$$\tilde{\sigma}^2_{w_d,N} = \left(\sigma_v^2(1-c_1) + \frac{c_1(c_2+1)(1-c_1)\sigma_v^4}{(c_2-1)\sigma_{P_d}^2}\right)\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}.$$

Consequently,

$$\begin{aligned}
\mathbb{E}\left[\exp\left(\jmath\Re\left(z^*\left(\Delta\mathbf{W}\right)_{i,j}\right)\right)|\mathbf{H},\mathbf{W},\mathbf{V}_2\right] &= \mathbb{E}\left[\exp\left(-\jmath\Re\left(z^*\tilde{\mathbf{w}}_i\mathbf{J}_j + z^*\mathbf{h}_i^{\#}\mathbf{V}_2\mathbf{W}\left(\mathbf{e}_j - \mathbf{J}_j\right)\right)\right)|\mathbf{W},\mathbf{v}_2\right] \\
&\times \exp\left(-\frac{1}{4}|z|^2\tilde{\sigma}^2_{w_d,N}\right).
\end{aligned}$$

Conditioning on $\mathbf{W}$ and $\mathbf{H}$, $\tilde{\mathbf{w}}_i\mathbf{J}_j + \mathbf{h}_i^{\#}\mathbf{V}_2\mathbf{W}\left(\mathbf{e}_j - \mathbf{J}_j\right)$ is a Gaussian random variable with mean equal to $\tilde{\mathbf{w}}_i\mathbf{J}_j$ and variance $\sigma^2_{w_m,N}$ given by:

$$\begin{aligned}
\sigma^2_{m_d,N} &= \mathbb{E}\left[\mathbf{h}_i^{\#}\mathbf{V}_2\mathbf{W}\left(\mathbf{e}_j - \mathbf{J}_j\right)\left(\mathbf{e}_j^{\mathrm{H}} - \mathbf{J}_j^{\mathrm{H}}\right)\mathbf{W}^{\mathrm{H}}\mathbf{V}_2^{\mathrm{H}}\left(\mathbf{h}_i^{\#}\right)^{\mathrm{H}}|\mathbf{W},\mathbf{H}\right] \\
&= \frac{\sigma_v^2}{N\sigma_{P_d}^2}\left[\left(\mathbf{H}^{\mathrm{H}}\mathbf{H}\right)^{-1}\right]_{i,i}\left(\mathbf{e}_j^{\mathrm{H}} - \mathbf{J}_j^{\mathrm{H}}\right)\mathbf{W}\mathbf{W}^{\mathrm{H}}\left(\mathbf{e}_j - \mathbf{J}_j\right).
\end{aligned}$$

Using corollary A.1, we can easily prove that:

$$\sigma^2_{m_d,N} - \tilde{\sigma}^2_{m_d,N} \longrightarrow 0 \text{ almost surely,}$$

where

$$\tilde{\sigma}^2_{m_d,N} = \frac{(1-c_1)\sigma^2_{w_d}\sigma^2_v}{\sigma^2_{P_d}} \left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}.$$

Conditioning only on $\mathbf{H}$, the conditional characteristic function satisfies:

$$\mathbb{E}\left[\exp\left(\jmath\Re\left(z^*\left(\Delta\mathbf{W}_d\right)_{i,j}\right)\right)|\mathbf{H}\right] - \mathbb{E}\left[\exp\left(-\jmath\Re\left(z^*\tilde{\mathbf{w}}_i\mathbf{J}_j\right)\right)\right]\exp\left(-\frac{1}{4}|z|^2\left(\tilde{\sigma}^2_{w_d,N}+\tilde{\sigma}^2_{m_d,N}\right)\right) \longrightarrow 0.$$

Giving the structure of the matrix $\mathbf{J}$, $\tilde{\mathbf{w}}_i\mathbf{J}_j$ involves the average of $\frac{1}{c_1}$ symmetric independent and identically distributed discrete random variables, and therefore,

$$\mathbb{E}\left[\exp\left(-\jmath\Re\left(z^*\tilde{\mathbf{w}}_i\right)\right)\right] = \sum_{i=1}^{\mathcal{Q}} p_i \exp\left(\jmath\Re\left(z^*\alpha_i\right)\right)$$

where $\mathcal{Q}$ is the set of all possible values of $\overline{\mathbf{W}}_{i,k} = c_1\sum_{i=1}^{\frac{1}{c_1}}\mathbf{W}_{i,k}$ and p_i is the probability that $\overline{\mathbf{W}}_{i,k}$ takes the value α_i. Consequently;

$$\mathbb{E}\left[\exp\left(\jmath\Re\left(z^*\left(\Delta\mathbf{W}_d\right)_{i,j}\right)\right)|\mathbf{H}\right] = \sum_{i=1}^{\mathcal{Q}} p_i \exp\left(\jmath\Re\left(z^*\alpha_i\right)\right)\exp\left(-\frac{1}{4}|z|^2\left(\tilde{\sigma}^2_{m_d,N}+\sigma^2_{w_d,N}\right)\right).$$

We conclude the proof by noting that

$$\tilde{\sigma}^2_{m_d,N} + \sigma^2_{w_d,N} = \sigma^2_{w_d}\left[(\mathbf{H}^{\mathrm{H}}\mathbf{H})^{-1}\right]_{i,i}\delta_d.$$

Part II

Performance analysis for multi-user receivers using random matrix theory

Introduction

In almost current digital communication applications, the channel is common to several transmitters, which have to comply with a predefined access protocol. A non comprehensive list of examples includes mobile phones transmissions to a base station, communication between an access point and terminals, or ground stations communicating with a satellite. To ensure the system flexibility with respect to the demand of users and to avoid outages, carrying out performance analysis that evaluate certain performance metric is often of great interest.

The early results on performance analysis of multi-user systems were conducted for academic contexts, where it is possible to get closed-form derivations. However, for practical situations taking into account general correlated channels, it is often impossible to get simple closed form-expressions. For this reason, the evaluation of the system performance has been often conducted by using massive Monte-Carlo simulations, which may not allow us to understand the effect of system parameters on the overall performance.

An other approach to evaluate the system performance is to consider the asymptotic case, when the system parameters grow to infinity with constant ratios. In such case, the mutual information has been shown to converge to a Gaussian random variable for the so called Kronecker correlated channels, [53, 97], for generally spatially correlated channels [76] and for general variance profile channels [55].

In this part, we will provide a framework for the evaluation of system performance for multi-user MIMO systems in the asymptotic regime. More explicitly, we will study the asymptotic behavior (asymptotic approximation and distribution) of the SNR at the output of the Wiener LMMSE receiver, under the assumption of perfect channel knowledge. In our study, we consider many variance profiles, where the entries of the channel matrix are assumed to be non i.i.d. This makes our results apply for a large variety of systems ranging from MIMO systems to CDMA uplink and downlink based systems. Unfortunately, this asymptotic analysis was found to fail to accurately estimate some performance metrics such as the BER and the outage probability. Indeed, the Gaussian distribution is unsuitable to approximate these metrics, since it allows negative values and its third central moment is zero, where the SNR is always positive and has nonzero third moment. To accurately approximate the BER and the outage probability, we propose to use the Generalized Gamma distribution instead of the Gaussian distribution.

This part is organized as follows: First, we review in chapter 3 recent mathematical results in the field of random matrix theory. Then, we present in chapter 4 the asymptotic results pertaining the fluctuations of the SINR. Finally, we devote chapter 5 to explain our method based on the Generalized Gamma approximation.

Notations: Along this part, except when mentioning, N will denote the size of the received vector while K will denote the number of interfering users. Moreover, the channel matrix will be denoted by $\mathbf{\Sigma}$.

Chapter 3

A brief Overview of random matrix theory

This chapter presents a general overview of recent developments of random matrix theory. In particular, It provides preliminary mathematical results that pave the way for the understanding of the two forthcoming chapters. For the sake of illustration, numerical results are given and used to draw some conclusions about the validity of the mathematical results.

3.1 Random matrix theory developments

Random matrix theory has matured into a field of applications in many disciplines of physics, engineering and mathematics. It was first introduced in 1928 by Wishart with the work on fixed-sizes matrices with Gaussian entries. The first asymptotic results on the limiting eigenvalue distributions of symmetric matrices were obtained by Wigner, and had found applications in nuclear physics. Since then, a great number of physicians have been interested in asymptotic spectral properties of random matrices, and very recently, there is an increasing interest in random matrix theory for the field of digital communications. Hereafter, we define some matrices that have retained a great attention in random matrix theory and we provide the obtained results that discussed their asymptotic behavior.

3.1.1 Resolvent and Gram matrices

Resolvent and Gram matrices play an important role in random matrix theory. If $\mathbf{\Sigma} = (\Sigma_{n,k})_{n=1,k=1}^{N,K}$ denotes a $N \times K$ random matrix then, $\mathbf{\Sigma}\mathbf{\Sigma}^{\mathrm{H}}$ and $\mathbf{Q}(z) = \left(\mathbf{\Sigma}\mathbf{\Sigma}^{\mathrm{H}} - z\rho\mathbf{I}_N\right)^{-1}$ are said to be the Gram and resolvent matrices corresponding to $\mathbf{\Sigma}$.

Since the size of the considered matrices is allowed to grow to infinity, the variance of the entries of $\mathbf{\Sigma}$ is often normalized by $\frac{1}{K}$ so as to ensure a finite energy per row and column of $\mathbf{\Sigma}$. In other words, it is often assumed that:

$$\Sigma_{n,k} = \frac{\sigma_{n,k} W_{n,k}}{\sqrt{K}}$$

where $(W_{n,k})_{n=1,k=1}^{N,K}$ is a sequence of i.i.d. random variables with zero mean and variance 1, and $(\sigma_{n,k}, 1 \leq n \leq N, 1 \leq k \leq K)$ is a bounded sequence of positive real numbers referred to as variance profile.

According to the structure of $\sigma_{n,k}$, three kind of variance profiles can be distinguished:

- The variance profile is said to be separable if there exist positive sequences d_n and $\tilde{d}_k$ such that:
$$\sigma_{n,k}^2 = d_n \tilde{d}_k.$$
- The variance profile is said to be limit if there exists a continuous function $\sigma^2(x, y)$ defined on $[0, 1]^2$ such that:
$$\sigma_{n,k}^2 = \sigma^2(n/N, k/K)$$
- If the variance profile is not separable, it is said to be general.

3.1.2 Asymptotic results for resolvent matrices

While results for the statistics of the resolvent matrix with fixed dimensions are often cumbersome and provide little insights, a number of powerful and appealing theorems in the asymptotic case offer simple expressions that approximate the spectral properties by deterministic functionals. These results are of great interest in the digital communication field, since the resolvent matrix is very often encountered in performance metric expressions. For instance, in MIMO systems, the Shannon's mutual information is, in most cases, expressed as:

$$I(\rho) = \mathbb{E} \log \det(\mathbf{I}_N + \rho \mathbf{\Sigma}\mathbf{\Sigma}^{\mathrm{H}})$$

where ρ represents the SNR and $\mathbf{\Sigma}$ the channel matrix. The expression of the mutual information can be further worked out to yield:

$$I(\rho) = \sum_{n=1}^{N} \mathbb{E} \log(1 + \rho \lambda_i)$$

where λ_i are the eigenvalues of the Gram matrix $\mathbf{\Sigma}\mathbf{\Sigma}^{\mathrm{H}}$. The resolvent matrix appears also in the MSE expression for the LMMSE receiver, given by, [105, equation (1.7)]:

$$\mathrm{MSE} = \frac{1}{K} \mathrm{Tr}(\mathbf{I}_N + \rho \mathbf{H}\mathbf{H}^{\mathrm{H}})^{-1}$$

which can be further expressed as:

$$\mathrm{MSE} = \frac{1}{K} \sum_{i=1}^{K} \frac{1}{1 + \rho \lambda_i(\mathbf{H}\mathbf{H}^{\mathrm{H}})}$$

3.1.2.1 Limit spectral measure

From the examples mentioned above, it emerges clearly that knowing asymptotic properties on the spectrum of the Gram matrix $\mathbf{H}\mathbf{H}^{\mathrm{H}}$, may help the understanding of the asymptotic behavior of certain performance metrics.

For that, we need to define an empirical function of eigenvalues of the Gram matrix and study its asymptotic behavior. The random quantity that is often considered is called the empirical spectral measure and is defined on the Borel sets of $\mathbb{R}+$ as:

$$\mu_n(A) = \frac{1}{K} \sum_{i=1}^{K} \delta_{\lambda_i}(A)$$

where $\delta_{\lambda_i}(A) = 1$ if $\lambda_i \in A$ and $\delta_{\lambda_i}(A) = 0$ if $\lambda_i \notin A$.

Asymptotic limits of the empirical spectral measure have been extensively investigated in many previous works. It has been shown that when the entries of $\mathbf{\Sigma}$ are zero mean with limit variance profile, the empirical distribution converges weakly to a limiting spectral measure μ that can be characterized via a functional equation, [45]. On the other hand, in case the entries of $\mathbf{\Sigma}$ are non-centered, the existence of a limiting spectral measure of the Gram matrix $\mathbf{\Sigma}\mathbf{\Sigma}^{\mathrm{H}}$ is not guaranteed except in some specific cases, [54].

3.1.2.2 Deterministic equivalents

Since the limiting spectral measure of Gram matrices does not always exist, [54] investigates on "non limit" deterministic equivalents that dictates the asymptotic behaviour of the spectral measure. Such deterministic equivalents are shown to be unique and to always exist for any variance profile, [54].

Hereafter, we recall their definition for general and separable variance profiles as well as some of their properties. But first, we shall introduce the following quantities:

Definition 3.1. *Let $(\sigma_{n,k}^2, 1 \leq n \leq N, 1 \leq k \leq K)$ be a general variance profile. Based on this variance profile, we define the following diagonal matrices:*

$$\begin{aligned} \mathbf{D}_k &= \mathrm{diag}\left(\sigma_{1k}^2, \cdots, \sigma_{Nk}^2\right), && 1 \leq k \leq K \\ \widetilde{\mathbf{D}}_n &= \mathrm{diag}\left(\sigma_{n1}^2, \cdots, \sigma_{nK}^2\right), && 1 \leq n \leq N \end{aligned}$$

In case the variance profile is separable, matrices $\mathbf{D}_k$ and $\widetilde{\mathbf{D}}_n$ are written as:

$$\mathbf{D}_k = \widetilde{d}_k \mathbf{D} \quad \text{and} \quad \widetilde{\mathbf{D}}_n = d_n \widetilde{\mathbf{D}}$$

where $\mathbf{D}$ and $\widetilde{\mathbf{D}}$ are the diagonal matrices:

$$\mathbf{D} = \mathrm{diag}(d_1, \ldots, d_N), \quad \widetilde{\mathbf{D}} = \mathrm{diag}(\tilde{d}_1, \ldots, \tilde{d}_K)\ . \tag{3.1}$$

Definition 3.2. *A complex function $t(z)$ belongs to class $\mathcal{S}$ if :*

- *$t(z)$ is analytical in the upper half $\mathbb{C}_+ = \{z \in \mathbb{C}; \Im(z) > 0\}$*
- *$t(z) \in \mathbb{C}_+$ for all $z \in \mathbb{C}_+$*
- *$\Im(z)|t(z)|$ is bounded over the upper half plane $\mathbb{C}_+$.*

We are now in position to introduce the deterministic equivalents for general and separable variance profiles : When the general variance profile is considered, the unique deterministic equivalents are the solution of $N + K$ system of equations:

Theorem 3.1. *[54, Theorem 2.4] Assume that the entries of the Gram matrix $\mathbf{\Sigma}$ are centered, then he system of $N + K$ functional equations:*

$$\begin{cases} t_{n,K}(z) &= \dfrac{-1}{z\left(1 + \frac{1}{K}\mathrm{Tr}(\widetilde{\mathbf{D}}_n(K)\widetilde{\mathbf{T}}_K(z))\right)}, \quad 1 \leq n \leq N \\ \tilde{t}_{k,K}(z) &= \dfrac{-1}{z\left(1 + \frac{1}{K}\mathrm{Tr}(\mathbf{D}_k(K)\mathbf{T}_K(z))\right)}, \quad 1 \leq k \leq K \end{cases} \tag{3.2}$$

where

$$\mathbf{T}_K(z) = \mathrm{diag}(t_{1,K}(z), \ldots, t_{N,K}(z)), \quad \widetilde{\mathbf{T}}_K(z) = \mathrm{diag}(\tilde{t}_{1,K}(z), \ldots, \tilde{t}_{K,K}(z))$$

admits a unique solution $(\mathbf{T}, \widetilde{\mathbf{T}})$ among the diagonal matrices for which the $t_{n,K}$'s and the $\tilde{t}_{k,K}$'s belong to class $\mathcal{S}$. Moreover, functions $t_{n,K}(z)$ and $\tilde{t}_{k,K}(z)$ admit an analytical continuation over $\mathbb{C} - \mathbb{R}_+$ which is real and positive for $z \in (-\infty, 0)$.

Moreover, $\mathbf{T}$ and $\widetilde{\mathbf{T}}$ can be computed numerically in an iterative way (Algorithm 1):

Algorithm 1 Iterative computation of $\mathbf{T}(z)$ and $\widetilde{\mathbf{T}}(z)$

$\mathbf{T} \leftarrow \mathbf{0}_{N\times N}$ // Initialization of $\mathbf{T}$
$\widetilde{\mathbf{T}} \leftarrow \mathbf{0}_{K\times K}$// Initialization of $\widetilde{\mathbf{T}}$
$\epsilon \leftarrow 10^{-3}$
repeat
  $\mathbf{T}^i \leftarrow \mathbf{T}$ // Previous value of $\mathbf{T}$
  $\widetilde{\mathbf{T}}^i \leftarrow \widetilde{\mathbf{T}}$ // Previous value of $\widetilde{\mathbf{T}}$
  for $n = 1$ **to** N **do**
    $\mathbf{T}(n,n) \leftarrow -\frac{1}{z\left(1+\frac{1}{K}\mathrm{Tr}\left(\widetilde{\mathbf{D}}_n\widetilde{\mathbf{T}}^i\right)\right)}$
  end for
  for $k = 1$ K **do**
    $\widetilde{\mathbf{T}}(k,k) \leftarrow -\frac{1}{z\left(1+\frac{1}{K}\mathrm{Tr}(\mathbf{D}_k\mathbf{T}^i)\right)}$
  end for
until $\frac{\|\mathbf{T}-\mathbf{T}^i\|}{\mathbf{T}^i} > \epsilon$

In [54], diagonal matrices $\mathbf{T}$ and $\widetilde{\mathbf{T}}$ are shown to reproduce the asymptotic behavior of resolvent matrices $\mathbf{Q}(z) = (\mathbf{\Sigma}\mathbf{\Sigma}^{\mathrm{H}} - z\mathbf{I}_N)$ and $\widetilde{\mathbf{Q}}(z) = (\mathbf{\Sigma}^{\mathrm{H}}\mathbf{\Sigma} - z\mathbf{I}_K)$. By asymptotic, we mean that both K and N go to $+\infty$, their ratio being bounded below and above as follows:

$$K \to +\infty, \quad \liminf \frac{K}{N} > 0 \quad \text{and} \quad \sup \frac{K}{N} < +\infty \ . \tag{3.3}$$

In the sequel, the notation "$K \to +\infty$" will refer to (3.3).

Under some mild assumptions, the quantities $\frac{1}{K}\mathrm{Tr}\mathbf{S}\left(\mathbf{Q}(z) - \mathbf{T}(z)\right)$ and $\frac{1}{K}\mathrm{Tr}\mathbf{S}\left(\widetilde{\mathbf{Q}}(z) - \widetilde{\mathbf{T}}(z)\right)$ converge almost surely to zero as $K \to +\infty$, where $\mathbf{S}$ is a test matrix with spectral bounded norm. More explicitly, this result can be stated as:

Theorem 3.2. *Let $\mathbf{\Sigma}$ a $N \times K$ random matrix, with entries verifying that:*

$$\Sigma_{n,k} = \frac{\sigma_{n,k}W_{n,k}}{\sqrt{K}}$$

where $W_{n,k}$ are i.i.d. complex random variables with zero mean, $\mathbb{E}W_{1,0}^2 = 0$ and $\mathbb{E}|W_{1,0}|^2 = 1$ and $\mathbb{E}|W_{1,0}|^8 < +\infty$. Assume also that the variance profile is bounded, i.e. there exists a real number $\sigma_{\max} < +\infty$ such that:

$$\sup_{K\geq 1} \max_{\substack{1\leq n\leq N\\ 0\leq k\leq K}} |\sigma_{nk}(K)| \leq \sigma_{\max} \ .$$

then, for every sequence $\mathbf{S}_K$ of $N \times N$ diagonal matrices and every sequence $\widetilde{\mathbf{S}}_K$ of $K \times K$ diagonal matrices with

$$\sup_K \max\left(\|\mathbf{S}_K\|, \|\widetilde{\mathbf{S}}_K\|\right) < +\infty \, ,$$

the following limits hold true almost surely:

$$\begin{aligned} \lim_{K\to+\infty} \frac{1}{K}\mathrm{Tr}\ \mathbf{S}_K\left(\mathbf{Q}_K(z) - \mathbf{T}_K(z)\right) &= 0, \qquad \forall z \in \mathbb{C} - \mathbb{R}_+, \\ \lim_{K\to+\infty} \frac{1}{K}\mathrm{Tr}\ \widetilde{\mathbf{S}}_K\left(\widetilde{\mathbf{Q}}_K(z) - \widetilde{\mathbf{T}}_K(z)\right) &= 0, \qquad \forall z \in \mathbb{C} - \mathbb{R}_+ \ . \end{aligned}$$

In case the variance profile is separable, the $N + K$ system of equations in theorem 3.1 simplifies to a system of two equations:

Theorem 3.3. *[54] Assume $\sigma_{nk}^2(K) = d_n(K)\tilde{d}_k(K)$. The system of two functional equations*

$$\begin{cases} \delta(z) &= \frac{1}{K}\mathrm{Tr}\left(\mathbf{D}\left(-z(\mathbf{I}_N + \tilde{\delta}(z)\mathbf{D})\right)^{-1}\right) \\ \tilde{\delta}(z) &= \frac{1}{K}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\left(-z(\mathbf{I}_K + \delta(z)\widetilde{\mathbf{D}})\right)^{-1}\right) \end{cases} \tag{3.4}$$

admits a unique solution $(\delta, \tilde{\delta}) \in \mathcal{S}^2$. Moreover, letting $z = -\rho \in (-\infty, 0)$, the system admits a unique point-wise solution $(\delta(-\rho), \tilde{\delta}(-\rho))$ such that $\delta(-\rho) > 0$, $\tilde{\delta}(-\rho) > 0$.

Moreover, in the separable variance profile case, the matrix functions $\mathbf{T}$ and $\widetilde{\mathbf{T}}$ defined by theorem 3.1 are given by

$$\mathbf{T} = -\frac{1}{z}(\mathbf{I} + \tilde{\delta}\mathbf{D})^{-1} \quad and \widetilde{\mathbf{T}} = -\frac{1}{z}(\mathbf{I} + \delta\widetilde{\mathbf{D}})^{-1}. \tag{3.5}$$

Similarly to the diagonal matrices $\mathbf{T}$ and $\widetilde{\mathbf{T}}$, δ and $\tilde{\delta}$ can be computed iteratively as the fixed point of a 2 dimensional system of equations (algorithm 2).

Algorithm 2 Iterative computation of δ and $\widetilde{\delta}$

$\delta \leftarrow 0$// Initialization of δ
$\widetilde{\delta} \leftarrow 0$// Initialization of $\widetilde{\delta}$
$\epsilon \leftarrow 10^{-3}$
repeat
  $\delta^i \leftarrow \delta$
  $\widetilde{\delta}^i \leftarrow \widetilde{\delta}$
  $\mathbf{T} \leftarrow -\frac{1}{z}(\mathbf{I}_N + \widetilde{\delta}^i\mathbf{D})^{-1}$
  $\widetilde{\mathbf{T}} \leftarrow -\frac{1}{z}(\mathbf{I}_K + \delta^i\widetilde{\mathbf{D}})^{-1}$
  $\delta \leftarrow \frac{1}{K}\mathrm{Tr}(\mathbf{D}\mathbf{T})$
  $\widetilde{\delta} \leftarrow \frac{1}{K}\mathrm{Tr}(\widetilde{\mathbf{D}}\widetilde{\mathbf{T}})$
until $\frac{|\delta - \delta^i|}{\delta^i} > \epsilon$

3.2 Numerical results

In this section, we study some numerical aspects regarding the number of required iterations for the computation of the deterministic equivalents, as well as their accuracy for finite size dimensions. We consider the separable case, with $\sigma_{i,j}^2 = \exp(-i/N)\exp(-j/K)$.

3.2.1 Number of iterations

The computation of the deterministic equivalents is in general performed by using the aforementioned algorithms algorithm. 1 and algorithm. 2 with the exception of some particular cases where closed-form expressions for the deterministic equivalents exist. In this section, we evaluate the number of required iterations in algorithm. 2. Fig. 3.1 plots $\frac{|\delta - \delta^i|}{\delta^i}$ for each iteration for $K = 16$, $N = 16$ and $\rho = 1$. We note that only few iterations are required to reach the convergence.

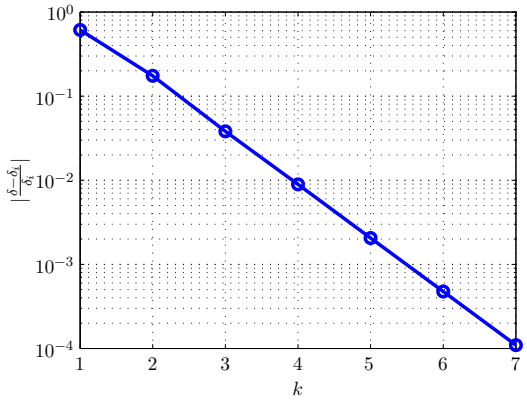
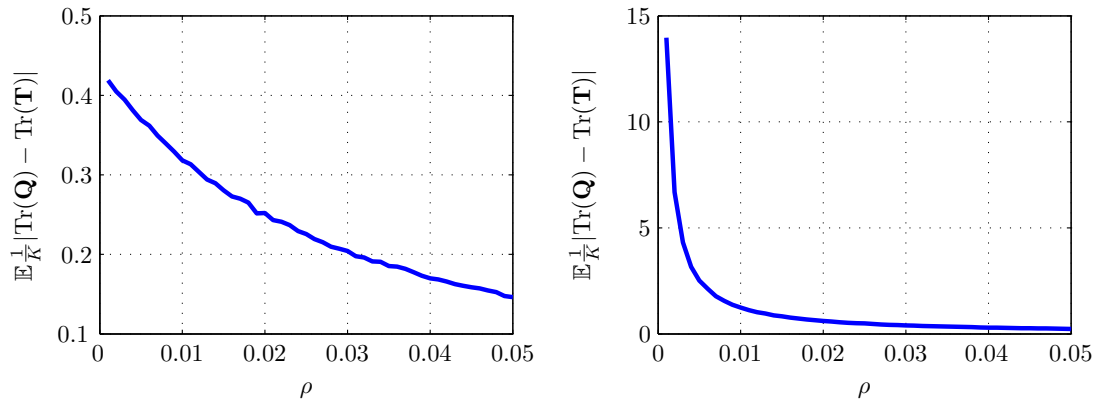


Figure 3.1: Required number of iterations

3.2.2 Accuracy of the deterministic equivalents

The almost sure convergence of $\frac{1}{K}\mathrm{Tr}(\mathbf{Q}(-\rho))$ to $\frac{1}{K}\mathrm{Tr}(\mathbf{T}(-\rho))$ stated in theorem 3.1 guarantees, that as long as the matrix dimensions N and K are large enough, the deterministic equivalent $\frac{1}{K}\mathrm{Tr}(\mathbf{T}(-\rho))$ reproduces accurately the asymptotic behaviour of the random quantity $\frac{1}{K}\mathrm{Tr}(\mathbf{Q}(-\rho))$. Nevertheless, it so happens in some cases that the values of N and K that yield a good accuracy, should reach higher values than those observed in usual cases. This is particularly the case when ρ is close to zero and $\frac{N}{K}$ approaches 1. Actually, in such case, the least eigenvalue of matrix $\rho\mathbf{I}_N + \mathbf{\Sigma}\mathbf{\Sigma}^{\mathrm{H}}$ is too small and thereby the largest eigenvalue of the resolvent matrix becomes too high. As a consequence, the error on the approximation of the trace of the resolvent matrix may be also high. A more rigorous discussion in a similar issue can be found in [35].

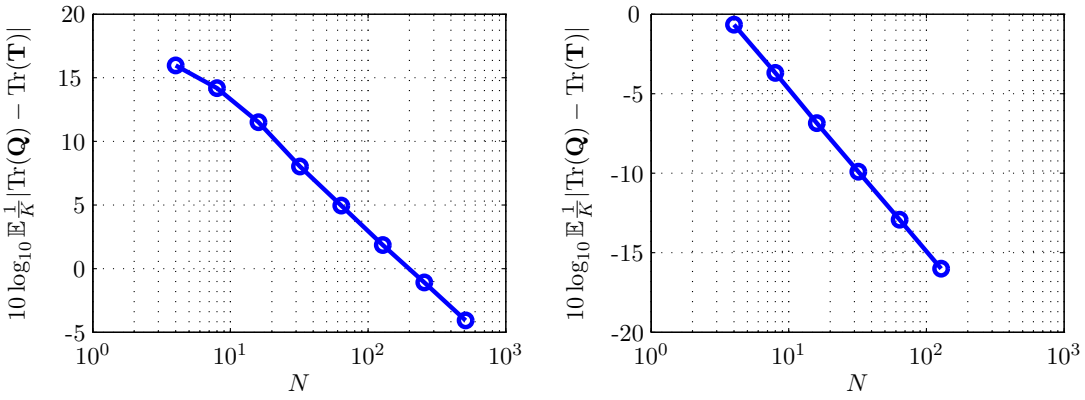


(a) $K = 16, N = 8$ (b) $K = 16, N = 16$

Figure 3.2: Accuracy of the deterministic equivalents

Fig. 3.2 plots the error $\frac{1}{K}|\mathrm{Tr}(\mathbf{Q}) - \mathrm{Tr}(\mathbf{T})|$, averaged over 10000 iterations, with respect to ρ for two scenarios corresponding respectively to $K = 16, N = 8$ and $K = N = 16$. We note that as ρ tends to zero, the scenario $K = N = 16$ exhibits higher errors. At the same level of error, values of K and N in the case $K = N$ need to be much higher than those in the case $K = 2N$.

Fig. 3.3 illustrates well this issue.



(a) $\rho = 0.001, K = N$ (b) $\rho = 0.001, K = 2N$

Figure 3.3: Accuracy of the deterministic equivalents for $K = N$ and $K = 2N$

3.3 Conclusion

This chapter provides a brief overview of recent developments in random matrix theory. In particular, new results about a class of deterministic equivalents approximating the resolvent matrix are presented, and numerical algorithms that allows their computation are described. It has been shown that only few number of iterations are required to get convergence, which is to be compared with the other alternative based on highly massive Monte Carlo iterations. Simulations proves also that the accuracy of the deterministic equivalents is almost guaranteed even for finite sizes, except in some special cases, in which it is required to get to higher system dimensions.

Chapter 4

Asymptotic performance of Linear receivers for multi-user systems

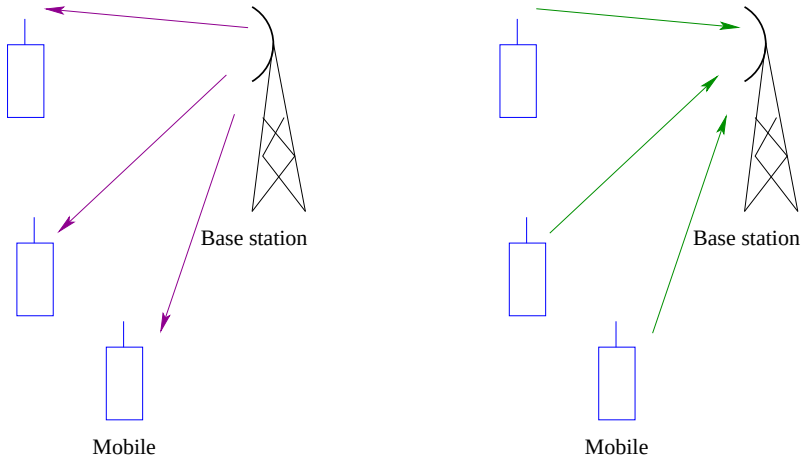
In the context of multi-user systems, the LMMSE receiver is often encountered. It is, indeed, the optimal linear receiver, in the sense that it achieves the lowest level of interference. It is thus of interest to analyse the SINR at its output, namely its first and second moments as well as its asymptotic distribution. While the first moment has been extensively investigated in the separable case [23] and also in the general variance profile case [104], the fluctuations of the SINR around its asymptotic approximation have not been studied, aside from some particular cases, when the entries of the channel are i.i.d. [103], or when the variance profile of the channel is separable [83].

In this chapter, we show that the fluctuations of the SINR for the general profile case, converge in distribution to the Gaussian law in the asymptotic regime, and that their variance decreases as $\frac{1}{K}$ where K is the dimension of the transmitted signal. We organize this chapter as follows: after a brief overview of mutli-user systems, first order results, whose presentation and understanding is compulsory to state the Central Limit Theorem (CLT) are recalled. The CLT which represents our main contribution is then provided. Finally, simulations and numerical illustrations are presented. Note that for the sake of clarity, we will keep the same notations as those in the previous chapter.

4.1 Multi-user systems

In many communication applications, it is of great economic interest to serve as much users as possible. Such situation can be encountered in cellular systems where a base station (BTS) serves many mobile terminals or also in radio and television broadcasting. The multi-access communication encompasses not only the case of a common receiver for all users (Uplink Link) but also the case of several receivers, each of which is interested in only one stream (Downlink Link). Fig. 4.1 illustrates these two configurations for cellular telephony:

In practice, the downlink transmissions are synchronized, in the sense that the BTS sends the data to the users at the same time. On the other hand, synchronizing transmissions in the uplink link requires closed-loop timing control or providing the transmitters with access to a common clock (such as the Global Positioning System), thus increasing the design of the uplink link. However, for sake of simplicity, the synchronization is often assumed at both links.



(a) Downlink Link (b) Uplink Link

Figure 4.1: Downlink and Uplink configurations for cellular telephony

4.1.1 Multiple access methods

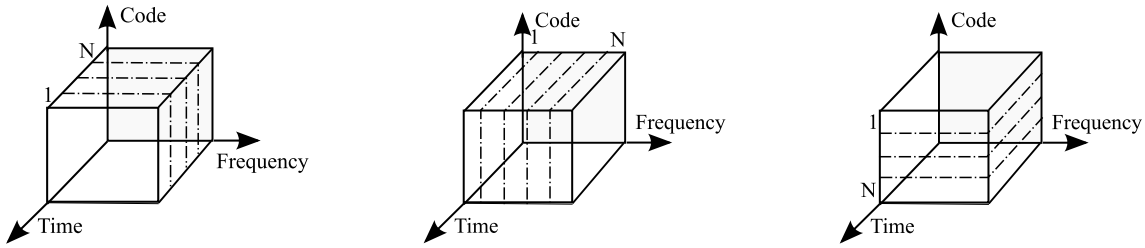
As mentioned in the previous section, the access to the radio channel in many communication applications is shared between several users. Since channel resources are limited, only a finite number of users could be supported simultaneously, the maximum being stipulated by the channel capacity. In order to serve as much users as possible, multiple access techniques that control the access to the radio channel, are regularly employed. Among these techniques, we can cite:

1. Frequency division multiple acess (FDMA): In FDMA, the total available bandwidth is shared between users, in such a way that users can access to the channel simultaneously but at different frequency ranges.

2. Time division multiple acess (TDMA): In TDMA, the time is divided into slots. Only one user is allowed to either transmit or receive in a slot. In practice, FDMA is usually used in conjunction with TDMA, (eg GSM).

3. Code division multiple access (CDMA): In CDMA, each user can send data at any time and by using the over-whole available bandwidth. Indeed, it is assigned a unique codeword which multiples the narrow-band message signal. More explicitly, At each time n, the symbol $s_k(n)$ corresponding to the k-th user is premultipled by a spreading sequence $\mathbf{c} = [c_1, \cdots, c_N]$, thus yielding the vector $[c_1 s_k(n), \cdots, c_N s_k(n)]$. Codewords between users being orthogonal, the receiver can eliminate interference and detect the user message signal. Fig. 4.2 illustrate the three aforementioned multiple access techniques.

The major advantage of CDMA over FDMA and TDMA techniques is its soft capacity, in the sense that there is no hard limit on the number of served users. This is to be compared with FDMA and TDMA systems where new users are rejected if no more free time/frequency slots are available. If a new user is added to a CDMA system, interference increases, but the user can still communicate as long as its SINR is above a certain threshold. This proves again the interest of studying the behavior of the SINR, an issue that is handled in the sequel.



(a) TDMA (b) FDMA (c) CDMA

Figure 4.2: Multiple access schemes

4.1.2 System model and LMMSE receiver

We consider multi-user applications, where the N dimensional received signal can be expressed as:

$$\mathbf{r} = \mathbf{\Sigma}\mathbf{s} + \mathbf{n} \tag{4.1}$$

$\mathbf{s} = [s_0, s_1, \cdots, s_K]^{\mathrm{T}}$ is the random vector of transmitted symbols by $K+1$ users, satisfying $\mathbb{E}\mathbf{s}\mathbf{s}^{\mathrm{H}} = \mathbf{I}_{K+1}$, the noise $\mathbf{n}$ is an independent AWGN, with covariance matrix $\mathbb{E}\mathbf{n}\mathbf{n}^{\mathrm{H}} = \rho\mathbf{I}_N$, where $\rho > 0$ and $\mathbf{\Sigma}$ are assumed to be known. The entries of $\mathbf{\Sigma}$ are assumed to verify:

$$\Sigma_{n,k} = \frac{\sigma_{n,k} W_{n,k}}{\sqrt{K}} \tag{4.2}$$

where $W_{n,k}$ are i.i.d. random variables. To detect the symbol s_0 corresponding to the first user, the LMMSE estimator is often proposed. It consists in searching the best filter $\mathbf{g}$ that minimizes the quadratic mean square error between $\widehat{s} = \mathbf{g}^{\mathrm{H}}\mathbf{r}$. In other words, $\mathbf{g}$ is given by:

$$\mathbf{g} = \min_{\mathbf{f}} \|\mathbf{f}^{\mathrm{H}}\mathbf{r} - s_0\|_2$$

Partition the channel matrix $\mathbf{\Sigma} = [\mathbf{y} \;\; \mathbf{Y}]$, where $\mathbf{y}$ is the first column of $\mathbf{\Sigma}$ and matrix $\mathbf{Y}$ has dimension $N \times K$, then vector $\mathbf{g}$ is given by:

$$\mathbf{g} = \left(\mathbf{\Sigma}\mathbf{\Sigma}^{\mathrm{H}} + \rho\mathbf{I}_N\right)^{-1}\mathbf{y} \tag{4.3}$$

Write the received vector $\mathbf{r}$ as $\mathbf{r} = s_0\mathbf{y} + \mathbf{r}_{\mathrm{in}}$ where $s_0\mathbf{y}$ is the relevant term and $\mathbf{r}_{\mathrm{in}}$ represents the so-called interference plus noise term, then the SINR is given by :

$$\beta_K = \frac{|\mathbf{g}^{\mathrm{H}}\mathbf{y}|^2}{\mathbb{E}|\mathbf{g}^{\mathrm{H}}\mathbf{r}_{\mathrm{in}}|^2} \tag{4.4}$$

Plugging (4.3) into (4.4), one can prove that the SINR β_K is given by the well-known expression:

$$\beta_K = \mathbf{y}^{\mathrm{H}}\left(\mathbf{Y}\mathbf{Y}^{\mathrm{H}} + \rho\mathbf{I}_N\right)^{-1}\mathbf{y}. \tag{4.5}$$

4.1.2.1 Possible Applications

Model (4.1) holds in many cases, including:

- CDMA transmissions on flat fading channels in the uplink direction : In this case, matrix $\mathbf{\Sigma}$ is given by :

$$\mathbf{\Sigma} = \mathbf{V}\mathbf{P}^{1/2} \tag{4.6}$$

where $\mathbf{V}$ is the $N \times (K+1)$ signature matrix assumed here to have random i.i.d. elements with mean zero and variance N^{-1}, and where $\mathbf{P} = \text{diag}(p_0, \ldots, p_K)$ is the users powers matrix. In this case, the variance profile is separable with $d_n = 1$ and $\tilde{d}_k = \frac{K}{N} p_k$. Note that elements of $\mathbf{V}$ are not Gaussian in general.

- Cellular MC-CDMA transmissions on frequency selective channels. In the uplink direction, the matrix $\mathbf{\Sigma}$ is written as:

$$\mathbf{\Sigma} = [\mathbf{H}_0 \mathbf{v}_0 \ \cdots \ \mathbf{H}_{K+1} \mathbf{v}_{K+1}] \ , \tag{4.7}$$

where $\mathbf{H}_k = \text{diag}(h_k(\exp(2\jmath\pi(n-1)/N);\ 1 \leq n \leq N)$ is the radio channel matrix of user k in the discrete Fourier domain (here N is the number of frequency bins) and $\mathbf{V} = [\mathbf{v}_0, \cdots, \mathbf{v}_K]$ is the $N \times (K+1)$ signature matrix with i.i.d. elements as in the CDMA case above. Modeling this time the channel transfer functions as deterministic functions, we have $\sigma_{nk}^2 = \frac{K}{N} |h_k(\exp(2\jmath\pi(n-1)/N))|^2$.
In the downlink direction, we have

$$\mathbf{\Sigma} = \mathbf{H}\mathbf{V}\mathbf{P}^{1/2} \tag{4.8}$$

where $\mathbf{H} = \text{diag}(h(\exp(2\jmath\pi(n-1)/N);\ 1 \leq n \leq N)$ is the radio channel matrix in the discrete Fourier domain, the $N \times (K+1)$ signature matrix $\mathbf{V}$ is as above, and $\mathbf{P} = \text{diag}(p_0, \ldots, p_K)$ is the matrix of the powers given to the different users. Model (4.8) coincides with the separable variance profile model with $d_n = \frac{K}{N} |h(\exp(2\jmath\pi(n-1)/N))|^2$ and $d_k = p_k$.

Aside from CDMA based systems, model (4.1) is valid in the case of multiple antenna transmissions with $K+1$ distant sources sending their signals toward an array of N antennas. The corresponding transmission model is $\mathbf{r} = \mathbf{\Xi}\mathbf{s} + \mathbf{n}$ where $\mathbf{\Xi} = \frac{1}{\sqrt{K}} \mathbf{H}\mathbf{P}^{\frac{1}{2}}$, matrix $\mathbf{H}$ is a $\times(K+1)$ random matrix with complex Gaussian elements representing the radio channel $\mathbf{P} = \text{diag}\,(p_0, \cdots, p_K)$ is the deterministic matrix of the powers given to the different sources, and $\mathbf{n}$ is the usual AWGN noise satisfying $\mathbb{E}\mathbf{n}\mathbf{n}^{\mathrm{H}} = \rho \mathbf{I}_N$. Partition $\mathbf{\Xi} = [\mathbf{x} \quad \mathbf{X}]$, then the SINR β at the output of the LMMSE estimator for the first element of vector $\mathbf{s}$ in the transmission model $\mathbf{r} = \mathbf{\Xi}\mathbf{s}$ is:

$$\beta = \mathbf{x}^{\mathrm{H}} \left(\mathbf{X}\mathbf{X}^{\mathrm{H}} + \rho \mathbf{I}_N\right)^{-1} \mathbf{x}.$$

Note that in this case, the entries of $\mathbf{X}$ and $\mathbf{x}$ are not necessarily independent since those of $\mathbf{H}$ are not. The assumption about the independence of entries of $\mathbf{X}$ was mentioned above when introducing the system model and is necessary to perform computations, since standard matrix theory tools are based in general on resolvent matrices stemming from matrices with i.i.d. random entries. Moreover, vector $\mathbf{x}$ should be also independent from $\mathbf{X}$ for technical purposes also.

In case the columns of $\mathbf{H}$ are independent and have the same eigenspace, one can further work out the expression of β_K and introduce new variables that verifies the independence assumption. Let $\mathbf{h}_k$ be the kth column of $\mathbf{H}$ and $\mathbf{C}_k$ its covariance matrix given by $\mathbf{C}_k = \mathbb{E}\mathbf{h}_k \mathbf{h}_k^{\mathrm{H}}$. Denote by $\mathbf{C}_k = \mathbf{U}_k \mathbf{\Lambda}_k \mathbf{U}_k^{\mathrm{H}}$ a spectral decomposition of $\mathbf{C}_k$ where $\mathbf{\Lambda}_k = \text{diag}\,(\lambda_{n,k}; 1 \leq n \leq N)$ is the matrix of eigenvalues. Since all matrices $\mathbf{C}_k$ have the same eigenspace, we can assume that matrices $\mathbf{U}_0, \cdots, \mathbf{U}_K$ are all equal to some matrix $\mathbf{U}$ a case considered, for instance, in [64] (note that sometimes they are all identified with the Fourier $N \times N$ matrix [64]). Let $\mathbf{\Sigma} = \mathbf{U}^{\mathrm{H}} \mathbf{\Xi}$, then matrix $\mathbf{\Sigma}$ is described by (4.2) where $W_{n,k}$ are standard Gaussian i.i.d., and $\sigma_{n,k}^2 = \lambda_{n,k} p_k$. Partition $\mathbf{\Sigma} = [\mathbf{y} \quad \mathbf{Y}]$ the the SINR β_K is given by:

$$\beta_K = \mathbf{x}^{\mathrm{H}} \left(\mathbf{X}\mathbf{X}^{\mathrm{H}} + \rho \mathbf{I}_N\right)^{-1} \mathbf{x} = \mathbf{y}^{\mathrm{H}} \left(\mathbf{Y}\mathbf{Y}^{\mathrm{H}} + \rho \mathbf{I}_N\right)^{-1} \mathbf{y}$$

It is also worth to say a few words about the particular separable case, which occurs when $\mathbf{\Lambda}_0 = \cdots = \mathbf{\Lambda}_K = \text{diag}\,(\lambda_1, \cdots, \lambda_N)$. In practice, this model corresponds to the well-known Kronecker model with correlations at reception [31] which can be expressed as:

$$\mathbf{\Xi} = \frac{1}{\sqrt{K}} \mathbf{\Psi}^{\frac{1}{2}} \mathbf{W} \mathbf{P} \tag{4.9}$$

where matrix $\mathbf{\Psi}$ is the $N \times N$ Hermitian non-negative matrix that captures the correlations at the receiver and $\mathbf{W}$ is a random matrix with i.i.d. standard Gaussian elements.

This model coincides with the separable variance profile model with $d_n = \lambda_n$ and $\widetilde{d}_k = p_k$.

4.2 Main assumptions

The expression of the SINR being established, we are now in position to present the first and second order results. But first, let us state the main assumptions.

A1. *The entries $\Sigma_{n,k}$ satisfy the same model as in chapter 3, i.e, $\Sigma_{n,k} = \frac{\sigma_{n,k} W_{n,k}}{\sqrt{K}}$. Moreover, $W_{n,k}$ are i.i.d. with $\mathbb{E}W_{1,0} = 0$, $\mathbb{E}W_{1,0}^2 = 0$, $\mathbb{E}|W_{1,0}|^2 = 1$ and $\mathbb{E}|W_{1,0}|^8 < +\infty$.*

A2. *There exists a real number $\sigma_{\max} < +\infty$ such that*

$$\sup_{K \geq 1} \max_{\substack{1 \leq n \leq N \\ 0 \leq k \leq K}} |\sigma_{nk}(K)| \leq \sigma_{\max} \ .$$

Similarly to chapter 3, denote by $\mathbf{D}_k$ and $\widetilde{\mathbf{D}}_n$ the diagonal matrices given by:

$$\mathbf{D}_k = \text{diag}\left(\sigma_{1k}^2(K), \cdots, \sigma_{Nk}^2(K)\right), \qquad 0 \leq k \leq K \tag{4.10}$$

$$\widetilde{\mathbf{D}}_n = \text{diag}\left(\sigma_{n1}^2(K), \cdots, \sigma_{nK}^2(K)\right), \qquad 1 \leq n \leq N \tag{4.11}$$

Then, the following assumptions may be considered in what follows:

A3. *The variance profile satisfies*

$$\liminf_{K \geq 1} \min_{0 \leq k \leq K} \frac{1}{K} \text{Tr} \mathbf{D}_k(K) > 0 \ .$$

Since $\mathbb{E}|W_{1,0}|^2 = 1$, one has $\mathbb{E}|W_{1,0}|^4 \geq 1$. The following is needed:

A4. *At least one of the following conditions is satisfied:*

$$\mathbb{E}|W_{1,0}|^4 > 1 \quad \textit{or} \quad \liminf_K \frac{1}{K^2} \text{Tr} \left(\mathbf{D}_0(K) \sum_{k=1}^{K} \mathbf{D}_k(K) \right) > 0 \ .$$

Assumptions **A1** and **A2** are required to be able to use Theorem 3.2, while assumption **A3** is technical and has already appeared in [55]. Assumption **A4** is necessary to get a non-vanishing variance Θ_K^2 in Theorem 4.3.

Similarly to chapter 3, we consider the following asymptotic regime:

$$K \to \infty, \quad , \liminf \frac{K}{N} > 0 \quad \text{and} \quad \sup \frac{K}{N} < +\infty.$$

and refer to it as $K \to \infty$.

4.3 First order results

From (4.5), it comes out that the SINR can be modeled as a quadratic form, and thus its behaviour can be studied by relying on well-known results in Random Matrix Theory. Let us first rewrite the SINR as:

$$\beta_K = \frac{1}{K}\mathbf{w}_0^{\mathrm{H}}\mathbf{D}_0^{\frac{1}{2}}\left(\mathbf{Y}\mathbf{Y}^{\mathrm{H}}+\rho\mathbf{I}\right)^{-1}\mathbf{D}_0^{\frac{1}{2}}\mathbf{w}_0 = \frac{1}{K}\mathbf{w}_0^{\mathrm{H}}\mathbf{D}_0^{\frac{1}{2}}\mathbf{Q}(-\rho)\mathbf{D}_0^{\frac{1}{2}}\mathbf{w}_0 \tag{4.12}$$

where the $N\times 1$ vector $\mathbf{w}_0$ is given by $\mathbf{w}_0 = [W_{1,0},\cdots,W_{N,0}]$ and the diagonal matrix $\mathbf{D}_0$ is given by (4.10). Recall that $\mathbf{w}_0$ and $\mathbf{Q}$ are independent and that $\|\mathbf{D}_0\| \leq \sigma^2_{\max}$ by **A2**. Furthermore, one can easily notice that $\|\mathbf{Q}(-\rho)\| = \|\left(\mathbf{Y}\mathbf{Y}^{\mathrm{H}}+\rho\mathbf{I}\right)^{-1}\| \leq \frac{1}{\rho}$.

Denote by $\mathbb{E}_Q$ the conditional expectation with respect to $\mathbf{Q}$, i.e., $\mathbb{E}_{\mathbf{Q}} = \mathbb{E}\left(.|\mathbf{Q}\right)$. From inequality (A.2) in appendix A.1, there exists a constant $C>0$ for which:

$$\begin{aligned}
\mathbb{E}\mathbb{E}_{\mathbf{Q}}\left|\beta_K - \frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{Q}(-\rho)\right|^4 &\leq \frac{C}{K^2}\left(\frac{N}{K}\right)^2\mathbb{E}\|\mathbf{D}_0\mathbf{Q}\|^4\left((\mathbb{E}|W_{1,0}|^4)^2+\mathbb{E}|W_{1,0}|^8\right)\\
&\leq \frac{C}{K^2}\left(\frac{N}{K}\right)^2\left(\frac{\sigma^2_{\max}}{\rho}\right)^4\left((\mathbb{E}|W_{1,0}|^4)^2+\mathbb{E}|W_{1,0}|^8\right)\\
&= \mathcal{O}\left(\frac{1}{K^2}\right).
\end{aligned}$$

By the Borel-Cantelli Lemma, we therefore have

$$\beta_K - \frac{1}{K}\mathrm{Tr}(\mathbf{D}_0\mathbf{Q}(-\rho)) \xrightarrow[K\to+\infty]{} 0 \quad \text{a.s.}$$

Using this result, simply apply Theorem 3.2 with $\mathbf{S}=\mathbf{D}_0$ (recall that $\|\mathbf{D}_0\| \leq \sigma^2_{\max}$) to obtain:

Theorem 4.1. *Let $\overline{\beta}_K = \frac{1}{K}\mathrm{Tr}(\mathbf{D}_0\mathbf{T}_K(-\rho))$ where $\mathbf{T}_k$ is given by Theorem 3.1 Assume* **A1** *and* **A2**. *Then,*

$$\beta_K - \overline{\beta}_K \xrightarrow[K\to+\infty]{} 0 \quad \textit{a.s.}$$

Theorem 4.1 can be adapted to the separable case. We recall that in this case, $\mathbf{D}_k = \widetilde{d}\mathbf{D}$ and $\widetilde{D}_n = d_n\widetilde{\mathbf{D}}$, where $\mathbf{D}$ and $\widetilde{\mathbf{D}}$ are given by:

$$\mathbf{D} = \mathrm{diag}\left(d_1,\cdots,d_N\right), \quad \widetilde{\mathbf{D}} = \mathrm{diag}\left(\widetilde{d}_1,\cdots,\widetilde{d}_K\right). \tag{4.13}$$

Noting $\mathbf{D}_0 = \widetilde{d}_0\mathbf{D}$ and that $\delta_K(-\rho)$ given by (3.4) coincides with $\frac{1}{K}\mathrm{Tr}(\mathbf{D}\mathbf{T})$, we get:

Theorem 4.2. *Assume that $\sigma^2_{n,k} = d_n\widetilde{d}_k$ and that* **A1** *and* **A2** *hold true. Then,*

$$\frac{\beta_K}{\widetilde{d}_0} - \delta_K(-\rho) \xrightarrow[K\to\infty]{} 0 \text{ a.s.}$$

where $\delta_K(-\rho)$ is given by (3.4) *in Theorem 3.3.*

4.4 Second order results: the CLT theorem

Beyond the convergence $\beta_K - \overline{\beta}_K \to 0$, a natural question arises regarding the accuracy of β_K for finite values of K. A first answer to this question consists in evaluating the MSE of the SINR $\mathbb{E}|\beta_K - \overline{\beta}_K|^2$ for large K. A further problem is the computation of outage probability, that is the probability for $\beta_K - \overline{\beta}_K$ to be below a certain level. Both problems can be addressed by establishing a Central Limit Theorem for $\beta_K - \overline{\beta}_K$, an issue that will be dealt with in this section.

Our approach in proving the CLT theorem is simple and powerful. It is based on the approximation of β_K by the sum of a martingale difference sequence and on the use of the CLT for martingales, [11]. We note that apart from the large random matrix theory context, such a technique has been used in [10] to establish a CLT on general quadratic forms of the type $\mathbf{z}^{\mathrm{H}}\mathbf{A}\mathbf{z}$ where $\mathbf{A}$ is a deterministic matrix and $\mathbf{z}$ is a random vector with i.i.d. elements. Hereafter, we state the CLT for the SINR β_K. Its proof is postponed to appendix B.1.

