POLITECNICO DI TORINO

Electronics and Telecommunications Department
Master of Science in Communications and Computer Networks Engineering

Master Thesis

Exploiting Channel Spatial Correlation in FDD Massive MIMO Systems

A Novel Enabling Technique

Host Institution:
Fraunhofer Heinrich Hertz Institute

Industrial Supervisors:
Dr. Renato Luis Garrido Cavalcante
Prof. Sławomir Stańczak

Academic Supervisors:
Prof. Roberto Garello
Prof. Monica Visintin

Candidate:
Lorenzo Miretti

April 2018

Contents

Chapter 1

Introduction

Multiple-input and multiple-output (MIMO) is a mature technology currently implemented in many modern broadband wireless communication standards, like in the fourth generation (4G) systems. By virtue of multiple antennas at the transmitters and receivers, such systems can achieve significant gains in terms of data rate and link reliability. These gains are typically quantified by the so called *multiplexing gain* and *diversity gain*, that in ideal conditions scale, respectively, as a function of the signal-to-noise ratio (SNR) according to $\min(N, M)\log(1 + \mathrm{SNR})$ and SNR^{NM}, where N and M denote the number of transmitting and receiving antennas. These benefits extends also to multi-user (MU) systems, where the MIMO theory enables advanced techniques for interference management, multiplexing and multiple access by adding a third *spatial* dimension to the classical time and frequency division techniques [1].

Motivated by the increased demand in data traffic envisioned for the fifth generation (5G) and beyond-5G wireless networks, in recent years a lot attention has been given by both the academic and industrial world to systems with antenna arrays equipped with a very large number of antennas, also called massive MIMO systems [2].

On top of the evident benefits in terms of multiplexing and diversity gains, the adoption of large-scale antenna arrays introduces additional properties that are deeply rooted in the law of large numbers and in the asymptotics of random matrix theory. Among them, a particularly interesting effect for MU-MIMO systems is that the channel vectors describing the propagation between single-antenna terminals and the base station (BTS) becomes asymptotically orthogonal as the number of antennas at the base station grows large. This fact has several advantages in terms of system performance, as it allows for multiple users to be jointly treated as if they belong to separated, non-interfering, single-user (SU) links, and thus classical linear techniques for SU-MIMO (e.g. maximum-ratio-transmission precoding) become optimal. Furthermore, another interesting effect of such large-scale systems is that the singular values distribution of the channel matrix grouping the users channel vectors becomes a deterministic function. Moreover, especially for the uplink (UL) transmission, the signal degradation due to thermal noise can be better

averaged out. In general, the benefits of massive MIMO can be summarized by observing that the performance of such systems depends less on the instantaneous realizations of statistical quantities and more on aggregated properties of the propagation.

Not surprisingly, all these advantages come with a price. For example, an immediate practical issue of large-scale arrays is the increased hardware complexity given by the large number of radio-frequency (RF) chains to be deployed. Furthermore, the more complex signal processing implies additional power consumption and need for higher computational capabilities at both ends. Moreover, as for the vast majority of MIMO system, the benefits promised by the massive MIMO technology rely on coherent transmission, which means that the system must acquire a sufficiently reliable estimate of the channel state (CSI). However, the acquisition of the CSI in the large-scale array regime is far from trivial, especially for the downlink (DL) channel, and it is one of the main limiting factors of Massive MIMO systems performance [3, 4].

This thesis focuses on the DL CSI acquisition problem in massive MIMO systems. In the following sections of this chapter we present a brief outline of the problem and of the main contributions given by this thesis.

1.1 Problem Overview

In massive MIMO systems conventional DL channel estimation techniques (e.g., those currently implement in 4G systems) require prohibitively large pilot sequences. In extreme cases, the training time can easily exceed the channel coherence time, making CSI acquisition practically impossible [2]. The main reason is that pilot sequences are typically designed to have a length comparable to the number of antennas of the BTS.

This pilot design choice is one of the main reasons why existing solutions for massive MIMO systems are typically based on the time division duplexing (TDD) mode. In the TDD mode, the small number of antennas at the terminals enables the acquisition of a reliable estimate of the UL channel within the channel coherence time, and the DL channel becomes immediately available to the BTS because of the phenomenon of channel reciprocity. However, for frequency division duplexing (FDD) systems, this reciprocity is not available. In this case, the available solutions for CSI acquisition typically rely on the existence of a lower dimensional representation of the channel vector in the large-scale array regime.

Current approaches for DL CSI acquisition in FDD systems can be divided into two main categories: methods based on compressed sensing (CS) and methods based on second-order statistics. Although promising, CS techniques (e.g. [5]) do not take into account the space-time correlation of the channel, which is often modeled by the well-known wide sense stationary (WSS) assumption. This correlation is exploited by approaches based on second order statistics, and they have been shown to reduce effectively the effort for DL CSI acquisition [6–8]. In this study we focus on this last category of algorithms.

1.2 Main Contributions

In this thesis we analyze the problem of DL CSI acquisition for massive MIMO systems operating in the frequency division duplexing (FDD) mode. We highlight the limitations of the classical pilot based channel estimation designs, currently implemented in the modern communication standards, when scaled-up to the large-scale array regime. We also point out how it is possible to effectively improve the performance of such systems by taking into account more advanced channel estimation techniques based on channel correlation properties. We finally propose, as a main contribution of this work, a novel technique for estimating the downlink (DL) channel spatial covariance matrix $\mathbf{R}^d$ from the uplink (UL) channel spatial covariance matrix $\mathbf{R}^u$, as a key enabling technology for a wide category of modern CSI acquisition algorithms that exploit channel correlation.

The main contributions given by this thesis can be summarized as follows:

- We start with Chapter 2 by presenting many approaches to channel modeling, with particular emphasis on the characterization of the spatial covariance matrix, which we show to be crucial for the development of the CSI acquisition algorithms proposed later in the thesis. A particularly interesting expression for the spatial covariance matrix is derived by taking into account also polarization effects of the propagation environment. Such an expression is not available in the literature.
- In Chapter 3 we review conventional CSI acquisition techniques, by giving insights on their fundamental performance limits and potentials. In particular, we highlight the advantages of covariance-aware systems when the channel fading is correlated, and how the knowledge of the spatial channel covariance matrix at the BTS can help in reducing the training overhead. Furthermore, we briefly outline possible connections between CS based channel estimation techniques and channel correlation. Finally, we review and partially extend popular Information Theory arguments that show that the channel coherence time is one of the main performance bottlenecks of massive MIMO systems, especially in case of independent fading. Overall, Chapter 3 aims at giving insights on the key role played by channel correlation in addressing the FDD massive MIMO DL CSI acquisition problem.
- Chapter 4 proposes novel algorithms for DL covariance estimation based on only UL measurements. The benefits of this approach are discussed. We show both theoretically and numerically how the proposed algorithms can be applied to realistic communication systems. In particular, unlike most current work in the literature, the proposed solutions take into account real-world 3D propagation environments and dual polarized antenna arrays. Furthermore, we show via simulations that our algorithms outperform current state-of-the art solutions in terms of accuracy and complexity, under different performance metrics.

1.3 Notation and Abbreviations

This list gives a brief overview of the main mathematical notation and abbreviations adopted in the text. Due to the large number of quantities occurring, some parts of the text may present a local notation not presented in this general list. In those cases, the reader should refer to the local explanation given in the specific section.

Notation:

- We adopt lower case boldface letters ($\mathbf{a}$) for vectors, with ith element denoted as a_i.
- We adopt upper case, boldface letters ($\mathbf{A}$) for matrices, with ijth entry denoted as a_{ij}.
- $(\cdot)^*$, $(\cdot)^T$, $(\cdot)^H$ and $(\cdot)^\dagger$ denote respectively the conjugate, the transpose, the Hermitian transpose, and the Moore–Penrose pseudo inverse operations.
- $\|\cdot\|_p$ and $\|\cdot\|_F$ denote respectively the p-norm over a generic L^p space and the Frobenius norm.
- $\Re[\cdot]$ and $\Im[\cdot]$ denote respectively the real and the imaginary parts.
- $\mathrm{tr}\{\cdot\}$ denotes the trace operator.
- $\mathrm{rank}\{\cdot\}$, $\mathrm{span}\{\cdot\}$, $\mathrm{Ker}\{\cdot\}$ denote respectively the rank, the column space, and the kernel of a given matrix.
- $\mathrm{diag}(\mathbf{a})$ denotes a diagonal matrix which diagonal elements are the elements of $\mathbf{a}$.
- $\mathrm{vec}(\mathbf{A})$ denotes a column vector obtained by stacking the columns of the matrix $\mathbf{A}$.
- $L^2[I]$ denotes the set of all square Lebesgue integrable functions over the domain $I \subset \mathbb{R}^n$.
- Given a Hilbert space, we denote by $x^{(i)} \rightharpoonup x$ a sequence $(x^{(i)})_{i\in\mathbb{N}}$ weakly convergent to a point x.
- Superscripts $(\cdot)^u$ and $(\cdot)^d$ indicate respectively UL and DL matrices, vectors, or functions when we need to emphasize the dependency on the carrier frequency.
- $\mathbb{E}[\cdot]$ denotes the expectation operation over a random quantity.

Acronyms:

3GPP	Third Generation Partnership Project.
4G	Fourth Generation.
5G	Fifth Generation.
APS	Angular Power Spectrum.
BTS	Base Transceiver Station.
CDF	Cumulative Density Function.
CS	Compressed Sensing.
CSI	Channel State Information.
DL	Downlink.
DoA	Direction of Arrival.
DoD	Direction of Departure.
DoF	Degrees of Freedom.
EAPM	Extrapolated Alternating Projection Method.
FDD	Frequency Division Duplexing mode.
GCS	Global Coordinate System.
GCSM	Geometric-based Stochastic Channel Model.
H-APS	APS for the Horizontal polarization.
LCS	Local Coordinate System.
LMMSE	Linear Minimum MSE estimator.
LS	Least Square estimator.
LOS	Line Of Sight path.
MIMO	Multiple Input Multiple Output system.
ML	Maximum Likelihood.
MRC	Maximum Ratio Combining.
MSE	Mean Square Error.
MU-MIMO	Multi-User MIMO.
NLOS	Non-LOS path.
NN	Neural Network.
OFDM	Orthogonal Frequency Division Multiplexing.
PDF	Probability Density Function.
PSD	Positive Semi-Definite matrix.
SE	Square Error.
SNR	Signal to Noise Ratio.
SU-MIMO	Single-User MIMO.
TDD	Time Division Duplexing mode.
UE	User Equipment.
UL	Uplink.
ULA	Uniform Linear Array.
UPA	Uniform Planar Array.
V-APS	APS for the Vertical polarization.
WSS	Wide-Sense Stationary process.
XPR	Cross Polarization Ratio