Theorem 4.3. *1. Assume that* **A2**, **A3** *and* **A4** *hold true. Let* $\mathbf{A}_K$ *and* $\mathbf{\Delta}_K$ *be the* $K \times K$ *matrices*

$$\begin{aligned} \mathbf{A}_K &= \left[\frac{1}{K} \frac{\frac{1}{K}\mathrm{Tr}\mathbf{D}_\ell \mathbf{D}_m \mathbf{T}(-\rho)^2}{\left(1+\frac{1}{K}\mathrm{Tr}\mathbf{D}_\ell \mathbf{T}(-\rho)\right)^2} \right]_{\ell,m=1}^{K} \quad \text{and} \\ \mathbf{\Delta}_K &= \mathrm{diag}\left(\left(1+\frac{1}{K}\mathrm{Tr}\mathbf{D}_\ell \mathbf{T}(-\rho)\right)^2 ; 1 \le \ell \le K \right) , \end{aligned} \tag{4.14}$$

where $\mathbf{T}$ *is defined in Theorem 3.1. Let* $\mathbf{g}_K$ *be the* $K \times 1$ *vector*

$$\mathbf{g}_K = \left[\frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{D}_1\mathbf{T}(-\rho)^2, \cdots, \frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{D}_K\mathbf{T}(-\rho)^2 \right]^{\mathrm{T}} .$$

Then the sequence of real numbers

$$\Theta_K^2 = \frac{1}{K}\mathbf{g}^{\mathrm{T}}(\mathbf{I}_K - \mathbf{A})^{-1}\mathbf{\Delta}^{-1}\mathbf{g} + (\mathbb{E}|W_{1,0}|^4 - 1)\frac{1}{K}\mathrm{Tr}\mathbf{D}_0^2\mathbf{T}(-\rho)^2 \tag{4.15}$$

is well defined and furthermore

$$0 < \liminf_K \Theta_K^2 \le \limsup_K \Theta_K^2 < +\infty .$$

2. Assume in addition **A1**. *Then the sequence* $\beta_K = \mathbf{y}^{\mathrm{H}}(\mathbf{Y}\mathbf{Y}^{\mathrm{H}} + \rho\mathbf{I})^{-1}\mathbf{y}$ *satisfies*

$$\frac{\sqrt{K}}{\Theta_K}\left(\beta_K - \overline{\beta}_K\right) \xrightarrow[K\to+\infty]{} \mathcal{N}(0,1)$$

in distribution where $\overline{\beta}_K = \frac{1}{K}\mathrm{Tr}\,\mathbf{D}_0\mathbf{T}_K$ *is defined in the statement of Theorem 4.1.*

Two remarks, which are worth pointing out, arise from Theorem 4.3. First, one can make a comparison the MSE related to the SINR β_K: $\mathrm{MSE}(\beta_K) = \mathbb{E}\left(\beta_K - \overline{\beta}_K\right)^2$, with the MSE related to Shannon's mutual information per transmit dimension $I(\rho) = \frac{1}{K}\log\det\left(\mathbf{I} + \rho\mathbf{\Sigma}\mathbf{\Sigma}^{\mathrm{H}}\right)$, (studied in [55], [77] for instance):

$$MSE(\beta_K) \quad \propto \quad \mathcal{O}\left(\frac{1}{K}\right) \qquad \text{while} \qquad MSE(I) \quad \propto \quad \mathcal{O}\left(\frac{1}{K^2}\right) .$$

The second point is about the achievability of the minimum of the MSE. Recall that the variance writes :

$$\Theta_K^2 = \frac{1}{K}\mathbf{g}^{\mathrm{T}}(\mathbf{I}_K - \mathbf{A})^{-1}\mathbf{\Delta}^{-1}\mathbf{g} + (\mathbb{E}|W_{1,0}|^4 - 1)\frac{1}{K}\mathrm{Tr}\,\mathbf{D}_0^2\mathbf{T}^2 \ .$$

As $\mathbb{E}|W_{1,0}|^2 = 1$, one clearly has $\mathbb{E}|W_{1,0}|^4 - 1 \geq 0$ with equality if and only if $|W_{1,0}| = 1$ with probability one. Moreover, we shall prove in the sequel (Section B.1.2) that $\liminf_K \frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{T}_K^2 > 0$. Therefore $(\mathbb{E}|W_{1,0}|^4 - 1)\frac{1}{K}\mathrm{Tr}\,\mathbf{D}_0^2\mathbf{T}^2$ is non-negative, and is zero if and only if $|W_{1,0}| = 1$ with probability one. As a consequence, Θ_K^2 is minimum with respect to the distribution of the W_{nk} if and only if these random variables have their values on the unit circle. In the context of CDMA and MC-CDMA, this is the case when the signature matrix elements are elements of a PSK constellation. In multi-antenna systems where model (4.1) holds (see next chapter), the W_{nk}'s are frequently considered as Gaussian, thus inducing a penalty on the SINR asymptotic MSE with respect to the unit norm case.

In the separable case, $\Theta_K^2 = \tilde{d}_0^2\Omega_K^2$ where Ω_K^2 is given by the following corollary.

Corollary 4.1. *Assume that* **A2** *is satisfied and that* $\sigma_{nk}^2 = d_n\tilde{d}_k$. *Assume moreover that*

$$\min\left(\liminf_K \frac{1}{K}\mathrm{Tr}(\mathbf{D}(K)), \liminf_K \frac{1}{K}\mathrm{Tr}(\widetilde{\mathbf{D}}(K))\right) > 0 \tag{4.16}$$

where $\mathbf{D}$ *and* $\widetilde{\mathbf{D}}$ *are given by* (4.13). *Let* $\gamma = \frac{1}{K}\mathrm{Tr}\mathbf{D}^2\mathbf{T}^2$ *and* $\tilde{\gamma} = \frac{1}{K}\mathrm{Tr}\widetilde{\mathbf{D}}^2\widetilde{\mathbf{T}}^2$. *Then the sequence*

$$\Omega_K^2 = \gamma\left(\frac{\rho^2\gamma\tilde{\gamma}}{1-\rho^2\gamma\tilde{\gamma}} + \left(\mathbb{E}|W_{1,0}|^4 - 1\right)\right) \tag{4.17}$$

satisfies $0 < \liminf_K \Omega_K^2 \leq \limsup_K \Omega_K^2 < +\infty$. *If, in addition,* **A1** *holds true, then:*

$$\frac{\sqrt{K}}{\Omega_K}\left(\frac{\beta_K}{\tilde{d}_0} - \delta_K\right) \xrightarrow[K\to+\infty]{} \mathcal{N}(0,1)$$

in distribution.

Note that condition (4.16) is the counterpart of Assumption **A3** in the case of a separable variance profile and suffices to establish $0 < \liminf_K(1-\rho^2\gamma\tilde{\gamma}) \leq \limsup_K(1-\rho^2\gamma\tilde{\gamma}) < 1$, (see for instance [53]), hence the fact that $0 < \liminf_K \Omega_K^2 \leq \limsup_K \Omega_K^2 < +\infty$. The remainder of the proof of Corollary 4.1 is postponed to Appendix B.2.

4.5 Application: MC-CDMA and multi-antenna based systems

We provide in this section a more explicit expression for δ_K which will be used in the simulations section to illustrate the SINR behaviour for the separable MIMO Model (4.9) and for MC-CDMA downlink model (4.8). By combining the tow equation in system (3.4) in Theorem 3.3, it turns out that $\delta = \delta_K(\rho)$ is the unique solution of the implicit equation:

$$\delta = \frac{1}{K}\sum_{n=0}^{N-1}\frac{d_n}{\rho + \frac{1}{K}d_n\sum_{k=1}^K \frac{p_k}{1+p_k\rho}}. \tag{4.18}$$

Recall that in the case of the separable MIMO model (4.9), $d_n = \lambda_n$ and $\widetilde{d}_k = p_k$, while in the case of MC-CDMA downlink model (4.8), $d_n = \frac{K}{N}|h(\exp(2\jmath\pi(n-1)/N))|^2$ and $\widetilde{d}_k = p_k$ again.

Here $\widetilde{d}_0 = p_0$ is the power of the user of interest (user 0), and therefore, $\frac{\beta_K}{\widetilde{d}_0}$ is the normalized SINR of this user. Notice that $\delta_K(\rho)$ is almost the same for all users, hence the normalized SINRs for all users are close to each other for large K. Their common deterministic approximation is given by (4.18) which is the discrete analogue of the integral equation (16) in [22].

As a direct application of Corollary 4.1, we can also provide expressions of γ and $\widetilde{\gamma}$ for the MIMO and MC-CDMA models. Using (3.1) and the fact that $\mathbf{T}(-\rho) = \frac{1}{\rho}\left(\mathbf{I} + \widetilde{\delta}(-\rho)\mathbf{D}\right)^{-1}$ and $\widetilde{\mathbf{T}}(-\rho) = \frac{1}{\rho}\left(\mathbf{I} + \delta(-\rho)\widetilde{\mathbf{D}}\right)^{-1}$ (See (3.5) in chapter 3), we get:

$$\begin{aligned}\gamma &= \frac{1}{K}\sum_{n=0}^{N-1}\left(\frac{d_n}{\rho + \rho d_n \tilde{\delta}}\right)^2 = \frac{1}{K}\sum_{n=0}^{N-1}\left(\frac{d_n}{\rho + \frac{1}{K}d_n\sum_{k=1}^{K}\frac{p_k}{1+p_k\delta}}\right)^2 \\ \tilde{\gamma} &= \frac{1}{K}\sum_{k=1}^{K}\left(\frac{p_k}{\rho + \rho p_k \delta}\right)^2\end{aligned}$$

4.6 Simulation results

In all simulations, the ratio $\frac{N}{K}$ is set to 2. The SNR is defined as $SNR \triangleq \frac{1}{\rho}$. Besides the power of the user of interest p_0 is set to 1.

4.6.1 The general (non necessarily separable) case

In this section, the accuracy of the Gaussian approximation is verified by simulation. In order to validate the results of Theorems 4.1 and 4.3 for practical values of K, we consider the example of a MC-CDMA transmission in the uplink direction. We recall that K is the number of interfering users in this context. In the simulation, the discrete time channel impulse response of user k is represented by the vector with $L = 5$ coefficients $\mathbf{g}_k = [g_{k,0}, \ldots, g_{k,L-1}]^{\mathrm{T}}$. These coefficients are assumed i.i.d. random variables with the complex Gaussian law $\mathcal{CN}(0, 1/L)$. We assume that a power control algorithm fixes the power received from user k to the value P_k. Setting the number of frequency bins to N, the channel matrix $\mathbf{H}_k$ for this user in the frequency domain (see Eq. (4.7)) is $\mathbf{H}_k = \mathrm{diag}(h_k(\exp(2\imath\pi(n-1)/N))$ where $h_k(z) = \frac{\sqrt{P_k}}{\|\mathbf{g}_k\|}\sum_{l=0}^{L-1} g_{k,l} z^{-l}$, and $\|\mathbf{g}_k\|$ is the Euclidean norm of $\mathbf{g}_k$. Concerning the distribution of the user powers P_k, we assume that these are arranged into five power classes with powers $P, 2P, 4P, 8P$ and $16P$ with relative frequencies given by the following table: The user of interest (User 0) is assumed to belong to Class 0, and

Table 4.1: Power classes and relative frequencies

Class	1	2	3	4	5
Power	P	$2P$	$4P$	$8P$	$16P$
Relative frequency	$\frac{1}{8}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{8}$	$\frac{1}{4}$

the Signal over Noise Ratio P/ρ for this user is set to 10 dB. In Figure 4.3, the evolution of $K\mathbb{E}(\beta_K - \bar{\beta}_K)^2/\Theta_K^2$ (where $\mathbb{E}(\beta_K - \bar{\beta}_K)^2$ is measured by simulation) is shown with respect to K. We note that this quantity is close to one for values of K as small as $K = 8$. In Figure 4.4, K is set to $K = 64$, and the SINR normalized MSE is plotted with respect to the SINR. This figure also shows that the Gaussian approximation yields a highly accurate estimation of the

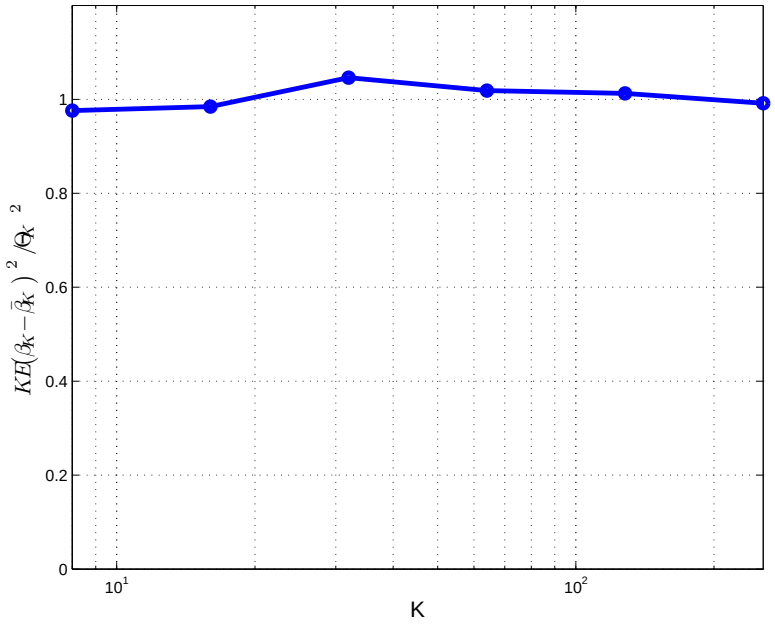


Figure 4.3: SINR normalized MSE vs K

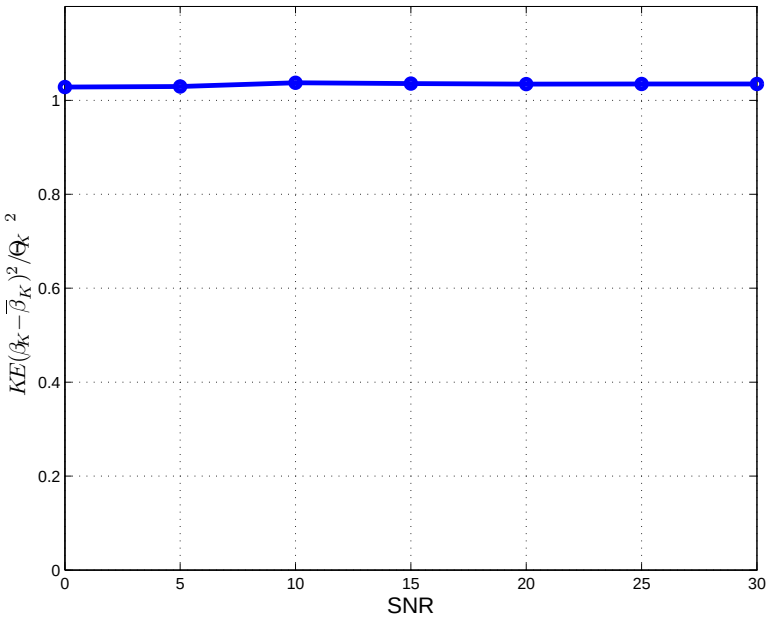


Figure 4.4: SINR normalized MSE vs SNR

normalized MSE. Figure 4.5 shows the histogram of $\sqrt{K}(\beta_K - \bar{\beta}_K)/\Theta_K$ for $N = 16$ and $N = 64$. This figure gives an idea of the similarity between the distribution of $\sqrt{K}(\beta_K - \bar{\beta}_K)/\Theta_K$ and $\mathcal{N}(0, 1)$. More rigorously, Figure 4.6 quantifies this similarity through a Quantile-Quantile test.

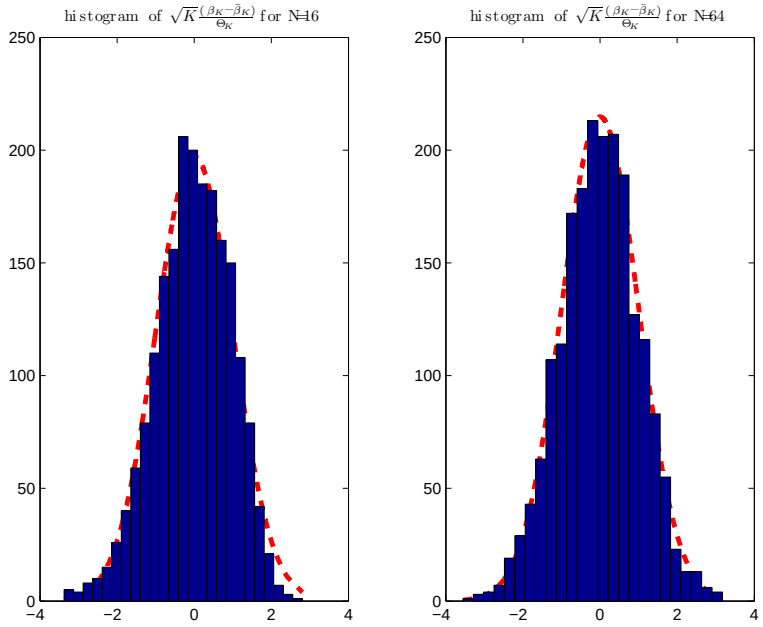


Figure 4.5: Histogram of $\sqrt{K}(\beta_K - \bar{\beta}_K)$ for $N = 16$ and $N = 64$.

4.6.2 The separable case

In order to test the results of Theorem 4.2 and Corollary 4.1, we consider the following MIMO model with exponentially decaying correlation at reception:

$$\mathbf{\Sigma} = \frac{1}{\sqrt{K}} \mathbf{\Psi}^{1/2} \mathbf{W} \mathbf{P}^{1/2}$$

where $\mathbf{\Psi} = [a^{m-n}]_{m,n=0}^{N-1}$ with $0 < a < 1$ is the covariance matrix that accounts for the correlations at the receiver side, $P = \text{diag}\,(p_0, \cdots, p_K)$ is the matrix of the powers given to the different users and W is a $N \times (K+1)$ matrix with Gaussian standard iid elements. Let $\mathbf{P_u}$ denote the vector containing the interfering users' powers. We set $\mathbf{P_u}$ (up to a permutation) to:

$$\mathbf{P_u} = \begin{cases} [4P \;\; 5P] & \text{if} \;\; K = 2 \\ [P \;\; P \;\; 2P \;\; 4P] & \text{if} \;\; K = 4 \\ [P \;\; P \;\; 2P \;\; 2P \;\; 2P \;\; 4P \;\; 4P \;\; 4P \;\; 8P \;\; 16P \;\; 16P \;\; 16P \;\;] & \text{if} \;\; K = 12 \end{cases}$$

if $K = 2^p (7 \geq p \geq 3)$, we assume that the interfering users are arranged into 5 classes according to their powers as shown in Table4.1. We set the SNR to 10dB and a to 0.1. We investigate in this section the accuracy of the Gaussian approximation in predicting the outage probability. In Fig.4.7, we compare the empirical 1% outage SIR with that given by the central limit theorem. We note that the Gaussian approximation tends to under estimate the 1% outage SIR. We also note that it yields a good accuracy for small values of α and for enough large values of N ($N \geq 64$).

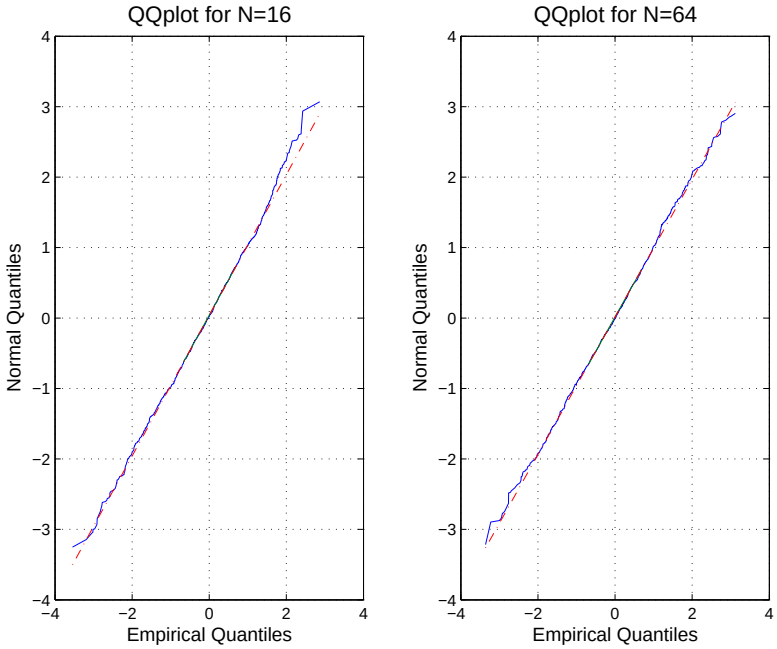


Figure 4.6: Q-Q test for $\sqrt{K}(\beta_K - \bar{\beta}_K)$, $N = 16$ and $N = 64$; in red dash doted line, 45 degree line.

4.7 Conclusion

This chapter establishes the asymptotic normality of the SINR at the output of the LMMSE receiver. The novelty of our approach is that it enables to address general variance profiles, corresponding to a wide range of systems, ranging from CDMA to MIMO based techniques. Simulations have been presented and have shown that the Gaussian approximation is valid, as long as the system dimensions are large enough. However, it has been shown that it cannot be suitable for computing outage probabilities for MIMO systems, since very large system dimensions are required. To approximate performance metrics, like the BER and outage probability, another technique should be used. This will be the subject of the next chapter.

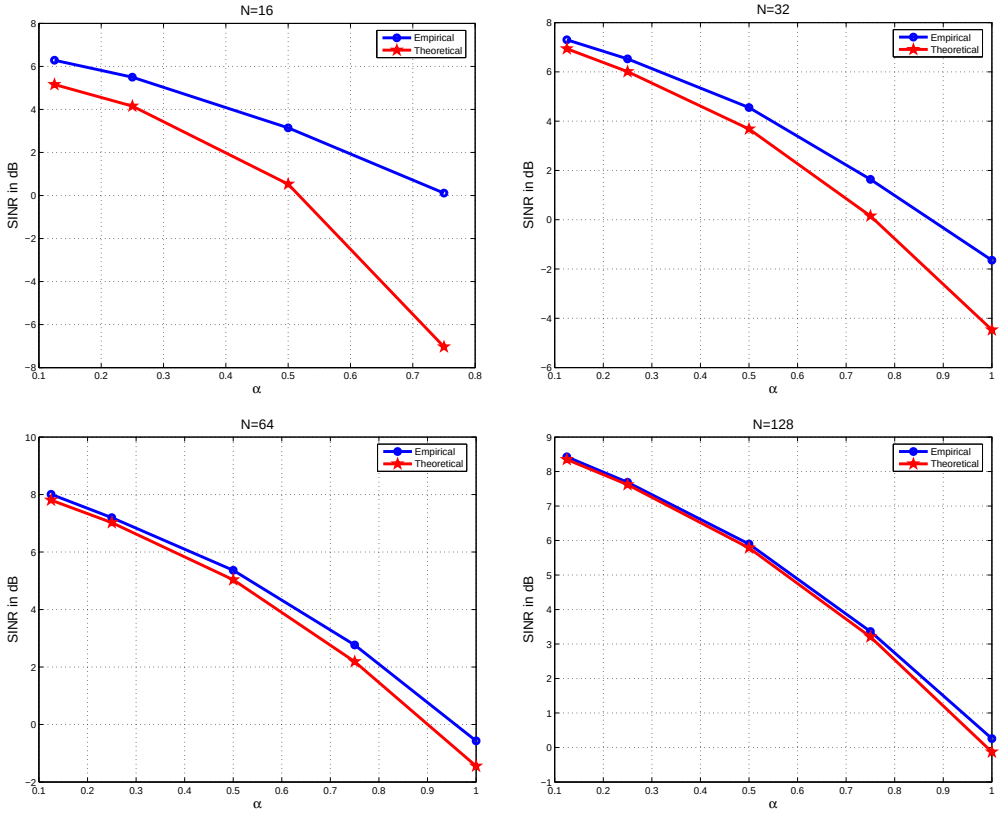


Figure 4.7: Comparison of the empirical and theoretical 1% outage SIR

Appendices

Appendix B

Proof of results in chapter 4

This appendix is devoted to the proof of theorem 4.3 and Corollary 4.1 in chapter 4.

B.1 Proof of theorem 4.3

In this section, we provide the details of the proof of Theorem 4.3. We begin with mathematical preliminaries.

B.1.1 Preliminaries

The following lemma gathers useful matrix results, whose proofs can be found in [58]:

Lemma B.1. *Assume* $\mathbf{X} = [x_{ij}]_{i,j=1}^N$ *and* $\mathbf{Y}$ *are complex* $N \times N$ *matrices. Then*

1. *For every* $i, j \leq N$, $|x_{ij}| \leq \|\mathbf{X}\|$. *In particular,* $\|\mathrm{diag}(\mathbf{X})\| \leq \|\mathbf{X}\|$.
2. $\|\mathbf{XY}\| \leq \|\mathbf{X}\| \, \|\mathbf{Y}\|$.
3. *For* $\rho > 0$, *the resolvent* $(\mathbf{X}\mathbf{X}^{\mathrm{H}} + \rho\mathbf{I})^{-1}$ *satisfies* $\|(\mathbf{X}\mathbf{X}^{\mathrm{H}} + \rho\mathbf{I})^{-1}\| \leq \rho^{-1}$.
4. *If* $\mathbf{Y}$ *is Hermitian non-negative, then* $|\mathrm{Tr}(\mathbf{XY})| \leq \|\mathbf{X}\|\mathrm{Tr}(\mathbf{Y})$.

Let $\mathbf{X} = \mathbf{U}\mathbf{\Lambda}\mathbf{V}^{\mathrm{H}}$ be a spectral decomposition of $\mathbf{X}$ where $\mathbf{\Lambda} = \mathrm{diag}(\lambda_1, \ldots, \lambda_n)$ is the matrix of singular values of $\mathbf{X}$. For a real $p \geq 1$, the Schatten ℓ_p-norm of $\mathbf{X}$ is defined as $\|\mathbf{X}\|_p = (\sum \lambda_i^p)^{1/p}$. The following bound over the Schatten ℓ_p-norm of a triangular matrix will be of help (for a proof, see [10], [79, page 278]):

Lemma B.2. *Let* $\mathbf{X} = [x_{ij}]_{i,j=1}^N$ *be a* $N \times N$ *complex matrix and let* $\widetilde{\mathbf{X}} = [x_{ij}\mathbf{1}_{i>j}]_{i,j=1}^N$ *be the strictly lower triangular matrix extracted from* $\mathbf{X}$. *Then for every* $p \geq 1$, *there exists a constant* C_p *depending on* p *only such that*

$$\|\widetilde{\mathbf{X}}\|_p \leq C_p \|\mathbf{X}\|_p \ .$$

The following lemma lists some properties of the resolvent $\mathbf{Q}$ and the deterministic approximation matrix $\mathbf{T}$. Its proof is postponed to Appendix B.1.4.

Lemma B.3. *The following facts hold true:*

1. *Assume* **A2**. *Consider matrices* $\mathbf{T}_K(-\rho) = \mathrm{diag}(t_1(-\rho), \ldots, t_N(-\rho))$ *defined by Theorem 3.1. Then for every* $1 \leq n \leq N$,

$$\frac{1}{\rho + \sigma_{\max}^2} \leq t_n(-\rho) \leq \frac{1}{\rho} \; . \tag{B.1}$$

2. *Assume in addition* **A1** *and* **A3**. *Let* $\mathbf{Q}_K(-\rho) = (\mathbf{Y}\mathbf{Y}^{\mathrm{H}} + \rho\mathbf{I})^{-1}$ *and let matrices* $\mathbf{S}_K$ *be as in the statement of Theorem 3.2. Then*

$$\sup_K \mathbb{E}\left|\mathrm{Tr}\ \mathbf{S}_K(\mathbf{Q}_K - \mathbf{T}_K)\right|^2 < \infty \; . \tag{B.2}$$

B.1.2 Proof of Theorem 4.3–(1)

We introduce the following notations. Assume that $\mathbf{X}$ is a real matrix, by $\mathbf{X} \succcurlyeq \mathbf{0}$ we mean $X_{ij} \geq 0$ for every element X_{ij}. For a vector $\mathbf{x}$, $\mathbf{x} \succcurlyeq \mathbf{0}$ is defined similarly. In the remainder of this appendix, $C = C(\rho, \sigma_{\max}^2, \liminf \frac{N}{K}, \sup \frac{N}{K}) < \infty$ denotes a positive constant whose value may change from line to line.

The following lemma, which directly follows from [55, Lemma 5.2 and Proposition 5.5], states some important properties of the matrices $\mathbf{A}_K$ defined in the statement of Theorem 4.3.

Lemma B.4. *Assume* **A2** *and* **A3**. *Consider matrices* $\mathbf{A}_K$ *defined by* (4.14). *Then the following facts hold true:*

1. *Matrix* $\mathbf{I}_K - \mathbf{A}_K$ *is invertible, and* $(\mathbf{I}_K - \mathbf{A}_K)^{-1} \succcurlyeq \mathbf{0}$.
2. *Element* (k,k) *of the inverse satisfies* $\left[(\mathbf{I}_K - \mathbf{A}_K)^{-1}\right]_{k,k} \geq 1$ *for every* $1 \leq k \leq K$.
3. *The maximum row sum norm of the inverse satisfies* $\limsup_K \left|\!\left|\!\left|(\mathbf{I}_K - \mathbf{A}_K)^{-1}\right|\!\right|\!\right|_\infty < \infty$.

Due to Lemma B.4–(1), Θ_K^2 is well defined. Let us prove that $\limsup_K \Theta_K^2 < \infty$. The first term of the right-hand side of (4.15) satisfies

$$\begin{aligned}\frac{1}{K}\mathbf{g}^{\mathrm{T}}(\mathbf{I}_K - \mathbf{A}_K)^{-1}\boldsymbol{\Delta}^{-1}\mathbf{g} &\leq \|\mathbf{g}\|_\infty \|(\mathbf{I}_K - \mathbf{A}_K)^{-1}\boldsymbol{\Delta}^{-1}\mathbf{g}\|_\infty \\ &\leq \|\mathbf{g}\|_\infty \left|\!\left|\!\left|(\mathbf{I}_K - \mathbf{A}_K)^{-1}\right|\!\right|\!\right|_\infty \|\boldsymbol{\Delta}^{-1}\mathbf{g}\|_\infty \leq \|\mathbf{g}\|_\infty^2 \left|\!\left|\!\left|(\mathbf{I}_K - \mathbf{A}_K)^{-1}\right|\!\right|\!\right|_\infty \end{aligned} \tag{B.3}$$

due to $\left|\!\left|\!\left|\boldsymbol{\Delta}^{-1}\right|\!\right|\!\right|_\infty \leq 1$. Recall that $\|\mathbf{T}\| \leq \rho^{-1}$ by Lemma B.3–(1). Therefore, any element of $\mathbf{g}$ satisfies

$$\frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{D}_k\mathbf{T}^2 \leq \frac{N}{K}\|\mathbf{D}_0\|\|\mathbf{D}_k\|\|\mathbf{T}\|^2 \leq \frac{N}{K}\frac{\sigma_{\max}^4}{\rho^2} \tag{B.4}$$

by **A2**, hence $\sup_K \|\mathbf{g}\| \leq C$. From Lemma B.4–(3) and (B.3), we then obtain

$$\limsup_K \frac{1}{K}\mathbf{g}^{\mathrm{T}}(\mathbf{I}_K - \mathbf{A}_K)^{-1}\boldsymbol{\Delta}^{-1}\mathbf{g} \leq C. \tag{B.5}$$

We can prove similarly that the second term in the right-hand side of (4.15) satisfies $\sup_K((\mathbb{E}|W_{1,0}|^4 - 1)\frac{1}{K}\mathrm{Tr}\mathbf{D}_0^2\mathbf{T}(-\rho)^2) \leq C$. Hence $\limsup_K \Theta_K^2 < \infty$.

Let us prove that $\liminf_K \Theta_K^2 > 0$. We have

$$\begin{aligned}
\frac{1}{K}\mathbf{g}^{\mathrm{T}}(\mathbf{I}_K - \mathbf{A}_K)^{-1}\boldsymbol{\Delta}^{-1}\mathbf{g} &\overset{(a)}{\geq} \frac{1}{K}\mathbf{g}^{\mathrm{T}}\mathrm{diag}\left((\mathbf{I}_K - \mathbf{A}_K)^{-1}\right)\boldsymbol{\Delta}^{-1}\mathbf{g} \\
&\overset{(b)}{\geq} \frac{1}{\left(1+\frac{N}{K}\frac{\sigma_{\max}^2}{\rho}\right)^2}\frac{1}{K}\sum_{k=1}^{K}\left(\frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{D}_k\mathbf{T}^2\right)^2 \\
&\overset{(c)}{\geq} \frac{1}{\left(1+\frac{N}{K}\frac{\sigma_{\max}^2}{\rho}\right)^2}\left(\frac{1}{K^2}\mathrm{Tr}\ \mathbf{D}_0\left(\sum_{k=1}^{K}\mathbf{D}_k\right)\mathbf{T}^2\right)^2 \\
&\overset{(d)}{\geq} \frac{1}{\left(1+\frac{N}{K}\frac{\sigma_{\max}^2}{\rho}\right)^2(\rho+\sigma_{\max}^2)^4}\left(\frac{1}{K^2}\mathrm{Tr}\ \mathbf{D}_0\sum_{k=1}^{K}\mathbf{D}_k\right)^2 \\
&\geq C\left(\frac{1}{K^2}\mathrm{Tr}\ \mathbf{D}_0\sum_{k=1}^{K}\mathbf{D}_k\right)^2 ,
\end{aligned}$$

where (a) follows from the fact that $(\mathbf{I}_K - \mathbf{A}_K)^{-1} \succcurlyeq \mathbf{0}$ (Lemma B.4–(1), and the straightforward inequalities $\boldsymbol{\Delta}^{-1} \succcurlyeq \mathbf{0}$ and $\mathbf{g} \succcurlyeq \mathbf{0}$), (b) follows from Lemma B.4–(2) and $\|\boldsymbol{\Delta}\| \leq (1+\frac{N}{K}\frac{\sigma_{\max}^2}{\rho})^2$, (c) follows from the elementary inequality $n^{-1}\sum x_i^2 \geq (n^{-1}\sum x_i)^2$, and (d) is due to Lemma B.3–(1). Similar derivations yield:

$$(\mathbb{E}|W_{1,0}|^4 - 1)\frac{1}{K}\mathrm{Tr}\mathbf{D}_0^2\mathbf{T} \quge \frac{\mathbb{E}|W_{1,0}|^4 - 1}{(\rho+\sigma_{\max}^2)^2}\left(\frac{1}{K}\mathrm{Tr}\mathbf{D}_0\right)^2 \quge C(\mathbb{E}|W_{1,0}|^4 - 1)$$

by **A3**. Therefore, if **A4** holds true, then $\liminf_K \Theta_K^2 > 0$ and Theorem 4.3–(1) is proved.

B.1.3 Proof of Theorem 4.3–(2)

Recall that the SINR β_K is given by Equation (4.12). The random variable $\frac{\sqrt{K}}{\Theta_K}(\beta_K - \overline{\beta}_K)$ can therefore be decomposed as

$$\begin{aligned}
\frac{\sqrt{K}}{\Theta_K}(\beta_K - \overline{\beta}_K) &= \frac{1}{\sqrt{K}\Theta_K}\left(\mathbf{w}_0^{\mathrm{H}}\mathbf{D}_0^{1/2}\mathbf{Q}\mathbf{D}_0^{1/2}\mathbf{w}_0 - \mathrm{Tr}(\mathbf{D}_0\mathbf{Q})\right) + \frac{1}{\sqrt{K}\Theta_K}\left(\mathrm{Tr}(\mathbf{D}_0(\mathbf{Q}-\mathbf{T}))\right) \\
&= U_{1,K} + U_{2,K} \ . \qquad \text{(B.6)}
\end{aligned}$$

Thanks to Lemma B.3–(2) and to the fact that $\liminf_K \Theta_K^2 > 0$, we have $\mathbb{E}U_{K,2}^2 < CK^{-1}$ which implies that $U_{K,2} \to 0$ in probability as $K \to \infty$. Hence, in order to conclude that

$$\frac{\sqrt{K}}{\Theta_K}(\beta_K - \overline{\beta}_K) \xrightarrow[K\to\infty]{} \mathcal{N}(0,1) \quad \text{in distribution} ,$$

it is sufficient by Slutsky's theorem to prove that $U_{1,K} \to \mathcal{N}(0,1)$ in distribution. The remainder of the section is devoted to this point.

Remark B.1. *Decomposition* (B.6) *and the convergence to zero (in probability) of* $U_{2,K}$ *yield the following interpretation: The fluctuations of* $\sqrt{K}(\beta_K - \overline{\beta}_K)$ *are mainly due to the fluctuations of*

vector $\mathbf{w}_0$*. Indeed the contribution of the fluctuations*[1] *of* $\frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{Q}$*, due to the random nature of* $\mathbf{Y}$*, is negligible.*

Denote by $\mathbb{E}_n$ the conditional expectation $\mathbb{E}_n[\,\cdot\,] = \mathbb{E}[\,\cdot \parallel W_{n,0}, W_{n+1,0}, \ldots, W_{N,0}, \mathbf{Y}]$. Put $\mathbb{E}_{N+1}[\,\cdot\,] = \mathbb{E}[\,\cdot \parallel \mathbf{Y}]$ and note that $\mathbb{E}_{N+1}(\mathbf{w}_0^{\mathrm{H}}\mathbf{D}_0^{1/2}\mathbf{Q}\mathbf{D}_0^{1/2}\mathbf{w}_0) = \mathrm{Tr}\mathbf{D}_0\mathbf{Q}$. With these notations at hand, we have:

$$U_{1,K} = \frac{1}{\Theta_K}\sum_{n=1}^{N}(\mathbb{E}_n - \mathbb{E}_{n+1})\frac{\mathbf{w}_0^{\mathrm{H}}\mathbf{D}_0^{1/2}\mathbf{Q}\mathbf{D}_0^{1/2}\mathbf{w}_0}{\sqrt{K}} \triangleq \frac{1}{\Theta_K}\sum_{n=1}^{N} Z_{n,K} \ . \tag{B.7}$$

Consider the increasing sequence of $\sigma-$fields

$$\mathcal{F}_{N,K} = \sigma(W_{N,0}, \mathbf{Y}) \ , \quad \cdots \ , \quad \mathcal{F}_{1,K} = \sigma(W_{1,0}, \cdots, W_{N,0}, \mathbf{Y}) \ .$$

Then the random variable $Z_{n,K}$ is integrable and measurable with respect to $\mathcal{F}_{n,K}$; moreover it readily satisfies $\mathbb{E}_{n+1}Z_{n,K} = 0$. In particular, the sequence $(Z_{N,K}, \ldots, Z_{1,K})$ is a martingale difference sequence with respect to $(\mathcal{F}_{N,K}, \cdots, \mathcal{F}_{1,K})$. The following CLT for martingales is the key tool to study the asymptotic behaviour of $U_{1,K}$:

Theorem B.1. *Let* $X_{N,K}, X_{N-1,K}, \ldots, X_{1,K}$ *be a martingale difference sequence with respect to the increasing filtration* $\mathcal{G}_{N,K}, \ldots, \mathcal{G}_{1,K}$*. Assume that there exists a sequence of real positive numbers* s_K^2 *such that*

$$\frac{1}{s_K^2}\sum_{n=1}^{N}\mathbb{E}\left[X_{n,K}^2 \| \mathcal{G}_{n+1,K}\right] \xrightarrow[K\to\infty]{} 1$$

in probability. Assume further that the Lyapunov condition holds:

$$\exists \alpha > 0, \quad \frac{1}{s_K^{2(1+\alpha)}}\sum_{n=1}^{N}\mathbb{E}\left|X_{n,K}\right|^{2+\alpha} \xrightarrow[K\to\infty]{} 0 \ ,$$

Then $s_K^{-1}\sum_{n=1}^{N} X_{n,K}$ *converges in distribution to* $\mathcal{N}(0,1)$ *as* $K \to \infty$*.*

Remark B.2. *This theorem is proved in [11], gathering Theorem 35.12 (which is expressed under the weaker Lindeberg condition) together with the arguments of Section 27 (where it is proved that Lyapunov's condition implies Lindeberg's condition).*

In order to prove that

$$U_{1,K} = \frac{1}{\Theta_K}\sum_{n=1}^{N} Z_{n,K} \xrightarrow[K\to\infty]{} \mathcal{N}(0,1) \quad \text{in distribution} \ , \tag{B.8}$$

we shall apply Theorem B.1 to the sum $\frac{1}{\Theta_K}\sum_{n=1}^{N} Z_{n,K}$ and the filtration $(\mathcal{F}_{n,K})$. The proof is carried out into four steps:

[1] In fact, one may prove that the fluctuation of $\frac{1}{K}\mathrm{Tr}\mathbf{D}_0(\mathbf{Q}-\mathbf{T})$ are of order K, i.e. $\mathrm{Tr}\mathbf{D}_0(\mathbf{Q}-\mathbf{T})$ asymptotically behaves as a Gaussian random variable. Such a speed of fluctuations already appears in [55], when studying the fluctuations of the mutual information.

Step 1 We first establish Lyapunov's condition. Due to the fact that $\liminf_K \Theta_K^2 > 0$, we only need to show that

$$\exists\ \alpha > 0, \quad \sum_{n=1}^{N} \mathbb{E}|Z_{n,K}|^{2+\alpha} \xrightarrow[K\to\infty]{} 0\ . \tag{B.9}$$

Step 2 We prove that $V_K = \sum_{n=1}^{N} \mathbb{E}_{n+1} Z_{n,K}^2$ satisfies

$$V_K - \left(\frac{\left(\mathbb{E}|W_{1,0}|^4 - 2\right)}{K} \mathrm{Tr}\left(\mathbf{D}_0^2 (\mathrm{diag}(\mathbf{Q}))^2\right) + \frac{1}{K}\mathrm{Tr}(\mathbf{D}_0\mathbf{Q}\mathbf{D}_0\mathbf{Q}) \right) \xrightarrow[K\to\infty]{} 0 \quad \text{in probability .} \tag{B.10}$$

Step 3 We first show that

$$\frac{1}{K}\mathrm{Tr}\mathbf{D}_0^2(\mathrm{diag}(\mathbf{Q}))^2 - \frac{1}{K}\mathrm{Tr}\mathbf{D}_0^2\mathbf{T}^2 \xrightarrow[K\to\infty]{} 0 \quad \text{in probability.} \tag{B.11}$$

In order to study the asymptotic behaviour of $\frac{1}{K}\mathrm{Tr}(\mathbf{D}_0\mathbf{Q}\mathbf{D}_0\mathbf{Q})$, we introduce the random variables $U_\ell = \frac{1}{K}\mathrm{Tr}(\mathbf{D}_0\mathbf{Q}\mathbf{D}_\ell\mathbf{Q})$ for $0 \le \ell \le K$ (the one of interest being U_0). We then prove that the U_ℓ's satisfy the following system of equations:

$$U_\ell = \sum_{k=1}^{K} c_{\ell k} U_k + \frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{D}_\ell\mathbf{T}^2 + \epsilon_\ell, \quad 0 \le \ell \le K, \tag{B.12}$$

where

$$c_{\ell k} = \frac{1}{K} \frac{\frac{1}{K}\mathrm{Tr}\mathbf{D}_\ell\mathbf{D}_k\mathbf{T}(-\rho)^2}{\left(1 + \frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{T}(-\rho)\right)^2}, \quad 0 \le \ell \le K,\ 1 \le k \le K \tag{B.13}$$

and the perturbations ϵ_ℓ satisfy $\mathbb{E}|\epsilon_\ell| \le CK^{-\frac{1}{2}}$ where we recall that C is independent of ℓ.

Step 4 We prove that $U_0 = \frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{Q}\mathbf{D}_0\mathbf{Q}$ satisfies

$$U_0 = \frac{1}{K}\mathrm{Tr}\mathbf{D}_0^2\mathbf{T}^2 + \frac{1}{K}\mathbf{g}^{\mathrm{T}}\left(\mathbf{I} - \mathbf{A}\right)^{-1}\mathbf{\Delta}^{-1}\mathbf{g} + \epsilon \tag{B.14}$$

with $\mathbb{E}|\epsilon| \le CK^{-\frac{1}{2}}$. This equation combined with (B.10) and (B.11) yields $\sum_n \mathbb{E}_{n+1} Z_{n,K}^2 - \Theta_K^2 \to 0$ in probability. As $\liminf_K \Theta_K^2 > 0$, this implies $\frac{1}{\Theta_K}\sum_n \mathbb{E}_{n+1} Z_{n,K}^2 \to 1$ in probability, which proves (B.8) and thus ends the proof of Theorem 4.3.

Write $\mathbf{B} = [b_{ij}]_{i,j=1}^N = \mathbf{D}_0^{1/2}\mathbf{Q}\mathbf{D}_0^{1/2}$ and recall from (B.7) that $Z_{n,K} = \frac{1}{\sqrt{K}}(\mathbb{E}_n - \mathbb{E}_{n+1})\mathbf{w}_0^{\mathrm{H}}\mathbf{B}\mathbf{w}_0$. We have

$$\mathbb{E}_n \mathbf{w}_0^{\mathrm{H}}\mathbf{B}\mathbf{w}_0 = \sum_{\ell=1}^{n-1} b_{\ell\ell} + \sum_{\ell_1,\ell_2=n}^{N} W_{\ell_1 0}^* W_{\ell_2 0} b_{\ell_1\ell_2}\ .$$

Hence

$$Z_{n,K} = \frac{1}{\sqrt{K}}\left(\left(|W_{n0}|^2 - 1\right) b_{nn} + W_{n0}^* \sum_{\ell=n+1}^{N} W_{\ell 0} b_{n\ell} + W_{n0} \sum_{\ell=n+1}^{N} W_{\ell 0}^* b_{\ell n} \right)\ . \tag{B.15}$$

Step 1: Validation of the Lyapunov condition The following inequality will be of help to check Lyapunov's condition.

Lemma B.5 (Burkholder's inequality)**.** *Let X_k be a complex martingale difference sequence with respect to the increasing sequence of σ–fields $\mathcal{F}_k$. Then for $p \geq 2$, there exists a constant C_p for which*

$$\mathbb{E}\left|\sum_k X_k\right|^p \leq C_p \left(\mathbb{E}\left(\sum_k \mathbb{E}\left[|X_k|^2 \| \mathcal{F}_{k-1}\right] \right)^{p/2} + \mathbb{E}\sum_k |X_k|^p \right) .$$

Recall Assumption **A1**. Eq. (B.15) yields:

$$\begin{aligned} |Z_{n,K}|^4 &\leq \frac{1}{K^2}\left(\frac{|W_{n0}|^2+1}{\rho\sigma_{\max}^2} + 2\left| W_{n0} \sum_{\ell=n+1}^{N} W_{\ell 0} b_{n\ell} \right| \right)^4 \\ &\leq \frac{2^3}{K^2}\left(\left(\frac{|W_{n0}|^2+1}{\rho\sigma_{\max}^2} \right)^4 + 2^4 \left| W_{n0} \sum_{\ell=n+1}^{N} W_{\ell 0} b_{n\ell} \right|^4 \right) \end{aligned} \tag{B.16}$$

where we use the fact that $|b_{nn}| \leq (\rho\sigma_{\max}^2)^{-1}$ (cf. Lemma B.1–(1)) and the convexity of $x \mapsto x^4$. Due to Assumption **A1**, we have:

$$\mathbb{E}\left(|W_{n0}|^2+1\right)^4 \leq 2^3\left(\mathbb{E}|W_{n0}|^8+1\right) < \infty . \tag{B.17}$$

Considering the second term at the right-hand side of (B.16), we write

$$\begin{aligned} \mathbb{E}\left| W_{n0} \sum_{\ell=n+1}^{N} W_{\ell 0} b_{n\ell} \right|^4 &= \mathbb{E}\,|W_{n0}|^4\, \mathbb{E}\left| \sum_{\ell=n+1}^{N} W_{\ell 0} b_{n\ell} \right|^4 , \\ &\overset{(a)}{\leq} C\left(\mathbb{E}\left(\sum_{\ell=n+1}^{N} (\mathbb{E}|W_{\ell 0}|^2)|b_{n\ell}|^2 \right)^2 + \sum_{\ell=n+1}^{N} (\mathbb{E}|W_{\ell 0}|^4)(\mathbb{E}|b_{n\ell}|^4) \right) , \\ &\overset{(b)}{\leq} C\left(\mathbb{E}\left(\sum_{\ell=n+1}^{N} |b_{n\ell}|^2 \right)^2 + \sum_{\ell=n+1}^{N} \mathbb{E}|b_{n\ell}|^2 \right) , \end{aligned}$$

where (a) follows from Lemma B.5 (Burkholder's inequality), the filtration being $\mathcal{F}_{N,K}, \ldots, \mathcal{F}_{n+1,K}$ and (b) follows from the bound $|b_{n\ell}|^4 \leq |b_{n\ell}|^2 \max\, |b_{n\ell}|^2 \leq |b_{n\ell}|^2(\sigma_{\max}^2\rho^{-1})^2$ (cf. Lemma B.1–(1)). Now, notice that

$$\sum_{\ell=n+1}^{N} |b_{n\ell}|^2 < \sum_{\ell=1}^{N} |b_{n\ell}|^2 = \left[\mathbf{D}_0^{1/2}\mathbf{Q}\mathbf{D}_0\mathbf{Q}\mathbf{D}_0^{1/2}\right]_{nn} \leq \|\mathbf{D}_0^{1/2}\mathbf{Q}\mathbf{D}_0\mathbf{Q}\mathbf{D}_0^{1/2}\| \leq \frac{\sigma_{\max}^4}{\rho^2} .$$

This yields $\mathbb{E}|W_{n0}\sum_{\ell=n+1}^{N} W_{\ell 0} b_{n\ell}|^4 \leq C$. Gathering this result with (B.17), getting back to (B.16), taking the expectation and summing up finally yields:

$$\sum_{n=1}^{N} \mathbb{E}|Z_{n,K}|^4 \leq \frac{C}{K} \xrightarrow[K\to\infty]{} 0$$

which establishes Lyapunov's condition (B.9) with $\alpha = 2$.

Step 2: Proof of (B.10) Eq. (B.15) yields:

$$\mathbb{E}_{n+1}Z_{n,K}^2 = \frac{1}{K}\left(\left(\mathbb{E}|W_{1,0}|^4-1\right)b_{nn}^2+\mathbb{E}_{n+1}\left(W_{n0}^*\sum_{\ell=n+1}^N W_{\ell 0}b_{n\ell}+W_{n0}\sum_{\ell=n+1}^N W_{\ell 0}^*b_{\ell n}\right)^2\right.$$
$$\left.+2b_{nn}\left(\mathbb{E}\,W_{1,0}^*|W_{1,0}|^2\right)\sum_{\ell=n+1}^N W_{\ell 0}b_{n\ell}+2b_{nn}\left(\mathbb{E}\,W_{1,0}|W_{1,0}|^2\right)\sum_{\ell=n+1}^N W_{\ell 0}^*b_{\ell n}\right) .$$

Note that the second term of the right-hand side writes:

$$\mathbb{E}_{n+1}\left(W_{n0}^*\sum_{\ell=n+1}^N W_{\ell 0}b_{n\ell}+W_{n0}\sum_{\ell=n+1}^N W_{\ell 0}^*b_{\ell n}\right)^2 = 2\sum_{\ell_1,\ell_2=n+1}^N W_{\ell_1 0}W_{\ell_2 0}^*b_{n\ell_1}b_{\ell_2 n} .$$

Therefore, $V_K=\sum_{n=1}^N \mathbb{E}_{n+1}Z_{n,K}^2$ writes:

$$V_K=\frac{\left(\mathbb{E}|W_{1,0}|^4-1\right)}{K}\sum_{n=1}^N b_{nn}^2+\frac{2}{K}\sum_{n=1}^N\sum_{\ell_1,\ell_2=n+1}^N W_{\ell_1 0}W_{\ell_2 0}^*b_{n\ell_1}b_{\ell_2 n}$$
$$+\frac{2}{K}\Re\left(\left(\mathbb{E}\,W_{1,0}^*|W_{1,0}|^2\right)\sum_{n=1}^N b_{nn}\sum_{\ell=n+1}^N W_{\ell 0}b_{n\ell}\right) ,$$

where $\Re$ denotes the real part of a complex number. We introduce the following notations:

$$\mathbf{R}=(r_{ij})_{i,j=1}^N \stackrel{\triangle}{=} (b_{ij}\mathbf{1}_{i>j})_{i,j=1}^N \qquad \text{and} \qquad \Gamma_K=\frac{1}{K}\sum_{n=1}^N b_{nn}\sum_{\ell=n+1}^N W_{\ell 0}b_{n\ell} .$$

Note in particular that $\mathbf{R}$ is the strictly lower triangular matrix extracted from $\mathbf{D}_0^{1/2}\mathbf{Q}\mathbf{D}_0^{1/2}$. We can now rewrite V_K as:

$$V_K=\frac{\left(\mathbb{E}|W_{1,0}|^4-1\right)}{K}\mathrm{Tr}\left(\mathbf{D}_0^2(\mathrm{diag}(\mathbf{Q}))^2\right)+\frac{2}{K}\mathbf{w}_0^{\mathrm{H}}\mathbf{R}\mathbf{R}^{\mathrm{H}}\mathbf{w}_0+2\Re\left(\Gamma_K\mathbb{E}W_{1,0}^*|W_{1,0}|^2\right) . \quad \text{(B.18)}$$

We now prove that the third term of the right-hand side vanishes, and find an asymptotic equivalent for the second one. Using Lemma B.1, we have:

$$\begin{aligned}
\mathbb{E}_{N+1}|\Gamma_K|^2 &= \frac{1}{K^2}\sum_{n,m=1}^N b_{nn}b_{mm}\sum_{\ell=1}^N b_{n\ell}b_{m\ell}^*\mathbf{1}_{\ell>n}\mathbf{1}_{\ell>m}=\frac{1}{K^2}\mathrm{Tr}\left(\mathrm{diag}(\mathbf{B})\mathbf{R}^{\mathrm{H}}\mathbf{R}\mathrm{diag}(\mathbf{B})\right)\\
&= \frac{1}{K^2}\mathrm{Tr}\left(\mathbf{D}_0^{1/2}\mathrm{diag}(\mathbf{Q})\mathbf{D}_0^{1/2}\mathbf{R}^{\mathrm{H}}\mathbf{R}\mathbf{D}_0^{1/2}\mathrm{diag}(\mathbf{Q})\mathbf{D}_0^{1/2}\right)\\
&\le \frac{1}{K^2}\|\mathbf{D}_0\|^2\|\mathbf{Q}\|^2\mathrm{Tr}(\mathbf{R}^{\mathrm{H}}\mathbf{R}) \le \frac{1}{K^2}\|\mathbf{D}_0\|^2\|\mathbf{Q}\|^2\mathrm{Tr}(\mathbf{B}^2) \le \frac{1}{K^2}\|\mathbf{D}_0\|^4\|\mathbf{Q}\|^2\mathrm{Tr}(\mathbf{Q}^2)\\
&\le \frac{1}{K}\|\mathbf{D}_0\|^2\|\mathbf{Q}\|^4 \le \frac{1}{K}\frac{\sigma_{\max}^4}{\rho^4} \xrightarrow[K\to\infty]{} 0 .
\end{aligned}$$

In particular, $\mathbb{E}|\Gamma_K|^2\to 0$ and

$$\Re\left(\left(\mathbb{E}W_{1,0}^*|W_{1,0}|^2\right)\Gamma_K\right)\xrightarrow[K\to\infty]{} 0 \quad \text{in probability} . \quad \text{(B.19)}$$

Consider now the second term of the right-hand side of Eq. (B.18). We prove that:

$$\frac{1}{K}\mathbf{w}_0^{\mathrm{H}}\mathbf{R}\mathbf{R}^{\mathrm{H}}\mathbf{w}_0 - \frac{1}{K}\mathrm{Tr}(\mathbf{R}\mathbf{R}^{\mathrm{H}}) \xrightarrow[K\to\infty]{} 0 \quad \text{in probability.} \tag{B.20}$$

By lemma A.1 ((A.2)) in appendix A.1

$$\mathbb{E}\left(\frac{1}{K}\mathbf{w}_0^{\mathrm{H}}\mathbf{R}\mathbf{R}^{\mathrm{H}}\mathbf{w}_0 - \frac{1}{K}\mathrm{Tr}(\mathbf{R}\mathbf{R}^{\mathrm{H}})\right)^2 \leq \frac{C}{K^2}(\mathbb{E}|W_{1,0}|^4)\mathrm{Tr}(\mathbf{R}\mathbf{R}^{\mathrm{H}}\mathbf{R}\mathbf{R}^{\mathrm{H}}) \ .$$

Notice that $\mathrm{Tr}(\mathbf{R}\mathbf{R}^{\mathrm{H}}\mathbf{R}\mathbf{R}^{\mathrm{H}}) = \|\mathbf{R}\|_4^4$ where $\|\mathbf{R}\|_4$ is the Schatten ℓ_4-norm of $\mathbf{R}$. Using Lemma B.2, we have:

$$\|\mathbf{R}\|_4^4 \leq C\|\mathbf{D}_0^{1/2}\mathbf{Q}\mathbf{D}_0^{1/2}\|_4^4 \leq NC\|\mathbf{D}_0^{1/2}\mathbf{Q}\mathbf{D}_0^{1/2}\|^4 \leq N\frac{C\sigma_{\max}^8}{\rho^4} \ .$$

Therefore,

$$\mathbb{E}\left(\frac{1}{K}\mathbf{w}_0^{\mathrm{H}}\mathbf{R}\mathbf{R}^{\mathrm{H}}\mathbf{w}_0 - \frac{1}{K}\mathrm{Tr}(\mathbf{R}\mathbf{R}^{\mathrm{H}})\right)^2 \leq C\frac{N}{K^2} \xrightarrow[K\to\infty]{} 0$$

which implies (B.20). Now, due to the fact that $\mathbf{B} = \mathbf{B}^{\mathrm{H}}$, we have

$$\begin{aligned}
\frac{2}{K}\mathrm{Tr}\,\mathbf{R}\mathbf{R}^{\mathrm{H}} &= \frac{2}{K}\sum_{n=1}^{N}\sum_{\ell=n+1}^{N}|b_{n\ell}|^2 \\
&= \frac{1}{K}\sum_{n,\ell=1}^{N}|b_{n\ell}|^2 - \frac{1}{K}\sum_{n=1}^{N}|b_{nn}|^2 \\
&= \frac{1}{K}\mathrm{Tr}\,\mathbf{D}_0\mathbf{Q}\mathbf{D}_0\mathbf{Q} - \frac{1}{K}\mathrm{Tr}\,\mathbf{D}_0^2(\mathrm{diag}(\mathbf{Q}))^2
\end{aligned} \tag{B.21}$$

Gathering (B.18–B.21), we obtain (B.10). Step 2 is proved.

Step 3: Proof of (B.11) **and** (B.12) We begin with some identities. Write $\mathbf{Q}(z) = [q_{ij}(z)]_{i,j=1}^N$ and $\widetilde{\mathbf{Q}}(z) = [\tilde{q}_{ij}(z)]_{i,j=1}^K$. Denote by $\mathbf{y}_k$ the column number k of $\mathbf{Y}$ and by $\boldsymbol{\xi}_n$ the row number n of $\mathbf{Y}$. Denote by $\mathbf{Y}^k$ the matrix that remains after deleting column k from $\mathbf{Y}$ and by $\mathbf{Y}_n$ the matrix that remains after deleting row n from $\mathbf{Y}$. Finally, write $\mathbf{Q}_k(z) = (\mathbf{Y}^k\big(\mathbf{Y}^k\big)^{\mathrm{H}} - z\mathbf{I})^{-1}$ and $\widetilde{\mathbf{Q}}_n(z) = (\mathbf{Y}_n^{\mathrm{H}}\mathbf{Y}_n - z\mathbf{I})^{-1}$. The following formulas can be established easily (see for instance [58, §0.7.3. and §0.7.4]):

$$q_{nn}(-\rho) = \frac{1}{\rho(1+\boldsymbol{\xi}_n\widetilde{\mathbf{Q}}_n(-\rho)\boldsymbol{\xi}_n^{\mathrm{H}})}, \quad \tilde{q}_{kk}(-\rho) = \frac{1}{\rho(1+\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k(-\rho)\mathbf{y}_k)}, \tag{B.22}$$

$$\mathbf{Q} = \mathbf{Q}_k - \frac{\mathbf{Q}_k\mathbf{y}_k\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k}{1+\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k\mathbf{y}_k} \tag{B.23}$$

Working out (B.23), it is possible to show the asymptotic equivalence between $\frac{1}{K}\mathrm{Tr}(\mathbf{D}_0^2\,(\mathrm{diag}(\mathbf{Q})))^2$ and $\frac{1}{K}\mathrm{Tr}\mathbf{D}_0^2\mathbf{T}^2$, but before going further, we shall state the following lemma which will be proved in section B.1.5.

Lemma B.6. *The following hold true:*

1. *(*Rank one perturbation inequality*) The resolvent* $\mathbf{Q}_k(-\rho)$ *satisfies* $|\mathrm{Tr}\mathbf{A}(\mathbf{Q}-\mathbf{Q}_k)| \leq \|\mathbf{A}\|/\rho$ *for any* $N \times N$ *matrix* $\mathbf{A}$*.*

2. *Let Assumptions* **A1**–**A3** *hold. Then,*

$$\max_{1\leq n\leq N} \mathbb{E}(q_{nn}(-\rho) - t_n(-\rho))^2 \leq \frac{C}{K} \ . \tag{B.24}$$

The same conclusion holds true if q_{nn} *and* t_n *are replaced with* $\tilde{q}_{kk}$ *and* $\tilde{t}_k$ *respectively.*

We are now in position to prove (B.11). First, notice that:

$$\begin{aligned} \mathbb{E}\left|q_{nn}^2 - t_n^2\right| &= \mathbb{E}\left|q_{nn} - t_n\right|\left(q_{nn} + t_n\right) \\ &\leq \sqrt{\mathbb{E}(q_{nn}-t_n)^2}\sqrt{\mathbb{E}(q_{nn}+t_n)^2} \quad \leq \quad \frac{2}{\rho}\sqrt{\mathbb{E}(q_{nn}-t_n)^2} \ . \end{aligned} \tag{B.25}$$

Now,

$$\begin{aligned} \frac{1}{K}\mathbb{E}\left|\mathrm{Tr}\,\mathbf{D}_0^2(\mathrm{diag}(\mathbf{Q})^2 - \mathbf{T}^2)\right| &\leq \frac{1}{K}\sum_{n=1}^N \sigma_{0,n}^4 \mathbb{E}\left|q_{nn}^2 - t_n^2\right| \quad \leq \quad \frac{\sigma_{\max}^4 N}{K}\max_{1\leq n\leq N}\mathbb{E}\left|q_{nn}^2 - t_n^2\right| \\ &\leq \frac{2\sigma_{\max}^4 N}{\rho K}\sqrt{\max_{1\leq n\leq N}\mathbb{E}(q_{nn}-t_n)^2} \quad \xrightarrow[K\to\infty]{} \quad 0 \ , \end{aligned}$$

where the last inequality follows from (B.25) together with Lemma B.6–(2). Convergence (B.11) is established.

We now establish the system of equations (B.12). Our starting point is the identity

$$\mathbf{Q} = \mathbf{T} + \mathbf{T}(\mathbf{T}^{-1} - \mathbf{Q}^{-1})\mathbf{Q} = \mathbf{T} + \frac{\rho}{K}\mathbf{T}\,\mathrm{diag}(\mathrm{Tr}\widetilde{\mathbf{D}}_1\widetilde{\mathbf{T}}, \ldots, \mathrm{Tr}\widetilde{\mathbf{D}}_N\widetilde{\mathbf{T}})\mathbf{Q} - \mathbf{T}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\mathbf{Q} \ .$$

Using this identity, we develop $U_\ell = \frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{Q}\mathbf{D}_\ell\mathbf{Q}$ as

$$\begin{aligned} U_\ell &= \frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{Q}\mathbf{D}_\ell\mathbf{T} + \frac{\rho}{K^2}\mathrm{Tr}\mathbf{D}_0\mathbf{Q}\mathbf{D}_\ell\mathbf{T}\mathrm{diag}(\mathrm{Tr}\widetilde{\mathbf{D}}_1\widetilde{\mathbf{T}}, \ldots, \mathrm{Tr}\widetilde{\mathbf{D}}_N\widetilde{\mathbf{T}})\mathbf{Q} - \frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{Q}\mathbf{D}_\ell\mathbf{T}\mathbf{Y}\mathbf{Y}^{\mathrm{H}}\mathbf{Q} \\ &\triangleq X_1 + X_2 - X_3 \ . \end{aligned} \tag{B.26}$$

Lemma B.3–(2) with $\mathbf{S} = \mathbf{D}_0\mathbf{D}_\ell\mathbf{T}$ yields:

$$X_1 = \frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{D}_\ell\mathbf{T}^2 + \epsilon_1 \tag{B.27}$$

where $\mathbb{E}|\epsilon_1| \leq \sqrt{\mathbb{E}\epsilon_1^2} \leq C/K$. Consider now the term $X_3 = \frac{1}{K}\sum_{k=1}^K \mathrm{Tr}\mathbf{D}_0\mathbf{Q}\mathbf{D}_\ell\mathbf{T}\mathbf{y}_k\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}$. Using (B.22) and (B.23), we have

$$\mathbf{y}_k^{\mathrm{H}}\mathbf{Q} = \left(1 - \frac{\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}\mathbf{y}_k}{1+\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}\mathbf{y}_k}\right)\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k = \rho\,\tilde{q}_{kk}\,\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k \ .$$

Hence

$$\begin{aligned} X_3 &= \frac{\rho}{K}\sum_{k=1}^K \tilde{q}_{kk}\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k\mathbf{D}_0\mathbf{Q}\mathbf{D}_\ell\mathbf{T}\mathbf{y}_k \\ &= \frac{\rho}{K}\sum_{k=1}^K \tilde{t}_k\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k\mathbf{D}_0\mathbf{Q}\mathbf{D}_\ell\mathbf{T}\mathbf{y}_k + \frac{\rho}{K}\sum_{k=1}^K(\tilde{q}_{kk}-\tilde{t}_k)\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k\mathbf{D}_0\mathbf{Q}\mathbf{D}_\ell\mathbf{T}\mathbf{y}_k \\ &\triangleq X_3' + \epsilon_2 \ . \end{aligned} \tag{B.28}$$

By Cauchy-Schwarz inequality,

$$\mathbb{E}|\epsilon_2| \leq \frac{\rho}{K}\sum_{k=1}^{K}\sqrt{\mathbb{E}(\tilde{q}_{kk}-\tilde{t}_k)^2}\sqrt{\mathbb{E}(\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k\mathbf{D}_0\mathbf{Q}\mathbf{D}_\ell\mathbf{T}\mathbf{y}_k)^2}\ .$$

We have $\mathbb{E}(\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k\mathbf{D}_0\mathbf{Q}\mathbf{D}_\ell\mathbf{T}\mathbf{y}_k)^2 \leq \sigma_{\max}^8\rho^{-6}\mathbb{E}\|\mathbf{y}_k\|^4 \leq C$. Using in addition Lemma B.6–(2), we obtain

$$\mathbb{E}|\epsilon_2| \leq \frac{C}{\sqrt{K}}\ .$$

Consider X_3'. From (B.22) and (B.23), we have $\mathbf{Q} = \mathbf{Q}_k - \rho\tilde{q}_{kk}\mathbf{Q}_k\mathbf{y}_k\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k$. Hence, we can develop X_3' as

$$\begin{aligned} X_3' &= \frac{\rho}{K}\sum_{k=1}^{K}\tilde{t}_k\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k\mathbf{D}_0\mathbf{Q}_k\mathbf{D}_\ell\mathbf{T}\mathbf{y}_k - \frac{\rho^2}{K}\sum_{k=1}^{K}\tilde{t}_k\tilde{q}_{kk}\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k\mathbf{D}_0\mathbf{Q}_k\mathbf{y}_k\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k\mathbf{D}_\ell\mathbf{T}\mathbf{y}_k \\ &\triangleq X_4 + X_5\ . \end{aligned} \tag{B.29}$$

Consider X_4. Notice that $\mathbf{y}_k$ and $\mathbf{Q}_k$ are independent. Therefore, by Lemma A.1 in appendix A.1, we obtain

$$\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k\mathbf{D}_0\mathbf{Q}_k\mathbf{D}_\ell\mathbf{T}\mathbf{y}_k = \frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{Q}_k\mathbf{D}_0\mathbf{Q}_k\mathbf{D}_\ell\mathbf{T} + \epsilon_3 = \frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{Q}\mathbf{D}_0\mathbf{Q}\mathbf{D}_\ell\mathbf{T} + \epsilon_3 + \epsilon_4$$

where $\mathbb{E}\epsilon_3^2 < CK^{-1}$ by Ineq. (A.3) in appendix A.1. Applying twice Lemma B.6–(1) to $\epsilon_4 = \frac{1}{K}(\mathrm{Tr}\mathbf{D}_k\mathbf{Q}_k\mathbf{D}_0\mathbf{Q}_k\mathbf{D}_\ell\mathbf{T} - \mathrm{Tr}\mathbf{D}_k\mathbf{Q}\mathbf{D}_0\mathbf{Q}\mathbf{D}_\ell\mathbf{T})$ yields $|\epsilon_4| < CK^{-1}$. Note in addition that $\sum\tilde{t}_k\mathbf{D}_k = \mathrm{diag}(\mathrm{Tr}\widetilde{\mathbf{D}}_1\widetilde{\mathbf{T}},\ldots,\mathrm{Tr}\widetilde{\mathbf{D}}_N\widetilde{\mathbf{T}})$. Thus, we obtain

$$\begin{aligned} X_4 &= \frac{\rho}{K^2}\mathrm{Tr}\left(\sum_{k=1}^{K}\tilde{t}_k\mathbf{D}_k\right)\mathbf{Q}\mathbf{D}_0\mathbf{Q}\mathbf{D}_\ell\mathbf{T} + \epsilon_5 \\ &= X_2 + \epsilon_5\ , \end{aligned} \tag{B.30}$$

where $\epsilon_5 = \epsilon_3 + \epsilon_4$, which yields $\mathbb{E}|\epsilon_5| \leq CK^{-\frac{1}{2}}$.
We now turn to X_5. First introduce the following random variable:

$$\epsilon_6 = \tilde{t}_k\tilde{q}_{kk}\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k\mathbf{D}_0\mathbf{Q}_k\mathbf{y}_k\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k\mathbf{D}_\ell\mathbf{T}\mathbf{y}_k - \tilde{t}_k\tilde{q}_{kk}\left(\frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{Q}_k\mathbf{D}_0\mathbf{Q}_k\right)\left(\frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{Q}_k\mathbf{D}_\ell\mathbf{T}\right)$$

Then

$$\begin{aligned} |\epsilon_6| \leq \frac{1}{\rho^2}\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k\mathbf{D}_0\mathbf{Q}_k\mathbf{y}_k\left|\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k\mathbf{D}_\ell\mathbf{T}\mathbf{y}_k - \frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{Q}_k\mathbf{D}_\ell\mathbf{T}\right| \\ + \frac{1}{\rho^2}\left|\mathbf{y}_k^{\mathrm{H}}\mathbf{Q}_k\mathbf{D}_0\mathbf{Q}_k\mathbf{y}_k^{\mathrm{H}} - \frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{Q}_k\mathbf{D}_0\mathbf{Q}_k\right|\frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{Q}_k\mathbf{D}_\ell\mathbf{T} \end{aligned}$$

and one can prove that $\mathbb{E}|\epsilon_6| < CK^{-\frac{1}{2}}$ with help of Lemma A.1 in appendix A.1, together with Cauchy-Schwarz inequality. In addition, we can prove with the help of Lemma B.6 that:

$$\begin{aligned} \tilde{t}_k\tilde{q}_{kk}\left(\frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{Q}_k\mathbf{D}_0\mathbf{Q}_k\right)\left(\frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{Q}_k\mathbf{D}_\ell\mathbf{T}\right) &= \tilde{t}_k^2\left(\frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{Q}\mathbf{D}_0\mathbf{Q}\right)\left(\frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{Q}\mathbf{D}_\ell\mathbf{T}\right) + \epsilon_7 \\ &= \tilde{t}_k^2\left(\frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{Q}\mathbf{D}_0\mathbf{Q}\right)\left(\frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{D}_\ell\mathbf{T}^2\right) + \epsilon_7 + \epsilon_8 \end{aligned}$$

where ϵ_7 and ϵ_8 are random variables satisfying $\mathbb{E}|\epsilon_7| < CK^{-\frac{1}{2}}$ by Lemma B.6, and $\max_{k,\ell} \mathbb{E}|\epsilon_8| \leq \max_{k,\ell} \sqrt{\mathbb{E}|\epsilon_8|^2} \leq CK^{-\frac{1}{2}}$ by Lemma B.3–(2). Using the fact that $\rho^2 \tilde{t}_k^2 = (1 + \frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{T})^{-2}$, we end up with

$$X_5 = -\frac{\rho^2}{K}\sum_{k=1}^{K} \tilde{t}_k^2 \left(\frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{Q}\mathbf{D}_0\mathbf{Q}\right)\left(\frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{D}_\ell\mathbf{T}^2\right) + \epsilon_9 = -\sum_{k=1}^{K} c_{\ell k}U_k + \epsilon_9 \tag{B.31}$$

where $c_{\ell k}$ is given by (B.13), and where $\mathbb{E}|\epsilon_9| < CK^{-\frac{1}{2}}$.

Plugging Eq. (B.27)–(B.31) into (B.26), we end up with $U_\ell = \sum_{k=1}^{K} c_{\ell k}U_k + \frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{D}_\ell\mathbf{T}^2 + \epsilon$ with $\mathbb{E}|\epsilon| < CK^{-\frac{1}{2}}$. Step 3 is established.

Step 4 : Proof of (B.14) We rely on results of Section B.1.2, in particular on Lemma B.4.

Define the following $(K+1) \times 1$ vectors:

$$\mathbf{u} = [U_k]_{k=0}^K,\ \mathbf{d} = \left[\frac{1}{K}\mathrm{Tr}\mathbf{D}_0\mathbf{D}_k\mathbf{T}^2\right]_{k=0}^K,\ \boldsymbol{\epsilon} = [\epsilon_k]_{k=0}^K\ ,$$

where the U_k's and ϵ_k's are defined in (B.12). Recall the definition of the $c_{\ell k}$'s for $0 \leq \ell \leq K$ and $1 \leq k \leq K$, define $c_{\ell 0} = 0$ for $0 \leq \ell \leq K$ and consider the $(K+1) \times (K+1)$ matrix $\mathbf{C} = [c_{\ell k}]_{\ell,k=0}^K$.

With these notations, System (B.12) writes

$$(\mathbf{I}_{K+1} - \mathbf{C})\,\mathbf{u} = \mathbf{d} + \boldsymbol{\epsilon}\ . \tag{B.32}$$

Let $\boldsymbol{\alpha} = \frac{1}{K}\mathrm{Tr}\,\mathbf{D}_0^2\mathbf{T}^2$ and $\boldsymbol{\beta} = (1 + \frac{1}{K}\mathrm{Tr}\,\mathbf{D}_0\mathbf{T})^2$. We have in particular

$$\mathbf{d} = \begin{bmatrix} \boldsymbol{\alpha} \\ \mathbf{g} \end{bmatrix},\quad \mathbf{C} = \begin{bmatrix} 0 & \frac{1}{K}\mathbf{g}^T\boldsymbol{\Delta}^{-1} \\ \mathbf{0} & \mathbf{A}^T \end{bmatrix}$$

(recall that $\mathbf{A}$, $\boldsymbol{\Delta}$ and $\mathbf{g}$ are defined in the statement of Theorem 4.3).

Consider a square matrix $\mathbf{X}$ which first column is equal to $[1, 0, \ldots, 0]^{\mathrm{T}}$, and partition $\mathbf{X}$ as $\mathbf{X} = \begin{bmatrix} 1 & \mathbf{x}_{01}^T \\ \mathbf{0} & \mathbf{X}_{11} \end{bmatrix}$. Recall that the inverse of $\mathbf{X}$ exists if and only if $\mathbf{X}_{11}^{-1}$ exists, and in this case the first row $[\mathbf{X}^{-1}]_0$ of $\mathbf{X}^{-1}$ is given by

$$\left[\mathbf{X}^{-1}\right]_0 = \begin{bmatrix} 1 & -\mathbf{x}_{01}^T\mathbf{X}_{11}^{-1} \end{bmatrix}$$

(see for instance [58]). We now apply these results to the system (B.32). Due to (B.32), U_0 can be expressed as

$$U_0 = [(\mathbf{I} - \mathbf{C})^{-1}]_0(\mathbf{d} + \boldsymbol{\epsilon})\ .$$

By Lemma B.4–(1), $(\mathbf{I}_K - \mathbf{A}^T)^{-1}$ exists hence $(\mathbf{I} - \mathbf{C})^{-1}$ exists,

$$\left[(\mathbf{I}_{K+1} - \mathbf{C})^{-1}\right]_0 = \left[1 \quad \frac{1}{K}\mathbf{g}^T\boldsymbol{\Delta}^{-1}(\mathbf{I}_K - \mathbf{A}^T)^{-1}\right]\ ,$$

and

$$U_0 = \boldsymbol{\alpha} + \frac{1}{K}\mathbf{g}^T\boldsymbol{\Delta}^{-1}\left(\mathbf{I} - \mathbf{A}^T\right)^{-1}\mathbf{g} + \epsilon_0 + \frac{1}{K}\mathbf{g}^T\boldsymbol{\Delta}^{-1}\left(\mathbf{I} - \mathbf{A}^T\right)^{-1}\boldsymbol{\epsilon}'$$

with $\boldsymbol{\epsilon}' = [\epsilon_1, \ldots, \epsilon_K]^T$. Gathering the estimates of Section B.1.2 together with the fact that $\|\mathbb{E}\boldsymbol{\epsilon}\|_\infty \leq CK^{-\frac{1}{2}}$, we get (B.14). Step 4 is established, so is Theorem 4.3.

B.1.4 Proof of Lemma B.3

Let us establish (B.1). The lower bound immediately follows from the representation

$$t_n = \frac{1}{\rho + \frac{1}{K}\sum_{k=1}^K \frac{\sigma_{nk}^2}{1+\frac{1}{K}\sum_{\ell=1}^N \sigma_{\ell k}^2 t_\ell}} \overset{(a)}{\geq} \frac{1}{\rho + \sigma_{\max}^2}$$

where (a) follows from **A2** and $t_\ell(-\rho) \geq 0$. The upper bound requires an extra argument: As proved in [54, Theorem 2.4], the t_n's are Stieltjes transforms of probability measures supported by $\mathbb{R}_+$, i.e. there exists a probability measure μ_n over $\mathbb{R}_+$ such that $t_n(z) = \int \frac{\mu_n(dt)}{t-z}$. Thus

$$t_n(-\rho) = \int_0^\infty \frac{\mu_n(dt)}{t+\rho} \leq \frac{1}{\rho} \ ,$$

and (B.1) is proved.

We now briefly justify (B.2). We have $\mathbb{E}\,|\mathrm{Tr}\,\mathbf{S}(\mathbf{Q}-\mathbf{T})|^2 = \mathbb{E}\,|\mathrm{Tr}\mathbf{S}(\mathbf{Q}-\mathbb{E}\mathbf{Q})|^2 + |\mathrm{Tr}\mathbf{S}(\mathbb{E}\mathbf{Q}-\mathbf{T})|^2$. In [55, Lemma 6.3] it is stated that $\sup_K \mathbb{E}\,|\mathrm{Tr}\mathbf{S}(\mathbf{Q}-\mathbb{E}\mathbf{Q})|^2 < \infty$. Furthermore, in the proof of [55, Theorem 3.3] it is shown that $\sup_K K\|\mathbb{E}\mathbf{Q}-\mathbf{T}\| < \infty$, hence $|\mathrm{Tr}\mathbf{S}(\mathbb{E}\mathbf{Q}-\mathbf{T})| \leq K\|\mathbf{S}(\mathbb{E}\mathbf{Q}-\mathbf{T})\| \leq K\|\mathbb{E}\mathbf{Q}-\mathbf{T}\|\|\mathbf{S}\| < \infty$ by Lemma B.1–(2). The result follows.

B.1.5 Proof of Lemma B.6

The proof of Part 1 can be found in [55, Proof of Lemma 6.3] (see also [93, Lemma 2.6]). Let us prove Part 2. Using the system of equations (3.2) defining matrices $\mathbf{T}$ and $\widetilde{\mathbf{T}}$ and (B.22), we get:

$$\begin{aligned} |q_{nn}(-\rho) - t_n(-\rho)| &= \frac{1}{\rho(1+\frac{1}{K}\mathrm{Tr}\widetilde{\mathbf{D}}_n\widetilde{\mathbf{T}})(1+\boldsymbol{\xi}_n\widetilde{\mathbf{Q}}_n\boldsymbol{\xi}_n^*)}\left|\boldsymbol{\xi}_n\widetilde{\mathbf{Q}}_n\boldsymbol{\xi}_n^* - \frac{1}{K}\mathrm{Tr}\widetilde{\mathbf{D}}_n\widetilde{\mathbf{T}}\right| \\ &\leq \frac{1}{\rho}\left|\boldsymbol{\xi}_n\widetilde{\mathbf{Q}}_n\boldsymbol{\xi}_n^* - \frac{1}{K}\mathrm{Tr}\widetilde{\mathbf{D}}_n\widetilde{\mathbf{T}}\right| \ . \end{aligned}$$

Hence,

$$\mathbb{E}(q_{nn}-t_n)^2 \leq \frac{2}{\rho}\mathbb{E}\left(\boldsymbol{\xi}_n\widetilde{\mathbf{Q}}_n\boldsymbol{\xi}_n^* - \frac{1}{K}\mathrm{Tr}\widetilde{\mathbf{D}}_n\widetilde{\mathbf{Q}}\right)^2 + \frac{2}{\rho K^2}\mathbb{E}\left(\mathrm{Tr}\widetilde{\mathbf{D}}_n(\widetilde{\mathbf{Q}}-\widetilde{\mathbf{T}})\right)^2 \leq \frac{C}{K}$$

by Lemma A.1 and Lemma B.3–(2), which proves (B.24).

B.2 Proof of Corollary 4.1

Recall that in the separable case, $\mathbf{D}_k = \tilde{d}_k\mathbf{D}$ and $\widetilde{\mathbf{D}}_n = d_n\widetilde{\mathbf{D}}$. Let $\tilde{\mathbf{d}}$ be the $K\times 1$ vector $\tilde{\mathbf{d}} = [\tilde{d}_k]_{k=1}^K$. In the separable case, Eq. (4.15) is written

$$\frac{\Theta^2}{\tilde{d}_0^2} = \frac{1}{K\tilde{d}_0^2}\mathbf{g}^T(\mathbf{I}-\mathbf{A})^{-1}\boldsymbol{\Delta}^{-1}\mathbf{g} + \gamma(\mathbb{E}|W_{1,0}|^4 - 1) \ , \tag{B.33}$$

where γ is defined in statement of the corollary. Here, vector $\mathbf{g}$ and matrix $\mathbf{A}$ are given by

$$\mathbf{g} = \gamma\tilde{d}_0\tilde{\mathbf{d}} \quad \text{and} \quad \mathbf{A} = \left[\frac{1}{K}\frac{\frac{1}{K}\mathrm{Tr}\mathbf{D}_\ell\mathbf{D}_m\mathbf{T}^2}{\left(1+\frac{1}{K}\mathrm{Tr}\mathbf{D}_\ell\mathbf{T}\right)^2}\right]_{\ell,m=1}^K = \frac{\gamma}{K}\boldsymbol{\Delta}^{-1}\tilde{\mathbf{d}}\tilde{\mathbf{d}}^T \ .$$

By the matrix inversion lemma [58], we have

$$\begin{aligned}\frac{1}{K\tilde{d}_0^2}\mathbf{g}^T(\mathbf{I}-\mathbf{A})^{-1}\mathbf{\Delta}^{-1}\mathbf{g} &= \frac{\gamma^2}{K}\tilde{\mathbf{d}}^T\left(\mathbf{\Delta}-\frac{\gamma}{K}\tilde{\mathbf{d}}\tilde{\mathbf{d}}^T\right)^{-1}\tilde{\mathbf{d}} \\ &= \frac{\gamma^2}{K}\tilde{\mathbf{d}}^T\left(\mathbf{\Delta}^{-1}+\frac{\gamma}{K}\frac{1}{1-\frac{\gamma}{K}\tilde{\mathbf{d}}^T\mathbf{\Delta}^{-1}\tilde{\mathbf{d}}}\mathbf{\Delta}^{-1}\tilde{\mathbf{d}}\tilde{\mathbf{d}}^T\mathbf{\Delta}^{-1}\right)\tilde{\mathbf{d}}\ .\end{aligned}$$

Noticing that

$$\frac{1}{K}\tilde{\mathbf{d}}^T\mathbf{\Delta}^{-1}\tilde{\mathbf{d}}=\frac{1}{K}\sum_{k=1}^{K}\frac{\tilde{d}_k^2}{\left(1+\frac{1}{K}\mathrm{Tr}\mathbf{D}_k\mathbf{T}\right)^2}=\frac{\rho^2}{K}\sum_{k=1}^{K}\tilde{d}_k^2\tilde{t}_k^2=\rho^2\tilde{\gamma}\ ,$$

we obtain

$$\frac{1}{K\tilde{d}_0^2}\mathbf{g}^T(\mathbf{I}-\mathbf{A})^{-1}\mathbf{\Delta}^{-1}\mathbf{g}=\gamma\frac{\rho^2\gamma\tilde{\gamma}}{1-\rho^2\gamma\tilde{\gamma}}\ .$$

Plugging this equation into (B.33), we obtain (4.17).

Chapter 5

Performance analysis for LMMSE detectors on general correlated MIMO channels

As we have seen in the previous chapter, the Gaussian approximation of the SINR leads to accurate 1% outage probabilites only in the case where the system dimensions are beyond 64. Such situation may be encountered in the case of CDMA based systems, but is hardly experienced when dealing with MIMO systems. The asymptotic BER based on the sole Gaussian approximation is also inaccurate for small system sizes and tends to under-estimate the experienced real value, [30]. The reasons behind this is that the Gaussian distribution allows negative values and has a zero central third moment, while the SINR is always positive and has a non zero third moment.

A more precise approximation of the BER or the outage probability is expected if one chooses to approximate the SINR probability distribution with a distribution 1) which is supported by $\mathbb{R}_+$ and 2) which is adjusted to the first three moments of the SINR instead of the first tow moments needed by the Gaussian approximation.

In this line of thought, Li, Paul, Narasimhan and Cioffi [82] proposed to use alternative parametrized distributions (Gamma and Generalized Gamma distributions) whose parameters are set to coincide with the asymptotic moments of the output SINR. This approach was derived for (transmit) correlated channels and asymptotic moments were provided for the special case of uncorrelated or equi-correlated channels. For the general correlated channel case, only limiting upper bounds for the first three asymptotic moments were provided. Based on Random Matrix Theory, and especially on the Gaussian mathematical tools elaborated in [53] and further used in [34], we derive closed-form expressions for the first three moments, generalizing the work of [82] to a general (receive) correlated channel. Using the generalized Gamma approximation, we provide closed-form expressions for the BER and numerical approximations for the outage probability.

We organize this chapter as follows: First, we present the system model and derive the SINR expression. Then, we review the Generalized Gamma approximation before providing the asymptotic central moments in the next section. Finally, we discuss in the last section the simulation results.

5.1 System Model

We consider an uplink transmission system, in which a base station equipped with N correlated antennas detects the symbols of a given user of interest in the presence of K interfering users. The N dimensional received signal writes:

$$\mathbf{r} = \mathbf{\Sigma}\mathbf{s} + \mathbf{n}$$

where $\mathbf{s} = [s_0, \cdots, s_K]^{\mathrm{T}}$ is the transmitted complex vector signal with size $K+1$ satisfying $\mathbb{E}\mathbf{s}\mathbf{s}^{\mathrm{H}} = \mathbf{I}_{K+1}$, and $\mathbf{\Sigma}$ is the $N \times (K+1)$ channel matrix. We assume that this matrix writes as

$$\mathbf{\Sigma} = \frac{1}{\sqrt{K}} \mathbf{\Psi}^{\frac{1}{2}} \mathbf{W} \mathbf{P}^{\frac{1}{2}},$$

where $\mathbf{\Psi}$ a $N \times N$ Hermitian nonnegative matrix that captures the correlations at the receiver, $\mathbf{P} = \mathrm{diag}\,(p_0, \cdots, p_K)$ is the deterministic matrix of the powers allocated to the different users and $\mathbf{W} = [\mathbf{w}_0, \cdots, \mathbf{w}_K]$ ($\mathbf{w}_k$ being the kth column) is a $N \times (K+1)$ complex Gaussian matrix with centered unit variance (standard) independent and identically distributed (i.i.d) entries.

To detect symbol s_0 and to mitigate the interference caused by users $1, \ldots, K$, the base station applies the LMMSE estimator. Along the same lines as in the previous chapter, one can prove that the experienced SINR is given by:

$$\beta_K = \mathbf{y}^{\mathrm{H}} \left(\frac{1}{K} \mathbf{\Psi}^{\frac{1}{2}} \widetilde{\mathbf{W}} \widetilde{\mathbf{P}} \widetilde{\mathbf{W}}^{\mathrm{H}} \mathbf{\Psi}^{\frac{1}{2}} + \rho \mathbf{I}_N \right)^{-1} \mathbf{y},$$

with $\widetilde{\mathbf{P}} = \mathrm{diag}\,(p_1, \cdots, p_K)$ and $\widetilde{\mathbf{W}} = [\mathbf{w}_1, \cdots, \mathbf{w}_K]$. Let $\mathbf{\Psi} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}$ be a spectral decomposition of $\mathbf{\Psi}$. Then, β_K writes:

$$\begin{aligned} \beta_K &= \frac{p_0}{K} \mathbf{w}_0^{\mathrm{H}} \mathbf{U} \mathbf{D}^{\frac{1}{2}} \left(\frac{1}{K} \mathbf{D}^{\frac{1}{2}} \mathbf{U}^{\mathrm{H}} \widetilde{\mathbf{W}} \widetilde{\mathbf{P}} \widetilde{\mathbf{W}}^{\mathrm{H}} \mathbf{U} \mathbf{D}^{\frac{1}{2}} + \rho \mathbf{I}_N \right)^{-1} \mathbf{D}^{\frac{1}{2}} \mathbf{U}^{\mathrm{H}} \mathbf{w}_0 \ , \\ &= \frac{p_0}{\rho K} \mathbf{z}^{\mathrm{H}} \mathbf{D}^{\frac{1}{2}} \left(\frac{1}{K\rho} \mathbf{D}^{\frac{1}{2}} \mathbf{Z} \tilde{\mathbf{D}} \mathbf{Z}^{\mathrm{H}} \mathbf{D}^{\frac{1}{2}} + \mathbf{I} \right)^{-1} \mathbf{D}^{\frac{1}{2}} \mathbf{z} \end{aligned}$$

where: $\mathbf{z} = \mathbf{U}^{\mathrm{H}} \mathbf{w}_0$ (resp. $\mathbf{Z} = \mathbf{U}^{\mathrm{H}} \widetilde{\mathbf{W}}$) is a $N \times 1$ vector with complex independent standard Gaussian entries (resp. $N \times K$ matrix with independent Gaussian entries).

The asymptotic behaviour of the SINR β_K has been studied in the previous chapter. The first and second deterministic moments, have been shown to admit deterministic approximations as $K, N \to \infty$, the ratio being bounded below by a positive constant and above by a finite constant. As it will be shown later, a deterministic approximation of the third central moment is required for the BER and outage probability approximation.

5.2 Bit Error Rate and outage probability approximations

5.2.1 A quick reminder of the generalised Gamma distribution

Recall that if a random variable X follows a generalized gamma distribution $G(\alpha, b, \xi)$, where α and b are respectively referred to as the shape and scale parameters, then:

$$\mathbb{E}X = \alpha b, \quad \mathrm{var}(X) = \alpha b^2 \quad \text{and} \quad \mathbb{E}(X - \mathbb{E}X)^3 = (\xi + 1)\alpha b^3 \ .$$

The probability density function (pdf) of the generalized Gamma distribution with parameters (α, b, ξ) does not have a closed form expression but its moment generating function (MGF) writes [38]:

$$\mathrm{MGF}(s) = \begin{cases} \exp(\frac{\alpha}{\xi-1}(1-(1-b\xi s)^{\frac{\xi-1}{\xi}})) & \text{if } \xi > 1, \\ (1-sb)^{-\alpha}, \ \ s < \frac{1}{b} & \text{if } \xi = 1, \\ \exp(\frac{\alpha}{1-\xi}((1-b\xi s)^{\frac{\xi-1}{\xi}} - 1)) & \text{if } \xi > 1. \end{cases}$$

5.2.2 BER approximation

Under QPSK constellations with Gray encoding and assuming that the noise at the LMMSE output is Gaussian, the BER is given by:

$$\mathrm{BER} = \mathbb{E}Q(\sqrt{\beta_K})$$

where $Q(x) = \frac{1}{\sqrt{2\pi}} \int_x^\infty e^{-t^2/2}\, dt$ and the expectation is taken over the distribution of the SINR β_K. Based on the asymptotic normality of the SINR, [103] and [86] proposed to use the limiting BER value given by:

$$\mathrm{BER} = \frac{1}{\sqrt{2\pi}} \int_{\sqrt{\overline{\beta}_K}}^\infty e^{-t^2/2} dt,$$

where $\overline{\beta}_K$ denotes an asymptotic deterministic approximation of the first moment of β_K. It was shown however in [82] that this expression is inaccurate since a Gaussian random variable allows negative values and has a zero third moment while the output SINR is always positive and has a non-zero third moment for finite system dimensions. To overcome these difficulties, Li *et al.* [82] approximate the BER by considering first that the SINR follows a Gamma distribution with scale α and shape b, these parameters being tuned by equating the first two moments of the Gamma distribution with the first two asymptotic moments of the SINR. However, the third asymptotic moment was shown to be different from the third moment of the Gamma distribution which only depends on the scale α and shape b. In light of this consideration, Li *et al.* [82] refine this approximation and consider that the SINR follows a generalized Gamma distribution which is adjusted by assuming that its first three moments equalize with the first three asymptotic moments of the SINR. As expected, this approximation has proved to be more accurate than the Gamma approximation, and so will be the one considered in this chapter. Next, we briefly review this technique, which we will rely on to provide accurate approximations for the BER and outage probability.

Let $\mathbb{E}_\infty(\beta_K)$, $\mathrm{var}_\infty(\beta_K)$ and $\mathrm{S}_\infty(\beta_K)$ denote respectively the deterministic approximations of the asymptotic central moments of β_K. Then, the parameters ξ, α and b are determined by solving:

$$\mathbb{E}_\infty(\beta_K) = \alpha b, \quad \mathrm{var}_\infty(\beta_K) = \alpha b^2 \quad \text{and} \quad \mathrm{S}_\infty(\beta_K) = (\xi+1)\alpha b^3,$$

thus giving the following values:

$$\alpha = \frac{(\mathbb{E}_\infty(\beta_K))^2}{\mathrm{var}_\infty(\beta_K)}, \quad \beta = \frac{\mathrm{var}_\infty(\beta_K)}{\mathbb{E}_\infty(\beta_K)} \quad \text{and} \quad \xi = \frac{\mathrm{S}_\infty(\beta_K)\mathbb{E}_\infty(\beta_K)}{(\mathrm{var}_\infty(\beta_K))^2} - 1.$$

Using the MGF, one can evaluate the BER by using the following relation [94], that holds for QPSK constellation:

$$\mathrm{BER} = \frac{1}{\pi} \int_0^{\frac{\pi}{2}} \mathrm{MGF}\left(-\frac{1}{2\sin^2\phi}\right) d\phi. \tag{5.1}$$

Note that similar expressions for the BER exist for other constellations and can be derived by plugging the following identity involving the function $Q(x)$ [94]:

$$Q(x) = \frac{1}{\pi}\int_0^{\frac{\pi}{2}} \exp\left(-\frac{x^2}{2\sin^2\theta}\right) d\theta$$

into the BER expression.

5.2.3 Outage probability approximation

Only the moment generation function (MGF) has a closed form expression. Knowing the MGF, one can compute numerically the cumulative distribution function by applying the saddle point approximation technique [17]. Denote by $K(y) = \log(\mathrm{MGF}(y))$ the cumulative generating function, by y the threshold SINR and by t_y the solution of $K'(t_y) = y$. Let w_0 and u_0 be given by: $w_0 = \mathrm{sign}(t_y)\sqrt{2\,(t_y y - K(t_y))}$ and $u_0 = t_y\sqrt{K"(t_y)}$. The saddle point approximate of the outage probability is given by:

$$\mathrm{P}_{out} = \Phi(w_0) + \phi(w_0)\left(\frac{1}{w_0} - \frac{1}{u_0}\right), \tag{5.2}$$

where $\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-t^2/2}\, dt$ and $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ denote respectively the standard normal cumulative distribution function and probability distribution function.

So far, we have presented the technique that will be used in simulations for the evaluation of the BER and outage probability. This technique is heavily based on the computation of the three first asymptotic moments of the SINR β_K, an issue that is handled in the next section.

5.3 Asymptotic moments

The first and second moments, being computed in the previous chapter, we devote this section to the asymptotic approximation of the third central moment. Our approach relies on the work in [53], from which several notations and results are borrowed. For the reader convenience, we will recall the required assumptions and some modified definitions for the deterministic equivalents.

5.3.1 Assumptions

Recall from Section 5.1 the various definitions $K, N, \mathbf{D}, \widetilde{\mathbf{D}}$. In the following, we assume that both K and N go to $+\infty$, their ratio being bounded below and above as follows:

$$0 \;<\; \ell^- = \liminf \frac{K}{N} \;\leq\; \ell^+ = \limsup \frac{K}{N} \;<\; +\infty\ .$$

In the sequel, the notation $K \to \infty$ will refer to this asymptotic regime. We will frequently write $\mathbf{D}_K$ and $\widetilde{\mathbf{D}}_K$ to emphasize the dependence in K, but may drop the subscript K as well. Assume the following mild conditions:

A1. *There exist real numbers $d_{\max} < \infty$ and $\tilde{d}_{\max} < \infty$ such that:*

$$\sup_K \|\mathbf{D}_K\| \leq d_{\max} \quad \text{and} \quad \sup_K \|\widetilde{\mathbf{D}}_K\| \leq \tilde{d}_{\max},$$

where $\|\mathbf{D}_K\|$ and $\|\widetilde{\mathbf{D}}_K\|$ are the spectral norms of $\mathbf{D}_K$ and $\widetilde{\mathbf{D}}_K$.

A2. *The normalized traces of $\mathbf{D}_K$ and $\widetilde{\mathbf{D}}_K$ satisfy:*

$$\inf_K \frac{1}{K}\mathrm{Tr}(\mathbf{D}_K) > 0 \quad \text{and} \quad \inf_K \frac{1}{K}\mathrm{Tr}(\widetilde{\mathbf{D}}_K) > 0.$$

5.3.2 Asymptotic moments computation

In this section, we provide closed form expressions for the first three asymptotic moments. But, first, we shall recall the definitions for the deterministic quantities that will be considered. Note that these definitions are different from those presented in chapter 3.

Proposition 5.1. *(cf. [53]) For every integer K and any $t > 0$, the system of equations in $(\delta, \tilde{\delta})$*

$$\begin{cases} \delta_K &= \frac{1}{K}\mathrm{Tr}\mathbf{D}_K \left(\mathbf{I} + t\tilde{\delta}_K \mathbf{D_K}\right)^{-1}, \\ \tilde{\delta}_K &= \frac{1}{K}\mathrm{Tr}\widetilde{\mathbf{D}}_K \left(\mathbf{I} + t\delta_K \widetilde{\mathbf{D}}_K\right)^{-1}, \end{cases}$$

admits a unique solution $\left(\delta_K(t), \tilde{\delta}_K(t)\right)$ satisfying $\delta_K(t) > 0$, $\tilde{\delta}_K(t) > 0$.

Let $\mathbf{T}$ and $\widetilde{\mathbf{T}}$ be the $N \times N$ and $K \times K$ diagonal matrices defined by:

$$\mathbf{T} = \left(\mathbf{I} + t\tilde{\delta_K}\mathbf{D}\right)^{-1} \text{ and } \widetilde{\mathbf{T}} = \left(\mathbf{I} + t\delta_K \widetilde{\mathbf{D}}\right)^{-1}.$$

Note that in particular: $\delta = \frac{1}{K}\mathrm{Tr}\mathbf{DT}$ and $\tilde{\delta} = \frac{1}{K}\mathrm{Tr}\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}$. Define also γ and $\tilde{\gamma}$ as $\gamma = \frac{1}{K}\mathrm{Tr}\mathbf{D}^2\mathbf{T}^2$ and $\tilde{\gamma} = \frac{1}{K}\mathrm{Tr}\widetilde{\mathbf{D}}^2\widetilde{\mathbf{T}}^2$. Finally, replace t by $\frac{1}{\rho}$ and introduce the following deterministic quantities:

$$\Omega_K^2 = \frac{\gamma}{\rho^2}\left(\frac{\gamma\tilde{\gamma}}{\rho^2 - \gamma\tilde{\gamma}} + 1\right),$$
$$\nu_K = \frac{2\rho^3}{K\left(\rho^2 - \gamma\tilde{\gamma}\right)^3}\left[\mathrm{Tr}\mathbf{D}^3\mathbf{T}^3 - \frac{\gamma^3}{\rho^3}\mathrm{Tr}\widetilde{\mathbf{D}}^3\widetilde{\mathbf{T}}^3\right].$$

As usual, the notation $\alpha_K = \mathcal{O}(\beta_K)$ means that $\alpha_K(\beta_K)^{-1}$ is uniformly bounded as $K \to \infty$. Then, the first three asymptotic moments are given by the following theorem:

Theorem 5.1. *Assuming that the matrices $\mathbf{D}$ and $\widetilde{\mathbf{D}}$ satisfy the conditions stated in* **A1** *and* **A2**, *then the following convergences hold true:*

1. *First asymptotic moment :*

$$\frac{\delta_K}{\rho} = \mathcal{O}(1) \quad \textit{and} \quad \mathbb{E}\left(\frac{\beta_K}{p_0}\right) - \frac{\delta_K}{\rho} \xrightarrow[K\to\infty]{} 0,$$

2. *Second asymptotic moment :*

$$\Omega_K = \mathcal{O}(1) \quad \textit{and} \quad K\mathbb{E}\left(\frac{\beta_K}{p_0} - \mathbb{E}\left(\frac{\beta_K}{p_0}\right)\right)^2 - \Omega_K^2 \xrightarrow[K\to\infty]{} 0,$$

3. *Third asymptotic moment:*

$$\nu_K = \mathcal{O}(1) \quad \textit{and} \quad K^2\mathbb{E}\left(\frac{\beta_K}{p_0} - \mathbb{E}\left(\frac{\beta_K}{p_0}\right)\right)^3 - \nu_K \xrightarrow[K\to\infty]{} 0.$$

The two first items of the theorem are proved in the previous chapter and can be also found in [63] (beware that the notations used in this article are the same as those in [53] and slightly differ from those used in [63]). Proof of the third item of the theorem is postponed to the appendix.

One can note that the third asymptotic moment is of order $\mathcal{O}(K^{-2})$. This is in accordance with the asymptotic normality of the SINR, where the third moment of $\sqrt{K}(\beta_K - \mathbb{E}(\beta_K))$ will vanish, as this quantity becomes closer to a Gaussian random variable. However, its value remains significant for small dimension systems.

5.4 Simulation results

In our simulations, we consider a MIMO system in the uplink direction. The base station is equipped with N receiving antennas and detects the symbols transmitted by a particular user in the presence of K interfering users. We assume that the correlation matrix $\mathbf{\Psi}$ is given by $\mathbf{\Psi}(i,j) = \sqrt{\frac{K}{N}} a^{|i-j|}$ with $0 \leq a < 1$. Recall that $\widetilde{\mathbf{P}}$ is the matrix of the interfering users' powers. We set $\widetilde{\mathbf{P}}$ (up to a permutation of its diagonal elements) to:

$$\widetilde{\mathbf{P}} = \begin{cases} \text{diag}([4P \;\; 5P]) & \text{if} \;\; K = 2 \\ \text{diag}([P \;\; P \;\; 2P \;\; 4P]) & \text{if} \;\; K = 4 \end{cases},$$

where P is the power of the user of interest. For $K = 2^p$ with $3 \leq p \leq 5$, we assume that the powers of the interfering sources are arranged into five classes as in Table 5.4. We investigate

Table 5.1: Power classes and relative frequencies

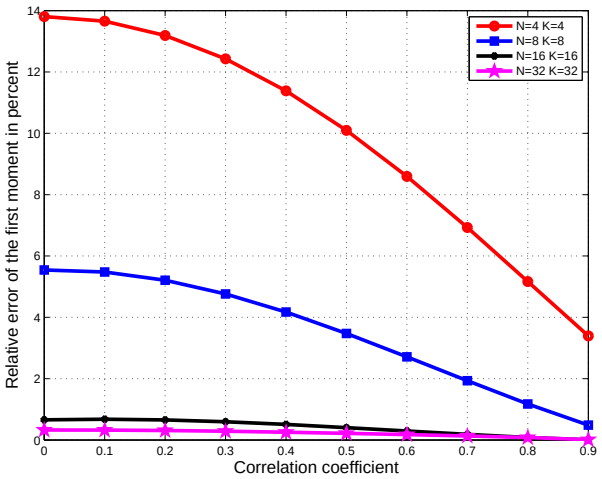
Class	1	2	3	4	5
Power	P	$2P$	$4P$	$8P$	$16P$
Relative frequency	1/8	1/4	1/4	1/8	1/4

the impact of the correlation coefficient a on the accuracy of the asymptotic moments when the input SNR is set to 15dB for $N = K$ (Fig. 5.1) and $N = 2K$ (Fig. 5.2). In these figures, the relative error on the estimated first three moments $\frac{|\mu_\infty - \mu|}{\mu}$ (μ_∞ and μ denote respectively the asymptotic and empirical moment) is depicted with respect to the correlation coefficient a. These simulations show that when the number of antennas is small, the asymptotic approximation of the second and third moments degrades for large correlation coefficients (a close to 1). Despite these discrepancies for a close to 1, simulations show that the BER and the outage probability are well approximated even for small system dimensions. Indeed, Figure 5.3 shows the evolution of the empirical BER and the theoretical BER predicted by (2.5) versus the input SNR for different values of a, K and N. In Figure 5.4, the saddle point approximate of the outage probability given by (5.2) is compared with the empirical one. In both Figures 5.3 and 5.4, 2000 channel realizations have been considered, and in Fig. 5.4, the input SNR has been set to 15 dB. These figures show that even for small system dimensions, the BER is well approximated for a wide range of SNR values. For high SNR values, the proposed approximation tends to underestimate the bit error rate. A possible reason might be that the first three moments are not sufficient to estimate accurately the bit error rate (BER). To get a more accurate bit error rate value, one should go beyond these moments and take into account the values of higher order moments.

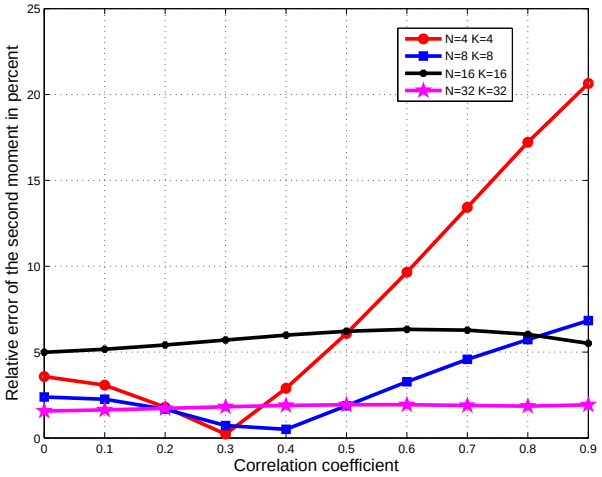
The outage probability is also well approximated except for small values of the SNR threshold that are likely to be in the tail of the asymptotic distribution.