Chapter 2

Channel Models

The goal of this chapter is to give a comprehensive analysis of the channel models adopted for the design of the algorithms presented throughout this study. Particular emphasis is put into the description of the second order statistics of the channel, in the form of spatial covariance matrix, because of its crucial role played in the invention presented in this study, and, more generally, in the problem of massive MIMO channel estimation. To widen as much as possible the domain of application of the proposed algorithms, we devote the core part of this chapter to the derivation of the expression of the spatial channel covariance matrix for several popular directional channel models. These models are presented in a bottom-up fashion, by treating in a constructive way different aspects with increasing complexity. The results are then merged to provide a unified description of the spatial channel covariance matrix, compliant with the design requirements of modern 4G and 5G systems. This hierarchical approach allows to highlight better specific characteristics of different system design choices and modeling philosophies. In brief, this chapter is structured as follows:

- In the introductory part, we outline the general system model, and we give a high-level overview of the main assumptions.
- As a first example, we analyze a simple narrow-band system in a 2D environment by using two different popular channel modeling philosophies, denoted here as "discrete scattering" and "continuous scattering" models.
- Secondly, we consider a wide-band OFDM system in a 2D environment, showing an interesting parallelism of the results derived for narrow-band systems.
- Thirdly, these models are extended to consider also 3D environments, polarization effects, non-isotropic antennas, and multi-antenna user equipment.
- Finally, the complete covariance models are summarized and discussed.

A rigorous proof of part of the propositions in this chapter requires results from measure theory. However, the main focus of this study is on the algorithms for channel estimation

and not on the channel models, so, for the sake of simplicity, we keep the proofs informal by making extensive use of the heuristic characterization of the generalized Dirac delta function δ, defined such that the following property holds:

$$\int_{-\infty}^{+\infty} \delta(x - x_0) f(x)\, dx = f(x_0)$$

where $f : \mathbb{R} \to \mathbb{C}$ is a continuous function and $x_0 \in \mathbb{R}$. We point out that δ must not be understood as a function, but as a shorthand that only acquires meaning when evaluated inside an integral operation.

2.1 General System Model

Let us consider a multi-user (MU) MIMO channel between a base station (BTS) with N antennas and K single-antenna user equipments (UE). Most of the work of this study is based on this scenario.

Assuming a flat-fading channel model, a typical discrete input-output relation describing the UL communication between the BTS and a UE is given by [1]

$$\mathbf{y}[m] = \mathbf{h}^u[m]x[m] + \bar{\mathbf{H}}^u[m]\bar{\mathbf{x}}[m] + \mathbf{w}[m], \tag{2.1}$$

where $m \in \mathbb{Z}$ is the discrete time index, $\mathbf{h}^u[m] \in \mathbb{C}^{N\times 1}$ is the UL time-variant channel vector between the UE and the BTS, $x[m] \in \mathbb{C}$ is the UE transmitted sequence, $\bar{\mathbf{H}}^u[m] \in \mathbb{C}^{N\times K-1}$ is the UL time-variant matrix collecting the channel vectors between the BTS and the interfering UEs, $\bar{\mathbf{x}}[m] \in \mathbb{C}^{K-1\times 1}$ is the vector collecting the sequences from the interfering UEs, $\mathbf{y}[m] \in \mathbb{C}^{N\times 1}$ is the received signal, and $\mathbf{w}[m] \sim \mathcal{CN}(0, \sigma_W^2 I_N)$ is a sample of the received white noise process. Similarly, a typical discrete input-output relation describing the DL communication between the UE and the BTS is given by

$$y[m] = \left(\mathbf{h}^d[m]\right)^T \mathbf{x}[m] + w[m], \tag{2.2}$$

where $\mathbf{h}^d[m] \in \mathbb{C}^{N\times 1}$ is the DL time-variant channel vector between the BTS and the UE, $\mathbf{x}[m] \in \mathbb{C}^{N\times 1}$ is the BTS transmitted sequence, $\mathbf{y}[m] \in \mathbb{C}$ is the received signal, and $w[m] \sim \mathcal{CN}(0, \sigma_W^2)$ is a sample of the received white noise process.

Although introduced for narrow-band communication systems, the flat-fading model can also be applied to describe the communication over a time-frequency slot in multi-carrier systems; e.g., orthogonal frequency division multiplexing (OFDM) based systems [1, Chapter 3.4].

2.1.1 Second-order Statistics

A classical channel model (e.g, see [9]) assumes the channel vector $\mathbf{h}[m]$ to be a zero-mean WSS process that is correlated both in time and in the spatial domain; i.e., with

spatio-temporal covariance matrix

$$\mathbf{R}_h[m,n] = \mathbb{E}\left[\mathbf{h}[m]\mathbf{h}[n]^H\right] = \mathbf{R}_h[m-n]. \tag{2.3}$$

An interesting condensed parameter that can be extracted from the spatio-temporal covariance matrix is the coherence interval M_c, which is defined as the time interval $m-n$ needed so that two channel vectors $\mathbf{h}[m]$ and $\mathbf{h}[n]$ can be considered uncorrelated. In many practical applications (for example, in the design of OFDM systems) the time correlation of the channel is often simplified by using the "block-fading" assumption, which assume the channel to be constant for an entire time frame of duration M_c [1]. In narrow-band systems, M_c is usually set equal to the channel coherence time T_c, expressed in discrete time. In wide-band OFDM systems instead, it is usually given by the product $M_c = T_c B_c$, where B_c denotes the coherence bandwidth.