5.5 Conclusion

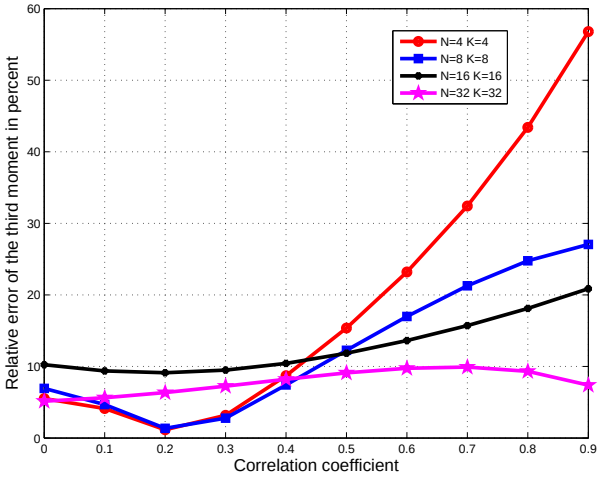
This chapter addresses the problem of BER and outage probability approximations for MIMO systems based on LMMSE receiver. Based on an asymptotic analysis, the BER and the outage probability can be shown to converge to deterministic values. However, for usual cases, these



(a) First moment of the SINR

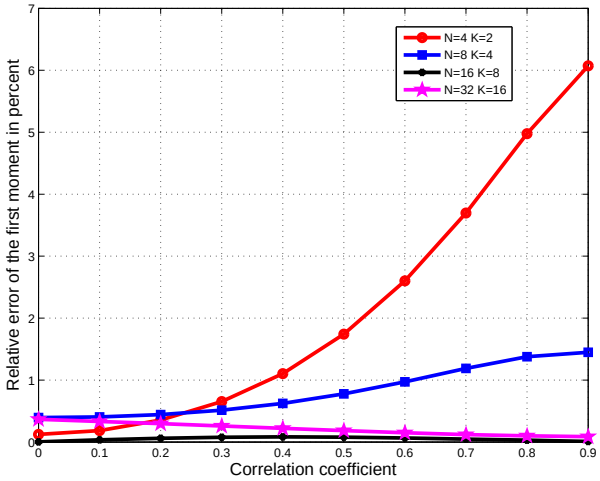


(b) Second moment of the SINR

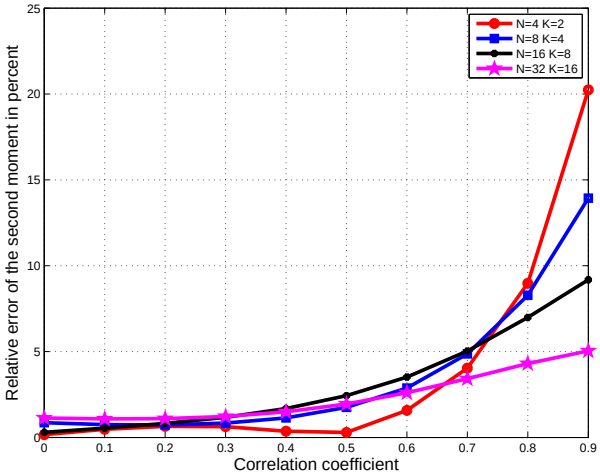


(c) Third moment of the SINR

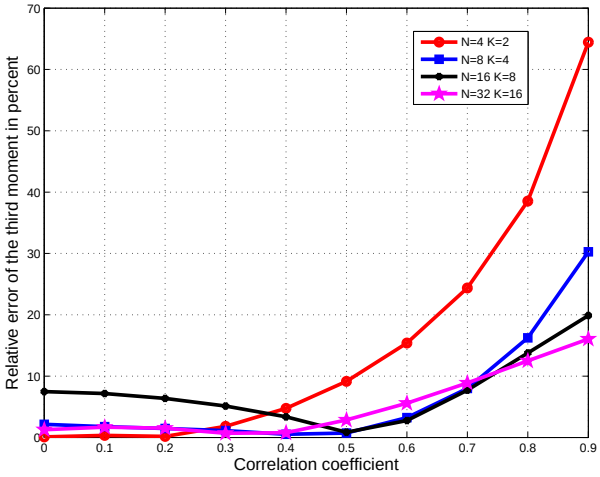
Figure 5.1: Absolute value of the relative error when $N = K$



(a) First moment of the SINR



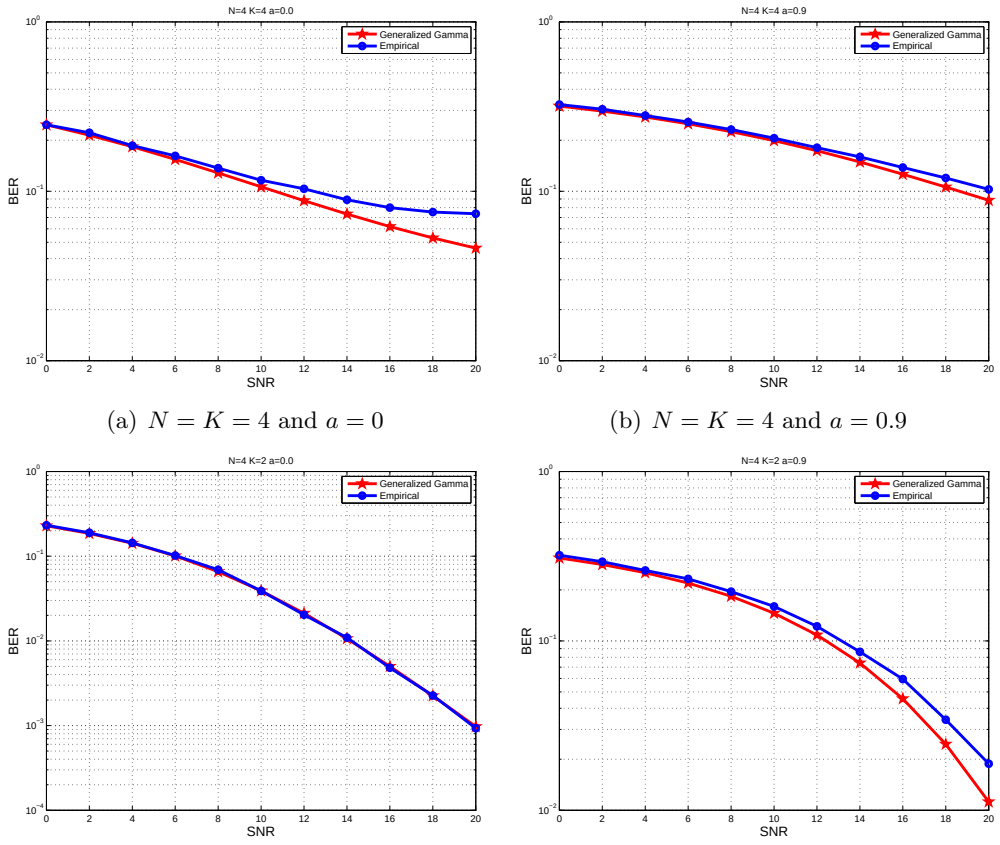
(b) Second moment of the SINR



(c) Third moment of the SINR

Figure 5.2: Absolute value of the relative error when $N = 2K$

values are not accurate, and cannot be used to get reliable approximations for these performance metrics. The reason behind these discrepancies is that the Gaussian distribution allows negative values and have zero third central moment, whereas the SINR is always positive and so is its third moment. To get around this problem, we employ the Generalized Gamma distribution instead the Gaussian one, and tune its three parameters in such away to equalize its three central moments, with the asymptotic ones. Our main contribution was to compute the asymptotic third moment, since the two first ones have been provided in the previous chapter. Simulations show that our method is efficient in approximating the BER and outage probability for a wide range of SNR values.



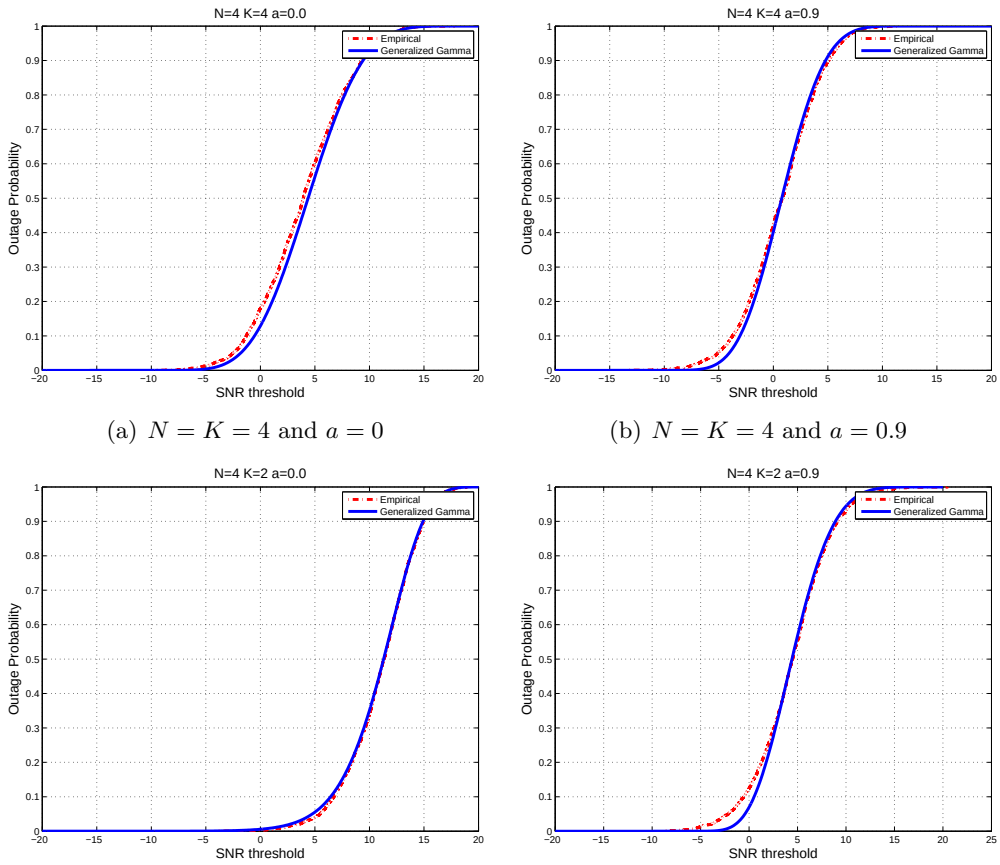
(a) $N = K = 4$ and $a = 0$

(b) $N = K = 4$ and $a = 0.9$

(c) $N = 2K = 4$ and $a = 0$

(d) $N = 2K = 4$ and $a = 0.9$

Figure 5.3: BER vs input SNR



(a) $N = K = 4$ and $a = 0$

(b) $N = K = 4$ and $a = 0.9$

(c) $N = 2K = 4$ and $a = 0$

(d) $N = 2K = 4$ and $a = 0.9$

Figure 5.4: Outage Probability vs SNR threshold

Appendices

Appendix C

Proof of results in chapter 5

In the sequel, we shall heavily rely on the results and techniques developed in [53]. In the sequel, $\mathbf{D}$ and $\widetilde{\mathbf{D}}$ are respectively $N \times N$ and $K \times K$ diagonal matrices which satisfy **A1** and **A2**, $\mathbf{Z}$ is a $N \times K$ matrix whose entries are i.i.d. standard complex Gaussian, $\mathbf{X}$ is a $N \times K$ matrix defined by:

$$\mathbf{X} = \mathbf{D}^{\frac{1}{2}}\mathbf{Z}\widetilde{\mathbf{D}}^{\frac{1}{2}} \ .$$

We shall often write $\mathbf{X} = [\mathbf{x}_1, \cdots, \mathbf{x}_K]$ where the $\mathbf{x}_j$'s are $\mathbf{X}$'s columns. We recall hereafter the mathematical tools that will be of constant use in the sequel.

C.1 Notations

Define the resolvent matrix $\mathbf{H}$ by:

$$\mathbf{H} = \left(\frac{t}{K}\mathbf{D}^{\frac{1}{2}}\mathbf{Z}\widetilde{\mathbf{D}}\mathbf{Z}^{\mathrm{H}}\mathbf{D}^{\frac{1}{2}} + \mathbf{I}_N\right)^{-1} = \left(\frac{t}{K}\mathbf{X}\mathbf{X}^{\mathrm{H}} + \mathbf{I}_N\right)^{-1} \ .$$

We introduce the following intermediate quantities:

$$\beta(t) = \frac{1}{K}\mathrm{Tr}(\mathbf{D}\mathbf{H}), \quad \alpha(t) = \frac{1}{K}\mathrm{Tr}(\mathbf{D}\mathbb{E}\mathbf{H}) \quad \text{and} \quad \overset{o}{\beta} = \beta - \alpha \ .$$

Matrix $\widetilde{\mathbf{R}}(t) = \mathrm{diag}\,(\tilde{r}_1, \cdots, \tilde{r}_K)$ is a $K \times K$ diagonal matrix defined by:

$$\widetilde{\mathbf{R}}(t) = \left(\mathbf{I} + t\alpha(t)\widetilde{\mathbf{D}}_K\right)^{-1} \ .$$

Let $\tilde{\alpha} = \frac{1}{K}\mathrm{Tr}(\widetilde{\mathbf{D}}\widetilde{\mathbf{R}})$. Then, matrix $\mathbf{R}(t) = \mathrm{diag}\,(r_1, \cdots, r_N)$ is a $N \times N$ matrix defined by:

$$\mathbf{R}(t) = (\mathbf{I} + t\tilde{\alpha}(t)\mathbf{D})^{-1} \, .$$

C.2 Mathematical Tools

The results below, of constant use in the proof of Theorem 5.1, can be found in [53].

C.2.1 Differentiation formulas

$$\frac{\partial H_{pq}}{\partial X_{ij}} = -\frac{t}{K}\left[\mathbf{X}^{\mathrm{H}}\mathbf{H}\right]_{jq} H_{pi} = -\frac{t}{K}\left[\mathbf{x}_j^{\mathrm{H}}\mathbf{H}\right]_{q} H_{pi}. \tag{C.1}$$

$$\frac{\partial H_{pq}}{\partial \overline{X_{ij}}} = -\frac{t}{K}\left[\mathbf{H}\mathbf{X}\right]_{pj} H_{iq} = -\frac{t}{K}\left[\mathbf{H}\mathbf{x}_j\right]_{p} H_{iq} \tag{C.2}$$

C.2.2 Integration by parts formula for Gaussian functionals

Let Φ be a $\mathcal{C}^1$ complex function polynomially bounded together with its derivatives, then:

$$\mathbb{E}\left[X_{ij}\Phi(\mathbf{X})\right] = d_i\tilde{d}_j\mathbb{E}\left[\frac{\partial\ \Phi(\mathbf{X})}{\partial\overline{X_{ij}}}\right]. \tag{C.3}$$

C.2.3 Poincaré-Nash inequality

Let $\mathbf{X}$ and Φ be as above, then:

$$\mathrm{Var}(\Phi(\mathbf{X})) \leq \sum_{i=1}^{N}\sum_{j=1}^{K} d_i\tilde{d}_j\mathbb{E}\left[\left|\frac{\partial\Phi(\mathbf{X})}{\partial X_{ij}}\right|^2 + \left|\frac{\partial\Phi(\mathbf{X})}{\partial\overline{X_{ij}}}\right|^2\right]. \tag{C.4}$$

C.2.4 Deterministic approximations and various estimations

Proposition C.1. *Let $(\mathbf{A}_K)$ and $(\mathbf{B}_K)$ be two sequences of respectively $N\times N$ and $K\times K$ diagonal deterministic matrices whose spectral norm are uniformly bounded in K, then the following hold true:*

$$\frac{1}{K}\mathrm{Tr}(\mathbf{AR}) = \frac{1}{K}\mathrm{Tr}(\mathbf{AT}) + \mathcal{O}(K^{-2}), \qquad \frac{1}{K}\mathrm{Tr}(\mathbf{B}\widetilde{\mathbf{R}}) = \frac{1}{K}\mathrm{Tr}(\mathbf{B}\widetilde{\mathbf{T}}) + \mathcal{O}(K^{-2}).$$

Proposition C.2. *Let $(\mathbf{A}_K)$, $(\mathbf{B}_K)$ and $(\mathbf{C}_K)$ be three sequences of $N\times N$, $K\times K$ and $N\times N$ diagonal deterministic matrices whose spectral norm are uniformly bounded in K. Consider the following functions:*

$$\Phi(\mathbf{X}) = \frac{1}{K}\mathrm{Tr}\left(\mathbf{AH}\frac{\mathbf{XBX}^{\mathrm{H}}}{K}\right), \quad \Psi(\mathbf{X}) = \frac{1}{K}\mathrm{Tr}\left(\mathbf{AHDH}\frac{\mathbf{XBX}^{\mathrm{H}}}{K}\right).$$

Then,

1. *the following estimations hold true:*

$$\mathrm{var}\,\Phi(\mathbf{X}),\ \mathrm{var}\,\Psi(\mathbf{X}),\ \mathrm{var}(\beta) \quad \textit{and} \quad \mathrm{var}\left(\frac{1}{K}\mathrm{Tr}\mathbf{AHCH}\right) \quad \textit{are} \quad \mathcal{O}(K^{-2})\ .$$

2. *the following approximations hold true:*

$$\mathbb{E}\left[\Phi(\mathbf{X})\right] = \frac{1}{K}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}\mathbf{B}\right)\frac{1}{K}\mathrm{Tr}\left(\mathbf{ADT}\right) + \mathcal{O}(K^{-2}), \tag{C.5}$$

$$\mathbb{E}\left[\Psi(\mathbf{X})\right] = \frac{1}{1-t^2\gamma\tilde{\gamma}}\left(\frac{1}{K^2}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}\mathbf{B}\right)\mathrm{Tr}(\mathbf{AD}^2\mathbf{T}^2) - \frac{t\gamma}{K^2}\mathrm{Tr}\left(\widetilde{\mathbf{D}}^2\widetilde{\mathbf{T}}^2\mathbf{B}\right)\mathrm{Tr}(\mathbf{ADT})\right) + \mathcal{O}(K^{-2}), \tag{C.6}$$

$$\mathbb{E}\frac{1}{K}\mathrm{Tr}\left[\mathbf{AHDH}\right] = \frac{1}{1-t^2\gamma\tilde{\gamma}}\frac{1}{K}\mathrm{Tr}(\mathbf{ADT}^2) + \mathcal{O}(K^{-2}). \tag{C.7}$$

Proofs of Propositions C.1 and C.2 are essentially provided in [53]. In the same vein, the following proposition will be needed.

Proposition C.3. *Let* $(\mathbf{A}_K)$, $(\mathbf{B}_K)$ *and* $(\mathbf{C}_K)$ *be three sequences of* $N \times N$, $K \times K$ *and* $N \times N$ *diagonal deterministic matrices whose spectral norm are uniformly bounded in* K. *Consider the following function:*

$$\varphi(\mathbf{X}) = \frac{1}{K}\mathrm{Tr}\left[\mathbf{CHAHAH}\frac{\mathbf{XBX}^{\mathrm{H}}}{K}\right] .$$

Then $\mathrm{var}\,\varphi(\mathbf{X}) = \mathcal{O}(K^{-2})$ *and* $\mathrm{var}\left(\frac{1}{K}\mathrm{Tr}\mathbf{AHAHAH}\right) = \mathcal{O}(K^{-2})$.

Proof of Proposition C.3 is essentially the same as the proof of Proposition C.2-1). It is provided for completeness and postponed to appendix C.4.

C.3 End of proof of Theorem 5.1

We are now in position to complete the proof of Theorem 5.1. Using the notations of [53], the SINR writes:

$$\beta_K = \frac{tp_0}{K}\mathbf{z}^{\mathrm{H}}\mathbf{D}^{\frac{1}{2}}\mathbf{H}(\mathbf{t})\mathbf{D}^{\frac{1}{2}}\mathbf{z},$$

where $t = \frac{1}{\rho}$. Hence, the third moment is given by:

$$\begin{aligned}
\mathbb{E}\left(\beta_K - \mathbb{E}\beta_K\right)^3 &= \frac{(tp_0)^3}{K^3}\mathbb{E}\left(\mathbf{z}^{\mathrm{H}}\mathbf{D}^{\frac{1}{2}}\mathbf{H}\mathbf{D}^{\frac{1}{2}}\mathbf{z} - \mathbb{E}\mathrm{Tr}\mathbf{DH}\right)^3 , \\
&= \frac{(tp_0)^3}{K^3}\mathbb{E}\left(\mathbf{z}^{\mathrm{H}}\mathbf{D}^{\frac{1}{2}}\mathbf{H}\mathbf{D}^{\frac{1}{2}}\mathbf{z} - \mathrm{Tr}\mathbf{DH} + \mathrm{Tr}\mathbf{DH} - \mathbb{E}\mathrm{Tr}\mathbf{DH}\right)^3 , \\
&= \frac{(tp_0)^3}{K^3}\left[\mathbb{E}\left(\mathbf{z}^{\mathrm{H}}\mathbf{D}^{\frac{1}{2}}\mathbf{H}\mathbf{D}^{\frac{1}{2}}\mathbf{z} - \mathrm{Tr}\mathbf{DH}\right)^3 + 3\mathbb{E}\left(\mathbf{z}^{\mathrm{H}}\mathbf{D}^{\frac{1}{2}}\mathbf{H}\mathbf{D}^{\frac{1}{2}}\mathbf{z} - \mathrm{Tr}\mathbf{DH}\right)^2\left(\mathrm{Tr}\mathbf{DH} - \mathbb{E}\mathrm{Tr}\mathbf{DH}\right)\right. \\
&\quad \left. + 3\mathbb{E}\left(\mathbf{z}^{\mathrm{H}}\mathbf{D}^{\frac{1}{2}}\mathbf{H}\mathbf{D}^{\frac{1}{2}}\mathbf{z} - \mathrm{Tr}\mathbf{DH}\right)\left(\mathrm{Tr}\mathbf{DH} - \mathbb{E}\mathrm{Tr}\mathbf{DH}\right)^2 + \mathbb{E}\left(\mathrm{Tr}\mathbf{DH} - \mathbb{E}\mathrm{Tr}\mathbf{DH}\right)^3\right] , \\
&= \frac{(tp_0)^3}{K^3}\left[\mathbb{E}\left(\mathbf{z}^{\mathrm{H}}\mathbf{D}^{\frac{1}{2}}\mathbf{H}\mathbf{D}^{\frac{1}{2}}\mathbf{z} - \mathrm{Tr}\mathbf{DH}\right)^3 + 3\mathbb{E}\left(\mathbf{z}^{\mathrm{H}}\mathbf{D}^{\frac{1}{2}}\mathbf{H}\mathbf{D}^{\frac{1}{2}}\mathbf{z} - \mathrm{Tr}\mathbf{DH}\right)^2\left(\mathrm{Tr}\mathbf{DH} - \mathbb{E}\mathrm{Tr}\mathbf{DH}\right)\right. \\
&\quad \left. + \mathbb{E}\left(\mathrm{Tr}\mathbf{DH} - \mathbb{E}\mathrm{Tr}\mathbf{DH}\right)^3\right] \qquad \text{(C.8)}
\end{aligned}$$

In order to deal with the first term of the right-hand side of (C.8), notice that if $\mathbf{M}$ is a deterministic matrix and $\mathbf{x}$ is a standard Gaussian vector, then:

$$\mathbb{E}\left(\mathbf{x}^{\mathrm{H}}\mathbf{M}\mathbf{x} - \mathrm{Tr}\mathbf{M}\right)^3 = \mathrm{Tr}(\mathbf{M}^3)\mathbb{E}\left(|x_1|^2 - 1\right)^3$$

(such an identity can be easily proved by considering the spectral decomposition of $\mathbf{M}$). Hence,

$$\begin{aligned}
\mathbb{E}\left(\mathbf{z}^{\mathrm{H}}\mathbf{D}^{\frac{1}{2}}\mathbf{H}\mathbf{D}^{\frac{1}{2}}\mathbf{z} - \mathrm{Tr}\mathbf{DH}\right)^3 = & \qquad \mathbb{E}\mathrm{Tr}\left(\mathbf{DH}\right)^3\mathbb{E}\left(|Z_{11}|^2 - 1\right)^3 , \\
&= 2\mathbb{E}\mathrm{Tr}\left(\mathbf{DHDHDH}\right) .
\end{aligned}$$

The second term of the right-hand side of (C.8) is uniformly bounded in K. Indeed:

$$\begin{aligned}
3\mathbb{E}\left(\mathbf{z}^{\mathrm{H}}\mathbf{D}^{\frac{1}{2}}\mathbf{H}\mathbf{D}^{\frac{1}{2}}\mathbf{z} - \mathrm{Tr}(\mathbf{DH})\right)^2 &= 3\mathbb{E}\left(|Z_{11}|^2 - 1\right)^2\mathrm{Tr}\mathbf{DHDH}\left(\mathrm{Tr}\mathbf{DH} - \mathbb{E}\mathrm{Tr}\mathbf{DH}\right) , \\
&\leq 3\sqrt{\mathrm{var}\left(\mathrm{Tr}\mathbf{DHDH}\right)}\sqrt{\mathrm{var}\left(\mathrm{Tr}\mathbf{DH}\right)}
\end{aligned}$$

which is $\mathcal{O}(1)$ according to Proposition C.2. It remains to deal with $\mathbb{E}\left(\mathrm{Tr}\mathbf{DH} - \mathbb{E}\mathrm{Tr}\mathbf{DH}\right)^3$, which can be proved to be uniformly bounded in K using concentration results for the spectral measure of random matrices [52] (see also [82, eq.(86)-(87)], where details are provided). Consequently, we end up with the following approximation:

$$K^2\mathbb{E}\left(\beta_K - \mathbb{E}\beta_K\right)^3 = \frac{(tp_0)^3}{K}\mathbb{E}\left(|Z_{11}|^2 - 1\right)^3 \mathbb{E}\mathrm{Tr}\mathbf{DHDHDH} + \mathcal{O}\left(K^{-1}\right)$$

which is deterministic but still depends on the distribution of the entries via the expectation operator $\mathbb{E}$. The rest of the proof is devoted to provide a deterministic approximation of $\mathbb{E}\mathrm{Tr}\left(\mathbf{DHDHDH}\right)$ depending on γ, $\tilde{\gamma}$, $\mathbf{T}$ and $\widetilde{\mathbf{T}}$.

Note that $\mathbf{H} = \mathbf{I} - \frac{t}{K}\mathbf{H}\mathbf{X}\mathbf{X}^{\mathrm{H}}$, thus:

$$\begin{aligned}\left[\mathbf{HDHDH}\right]_{pp} &= \left[\mathbf{HDHD}\right]_{pp} - t\left[\mathbf{HDHDH}\frac{\mathbf{X}\mathbf{X}^{\mathrm{H}}}{K}\right]_{pp}, \\ &= \left[\mathbf{HDHD}\right]_{pp} - \frac{t}{K}\sum_{j=1}^{K}\left[\mathbf{HDHDH}\mathbf{x}_j\right]_p\overline{X_{pj}}. \end{aligned} \tag{C.9}$$

Let us deal with the second term of (C.9). We have:

$$\mathbb{E}\frac{1}{K}\left[\mathbf{HDHDH}\mathbf{x}_j\right]_p\overline{X_{pj}} = \frac{1}{K}\sum_{k=1}^{N}\mathbb{E}\left(\left[\mathbf{HDHDH}\right]_{pk}X_{kj}\overline{X_{pj}}\right).$$

Using the integration by part formula (C.3), we get:

$$\begin{aligned}\mathbb{E}\left[\mathbf{HDHDH}\mathbf{x}_j\right]_p\overline{X_{pj}} &= \sum_{k=1}^{N} d_k\tilde{d}_j\delta(p-k)\mathbb{E}\left[\mathbf{HDHDH}\right]_{pk} + \sum_{k=1}^{N} d_k\tilde{d}_j\mathbb{E}\left[\overline{X_{pj}}\sum_{\ell,m=1}^{N}\frac{\partial\left[H_{p\ell}d_\ell d_m H_{\ell m}H_{mk}\right]}{\partial\overline{X_{kj}}}\right], \\ &= d_p\tilde{d}_j\mathbb{E}\left[\mathbf{HDHDH}\right]_{pp} - \frac{t}{K}\sum_{k,\ell,m=1}^{N} d_k\tilde{d}_j d_m d_\ell\mathbb{E}\left[\overline{X_{pj}}\left[\mathbf{H}\mathbf{x}_j\right]_p H_{k\ell}H_{\ell m}H_{mk}\right] \\ &\quad - \frac{t}{K}\sum_{k,\ell,m=1}^{N} d_k\tilde{d}_j d_m d_\ell\mathbb{E}\left[\overline{X_{pj}}H_{p\ell}\left[\mathbf{H}\mathbf{x}_j\right]_\ell H_{km}H_{mk}\right] \\ &\quad - \frac{t}{K}\sum_{k,\ell,m=1}^{N} d_k\tilde{d}_j d_m d_\ell\mathbb{E}\left[H_{p\ell}H_{\ell m}\left[\mathbf{H}\mathbf{x}_j\right]_m H_{kk}\right]. \\ &= d_p\tilde{d}_j\mathbb{E}\left[\mathbf{HDHDH}\right]_{pp} - \frac{t}{K}\tilde{d}_j\mathbb{E}\left[\left[\mathbf{H}\mathbf{x}_j\right]_p\overline{X_{pj}}\mathrm{Tr}\left(\mathbf{DHDHDH}\right)\right] \\ &\quad - \frac{t}{K}\tilde{d}_j\mathbb{E}\left[\left[\mathbf{HDH}\mathbf{x}_j\right]_p\overline{X_{pj}}\mathrm{Tr}\left(\mathbf{DHDH}\right)\right] - \frac{t}{K}\tilde{d}_j\mathbb{E}\left[\left[\mathbf{HDHDH}\mathbf{x}_j\right]_p\overline{X_{pj}}\mathrm{Tr}\left(\mathbf{DH}\right)\right]. \end{aligned}$$

Substituting in the last term $\frac{1}{K}\mathrm{Tr}\mathbf{DH} = \overset{o}{\beta} + \alpha$ where $\overset{o}{\beta} = \beta - \alpha$, we get:

$$\begin{aligned}\mathbb{E}\left[\mathbf{HDHDH}\mathbf{x}_j\right]_p\overline{X_{pj}} &= d_p\tilde{d}_j\mathbb{E}\left[\mathbf{HDHDH}\right]_{pp} - \frac{t}{K}\tilde{d}_j\mathbb{E}\left[\left[\mathbf{H}\mathbf{x}_j\right]_p\overline{X_{pj}}\mathrm{Tr}\left(\mathbf{DHDHDH}\right)\right] \\ &\quad - \frac{t}{K}\tilde{d}_j\mathbb{E}\left[\left[\mathbf{HDH}\mathbf{x}_j\right]_p\overline{X_{pj}}\mathrm{Tr}\left(\mathbf{DHDH}\right)\right] - t\tilde{d}_j\mathbb{E}\left[\left[\mathbf{HDHDH}\mathbf{x}_j\right]_p\overline{X_{pj}}\,\overset{o}{\beta}\right] \\ &\quad - t\tilde{d}_j\mathbb{E}\left[\left[\mathbf{HDHDH}\mathbf{x}_j\right]_p\overline{X_{pj}}\right]\alpha. \end{aligned}$$

Therefore, we have:

$$\left(1+t\alpha\tilde{d}_j\right)\mathbb{E}\left[\left[\mathbf{HDHDHx}_j\right]_p\overline{X_{pj}}\right]=d_p\tilde{d}_j\mathbb{E}\left[\mathbf{HDHDH}\right]_{pp}-\frac{t}{K}\mathbb{E}\left[\left[\mathbf{Hx}_j\right]_p\overline{X_{pj}}\tilde{d}_j\mathrm{Tr}\left[\mathbf{DHDHDH}\right]\right]$$
$$-\frac{t}{K}\tilde{d}_j\mathbb{E}\left[\left[\mathbf{HDHx}_j\right]_p\overline{X_{pj}}\mathrm{Tr}\left[\mathbf{DHDH}\right]\right]-t\tilde{d}_j\mathbb{E}\left[\left[\mathbf{HDHDHx}_j\right]_p\overline{X_{pj}}\overset{o}{\beta}\right].$$

Multiplying the right hand and the left hand sides by $\tilde{r}_j=\frac{1}{1+t\alpha\tilde{d}_j}$, we get:

$$\mathbb{E}\left[\mathbf{HDHDHx}_j\right]_p\overline{X_{pj}} \quad = \quad \tilde{r}_jd_p\tilde{d}_j\mathbb{E}\left[\mathbf{HDHDH}\right]_{pp}-\frac{t}{K}\tilde{r}_j\mathbb{E}\left[\left[\mathbf{Hx}_j\right]_p\overline{X_{pj}}\tilde{d}_j\mathrm{Tr}\left[\mathbf{DHDHDH}\right]\right]$$
$$-\frac{t}{K}\tilde{d}_j\tilde{r}_j\mathbb{E}\left[\left[\mathbf{HDHx}_j\right]_p\overline{X_{pj}}\mathrm{Tr}\left[\mathbf{DHDH}\right]\right]-t\tilde{d}_j\tilde{r}_j\mathbb{E}\left[\left[\mathbf{HDHDHx}_j\right]_p\overline{X_{pj}}\overset{o}{\beta}\right]. \quad \text{(C.10)}$$

Plugging (C.10) into (C.9), we obtain:

$$\mathbb{E}\left[\mathbf{HDHDH}\right]_{pp}=\mathbb{E}\left[\mathbf{HDHD}\right]_{pp}-\sum_{j=1}^{K}\frac{t}{K}\tilde{r}_jd_p\tilde{d}_j\mathbb{E}\left[\mathbf{HDHDH}\right]_{pp}+\frac{t^2}{K^2}\sum_{j=1}^{K}\tilde{r}_j\mathbb{E}\left[\mathbf{Hx}_j\right]_p\overline{X_{pj}}\tilde{d}_j\mathrm{Tr}\left[\mathbf{DHDHDH}\right]$$
$$+\frac{t^2}{K^2}\sum_{j=1}^{K}\tilde{d}_j\tilde{r}_j\mathbb{E}\left[\mathbf{HDHx}_j\right]_p\overline{X_{p,j}}\mathrm{Tr}\left[\mathbf{DHDH}\right]+\frac{t}{K}\sum_{j=1}^{K}\tilde{d}_j\tilde{r}_j\mathbb{E}\left[\mathbf{HDHDHx}_j\right]_p\overline{X_{p,j}}\overset{o}{\beta},$$
$$=\mathbb{E}\left[\mathbf{HDHD}\right]_{pp}-t\tilde{\alpha}d_p\mathbb{E}\left[\mathbf{HDHDH}\right]_{pp}+\frac{t^2}{K^2}\mathbb{E}\mathrm{Tr}(\mathbf{DHDHDH})\left[\mathbf{HX}\widetilde{\mathbf{R}}\widetilde{\mathbf{D}}\mathbf{X}^{\mathrm{H}}\right]_{pp}$$
$$+\frac{t^2}{K^2}\mathbb{E}\mathrm{Tr}\left[\mathbf{DHDH}\right]\left[\mathbf{HDHX}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{X}^{\mathrm{H}}\right]_{pp}+\frac{t^2}{K}\mathbb{E}\overset{o}{\beta}\left[\mathbf{HDHDHX}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{X}^{\mathrm{H}}\right]_{pp}.$$

Hence,

$$(1+t\tilde{\alpha}d_p)\mathbb{E}\left[\mathbf{HDHDH}\right]_{pp}=\mathbb{E}\left[\mathbf{HDHD}\right]_{pp}+\frac{t^2}{K^2}\mathbb{E}\mathrm{Tr}\left[\mathbf{DHDHDH}\right]\left[\mathbf{HX}\widetilde{\mathbf{R}}\widetilde{\mathbf{D}}\mathbf{X}^{\mathrm{H}}\right]_{pp}$$
$$+\frac{t^2}{K^2}\mathbb{E}\mathrm{Tr}\left[\mathbf{DHDH}\right]\left[\mathbf{HDHX}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{X}^{\mathrm{H}}\right]_{pp}+\frac{t^2}{K}\mathbb{E}\overset{o}{\beta}\left[\mathbf{HDHDHX}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{X}^{\mathrm{H}}\right]_{pp}.$$

Multiplying the left and right hand sides by $r_p=\frac{1}{1+t\tilde{\alpha}d_p}$, we get:

$$\mathbb{E}\left[\mathbf{HDHDH}\right]_{pp}=r_p\mathbb{E}\left[\mathbf{HDHD}\right]_{pp}+\frac{t^2}{K^2}r_p\mathbb{E}\mathrm{Tr}\left[\mathbf{DHDHDH}\right]\left[\mathbf{HX}\widetilde{\mathbf{R}}\widetilde{\mathbf{D}}\mathbf{X}^{\mathrm{H}}\right]_{pp}$$
$$+\frac{t^2}{K^2}r_p\mathbb{E}\mathrm{Tr}\left[\mathbf{DHDH}\right]\left[\mathbf{HDHX}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{X}^{\mathrm{H}}\right]_{pp}+\frac{t^2}{K}r_p\mathbb{E}\overset{o}{\beta}\left[\mathbf{HDHDHX}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{X}^{\mathrm{H}}\right]_{pp}. \quad \text{(C.11)}$$

Multiplying by d_p, summing over p and dividing by K, we obtain:

$$\mathbb{E}\frac{1}{K}\mathrm{Tr}\left[\mathbf{DHDHDH}\right]=\mathbb{E}\frac{1}{K}\sum_{p=1}^{K}d_p\left[\mathbf{HDHDH}\right]_{pp},$$
$$=\frac{1}{K}\sum_{p=1}^{K}r_pd_p\mathbb{E}\left[\mathbf{HDHD}\right]_{pp}+\frac{t^2}{K^3}\mathbb{E}\,\mathrm{Tr}\left(\mathbf{DHDHDH}\right)\mathrm{Tr}\left(\mathbf{DRHX}\widetilde{\mathbf{R}}\widetilde{\mathbf{D}}\mathbf{X}^{\mathrm{H}}\right)$$
$$+\frac{t^2}{K^3}\mathbb{E}\mathrm{Tr}\left(\mathbf{DHDH}\right)\mathrm{Tr}\left(\mathbf{DRHDHX}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{X}^{\mathrm{H}}\right)$$
$$+\frac{t^2}{K^2}\mathbb{E}\ \overset{o}{\beta}\,\mathrm{Tr}\left(\mathbf{DRHDHDHX}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{X}^{\mathrm{H}}\right),$$
$$\overset{\triangle}{=}\chi_1+\chi_2+\chi_3+\chi_4, \quad \text{(C.12)}$$

where:

$$\chi_1 = \frac{1}{K}\mathbb{E}\mathrm{Tr}\left(\mathbf{DRHDHD}\right) ,$$
$$\chi_2 = \frac{t^2}{K}\mathbb{E}\mathrm{Tr}\left(\mathbf{DHDHDH}\right)\frac{1}{K}\mathrm{Tr}\left(\mathbf{DRH}\frac{\mathbf{X}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{X}^{\mathrm{H}}}{K}\right) ,$$
$$\chi_3 = \frac{t^2}{K}\mathbb{E}\mathrm{Tr}\left(\mathbf{DHDH}\right)\frac{1}{K}\mathrm{Tr}\left(\mathbf{DRHDH}\frac{\mathbf{X}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{X}^{\mathrm{H}}}{K}\right) ,$$
$$\chi_4 = \frac{t^2}{K}\mathbb{E}\ \overset{o}{\beta}\ \mathrm{Tr}\left(\mathbf{DRHDHDH}\frac{\mathbf{X}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{X}^{\mathrm{H}}}{K}\right) .$$

According to Proposition C.2, $\mathrm{var}\frac{1}{K}\mathrm{Tr}\left(\mathbf{DRHDHDH}\frac{\mathbf{X}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{X}^{\mathrm{H}}}{K}\right)$ is of order $\mathcal{O}(K^{-2})$. Similarly, $\mathrm{var}(\beta) = \mathcal{O}(K^{-2})$. Hence, using Cauchy-Schwartz inequality, we get the estimation $\chi_4 = \mathcal{O}(K^{-2})$. It remains to work out the expressions involved in χ_1, χ_2 and χ_3 by removing the terms with expectation and replacing them with deterministic equivalents.

Since $\mathrm{var}\frac{1}{K}\mathrm{Tr}\left(\mathbf{DRH}\frac{\mathbf{X}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{X}^{\mathrm{H}}}{K}\right) = \mathcal{O}(K^{-2})$ by Proposition C.2 and $\mathrm{var}(\frac{1}{K}\mathrm{Tr}\mathbf{DHDHDH}) = \mathcal{O}(K^{-2})$ by Proposition C.3, we have:

$$\begin{aligned}\chi_2 &= \frac{t^2}{K}\mathbb{E}\mathrm{Tr}\left(\mathbf{DHDHDH}\right)\mathbb{E}\left(\frac{1}{K}\mathrm{Tr}\left[\mathbf{DRH}\frac{\mathbf{X}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{X}^{\mathrm{H}}}{K}\right]\right) + \mathcal{O}(K^{-2}),\\ &\overset{(a)}{=} \frac{t^2}{K}\mathbb{E}\mathrm{Tr}\left(\mathbf{DHDHDH}\right)\frac{1}{K}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\right)\frac{1}{K}\mathrm{Tr}\left(\mathbf{DRDT}\right) + \mathcal{O}(K^{-2}),\\ &\overset{(b)}{=} \frac{t^2}{K}\mathbb{E}\mathrm{Tr}\left(\mathbf{DHDHDH}\right)\gamma\tilde{\gamma} + \mathcal{O}(K^{-2}) .\end{aligned} \tag{C.13}$$

where (a) follows from Proposition C.2-2) and (b), from Proposition C.1. Similar arguments yield:

$$\begin{aligned}\chi_3 &= \frac{t^2}{K}\mathbb{E}\mathrm{Tr}\left(\mathbf{DHDH}\right)\mathbb{E}\left(\frac{1}{K}\mathrm{Tr}\left[\mathbf{DRHDH}\frac{\mathbf{X}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\mathbf{X}^{\mathrm{H}}}{\mathbf{K}}\right]\right) + \mathcal{O}(K^{-2}),\\ &= \frac{t^2\gamma}{(1-t^2\gamma\tilde{\gamma})^2}\left[\frac{1}{K}\mathrm{Tr}\left(\widetilde{\mathbf{D}}\widetilde{\mathbf{T}}\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\right)\frac{1}{K}\mathrm{Tr}\left(\mathbf{DRD}^2\mathbf{T}^2\right) - \frac{t\gamma}{K}\mathrm{Tr}\left(\widetilde{\mathbf{D}}^2\widetilde{\mathbf{T}}^2\widetilde{\mathbf{D}}\widetilde{\mathbf{R}}\right)\frac{1}{K}\mathrm{Tr}(\mathbf{DRDT})\right] + \mathcal{O}(K^{-2}) ,\\ &= \frac{t^2\gamma}{(1-t^2\gamma\tilde{\gamma})^2}\left[\frac{\tilde{\gamma}}{K}\mathrm{Tr}(\mathbf{D}^3\mathbf{T}^3) - \frac{t\gamma^2}{K}\mathrm{Tr}(\widetilde{\mathbf{D}}^3\widetilde{\mathbf{T}}^3)\right] + \mathcal{O}(K^{-2})\end{aligned} \tag{C.14}$$

and

$$\begin{aligned}\chi_1 &= \frac{1}{1-t^2\gamma\tilde{\gamma}}\frac{1}{K}\mathrm{Tr}\left(\mathbf{D}^2\mathbf{RDT}^2\right) + \mathcal{O}(K^{-2})\\ &= \frac{1}{1-t^2\gamma\tilde{\gamma}}\frac{1}{K}\mathrm{Tr}(\mathbf{D}^3\mathbf{T}^3) + \mathcal{O}(K^{-2}).\end{aligned} \tag{C.15}$$

Plugging (C.14), (C.13) and (C.15) into (C.12), we obtain:

$$\frac{1}{K}\mathbb{E}\mathrm{Tr}(\mathbf{DHDHDH}) = \frac{1}{K(1-t^2\gamma\tilde{\gamma})^3}\mathrm{Tr}\mathbf{D}^3\mathbf{T}^3 - \frac{t^3\gamma^3}{K(1-t^2\gamma\tilde{\gamma})^3}\mathrm{Tr}\widetilde{\mathbf{T}}^3\widetilde{\mathbf{D}}^3 + \mathcal{O}(K^{-2}).$$

Hence,

$$K^2\mathbb{E}\left(\frac{\beta_K}{p_0}-\mathbb{E}\frac{\beta_K}{p_0}\right)^3=\frac{\rho^3}{K\left(\rho^2-\gamma\tilde{\gamma}\right)^3}\left[\mathrm{Tr}\mathbf{D}^3\mathbf{T}^3-\frac{\gamma^3}{\rho^3}\mathrm{Tr}\widetilde{\mathbf{D}}^3\widetilde{\mathbf{T}}^3\right]\mathbb{E}\left(|Z_{11}|^2-1\right)^3+\mathcal{O}\left(\frac{1}{K}\right),$$
$$=\frac{2\rho^3}{K\left(\rho^2-\gamma\tilde{\gamma}\right)^3}\left[\mathrm{Tr}\mathbf{D}^3\mathbf{T}^3-\frac{\gamma^3}{\rho^3}\mathrm{Tr}\widetilde{\mathbf{D}}^3\widetilde{\mathbf{T}}^3\right]+\mathcal{O}\left(\frac{1}{K}\right).$$

The fact that $\nu_K=\frac{2\rho^3}{K(\rho^2-\gamma\tilde{\gamma})^3}\left[\mathrm{Tr}\mathbf{D}^3\mathbf{T}^3-\frac{\gamma^3}{\rho^3}\mathrm{Tr}\widetilde{\mathbf{D}}^3\widetilde{\mathbf{T}}^3\right]$ is of order $\mathcal{O}(1)$ is straightforward and its proof is omitted. Proof of Theorem 5.1 is completed.

C.4 Proof of Proposition C.3

The proof mainly relies on Poincaré-Nash inequality. Using the Poincaré-Nash inequality, we have:

$$\mathrm{var}(\varphi(\mathbf{X}))\leq\sum_{i=1}^{N}\sum_{j=1}^{K}d_i\tilde{d}_j\mathbb{E}\left|\frac{\partial\varphi}{\partial X_{ij}}\right|^2+\sum_{i=1}^{N}\sum_{j=1}^{K}d_i\tilde{d}_j\mathbb{E}\left|\frac{\partial\varphi}{\partial\overline{X_{ij}}}\right|^2.$$

We only deal with the first term of the last inequality (the second term can be handled similarly). We have $\varphi(\mathbf{X})=\frac{1}{K^2}\sum_{p,r,s,t=1}^{N}\sum_{u=1}^{K}c_{pp}H_{pr}A_{rr}H_{rs}A_{ss}H_{st}X_{tu}B_{uu}X^*_{pu}$. After straightforward calculations using the differentiation formula (C.1), we get that:

$$\frac{\partial\varphi}{\partial X_{ij}}=\phi_{ij}^{(1)}+\phi_{ij}^{(2)}+\phi_{ij}^{(3)}+\phi_{ij}^{(4)},$$

where:

$$\phi_{ij}^{(1)}=-\frac{t}{K^3}\left[\mathbf{X}^{\mathrm{H}}\mathbf{HAHAHXBX}^{\mathrm{H}}\mathbf{CH}\right]_{ji},\qquad\phi_{ij}^{(2)}=-\frac{t}{K^3}\left[\mathbf{X}^{\mathrm{H}}\mathbf{HAHXBX}^{\mathrm{H}}\mathbf{CHAH}\right]_{ji},$$
$$\phi_{ij}^{(3)}=-\frac{t}{K^3}\left[\mathbf{X}^{\mathrm{H}}\mathbf{HXBX}^{\mathrm{H}}\mathbf{CHAHAH}\right]_{ji},\qquad\phi_{ij}^{(4)}=\frac{1}{K^2}\left[\mathbf{BX}^{\mathrm{H}}\mathbf{CHAHAH}\right]_{ji}.$$

Hence, $\left|\frac{\partial\varphi}{\partial X_{ij}}\right|^2\leq 4\left(\left|\phi_{ij}^{(1)}\right|^2+\left|\phi_{ij}^{(2)}\right|^2+\left|\phi_{ij}^{(3)}\right|^2+\left|\phi_{ij}^{(4)}\right|^2\right)$ and

$$\sum_{i=1}^{N}\sum_{j=1}^{K}d_i\tilde{d}_j\mathbb{E}\left[\left|\frac{\partial\varphi}{\partial X_{ij}}\right|^2\right]\leq\frac{4t^2}{K^6}\mathbb{E}\mathrm{Tr}\left(\mathbf{DHCXBX}^{\mathrm{H}}\mathbf{HAHAHX}\widetilde{\mathbf{D}}\mathbf{X}^{\mathrm{H}}\mathbf{HAHAHXBX}^{\mathrm{H}}\mathbf{CH}\right)$$
$$+\frac{4t^2}{K^6}\mathbb{E}\mathrm{Tr}\left(\mathbf{DHAHCXBX}^{\mathrm{H}}\mathbf{HAHX}\widetilde{\mathbf{D}}\mathbf{X}^{\mathrm{H}}\mathbf{HAHXBX}^{\mathrm{H}}\mathbf{CHAH}\right)$$
$$+\frac{4t^2}{K^6}\mathbb{E}\mathrm{Tr}\left(\mathbf{DHAHAHCXBX}^{\mathrm{H}}\mathbf{HX}\widetilde{\mathbf{D}}\mathbf{X}^{\mathrm{H}}\mathbf{HXBX}^{\mathrm{H}}\mathbf{CHAHAH}\right)$$
$$+\frac{4}{K^4}\mathbb{E}\mathrm{Tr}\left(\mathbf{DHAHAHCXB}\widetilde{\mathbf{D}}\mathbf{BX}^{\mathrm{H}}\mathbf{CHAHAH}\right).$$

We only prove that the first term of the right hand side is of order K^{-2}; the other terms being handled similarly. Using Cauchy-Schwartz inequality, we get:

$$\begin{aligned}
4\sum_{i=1}^{N}\sum_{j=1}^{K} d_i\tilde{d}_j\mathbb{E}\left|\phi_{ij}^1\right|^2 &\leq \frac{4t^2 d_{\max}\|\mathbf{H}\|^2\|\mathbf{C}\|^2}{K^6}\mathbb{E}\mathrm{Tr}\left((\mathbf{HA})^2\mathbf{H}\widetilde{\mathbf{D}}\mathbf{X}^{\mathrm{H}}\mathbf{H}(\mathbf{AH})^2(\mathbf{XBX}^{\mathrm{H}})^2\right),\\
&\leq \frac{4t^2}{K^6}d_{\max}\|\mathbf{H}\|^2\|\mathbf{C}\|^2\left(\mathbb{E}\mathrm{Tr}(\mathbf{HA})^2\mathbf{H}\widetilde{\mathbf{D}}\mathbf{X}^{\mathrm{H}}\mathbf{H}(\mathbf{AH})^2(\mathbf{HA})^2\mathbf{H}\widetilde{\mathbf{D}}\mathbf{X}^{\mathrm{H}}\mathbf{H}(\mathbf{AH})^2\right)^{\frac{1}{2}}\\
&\times\left(\mathbb{E}\mathrm{Tr}(\mathbf{XBX}^{\mathrm{H}})^4\right)^{\frac{1}{2}}\\
&\leq \frac{4t^2}{K^2}d_{\max}\|\mathbf{H}\|^8\|\mathbf{C}\|^2\|\mathbf{A}\|^4\sqrt{\mathbb{E}\frac{1}{K}\left(\frac{\mathbf{X}\widetilde{\mathbf{D}}\mathbf{X}^{\mathrm{H}}}{K}\right)^2}\sqrt{\mathbb{E}\frac{1}{K}\left(\frac{\mathbf{XBX}^{\mathrm{H}}}{K}\right)^4},
\end{aligned}$$

where the first inequality follows by using the fact that $|\mathrm{Tr}\mathbf{AB}| \leq \|\mathbf{B}\|\mathrm{Tr}(\mathbf{A})$, $\mathbf{A}$ being hermitian non-negative matrix and the second follows by applyig twice Cauchy-Schwartz inequalities: $\mathrm{Tr}(\mathbf{AB}) \leq \sqrt{\mathrm{Tr}(\mathbf{AA}^{\mathrm{H}})}\sqrt{\mathrm{Tr}(\mathbf{BB}^{\mathrm{H}})}$ and $\mathbb{E}XY \leq \sqrt{\mathbb{E}X^2}\sqrt{\mathbb{E}Y^2}$. We end up the proof of the first statement by using the fact that $\frac{1}{K}\mathbb{E}\left[\frac{1}{K}\mathrm{Tr}\left(\frac{1}{K}\mathbf{XB}_K\mathbf{X}^{\mathrm{H}}\right)^n\right]$ is uniformly bounded in K whenever $\mathbf{B}_K$ is a sequence of diagonal matrices with uniformly bounded spectral norm and n is a given integer.

The second statement follows from the resolvent identity:

$$\frac{1}{K}\mathrm{Tr}\mathbf{AHAHAH} = \frac{1}{K}\mathrm{Tr}\mathbf{AHAHA} - \frac{t}{K}\mathrm{Tr}\mathbf{AHAHAHXX}^{\mathrm{H}}.$$

According to the first part of the proposition,

$$\mathrm{var}\left(\frac{1}{K}\mathrm{Tr}\mathbf{AHAHAHXX}^{\mathrm{H}}\right) = \mathcal{O}(K^{-2})\ .$$

Now, $\mathrm{Tr}\mathbf{AHAHA} = \mathrm{Tr}\mathbf{A^2HAH}$ and $\mathrm{var}\frac{1}{K}\mathrm{Tr}\mathbf{A^2HAH} = \mathcal{O}(K^{-2})$ by Proposition C.2-1). Hence, applying inequality $\mathrm{var}(X+Y) \leq \mathrm{var}(X) + \mathrm{var}(Y) + 2\sqrt{\mathrm{var}(X)\mathrm{var}(Y)}$ yields the desired result. Proof of Proposition C.3 is completed.

Part III

Blind and semi-blind equalization techniques

Introduction

Blind equalization techniques can be classified into two categories. One is the indirect blind approach for which a priori estimation of the channel is required. The other approach namely, the direct blind method, estimates the optimum linear equalizer by using the second order statistics of the data without involving any estimation of the channel filter.

Unlike training based techniques, blind methods do not rely on any pilot symbols, thereby allowing higher spectral efficiency. However, the practical implementation of blind methods still faces several challenges that are yet to be fulfilled. As a matter of fact, most of the conventional blind methods involves high computational complexity and are too sensitive to channel order over-estimation. Finding out solutions that are robust to the estimation errors of the channel order while keeping a low-complexity is one of the most active research topics in the area of blind channel equalization. Among the most recent findings in this direction, two works have retained our attention.

- The first work proposes a direct blind method based on the truncation of the covariance matrix [60]. This method exhibits a high robustness to channel order over-estimation, and can be implemented by using low complexity adaptive algorithms. Nevertheless, being based on zero delay equalization, this method is too sensitive to the variance of the first channel coefficient.

- The second work deals with the use of the ℓ_p quasi-norms for making conventional blind methods more robust to channel order over-estimation [9]. This was only shown by using simulations, but have not been so far theoretically investigated.

In between the blind and the training based techniques, we have the class of semi-blind methods which jointly exploit the blind and the training based criteria or any side information about the system. Since more information is used, this class of method, exhibits better channel estimation quality. These methods can be optimal, in the sense that they minimize a maximum likelihood criteria, or also suboptimal and thus involves lower computational complexity. In this part, we consider suboptimal semi-blind regularized techniques, in which the blind criterion is linearly combined with the training based one by using a regularizing coefficient [15]. The setting of this coefficient is highly determinant on the obtained channel estimation quality. In light of this consideration, the work in [15] has proposed to set the regularizing coefficient so as to minimize the asymptotic mean square error. It was conjectured that no local minima exists, and thus the optimal regularizing coefficient is unique. Moreover, it is estimated by using iterative optimization algorithms since it does not have a closed-form expression, in general. Our contributions in this part concern blind and semi-blind techniques. For blind methods, our work has been mainly to:

- extend the truncation covariance based blind method to nonzero delay equalization

- and to provide well founded arguments that justify the relevance of the use of ℓ_p quasi-norms in conventional blind methods.

These contributions are covered by chapter 7 and 8. As for semi-blind techniques, we prove in chapter 9 the quasi-convexity of the asymptotic MSE, thereby establishing the existence of a unique minimum. In chap 10, we propose to use a regularizing matrix instead of a regularizing constant. Interestingly, in this case, we show that a closed-form expression of the optimal regularizing matrix exists, thereby avoiding the need of iterative optimization algorithms.

Notations: Along this part, except when mentioning, N, L and M denote respectively the number of receiving antenna, the channel order and the window size.

Chapter 6

Overview of Blind and semi-blind channel estimation techniques

This chapter presents an overview on both indirect blind and semi-blind channel estimation techniques. It is organized as follows: after presenting the system model, we provide an overview of linear prediction and subspace based methods. Then, we make a brief literature review on optimal and suboptimal semi-blind techniques. Finally, we end up our chapter by providing some numerical results about subspace and linear prediction techniques.

6.1 Blind equalization techniques

In this chapter, we only review indirect blind techniques. We leave the details about direct blind methods to the next chapter, where we enhance the performance of a recently proposed method by extending it to the case of nonzero delay equalization. Before presenting the indirect blind methods, we shall introduce the system model.

We consider a $N \times K$ MIMO system, for which the received signal vector is given by:

$$\mathbf{y}(k) = [\mathbf{H}(z)]\,\mathbf{s}(k) + \mathbf{b}(k), \quad k \in \mathbb{Z} \tag{6.1}$$

$$= \sum_{l=0}^{L} \mathbf{H}(l)s(k-l) + \mathbf{b}(k) \tag{6.2}$$

where $\mathbf{H}(k)$ are the $N \times K$ unknown system parameters, while $\mathbf{s}(k)$ and $\mathbf{b}(k)$ represent respectively the unknown K input process and the additive Gaussian noise.

Let $\mathbf{y}_M(k) = [\mathbf{y}^{\mathrm{T}}(k), \mathbf{y}^{\mathrm{T}}(k-1), \cdots, \mathbf{y}^{\mathrm{T}}(k-M+1)]^{\mathrm{T}}$, where M is the size of the sliding window.

Reformulating (6.2) in a matrix form, we will get:

$$\mathbf{y}_M(k) = \boldsymbol{\mathcal{I}}_M(\mathbf{h})\mathbf{s}_M(k) + \mathbf{b}_M(k)$$

where $\boldsymbol{\mathcal{I}}_M(\mathbf{h})$ is a $NM \times K(M+L)$ block-Toeplitz matrix given by:

$$\boldsymbol{\mathcal{I}}_M(\mathbf{H}) = \begin{bmatrix} \mathbf{H}(0) & \mathbf{H}(1) & \cdots & \mathbf{H}(L) & \mathbf{0} & \cdots & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{H}(0) & \mathbf{H}(1) & \cdots & \mathbf{H}(L) & \mathbf{0} & \cdots & \mathbf{0} \\ \vdots & \ddots & \ddots & \ddots & \ddots & & & \vdots \\ \mathbf{0} & \cdots & \cdots & \mathbf{0} & \mathbf{H}(0) & \mathbf{H}(1) & \cdots & \mathbf{H}(L) \end{bmatrix}$$

and $\mathbf{s}_M(k) = [\mathbf{s}^{\mathrm{T}}(k), \cdots, \mathbf{s}^{\mathrm{T}}(k-M+1-L)]^{\mathrm{T}}$, $\mathbf{b}_M(k) = [\mathbf{b}^{\mathrm{T}}(k), \cdots, \mathbf{b}^{\mathrm{T}}(k-M+1)]^{\mathrm{T}}$ and $\mathbf{h} = \mathrm{vec}\left([\mathbf{H}^{\mathrm{T}}(0), \cdots, \mathbf{H}^{\mathrm{T}}(L)]^{\mathrm{T}}\right)$. Let $\mathbf{Y}_M$ be the data matrix given by all vector $\mathbf{y}_M(k)$, i.e.,

$$\mathbf{Y}_M = [\mathbf{y}_M(M), \cdots, \mathbf{y}_M(n)] = \boldsymbol{\mathcal{I}}_M(\mathbf{h})\mathbf{S}_M + \mathbf{B}_M$$

where $\mathbf{s} = [\mathbf{s}^{\mathrm{T}}(1-L), \cdots, \mathbf{s}^{\mathrm{T}}(n)]$ and $\mathbf{S}_M$ is the $K(M+L) \times (n-M+1)$ block Toeplitz matrix given by:

$$\mathbf{S}_M = \begin{bmatrix} \mathbf{s}(M) & \cdots & \mathbf{s}(k) \\ \mathbf{s}(M-1) & \cdots & \mathbf{s}(k-1) \\ \vdots & \vdots & \vdots \\ \mathbf{s}(1-L) & \cdots & \mathbf{s}(k-M+1-L) \end{bmatrix}$$

Assume that both the channel input signal and the channel noise are white with zero mean with variances equal to 1 and σ_b^2. The channel output covariance matrix is therefore given by:

$$\mathbf{R}_M = \mathbb{E}\left[\mathbf{y}_M(k)\mathbf{y}_M^{\mathrm{H}}(k)\right] \tag{6.3}$$

$$= \boldsymbol{\mathcal{I}}_M(\mathbf{h})\boldsymbol{\mathcal{I}}_M^{\mathrm{H}}(\mathbf{h}) + \sigma_b^2\mathbf{I}_{MN} \tag{6.4}$$

In practice, the covariance matrix is estimated as:

$$\widehat{\mathbf{R}}_M = \frac{1}{n-M+1}\sum_{k=M}^{n}\mathbf{y}_M(k)\mathbf{y}_M^{\mathrm{H}}(k) \tag{6.5}$$

$$= \frac{1}{n-M+1}\mathbf{Y}_M\mathbf{Y}_M^{\mathrm{H}} \tag{6.6}$$

6.2 Indirect blind equalization techniques

6.2.1 Subspace based methods

The first use of blind methods for digital communications can be traced back to the work of Y. Sato where a simple linear equalizer for Pulse Amplitude Modulation symbols is described [91]. Nevertheless, it was not until 1980 that the principle of blind deconvolution techniques were first established in [7].

In particular, it was revealed in [7] that second order statistics are not sufficient to identify non-minimum phase signals. Since the channel of a SISO system is likely to be nonminimum phase, the identification problem was first addressed using higher order statistics of the channel output [42, 43]. This translates into slow-convergence for online methods or unreasonable data length required for off-line methods.

For this reason, the proposal of Tong *et al* [101], showing how to identify the channel by the sole use of second-order statistics was considered as a major contribution. Their method relies on oversampling the received signal thus leading to the so-called *fractionally-spaced equalization*.

Soon after, came the work of Moulines [75] *et al* , proving that under some mild technical assumptions, a SIMO channel transfer function can be retrieved by using a singular value decomposition (SVD) of the covariance matrix. It was shown later that this problem corresponds to the estimation of a minimal polynomial basis of a rational subspace, a problem often encountered in multivariate linear system theory [62]. By borrowing the tools of multivariate system theory, the subspace method was finally extended to MIMO systems in [74].

Let us now go into the details of this method. As previously mentioned, the covariance matrix can be decomposed into two orthogonal subspaces, namely the signal subspace corresponding to

nonzero singular values of $\boldsymbol{\mathcal{I}}(\mathbf{h})$ and the noise subspace corresponding to the null space of $\boldsymbol{\mathcal{I}}(\mathbf{h})$. Actually, by assuming that $NM \geq K(M+L)$ and that the channels have no zeros in common, we have Rank$(\boldsymbol{\mathcal{I}}(\mathbf{h}) = K(M+L)$. Let $\lambda_0 \geq \lambda_1 \geq \cdots \geq \lambda_{MN-1}$ denote the eigenvalues of $\mathbf{R}_M$. We have:

$$\begin{aligned} \lambda_i &> \sigma^2 \quad \text{for} \quad i = 0, \cdots, K(M+L)-1 \\ \lambda_i &= \sigma^2 \quad \text{for} \quad i = K(M+L), \cdots, MN-1 \end{aligned}$$

The covariance matrix $\mathbf{R}_M$ can be expressed as:

$$\mathbf{R}_M = \mathbf{U}_s \text{diag}\left(\lambda_0, \cdots, \lambda_{K(M+L)-1}\right) \mathbf{U}_s^{\text{H}} + \sigma^2 \mathbf{U}_b \mathbf{U}_b^{\text{H}}$$

where the columns of the matrix $\mathbf{U}_s$ span the signal subspace, while the columns of $\mathbf{U}_b$ span its orthogonal complement referred to as noise subspace. The subspace based technique consists in determining the factorization $\mathbf{Y}_M = \boldsymbol{\mathcal{I}}_M(\mathbf{h})\mathbf{S}_M$ that minimizes the distance between the vector space Range$(\mathbf{Y}_M)$ and Range$(\boldsymbol{\mathcal{I}}(\mathbf{h}))$ and under the constraint that $\boldsymbol{\mathcal{I}}_M(\mathbf{h})$ is block-Toeplitz. In other words, the subspace channel estimate is given by:

$$\widehat{\mathbf{h}} = \arg \min_{\mathbf{h}} \min_{\mathbf{A}} \|\boldsymbol{\mathcal{I}}_M(\mathbf{h}) - \mathbf{U}_s \mathbf{A}\|. \tag{6.7}$$

Minimizing first over $\mathbf{A}$, we get that the minimum is obtained when:

$$\mathbf{A} = \widehat{\mathbf{U}}_s^{\#} \boldsymbol{\mathcal{I}}_M(\mathbf{h})$$

so that (6.7) becomes:

$$\begin{aligned} \widehat{\mathbf{h}} &= \arg \min_{\mathbf{h}} \|\boldsymbol{\mathcal{I}}_M(\mathbf{h}) - \boldsymbol{\Pi}_{\text{sub}} \boldsymbol{\mathcal{I}}_M(\mathbf{h})\|^2 \\ &= \min_{\mathbf{h}} \|\boldsymbol{\Pi}_{\text{sub}} \boldsymbol{\mathcal{I}}_M(\mathbf{h})\|^2 \end{aligned}$$

where $\boldsymbol{\Pi}_{\text{sub}} = \mathbf{U}_b \mathbf{U}_b^{\text{H}}$ is the orthogonal projection matrix onto the noise subspace. Let us split the projection matrix $\boldsymbol{\Pi}_{\text{sub}}$ into M matrices such that $\boldsymbol{\Pi}_{\text{sub}} = [\boldsymbol{\Pi}_0, \boldsymbol{\Pi}_1, \cdots, \boldsymbol{\Pi}_{M-1}]$. Each matrix $\boldsymbol{\Pi}_i$ is $NM \times N$. We have, then:

$$\|\boldsymbol{\Pi}_{\text{sub}} \boldsymbol{\mathcal{I}}_M(\mathbf{h})\|^2 = \|\boldsymbol{\mathcal{D}}(\boldsymbol{\Pi}_{\text{sub}}) \begin{bmatrix} \mathbf{H}(0) \\ \vdots \\ \mathbf{H}(L) \end{bmatrix} \|^2,$$

where

$$\boldsymbol{\mathcal{D}}(\boldsymbol{\Pi}_{\text{sub}}) = \begin{bmatrix} \boldsymbol{\Pi}_0 & & \mathbf{0} \\ & \ddots & \\ \vdots & & \boldsymbol{\Pi}_0 \\ & & \\ \boldsymbol{\Pi}_{M-1} & & \vdots \\ & \ddots & \\ \mathbf{0} & & \boldsymbol{\Pi}_{M-1} \end{bmatrix} \tag{6.8}$$

The subspace channel estimate can be then derived by solving;

$$\widehat{\mathbf{h}} = \min_{\|\mathbf{h}\|=1} \mathbf{h}^{\text{H}} \mathbf{Q} \mathbf{h}. \tag{6.9}$$

where

$$\mathbf{Q} = \left(\mathbf{I}_K \otimes \boldsymbol{\mathcal{D}}(\boldsymbol{\Pi}_{\text{sub}})^{\text{H}}\right)\left(\mathbf{I}_K \otimes \boldsymbol{\mathcal{D}}(\boldsymbol{\Pi}_{\text{sub}})\right). \tag{6.10}$$

Note that the subspace method allows to identify the channel $\mathbf{H}(z)$ up to an invertible matrix ambiguity matrix $\mathbf{R}$. However, in many practical cases, the ambiguity matrix could be further simplified by referring to other assumptions such as having a channel matrix with known structure.

6.2.2 Linear Prediction based methods

The concept of linear prediction for SIMO systems was first introducted by Slock *et al* in [95], where a linear prediction approach was provided, when the channel order is known and the number of receiving antenna is equal to 2. This approach was afterwards generalized in [73], to encompass arbitrary number of receiving antennas and unknown channel order. The basic idea behind the generalized method arises from the observation that the system output can by fully described by an auto-regressive (AR) process, under some mild assumptions regarding the channel conditions.

In contrast to blind subspace and cross relation based methods which tend to exhibit high sensitivity if the channel order is incorrectly estimated, the linear prediction technique is found to be robust to channel order over-estimation [73, 72]. We provide hereafter insights about the linear prediction algorithm. For simplicity, we will consider the noiseless case, (when $\mathbf{b}(k) = 0$).

The principle of the linear prediction algorithm is tighly related to the generalized Bezout identity, [62]. This identity says that, if the channel is such that $\mathbf{H}(z)$ is full rank, i.e.,

$$\text{Rank}(\mathbf{H}(z)) = K \quad \forall z \neq 0$$

then there exists a $K \times M$ filter $\mathbf{G}(z)$ such that

$$\mathbf{G}(z)\mathbf{H}(z) = \mathbf{I}_K.$$

This implies that:

$$[\mathbf{G}(z)]\,\mathbf{y}(k) = \mathbf{s}(k)$$

In [73], it has been shown that $\mathbf{y}(k)$ is a finite AR process with order less than L. Moreover, its innovation noise is given by: $\mathbf{i}(k) = \mathbf{H}(0)\mathbf{s}(k)$. Therefore, there exists a linear prediction filter $\mathbf{P} = [\mathbf{P}_1, \cdots, \mathbf{P}_M]$ such that:

$$\mathbf{y}(k) = -\sum_{i=1}^{M} \mathbf{P}_i \mathbf{y}(k-i) + \mathbf{H}(0)\mathbf{s}(k)$$

Since the innovation noise is orthogonal to the space spanned by $\mathcal{H} = [\mathbf{y}(k-1), \cdots, \mathbf{y}(k-M)]$, the filter coefficient $\mathbf{P}_i$ can be determined as the ones that verify :

$$\mathbb{E}\left([\mathbf{P}(z)]\,\mathbf{y}(k)\,[\mathbf{y}^{\text{H}}(k-1), \cdots, \mathbf{y}^{\text{H}}(k-M)]\right) = 0$$

where $\mathbf{P}(z) = \mathbf{I}_N + \sum_{i=1}^{M} \mathbf{P}_i z^{-i}$ i.e,

$$[\mathbf{P}_1, \cdots, \mathbf{P}_M] = -\,[\mathbf{R}(1), \cdots, \mathbf{R}(M)]\,\mathbf{R}_{M-1}^{\#} \tag{6.11}$$

where $\mathbf{R}(i) = \mathbb{E}\,[\mathbf{y}(k)\mathbf{y}^{\text{H}}(k-i)]$ and $\mathbf{R}_{M-1} = [\mathbf{y}^{\text{H}}(k-1), \cdots, \mathbf{y}^{\text{H}}(k-M)]^{\text{H}}\,[\mathbf{y}^{\text{H}}(k-1), \cdots, \mathbf{y}^{\text{H}}(k-M)]$.

Also, by virtue of the orthogonality of the innovation noise and the space $\mathcal{H}$, the covariance of the innovation noise is the same as its covariance with the received signal $\mathbf{y}(k)$:

$$\mathbf{D} = \mathbb{E}\left[\mathbf{i}(k)\mathbf{i}^{\mathrm{H}}(k)\right] = \mathbb{E}\left[\mathbf{i}(k)\mathbf{y}^{\mathrm{H}}(k)\right] = \mathbf{H}(0)\mathbf{H}^{\mathrm{H}}(0) \tag{6.12}$$

Hence, the zero delay martix $\mathbf{H}(0)$ can be determined up to a nonsingular matrix ambiguity by using the covariance matrix of the innovation noise. The latter quanitity may be estimated by expanding the expression of the innovation noise, thus leading to:

$$\mathbf{D} = \mathbb{E}\left[\mathbf{i}(k)\mathbf{y}^{\mathrm{H}}(k)\right] = \mathbf{R}(0) + \sum_{i=1}^{M} \mathbf{P}_i \mathbf{R}^{\mathrm{H}}(i) \tag{6.13}$$

To sum up, the linear prediction algorithm consists in the following steps:

- First the linear prediction filter is estimated using (6.11).
- Matrix $\mathbf{D}$ is then derived using (6.13), $\mathbf{H}(0)$ can be determined up to an ambiguity matrix.
- Finally, the channel filter $\mathbf{H}(z)$ is obtained by using the following equality:

$$\mathbf{P}(z)\mathbf{H}(z) = \mathbf{H}(0) \tag{6.14}$$

As a matter of fact, (6.14) is equivalent to:

$$\delta(\mathbf{P}) \begin{bmatrix} \mathbf{H}(0) \\ \vdots \\ \mathbf{H}(L) \end{bmatrix} = \begin{bmatrix} \mathbf{H}(0) \\ \mathbf{0} \\ \vdots \\ \mathbf{0} \end{bmatrix}$$

where $\delta(\mathbf{P})$ denotes the $N(M+L+1) \times N(L+1)$ block-Toeplitz matrix given by:

$$\delta(\mathbf{P}) = \begin{pmatrix} \mathbf{I}_N & & & \mathbf{0} \\ \mathbf{P}_1 & \ddots & & \\ \vdots & \ddots & \ddots & \\ \mathbf{P}_M & \vdots & \ddots & \mathbf{I}_N \\ & \ddots & \vdots & \mathbf{P}_1 \\ & & \ddots & \vdots \\ \mathbf{0} & & & \mathbf{P}_M \end{pmatrix}.$$

Denote by $\mathbf{\Pi}_{\mathrm{lin}}$ the orthogonal projector on the null column space spanned by $\mathbf{D}$. Therefore, we have:

$$\mathbf{\Pi}_{\mathrm{lin}} \otimes \delta(\mathbf{P}) \begin{bmatrix} \mathbf{H}(0) \\ \vdots \\ \mathbf{H}(L) \end{bmatrix} = \mathbf{0}$$

thus giving:

$$\mathbf{I}_K \otimes \left(\mathbf{\Pi}_{\mathrm{lin}} \otimes \delta(\mathbf{P})\right) \mathbf{h} = \mathbf{0} \tag{6.15}$$

Once the channel is estimated, the recovering of the transmitted sequence can be performed by the viterbi equalizer in case optimal symbol detection is considered, or alternatively by means of suboptimal equalization techniques like MMSE or zero forcing.

MMSE equalizer The MMSE equalizer with delay τ is the one that performs the projection of $\mathbf{s}(k-\tau)$ on the space spanned by the received signal vectors. For SIMO systems, it can be expressed as:

$$\mathbf{v}_\tau = \arg\min_{\mathbf{v}} \mathbb{E}\left(\|s(k-\tau) - \mathbf{v}^{\mathrm{H}}\mathbf{y}_M(k)\|\right) = \mathbf{R}_M^{-1}\mathbf{g}_\tau \tag{6.16}$$

where $\mathbf{g}_\tau$ is the vector given by :

$$\mathbf{g}_\tau = \mathbb{E}(\mathbf{y}_M(k)s^*(k-\tau))\boldsymbol{\mathcal{I}}(\mathbf{h})(:,\tau+1)$$

Zero forcing equalizer When the window size is large enough, matrix $\boldsymbol{\mathcal{I}}(\mathbf{h})$ is invertible. The zero forcing equalizer with delay τ can be thus given by the $\tau+1$ column of the pseudo inverse matrix of $\boldsymbol{\mathcal{I}}(\mathbf{h})$. It has actually the following expression:

$$\mathbf{v}_\tau = \mathbf{U}_s \mathrm{diag}\left((\lambda_0-\sigma^2)^{-1},\cdots,(\lambda_{M+L-1}-\sigma^2)^{-1}\right)\mathbf{U}_s^{\mathrm{H}}\boldsymbol{\mathcal{I}}(\mathbf{h})(:,\tau+1) \tag{6.17}$$

In case the first channel coefficient is too low, the equalization process tends to estimate the $\tau+1$ delay signal sequence instead of the τ delay signal. If the delay is set at τ, the equalizer will fail to yield a good estimate of the channel. An appropriate solution would be to consider that the estimation is performed up to a certain unknown delay. In practice this issue can be dealt with by a priori synchronization process. It might be relevant to always consider the effective delay that minimizes the errors between the equalized signal and the delay source signal sequences.

6.3 Semi-blind estimation techniques

In practice, all the digital communication systems based on coherent detection use known symbols for either synchronization or channel estimation purposes. Future generation of communication systems do not seem to do without the transmission of these known symbols, either. Using only blind methods for channel estimation seem to be far in advance of the current state-of-the-art in the areas of communication technologies. Another interesting alternative consists in using all the information that one can get from the received frame. Since both data and training symbols are taken into account, this alternative is referred to as semi-blind channel estimation. It can yield optimal channel estimates if we use all the information that we can get about the channel. Such information is available from the received symbols corresponding to the training and data periods.

We assume hereafter that each frame is composed of a training period, corresponding to the transmission of m training symbols, and a data period, in which $n+M$ data symbols are transmitted. We add in general L samples so as to ensure that observations depend either on data or training symbols.

The received vector corresponding to training and data samples, is then given by:

$$\mathbf{y}_{n+m+L} = \boldsymbol{\mathcal{I}}_{n+m+L}(\mathbf{h})\mathbf{s}_{n+m+L} + \mathbf{b}_{n+m+L}$$

where $\mathbf{y}_{n+m+L} = [\mathbf{y}^{\mathrm{T}}(n+m+L),\cdots,\mathbf{y}^{\mathrm{T}}(1)]^{\mathrm{T}}$.

6.3.1 Optimal methods

In the literature, four methods that take different initial assumptions have been considered as optimal. They are:

1. Deterministic Maximum Likelihood (DML) [33] : In DML both the channel and the unknown symbols are considered as deterministic unknowns. Assuming that the noise is Gaussian, the DML criterion for SIMO systems can be written as:

$$\min_{\mathbf{h},\mathbf{s}_I} \|\mathbf{y}_{n+m+L} - \boldsymbol{\mathcal{I}}_{n+m+L}(\mathbf{h})\mathbf{s}\|$$

   where $\mathbf{s}_I$ are the unknown symbols in vector $\mathbf{s}$.

2. Gaussian Maximum Likelihood (GML) [32] : In GML, the input uknown symbols are assumed to be Gaussian distributed, whereas the channel and the noise variance are the parameters to be estimated. The GML criterion is determined by maximizing the probability density function of the observations.

3. Maximum Likelihood with finite alphabet constraints on the input symbols [88] FA-ML: The FA-ML considers that the unknown input symbols are deterministic and belong to a finite alphabet. The FA-ML criterion is then similar to that of DML and is given by:

$$\min_{\mathbf{h},\mathbf{s}_I \in \mathcal{A}} \|\mathbf{y}_{n+m+L} - \boldsymbol{\mathcal{I}}_{n+m+L}(\mathbf{h})\mathbf{s}\|$$

   where the elements of $\mathbf{s}_I$ are now constrained to belong to the finite alphabet $\mathcal{A}$.

4. Stochastic Maximum Likelihood (SML) [25, 8] : SML considers that the unknown input symbols are i.i.d random variables with zero mean and with values belonging to a discrete alphabet. The parameters to be estimated are the channel and the noise variance. The SML criterion is obtained by maximizing the conditional probability of the observations given the unknown input symbols. It is given by:

$$\min_{\mathbf{h},\sigma_b^2} \sum_{\mathbf{s}_I \in \mathcal{A}} \exp\left[-\frac{1}{\sigma_b^2}\|\mathbf{y}_{n+m+L} - \boldsymbol{\mathcal{I}}_{n+m+L}(\mathbf{h})\mathbf{s}\|^2\right]$$

   where σ_b^2 is the variance to be estimated. The direct optimization of the SML is highly expensive in terms of computing time. The Expectation-Maximization (EM) iterative algorithm was proposed to solve the SML criterion by using the Hidden Markov Model framework. The E-M algorithm was shown to converge to the SML solution provided a good initialization. In the semi-blind case, the initialization can be taken as the one given by the least square estimate.

6.3.2 Suboptimal methods

Among all these methods, the SML technique is the one that fits the best to the theoretical SIMO model. However from a practical point of view, the implementation of the E-M algorithm induces a high computational complexity that grows exponentially with the channel length and the receiving antenna-array size.

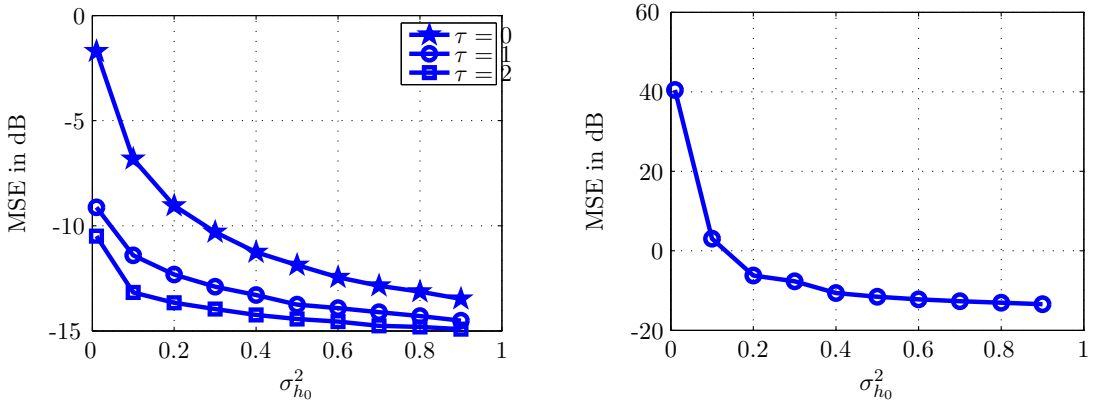
A suboptimal alternative consists in minimizing a weighted sum of the training and the blind cost functions. It was proposed in [48] and was referred to as regularized semi-blind channel estimation method. This technique exhibits low computational complexity, but its performance is strongly influenced by the tunining of the regularizing constant. Based on an asymptotic analysis, the work in [15] proposes to set the regularizing constant in such a way to minimize the asymptotic channel estimation error, while the asymptotic regime refers to m, n growing to infinity at the same rate. More insights about this method will be provided in chapter 10.

6.4 Numerical results

The linear prediction and subspace based method have different strengths and weaknesses regarding several practical issues. For instance, the linear prediction technique is sensitive to the power of the first coefficient but is robust to channel order over-estimation, while its subspace counterpart exhibit high errors when the channel is over-estimated but outperforms the linear prediction technique in case the channel order is known. Hereafter, we investigate on these practical issues by using simulations. We consider a SIMO model with $N = 4$, $L = 3$, $M = 11$ and $n = 600$. We measure the experienced MSE in the symbol estimation, when using the MMSE equalizer described by (6.16). Moreover, in all experiments, we use BPSK constellation and we set the SNR to 15 dB.

6.4.1 Sensitivity to the power of the first channel coefficient

Fig. 6.1 shows the experienced MSE in dB for linear prediction and subspace based methods with respect to the variance of the first channel coefficient. For the subspace based method, the three delays $\tau = 0$, $\tau = 1$ and $\tau = 2$ have been considered. As for the linear prediction method, only the delay $\tau = 0$ is investigated, since the performance of this method depends to a large extent on the zero delay equalizer. The other equalizers exhibit almost the same performance.



(a) Subspace based method

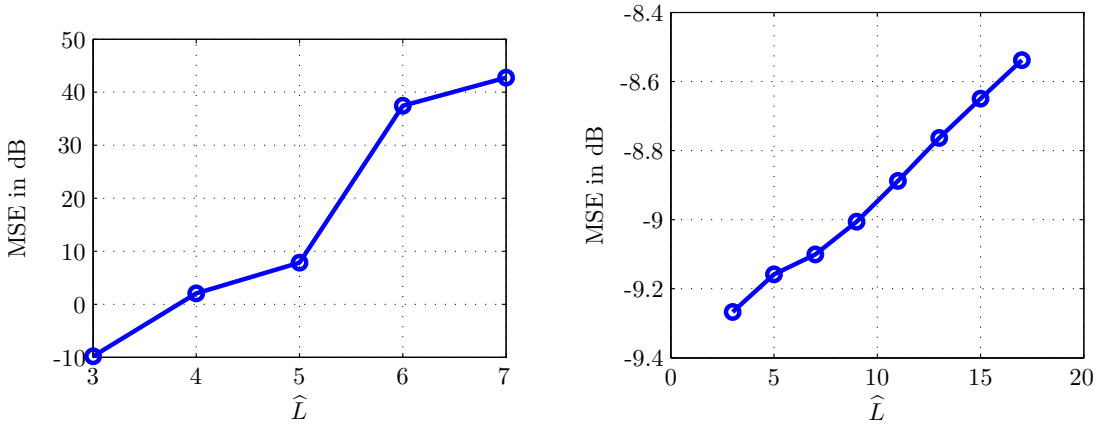
(b) Linear prediction based method

Figure 6.1: Sensitivity to the variance of $\mathbf{h}(0)$

We note that both methods are sensitive to the variance of the first coefficient but the linear prediction is the one that exhibits the largest errors.

6.4.2 Channel order over-estimation

Fig. 6.2 displays the MSE for the linear prediction and subspace based method with respect to the overestimated channel order. We note that the linear prediction method is robust to channel order over-estimation, whereas the subspace based method, exhibits a high MSE even when the channel order is overestimated by 1 or 2 taps.



(a) Subspace based method

(b) Linear prediction based method

Figure 6.2: Sensitivity to the channel order over-estimation

6.5 Conclusion

In this chapter, we have reviewed a selection of well known results about subspace and linear prediction methods. We have particularly pointed out practical considerations regarding the high sensitivity of subspace based methods to channel order over-estimation and that of linear prediction techniques to the variance of the first channel coefficient. We have also reviewed semi-blind techniques, which can be regarded as intermediate solutions between training and blind based techniques.

Chapter 7

Blind Nonzero Delay MMSE Equalizer for SIMO FIR Systems

Despite being proposed several decades ago, blind methods failed so far to be commercially deployed for civil communication products. This is attributed to several practical difficulties, limiting their use in practice to specific applications where training data is unavailable. Actually, several blind methods and in particular indirect blind approaches are based on singular value decompositions of the output correlation matrix, thus making their real-time implementation difficult, [114]. They also require a-priori reliable channel order estimation, otherwise symbol detection is not possible.

On the other hand, direct blind methods, can be implemented by using adaptive algorithms that exhibit low-cost computational complexity, thereby making them more appropriate for real-time implementations. However, the major challenge that still faces the research community is to make them robust to channel order over-modeling while ensuring a reasonable complexity.

Many previous works investigated this issue. The techniques proposed so far either require the estimation of several equalizers like the mutually referenced filter based method [29], or involve the estimation of a unique equalizer such as linear prediction based techniques [24, 114]. The latter methods are robust to channel order overmodeling but their adaptive implementations are either expensive (e.g RLS like) or slowly convergent (e.g. LMS like).

Later, new methods based on the truncation of the covariance matrix have been proposed [60, 61]. These techniques yield zero delay equalizers through performing an appropriate truncation to the covariance matrix. Despite their high robustness to channel overmodeling, these methods do not always yield satisfactory results, since they involve zero delay equalization and thus are sensitive to the value of the first channel coefficient.

In this chapter, we propose to generalize the methods in [60] to nonzero delay equalization. We show by using simulations that our method allows significant performance improvement, while maintaining robustness to channel order over-modeling.

7.1 System model

We consider the discrete time (SIMO) model, with N outputs, given by:

$$\mathbf{y}(k) = \sum_{l=0}^{L} \mathbf{h}(l)s(k-l) + \mathbf{b}(k)$$

where $s(k)$ denotes the transmitted symbol sequence and $\mathbf{h}(l)$ refers to the $N \times 1$ channel vector corresponding to the l-th tap. $\mathbf{b}(k)$ denotes the white noise sequence with variance σ_b^2. Stacking M successive observations of the received signal $\mathbf{y}(k)$ into a single vector, we get:

$$\begin{aligned}\mathbf{y}_M(k) &= [\mathbf{y}^{\mathrm{T}}(k), \cdots, \mathbf{y}^{\mathrm{T}}(k-M+1)]^{\mathrm{T}} \\ &= \boldsymbol{\mathcal{I}}_M(\mathbf{h})\mathbf{s}_M(k) + \mathbf{b}_M(k)\end{aligned}$$

where $s_M(k) = [s(k), \cdots, s(k-M-L+1)]^{\mathrm{T}}$ and $\mathbf{b}_M(k) = [\mathbf{b}^{\mathrm{T}}(k), \cdots, \mathbf{b}^{\mathrm{T}}(k-M+1)]^{\mathrm{T}}$. The matrix $\boldsymbol{\mathcal{I}}_M(\mathbf{h})$ is the $NM \times M+L$ block-Toeplitz matrix given by:

$$\boldsymbol{\mathcal{I}}_M(\mathbf{h}) = \begin{bmatrix} \mathbf{h}(0) & \cdots & \mathbf{h}(L) & & \mathbf{0} \\ & \ddots & & \ddots & \\ \mathbf{0} & & \mathbf{h}(0) & \cdots & \mathbf{h}(L) \end{bmatrix}. \tag{7.1}$$

In the sequel, we will consider the following additive assumptions:

A1. *The transmitted signal $s(k)$ is assumed to be an independent and identically distributed zero mean unit power process.*

A2. *The polynomial $\mathbf{h}(z) = \sum_{k=0}^{L} \mathbf{h}_k z^{-k}$ verifies:*

$$\mathbf{h}(z) \neq 0 \text{ for all } z \neq 0,$$

or equivalently, the sub-channels corresponding to the receiving antennas do not share any zero in common.

7.2 Problem statement: MMSE equalization

We recall hereafter the equalization technique that has been proposed in [60]. The main features of this technique are its highest robustness to channel overmodeling and also its low computational complexity as compared to other proposed techniques.

As we will see below, this method is based on the fact that the τ-delay MMSE equalizer belongs to a certain vector space that depends solely on the covariance matrix, and whose dimension is equal to $\tau+1$. As a consequence, if $\tau = 0$, the zero-delay MMSE equalizer can be estimated up to a scalar ambiguity.

For the reader's convenience, we provide in the sequel an overview on the main results derived in [60]. Let us first recall that the τ-delay linear MMSE equalizer ($\tau \in \{0, \cdots, M+L-1\}$) is the optimal linear filter that extracts $s(k-\tau)$ in the least square sense.

More explicitly, the linear MMSE equalizer vector $\mathbf{v}_\tau$ is given by:

$$\mathbf{v}_\tau = \arg\min_{\mathbf{v}} \mathbb{E}\left(\|s(k-\tau) - \mathbf{v}^{\mathrm{H}}\mathbf{y}_M(k)\|^2\right),$$

which leads to:

$$\mathbf{v}_\tau = \mathbf{R}_M^{-1}\mathbf{g}_\tau, \tag{7.2}$$

where

$$\mathbf{R}_M = \mathbb{E}\left(\mathbf{y}_M(k)\mathbf{y}_M^{\mathrm{H}}(k)\right) = \boldsymbol{\mathcal{I}}_M(\mathbf{h})\boldsymbol{\mathcal{I}}_M^{\mathrm{H}}(\mathbf{h}) + \sigma_b^2\mathbf{I}_{MN}$$

and

$$\mathbf{g}_\tau = \mathbb{E}\left(\mathbf{y}_M(k)s^*(t-\tau)\right) = \boldsymbol{\mathcal{I}}_M(\mathbf{h})(:, \tau+1).$$

The covariance matrix $\mathbf{R}_M$ is estimated in practice as:

$$\widehat{\mathbf{R}}_M = \frac{1}{n}\sum_{k=1}^{n} \mathbf{y}_M(k)\mathbf{y}_M^{\mathrm{H}}(k)$$

where n denotes the sample size. One can note that the linear MMSE equalizer belongs to the signal subspace, i.e., Range($\boldsymbol{\mathcal{I}}_M(\mathbf{h})$) and thus can be written as:

$$\mathbf{v}_\tau = \mathbf{W}\widetilde{\mathbf{v}}_\tau,$$

where $\mathbf{W}$ denotes the signal subspace basis vectors. Along the same lines as in [60], we can prove the following result:

Theorem 7.1. *Let $\underline{\mathbf{R}}_{M,\tau}$ be the matrix given by the last $MN-(\tau+1)N$ rows of $\mathbf{R}_M$. Then, assuming* **A1**, **A2**, *and that $M > L+1+\tau$, the kernel of matrix $\mathbf{W}^{\mathrm{H}}\underline{\mathbf{R}}_{M,\tau}^{\mathrm{H}}\underline{\mathbf{R}}_{M,\tau}\mathbf{W}$ has dimension $\tau+1$ and contains all the t-delay equalizers $\mathbf{v}_t$, $t \in \{0,\cdots,\tau\}$.*

The proof of this theorem has not been provided before, but it relies on the same technique used in [60]. For the sake of completeness, we provide hereafter the proof of this theorem:

Proof. Let $\mathbf{R}_M = \mathbf{W}\boldsymbol{\Gamma}\mathbf{W}^{\mathrm{H}} + \sigma_b^2\mathbf{U}\mathbf{U}^{\mathrm{H}}$ be the eigenvalue decomposition of $\mathbf{R}_M$, where $\mathbf{W}$ and $\mathbf{U}$ are the eigenvectors that span, respectively, the signal and the noise subspace. Since, the columns of $\mathbf{W}$ and $\mathbf{U}$ are orthogonal, we have:

$$\mathbf{R}_M\mathbf{W} = \mathbf{W}\boldsymbol{\Gamma}.$$

Hence, Range($\mathbf{R}_M\mathbf{W}$) = Range($\mathbf{W}$) = Range($\boldsymbol{\mathcal{I}}_M(\mathbf{h})$). Therefore, there exists a nonsingular matrix $\mathbf{P}$ such that $\mathbf{R}_M\mathbf{W} = \boldsymbol{\mathcal{I}}_M(\mathbf{h})\mathbf{P}$.

As a consequence,

$$\underline{\mathbf{R}}_{M,\tau}\mathbf{W} = \left[0_{(M-\tau-1)N\times\tau+1}\ \boldsymbol{\mathcal{I}}_{M-\tau-1}(\mathbf{h})\right]\mathbf{P}.$$

We end up the proof by noting that if $M-\tau-1 > L$, $\boldsymbol{\mathcal{I}}_{M-\tau-1}(\mathbf{h})$ is full column rank [75] and thus dim(null($\underline{\mathbf{R}}_{M,\tau}\mathbf{W}$)) $= \tau+1$. □

Example: Zero Delay Equalizer

Let $\mathbf{v}_0$ denote the zero delay equalizer. There exists a vector $\widetilde{\mathbf{v}}_0$ such that $\mathbf{v}_0 = \mathbf{W}\widetilde{\mathbf{v}}_0$, with

$$\mathbf{R}_M\mathbf{W}\widetilde{\mathbf{v}}_0 = \mathbf{g}_0 = \begin{bmatrix}\mathbf{h}(0)\\ \mathbf{0}\\ \vdots\\ \mathbf{0}\end{bmatrix}.$$

Let $\underline{\mathbf{R}}_{M,0}$ be the matrix given by the last $MN-M$ rows of $\mathbf{R}_M$. The following result has been shown in [60] and can be easily deduced from Theorem 7.1.

Corollary 7.1. *Assuming that $M > L+1$, the solution of*

$$\underline{\mathbf{R}}_{M,0}\mathbf{W}\widetilde{\mathbf{v}} = \mathbf{0},$$

is unique and corresponds to the desired zero-delay equalizer vector $\mathbf{v}_0$ up to a scalar ambiguity.

In other words, this result states that the zero delay equalizer can be determined as the intersection line between the range space of $\mathbf{W}$ and the kernel of the matrix $\underline{\mathbf{R}}_{M,0}$. The algorithm that has been proposed in [60] can be summarized into the following steps:

Algorithm 3 SIMO blind MMSE equalization algorithm

$\widehat{\mathbf{R}}_M \longleftarrow \frac{1}{n}\sum_{k=1}^{n} \mathbf{y}_M(k)\mathbf{y}_M^{\mathrm{H}}(k)$
$[\mathbf{W}, \mathbf{\Gamma}] \longleftarrow \mathrm{eigs}(\widehat{\mathbf{R}}_M, M+L)$
$\mathbf{T} \longleftarrow \widehat{\mathbf{R}}_M(N+1:end,:)\mathbf{W}$
$\mathbf{Q} \longleftarrow \mathbf{T}^{\mathrm{H}}\mathbf{T}$
$\widetilde{\mathbf{v}}_0$ the least eigenvector of $\mathbf{Q}$
$\mathbf{v} \longleftarrow \mathbf{W}\widetilde{\mathbf{v}}_0$

In the nonzero delay case, this intersection becomes a vector space of dimension $\tau+1$, that contains all the t delay equalizers, $t \in \{0, \cdots, \tau\}$. The main issue is how to select in this vector space the direction of the desired τ delay equalizer. In [60], a two-step approach is proposed: first, the zero delay equalizer is estimated and the transmitted symbols are decoded by performing a hard decision on the equalized signal. After that, the estimated symbols are reused to reestimate a nonzero delay equalizer according to (7.2).

However, one may expect that this technique will not provide good performance, since as soon as the result of the first step is bad, all the process that comes after, will be affected.

7.3 NonZero delay equalization

The performance of the zero delay equalizer is poor when the energy of the first channel coefficient $\mathbf{h}(0)$ is low. In average, the nonzero delay equalizer has better performance since it depends on more than one channel tap and thus can benefit from the channel path diversity. Motivated by this well known result [84], we propose in this section to extend the work of [60] to the nonzero delay case.

Let $\mathcal{F}_\tau$ denote the vector space given by the intersection between the kernel of $\underline{\mathbf{R}}_{M,\tau}$ and the range of $\mathbf{W}$, and $\mathbf{B}_\tau = \{\tilde{\mathbf{v}}_0, \cdots, \tilde{\mathbf{v}}_\tau\}$ a basis of $\mathcal{F}_\tau$. Then, obviously, we have $\mathcal{F}_{\tau-1} \subset \mathcal{F}_\tau$. More particularly, a basis of $\mathcal{F}_\tau$ could be given as the union of $\mathbf{B}_{\tau-1}$ and another vector $\tilde{\mathbf{v}}_\tau$ which cannot be written as a linear sum of elements of $\mathbf{B}_{\tau-1}$.

The difficulty here is to select the right direction $\tilde{\mathbf{v}}_\tau$, which corresponds to the τ-delay equalizer. To solve this problem, we will make use of an approximate orthogonality relation between the equalizer vectors, which is accurate as long as the mean square symbol estimation error is low. Indeed, since $\mathbf{v}_t^{\mathrm{H}}\mathbf{y}_M(k) \simeq s(k-t)$ and $\mathbf{v}_{t'}^{\mathrm{H}}\mathbf{y}_M(k) \simeq s(k-t')$, and since the input symbols are i.i.d, we get:

$$0 = \mathbb{E}\left(s(k-t)s^*(k-t')\right) \simeq \mathbf{v}_t^{\mathrm{H}}\mathbf{R}_M\mathbf{v}_{t'}. \tag{7.3}$$

More explicitely, if $\mathbf{R}_M = \mathbf{W}\mathbf{\Gamma}\mathbf{W}^{\mathrm{H}} + \sigma_b^2\mathbf{U}\mathbf{U}^{\mathrm{H}}$ denotes the eigenvalue decomposition of $\mathbf{R}$, then, since $\mathbf{v}_t$ belongs to the range space of $\mathbf{W}$, (7.3) is equivalent to:

$$\mathbf{v}_t^{\mathrm{H}}\mathbf{W}\mathbf{\Gamma}\mathbf{W}^{\mathrm{H}}\mathbf{v}_{t'} \simeq 0.$$

Let $\breve{\mathbf{v}}_\tau = \mathbf{\Gamma}^{\frac{1}{2}}\mathbf{W}^{\mathrm{H}}\mathbf{v}_\tau$, i.e, $\mathbf{v}_\tau = \mathbf{W}^{\mathrm{H}}\mathbf{\Gamma}^{\frac{-1}{2}}\breve{\mathbf{v}}_\tau$, then, it is clear that $\breve{\mathbf{v}}_\tau$ is approximatively orthogonal to $\breve{\mathbf{v}}_{\tau'}$, for $\tau' \neq \tau$. By performing the variable change $\breve{\mathbf{v}}_\tau = \mathbf{\Gamma}^{\frac{1}{2}}\mathbf{W}^{\mathrm{H}}\mathbf{v}_\tau$, the desired direction

can be selected to be the one that is orthogonal to the range space of $\mathcal{F}_{\tau-1}$. To sum up, our algorithm consists in the following steps:

Algorithm 4 Proposed algorithm: Non zero-delay equalizer for SIMO systems

$\widehat{\mathbf{R}}_M \longleftarrow \frac{1}{n}\sum_{k=1}^{n}\mathbf{y}_M(k)\mathbf{y}_M^{\mathrm{H}}(k)$ // Estimation of the covariance matrix
$[\mathbf{W},\mathbf{\Gamma}] \longleftarrow \text{eigs}\left(\widehat{\mathbf{R}}_M, M+L\right)$ // Estimation of the signal subspace
$\mathbf{T}_\tau \longleftarrow \widehat{\mathbf{R}}_{M,\tau}\mathbf{W}\mathbf{\Gamma}^{-\frac{1}{2}}$
$\mathbf{E} = [\mathbf{e}_0,\cdots,\mathbf{e}_\tau] \longleftarrow \text{eigs}\left(\mathbf{T}_\tau^{\mathrm{H}}\mathbf{T}_\tau, \tau+1, 0\right)$ // Estimation of the $\tau+1$ least eigenvectors of $\mathbf{T}_\tau^{\mathrm{H}}\mathbf{T}_\tau$
$\mathbf{T}_{\tau-1} \longleftarrow \widehat{\mathbf{R}}_{M,\tau-1}\mathbf{W}\mathbf{\Gamma}^{-\frac{1}{2}}$
$\mathbf{F} = [\mathbf{f}_0,\cdots,\mathbf{f}_{\tau-1}] \longleftarrow \text{eigs}\left(\mathbf{T}_{\tau-1}^{\mathrm{H}}\mathbf{T}_{\tau-1}, \tau, 0\right)$ // Estimation of the τ least eigenvectors of $\mathbf{T}_{\tau-1}^{\mathrm{H}}\mathbf{T}_{\tau-1}$
$\breve{\mathbf{v}}_\tau = \text{eigs}(\mathbf{E}^{\mathrm{H}}\mathbf{F}\mathbf{F}^{\mathrm{H}}\mathbf{E})$ // Selection of the direction that is contained in the range of $\mathbf{E}$ and that is orthogonal to the range of $\mathbf{F}$
$\mathbf{v}_\tau \longleftarrow \mathbf{W}^{\mathrm{H}}\mathbf{\Gamma}^{-\frac{1}{2}}\breve{\mathbf{v}}_\tau$

Before going further, some remarks should be pointed out:

Remark 7.1. *One can note that if $\tau = 0$, our algorithm will be reduced to the following steps:*

- *Computation of the autocorrelation matrix.*
- *Estimation of $\mathbf{T}_0 = \mathbf{R}_0\mathbf{W}\mathbf{\Gamma}^{-\frac{1}{2}}$ and $\breve{\mathbf{v}}_0$ the least eigenvector of $\mathbf{T}_0^{\mathrm{H}}\mathbf{T}_0$. The zero delay equalizer is therefore given by:*

$$\mathbf{v}_0 = \mathbf{W}^{\mathrm{H}}\mathbf{\Gamma}^{-\frac{1}{2}}\breve{\mathbf{v}}_0.$$

Remark 7.2. *In case of channel overmodeling, the matrix $\mathbf{\Gamma}^{-\frac{1}{2}}$ plays an important role in ensuring the robustness of the proposed method. Indeed, at moderate or high SNR, it discards the eigenvectors that lie in the noise subspace because they are weighted by $\frac{1}{\sigma_b}$.*

Remark 7.3. *In this work, we did not seek efficiency in terms of computational cost. Indeed, we can use the relation between $\mathbf{T}_\tau$ and $\mathbf{T}_{\tau-1}$ to eliminate one of the SVDs and reduce the algorithm's complexity. Also, as in [60], we can use efficient subspace tracking techniques in an adaptive scheme in order to reach linear or close to linear complexity per iteration.*

7.4 Simulation results

In all our simulations, we consider a SIMO model with $N = 4$ receiving antennas and $L+1 = 4$ channel coefficients chosed randomly according to the Rayleigh distribution. The input signal is drawn from the BPSK constellation and the temporal window M is set to 11. We measure the system performance by using the mean square error (MSE) given by:

$$\text{MSE} = 1 - \mathbb{E}\frac{|\widehat{\mathbf{s}}^{\mathrm{H}}\mathbf{s}|^2}{\|\mathbf{s}\|^2\|\widehat{\mathbf{s}}\|^2}.$$

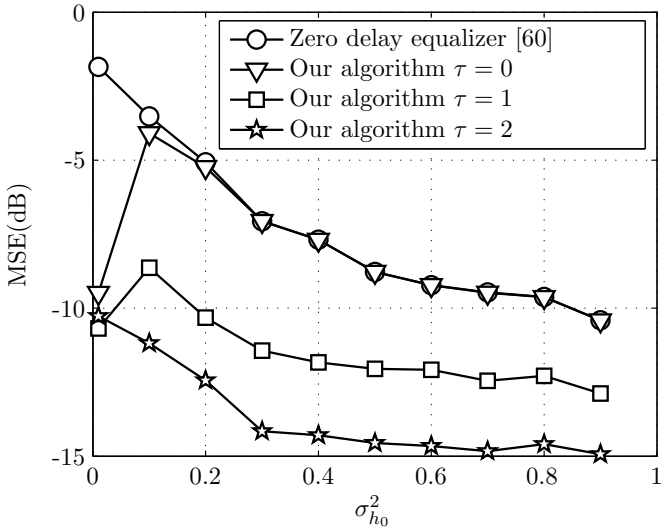


Figure 7.1: Sensitivity to the variance of $\mathbf{h}(0)$.

7.4.1 Sensitivity to the variance of $\mathbf{h}(0)$

For small values of the first channel coefficient, the equalization process tends to estimate the $\tau + 1$ delay signal sequence. Therefore, as mentioned in chapter 6, we consider that our method estimates the transmitted sequence up to a certain unknown delay. This assumes that a priori synchronization process has been used to properly estimate the effective delay corresponding to the one that errors between the equalized signal and the delay source signal sequences. Moreover, we expect that even for a non zero delay equalizer, our method should exhibit some sensitivity towards the variance of the first channel coefficient, because the matrices $\mathbf{T}_\tau$ and $\mathbf{T}_{\tau-1}$ may have low singular values thus introducing wrong directions.

Fig. 7.1 displays the MSE of our algorithm, and compares it with that of the algorithm in [60]. The MSE vs the variance of the channel $\sigma^2_{h_0}$ is represented for different values of τ.We note that when the variance of $\mathbf{h}(0)$ is too low, the proposed algorithm exhibits a small degradation in the mean square error performance, as compared to the algorithm in [60]. Besides, for $\tau = 0$ and very low channel coefficient variance $\sigma^2_{\mathbf{h_0}} = 0.01$, our algorithm is able to switch to the delay $\tau = 1$, thus explaining its good performance in this case.

7.4.2 Robustness to channel order overmodeling

We investigate in this section the robustness of our algorithm to channel order overmodeling. Fig. 7.2 compares the MSE with respect to the estimated channel order for the proposed equalization process (when $\tau = 2$) and the zero-delay equalizer that is proposed in [60], when the SNR is set to 0 dB and 10 dB, respectively. We note that like the zero delay equalizer, our algorithm is robust to the channel order overestimation, even at low SNR values.

7.4.3 Effect of the delay

In this section, we investigate the effect of the equalizer delay on the MSE performance of our algorithm. Fig. 7.3 displays the MSE with respect to the equalizer delay at different SNR values. We can see that gain of almost 6 dB in MSE can be otained by increasing the equalizer delay.

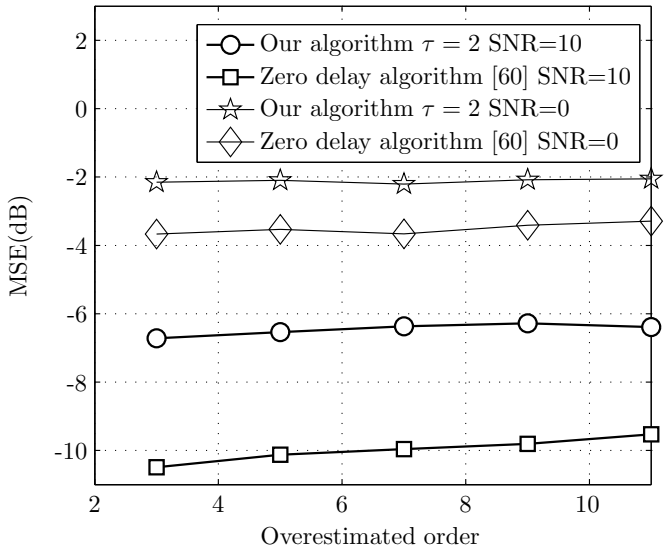


Figure 7.2: Evaluation of the robustness of the proposed algorithm.

Moreover, as expected, when τ is set higher than $L+1 = 4$, the performance enhancement is not significant, in other words, the wide range of gain is approximately achieved when $\tau \geq L + 1$.

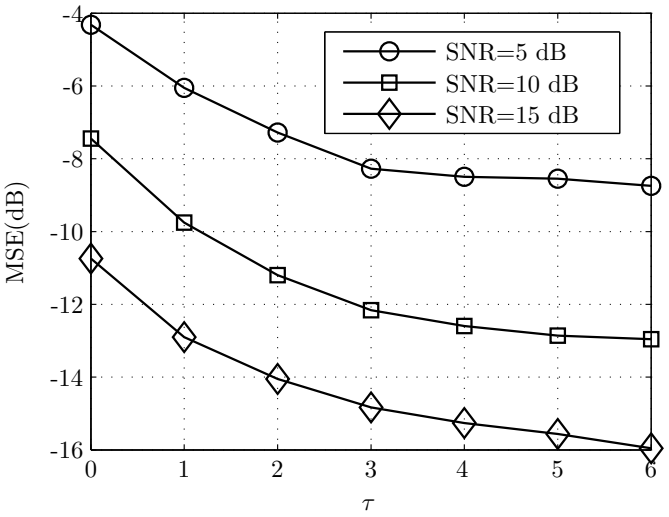


Figure 7.3: MSE with respect to the equalizer delay for different SNR values.

7.4.4 Iterative decoding

As we have previously mentioned, the proposed algorithm is somehow sensitive to the variance of the first channel coefficient (see Fig. 7.1). In order to further enhance the bit error rate performance, we propose to use iterative decoding that makes use of the estimated a posteriori probabilities. Fig. 7.4 summarizes the iterative process. Given the τ delay equalizer estimate,

we compute the variance of the noise $\widehat{\sigma}_n^2 = 1 - \widehat{\mathbf{v}}_\tau^{\mathrm{H}} \widehat{\mathbf{R}}_M \mathbf{v}_\tau$. We also estimate the transmitted signal sequence $\mathbf{s}_{k-\tau} \triangleq [s(M-\tau), \cdots, s(k-\tau)]$ by $\mathbf{v}_\tau^{\mathrm{H}} \mathbf{Y}_M(k)$, where $\mathbf{Y}_M(k) = [\mathbf{y}_M(M), \cdots, \mathbf{y}_M(k)]$. For BPSK constellation, it can be easily shown that the a posteriori probability on the transmitted bits is given by:

$$p_k = \mathrm{APP}(s_k = 1) \triangleq \mathrm{P}(s_k = 1 | \widehat{s}_k) = \frac{1}{1 + \exp(-\frac{2\widehat{s}_k}{\widehat{\sigma}_k^2})}.$$

The vector $\mathbf{g}_\tau$ can be estimated by using hard decision on the equalized signal as in [60]. However, this is not optimal in the sense that the non reliable entries will have the same contribution as the reliable ones. Using the a posteriori probabilities, the vector $\mathbf{g}_\tau$ can be estimated as:

$$\begin{aligned} \widehat{\mathbf{g}}_\tau &= \frac{1}{K} \sum_{k=1}^{n} \mathbf{y}_M(k) \mathbb{E}\left[\widehat{s}_{k-\tau}\right] \\ &= \frac{1}{K} \sum_{k=1}^{n} \mathbf{y}_M(k) \left(2p_{k-\tau} - 1\right). \end{aligned}$$

Fig. 7.6 and Fig. 7.5 display the bit error rate performance of our algorithm for $\tau = 6$ with the soft iterative as well as the hard iterative processing [60]. Fig. 7.6 displays also the bit error rate performance of the zero delay MMSE equalizer and shows that a large gain is obtained by using our technique. In the legend 'ideal MMSE' refers to the genie MMSE equalizer which exactly knows the correlation matrix $\mathbf{R}_M$ and the correlation vector $\mathbf{g}_\tau$. We note that with soft iterative processing, the performance of our algorithm can become very close to that of the ideal receiver.

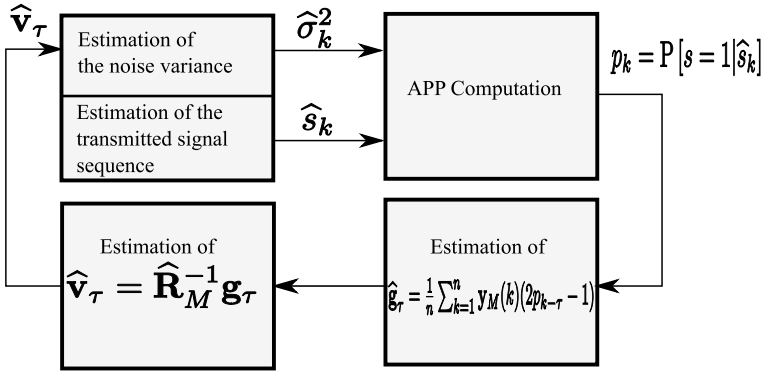


Figure 7.4: Iterative decoding for BPSK constellation.