The core part of this work is based on the properties of the spatial channel covariance matrix $\mathbf{R} := \mathbf{R}_h[0]$. To exploit spatial diversity, MIMO systems are typically designed with an inter-antenna spacing sufficiently large so that the antennas can be considered uncorrelated [9]. The minimum spacing is often assumed to be equal to the coherence length $d := \lambda/2$, where λ denotes the carrier wavelength. However, the coherence length is derived by assuming the "one-ring" model [9], which often does not correspond to real scenarios. Therefore, as in 3GPP models, in this work we take into account the spatial correlation given by real scattering environments.

According to the WSS assumption, $\mathbf{R}_h$ (and thus $\mathbf{R}$) is invariant over time. In practice, it is a slowly-varying quantity: channels can be safely assumed to be WSS just over a certain window of time T_{WSS}, which in usual scenarios is several order of magnitude larger than the channel coherence time T_c [8, 10]. In this work we assume T_{WSS} to be large enough for the time scale requirements of the proposed algorithms.

2.1.2 Channel Reciprocity

In general, the channel vector $\mathbf{h}[m]$ depends on the carrier frequency. In time division duplexing (TDD) systems, since the UL and the DL channels share the same carrier frequency, it is typically assumed that $\mathbf{h}^u[m] \equiv \mathbf{h}^d[m]$. In contrast, in frequency division duplexing (FDD) systems, where the UL and the DL channels are typically separated by 10-100 MHz, this assumption cannot be made.

However, by representing the channel with directional channel models, a weaker form of channel reciprocity in the angular domain exists. For example, by considering a simple 2D model (more general models are described later), we can relate the UL and DL channel covariance matrices by

$$\mathbf{R}^u = \int_{-\pi}^{\pi} \rho(\theta)\mathbf{a}^u(\theta)\mathbf{a}^u(\theta)^H d\theta, \tag{2.4}$$

$$\mathbf{R}^d = \int_{-\pi}^{\pi} \rho(\theta)\mathbf{a}^d(\theta)\mathbf{a}^d(\theta)^H d\theta, \tag{2.5}$$

where $\mathbf{R}^u$ and $\mathbf{R}^d$ are given in terms of an angular power spectrum (APS) $\rho : [-\pi, \pi] \to \mathbf{R}^+$ describing the received or transmitted power in a given physical direction θ, and the antenna array responses $\mathbf{a}^u : [-\pi, \pi] \to \mathbf{C}^{N\times 1}$, $\mathbf{a}^d : [-\pi, \pi] \to \mathbf{C}^{N\times 1}$. (Hereafter, integrals involving matrix-valued functions should be understood coordinate-wise.) The angular reciprocity is modeled by assuming that the APS, unlike the array response, is frequency invariant. On top of the formal derivation from the channel models in the next sections, this assumption is further motivated by several measurement campaigns (see for example [9, 11]), where, for typical duplex gaps, the APS is shown to exhibit strong frequency correlation properties.

2.2 Narrow-band Systems

2.2.1 Discrete Scattering Model

Let us consider a narrow-band system in a 2D (azimuth-only) scenario. A classical expression for a realization of the channel vector $\mathbf{h} := \mathbf{h}[m_0]$ at an arbitrary time m_0 is given by [8]:

$$\begin{aligned} \mathbf{h^d} &= \sqrt{\frac{\alpha}{N_p}} \sum_{i=1}^{N_p} e^{j\varphi_i^d} \mathbf{a}^d(\theta_i) \\ \mathbf{h^u} &= \sqrt{\frac{\alpha}{N_p}} \sum_{i=1}^{N_p} e^{j\varphi_i^u} \mathbf{a}^u(\theta_i) \end{aligned} \tag{2.6}$$

where the DL and UL channel vectors $\mathbf{h}^d, \mathbf{h}^u \in \mathbb{C}^{N\times 1}$ are expressed in term of their multipath components according to the following model:

- N_p is the total number of paths, and it is generally assumed to be very large.
- θ_i is the direction of arrival (DoA) or the direction of departure (DoD) of path i respectively for the UL and for the DL. It is notably assumed to be frequency invariant, according to the angular reciprocity assumption. θ_i is independently drawn from a generic probability density function $f : [-\pi, \pi] \longrightarrow \mathbb{R}$.
- $\alpha > 0$ is the path loss, and it is again assumed to be equal for UL and DL.
- $\mathbf{a}^d, \mathbf{a}^u : [-\pi, \pi] \longrightarrow \mathbb{C}^{N\times 1}$ are the BTS array responses. In FDD systems, they are different from DL and UL.
- φ_i^d, φ_i^u are the phase shift terms of each path. They are generally assumed to be i.i.d and uniformly distributed in $[-\pi, \pi]$. In FDD systems UL and DL coefficients are assumed to be uncorrelated.

To reflect the considerations given in Section 2.1.1, the time evolution of the channel is modeled as follows:

- Fast-varying parameters φ_i^d, φ_i^u and θ_i are drawn independently and kept fixed at intervals corresponding to the coherence time T_c (block-fading assumption).

- Slowly-varying parameters α and f are assumed constant over a WSS window T_{WSS}, with $T_{WSS} \gg T_c$. Moreover, adjacent WSS windows must experience similar parameters.