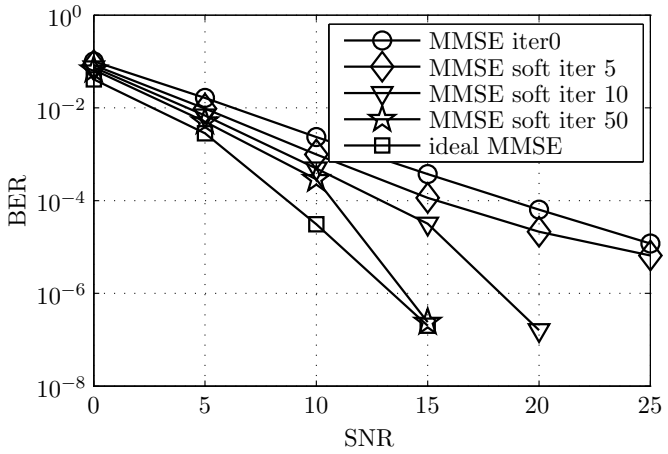


Figure 7.5: Bit error rate with soft iterative processing.

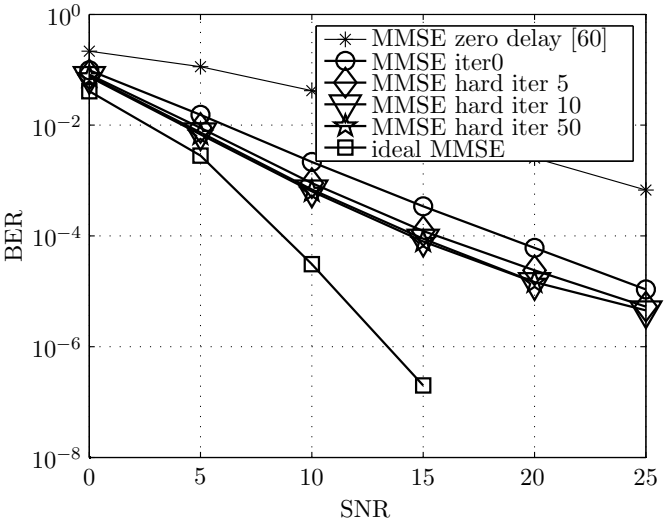


Figure 7.6: Bit error rate with hard iterative processing.

7.5 Conclusion

In this chapter, we have generalized a recently proposed zero delay blind equalization to arbitrary delay equalization. Unlike the original method, our technique is much less sensitive to the variance of the first channel coefficient and it inherits the same interesting properties regarding the robustness to channel over-modeling. Moreover, by using an iterative decoding algorithm, our technique can reach the performance of an ideal equalizer.

Chapter 8

Robustness of Blind subspace based technique using ℓ_p quasi-norms

One of the most major shortcomings of blind methods is their sensitivity to channel order over-modeling. As long as the channel order is correctly estimated, the channel can be uniquely identified using blind methods, but once an error on the estimation of the channel order occurs, identifiability is no longer possible for many existing blind methods. This is for instance the case of conventional subspace-based methods, which are known to exhibit a significant sensitivity to channel order over-modeling [74]. Actually, in the noiseless case, the channel can be identified as the vector that spans the 1-dimensional kernel of a matrix denoted by $\mathbf{Q}$ that can be estimated by using solely second-order statistics, (See chapter 6). But when the channel order is overestimated, the kernel of the matrix $\mathbf{Q}$ is no longer a line, but rather a vector space whose dimension depends on the overestimated order, thereby raising a new issue: how to estimate the right direction among all the vectors that span the kernel of $\mathbf{Q}$?

To deal with this problem, a large effort has been devoted to either add to conventional subspace techniques a feature that estimates efficiently the channel order [47], or to propose new methods that are robust to channel order over-modeling. In this context, a new technique for blind channel estimation for sparse channels has been recently proposed. This technique takes into account the sparsity criterion so as to select among all possible vectors, the vector that exhibits the lowest ℓ_p, $0 < p \leq 1$ quasi-norm. It was noted that using this technique, cross relation as well as blind deterministic maximum likelihood based methods become robust to channel-order over-modeling as far as sparse channels are concerned [1, 9]. However, for non-sparse channels, no results are available so far, to the best of our knowledge. Yet, we strongly believe that introducing likewise a sparsity criterion shall enhance the channel identifiability probability. Actually, in this case, one can note that over-modeling the channel is equivalent to zero-padding the channel vector, thus making it artificially sparse. Moreover, as far as subspace methods are concerned, it can be shown that the zero-padded channel vector is the one that exhibits the most sparsity. In light of this consideration, we claim that selecting the vector vector that minimizes the ℓ_p quasi-norm should often yield the desired channel vector response (up to a scalar ambiguity).

In this chapter, we propose to study the robustness of subspace based methods using ℓ_p quasi-norms for non-sparse channels. We derive necessary and sufficient condition for channel identifiability when considering the ℓ_1 norm as well as a sufficient condition when considering the ℓ_p quasi-norm $0 < p < 1$. Then, we derive lower bounds on the probability that these conditions are satisfied. Using these lower bounds, we study the effect of the system parameters on the

channel identifiability probability. For instance, we note that in the ℓ_1 problem, increasing the number of antennas improves significantly the channel identifiability probability, in contrast to increasing the number of channel coefficients, which tends to reduce it.

8.1 System model and problem statement

We consider the same SIMO system model as in chapter 7.

$$\mathbf{y}_M(k) = \boldsymbol{\mathcal{I}}_M(\mathbf{h})\mathbf{s}_M(k) + \mathbf{b}_M(k)$$

As it has been shown in chapter 6, the blind subspace estimator is given by:

$$\widehat{\mathbf{h}} = \min_{\|\mathbf{f}\|_2=1} \mathbf{f}^{\mathrm{H}}\mathbf{Q}\mathbf{f},$$

with

$$\mathbf{Q} = \boldsymbol{\mathcal{D}}(\boldsymbol{\Pi})^{\mathrm{H}}\boldsymbol{\mathcal{D}}(\boldsymbol{\Pi})$$

where $\boldsymbol{\mathcal{D}}(\boldsymbol{\Pi})$ has been defined in (6.8) in chapter 6. In case of channel over-modeling, the kernel of matrix $\mathbf{Q}$ is a vector space with dimension equal to $\delta = L' - L$ (L' being the overestimated order), which is spanned by the channel vector as well as all its $\delta - 1$ delayed copies [74]. In other words, the kernel is spanned by the following $(\delta + L + 1)N \times (\delta + 1)$ (N being the number of receiving antennas) block-Toeplitz matrix $\mathbf{H}$:

$$\mathbf{H} = \begin{bmatrix} \mathbf{h}_0 & \cdots & \mathbf{0} \\ \vdots & \ddots & \vdots \\ \mathbf{h}_L & \ddots & \mathbf{h}_0 \\ \vdots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{h}_L \end{bmatrix}$$

One can note that the columns of matrix $\mathbf{H}$ represent almost surely the sparsest vectors of the kernel of matrix $\mathbf{Q}$. In fact, any linear combination of vectors of $\mathbf{H}$ will yield almost surely vectors that are less sparse as they contain less zeros. Hence, in the noiseless case, the channel vector can be selected as the one that solves the following combinatorial optimization problem.

$$(P_0) \qquad \min_{\mathbf{x}, x_1=1, \mathbf{Q}\mathbf{x}=\mathbf{0}} \|\mathbf{x}\|_0 \tag{8.1}$$

where x_1 denotes the first entry of $\mathbf{x}$, $\|.\|_0$ is the ℓ_0 quasi-norm that returns the number of coefficients where the vector is not equal to zero. However, solving (P_0) requires generally an intractable combinatorial search, thus reducing its interest for real-time applications.

An alternative is to consider the optimization problem:

$$(P_p) \qquad \min_{\mathbf{x}, \mathbf{x}_1=1, \mathbf{Q}\mathbf{x}=\mathbf{0}} \|\mathbf{x}\|_p \tag{8.2}$$

where $\|\mathbf{x}\|_p$ denotes the ℓ_p quasi-norm: $\|\mathbf{x}\|_p = (\sum_i |\mathbf{x}_i|^p)^{\frac{1}{p}}$. It should be mentioned that this approach has been extensively studied by the compressed sensing theory [19] and applied to many fields like image processing [12] and communication systems [18, 44]. For all these applications, the problem is usually put under the form:

$$\min_{\mathbf{x}, \boldsymbol{\Phi}\mathbf{x}=\mathbf{y}} \|\mathbf{x}\|_p$$

where $\mathbf{\Phi}$ is a matrix independently distributed from vector $\mathbf{x}$. This is different from our case, since matrix $\mathbf{Q}$ is a function of vector $\mathbf{h}$. Therefore, all the theoretical results that have been derived in compressed sensing theory should be adapted to our context, and cannot be applied directly. This will be dealt with in next section.

8.2 Conditions for channel identifiability

Taking into account the structure of our problem, we can deduce that (P_p) is equivalent to:

$$\min_{\mathbf{s},\mathbf{s_1}=\mathbf{1}} \|\mathbf{Hs}\|_p^p \Leftrightarrow \min_{\mathbf{s}} \| \begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_L \\ \vdots \\ \mathbf{0} \end{bmatrix} + \widetilde{\mathbf{H}}\mathbf{s}\|_p^p$$

where $\widetilde{\mathbf{H}}$ is a $(\delta+L)N \times \delta$ block Toeplitz matrix having the same shape as $\mathbf{H}$. Before proceeding, we shall partition $\widetilde{\mathbf{H}} = \left[\begin{array}{c} \mathbf{A} \\ \hline \mathbf{B} \end{array}\right]$, where $\mathbf{A}$ (resp. $\mathbf{B}$) represents the first NL (resp. the last $N(\delta+1)$) rows of $\widetilde{\mathbf{H}}$.

$$\mathbf{B} = \begin{bmatrix} \mathbf{h}_L & \mathbf{h}_{L-1} & \cdots & \mathbf{h}_{L-\delta+1} \\ \mathbf{0} & \mathbf{h}_L & & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ \mathbf{0} & \cdots & \mathbf{0} & \mathbf{h}_L \end{bmatrix}$$

$$\mathbf{A} = \begin{bmatrix} \mathbf{h}_0 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{h}_1 & \ddots & & \\ \vdots & \ddots & & \mathbf{h}_0 \\ \vdots & \vdots & & \vdots \\ \mathbf{h}_{L-1} & \mathbf{h}_{L-2} & \cdots & \mathbf{h}_{L-\delta} \end{bmatrix}.$$

These matrices will play a key role in formulating the channel identifiability conditions.

8.2.1 ℓ_1 norm

Unlike the ℓ_p quasi-norm,$(p < 1)$, the ℓ_1 norm is convex. So in this case, it is possible to derive a necessary and sufficient condition for channel identifiability, which can be stated by the following theorem. For simplicity, we consider here the real case, i.e $\mathbf{h} \in \mathbb{R}^{N(L+1)}$.

Theorem 8.1. *Necessary and sufficient condition*

Let $\mathbf{v} = \left[\text{sign}(\mathbf{h}_1)^{\mathrm{T}}, \cdots, \text{sign}(\mathbf{h}_L)^{\mathrm{T}}\right]^{\mathrm{T}}$ *and assume that* $L > \delta \geq 1$*, then the necessary and sufficient condition for channel identifiability can be expressed as:*

$$\frac{|\mathbf{v}^{\mathrm{T}}\mathbf{As}|}{\|\mathbf{Bs}\|_1} \leq 1 \quad \forall \mathbf{s} \in \mathbb{R}^{\delta}.$$

Proof. The proof will rely on the following mathematical results about the optimization of convex functions.

Definition 8.1. *Let $f : \mathbb{R}^n \to \mathbb{R}$ be a real-valued function. The directional derivative of f at $\mathbf{x}_0$ is given by:*

$$f'(x_0, y) = \inf_{t>0} \frac{f(\mathbf{x}_0 + ty) - f(\mathbf{x}_0)}{t}$$

Theorem 8.2. *Let f be a convex function defined on a convex set X and $\mathbf{x}_0 \in X$ be a point where f is finite. Then $\mathbf{x}_0$ is a global minimum point of f if and only if the following conditions holds:*

$$f'(\mathbf{x}_0, \mathbf{x} - \mathbf{x}_0) \geq 0 \ \forall \mathbf{x} \in X$$

Let $f(\mathbf{s}) = \| \begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_L \\ \mathbf{0} \\ \mathbf{0} \end{bmatrix} + \widetilde{\mathbf{H}}\mathbf{s}\|$. Then $f(\mathbf{s})$ can be also given by:

$$f(s) = \|\widetilde{\mathbf{h}} + \mathbf{A}\mathbf{s}\|_1 + \|\mathbf{B}\mathbf{s}\|_1$$

where $\widetilde{\mathbf{h}} = \begin{bmatrix} \mathbf{h}_1 \\ \vdots \\ \mathbf{h}_L \end{bmatrix}$.

Using definition 8.1, the directional derivative of f at $\mathbf{s} = \mathbf{0}$ is given by:

$$\begin{aligned} f'(\mathbf{0}, \mathbf{y}) &= \inf_{t>0} \frac{\|\widetilde{\mathbf{h}} + t\mathbf{A}\mathbf{y}\|_1 + t\|\mathbf{B}\mathbf{y}\|_1 - \|\widetilde{\mathbf{h}}\|_1}{t} \\ &= \mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{y} + \|\mathbf{B}\mathbf{y}\|_1 \end{aligned}$$

Using theorem 8.2, we conclude that $\mathbf{0}$ is a global minimum of f if and only if, for all $\mathbf{y} \in \mathbb{R}^{\delta}$, we have:

$$f'(\mathbf{0}, \mathbf{y}) \geq 0.$$

It can be easily verified that this condition holds if and only if:

$$\frac{|\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{y}|}{\|\mathbf{B}\mathbf{y}\|_1} \leq 1,$$

thus concluding the proof of theorem 8.1. □

8.2.2 ℓ_p quasi-norm

Since the ℓ_p quasi-norm is a non-convex function, the problem might have many local minima. Nevertheless, we still can find a sufficient condition that ensures that the channel can be identified as a local minimum of (8.2). This result is stated in the following theorem:

Theorem 8.3. *Sufficient condition*
Let $\mathbf{v} = p \begin{bmatrix} \text{sign}(\mathbf{h}_1) \\ \vdots \\ \text{sign}(\mathbf{h}_L) \end{bmatrix} \bullet \begin{bmatrix} |\mathbf{h}_1|^{p-1} \\ \vdots \\ |\mathbf{h}_L|^{p-1} \end{bmatrix}$ *where* $\bullet$ *denotes the Hadamard (element by element) product. If the following condition is satisfied:*

$$\frac{|\mathbf{v}^{\mathrm{T}}\mathbf{A}\mathbf{s}|}{\|\mathbf{B}\mathbf{s}\|_1} \leq 1 \ \ \forall \mathbf{s} \in \mathbb{R}^{\delta} \tag{8.3}$$

then the channel can be identified as a local minimum of (8.2).

Proof. To prove that (8.3) implies channel identifiability, it suffices to prove that there exists a neighbourhood V_ϵ for $\epsilon > 0$ such that

$$\|\widetilde{\mathbf{h}} + \mathbf{A}\mathbf{x}\|_p^p + \|\mathbf{B}\mathbf{x}\|_p^p \geq \|\widetilde{\mathbf{h}}\|_p^p$$

for all $\mathbf{x} \in V_\epsilon$. Denote by $\widetilde{h}_i$ the i-th entry of $\widetilde{\mathbf{h}}$ and by $\mathbf{e}_i$ the vector with 1 on the i-th component and 0 elsewhere. Let $s_i = \text{sign}(\widetilde{h}_i)$. Then, it can be easy to see that:

$$\begin{aligned} |\widetilde{h}_i + \mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}|^p &= ||\widetilde{h}_i| + s_i\mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}|^p \\ &= |\widetilde{h}_i|^p \times |1 + \frac{s_i\mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}}{|\widetilde{h}_i|}|^p \end{aligned}$$

As a consequence, the Taylor approximation of $|\widetilde{h}_i + \mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}|^p$ when $\mathbf{x}$ lies in a neighborhood of $\mathbf{0}$ can be given by:

$$|\widetilde{h}_i + \mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}|^p = |\widetilde{h}_i|^p \left(1 + p\frac{s_i\mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}}{|\widetilde{h}_i|} + p(p-1)\frac{|\mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}|^2}{|\widetilde{h}_i|^2}\right) + \mathrm{o}(\|\mathbf{x}\|^3)$$

Hence,

$$\frac{|\widetilde{h}_i + \mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}|^p - |\widetilde{h}_i|^p - ps_i|\widetilde{h}_i|^{p-1}\mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}}{ps_i|\widetilde{h}_i|^{p-1}\mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}} \xrightarrow[\mathbf{x}\to\mathbf{0}]{} 0$$

As a consequence, for all $\epsilon > 0$ there exists a neighborhood $V_{\epsilon,i}$ of $\mathbf{0}$ such that;

$$\left|\frac{|\widetilde{h}_i + \mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}|^p - |\widetilde{h}_i|^p - ps_i|\widetilde{h}_i|^{p-1}\mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}}{ps_i|\widetilde{h}_i|^{p-1}\mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}}\right| \leq \epsilon \ \forall \mathbf{x} \in V_{\epsilon,i}$$

Therefore,

$$||\widetilde{h}_i| + s_i\mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}|^p \geq |\widetilde{h}_i|^p + p|\widetilde{h}_i|^{p-1}s_i\mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x} - \epsilon p|\widetilde{h}_i|^{p-1}|s_i\mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}|$$

Let $\widetilde{\mathbf{V}}_\epsilon = \cap_{i=\{1,\cdots,LN\}}$, then for all $\mathbf{x} \in \widetilde{\mathbf{V}}_\epsilon$, we have:

$$\sum_{i=1}^{LN} ||\widetilde{h}_i| + s_i\mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}|^p \geq \sum_{i=1}^{LN} |\widetilde{h}_i|^p + p|\widetilde{h}_i|^{p-1}s_i\mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x} - \epsilon p|\widetilde{h}_i|^{p-1}|s_i\mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x}| \tag{8.4}$$

Without loss of generality, we can assume that $\sum_{i=1}^{LN} p|\widetilde{h}_i|^{p-1}s_i\mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x} \leq 0$ since, if $\sum_{i=1}^{LN} p|\widetilde{h}_i|^{p-1}s_i\mathbf{e}_i^{\mathrm{T}}\mathbf{A}\mathbf{x} \geq 0$ it can be easily shown that for ϵ small enough, we have:

$$\|\widetilde{\mathbf{h}} + \mathbf{A}\mathbf{x}\|^p + \|\mathbf{B}\mathbf{x}\|_p^p \geq \|\widetilde{\mathbf{h}}\|_p^p.$$

Let $\epsilon = \epsilon^{'} \frac{|\sum_{i=1}^{LN} p|\widetilde{h}_i|^{p-1} s_i \mathbf{e}_i^{\mathrm{T}} \mathbf{A}\mathbf{x}|}{\sum_{i=1}^{LN} p|\widetilde{h}_i|^{p-1}|\mathbf{e}_i^{\mathrm{T}} \mathbf{A}\mathbf{x}|}$, where $\epsilon^{'} > 0$. Then (8.4) becomes:

$$\sum_{i=1}^{LN} ||\widetilde{h}_i| + s_i \mathbf{e}_i^{\mathrm{T}} \mathbf{A}\mathbf{x}|^p \geq \sum_{i=1}^{LN} |\widetilde{h}_i|^p + p|\widetilde{h}_i|^{p-1} s_i \mathbf{e}_i^{\mathrm{T}} \mathbf{A}\mathbf{x} - \epsilon^{'} |p \sum_{i=1}^{LN} |\widetilde{h}_i|^{p-1} s_i \mathbf{e}_i^{\mathrm{T}} \mathbf{A}\mathbf{x}|$$

Since $\sum_{i=1}^{LN} p|\widetilde{h}_i|^{p-1} s_i \mathbf{e}_i^{\mathrm{T}} \mathbf{A}\mathbf{x} < 0$,

$$\sum_{i=1}^{LN} p|\widetilde{h}_i|^{p-1} s_i \mathbf{e}_i^{\mathrm{T}} \mathbf{A}\mathbf{x} = -|\sum_{i=1}^{LN} p|\widetilde{h}_i|^{p-1} s_i \mathbf{e}_i^{\mathrm{T}} \mathbf{A}\mathbf{x}|,$$

thus yielding:

$$\sum_{i=1}^{LN} ||\widetilde{h}_i| + s_i \mathbf{e}_i^{\mathrm{T}} \mathbf{A}\mathbf{x}|^p \geq \sum_{i=1}^{LN} |\widetilde{h}_i|^p - (1 + \epsilon^{'})|\sum_{i=1}^{LN} p|\widetilde{h}_i|^{p-1} s_i \mathbf{e}_i^{\mathrm{T}} \mathbf{A}\mathbf{x}| \tag{8.5}$$

On the other hand, as $\mathbf{x}$ tends to $\mathbf{0}$, we have $\frac{\|\mathbf{B}\mathbf{x}\|_p^p}{\|\mathbf{B}\mathbf{x}\|_1}$ tends to infinity. This implies that there exists a neighborhood of $\mathbf{0}$, $V_{\epsilon'}$ such that:

$$\|\mathbf{B}\mathbf{x}\|_p^p \geq (1 + \epsilon^{'})\|\mathbf{B}\mathbf{x}\|_1 \forall \mathbf{x} \in V_{\epsilon'}$$

Since by (8.3), $\|\mathbf{B}\mathbf{x}\|_1$ satisfies:

$$\|\mathbf{B}\mathbf{x}\|_1 \geq |\sum_{i=1}^{LN} p|\widetilde{h}_i|^{p-1} s_i \mathbf{e}_i^{\mathrm{T}} \mathbf{A}\mathbf{x}|$$

we get:

$$\|\mathbf{B}\mathbf{x}\|_p^p \geq (1 + \epsilon^{'})\|\mathbf{B}\mathbf{x}\|_1 \geq (1 + \epsilon^{'})|\sum_{i=1}^{LN} p|\widetilde{h}_i|^{p-1} s_i \mathbf{e}_i^{\mathrm{T}} \mathbf{A}\mathbf{x}| \tag{8.6}$$

Plugging (8.6) into (8.5), we have:

$$\sum_{i=1}^{LN} ||\widetilde{h}_i| + s_i \mathbf{e}_i^{\mathrm{T}} \mathbf{A}\mathbf{x}|^p + \|\mathbf{B}\mathbf{x}\|_p^p \geq \sum_{i=1}^{LN} |\widetilde{h}_i|^p \ \forall \mathbf{x} \in \widetilde{V}_\epsilon \cap V_{\epsilon'}$$

which concludes the proof of theorem 8.3. □

8.3 Probabilistic Analysis

In this section we will study the effect of the system parameters on the channel identifiability probability. We assume that the channel coefficients are drawn from the Gaussian distribution with mean 0 and variance $\frac{1}{L+1}$. To determine a lower bound on the channel identifiability probability, we will rely on the techniques derived in [51, 50]. Actually, in the same way as [51], we recast the probability conditions in an other form as stated by the following theorem:

Theorem 8.4. *Let* $\mathbf{d}^*$ *be the value that minimizes :*

$$\begin{aligned} &\min \ \|\mathbf{d}\|_\infty \\ &\text{subject to } \mathbf{B}^{\mathrm{T}}\mathbf{d} = \mathbf{A}^{\mathrm{T}}\mathbf{v} \end{aligned}$$

Then, the channel can be identified if and only if $\|\mathbf{d}\|_\infty \leq 1$.

The new formulation given by theorem 8.4 is interesting in the sense that it allows geometric interpretation of the channel identifiability condition. Indeed, it follows from theorem 8.4 that the channel identifiability holds for a given channel realization if and only if there is a vector $\mathbf{d}$ on the cube $Q = [-1, 1]^{\delta N}$ such that $\mathbf{B}^{\mathrm{T}}\mathbf{d} = \mathbf{A}^{\mathrm{T}}\mathbf{v}$, i.e, $\mathbf{A}^{\mathrm{T}}\mathbf{v}$ belongs to the image of the cube generated by $\mathbf{B}^{\mathrm{T}}$. Since $\mathrm{rank}(\mathbf{B}) = \delta$ almost surely, the channel identifiability will hold if the following conditions are satisfied:

- The image of the cube by $\mathbf{B}^{\mathrm{T}}$ contains a ball of radius α
- The vector $\mathbf{A}^{\mathrm{T}}\mathbf{v}$ satisfies $\|\mathbf{A}^{\mathrm{T}}\mathbf{v}\|_2 \leq \alpha$.

Let $\mathcal{P}$ denote the probability that the channel identifiability holds and E_α^1 and E_α^2 be the events given by:

$$E_\alpha^1 = \{\text{The image of the cube by } \mathbf{B}^{\mathrm{T}} \text{contains a ball of radius } \alpha\},$$
$$E_\alpha^2 = \{\|\mathbf{A}^{\mathrm{T}}\mathbf{v}\|_2 \leq \alpha\}.$$

Then, $\mathcal{P}$ can be lower bounded as:

$$\mathcal{P} \geq \mathbb{P}\left\{\bigcup_\alpha E_\alpha^1 \cap E_\alpha^2\right\} \geq \max_\alpha \mathbb{P}\left(E_\alpha^1 \cap E_\alpha^2\right).$$

8.3.1 ℓ_1 norm

In the following, we propose to determine a lower bound on the probability of the events E_α^1 and E_α^2, while considering the ℓ_1 norm minimization. We will consider first the relatively easy case $\delta = 1$ and after that the more general case $\delta \leq \min(N, L-1)$. For $\delta \geq N$, we have not been able to derive a lower bound on the probability of channel identifiability, but we conjecture that the effect of the system parameters remains the same.

Before going any further, let us, first, write the event E_α^1 in an other equivalent way [51]:

$$E_\alpha^1 = \left\{\min_{\mathbf{x}} \frac{\|\mathbf{B}\mathbf{x}\|_1}{\|\mathbf{x}\|_2} \geq \alpha\right\}.$$

8.3.1.1 Case when $\delta = 1$

When $\delta = 1$, it is easy to see that $\mathbf{v}^{\mathrm{T}}\mathbf{A}$ is a real standard Gaussian random variable with mean 0 and variance $\frac{LN}{L+1}$. Hence we have;

$$\mathbb{P}\left(E_\alpha^2\right) = \mathbb{P}\left(|\mathbf{v}^{\mathrm{T}}\mathbf{A}| \leq \alpha\right) \tag{8.7}$$
$$= \mathbb{P}\left(|\mathbf{v}^{\mathrm{T}}\mathbf{A}|^2 \leq \alpha^2\right) = \frac{1}{\sqrt{\pi}}\gamma\left(\frac{1}{2}, \frac{\alpha^2(L+1)}{2LN}\right) \tag{8.8}$$

where $\gamma(a, x)$ is the lower incomplete gamma function given by:

$$\gamma(a, x) = \int_0^x \exp(-t)t^{a-1}dt.$$

On the other hand, using standard concentration inequalities for normal variables [67], we show that, for every $\epsilon \in [0, 1]$, we have :

$$\mathbb{P}\left(\|\mathbf{h}_L\|_1 \geq N\sqrt{\frac{2}{\pi(L+1)}}(1-\epsilon)\right) \geq 1 - \exp\left(-\frac{N\epsilon^2}{\pi}\right). \tag{8.9}$$

Since for $\delta = 1$, the events E_α^1 and E_α^2 are independent, we get after combining (8.9) and (8.8), and setting $\alpha = N\sqrt{\frac{2}{\pi(L+1)}}(1-\epsilon)$, the following theorem:

Theorem 8.5. *For $\delta = 1$, the probability $\mathcal{P}$ that channel identifiability occurs is greater than:*

$$\mathcal{P} \geq \max_{\epsilon\in[0,1]} \left(1 - \exp\left(-\frac{N\epsilon^2}{\pi}\right)\right) \frac{1}{\sqrt{\pi}}\gamma\left(\frac{1}{2}, \frac{2N}{L}(1-\epsilon)^2\right) \tag{8.10}$$

Remark 8.1. *Under the assumption that the random variables $\|\mathbf{v}^{\mathrm{T}}\mathbf{A}\|_2$ and $\|\mathbf{h}_L\|_1 = \min_{\mathbf{x}} \frac{\|\mathbf{B}\mathbf{x}\|_1}{\|\mathbf{x}\|_2}$ are concentrated around their expected values with high probability (this assumption is valid in general for standard random distributions), one can understand intuitively the effect of the system parameters N and L on the probability for channel identifiability. Actually, given that $\mathbb{E}\|\mathbf{h}_L\|_1 = N\sqrt{\frac{2}{\pi(L+1)}}$, we deduce that we can find, a ball of radius r_1 of order $\mathcal{O}\left(\frac{N}{\sqrt{L}}\right)$ that is contained in the image of the cube Q by $\mathbf{B}^{\mathrm{T}}$ with a high probability. In the same way, given that the expected value of $|\mathbf{v}^{\mathrm{T}}\mathbf{A}|$ is of the order $\mathcal{O}\left(\sqrt{N}\right)$, we can find a ball of radius $r_2 = \mathcal{O}\left(\sqrt{N}\right)$ that contains the vector $\mathbf{v}^{\mathrm{T}}\mathbf{A}$, with a high probability. Since channel identifiability occurs when $r_1 \geq r_2$, we deduce that as N increases, and L decreases, channel identifiability should be more likely to occur.*

8.3.1.2 Case when $\delta > 1$ and $\delta \leq \min(L-1, N)$

When $\delta > 1$, the problem becomes more difficult, since $\mathbf{v}^{\mathrm{T}}\mathbf{A}$ is no longer Gaussian and $\min_{\mathbf{x}} \frac{\|\mathbf{B}\mathbf{x}\|_1}{\|\mathbf{x}\|_2}$ has no closed-form expression. Besides E_α^1 and E_α^2 are no longer independent, thus making our computations less tighter. But, as we can see later, even if the lower bound probability is too loose, one can still draw conclusions about the effect of the system parameters on the channel identifiability probability.

Let us now deal with the probability of the event E_α^1.

$$\mathbb{P}\left(E_\alpha^1\right) = \left\{\min_{\mathbf{x}} \frac{\|\mathbf{B}\mathbf{x}\|_1}{\|\mathbf{x}\|_2} \geq \alpha\right\}.$$

Since $\delta \leq \min(L-1, N)$, it can be shown that:

$$\min_{\mathbf{x}} \frac{\|\mathbf{B}\mathbf{x}\|_1}{\|\mathbf{x}\|_2} \geq \min_{\mathbf{x}} \frac{\|\widetilde{\mathbf{B}}\mathbf{x}\|_1}{\|\mathbf{x}\|_2} \tag{8.11}$$

where $\widetilde{\mathbf{B}} = [\mathbf{h}_L, \cdots, \mathbf{h}_{L-\delta+1}]$. Consequently,

$$\mathbb{P}\left(E_\alpha^1\right) \geq \mathbb{P}\left\{\min_{\mathbf{x}} \frac{\|\widetilde{\mathbf{B}}\mathbf{x}\|_1}{\|\mathbf{x}\|_2} \geq \alpha\right\}. \tag{8.12}$$

To determine a lower bound on the probability $\mathbb{P}\left(E_\alpha^1\right)$, we will use the following result:

Theorem 8.6. *[21] Let Φ be a $N \times \delta$ Gaussian matrix, with iid entries, i.e, $\phi_{i,j} \sim \mathcal{N}(0, \sigma^2)$. Let $1 > \kappa > 0$ and choose $\eta, \epsilon > 0$ such that $\kappa = \frac{\eta+\epsilon}{1-\epsilon}$. Then*

$$\|\Phi x\|_1 \geq N\sigma\sqrt{\frac{2}{\pi}}(1-\kappa)\|\mathbf{x}\|_2$$

holds uniformly for $\mathbf{x} \in \mathbb{R}^{\delta}$ *with probability exceeding*

$$1 - (1 + 2/\epsilon)^{\delta} \exp\left(-\frac{\eta^2 N}{2c^2}\right)$$

where $c = (31/40)^{\frac{1}{4}} (1.13 + \sqrt{\pi})$.

Applying theorem 8.6, we get that for every $1 \geq \kappa > 0$, and $\alpha^* = N\sqrt{\frac{2}{\pi(L+1)}}(1-\kappa)$ we have :

$$\mathbb{P}\left(E^1_{\alpha^*}\right) \geq 1 - (1 + 2/\epsilon)^{\delta} \exp\left(-\frac{\eta^2 N}{2c^2}\right) \tag{8.13}$$

where ϵ and η are positive reals satisfying $\kappa = \frac{\eta+\epsilon}{1-\epsilon}$.

Remark 8.2. *Note that in contrast to* ϵ*, increasing* η *improves the lower bound probability. Consequently, the values of* η *and* ϵ *can be set so as to maximize the lower bound probability.*

According to Markov inequality, $\mathbb{P}\left(E^2_{\alpha}\right)$ can be written as:

$$\begin{aligned}
\mathbb{P}\left(E^2_{\alpha^*}\right) &= \mathbb{P}\left\{\|\mathbf{A}^{\mathrm{T}}\mathbf{v}\|_2 \leq \alpha^*\right\} \\
&\geq 1 - \frac{\mathbb{E}\left(\|\mathbf{A}^{\mathrm{T}}\mathbf{v}\|_2^2\right)}{\left(\alpha^*\right)^2} \\
&\geq 1 - \frac{N\delta(2L - \delta + 1)}{2\left(\alpha^*\right)^2 (L+1)} \\
&\geq 1 - \frac{\pi\delta\left(2L - \delta + 1\right)}{4N\left(1-\kappa\right)^2}.
\end{aligned} \tag{8.14}$$

Using (8.9) and (8.14), the lower bound on the channel identifiability can be lower bounded by:

$$\mathcal{P} \geq \max_{\substack{\kappa,\epsilon,\eta \\ \kappa = \frac{\eta-\epsilon}{1+\epsilon}}} 1 - (1 + 2/\epsilon)^{\delta} \exp\left(-\frac{\eta^2 N}{2c^2}\right) - \frac{\pi\delta\left(2L - \delta + 1\right)}{4N\left(1-\kappa\right)^2}. \tag{8.15}$$

Although the provided bound is not tight, it provides information about the impact of the system parameters on the channel identifiability probability. One can obviously see that increasing the number of antennas N improves the channel identifiability probability, in contrast to the system parameters L and δ which tend to decrease it.

8.3.2 ℓ_p quasi-norm

In this section, we will consider, only the case when $\delta = 1$. Let $\mathbf{h}_i = [h_{i,1}, \cdots, h_{i,N}]$ and $s_{i,j} = \mathrm{sign}(h_{i,j})\mathrm{sign}(h_{i-1,j})$. Then, the probability that E^2_{α} is satified can be expressed as:

$$\mathbb{P}\left(E^2_{\alpha}\right) = \mathbb{P}\left\{\left| p \sum_{k=1}^{L} \sum_{j=1}^{N} |h_{i,j}|^{p-1} |h_{i-1,j}| \, s_{i,j} \right| \leq \alpha \right\}.$$

One can note that as p tends to zero, $p \sum_{k=1}^{L} \sum_{j=1}^{N} |h_{i,j}|^{p-1} |h_{i-1,j}| \, s_{i,j}$ converges almost surely to zero, thus implying that $\mathbb{P}\left(E^2_{\alpha}\right)$ tends to one as p tends to zero.

Let us deal with the complementary event of E_α^2. According to Markov inequality, $\mathbb{P}\left({}^cE_\alpha^2\right)$ can be upper bounded by:

$$\mathbb{P}\left({}^cE_\alpha^2\right) \leq \frac{\mathbb{E}\left|p\sum_{i=1}^{L}\sum_{j=1}^{N}|h_{i,j}|^{p-1}|h_{i-1,j}|s_{i,j}\right|^{\frac{1}{2}}}{\alpha^{\frac{1}{2}}}.$$

To remove the expectation on the Rademacher sequence (i.i.d random variables that get +1 or -1 with equal probabilities) $s_{i,j}$, we will use the Kintchine's inequality that can be stated as follows:

Lemma 8.1. *[37] Let* $\mathbf{x}$ *be a vector in* $\mathbb{R}^n$*, and* s_j *a Rademacher sequence. Then, we have:*

$$\mathbb{E}\sqrt{\left|\sum_{j=1}^{n}s_j\mathbf{x}_j\right|} \leq C\left[\sum_{j=1}^{n}x_j^2\right]^{\frac{1}{4}}$$

where $C = \frac{\Gamma(\frac{3}{4})}{2^{\frac{3}{4}}\Gamma(\frac{3}{2})}$.

Applying the Kintchine's inequality, we get:

$$\mathbb{E}\left|\sum_{i=1}^{L}\sum_{j=1}^{N}|h_{i,j}|^{p-1}|h_{i-1,j}|s_{i,j}\right|^{\frac{1}{2}} \leq C\sum_{i=1}^{L}\mathbb{E}\left[\sum_{j=1}^{N}|h_{i,j}|^{2(p-1)}|h_{i-1,j}|^2\right]^{\frac{1}{4}} \tag{8.16}$$

Since $p<1$, $x\mapsto|x|^p$ is a concave function when $x>0$, we can therefore prove, using Jensen inequality, that:

$$\sum_{j=1}^{N}|h_{i,j}|^{2(p-1)}|h_{i-1,j}|^2 \leq \left(\sum_{j=1}^{N}|h_{i-1,j}|^2\right)^p\left(\sum_{j=1}^{N}\left|\frac{h_{i-1,j}}{h_{i,j}}\right|^2\right)^{1-p} \tag{8.17}$$

$$\overset{(a)}{\leq} p\sum_{j=1}^{N}|h_{i-1,j}|^2 + (1-p)\sum_{j=1}^{N}\frac{|h_{i-1,j}|^2}{|h_{i,j}|^2} \tag{8.18}$$

where the second inequality referred to as (a) follows from the fact that $a^pb^q \leq pa+qb$ whenever a and b are positive and $p+q=1$. Combining (8.16) and (8.18), we get :

$$\begin{aligned}\mathbb{E}\left[p\sum_{i=1}^{L}\left|\sum_{j=1}^{N}|h_{i,j}|^{p-1}|h_{i-1,j}|s_{i,j}\right|\right]^{\frac{1}{2}} &\leq C\sqrt{p}\sum_{i=1}^{L}\mathbb{E}\left(p\sum_{j=1}^{N}|h_{i-1,j}|^2\right)^{\frac{1}{4}} + \sqrt{p}\,(1-p)^{\frac{1}{4}}\sum_{j=1}^{N}\mathbb{E}\frac{\sqrt{|h_{i-1,j}|}}{\sqrt{|h_{i,j}|}}\\ &\leq CL\sqrt{p}\left(\frac{2p}{L+1}\right)^{\frac{1}{4}}\frac{\Gamma(N/2+\frac{1}{4})}{\Gamma(N/2)} + \frac{CLN\sqrt{p}(1-p)^{\frac{1}{4}}}{\pi}\Gamma(3/4)\Gamma(\frac{1}{4}).\end{aligned}$$

Consequently, for $\alpha = N\sqrt{\frac{2}{\pi(L+1)}}(1-\epsilon)$

$$\mathbb{P}\left(E_\alpha^2\right) \geq 1 - \frac{A_1p^{\frac{3}{4}}}{\sqrt{1-\epsilon}} - \frac{A_2\sqrt{p}\,(1-p)^{\frac{1}{4}}}{\sqrt{1-\epsilon}} \tag{8.19}$$

where

$$A_1 = \frac{2^{\frac{1}{4}} L\Gamma(N/2+1/4)\Gamma(3/4)}{\Gamma(N/2)\sqrt{N}\pi^{\frac{1}{4}}}$$
$$A_2 = \frac{L(L+1)^{\frac{1}{4}}\sqrt{N}\Gamma(1/4)(\Gamma(3/4))^2}{\pi^{\frac{5}{4}}}.$$

We note that unlike A_1, A_2 tends to increase with N. Hence, the lower bound probability does not always decrease with N. Combining (8.13) and (8.19), it can be proved that the channel identifiability is lower bounded by:

$$\mathcal{P} \geq \max_{\epsilon} \left(1 - \exp(-\frac{N\epsilon^2}{\pi})\right)\left(1 - \frac{A_1 p^{\frac{3}{4}}}{\sqrt{1-\epsilon}} - \frac{A_2\sqrt{p}\,(1-p)^{1/4}}{\sqrt{1-\epsilon}}\right). \tag{8.20}$$

8.4 Simulation results

8.4.1 ℓ_1 norm

We present here simulation results for the ℓ_1 norm. Fig. 8.1 displays the effect of the system parameters L and N on the lower bound probability that we have computed by maximizing (8.10) numerically. We note that, as expected, increasing the number of antennas, tends to

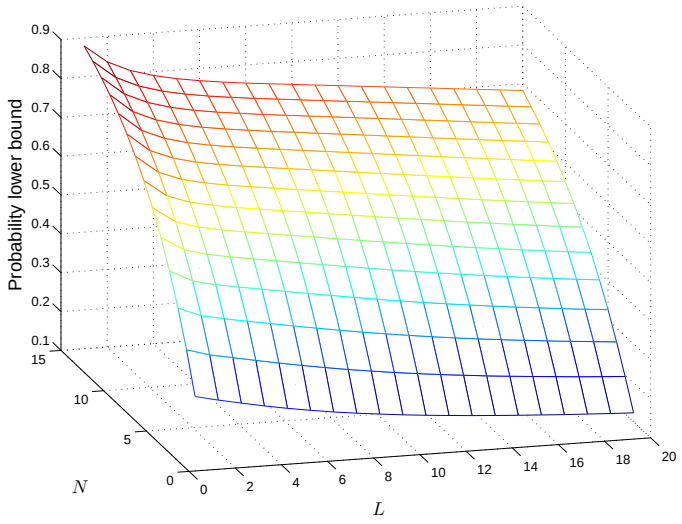


Figure 8.1: Impact of the system parameters L and N on the lower bound probability.

enhance the channel identifiability probability.

8.4.2 ℓ_p quasi-norm

For the ℓ_p quasi-norm, we study the effect of the parameter p on the lower bound probability. We set the system parameters M and L to 6 and 3, and we vary p from 10^{-3} to 10^{-6}. Fig. 8.2 displays the lower bound with respect to p. We note that as p tends to zero, the lower bound probability increases.

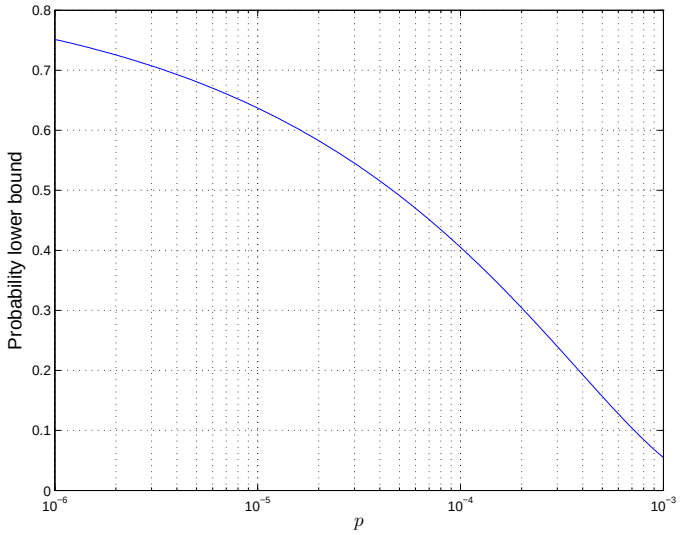


Figure 8.2: Lower bound probability with respect to p.

8.5 Conclusion

This chapter analyses the robustness of certain blind channel identification methods when using ℓ_p quasi-norms. Necessary and sufficient conditions of channel identifiability are provided. Lower bounds on the channel identifiability probability are derived thereby providing some useful insights on the impact of the number of sensors M, the channel length L and the quasi-norm parameter p on the identifiability conditions.

Chapter 9

Quasi-convexity of the asymptotic channel MSE in scalar regularized semi-blind estimation

As we have seen in chapter 6, semi blind estimation techniques use jointly information from data and training samples. An optimal semi-blind approach will be obtained if all the available information about the channel parameter is taken into account. In the literature, different methods using the Maximum likelihood criteria, have been proposed. Although being optimal, these methods involve in general high computational complexity thus restricting their interest to only theoretical issues.

Suboptimal approaches such as blind or training based techniques, retain only information provided by the data or the training samples, to build one single criterion. In contrast to blind estimation methods, training based techniques are more sensitive to noise and entail inefficient bandwidth utilization. However, blind methods are more complex, estimate the channel only up to a scalar ambiguity and are often non-robust to modelization errors (e.g channel order overestimation errors). For these reasons, it might be interesting to combine linearly both criteria so as to resolve the drawbacks inherent to blind and training based techniques. Such approaches are referred to as regularized semi-blind techniques in reference to the regularizing constant which parametrizes the channel estimate. Despite exhibiting low computational complexity as compared to the EM algorithm used for SML technique (See section 6.3 in chapter 6), semi blind approaches are strongly influenced by the tuning of the regularizing constant [15]. To deal with this issue, the work in [15] proposes to set the regularizing constant to the value that minimizes the asymptotic mean square error. The statement of the uniqueness of such minimum has been mentioned in [16, 15] and supported by simulations. To the best of our knowledge, this statement has not been proved in any previous work.

In this chapter, we demonstrate that the asymptotic MSE function is quasi-convex which provides a guarantee that numerical optimization always leads to the unique desired optimal weighing parameter value. Moreover, one can observe that regularized approaches has been used in many other fields, [59, 1], and thus, we believe that the result given in this chapter might be extended and adapted to other problems, where an optimal weighting coefficient is needed.

9.1 Expression of the Asymptotic MSE

In this section, we review the expression for the asymptotic MSE in regularized semi blind methods. As mentioned earlier, semi blind methods employ in addition to training symbols, a blind criterion that is based on the statistical properties of the received signal in the data period. In most cases, this criterion can be put under the form:

$$\min_{\|\mathbf{h}\|=1} \mathbf{h}^{\mathrm{H}}\mathbf{Q}\mathbf{h}$$

where $\mathbf{h}$ is the channel parameter and $\mathbf{Q}$ is a matrix that depends solely on the statistical properties of the received signal.

Under some mild assumptions, it was shown in [15] that the asymptotic MSE is proportional to:

$$\mathrm{MSE} \propto \mathrm{Tr}\left\{ \left(\mathbf{I}+\lambda\mathbf{Q}\right)^{-1} \left(\mathbf{I}+\lambda^{\mathbf{2}}\gamma\boldsymbol{\mathcal{M}}(\mathbf{h})\right) \left(\mathbf{I}+\lambda\mathbf{Q}\right)^{-1} \right\}$$

where $\lambda > 0$ is the regularizing coefficient, $\gamma = \frac{n}{m}$ (n and m being respectively the length of the data and training sequences) and $\boldsymbol{\mathcal{M}}(\mathbf{h})$ is a hermitian matrix that has the same row and column space as $\mathbf{Q}$ (meaning that if $\mathbf{Q} = \mathbf{U}\mathbf{D}\mathbf{U}^{\mathrm{H}}$ is the eigenvalue decomposition of $\mathbf{Q}$, $\boldsymbol{\mathcal{M}}(\mathbf{h})$ writes as $\boldsymbol{\mathcal{M}}(\mathbf{h}) = \mathbf{U}\mathbf{A}\mathbf{U}^{\mathrm{H}}$, where $\mathbf{A}$ is a given hermitian matrix.). Using the eigenvalue decomposition of $\mathbf{Q}$ it can be easily verified that the MSE is proportional to:

$$\mathrm{MSE} \propto \sum_i \frac{1+\lambda^2\gamma a_{i,i}}{(1+\lambda d_{i,i})^2} \tag{9.1}$$

where $a_{i,i} > 0$ (resp. $d_{i,i} > 0$) denote the diagonal elements of $\mathbf{A}$ (resp. the non zero diagonal elements of $\mathbf{D}$).

To conclude about the quasi-convexity of the MSE, it amounts to demonstrating that of the sum of quadratic fractions in the form $\sum_{i=1}^{q} \frac{1+c_i x^2}{(1+d_i x)^2}$ defined on the positive real axis $\mathbb{R}^+$, c_i and d_i being strictly positive scalars. This issue will be dealt with in the sequel, but before, we shall provide some mathematical preliminaries.

9.2 Mathematical preliminaries

For the reader convenience, we devote this section to recalling the definition and some useful results about quasi-convex functions. We refer the reader to [14] for further information.

Definition 9.1. *A real valued function f is said to be quasi-convex if its domain of definition and all its sublevel sets:*

$$S_\alpha = \{x \in \mathbf{dom} f | f(x) \leq \alpha\}$$

for $\alpha \in \mathbb{R}$, are convex, where $\mathbf{dom} f$ denotes the set over which the function f is defined.

Examples of quasi-convex functions To illustrate this concept, we provide in fig 9.1 some examples of quasi-convex functions. One can see that a concave and also a non convex function can be also quasi-convex. Like convex functions, quasi-convex functions satisfy a modified Jensen inequality which is given by:

Theorem 9.1. *A function f is quasi-convex if and only if $\mathbf{dom} f$ is convex and for all x, $y \in \mathbf{dom}(f)$ and $0 \leq \theta \leq 1$*

$$f(\theta x + (1-\theta)y) \leq \max\left\{f(x), f(y)\right\}.$$

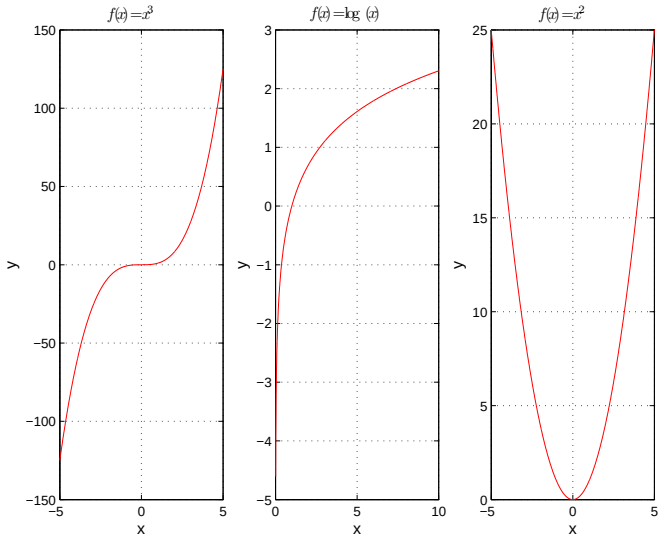


Figure 9.1: Examples of quasi-convex functions

Clearly, the quasi-convexity generalizes the notion of convexity in the sense that the class of quasi-convex functions is larger than and includes the class of convex functions. Also, in most cases, quasi-convex functions inherit the nice properties of convex functions including the absence of local minimum as stated in the following theorem.

Theorem 9.2. *Let f be a quasi-convex function. Then every local minimum is a global minimum or f is constant in a neighborhood of this local minimum.*

Consequently, if a quasi-convex function f is non constant over any given interval (which is the case for the sum of quadratic functions we consider), then each local minimum is also a global minimum. Moreover, this global minimum (whenever it exists) is unique for real valued functions. To prove the non existence of local minima besides the global one, we use often the following second-order condition:

Theorem 9.3. *Let f be a real function which is twice derivable. If f satisfies:*

$$\forall c \quad \text{such that} \quad f'(c) = 0, \qquad\qquad f''(c) > 0,$$

then, f is quasi-convex, and each local minimum is a global minimum.

9.3 Quasi-convexity of the asymptotic MSE

The quasi-convexity of the asymptotic MSE is a by-product of the following theorem:

Theorem 9.4. *Let c_i, d_i be two sequences of $q \in \mathbb{N}^*$ stricly positive reals. Then the derivative of*

$$F_q(x) = \sum_{i=1}^{q} \frac{1 + c_i x^2}{(1 + d_i x)^2}$$

has a unique positive zero x_0 with $F_q^{(2)}(x_0) > 0$. Consequently, $F_q(x)$ is a quasi-convex function when its domain of definition is restricted to $\mathbb{R}^+$ and hence has a unique local (global) minimum on the positive real axis.

In the sequel we will omit the index q for notational simplicity so that F_q will be referred to as F. To prove theorem 9.4, we proceed in the following steps:

- First, we show that the number of positive real values of $F^{(k+1)}$ is larger or equal than that of $F^{(k)}$, where $F^{(k)}$ denotes the k-th derivative of F.
- We introduce the function G_k which has the same number of zeros as $F^{(k)}$ and prove that it converges uniformly to G_∞, over a compact set that contains all the zeros of $F^{(k)}$.
- Then we prove that G_∞ has a unique positive zero in that compact set.
- By applying Hurwitz theorem [90], we conclude that for large values of k, G_k is zero only once and thus will be also the case of $F^{(k)}$. Finally, we prove that the second derivative of F is strictly positive when evaluated at the zero argument of F. Fig. 9.2 illustrates the shape of function F and its first and second order derivatives for $q = 3$.

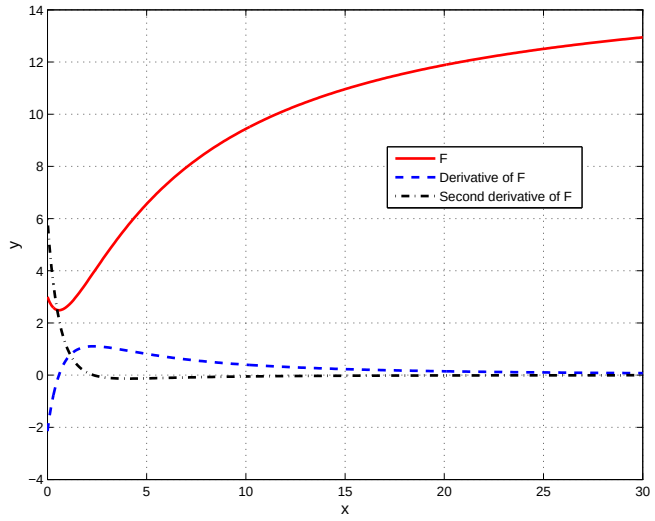


Figure 9.2: Function F and its first and second order derivatives

The details of all these steps and their proofs will be provided in the next section.

9.4 Analysis and properties of F

As we have seen previously, the proof of our main theorem can be carried out in 4 steps, whose proof are sequentially provided in the sequel:

9.4.1 Closed-form expressions for the derivatives of F

In this subsection, we provide a closed form expression for the k-th derivative of function F. We also show that the number of zeros of the k-th derivative is increasing with k.

Lemma 9.1. *The k-th derivative of $F(x)$ $(k > 0)$ can be put on the following expression :*

$$F^{(k)}(x) = (-1)^{k+1} \sum_{i=1}^{q} \frac{b_{i,k}x - a_{i,k}}{(1 + d_i x)^{k+2}} \tag{9.2}$$

where $a_{i,k}$ and $b_{i,k}$ are sequences of positive reals given by:

$$\begin{cases} b_{i,k} &= 2k!c_i d_i^{k-1} \\ a_{i,k} &= 2k!c_i d_i^{k-1}\left(\frac{k+1}{2}\frac{d_i}{c_i} + \frac{k-1}{2d_i}\right) \end{cases}$$

Proof. See Appendix D. □

Given the previous expressions of $F^{(k)}$, we are able to prove our first step result concerning the increasing number of zeros of $F^{(k)}$. We have the following lemma:

Lemma 9.2. *Let Z_k denote the number of zeros of the k-th derivative of F given by (9.2). Then $Z_{k+1} \geq Z_k$.*

Proof. Let $x_1, \cdots, x_{Z_k}$ denote the zeros of the k-th derivative $F^{(k)}$ in $[0, \infty[$. Therefore, using Rolle's Theorem [26], $F^{(k+1)}$ has at least $Z_k - 1$ zeros $y_1, \cdots, y_{Z_k-1}$ where $x_i \leq y_i \leq x_{i+1}, i \in \{1, \cdots, Z_k - 1\}$. Since $\lim_{x\to+\infty} F^{(k)}(x) = 0$, there exist at least one zero of $F^{(k+1)}$ in $\left[x_{Z_k}, \infty\right[$. Consequently, the number of zeros of $F^{(k+1)}$ is at least equal to $Z^{(k)}$, i.e $Z_{k+1} \geq Z_k$. □

9.4.2 Uniform equivalence of G_k

In this step, we introduce an alternative function G_k that has the same number of positive valued zeros as $F^{(k)}$ and we provide its asymptotic equivalent expression. For that, let us start by providing a useful approximation of coefficient $a_{i,k}$ that will be used later to build the function G_k.