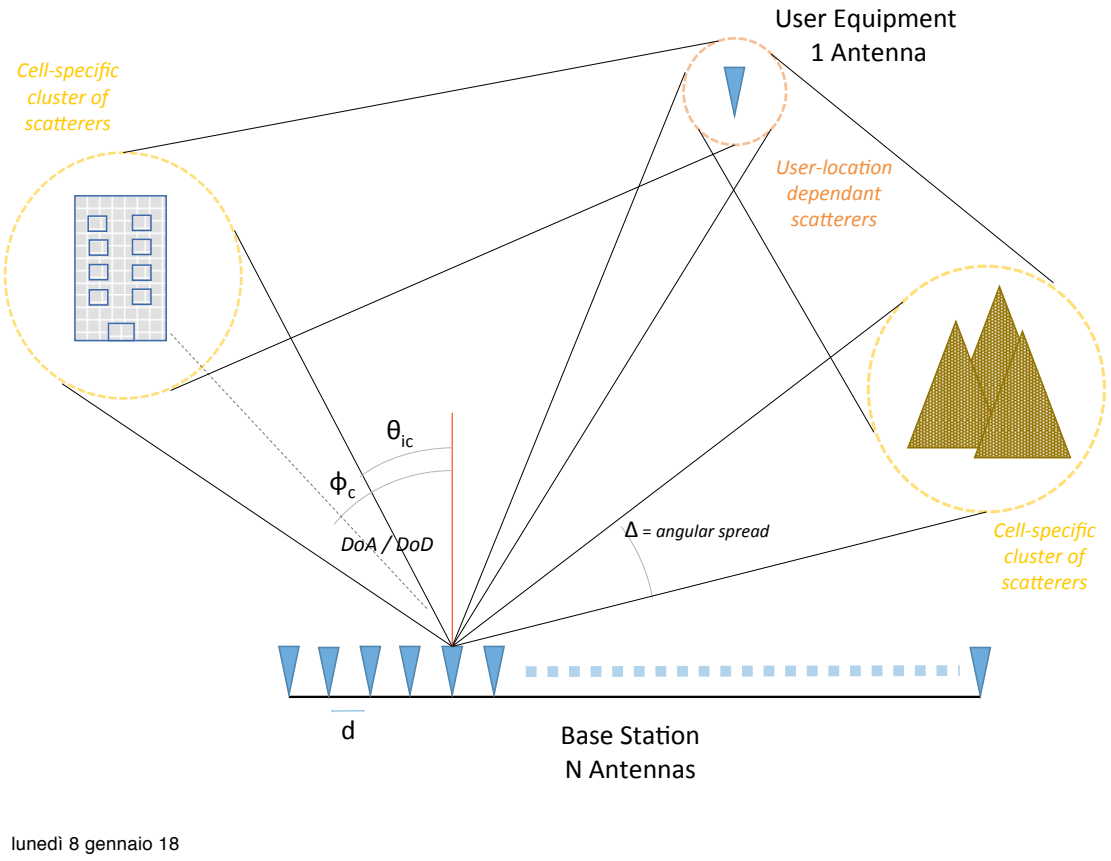


Figure 2.1: Massive MIMO channel in presence of a GSCM-like scattering environment.

The Geometry-based Stochastic Channel Model: In many applications a very effective model for a typical cellular environment is the so called "geometry-based stochastic channel model" (GSCM, Figure 2.1) [9], an extension of (2.6) where the paths are

clustered together as in the following expressions:

$$\begin{aligned}\mathbf{h}^d &= \frac{1}{\sqrt{N_p}} \sum_{c=1}^{N_c} \sum_{i=1}^{N_p} \sqrt{\alpha_c} e^{j\varphi_{ic}^d} \mathbf{a}^d(\theta_{ic}) \\ \mathbf{h}^u &= \frac{1}{\sqrt{N_p}} \sum_{c=1}^{N_c} \sum_{i=1}^{N_p} \sqrt{\alpha_c} e^{j\varphi_{ic}^u} \mathbf{a}^u(\theta_{ic})\end{aligned} \tag{2.7}$$

- N_c denotes the number of clusters and N_p the associated number of subpaths.
- θ_{ic} is the DoA/DoD of subpath i belonging to cluster c. It is randomly drawn from the continuous pdf $f_c(\theta)$ characterized by a main DoA/DoD ϕ_c and an angular spread Δ_c. Formal definition of these parameters will be clarified in the following.
- $\alpha_c > 0$ is the average power of all the subpaths of cluster c, and it assumed to be equal for UL and DL. The overall path loss is $\alpha := \sum_{c=1}^{N_c} \alpha_c$.
- $\mathbf{a}^d, \mathbf{a}^u : [-\pi, \pi] \longrightarrow \mathbb{C}^{N \times 1}$ are the BTS array responses as in (2.6).
- φ_{ic} are again assumed i.i.d uniformly distributed as in (2.6).

Reasonable models for $f_c(\theta)$ [12] in a typical cellular environment are:

- *Uniform distribution*
$$f_c(\theta) \sim \mathcal{U}\left[\phi_c - \frac{\Delta_c}{2}, \phi_c + \frac{\Delta_c}{2}\right]$$
- *Gaussian distribution*
$$f_c(\theta) \sim \mathcal{N}\left(\phi_c, \Delta_c^2\right)$$
- *Laplacian distribution*
$$f_c(\theta) = \frac{1}{\Delta_c \sqrt{2}} e^{-\frac{\sqrt{2}}{\Delta_c}|x - \phi_c|}$$

In [12] it is suggested that the Uniform distribution may be safely assumed for the line-of-sight (LOS) cluster (user position dependent cluster), while for the others a Gaussian or Laplacian distribution may better fit.

GSCM is particularly suitable for outdoor environments, where the clusters have the physical meaning of macro-objects (like buildings) responsible for the main reflections in the cell. For this reason, a model very similar to (2.7) is implemented in many simulators; e.g. the ones compliant with the 3GPP technical document [13]. The main difference between (2.7) and the model proposed by 3GPP is that, in the latter, only the main DoA/DoD of the clusters are drawn statistically, while the correspondent subpaths are computed deterministically with tabular values. This is done mainly to reduce the complexity of the simulation. In this work, instead, the proposed model is kept more general to not confine subpaths angles into a pre-defined grid, which is likely to be unrealistic.

The next proposition is the first result justifying that a continuous model for the angular power spectrum should be considered in channel estimation algorithms and simulations. Although Proposition 2.1 is based on arguably simple channel models, we show later that similar results hold for more advanced channel models considering 3D arrays with cross-polarization. This observation will play a key role in the algorithms for channel estimation considered in the next chapters.

Proposition 2.1. *For both models (2.6) and (2.7), the spatial covariance matrices* $\mathbf{R}^u = \mathbb{E}[\mathbf{h}^u(\mathbf{h}^u)^H]$ *and* $\mathbf{R}^d = \mathbb{E}[\mathbf{h}^d(\mathbf{h}^d)^H]$ *can be expressed as follows:*

$$\mathbf{R}^u = \int_{-\pi}^{\pi} \rho(\theta)\mathbf{a}^u(\theta)\mathbf{a}^u(\theta)^H d\theta, \tag{2.8}$$

$$\mathbf{R}^d = \int_{-\pi}^{\pi} \rho(\theta)\mathbf{a}^d(\theta)\mathbf{a}^d(\theta)^H d\theta, \tag{2.9}$$

where the function $\rho : [-\pi, \pi] \longrightarrow \mathbb{R}^+$*, here denominated as "angular power spectrum" (APS), is defined as*

$$\rho(\theta) := \alpha f(\theta)$$

for model (2.6), and

$$\rho(\theta) := \sum_{c=1}^{N_c} f_c(\theta)\alpha_c$$

for model (2.7). [1]

Proof. Let us drop for simplicity the UL/DL superscripts. By recalling channel model (2.6) and computing the conditional average over all the i.i.d random phases $\boldsymbol{\varphi}|\boldsymbol{\theta}$, with $\boldsymbol{\varphi} = [\varphi_1 \dots \varphi_{N_p}]$, and $\boldsymbol{\theta} = [\theta_1 \dots \theta_{N_p}]$, we obtain:

$$\begin{aligned}
\mathbf{R}' :&= \mathbb{E}_{\boldsymbol{\varphi}|\theta}[\mathbf{h}\mathbf{h}^H] \\
&= \frac{\alpha}{P} \sum_{i=1}^{N_p} \sum_{l=1}^{N_p} \mathbb{E}_{\varphi_i,\varphi_l|\theta} \left[e^{j\varphi_i} e^{-j\varphi_l}\right] \mathbf{a}(\theta_i)\mathbf{a}^H(\theta_l) \\
&= \frac{1}{P} \sum_{i}^{N_p} \mathbf{a}(\theta_i)\mathbf{a}^H(\theta_i),
\end{aligned}$$

where the last equality comes from

$$\mathbb{E}_{\varphi_i,\varphi_l|\theta} \left[e^{j\varphi_i} e^{-j\varphi_l}\right] = \delta_{il},$$

[1] The name "angular power spectrum" comes from its physical interpretation as a power density:

$$\int_{-\pi}^{\pi} \rho(\theta)d\theta = \alpha. \tag{2.10}$$

because, for $i \neq l$,

$$\begin{aligned}\mathbb{E}_{\varphi_i,\varphi_l|\theta}\left[e^{j\varphi_i}e^{-j\varphi_l}\right] &= \mathbb{E}_{\varphi_i|\theta}\left[e^{j\varphi_i}\right]\mathbb{E}_{\varphi_l|\theta}\left[e^{-j\varphi_l}\right] \\ &= \left(\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j\varphi}d\varphi\right)\left(\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-j\varphi'}d\varphi'\right) \\ &= 0\end{aligned}$$

By averaging now over all the i.i.d. random DoA/DoD $\boldsymbol{\theta}$:

$$\begin{aligned}\mathbf{R} &= \mathbb{E}_{\boldsymbol{\theta}}[\mathbf{R}'] \\ &= \frac{\alpha}{N_p}\sum_{i}^{N_p}\mathbb{E}_{\theta_i}[\mathbf{a}(\theta_i)\mathbf{a}^H(\theta_i)] \\ &= \alpha\int_{-\pi}^{\pi} f(\theta)\mathbf{a}(\theta)\mathbf{a}^H(\theta)d\theta \\ &= \int_{-\pi}^{\pi}\rho(\theta)\mathbf{a}(\theta)\mathbf{a}^H(\theta)d\theta.\end{aligned}$$

The expression for model (2.6) is now proved.