The Stirling formulae [57] provides us an equivalent [2] for $k!$:

$$k! \sim \sqrt{2\pi k}\left(\frac{k}{e}\right)^k$$

we can easily show that:

$$\begin{aligned} a_{i,k} &\sim \sqrt{2\pi}k^{k+\frac{3}{2}}e^{-k}c_i d_i^{k-1}\left(\frac{d_i}{c_i} + \frac{1}{d_i}\right) \\ &\sim \sqrt{2\pi}k^{k+\frac{3}{2}}e^{-k}c_i d_i^{k+2}\left(\frac{1}{c_i d_i^2} + \frac{1}{d_i^4}\right). \end{aligned} \tag{9.3}$$

We recall that the overall quasi-convexity proof is based on studying the zeros of the function $F^{(k)}$ as k goes to infinity. Actually, one can show[3] that these zeros belong to the interval $\left[V_{\min}^k, V_{\max}^k\right]$, where $V_{\min}^k = \min_{i\in\{1,\cdots q\}} \frac{a_{i,k}}{b_{i,k}}$ and $V_{\max}^k = \max_{i\in\{1,\cdots q\}} \frac{a_{i,k}}{b_{i,k}}$. A lower bound for $V_{\min}^k$ and an upper bound for $V_{\max}^k$ can be easily computed and are given by:

$$V_{\min}^k \geq k\tau_{\min} \tag{9.4}$$

$$V_{\max}^k \leq k\tau_{\max} \tag{9.5}$$

[2] Equivalence here means that $\frac{k!}{\sqrt{2\pi k}\left(\frac{k}{e}\right)^k} \xrightarrow[k\to\infty]{} 1$

[3] Outside this interval, all the terms in the sum given by (9.2) have the same sign and hence $F^{(k)}$ cannot be zero.

where $\tau_{\min} = \min_{i\in\{1,\cdots q\}} \frac{d_i}{2c_i}$ and $\tau_{\max} = \max_{i\in\{1,\cdots q\}} \frac{d_i}{c_i} + \frac{1}{2d_i}$. The difficulty that we face is that the zeros of $F^{(k)}$ are of order k, thus making the analysis of the asymptotic behavior of function $F^{(k)}$ somehow delicate. To deal with this difficulty, another function, denoted by G_k, and which brings back those zeros to a given fixed interval is introduced. This function will be studied over the interval of interest $[\tau_{\min}, \tau_{\max}]$.

Function G_k is defined as:

$$G_k(x) = (-1)^{k+1}\sqrt{\frac{k}{2\pi}}e^k x^{k+2} F^{(k)}(kx) \tag{9.6}$$

One can easily note that over $[\tau_{\min}, \tau_{\max}]$, $G_k(x)$ have the same number of zeros as $F^{(k)}$.

Substituting $F^{(k)}$ by its expression in (9.2), G_k writes as:

$$\begin{aligned} G_k(x) &= \sqrt{\frac{k}{2\pi}}e^k x^{k+2} \sum_{i=1}^{q} \frac{kb_{i,k}x - a_{i,k}}{(1+kd_i x)^{k+2}} \\ &= \sqrt{\frac{k}{2\pi}}e^k \sum_{i=1}^{q} \frac{a_{i,k}\left(\frac{kb_{i,k}}{a_{i,k}}x - 1\right)}{k^{k+2}d_i^{k+2}\left(\frac{1}{kd_i x} + 1\right)^{k+2}} \\ &\triangleq \sum_{i=1}^{q} \frac{g_{i,k}(x)}{h_{i,k}(x)} \end{aligned}$$

where $g_{i,k}(x) \triangleq \frac{\sqrt{k}e^k}{\sqrt{2\pi}d_i^{k+2}k^{k+2}} a_{i,k}\left(\frac{kb_{i,k}}{a_{i,k}}x - 1\right)$ and $h_{i,k}(x) = (\frac{1}{kd_i x} + 1)^{k+2}$.

In the following, we extend the domain of the function G_k to the rectangle R of $\mathbb{C}$ given by:

$$\mathcal{R}_\epsilon = \{z = x + iy, x \in [\tau_{\min}, \tau_{\max}], -\epsilon \le y \le \epsilon\}$$

where ϵ is a constant real that will be specified later. Over this domain, the asymptotic equivalent of G_k is given by the following theorem:

Theorem 9.5. *In the rectangle $\mathcal{R}_\epsilon$, G_k converges uniformly to G_∞ given by:*

$$G_\infty(z) = \sum_{i=1}^{q} c_i \left(\frac{1}{c_i d_i^2} + \frac{1}{d_i^4}\right)(V_{\infty,i} z - 1)\, e^{-\frac{1}{d_i z}}$$

where $V_{\infty,i} = \frac{2}{\frac{d_i}{c_i} + \frac{1}{d_i}}$.

Proof. See Appendix D.2. □

9.4.3 Zeros of the uniform limit of G_k

In this section, we prove that G_∞ has a unique positive real zero. This is a by-product of the following theorem:

Theorem 9.6. *Let a_i b_i and α_i three sequences of q strictly positive real scalars. Let f be the function given by:*

$$f(x) = \sum_{i=1}^{q} (a_i x - b_i) e^{-\alpha_i x} \tag{9.7}$$

Then f admits a unique real positive zero.

Proof. See Appendix D.3. □

By defining $f(z) = zG_\infty(\frac{1}{z})$ and applying Theorem9.6, we conclude that $G_\infty(z)$ has a unique real positive zero.

9.4.4 Application of Hurwitz theorem

To prove that from a certain range of k G_k is zero only once at the real positive axis, we will rely on the following known result in complex analysis, [90]:

Theorem 9.7. *Let $f_k(z)$ be a sequence of analytic functions in a compact $\mathbf{C}$. Assume that f_k converges uniformly to f in $\mathbf{C}$. Assume also that f has no zeros on the frontier $\partial\mathbf{C}$ of $\mathbf{C}$. Then, there exists $k_0 \in \mathbb{N}$ such that $\forall k \geq k_0$, f and f_k have the same number of zeros in $\mathbf{C}$.*

Applying this theorem, we can deduce that, G_k will have a unique zero value in $\mathcal{R}_\epsilon$ as G_∞, where ϵ is chosen so that G_∞ has no zeros on the frontier of $\mathcal{R}_\epsilon$ and has no complex zeros besides its real positive zero. Since the number of zeros of G_k is increasing with respect to k, we conclude that all G_k and hence all $F^{(k)}$ have only a unique positive zero.

Let x_z be the unique positive zero argument of $F^{(1)}$. Since $F^{(1)}$ is negative in a neighborhood of zero, and that $F^{(1)}$ has no zeros for $x \leq x_z$, $F^{(1)}$ is negative in the interval $[0, x_z]$. Therefore the function F is decreasing in $[0, x_z]$.

Since $F^{(1)}$ is positive for large value of x, $F^{(1)}$ must change its sign at x_z, and hence it is positive in the interval $[x_z, \infty[$. Consequently, F is increasing in $[x_z, \infty[$.

To sum up, we have established that in $[0, x_z]$, F is decreasing and in $[x_z, \infty[$ F is increasing. This guarantees that x_z is a minimum for F and hence $F^{(2)}(x_z) \geq 0$. In fact, $F^{(2)}(x_z)$ is strictly positive, since $F^{(1)}(x_z) = 0$ and $\lim_{x\to\infty} F^{(1)}(x) = 0$ means that there exists $y_z \in]x_z, \infty[$ such that $F^{(2)}(y_z) = 0$ and hence $F^{(2)}(x_z) \neq 0$ (because $F^{(2)}$ has a unique zero).

9.5 Conclusion

In this chapter, we have provided a rigorous proof for the quasi-convexity of the asymptotic MSE of the regularized semi-blind channel estimate. More generally we have proved that any function given by a finite sum of quadratic fractions $\frac{1+cx^2}{(1+dx)^2}, c, d > 0$ is a unimodal function over $\mathbb{R}^+$. As far as semi-blind estimation is considered, the previous result guarantees the absence of non-desired local minima of the MSE function when optimized with respect to the weighting coefficient. However, the optimal value of the regularizing scalar does not have closed-form expression, and hence is estimated by using iterative numerical optimization. To avoid the use of numerical algorithms, we propose in the next chapter a new technique where the regularizing constant has closed-form expression.

Appendices

Appendix D

Proof of results of chapter 9

This appendix provides the proof of the results stated in chapter9.

D.1 Proof of lemma 9.1

Lemma 9.1 can be proved easily by induction on k. For $k = 1$, we have:

$$\begin{aligned} F^{(1)}(x) &= \sum_{i=1}^{q} \frac{2c_i x(1+d_i x)^2 - 2d_i(1+c_i x^2)(1+d_i x)}{(1+d_i x)^4} \\ &= \sum_{i=1}^{q} \frac{(1+d_i x)(2c_i x(1+d_i x) - 2d_i(1+c_i x^2))}{(1+d_i x)^4} \\ &= \sum_{i=1}^{q} 2\frac{c_i x - d_i}{(1+d_i x)^3} \end{aligned}$$

Let $k \in \mathbb{N}^*$. Assume that the result is true until order k. Hence, $F^{(k)}$ can be written as:

$$F^{(k)}(x) = (-1)^{k+1} \sum_{i=1}^{q} \frac{b_{i,k}x - a_{i,k}}{(1+d_i x)^{k+2}}$$

Therefore,

$$\begin{aligned} F^{(k+1)}(x) &= (-1)^{k+1} \left(\sum_{i=1}^{q} \frac{b_{i,k}(1+d_i x)^{k+2}}{(1+d_i x)^{2k+4}} \right. \\ &\quad \left. - \frac{(k+2)d_i(1+d_i x)^{k+1}(b_{i,k}x - a_{i,k})}{(1+d_i x)^{2k+4}} \right) \\ &= (-1)^{k+1} \sum_{i=1}^{q} \frac{b_{i,k} + (k+2)d_i a_{i,k} - (k+1)d_i b_{i,k} x}{(1+d_i x)^{k+3}} \\ &= (-1)^{k+2} \sum_{i=1}^{q} \frac{(k+1)b_{i,k}d_i x - (b_{i,k} + (k+2)d_i a_{i,k})}{(1+d_i x)^{k+3}} \\ &= (-1)^{k+2} \sum_{i=1}^{q} \frac{b_{i,k+1}x - a_{i,k+1}}{(1+d_i x)^{k+3}} \end{aligned}$$

where $b_{i,k+1} = (k+1)b_{i,k}d_i$ and $a_{i,k+1} = (b_{i,k} + (k+2)d_i a_{i,k})$. Since $b_{i,k} = 2k!d_i^{k-1}c_i$, we get $b_{i,k+1} = 2(k+1)!d_i^k c_i$.

Also,

$$\begin{aligned} a_{i,k+1} &= b_{i,k} + (k+2)d_i a_{i,k} \\ &= 2k!c_i d_i^{k-1} + 2k!(k+2)d_i^k c_i \left(\frac{k+1}{2}\frac{d_i}{c_i} + \frac{k-1}{2d_i}\right) \\ &= 2(k+1)!c_i d_i^k \left(\frac{1}{(k+1)d_i} + \frac{k+2}{k+1}\left(\frac{k+1}{2}\frac{d_i}{c_i} + \frac{k-1}{2d_i}\right)\right) \\ &= 2(k+1)!c_i d_i^k \left(\frac{k}{2d_i} + \frac{k+2}{2}\frac{d_i}{c_i}\right). \end{aligned}$$

D.2 Proof of theorem 9.5

In this lemma, we propose to find the uniform-limit function for the function $G_k(z) = \sum_{i=1}^n \frac{h_{i,k}(z)}{g_{i,k}(z)}$ in the rectangle $\mathcal{R}_\epsilon$. For that, we will first begin by finding the uniform limit functions of $h_{i,k}$ and $g_{i,k}$.

Lemma D.1. *In the rectangle $\mathcal{R}_\epsilon$, the sequence of functions $(h_{i,k})_k$ converges uniformly to $h_{i,\infty}$ given by:*

$$h_{i,\infty}(z) = e^{\frac{1}{d_i z}}$$

Also, the sequence of functions $(g_{i,k})_k$ converges uniformly to $g_{i,\infty}$ given by:

$$g_{i,\infty} = c_i \left(\frac{1}{c_i d_i^2} + \frac{1}{d_i^4}\right)(V_{\infty,i} z - 1)$$

where $V_{\infty,i} = \lim_{k\to\infty} \frac{kb_{i,k}}{a_{i,k}} = \frac{2}{\frac{d_i}{c_i} + \frac{1}{d_i}}$.

Proof. The uniform convergence of $h_{i,k}$ to $h_{i,\infty}$ is a by-product of the following known result:

Lemma D.2. *Over a compact set the sequence function $(1 + \frac{z}{n})^n$ converges uniformly to e^z.*

The uniform convergence of $g_{i,k}$ to $g_{i,\infty}$ is obtained by using the asymptotic equivalent of $a_{i,k}$ given in (9.3). □

The uniform convergence of $h_{i,k}$ to $h_{i,\infty}$ and of $g_{i,k}$ to $g_{i,\infty}$ does not ensure the uniform convergence of $\frac{g_{i,k}}{h_{i,k}}$ to $\frac{g_{i,\infty}}{h_{i,\infty}}$.

Other extra conditions are needed as it will be noticed in the following lemma:

Lemma D.3. *Let f_k and g_k denote sequences of continuous functions over a compact C. Assume that g_k is bounded over C away from zero uniformly in k and in z, i.e there exists a constant M such that:*

$$\forall k \in \mathbb{N}, \forall z \in C \; |g_k(z)| > M.$$

Assume also that f_k and g_k converge uniformly to f_∞ and g_∞. Then, $\frac{f_k}{g_k}$ converges uniformly to $\frac{f_\infty}{g_\infty}$ over the compact C.

Proof. Since f_k and g_k are continuous, their uniform limits f_∞ and g_∞ are also continuous. Therefore, there exists constant reals M_f, M_g such that:

$$\forall z \in C,\ |f_\infty(z)| \leq M_f \quad \text{and} \quad |g_\infty(z)| \leq M_g.$$

Since for all $k \in \mathbb{N}$, $|g_k(z)| > M$, we have $|g_\infty(z)| > M$ To prove the uniform convergence of $\frac{f_k}{g_k}$ towards $\frac{f_\infty}{g_\infty}$, it is sufficient to prove that $\sup_{z\in C}\left|\frac{f_k}{g_k} - \frac{f}{g}\right|$ converges to zero as k tends to infinity.

We have:

$$\begin{aligned}
\sup_{z\in C}\left|\frac{f_k}{g_k} - \frac{f}{g}\right| &= \sup_{z\in C}\left|\frac{f_k g - f g_k}{g_k g}\right| \\
&\leq \sup_{z\in C}\left|\frac{f_k g - f g_k}{M^2}\right| \\
&\leq \frac{1}{M^2}\left(\sup_{z\in C}|f_k g - fg| + \sup_{z\in C}|fg - fg_k|\right) \\
&\leq \frac{1}{M^2}\left(M_g \sup_{z\in C}|f_k - f| + M_f \sup_{z\in C}|g_k - g|\right) \xrightarrow[k\to\infty]{} 0
\end{aligned}$$

which proves that $\frac{f_k}{g_k}$ converges uniformly to $\frac{f_\infty}{g_\infty}$. □

Since $|h_{i,k}(z)| > 1$ over $\mathcal{R}_\epsilon$, $h_{i,k}$ satisfies the condition of lemma D.3. Applying this lemma on the functions $g_{i,k}$ and $h_{i,k}$, we prove that $\frac{g_{i,k}}{h_{i,k}}$ converges uniformly to $\frac{g_{i,\infty}}{h_{i,\infty}}$. Consequently, $G_k(z) = \sum_{i=1}^{n} \frac{g_{i,k}(z)}{h_{i,k}(z)}$ converges uniformly over $\mathcal{R}_\epsilon$ to $G_\infty(z) = \sum_{i=1}^{n} \frac{g_{i,\infty}(z)}{h_{i,\infty}(z)}$.

D.3 Proof of Theorem 9.6

The proof is performed by induction on q. For $q = 1$, the result is straightforward. Let $q \in \mathbb{N}^*$ be a given integer, and assume that the result holds true for all $k \leq q$, and all functions f of the form given by (9.7). Assume that there exists a_i, b_i and α_i three sequences of $q+1$ strictly positive real scalars such that the function

$$f(x) = \sum_{i=1}^{q+1}(a_i x - b_i)e^{-\alpha_i x}$$

admits more than one positive zero. Let x_1 be the first smallest zero of f on $\mathbb{R}^+$,

Without loss of generality, we can assume that all the α_i are two by two different and that $\alpha_{q+1} = \min_{1\leq i\leq q+1} \alpha_i$. Since f is strictly negative in zero and is positive for large values of x, f should change its sign at at least one zero. In the following we will consider only the case when f changes its sign at x_1. The other case could be treated in the same way. Let x_2 be the second smallest zero of f on $\mathbb{R}^+$. Under this condition, we distinuish the following cases:

- f changes its sign at x_1 and at x_2.
- f changes its sign only at x_1.

For the both cases, we can prove that the second derivative of

$$\begin{aligned} g_m(x) &= e^{(\alpha_{q+1}-\frac{1}{m})x} f(x) \\ &= \sum_{i=1}^{q} (a_i x - b_i) e^{-(\alpha_i - \alpha_{q+1} + \frac{1}{m})x} + (a_{q+1}x - b_{q+1})\, e^{-\frac{x}{m}}, \ \text{for}\, m \in \mathbb{N}^* \end{aligned}$$

has three zeros. More particularly, we have the following:

Case 1: f changes its sign at x_1 and at x_2

Since $f(0) < 0$, $f(x) < 0$ for $x \leq x_1$. Therefore, for $x \geq x_2$ and in the vicinity of x_2, $f(x) < 0$ for $x \geq x_2$. Since $f(x) > 0$ for x large enough, f should have a third zero $x_3 > x_2$.

For all integers m, we note that f and g_m have the same number of zeros. Using Rolles theorem, it can be proved that the derivative of g_m which we denote $g_m^{(1)}$ and which is given by:

$$\begin{aligned} g_m^{(1)}(x) = \sum_{i=1}^{q} \left[-(\alpha_i - \alpha_{q+1} + \frac{1}{m}) a_i x + b_i(\alpha_i - \alpha_{q+1} + \frac{1}{m}) + a_i \right] e^{-(\alpha_i - \alpha_{q+1} + \frac{1}{m})x} \\ + a_{q+1} e^{-\frac{1}{m}x} - \frac{1}{m}\,(a_{q+1}x - b_{q+1})\, e^{-\frac{x}{m}} \end{aligned} \tag{D.1}$$

has at least three zeros, since $g_m^{(1)}(x)$ tends to zero as x tends to infinity.

Also again by using the Rolle's theorem, we conclude that the second derivative of g_m denoted by $g_m^{(2)}(x)$ has at least two zeros.

Case 2: f changes its sign at only one zero In this case, we can also prove that the first derivative of g_m has three zeros. Actually, at x_2, the first derivative of g_m must be also zero, since x_2 is a local minimum for f and hence for g_m. As g_m tends to zero when x tends to infinity, $g_m^{(1)}$ has two zeros between $]x_1, x_2[$ and $]x_2, \infty[$. Consequently, in total, $g_m^{(1)}$ has at least three zeros, and therefore, the second derivative of g_m has at least two zeros.

Taking the derivative of (D.1), $g_m^{(2)}(x)$ writes as:

$$\begin{aligned} g_m^{(2)}(x) &= \sum_{i=1}^{q} (\alpha_i - \alpha_{q+1} + \frac{1}{m}) \left[a_i(\alpha_i - \alpha_{q+1} + \frac{1}{m})x - 2a_i - b_i(\alpha_i - \alpha_{q+1} + \frac{1}{m}) \right] e^{-(\alpha_i - \alpha_{q+1} + \frac{1}{m})x} \\ &\quad -\frac{2}{m} a_{q+1} e^{-\frac{1}{m}x} + \frac{1}{m^2}(a_{q+1}x - b_{q+1}) e^{-\frac{x}{m}} \end{aligned}$$

Extending the definition domain of $g_m^{(2)}$ to $\mathbb{C}^+ = \{z = x + iy, x > 0\}$, we note that for every compact in $\mathbb{C}^+$, $g_m^{(2)}$ converges uniformly to g_∞ given by:

$$g_\infty(z) = \sum_{i=1}^{q} (\alpha_i - \alpha_{q+1})\, (a_i(\alpha_i - \alpha_{q+1})z - 2a_i - b_i(\alpha_i - \alpha_{q+1}))\, e^{-(\alpha_i - \alpha_{q+1})z}$$

Let $\mathcal{C}$ be the contour corresponding to the rectangle

$$R_\epsilon = \left\{ x + iy, x \in \left[\inf_{m,i} \frac{2a_i + b_i(\alpha_i - \alpha_{q+1}) + \frac{1}{m}}{a_i(\alpha_i - \alpha_{q+1} + \frac{1}{m})}, \sup_{m,i} \frac{2a_i + b_i(\alpha_i - \alpha_{q+1}) + \frac{1}{m}}{a_i(\alpha_i - \alpha_{q+1} + \frac{1}{m})} \right], y \in [-\epsilon, \epsilon] \right\},$$

ϵ is chosen such that $|g_\infty|$ is bounded above zero in $\mathcal{C}$ and has no complex valued zeros. Then referring to Hurwitz theorem, g_∞ and $g_m^{(2)}$ will have the same number of zeros in R_ϵ for large enough values of m, which is in contradiction with the induction assumption.

Chapter 10

Matrix regularized semi-blind channel estimation techniques

As mentioned in previous chapters, the optimization of the regularizing constant in semi-blind estimation techniques is mandatory, otherwise, the expected improvement in performance could not be achieved. Given its impact on the channel estimation quality, the regularizing constant has been proposed to be set in such a way that it minimizes the asymptotic MSE, [15]. Since the optimal regularizing constant has no closed-form expression, [15] proposes to employ iterative algorithms that converge to the optimal solution. On the other hand, [66] proposes to evaluate the asymptotic MSE at finite discrete possible values for the regularizing constant and keep thereafter the value that exhibits the least channel estimation error. In [111], an explicit formula was given by assuming that the minimization of the semi-blind cost can be transformed into a Weighted-Least-Square (WLS)[106] minimization problem.

In this chapter, we propose to use a regularizing matrix instead of a regularizing constant. In contrast to preliminary predictions, the optimization of the proposed method is less complicated and yields a closed-form expression for the regularizing matrix. It is also shown to slightly outperform the conventional regularized scheme proposed in [15], while avoiding the need for iterative computations. This technique will be first applied for SIMO systems, and then generalized to the context of MIMO-OFDM.

10.1 Matrix regularized semi-blind estimation for SIMO systems

We consider a SIMO system with N receiving antennas, operating over a frequency selective channel. We assume the following discrete time model:

$$\mathbf{y}(k) = \sum_{l=0}^{L} \mathbf{h}_l s_{k-l} + \mathbf{b}_k$$

where s_k are assumed to be i.i.d complex circular random variables with $\mathbb{E}s_k = 0$ and $\mathbb{E}|s_k|^2 = 1$.

During the training period, the received gets m samples $\mathbf{y} = [\mathbf{y}^{\mathrm{T}}(1), \cdots, \mathbf{y}^{\mathrm{T}}(m)]^{\mathrm{T}}$ that depend on the training sequence. The vector $\mathbf{y}$ can be written as:

$$\mathbf{y} = \mathbf{S}\mathbf{h} + \mathbf{b}$$

where $\mathbf{b} = [\mathbf{b}^{\mathrm{T}}(1), \cdots, \mathbf{b}^{\mathrm{T}}(m)]^{\mathrm{T}}$, $\mathbf{S} = [\mathbf{S}_1, \cdots, \mathbf{S}_m]^{\mathrm{T}}$, $\mathbf{S}_k^{\mathrm{T}} = [s(k), \cdots, s(k-L)] \otimes \mathbf{I}_N$ and $\mathbf{h} = [\mathbf{h}^{\mathrm{T}}(0), \cdots, \mathbf{h}^{\mathrm{T}}(L)]^{\mathrm{T}}$. The least square estimate of $\mathbf{h}$ is therefore given by:

$$\begin{aligned}\widehat{\mathbf{h}} &= (\mathbf{S}^{\mathrm{H}}\mathbf{S})^{-1}\mathbf{S}^{\mathrm{H}}\mathbf{y} \\ &= \mathbf{S}^{\#}\mathbf{y}\end{aligned}$$

where it is assumed that the training sequence is properly chosen to avoid the degeneracy of $\mathbf{S}^{\mathrm{H}}\mathbf{S}$.

Stacking M observations of vector $\mathbf{y}(k)$ into a $NM \times 1$ vector $\mathbf{y}_M(k)$, we will get:

$$\mathbf{y}_M(k) = \boldsymbol{\mathcal{I}}_M(\mathbf{h})\mathbf{s}_M(k) + \mathbf{b}_M(k)$$

where $\boldsymbol{\mathcal{I}}_M(\mathbf{h})$ is the $NM \times L + M$ block-Toeplitz matrix given by (7.1). Let n denote the length of the data sequence, and $\widehat{\mathbf{R}}_M$ the estimated covariance matrix. The most common blind estimation techniques evaluate the channel up to a scalar ambiguity by solving the following minimization problem:

$$\min_{\|\mathbf{h}\|=1} \mathbf{h}^{\mathrm{H}}\widehat{\mathbf{A}}_n\mathbf{h} \tag{10.1}$$

where $\widehat{\mathbf{A}}_n$ is an estimated matrix of $\mathbf{A}$, $\mathbf{A}$ being a matrix that depends on the considered blind estimation technique. Note that the matrix $\mathbf{A}$ is singular and $\mathbf{h}$ is the unique vector (up to a scalar factor) generating its kernel.

10.1.1 Semi-blind estimation

In the conventional regularized semi-blind estimation technique, the blind criterion is combined linearly with the training sequence criterion, thus leading to the following cost function:

$$C(\mathbf{f}, \alpha) = \|\mathbf{y} - \mathbf{S}\mathbf{f}\|^2 + \alpha n \mathbf{f}^{\mathrm{H}}\widehat{\mathbf{A}}_n\mathbf{f}, \tag{10.2}$$

where α is a regularizing constant. The semi-blind estimator that minimizes (10.2) is given by:

$$\widehat{\mathbf{h}}_\alpha(\mathbf{A}) = \left(\mathbf{S}^{\mathrm{H}}\mathbf{S} + \alpha n\widehat{\mathbf{A}}_n\right)^{-1}\mathbf{S}^{\mathrm{H}}\mathbf{y}.$$

The optimization of the regularizing constant cannot be performed directly and requires in general the use of iterative algorithms. In this work, we propose to minimize the following cost function that is given by:

$$C(\mathbf{f}, \boldsymbol{\Lambda}) = \|\mathbf{y} - \mathbf{S}\mathbf{f}\|^2 + n\mathbf{f}^{\mathrm{H}}\widehat{\mathbf{P}}_n\boldsymbol{\Lambda}\widehat{\mathbf{P}}_n\mathbf{f}, \tag{10.3}$$

where $\boldsymbol{\Lambda}$ is a regularizing matrix assumed to be hermitian, and $\widehat{\mathbf{P}}_n$ is an estimate of the orthogonal projector $\mathbf{P}$ onto the space spanned by the columns of $\mathbf{A}$. In this case, the semi-blind estimator that minimizes (10.3) is given by:

$$\widehat{\mathbf{h}}_{\boldsymbol{\Lambda}}(\mathbf{A}) = \left(\mathbf{S}^{\mathrm{H}}\mathbf{S} + n\widehat{\mathbf{P}}_n\boldsymbol{\Lambda}\widehat{\mathbf{P}}_n\right)^{-1}\mathbf{S}^{\mathrm{H}}\mathbf{y}.$$

Interestingly, we show in this chapter that a closed-form expression for the optimum regularizing matrix $\boldsymbol{\Lambda}$ exists, thus avoiding the need for iterative algorithms.

10.1.2 Asymptotic MSE

The expression of the asymptotic MSE has been derived in [15] when the subspace blind criterion is considered.Hereafter, we show that, under some mild assumptions, the results established in [15] can be extended to any blind method characterized by (10.1). In particular we can prove along the same lines of [15], the following results:

Theorem 10.1. *Let* $\gamma = \lim_{n\to\infty} \frac{n}{m}$. *For any matrix* $\mathbf{A}$ *that verifies:*

- $\mathbf{Ah} = 0$,
- $\delta\mathbf{A} \triangleq \widehat{\mathbf{A}}_n - \mathbf{A} = \mathcal{O}_p\left(n^{-\frac{1}{2}}\right)$.

we have

$$\widehat{\mathbf{h}}_\alpha(\mathbf{A}) - \mathbf{h} = \left(\mathbf{R_{SS}} + \gamma\alpha\mathbf{A}\right)^{-1}\left(\mathbf{R_{SV}} - \gamma\alpha\delta\mathbf{Ah}\right) + \mathcal{O}_p(\frac{1}{n}),$$

$$\widehat{\mathbf{h}}_{\mathbf{\Lambda}}(\mathbf{A}) - \mathbf{h} = \left(\mathbf{R_{SS}} + \gamma\mathbf{P\Lambda P}\right)^{-1}\left(\mathbf{R_{SV}} - \gamma\delta\left(\mathbf{P\Lambda P}\right)\mathbf{h}\right) + \mathcal{O}_p(\frac{1}{n}),$$

where $\mathbf{R_{SS}} = \frac{1}{m}\mathbf{S}^{\mathrm{H}}\mathbf{S}$ *and* $\mathbf{R_{SV}} = \frac{1}{m}\mathbf{S}^{\mathrm{H}}\mathbf{v}$.

Theorem 10.2. $\sqrt{m}\left(\widehat{\mathbf{h}}_\alpha(\mathbf{A}) - \mathbf{h}\right)$ *is asymptotically normal with covariance matrix* $\Gamma_\alpha\left(\mathbf{h}\right)$ *given by:*

$$\Gamma_\alpha(\mathbf{h}) = \left(\mathbf{I} + \alpha\gamma\mathbf{A}\right)^{-1}\left(\sigma_b^2\mathbf{I} + \alpha^2\gamma^2 \lim_{\substack{n\to\infty \\ \frac{n}{m}\to\gamma}} m\mathrm{Cov}(\delta\mathbf{Ah})\right)\left(\mathbf{I} + \alpha\gamma\mathbf{A}\right)^{-1}. \tag{10.4}$$

Also, $\sqrt{m}\left(\widehat{\mathbf{h}}_{\mathbf{\Lambda}}(\mathbf{A}) - \mathbf{h}\right)$ *is asymptotically normal with covariance matrix* $\Gamma_{\mathbf{\Lambda}}\left(\mathbf{h}\right)$ *given by:*

$$\Gamma_{\mathbf{\Lambda}}(\mathbf{h}) = \left(\mathbf{I} + \gamma\mathbf{P\Lambda P}\right)^{-1}\left(\sigma_b^2\mathbf{I} + \gamma^2 \lim_{\substack{n\to\infty \\ \frac{n}{m}\to\gamma}} m\mathrm{Cov}(\delta(\mathbf{P\Lambda Ph}))\right)\left(\mathbf{I} + \gamma\mathbf{P\Lambda P}\right)^{-1}. \tag{10.5}$$

Proof. : The asymptotic normality can be proved along the same lines as in [15]. Then, the expression for the covariance matrix is easily deduced from Theorem 10.1. □

We can see from (10.5) that the asymptotic MSE expression depends on the covariance of $\delta\mathbf{P\Lambda P}$. This quantity can be given as:

Theorem 10.3. *The covariance matrix of* $\delta\mathbf{P\Lambda Ph}$ *is given by:*

$$\mathrm{Cov}\delta\left(\mathbf{P\Lambda Ph}\right) = \mathbf{P\Lambda A}^{\#}\mathrm{Cov}(\delta\mathbf{Ah})\mathbf{A}^{\#}\mathbf{\Lambda P} + \mathcal{O}_p(\frac{1}{n^2}).$$

Proof. Let $\mathbf{P}_\perp$ be the orthogonal projector onto the null space of $\mathbf{A}$. Thus, $\mathbf{P} = \mathbf{I} - \mathbf{P}_\perp$. Using standard perturbation formulae [96], we get:

$$\widehat{\mathbf{P}}_\perp = \mathbf{P}_\perp - \mathbf{P}_\perp\delta\mathbf{A}\mathbf{A}^{\#} - \mathbf{A}^{\#}\delta\mathbf{A}\mathbf{P}_\perp + \mathcal{O}_p(\|\delta\mathbf{A}\|^2).$$

Assuming that $\delta\mathbf{A} = \mathcal{O}_p(\frac{1}{\sqrt{n}})$, which is valid if $\widehat{\mathbf{A}}_n$ is related to $\widehat{\mathbf{R}}_n$ through an infinitely differentiable mapping, we get:

$$\delta\mathbf{Ph} = (\widehat{\mathbf{P}} - \mathbf{P})\mathbf{h} = \mathbf{A}^{\#}\delta\mathbf{Ah} + \mathcal{O}_p(\frac{1}{n}),$$

thus leading to:

$$\mathrm{Cov}(\delta(\mathbf{P\Lambda P})\mathbf{h}) = \mathbf{P\Lambda}\mathrm{Cov}(\delta\mathbf{Ph})\mathbf{\Lambda P} \tag{10.6}$$

$$= \mathbf{P\Lambda A}^{\#}\mathrm{Cov}(\delta\mathbf{Ah})\mathbf{A}^{\#}\mathbf{\Lambda P} + \mathcal{O}_p(\frac{1}{n^2}). \tag{10.7}$$

□

Corollary 10.1. *Let* $\mathbf{\Sigma}_\infty = \frac{\gamma}{\sigma_b^2}\lim_{\substack{n,\to\infty \\ \frac{n}{m}\to\gamma}} m\mathrm{Cov}(\delta\mathbf{Ah})$. *Hence,*

$$\Gamma_{\mathbf{\Lambda}}(\mathbf{h}) = \sigma_b^2(\mathbf{I}+\gamma\mathbf{P\Lambda P})^{-1}\Big(\mathbf{I}+\gamma\mathbf{P\Lambda A}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}\mathbf{\Lambda P}\Big)(\mathbf{I}+\gamma\mathbf{P\Lambda P})^{-1}.$$

10.1.3 Optimal regularizing matrix of $\mathbf{\Lambda}$

In the following theorem, we give the closed-form expression for the optimal regularizing matrix $\mathbf{\Lambda}$ that minimizes the trace of $\Gamma_{\mathbf{\Lambda}}(\mathbf{h})$.

Theorem 10.4. *Assuming that the rank of* $\mathbf{A}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}$ *is equal to the rank of* $\mathbf{A}$*, the optimal regularizing matrix* $\mathbf{\Lambda}$ *is given by:*

$$\mathbf{\Lambda} = \left(\mathbf{A}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}\right)^{\#}.$$

Proof. : To determine the optimal regularizing matrix $\mathbf{\Lambda}$, we optimize the asymptotic channel estimation error first with respect to the eigenvalues of $\mathbf{P\Lambda P}$ and then with respect to the eigenvectors.

Optimization with respect to the eigenvalues

Consider the eigenvalue decomposition of $\mathbf{P\Lambda P} = \mathbf{U}^{\mathrm{H}}\mathbf{DU}$, where $\mathbf{D} = \mathrm{diag}(d_1, \cdots, d_{N(L+1)})$. Without loss of generality, we assume that $\mathbf{P}$ has a unique zero eigenvalue, i.e, $d_1 = 0$. Then, we have:

$$\mathrm{Tr}(\Gamma_{\mathbf{\Lambda}}(\mathbf{h})) = \sum_{i=2}^{N(L+1)} \frac{1+\gamma d_i^2\left[\mathbf{UA}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}\mathbf{U}^{\mathrm{H}}\right]_{i,i}}{(1+\gamma d_i)^2}.$$

Through simple calculations, we can find that the derivative of $\mathrm{Tr}(\Gamma_{\mathbf{\Lambda}}(\mathbf{h}))$ is given by:

$$\frac{\partial\mathrm{Tr}(\Gamma_{\mathbf{\Lambda}}(\mathbf{h}))}{\partial d_i} = \sum_{i=2}^{N(L+1)} \frac{d_i\left[\mathbf{UA}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}\mathbf{U}^{\mathrm{H}}\right]_{i,i} - 1}{(1+\gamma d_i)^3}.$$

Hence, the gradient is equal to zero when $d_i = \frac{1}{\left[\mathbf{UA}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}\mathbf{U}^{\mathrm{H}}\right]_{i,i}}$. Moreover, one can easily show that in this case, the Hessian matrix is strictly positive.

Consequently, the optimal eigenvalues of $\mathbf{P\Lambda P}$ are given by:

$$d_i = \frac{1}{\left[\mathbf{UA}^{\#}\Sigma_\infty\mathbf{A}^{\#}\mathbf{U}^{\mathrm{H}}\right]_{i,i}}.$$

One can note that under the assumption that the rank of $\mathbf{A}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}$ is equal to that of $\mathbf{A}$, we have

$$\left[\mathbf{UA}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}\mathbf{U}^{\mathrm{H}}\right]_{i,i} \neq 0 \quad \forall\, i \geq 2.$$

Optimization with respect to the eigenvector basis

In the following we prove that the optimal basis of eigenvectors is the one that makes $\mathbf{UA}^{\#}\mathbf{\Sigma}_\infty\mathbf{A}^{\#}\mathbf{U}^{\mathrm{H}}$ diagonal. But before tackling the proof, we shall recall the following results:

Definition 10.1. *[71] For two vectors* $\mathbf{x}$, $\mathbf{y} \in \mathbb{R}^n$ *with descending ordered components* $x_1 \geq x_2 \cdots \geq x_n \geq 0$ *and* $y_1 \geq y_2 \cdots \geq y_n \geq 0$, *we say that* $\mathbf{x}$ *is weakly majorized by* $\mathbf{y}$ *and write* $\mathbf{x} \preceq_w \mathbf{y}$ *when:*

$$\sum_{k=1}^{m} x_k \leq \sum_{k=1}^{m} y_k \text{ for all } \ m = 1, \cdots, n.$$

Theorem 10.5. *[71] Let* $\mathbf{A}$ *be a* $n \times n$ *hermitian matrix. Then the descending ordered vector of diagonal entries of* $\mathbf{A}$ *is weakly majorized by the descending ordered vector of eigenvalues.*

Definition 10.2. *[71] A real-valued function* f *defined on* $\mathbb{R}^n$ *is said to be Schur-convex (resp. Schur-concave) on* $\mathbb{R}^n$ *if* $x \preceq_w y \Longrightarrow f(x) \leq f(y)$ *(resp. if* $x \preceq_w y \Longrightarrow f(x) \geq f(y)$ *)*

Proposition 10.1. *If* $g : \mathbb{R} \to \mathbb{R}$ *is convex (resp. concave) then*

$$f(x_1, \cdots, x_n) = \sum_{k=1}^{n} g(x_k)$$

is Schur-convex (resp. Schur-concave) on $\mathbb{R}$.

Substituting d_i by their optimal values, the asymptotic estimation error becomes:

$$\mathrm{Tr}\Gamma_{\mathbf{\Lambda}}(\mathbf{h}) \ = \ \sum_{i=2}^{N(L+1)} \frac{1}{1 + \frac{\gamma}{\left[\mathbf{U}\mathbf{A}^{\#}\Sigma_{\infty}\mathbf{A}^{\#}\mathbf{U}^{\mathrm{H}}\right]_{i,i}}}. \tag{10.8}$$

By proposition 10.1, $\mathbf{x} := (x_1, \cdots, x_{N(L+1)}) :\to \sum_{i=1}^{N(L+1)} \frac{1}{1+\frac{\gamma}{x_i}}$ is Schur-concave since $x \to \frac{1}{1+\frac{\gamma}{x}}$ is concave. Consider the eigenvalue decomposition of $\mathbf{A}^{\#}\mathbf{\Sigma}_{\infty}\mathbf{A}^{\#} = \mathbf{V}\Delta\mathbf{V}^{\mathrm{H}}$ then, according to theorem 10.5, the vector of the diagonal elements of $\mathbf{U}\mathbf{A}^{\#}\mathbf{\Sigma}_{\infty}\mathbf{A}^{\#}\mathbf{U}^{\mathrm{H}}$ is weakly majorized by the diagonal elements of $\Delta = \mathbf{V}^{\mathrm{H}}\mathbf{A}^{\#}\mathbf{\Sigma}_{\infty}\mathbf{A}^{\#}\mathbf{V}$. Using the definition of Schur-Concave functions, we conclude that the minimum of the channel estimation error is achieved when $\mathbf{V} = \mathbf{U}^{\mathrm{H}}$. Consequently, the optimal regularizing matrix $\mathbf{\Lambda} = \left(\mathbf{A}^{\#}\mathbf{\Sigma}_{\infty}\mathbf{A}^{\#}\right)^{\#}$. □

10.1.4 Application: subspace semi-blind optimal regularized estimation

In this section, we consider the case when the blind subspace criterion is considered. As we have seen in section 6.2.1 chapter 6, the blind criterion can be expressed as:

$$\mathbf{h}^{\mathrm{H}}\mathbf{Q}\mathbf{h}$$

where $\mathbf{Q} = \boldsymbol{\mathcal{D}}^{\mathrm{H}}(\Pi)\boldsymbol{\mathcal{D}}(\Pi)$, $\boldsymbol{\mathcal{D}}(\Pi)$ being defined by (6.8) (refer to chapter 6 for more details). In this case, the covariance of $\delta\mathbf{Q}\mathbf{h}$ can be expressed as:

Theorem 10.6. *[15] The covariance of* $\delta\mathbf{Q}\mathbf{h}$ *is given by:*

$$\mathrm{Cov}\left(\delta\mathbf{Q}\mathbf{h}\right) = \frac{\sigma_b^2}{n}\boldsymbol{\mathcal{M}}(\mathbf{h}) + \mathcal{O}(\frac{1}{n^2}),$$

where

$$\boldsymbol{\mathcal{M}}(\mathbf{h}) = \sum_{|\tau| \leq M} \boldsymbol{\mathcal{D}}^{\mathrm{H}}\left(\Pi\right)\left[\left(\boldsymbol{\mathcal{R}}_{\Pi_M}(\tau) + \sigma_b^2\boldsymbol{\mathcal{R}}_{\Phi_M}(\tau)\right) \otimes \boldsymbol{\mathcal{R}}_{\Pi}(\tau)\right]\boldsymbol{\mathcal{D}}\left(\Pi\right),$$

and

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\Pi}}(\tau) \triangleq \boldsymbol{\Pi}\mathbf{J}_r^{\tau N}\boldsymbol{\Pi}, \tag{10.9}$$

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\Pi}_M}(\tau) \triangleq \boldsymbol{\Pi}_M^{\mathrm{T}}\mathbf{J}_p^{\tau}\boldsymbol{\Pi}_M^{*}, \tag{10.10}$$

$$\boldsymbol{\mathcal{R}}_{\boldsymbol{\Phi}_M}(\tau) \triangleq \boldsymbol{\Phi}_M^{\mathrm{T}}\mathbf{J}_r^{\tau N}\boldsymbol{\Phi}_M^{*}, \tag{10.11}$$

$$\boldsymbol{\Pi}_M \triangleq \boldsymbol{\mathcal{I}}_M(\mathbf{h})^{\mathrm{H}}\left[\boldsymbol{\mathcal{I}}_M(\mathbf{h})\boldsymbol{\mathcal{I}}_M(\mathbf{h})^{\mathrm{H}}\right]^{\#}\boldsymbol{\mathcal{I}}_M(\mathbf{h})^{\mathrm{H}}, \tag{10.12}$$

$$\boldsymbol{\Phi}_M \triangleq \left[\boldsymbol{\mathcal{I}}_M(\mathbf{h})\boldsymbol{\mathcal{I}}_M(\mathbf{h})^{\mathrm{H}}\right]^{\#}\boldsymbol{\mathcal{I}}_M(\mathbf{h}), \tag{10.13}$$

where $\mathbf{J}_p(\tau) \triangleq \begin{bmatrix} \vdots & \mathbf{I}_{p-\tau} \\ \mathbf{0}_\tau & \cdots \end{bmatrix}$ *for* $\tau \geq 0$ *and* $\mathbf{J}_p^{\tau} \triangleq \left(\mathbf{J}_p^{-\tau}\right)^{\mathrm{T}}$ *if* $\tau < 0$. *Moreover, we have in this case:*

$$\boldsymbol{\Sigma}_\infty = \boldsymbol{\mathcal{M}}(\mathbf{h}).$$

10.1.5 Practical implementation

From the expression of $\boldsymbol{\Sigma}_\infty$, we note that the optimal value of $\boldsymbol{\Lambda}$ depends on the channel statistics which are expressed through $\boldsymbol{\mathcal{R}}_{\boldsymbol{\Pi}(\tau)}$ and also on the current channel value via the terms $\boldsymbol{\mathcal{R}}_{\boldsymbol{\Phi}_M}(\tau)$ and $\boldsymbol{\mathcal{R}}_{\boldsymbol{\Pi}_M}(\tau)$. Assuming that the channel $\mathbf{h}$ has no zeros in common, we can assume that $\boldsymbol{\Pi}_M$ is equal to the identity matrix thus removing the dependence of the term $\boldsymbol{\mathcal{R}}_{\boldsymbol{\Pi}_M}(\tau)$ on the channel. To deal with the term $\boldsymbol{\mathcal{R}}_{\boldsymbol{\Phi}_M}(\tau)$, the work in [15] suggests to substitute the unkown channel vector by an-other estimate, while the work in [65] and [66] proposes to just remove it since it is of order σ_b^4 in the expression of the asymptotic channel estimation error and thus could be removed as far as high SNR values are considered.

It has been shown that even for low SNR values, regularization is quite well performed while considering this assumption. Therefore, in this article, we will assume that:

$$\boldsymbol{\Sigma}_\infty \simeq \boldsymbol{\mathcal{D}}^{\mathrm{H}}(\Pi)\,\mathbf{M}(\mathbf{h})\boldsymbol{\mathcal{D}}(\Pi), \tag{10.14}$$

where $\mathbf{M}(\mathbf{h}) = \sum_{|\tau|\leq M}\boldsymbol{\mathcal{R}}_{\Pi_M}(\tau)\otimes\boldsymbol{\mathcal{R}}_{\Pi}(\tau)$. Also, we can prove that in this case, the optimal value of $\boldsymbol{\Lambda}$ is given by:

$$\boldsymbol{\Lambda} = \mathbf{Q}\boldsymbol{\Sigma}_\infty^{\#}\mathbf{Q}.$$

To estimate $\boldsymbol{\Sigma}_\infty$, we replace $\boldsymbol{\mathcal{D}}(\Pi)$, $\boldsymbol{\mathcal{R}}_{\Pi_M}(\tau)$ and $\boldsymbol{\mathcal{R}}_{\Pi}(\tau)$ by their respective estimates obtained from the estimated covariance matrix. Since the estimate $\widehat{\boldsymbol{\Sigma}}_\infty$ of $\boldsymbol{\Sigma}_\infty$ is an ill-conditioned matrix, we compute instead the following estimate given by:

$$\boldsymbol{\Lambda} = \widehat{\mathbf{Q}}_n\left(\widehat{\Sigma}_\infty + \sigma_b^2\mathbf{I}_{N(L+1)}\right)^{-1}\widehat{\mathbf{Q}}_n.$$

10.1.6 Simulation results

10.1.6.1 Asymptotic analysis

As it has been already mentioned, the derived asymptotic results hold in the asymptotic regime defined as $n \to \infty$ and $m \to \infty$ while $\frac{n}{m} \to \gamma$. In this section, we assess the accuracy of the derived results for finite data and training periods. Our simulations are conducted in the cases of small and also large training and data periods ($n = 104$, $m = 26$ and $n = 1040$, $m = 260$).

For each case, we set the channel to a fixed value and estimate the Mean Square Error (MSE) given by :

$$\text{MSE} = m\|\widehat{\mathbf{h}} - \mathbf{h}\|^2,$$

where the normalization by m is introduced in order to allow the comparison between the figures pertaining to different choices for m and n. More particularly, we estimate the empirical MSE for the least sqaure estimator and for both regularization-based estimators when the regularizing coefficient or matrix is set to its theoretical value or is estimated through the minimization of the channel matrix error. We compare these values to the theoretical MSE which is given by the trace of $\Gamma_\alpha(\mathbf{h})$ or $\Gamma_\Lambda(\mathbf{h})$, depending on the considered estimator. Figs. 10.1 and 10.2 display the obtained results. In the legend, 'Emp.MSE.conv.th.reg' and 'Emp.MSE.conv.est.reg' stand for the empirical MSE for the conventional regularized estimator when setting the regularizing coefficient to its optimal value and when estimating numerically the optimum regularizing coefficient, respectively. 'The.MSE.conv.th.reg' stands for the theoretical MSE of the conventional regularized estimator when the regularizing coefficient is set to its theoretical optimum value, and 'Emp.MSE.least.square' stands for the training-based least square estimator. Similar notation is used for the optimum regularized estimator. We note that even for relatively small system dimensions, optimizing the asymptotic results lead approximatively to the expected mean square error.

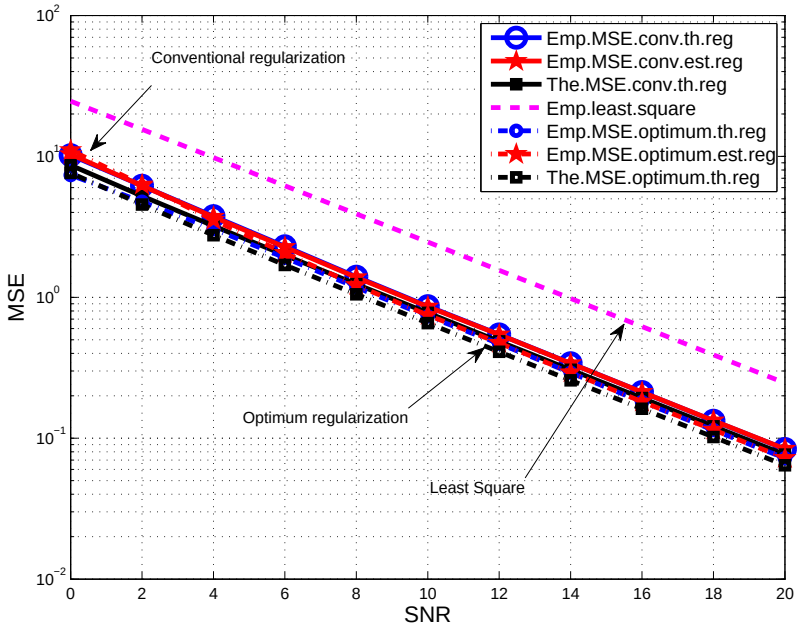


Figure 10.1: MSE vs SNR for small system dimensions ($N = 4$, $n = 104$, $m = 26$).

10.1.6.2 Bit error rate (BER)

In this section, we compare the BER performance of the least square based receiver with that of the semi-blind regularization-based estimators and that of a genie receiver which knows exactly the channel. We set the number of receiving sensors N to 6 and L to 4. We also assume that the number of training symbols is equal to 26 and that of data symbols is equal to 464. We consider the case when the channel coefficients are Rayleigh distributed with exponential

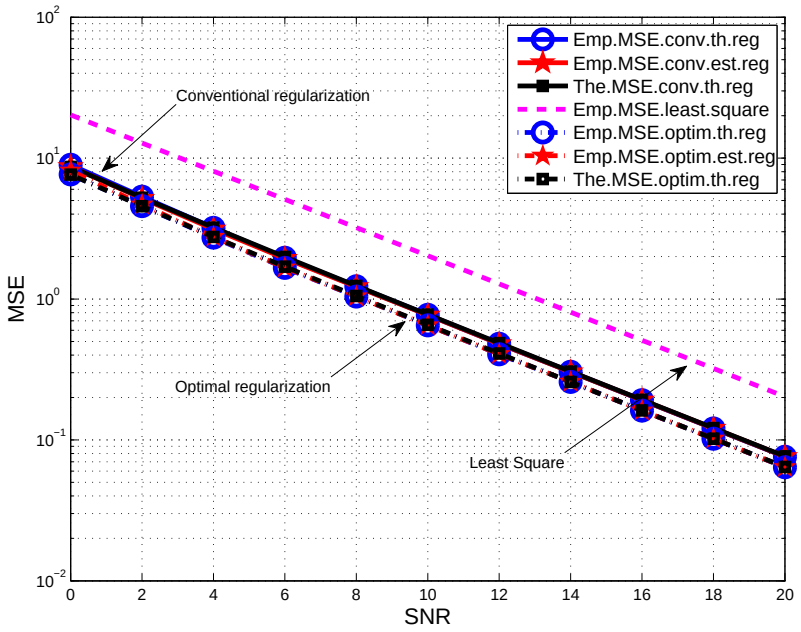


Figure 10.2: MSE vs SNR for large system dimensions ($N = 4$, $n = 1040$, $m = 260$).

decaying profile (the decay factor is taken to be 0.2). Fig. 10.3 illustrates the BER obtained over 1000 realizations when a MLSE estimation of the data symbols is performed using the Viterbi algorithm. In this context, we achieve almost the same performance as the conventional receiver with a gain of 0.8 dB over the least square estimation.

10.1.6.3 Implementation complexity

It is mentioned in [15] that the localization of the optimal regularizing scalar α requires three iterations, each of which needs the inversion of a matrix of order $N(L+1)$. However, the proposed method requires only one iteration in which one inversion matrix of order $N(L+1)$ is performed. Hence, we believe that our method exhibits lower complexity since it does not require iterations.

10.2 Matrix regularized semi-blind estimation for MIMO-OFDM systems

In this section, we extend the matrix regularized semi-blind estimation technique to the context of MIMO-OFDM systems. Exceptionally, in this part, we may use different notations from those considered earlier.