Let us now consider the GSCM channel model (2.7). By introducing the notation

$$\mathbf{h} = \sum_{c=1}^{N_c}\mathbf{h}_c, \quad \mathbf{h}_c = \sqrt{\frac{\alpha_c}{N_p}}\sum_{i=1}^{N_p} e^{j\varphi_{ic}}\mathbf{a}(\theta_{ic}), \tag{2.11}$$

it is possible to decompose the overall covariance matrix as

$$\mathbf{R} = \sum_{c=1}^{N_c}\mathbf{R}_{\mathbf{c}} + \sum_{c=1}^{N_c}\sum_{c'\neq c}\mathbf{R}_{cc'}$$

where $\mathbf{R}_c := \mathbb{E}[\mathbf{h}_c(\mathbf{h}_c)^H]$ denote the intra-cluster covariance matrix of cluster c, and $\mathbf{R}_{cc'} := \mathbb{E}[\mathbf{h}_c(\mathbf{h}_{c'})^H]$ denotes the inter-cluster covariance matrix between clusters c and c', $c \neq c'$. By observing that $\mathbf{h}_c$ has the same expression as the channel model (2.6), and by proceeding on the same lines of the derivation of its covariance matrix, it is immediately possible to write

$$\mathbf{R}_c = \alpha_c\int_{-\pi}^{\pi} f_c(\theta)\mathbf{a}(\theta)\mathbf{a}^H(\theta)d\theta,$$

$$\mathbf{R}_{cc'} = \mathbf{0},$$

where the last equation comes from

$$\mathbb{E}_{\varphi_{ic},\varphi_{lc'}|\theta}\left[e^{j\varphi_{ic}}e^{-j\varphi_{lc'}}\right] = \mathbb{E}_{\varphi_{ic}|\theta}\left[e^{j\varphi_{ic}}\right]\mathbb{E}_{\varphi_{lc'}|\theta}\left[e^{-j\varphi_{lc'}}\right] = 0, \quad \forall i, l.$$

Finally, $\mathbf{R}$ is given by

$$\begin{aligned}\mathbf{R} = \sum_{c=1}^{N_c} \mathbf{R_c} &= \int_{-\pi}^{\pi} \left(\sum_{c=1}^{N_c} \alpha_c f_c(\theta) \right) \mathbf{a}(\theta)\mathbf{a}^H(\theta) d\theta \\ &= \int_{-\pi}^{\pi} \rho(\theta)\mathbf{a}(\theta)\mathbf{a}^H(\theta) d\theta.\end{aligned}$$

The expression in (2.7) is proved, and the proof is complete. □

2.2.2 Continuous Scattering Model

Most of the simulators and the algorithms available in literature for massive MIMO channel estimation and also those for problems related to, for example, DoA estimation (e.g. MUSIC [14]) rely on the discrete scattering model considered in Section 2.2.1. However more recently some authors (see, for example, [10]) started considering generalizations of the discrete scattering model described in Section 2.2.1, by assuming the channel vector to be formed by a superposition of a continuum of array responses, weighted by a continuous function ρ. More precisely, in these generalizations, the downlink and uplink channels are given by, respectively,

$$\begin{aligned}\mathbf{h^d} &= \int_{-\pi}^{\pi} \sqrt{\rho(\theta)} z^d(\theta)\mathbf{a}^d(\theta) d\theta \text{ and} \\ \mathbf{h^u} &= \int_{-\pi}^{\pi} \sqrt{\rho(\theta)} z^u(\theta)\mathbf{a}^u(\theta) d\theta,\end{aligned} \tag{2.12}$$

where:

- $\rho : [-\pi, \pi] \longrightarrow \mathbb{R}^+$ describes the received or transmitted power per unit angle. Because of its equivalent physical meaning of the APS defined in Section 2.2.1, it is here denoted with the same name. Angular reciprocity is here modeled assuming equal APS for UL and DL.
- $\mathbf{a}^d, \mathbf{a}^u : [-\pi, \pi] \longrightarrow \mathbb{C}^{N\times 1}$ are the BTS array responses as in (2.6).
- $z^d(\theta), z^u(\theta)$ are the white unitary power complex random processes in the angular domain modeling the small scale fading. In FDD systems, UL and DL processes are uncorrelated. They are generally assumed to be circularly symmetric Gaussian.

The underlying assumption of this continuous scattering model is that the scattering environment might be continuous in nature, so developing algorithms relying on the most generic model possible (as in [10]) ensures their resilience to model mismatches.

To reflect the considerations given in Section 2.1.1, we model the time evolution of the channel as follows:

- The realizations of the random processes $z^d(\theta)$ and $z^u(\theta)$ are drawn independently, and they are kept fixed at intervals corresponding to the coherence time T_c (block-fading assumption).
- A realization of the random process $\rho(\theta)$ is kept fixed for a WSS window T_{WSS}, with $T_{WSS} \gg T_c$. Moreover, adjacent WSS windows must experience correlated realizations.

We now sketch a proof that the expressions for the covariance matrices for the channel model in this subsection are similar to those considering the discrete scattering model in Section 2.2.1.