10.2.1 System model

We consider a MIMO-OFDM system with K transmitting antennas and N receiving antennas . We assume that each OFDM symbol is composed of Q subcarriers, obtained from the Inverse Discrete Fourier Transform (IDFT). Let $\mathbf{x}(n,k) = [x_1(n,k), \cdots, x_K(n,k)]$ be the n-th transmitted sample of the k-th OFDM symbol after the IDFT module, and $\mathbf{y}(n,k) = [y_1(n,k), \cdots, y_N(N,k)]$

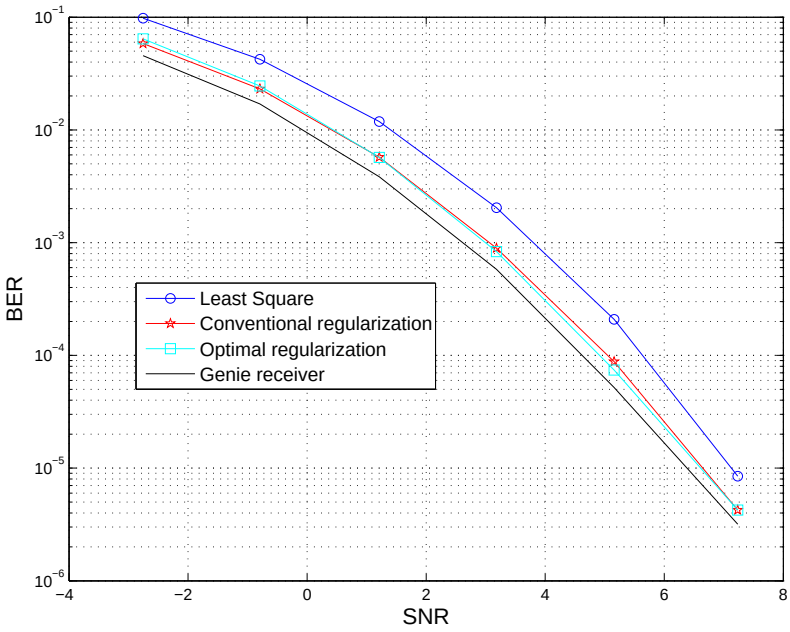


Figure 10.3: BER vs SNR with semi-blind regularization.

be the received signal before the Discrete Fourier Transform (DFT) module. (see fig. 10.4). Assuming that the channel length L is less than the cyclic prefix length μ, the linear convolution between the channel and the transmitted signal is transformed into a circular convolution as follows:

$$\mathbf{y}(n,k) = \sum_{l=0}^{L} \mathbf{H}(l)\mathbf{x}(n-l,k) + \mathbf{b}(n,k), \ n = 0, \cdots, Q-1 \tag{10.15}$$

where $\mathbf{H}(l)$ is the $N \times K$ channel response matrix at time l and $\mathbf{b}(n,k)$ is the noise vector. Moreover, if $n < 0$, due to the circular convolution, the value $\mathbf{x}(n,k)$ is set to $\mathbf{x}(Q+n,k)$. Let $\mathbf{H} = [\mathbf{H}^{\mathrm{T}}(0), \cdots, \mathbf{H}^{\mathrm{T}}(L)]^{\mathrm{T}}$ and $\mathbf{h} = \mathrm{vec}(\mathbf{H})$. Stacking $M+1$ observations $\mathbf{y}(n,k)$ in the $N(M+1)$ vector $\mathbf{y}_M(n,k) = [\mathbf{y}^{\mathrm{T}}(n,k), \cdots, \mathbf{y}^{\mathrm{T}}(n-M,k)]^{\mathrm{T}}$, we get for $M \leq n \leq Q-1$

$$\mathbf{y}_M(n,k) = \boldsymbol{\mathcal{I}}_M(\mathbf{H})\mathbf{x}_M(n,k) + \mathbf{b}_M(n,k) \tag{10.16}$$

where $\boldsymbol{\mathcal{I}}_M(\mathbf{H})$ is the $(M+1)N \times (M+L+1)K$ block Toeplitz matrix with the first block row given by $[\mathbf{H}(0), \cdots, \mathbf{H}(L), \mathbf{0}, \cdots, \mathbf{0}]$.

10.2.2 Blind criterion

Assuming that $\mathbb{E}|\mathbf{x}(n,k)|^2 = 1$ the covariance matrix of the received signal can be expressed as:

$$\mathbf{R}_M = \mathbb{E}\mathbf{y}_M(n,k)\mathbf{y}_M^{\mathrm{H}}(n,k) = \boldsymbol{\mathcal{I}}_M(\mathbf{H})\boldsymbol{\mathcal{I}}_M^{\mathrm{H}}(\mathbf{H}) + \sigma_b^2\mathbf{I}_{(M+1)N}$$

where σ_b^2 denotes the noise variance. In practice, it is estimated as:

$$\widehat{\mathbf{R}}_M = \frac{1}{(Q-M)n_g} \sum_{n=M}^{Q-1} \sum_{k=0}^{n_g-1} \mathbf{y}_M(n,k)\mathbf{y}_M^{\mathrm{H}}(n,k)$$

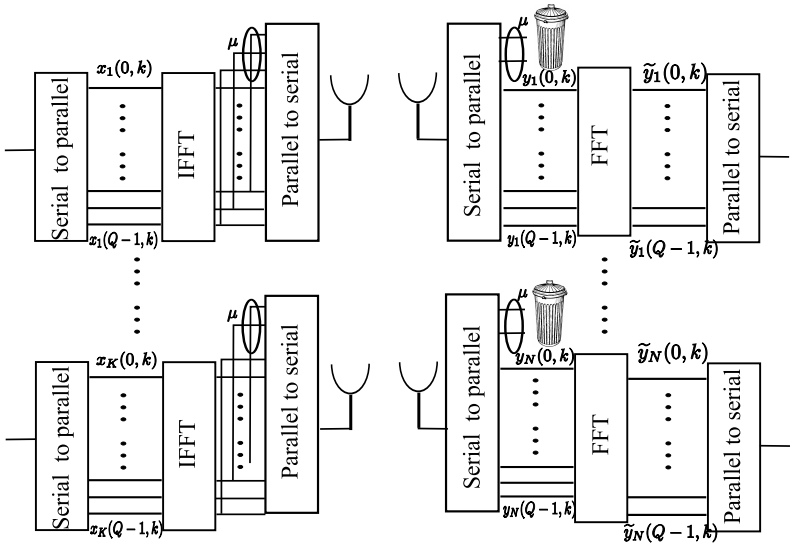


Figure 10.4: System Model

where n_g is the number of OFDM symbols in the considered frame.

In the same way as the case of SIMO systems, we will consider blind methods that are based on the minimization problem of a given quadratic form. In particular, we assume that the channel is evaluated up to a matrix ambiguity by solving:

$$\min_{\|\mathbf{h}\|=1} \mathbf{h}^{\mathrm{H}} \widehat{\mathbf{B}}_Q^{\mathrm{H}} \widehat{\mathbf{B}}_Q \mathbf{h}$$

where $\widehat{\mathbf{B}}_Q$ is an estimated matrix of $\mathbf{B}$, $\mathbf{B}$ being a matrix that depends on the considered blind technique.

10.2.3 Least square channel estimation

We assume that the channel estimation at the receiver side is conducted over n_g OFDM symbols, each OFDM symbol containing k_p pilot samples (i.e pilot subcarriers). Obviously, the least-square chanel estimation is possible only when $n_p = n_g k_p \geq (L+1)K$.

For MIMO-OFDM systems, optimal pilot sequences and optimal pilot placement of the pilot tones with respect to the MSE of the least-square channel estimate were derived in [4]. We design our pilot sequence according to [4] as follows:

Let $\widetilde{\mathbf{y}}_i(k)$ denotes the $k_p \times 1$ frequency domain vector of pilot samples received at time k by the i-th antenna and let $\widetilde{\mathbf{y}}_i = [\widetilde{\mathbf{y}}_i^{\mathrm{T}}(0), \cdots, \widetilde{\mathbf{y}}_i^{\mathrm{T}}(n_g-1)]^{\mathrm{T}}$. Then, $\widetilde{\mathbf{y}}_i$ satisfies:

$$\widetilde{\mathbf{y}}_i = \mathbf{A}\widetilde{\mathbf{h}}_i + \widetilde{\mathbf{b}}_i$$

where $\mathbf{A}$ is a $n_p \times (L+1)K$ matrix that depends on the pilot symbols and chosen to be orthogonal, $\mathbf{A}^{\mathrm{H}}\mathbf{A} = n_p \mathbf{I}_{(L+1)K}$, $\widetilde{\mathbf{b}}_i$ is the noise vector with respect to the i-th receiving antenna and $\widetilde{\mathbf{h}}_i$ is the channel vector response associated to the i-th receiving antenna given by:

$$\widetilde{\mathbf{h}}_i = [h_{i,1}(0), \cdots, h_{i,1}(L), \cdots, h_{i,K}(0), \cdots, h_{i,K}(L)]^{\mathrm{T}}$$

Defining:

$$\mathbf{Y} = [\widetilde{\mathbf{y}}_1, \cdots, \widetilde{\mathbf{y}}_N],$$
$$\widetilde{\mathbf{H}} = \left[\tilde{\mathbf{h}}_1, \cdots, \tilde{\mathbf{h}}_N\right],$$
$$\widetilde{\mathbf{V}} = [\widetilde{\mathbf{v}}_1, \cdots, \widetilde{\mathbf{v}}_N],$$

we have

$$\mathbf{Y} = \mathbf{A}\widetilde{\mathbf{H}} + \widetilde{\mathbf{V}}.$$

Let $\widetilde{\mathbf{h}} = \mathrm{vec}(\widetilde{\mathbf{H}})$ [4]. The least-square estimate minimizes the following criterion:

$$\min_{\tilde{\mathbf{h}}} \|\mathrm{vec}(\mathbf{Y}) - \mathbf{I}_N \otimes \mathbf{A}\tilde{\mathbf{h}}\|^2 = \min_{\mathbf{h}} \|\mathrm{vec}(\mathbf{Y}) - \widetilde{\mathbf{A}}\mathbf{h}\|^2, \tag{10.17}$$

where $\widetilde{\mathbf{A}} = (\mathbf{I}_N \otimes \mathbf{A})\,\mathbf{E}$ and $\mathbf{h} = \mathrm{vec}(\mathbf{H})$, $\mathbf{E}$ being the permutation matrix that transforms $\tilde{\mathbf{h}}$ into $\mathbf{h}$.

10.2.4 Matrix regularized channel estimation

Similarly to the case of SIMO systems, we consider to minimize with respect to $\mathbf{f}$ the following cost function:

$$C(\mathbf{f}, \Lambda) = \|\mathrm{vec}(\mathbf{Y} - \widetilde{\mathbf{A}}\mathbf{f}\|^2 + (Qn_g)\mathbf{f}^{\mathrm{H}}\widehat{\mathbf{P}}_Q\mathbf{\Lambda}\widehat{\mathbf{P}}_Q\mathbf{f}$$

where $\widehat{\mathbf{P}}_Q$ is the estimate of $\mathbf{P}$, the orthogonal projector on the space spanned by $\mathbf{B}_Q$ and $\mathbf{\Lambda}$ is a hermitian regularizing matrix. The semi-blind channel estimate is theorefore given by:

$$\widehat{\mathbf{h}}_{\mathbf{\Lambda}} = \left(\widetilde{\mathbf{A}}^{\mathrm{H}}\widetilde{\mathbf{A}} + Qn_g\widetilde{\mathbf{P}}_Q\mathbf{\Lambda}\widehat{\mathbf{P}}_Q\right)^{-1}\mathbf{A}^{\mathrm{H}}\mathrm{vec}(\mathbf{Y})$$

In the same way as previously mentioned, matrix $\mathbf{\Lambda}$ is chosen in such a way that it minimizes the asymptotic MSE. By asymptotic, we mean that Qn_g and n_P tend to infinity and $\frac{Kn_g}{n_P} \to \gamma$. Using the same approach as in the case of SIMO systems, it can be shown that the optimal regularizing matrix is given by:

$$\mathbf{\Lambda}_{\mathrm{op}} = \left(\mathbf{B}_Q^{\#}\mathbf{\Sigma}_\infty\mathbf{B}_Q^{\#}\right)^{\#} \tag{10.18}$$

where

$$\mathbf{\Sigma}_\infty = \frac{1}{\sigma_b^2}\lim_{\substack{Qn_g\to\infty \\ \frac{Qn_g}{n_p}\to\gamma}} Qn_g\mathrm{Cov}(\widehat{\mathbf{B}}_Q\mathbf{h}).$$

10.2.5 Semi-blind techniques based on Subspace and Linear prediction criteria

We consider the subspace and linear prediction techniques described in 6.2.1 and 6.2.2 in chapter 6. Keeping the same notations as in chapter 6, matrix $\mathbf{B}_Q$ is given by:

$$\begin{cases} \mathbf{B}_Q = \mathbf{B}_{\mathrm{sub}} = \mathbf{I}_K \otimes \boldsymbol{\mathcal{D}}(\mathbf{\Pi}_{\mathrm{sub}}) \text{ for the blind subspace criterion} & (10.19) \\ \mathbf{B}_Q = \mathbf{B}_{\mathrm{lin}} = \mathbf{I}_K \otimes (\mathbf{\Pi}_{\mathrm{lin}} \otimes \delta(\mathbf{P})) \text{ for the linear prediction criterion} & (10.20) \end{cases}$$

[4] Note that $\tilde{\mathbf{h}}$ and $\mathbf{h}$ are the same up to a permutation.

Let $\mathbf{P}_{\text{sub}}$ and $\mathbf{P}_{\text{lin}}$ denote respectively the orthogonal projectors onto the column range space of $\boldsymbol{\mathcal{D}}(\boldsymbol{\Pi}_{\text{sub}})$ and $\boldsymbol{\Pi}_{\text{lin}} \otimes \delta(\mathbf{P})$. Denoting by $\widehat{\mathbf{P}}_{\text{sub}}$ and $\widehat{\mathbf{P}}_{\text{lin}}$ their estimates, the semi-blind channel estimates based on the subspace and the linear prediction criteria are given by:

$$\begin{cases} \widehat{\mathbf{h}}_{\text{sub}} = \left(\widetilde{\mathbf{A}}^{\text{H}}\mathbf{A} + Qn_g\left(\mathbf{I}_K \otimes \widehat{\mathbf{P}}_{\text{sub}}\right)\boldsymbol{\Lambda}\left(\mathbf{I}_K \otimes \widehat{\mathbf{P}}_{\text{sub}}\right)\right)^{-1}\widetilde{\mathbf{A}}^{\text{H}}\text{vec}(\mathbf{Y}) & (10.21) \\ \widehat{\mathbf{h}}_{\text{lin}} = \left(\widetilde{\mathbf{A}}^{\text{H}}\mathbf{A} + Qn_g\left(\mathbf{I}_K \otimes \widehat{\mathbf{P}}_{\text{lin}}\right)\boldsymbol{\Lambda}\left(\mathbf{I}_K \otimes \widehat{\mathbf{P}}_{\text{lin}}\right)\right)^{-1}\widehat{\mathbf{A}}^{\text{H}}\text{vec}\mathbf{Y} & (10.22) \end{cases}$$

To determine the optimal values for $\boldsymbol{\Lambda}_{\text{lin}}$ and $\boldsymbol{\Lambda}_{\text{sub}}$, we need to compute $\boldsymbol{\Lambda}_\infty$ for both blind criteria. This can be carried out by following the same lines as in [65]. We obtain thus the following theorem:

Theorem 10.7. *The asymptotic covariance matrix of the channel estimation error $\delta\mathbf{h} = \widehat{\mathbf{h}} - \mathbf{h}$ is given by:*

$$\begin{aligned} \text{Cov}(\delta\mathbf{h}) &= \lim_{\substack{Qn_g\to\infty \\ \frac{Qn_g}{n_p}\to\gamma}} n_p\mathbb{E}\left(\delta\mathbf{h}\delta\mathbf{h}^{\text{H}}\right) \\ &= \sigma_b^2\mathbf{M}^{-1}\left(\mathbf{I}_{NK(L+1)} + \gamma\mathbf{C}\right)\mathbf{M}^{-1} \end{aligned}$$

where:

$$\begin{aligned} \mathbf{M} &= \left(\mathbf{I}_{(L+1)NK} + \gamma\left(\mathbf{I}_K \otimes \mathbf{P}\right)\boldsymbol{\Lambda}\left(\mathbf{I}_K \otimes \mathbf{P}\right)\right) \\ \mathbf{C} &= \left(\mathbf{I}_K \otimes \mathbf{P}\right)\boldsymbol{\Lambda}\mathbf{B}^{\#}\boldsymbol{\Sigma}_\infty\mathbf{B}^{\#}\boldsymbol{\Lambda}\left(\mathbf{I}_K \otimes \mathbf{P}\right) \end{aligned}$$

and $\mathbf{P} = \mathbf{P}_{\text{sub}}$ (resp. $\mathbf{P} = \mathbf{P}_{\text{lin}}$), $\mathbf{B} = \mathbf{B}_{\text{sub}}$ (resp. $\mathbf{B} = \mathbf{B}_{\text{lin}}$) and $\boldsymbol{\Sigma}_\infty = \boldsymbol{\Sigma}_{\infty,\text{sub}}$ (resp. $\boldsymbol{\Sigma}_\infty = \boldsymbol{\Sigma}_{\infty,\text{lin}}$) to refer to the subspace-based semi-blind estimator (resp. linear prediction-based semi-blind estimator).

Let $\mathbf{J}_p^\tau \triangleq \begin{bmatrix} \vdots & \mathbf{I}_{p-\tau} \\ \mathbf{0}_\tau & \cdots \end{bmatrix}$ for $\tau \geq 0$ and $\mathbf{J}_p^\tau \triangleq \left(\mathbf{J}_p^{-\tau}\right)^{\text{T}}$ if $\tau < 0$.

For the linear prediction based estimator, we can prove that $\boldsymbol{\Sigma}_\infty$ is given by:

$$\begin{aligned} \boldsymbol{\Sigma}_{\infty,\text{lin}} = \mathbf{I}_K \otimes \text{diag}\Big(&\overline{\mathbf{P}}_{\text{lin}}{\overline{\mathbf{P}}_{\text{lin}}}^{\text{H}}, \left(\mathbf{I}_{M+L-2} \otimes \overline{\mathbf{P}}_{\text{lin}}\right)\boldsymbol{\mathcal{J}}_{\text{lin}} \\ &\left(\mathbf{I}_{M+L-2} \otimes {\overline{\mathbf{P}}_{\text{lin}}}^{\text{H}}\right)\Big) + \mathcal{O}(\sigma^2) \end{aligned}$$

where:

$$\begin{aligned} \overline{\mathbf{P}}_{\text{lin}} &= \boldsymbol{\Pi}_{\text{lin}}\left[\mathbf{I}_N \quad \mathbf{P}\right] \\ \boldsymbol{\mathcal{J}}_{\text{lin}} &= \left(\sum_{k=-M}^{M}\mathbf{J}_{M+L-2}^k \otimes \mathbf{J}_{M+1}^k \otimes \mathbf{I}_N\right). \end{aligned}$$

For the subspace-based semi-blind estimator, $\boldsymbol{\Sigma}_\infty$ is given by:

$$\boldsymbol{\Sigma}_{\infty,\text{sub}} = \mathbf{I}_K \otimes \left(\sum_{k=-M}^{M}\mathbf{J}_{M+1}^k \otimes \boldsymbol{\Pi}_{\text{sub}}\left(\mathbf{J}_{M+1}^k \otimes \mathbf{I}_N\right)\boldsymbol{\Pi}_{\text{sub}}\right) + \mathcal{O}(\sigma^2).$$

Note that if we neglect the terms in $\mathcal{O}(\sigma^2)$, we end-up with expressions that are independent from the system parameters, $\mathbf{h}$ and σ^2. Hence, the implementation of the optimal weighting matrix in (10.18) can be achieved without resorting to iterative techniques.

10.2.6 Simulations

In all the simulations described below, we consider a MIMO-OFDM system with $K = 2$ transmit antennas and $N = 4$ receive antennas. The length of the cyclic prefix is 20, and the block size M is equal to 10. We assume also a Rayleigh channel model with a $L + 1 = 5$ tap MIMO-FIR filter where each tap is represented by a 2×4 random matrix whose elements are i.i.d complex Gaussian variables with zero mean and variance equal to $\frac{1}{L+1}$.

The empirical MSE is given by:

$$\text{MSE} = \frac{1}{N_T} \sum_{k=1}^{N_T} \|\widehat{\mathbf{h}}_k - \mathbf{h}_k\|^2,$$

where N_T is the total number of Monte Carlo iterations.

10.2.6.1 Accuracy of the channel estimation error

In this experiment, we set the number of the subcarriers Q to 2048, of which 16 subcarriers are devoted to training. We estimate empirically the MSE over $N_T = 100$ iterations. Fig. 10.5 displays in the same graph the empirical MSE as well as the theoretical MSE for the linear prediction and subspace-based semi-blind estimators. At low SNR, we observe that the theoretical MSE underestimates the real MSE since it does not take into account the term of order σ^2 in the expression of $\mathbf{\Sigma}_\infty$. However at moderate and high SNR, we obtain a good match between theoretical and empirical results.

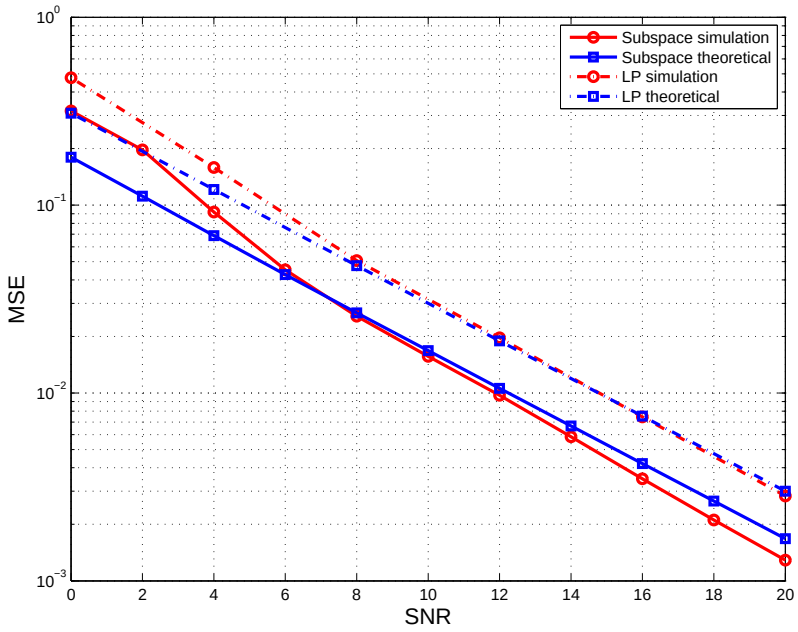


Figure 10.5: Theoretical and empirical MSE versus SNR

10.2.6.2 Impact of channel overmodeling

In this experiment, the channel estimation performance in terms of MSE is investigated. The simulation is undertaken based on 100 Monte Carlo runs of the transmission of $n_g = 4$ OFDM

symbols on $Q = 512$ subcarriers, of which $n_p = 16$ subcarriers are used for training. The channel is set to be the same for all the SNR range. We propose to investigate the impact of the over-modeling of the channel on the performance of the semi-blind estimators. We note that if the estimated channel order $\widehat{L}$ is equal to L, the semi-blind subspace estimator outperforms the linear prediction based estimator. But, once $\widehat{L}$ is greater than L, the linear prediction based estimator becomes better, as we can see on fig. 10.6.

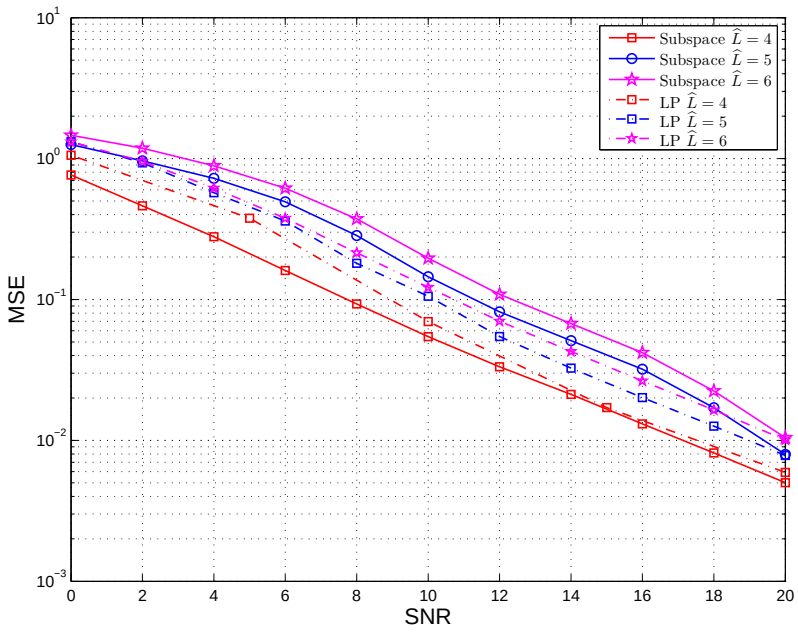


Figure 10.6: Impact of channel overmodeling.

10.2.6.3 Mean square error with respect to the number of pilots

In this section, we investigate the impact of the number of pilots on the mean square error. Fig. 10.7 compares the MSE of the least square channel estimator with that of the semi-blind estimators, when the number of carriers Q is set to 256 and the SNR is set to 10 dB. We also assume that the channel estimation is conducted over $n_g = 4$ OFDM symbols. We note that for equal MSE both the linear prediction and subspace estimator, employ at least 20 pilot symbols less than the least square estimator, thereby saving 5 subcarriers per OFDM symbol.

10.2.6.4 Semi-blind estimation with iterative decoding

At the receiver side, we perform iterative symbol detection and channel decoding. The receiver consists of a MIMO detector and a SISO channel decoder that exchange extrinsic soft information with each oter, so as to maximize the aposteriori probabilities. We set the number of iterations between the MIMO detector and the SISO decoder to 2, [100]. We assume that the binary information data are encoded by a rate 1/2 NRNSC code with constraint length set to 5 defined in octal form (037021). Throughout our simulations, we assume that each frame is composed of $n_g = 4$ OFDM symbols with $Q = 512$ subcarriers. We also consider $16 - QAM$ constellations with Gray labeling and we assume that the channel length has been over-estimated, $\widehat{L} = 6$. Fig.

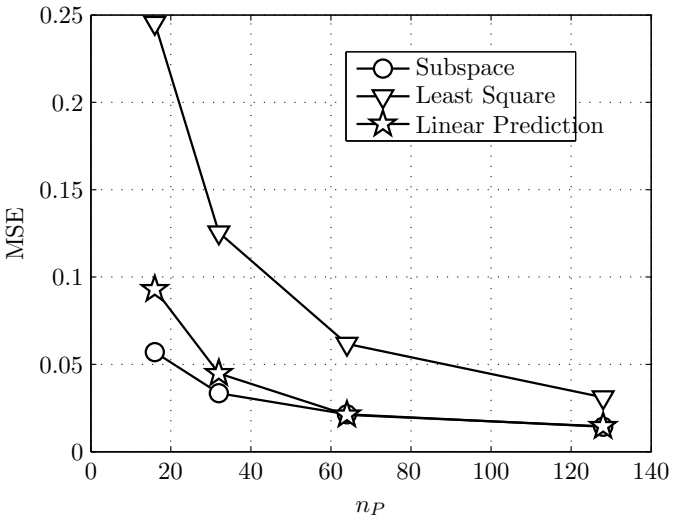


Figure 10.7: Impact of the number of pilots on the mean square error

10.8 displays the BER for the Least square and the two semi-blind based schemes. We note that both semi-blind schemes achieve almost the same performance and outperform the least-square based scheme by about 1dB.

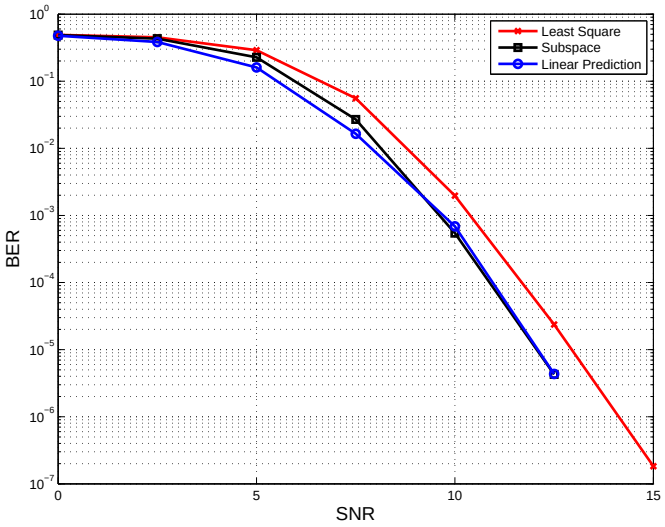


Figure 10.8: BER versus SNR

10.3 Conclusion

In this chapter, we have proposed a new regularization based method for semi-blind channel estimation. Our technique is based on the use of a regularizing matrix instead of a scalar, thus providing more degrees of freedom in the setting of the regularizing constant. Interestingly, we show that in this case, a closed-form expression for the optimal regularizing matrix exists,

thereby avoiding the need for iterative computations. Being initially proposed for SIMO systems, our method has been then extended to the MIMO-OFDM context. As an application, we have investigated the performance of subspace and linear prediction based estimators for which closed-form expressions for the regularizing matrix has been provided.

Conclusion

In this thesis, three principal issues have been investigated. The first one deals with performance analysis for training based systems while taking into account imperfect channel estimation. In particular, we have made a performance comparison in terms of bit error rate and outage probability between three training based schemes, namely, the time-division multiplexing scheme and the conventional and the data-dependent superimposed training schemes. While closed-form expressions for these performance indexes have been provided only in the asymptotic regime, the conducted simulations strongly support their accuracy for usual system dimensions.

Still about performance analysis issues, we have established in the second part the asymptotic normality of the SINR at the output of the LMMSE receiver, in the context of multi-user systems. Our results apply for a large variety of systems ranging from (receive) correlated MIMO systems to CDMA (uplink and downlink) based systems. This asymptotic analysis provides insights about the asymptotic behaviour of the SINR but is not sufficient to accurately approximate the BER and the outage probability. This can be principally attributed to the fact that the Gaussian distribution allows negative values and has a zero central third moment, while the SINR always takes positive values and thus has a non zero third moment. In light of this consideration, we have proposed to approximate the distribution of the SINR by the Generalized Gamma distribution and to tune its parameters in such a way that its first three moments equate the first three asymptotic moments of the SINR.

The third part of our report have dealt with blind and semi-blind techniques. It is common knowledge that most of the conventional blind methods yield a good performance as long as the channel order is correctly estimated. Otherwise, the promised performance gains could not be achieved. In this context, we have proposed a blind equalization technique that combines high performance with robustness towards channel order over-modeling. The proposed technique is the extension of an earlier method based on the truncation of the covariance matrix. Unlike the original approach, our technique can handle non-zero delay equalization, thereby providing higher performance while inheriting the same interesting properties regarding the robustness to channel order over-modeling.

On the other hand, the use of ℓ_p quasi-norms has been found to confer robustness to subspace blind techniques. Motivated by these interesting results, we have provided in our work theoretical arguments that strongly support the efficiency of the ℓ_p quasi-norms and provide useful insights on the impact of the number of sensors and the channel length on the identifiability condition.

The second issue dealt with in the third part of our report has concerned regularized semi-blind techniques. For these methods, the setting of the regularizing coefficient can considerably affect the channel estimation quality. As far as scalar regularization is considered, we have proved that the asymptotic MSE has a unique minimum, thereby showing the existence of only one optimal value for the regularizing coefficient. It should be noted that such result was only observed by simulations but to the best of our knowledge has not been so far established. As a matter of fact, the asymptotic MSE is not always convex and the eventual zeros of its

derivative do not have in general closed-form expressions. The quasi-convexity of the MSE has been conducted by using, in an original way, elementary tools of mathematical analysis.

The second problem that is faced in regularized semi-blind techniques is the optimization process of the asymptotic MSE. Actually, while our proof establishes the existence of a unique minimum, it does not tell us how to estimate it. In this regard, we propose to substitute the regularizing scalar by a regularizing matrix. Interestingly, the optimal regularizing matrix is unique and has a closed-form expression, thereby avoiding the need for iterative optimization algorithms. Being first proposed for SIMO systems with subspace blind criterion, the proposed matrix regularizing method has been extended in a second step to MIMO-OFDM systems using subspace and linear prediction criteria.

Since a PHD thesis opens more problems than it solves, we end up our report by presenting some perspectives. Like any other research work, our contribution does not pretend to be exhaustive in any way. Many questions are deliberately left without answer because of lack of time. Despite their importance, these questions are just briefly or not mentioned at all in the present report, but could be pursued as possible research directions. Hereafter, we provide more insights about these questions in the hope that they could serve to any interested reader.

1. First Part: In the first part of our report, the carried out performance analysis is only valid for i.i.d Rayleigh channels. Such assumption significantly simplifies calculations but hardly holds in practice. Extending our performance analysis to more complicated channel models could be pointed out as a possible research direction.

2. Second Part: When only the asymptotic normality of the SINR can be established, we show that accurate approximation of the BER and the outage probability can be obtained by using generalized gamma approximation. Our work has focused on receive correlated channels but it might be interesting to extend it to general channel profiles. Also, the bit error rate approximation can be derived by using the Taylor approximation as in the first part. Comparison between the two approaches (Taylor approximation and Generalized Gamma approximation) could be a topic for a future work.

3. Third part: Several possible research directions can be further pursued. As for our proposed robust blind MMSE method, future work should focus on implementing it using adaptive algorithms. In addition, It might be interesting to make comparisons with other robust blind techniques in terms of BER performance and computational complexity. Also, Many issues related to our work about the ℓ_p quasi-norms should be further investigated. Along our study, only the noiseless case has been considered. A more comprehensive work should take into account the experienced Gaussian noise, and also investigate the MSE in case the ℓ_p quasi-norm fails to correctly identify the channel. Although only non-sparse channels have been investigated, we believe that it is not so difficult to extend our work to the case of sparse channels.

   Concerning our contributions on regularized semi-blind techniques, two issues need to be further investigated.

   - During our simulations, we have noticed that for usual system dimensions (small available samples and relatively high window sizes), the estimation of some blind variables (like the noise projector) may not be accurate. We have also noted that a considerable gain can be achieved if these quantities are well estimated.

     A possible way to improve their estimation is to assume that the sample and the window sizes grow to infinity with a constant pace.

- The major effort in the field of regularized semi-blind techniques has been widely focused on estimating the channel. Semi-blind equalizer estimation techniques that estimate directly the equalizer in the same way as direct blind methods, have not been yet considered.

Bibliography

[1] A. Aissa-El-Bey and K. Abed-Meraim. Blind Identification of Sparse SIMO Channels using Maximum A Posteriori Approach. In *Proceedings of the EUropean SIgnal Processing COnference*, Lausanne Switzerland, August 2008.

[2] E. K. S. Au, C. Wang, S. Sfar, R. D. Murch, W. H. Mow, V. K. N. Lau, R. S. Cheng, and K. B. Letaief. Error Probability for MIMO Zero-Forcing Receiver with Adaptive Power Allocation in the Presence of Imperfect Channel State Information. *IEEE Trans. on Wireless Communications*, 6(4), April 2007.

[3] Z.D. Bai and J.W. Silverstein. No eigenvalues outside the support of the limiting spectral distribution of large dimensional sample covariance matrices. *Annals of Probability*, 26(1):316–345, 1998.

[4] I. Barhumi, G. Leus, and M. Moonen. Optimal Training Design for MIMO-OFDM Systems in Mobile Wireless Channels. *IEEE Trans. on Signal Processing*, 51(6), June 2003.

[5] J. Baxter, R. Jones, M. Lin, and J. Olsen. SLLN for Weighted Independent Identically Distributed Random Variables. *Journal of Theoretical Probability*, 1:165–181, 2004.

[6] N. C. Beaulieu. The Evaluation of Error Probabilities for Intersymbol and Cochannel Interference. *IEEE Trans. on Wireless Communications*, 39(12):1740–1749, December 1991.

[7] Albert Benveniste, Maurice Goursat, and Gabriel Ruget. Robust identification of a nonminimum phase system: Blind adjustment of a linear equalizer in data communications. *IEEE Trans. on Automatic Control*, AC-25(3):385–399, June 1980.

[8] L. Berriche and K. Abed-Meraim. Stochastic Maximum Likelihood Methods for Semi-Blind Channel Equalization. In *Proceedings of the IEEE Personal Indoor Mobile Radio Communications Conference*, Berlin Germany, September 2005.

[9] A. Aissa El Bey and K. Abed-Meraim. Blind SIMO Channel Identification using a Sparsity Criterion. In *Proceedings of the IEEE Workshop on Signal Processing Advances in Wireless Communications*, Recife Brazil, July 2008.

[10] R.J. Bhansali, L. Giraitis, and P.S. Kokoszka. Convergence of quadratic forms with non-vanishing diagonal. *Stat. Probab. Letters*, 77:726–734, 2007.

[11] P. Billingsley. *Probability and measure*. John Wiley, 3nd edition, 1995.

[12] P. Blomgren and T. F. Chan. Color TV: Total Variation Methods for Restoration of Vector-Valued Images. *IEEE Transactions on Image Processing*, 7:304–309, March 1998.

[13] P. Bohlin and M. Tapio. Performance Evaluation of MIMO Communication Systems based on Superimposed Pilots. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 425–428, May 2004.

[14] S. Boyd and L. Vandenberghe. *Convex Optimization*. Cambridge University Press, 2004.

[15] V. Buchoux, O. Cappé, E. Moulines, and A. Gorokhov. On the performance of semi-blind subspace-based channel estimation. *IEEE Trans. on Signal Processing*, 48(6):1750–1759, June 2000.

[16] V. Buchoux, E. Moulines, and O. Cappé. Estimation Semi-autodidacte par Méthode Sous-espace pour Systèmes de réception Multicapteurs. In *Proceedings of the GRETSI Conference on Signal Processing*, September 1999.

[17] R. W. Butler and A. T. A. Wood. Saddlepoint Approximation for Moment Generating Functions of Truncated Random Variables. *The Annals of Statistics*, 32(6):2712–2730, 2004.

[18] E. J. Candès and T. Tao. Decoding by Linear Programming. *IEEE Trans. on Information Theory*, 51(12), December 2005.

[19] E. J. Candès and M. B. Wakin. An Introduction to Compressive Sampling. *IEEE Signal Processing Magazine*, 25(2):21–30, 2008.

[20] J.K. Cavers. An Analysis of Pilot Symbol Assisted Modulation for Rayleigh Fading Channels. *IEEE Trans. on Vehicular Technology*, 40:686–693, November 1991.

[21] Rick Chartrand and Valentina Staneva. Restricted isometry properties and nonconvex compressive sensing. *Inverse Problems*, 24(035020):1–14, 2008.

[22] J.-M. Chaufray, W. Hachem, and Ph. Loubaton. Asymptotic Analysis of Optimum and Sub-Optimum CDMA Downlink MMSE Receivers. *IEEE Trans. on Information Theory*, 50(11):2620–2638, November 2004.

[23] J.M. Chaufray, W. Hachem, and Ph. Loubaton. Asymptotical Analysis of Optimum and Sub-Optimum CDMA Downlink MMSE Receivers. In *Proceedings of the IEEE International Symposium on Information Theory*, Lausanne, Switzerland, July 2002. can be downloaded at *http://www-syscom.univ-mlv.fr/˜loubaton/index.html*.

[24] F. Chen, S. Kwong, and C. Kok. Blind MMSE Equalization of FIR/IIR Channels Using Oversampling and Multichannel Linear Prediction. *ETRI Journal*, 31(2):162–172, April 2009.

[25] H. A. Cirpan and M. K. Tsatsanis. Stochastic Maximum-Likelihood Methods for Semi-Blind Channel Estimation . *IEEE Signal Processing Letters*, 5(1):21–24, January 1998.

[26] C. R. J. Clapham. *Introduction to Mathematical Analysis*. Routledge, 1973.

[27] M. Codray and P. Bohlin. Training Based Mimo Systems: Part I: Performance Comparison. *IEEE Trans. on Signal Processing*, 55(11):5464–5476, November 2007.

[28] ETSI Normalization Committee. Channel Models for HIPERLAN/2 in different indoor scenarios. Norme ETSI, document 3ERI085B, European Telecommunications Standards Institute, Sophia-Antipolis, Valbonne, France, 1998.

[29] D. Gesbert and P. Duhamel and S. Mayrargue. On-Line Blind Multichannel Equalization based on Mutually Referenced Filter. *IEEE Trans. on Signal Processing*, 45(9), September 2006.

[30] D. Guo and S. Verdu and L. K. Rasmussen. Asymptotic Normality of Linear Multiuser Receiver Outputs. *IEEE Trans. on Information Theory*, 48(12):3080–3095, December 2002.

[31] Shiu D.-S., G.J. Foschini, M.J. Gans, and J.M. Kahn. Fading correlation and its effect on the capacity of multielement antenna systems. *IEEE Trans. on Communications*, 48(3):502–513, March 2000.

[32] E. de Carvalho and D. T. M Slock. Semi-Blind Maximum-Likelihood Estimation with Gaussian Prior for the Symbols using Soft Decisions. In *Proceedings of the IEEE Vehicular Technology Conference*, Ottawa Canada, May 1998.

[33] Elisabeth de Carvalho. *Blind and Semi-Blind Multichannel Estimation and Equalization for Wireless Communications*. PhD thesis, ENST Paris, 1999.

[34] J. Dumont, W. Hachem, S. Lasaulce, P. Loubaton, and J. Najim. On the capacity achieving covariance matrix for rician mimo channels: an asymptotic approach. *IEEE Trans. on Information Theory*, 2008. Accpeted for publication in IEEE Trans. on Information Theory.

[35] J. Dumont, W. Hachem, and J. Najim. High SNR Approximations of the Capacity of MIMO Correlated Rician Channels: A Large System Approach. In *Proceedings of the IEEE International Symposium on Information Theory*, pages 536–540, Nice, France, June 2007.

[36] O. Fonseca and I. N. Psaromiligkos. Approximation of the Bit Error Rate of BPSK Transmissions over Frequency Selective Channels. In *IEEE International Conference on Wireless and Mobile Computing, Networking and Communications*, pages 217–222, Montréal Canada, August 2005.

[37] S. Foucart and M. J. Lai. Sparse Recovery with Pre-Gaussian Random Matrices. 2009.

[38] H. U. Gerber. From the generalized gamma to the generalized negative binomial distribution. *Insurance: Mathematics and Economics*, 10(4):303–309, 1992.

[39] M. Ghogho, D. McLernon, E. Alameda-Hernandez, and A. Swami. Channel Estimation and Symbol Detection for Block Transmission using Data-Dependent Superimposed Training. *IEEE Signal Processing Letters*, 12(3):226–229, March 2005.

[40] M. Ghogho and A. Swami. Channel Estimation for MIMO Using Data-Dependent Superimposed Training. In *42 rd Annual Allerton Conference on Communications, Control and Computing*, Illinois, USA, September 2004.

[41] M. Ghogho, T. Whitworth, A. Swami, and D. McLernon. Full-Rank and Rank-Deficient Precoding . *IEEE Trans. on Signal Processing*, 57(11):4433–4442, November 2009.

[42] G.B. Giannakis, Y. Imouye, and J. Mendel. Cumulant-Based Identification of Multichannel Moving Average Models. *IEEE Trans. on Automatic Control*, 34:783–787, 1989.

[43] G.B. Giannakis and J. Mendel. Identification of Non-Minimum Phase Systems Using Higher-Order Statistics. *IEEE Trans. on Acoustics, Speech and Signal Processing*, 37:360–377, 1989.

[44] A. C. Gilbert and J. A. Tropp. Applications of Sparse Approximation in Communications. In *Proceedings of the IEEE International Symposium on Information Theory*, pages 1000–1004, Adelaide Australia, September 2005.

[45] V. L. Girko. *Theory of Random Determinants*, volume 45 of *Mathematics and its Applications (Soviet Series)*. Kluwer Academic Publishers Group, Dordrecht, 1990.

[46] D. A. Gore, R. W. Heath, and A. J. Paulraj. Transmit Selection in Spatial Multiplexing Systems. *IEEE Communication Letters*, 6(11):491–493, November 2002.

[47] A. Gorokhov, M. Kristensson, and B. Ottersten. Robust Blind Second-Order Deconvolution. *IEEE Signal Processing Letters*, 6(1):13–16, January 1999.

[48] A. Gorokhov and P. Loubaton. Semi-blind second order identification of convolutive channels. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 5, pages 3905–3908, Munich, Germany, April 1997.

[49] I. S. Gradshteyn and I. M. Ryzhik. *Table Of Integrals, Series and Products*. Academic Press, 7th edition, 2007.

[50] R. Gribonval and K. Schnass. Dictionary Identifiability from Few Training Samples. In *Proceedings of the EUropean SIgnal Processing COnference*, Lausanne Switzerland, August 2008.

[51] R. Gribonval and K. Schnass. Dictionary Identification- Sparse Matrix-Factorisation via ℓ_1-Minimisation. *arxiv preprint 0904.4774*, April 2009.

[52] A. Guillonnet and O. Zeitouni. Concentration of the Spectral Measure for Large Matrices. *Electronic Communications in Probability*, pages 119–136, 2000.

[53] W. Hachem, O. Khorunzhiy, Ph. Loubaton, J. Najim, and L. Pastur. A New Approach for Capacity Analysis of Large Dimensional Multi-Antenna Channels. *IEEE Trans. on Information Theory*, 54(9), September 2008.

[54] W. Hachem, P. Loubaton, and J. Najim. Deterministic equivalents for certain functionals of large random matrices. *Ann. Appl. Probab.*, 17(3):875–930, 2007.

[55] W. Hachem, P. Loubaton, and J. Najim. A CLT for Information Theoretic Statistics of Gram Random Matrices With a Given Variance Profile. *Annals of Probability*, 18(6):2071–2130, 2008.

[56] B. Hassibi and B. M. Hochwald. How Much Training is Needed in Multiple-Antenna Wireless Links. *IEEE Trans. on Information Theory*, 49(4):951–963, April 2003.

[57] J. Havil. *Gamma: Exploring Euler's constant*. Princeton, NJ: Princeton University Press, 2003.

[58] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge Univ. Press, 1994.

[59] T. Ben Jabeur, K. Abed-meraim, and H. Boujemaa. Blind Channel Shortening in OFDM System using Nulltones and Cyclic Prefix. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 3041–3044, Las Vegas USA, March 2008.

[60] I. Kacha, K. Abed-Meraim, and A. Belouchrani. Fast Adaptive Blind MMSE Equalizer for Multichannel FIR Systems. *Eurasip Journal on Signal Processing*, 2006(Article ID 14827):1–17, 2006.

[61] I. Kacha, K. Abed-Meraim, and A. Belouchrani. A Low-Cost Adaptive Algorithm for Blind Equalization without Channel Order Estimation. *International Symposium on Communications, Control and Signal Processing*, pages 1384–1388, March 2008.

[62] T. Kailath. *Linear systems*. Prentice Hall, Englewood Cliffs, New Jersey, USA, 1980.

[63] A. Kammoun, M. Kharouf, W. Hachem, and J. Najim. A Central Limit Theorem for the SINR at the LMMSE estimator output for large dimensional signals. *IEEE Trans. on Information Theory*, 55, November 2009.

[64] J.H. Kotecha and A.M. Sayeed. Transmit signal design for optimal estimation of correlated MIMO channels. *IEEE Trans. on Signal Processing*, 52(2):546–557, February 2004.

[65] S. Lasaulce. *Channel Estimation and Multiuser Detection for TD-CDMA Systems*. PhD thesis, ENST Paris, 2001.

[66] S. Lasaulce, P. Loubaton, and E. Moulines. A Semi-Blind Channel Estimation Technique Based on Second Order Blind Method for CDMA Systems. *IEEE Trans. on Signal Processing*, 51(7):1894–1903, 2003.

[67] M. Ledoux and M. Talagrand. *Probability in Banach Spaces. Isoperimetry and processes*. Springer-Verlag, 1991.

[68] X. Ma, L. Yang, and G. B. Giannakis. Optimal Training for MIMO Frequency-Selective Fading Channels. *IEEE Trans. on Wireless Communications*, 4(2):453–466, March 2005.

[69] J. R. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley, 3rd edition, 2007.

[70] J. H. Manton, I. V. Mareels, and Y. Hua. Affine precoders for reliable communication. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 2749–2752, June 2000.

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

[72] K. Meraim, P. Duhamel, D. Gesbert, P. Loubaton, S. Mayrargue, E. Moulines, and D. Slock. Prediction Error Methods for Time-Domain Blind Identification of Multichannel FIR Filters. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 2, pages 1968–1971, Detroit, USA, May 1995.

[73] K. Meraim, E. Moulines, and P.Loubaton. Prediction Error Methods for Second-Order Blind Identification. *IEEE Trans. on Signal Processing*, 45(3):694–705, March 1997.

[74] K. Abed Meraim, Ph. Loubaton, and E. Moulines. A subspace algorithm for certain blind identification problems. *IEEE Trans. on Information Theory*, 43(3), March 1997.

[75] Eric Moulines, Pierre Duhamel, Jean-François Cardoso, and Sylvie Mayrargue. Subspace Methods for the Blind Identification of Multichannel FIR Filters. *IEEE Trans. on Signal Processing*, 43(2):516–525, February 1995.

[76] A.L. Moustakas and S.H. Simon. On the Outage Capacity of Correlated Multiple-Path MIMO Channels. *IEEE Trans. on Information Theory*, 53(11), November 2007.

[77] A.L. Moustakas, S.H. Simon, and A.M. Sengupta. MIMO capacity through correlated channels in the presence of correlated interferers and noise: A (not so) large N analysis. *IEEE Trans. on Information Theory*, 49(10):2545–2561, October 2003.

[78] J. Murphy. Binary Error Rate Caused by Intersymbol Interference and Gaussian Noise. *IEEE Trans. on Wireless Communications*, 21(9):1039–1046, September 1973.

[79] N.K. Nikolski. *Operators, Functions and Systems: An Easy Reading. Vol. 2: Model Operators and Systems.* Mathematical Surveys and Monographs. AMS, 2002.

[80] S. Ohno and G. B. Giannakis. Superimposed Training on Redundant Precoding for Low-complexity Recovery of Block Transmissions. In *Proceedings of the IEEE International Conference on Communications*, Helsinki, Finland, June 1001.

[81] S. Ohno and G. B. Giannakis. Optimal Training and Redundant Precoding for Block Transmissions with Applications to Wireless OFDM. *IEEE Trans. on Communications*, 50(12):2113–2123, December 2002.

[82] P. Li and D. Paul and R. Narasimhan and J. Cioffi. On the Distribution of SINR for the MMSE MIMO Receiver and Performance Analysis. *IEEE Trans. on Information Theory*, 52(1):271–286, January 2006.

[83] G.-M. Pan, M.-H Guo, and W. Zhou. Asymptotic distributions of the Signal-to-Interference Ratios of LMMSE detection in multiuser communications. *Ann. Appl. Probab.*, 17(1):181–206, 2007.

[84] C. Papadias. *Methods for Blind Equalization and Identification Of Linear Channels.* PhD thesis, ENST Paris, March 1995.

[85] D. H. Pham and J. Manton. Orthogonal Superimposed Training on Linear Precoding: A New Affine Precoder Design. In *Proceedings of the IEEE Workshop on Signal Processing Advances in Wireless Communications*, New York City USA, June 2005.

[86] H. Poor and S. Verdú. Probability of Error in MMSE Multiuser Detection. *IEEE Trans. on Information Theory*, 43(3):858–871, May 1997.

[87] J. G. Proakis. *Digital Communications.* Mc Graw Hill, New York, USA, third edition, 1995.

[88] V. Radinov and S. Mayrargue. Semi-blind Approach to Second Order Identification of SIMO-FIR Channel Driven by Finite-Alphabet Sequence. In *Proceedings of the International Conference on Digital Signal Processing*, volume 1, pages 115–118, Santorini Greece, July 1997.

[89] S. Redl, M. K. Weber, and M. Oliphant. *An Introduction to GSM*. Boston: Artech House, first edition, 1995.

[90] R. Remmert. *Theory of Complex Functions*, volume 22 of *Graduate Texts in Math.* Springer-Verlag, 1991.

[91] Y. Sato. A Method of Self-Recovering Equalization for Multi-Level Amplitude Modulation. *IEEE Trans. on Communications*, 6:679–682, June 1975.

[92] J. W. Silverstein. The smallest Eigenvalue of a Large Dimensional Wishart Matrix. *Annals of Probability*, 13(4):1364–1368, 1985.

[93] J.W. Silverstein and Z.D. Bai. On the empirical distribution of eigenvalues of a class of large dimensional random matrices. *J. Multivariate Anal.*, 54(2):175–192, 1995.

[94] Marvin K. Simon and Mohamed-Slim Alouini. *Digital Communication over Fading Channels*. John Wiley, New Jersey, 2004.

[95] D. T. M. Slock. Blind Fractionally-Spaced Equalization, Perfect-Reconstruction Filter Banks and Multichannel Linear Prediction. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, volume 4, pages 585–588, Adelaid, Australia, April 1994.

[96] G. W. Stewart and J. G. Sun. *Matrix Perturbation Theory*. Academic Press, 1990.

[97] G. Taricco. Asymptotic Mutual Information Statistics of Separately Correlated Rician Fading MIMO Channels. *IEEE Trans. on Information Theory*, 54(8):3490–3504, August 2008.

[98] I.E. Telatar. Capacity of Multi-Antenna Gaussian Channels. published in *European Transactions on Telecommunications*, 10(6):585–595, Nov/Dec 1999. Technical Memorandum, Bell Laboratories, Lucent Technologies, October 1995.

[99] E. Thomas and L. B. Milstein. Coherent DS-CDMA Performance in Nakagami Multipath Fading. *IEEE Trans. on Communications*, 43(2/3/4):1134–1143, 1995.

[100] A. M. Tonello. Space-Time Bit-Interleaved Coded Modulation with an Iterative Decoding Strategy. *Proceedings of the IEEE Vehicular Technology Conference*, 1:473–478, 2000.

[101] L. Tong, G. Xu, and T. Kailath. Blind Identification and Equalization Based on Second-Order Statistics: A Time Domain Approach. *IEEE Trans. on Information Theory*, 40(2):340–349, March 1994.

[102] N. N. Tran, D. H. Pham, H. D. Tuan, and H. H. Nguyen. Orthogonal Affine Precoding and Decoding for Channel Estimation and Source Detection in MIMO Frequency-Selective Fading Channels. *IEEE Trans. on Signal Processing*, 57(3):1151–1162, March 2009.

[103] David N. C. Tse and Ofer Zeitouni. Linear multiuser receivers in random environments. *IEEE Trans. on Information Theory*, 46(1):171–188, January 2000.

[104] A.M. Tulino, L. Li, and S. Verdú. Spectral Efficiency of Multicarrier CDMA. *IEEE Trans. on Information Theory*, 51(2):479–505, February 2005.

[105] Antonia M. Tulino and Sergio Verdú. *Random Matrix Theory and Wireless Communications*. Now, 2004.

[106] J. Valyon and G. Horvath. Extended Least-Squares LS-SVM. *International Journal of Computational Intelligence*, 3(3):234–242, 2007.

[107] H. Vikalo, B. Hassibi, B. Hochwald, and T. Kailath. On the Capacity of Frequency-Selective Channels in Training-based Transmission Schemes. *IEEE Trans. on Signal Processing*, 52(9):2572–2583, September 2004.

[108] A. Vosoughi and A. Scaglione. Channel Estimation for Precoded MIMO System. In *Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing*, pages 409–412, Montreal, Quebec, Canada, 2004.

[109] A. Vosoughi and A. Scaglione. Everything you Always Wanted to Know about Training: Guidelines Derived using the Affine Precoding Framework and CRB. *IEEE Trans. on Signal Processing*, 54(3):940–954, March 2006.

[110] A. Vosoughi and A. Scaglione. On the Effect of Receiver Estimation Error upon Channel Mutual Information. *IEEE Trans. on Signal Processing*, 54(2):459–472, February 2006.

[111] F. Wan, W.-P. Zhu, and M. N. S. Swamy. A Semi-Blind Channel Estimation Approach for MIMO-OFDM Systems. *IEEE Trans. on Signal Processing*, 56(7):2821–2834, July 2008.

[112] C. Wang, E. K. S. Au, R. D. Murch, W. H. Mow, R. S. Cheng, and V. Lau. On the Performance of the MIMO Zero-Forcing Receiver in the Presence of Channel Estimation Error. *IEEE Trans. on Wireless Communications*, 6(3), March 2007.

[113] J. Winters, J. Salz, and R. Gitlin. The Impact of Antenna Diversity on the Capacity of Wireless Communication Systems. *IEEE Trans. on Communications*, 42:1740–1751, February 1994.

[114] X. Li and H. Fan. Linear Prediction Methods for Blind Fractionally Spaced Equalization. *IEEE Trans. on Signal Processing*, 48(6):1667–1675, June 2000.

[115] R. Xu and F. C. M Lau. Performance analysis for MIMO systems using zero forcing detector over Rice fading channel. In *Proceedings of the IEEE International Symposium on Circuits and Systems*, volume 5, pages 4955–4958, Kobe, Japan, May 2005.