Proposition 2.2. *The covariance matrices* $\mathbf{R^u}$ *and* $\mathbf{R}^d$ *for the channel model in* 2.12 *are given by*

$$\mathbf{R}^u = \int_{-\pi}^{\pi} \rho(\theta)\mathbf{a}^u(\theta)\mathbf{a}^u(\theta)^H d\theta, \tag{2.13}$$

$$\mathbf{R}^d = \int_{-\pi}^{\pi} \rho(\theta)\mathbf{a}^d(\theta)\mathbf{a}^d(\theta)^H d\theta. \tag{2.14}$$

Proof. (Informal) Let us drop for simplicity the UL/DL superscripts. By recalling the channel model in (2.12), we obtain

$$\begin{aligned}\mathbf{R} &= \mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right] \\ &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi} \sqrt{\rho(\theta)}\sqrt{\rho(\theta')}\mathbb{E}\left[z(\theta)z^*(\theta')\right]\mathbf{a}(\theta)\mathbf{a}(\theta')^H d\theta d\theta' \\ &= \int_{-\pi}^{\pi} \rho(\theta)\mathbf{a}(\theta)\mathbf{a}^H(\theta)d\theta,\end{aligned}$$

where the last equality comes from $\mathbb{E}[z(\theta)z^*(\theta')] = \delta(\theta - \theta')$, by definition, and the heuristic interpretation of the Dirac function. The proof is now complete. □

2.3 Wide-band OFDM Systems

2.3.1 Discrete Scattering Model

Let us consider a wide-band channel in an under-spread environment; i.e. with delay spread $T_s \ll T_c$, an assumption that is typically done while designing an OFDM system [1, Chapter 3.4]. Let us extend the narrow-band GSCM discrete scattering model (2.7) by using the same approach proposed in [9, Chapter 6] and in the 3GPP technical document [13], denoted as "tapped delay line." After sampling, and by denoting with $l \in \mathbb{N}$ the discrete time index of the lth tap, the sampled impulse response in the delay

domain is given by:

$$\begin{aligned}\mathbf{h}^d[l] &= \sum_{c=1}^{N_c} \mathbf{h}_c^d \delta[l - l_c], \\ \mathbf{h}^u[l] &= \sum_{c=1}^{N_c} \mathbf{h}_c^u \delta[l - l_c],\end{aligned} \tag{2.15}$$

where

$$\begin{aligned}\mathbf{h}_c^d &:= \sqrt{\frac{\alpha_c}{N_p}} \sum_{i=1}^{N_p} e^{j\varphi_{ic}^d} \mathbf{a}^d(\theta_{ic}), \\ \mathbf{h}_c^u &:= \sqrt{\frac{\alpha_c}{N_p}} \sum_{i=1}^{N_p} e^{j\varphi_{ic}^u} \mathbf{a}^u(\theta_{ic}).\end{aligned}$$

- N_c denotes the number of clusters and N_p the associated number of subpaths.
- $l_c \in \mathbb{N}$ denotes the discrete time delay of all the subpaths belonging to cluster c, which are assumed to be unresolvable in the delay domain after sampling. Finally, l_c is assumed to be equal for UL and DL.
- θ_{ic} is the DoA/DoD of subpath i belonging to cluster c. It is randomly drawn from the continuous pdf $f_c(\theta)$ with parameters ϕ_c and Δ_c as defined for the model in (2.7).
- $\alpha_c > 0$ is the average power of all the subpaths of cluster c, and it assumed to be equal for UL and DL. The total impulse response power is defined as $\alpha := \sum_{c=1}^{N_c} \alpha_c$.
- $\mathbf{a}^d, \mathbf{a}^u : [-\pi, \pi] \longrightarrow \mathbb{C}^{N \times 1}$ are the BTS array responses as in (2.6).
- φ_{ic} are again assumed i.i.d uniformly distributed as in (2.6).

To reflect the considerations given in Section 2.1.1, we model the time evolution of the channel as follows:

- The fast-varying parameters φ_{ic}^d, φ_{ic}^u, and θ_{ic} are drawn independently, and they are kept fixed at intervals corresponding to the coherence time T_c (block-fading assumption).
- The slowly time-varying parameters α_c, l_c, and f_c are assumed constant over a WSS window T_{WSS}, with $T_{WSS} \gg T_c$. Moreover, adjacent WSS windows must experience similar parameters.

The channel vector in the sub-carrier domain is then given by [1, Chapter 3.4]:

$$\begin{aligned}\tilde{\mathbf{h}}^d[k] &= \sum_{l=0}^{L-1} \mathbf{h}^d[l] e^{-j\frac{2\pi kl}{N_s}}, \\ \tilde{\mathbf{h}}^u[k] &= \sum_{l=0}^{L-1} \mathbf{h}^u[l] e^{-j\frac{2\pi kl}{N_s}},\end{aligned} \tag{2.16}$$

where L is the impulse response length, N_s is the chosen OFDM block length, and $k = 0 \dots (N_s - 1)$ is the sub-carrier index.

Proposition 2.3. *By assuming the discrete multi-path model (2.15), the space-frequency correlation matrices depend just on the sub-carrier indexes distance and not on the absolute sub-carrier indexes:*

$$\begin{aligned}\mathbb{E}\left[\tilde{\mathbf{h}}^d[k]\left(\tilde{\mathbf{h}}^d[k']\right)^H\right] &= \tilde{\mathbf{R}}^d[k-k'], \\ \mathbb{E}\left[\tilde{\mathbf{h}}^u[k]\left(\tilde{\mathbf{h}}^u[k']\right)^H\right] &= \tilde{\mathbf{R}}^u[k-k'].\end{aligned} \tag{2.17}$$

Furthermore, the spatial covariance matrices in the sub-carrier domain have the following expression, for every sub-carrier k:

$$\begin{aligned}\tilde{\mathbf{R}}^d := \tilde{\mathbf{R}}^d[0] &= \int_{-\pi}^{\pi} \rho(\theta)\mathbf{a}^d(\theta)\mathbf{a}^d(\theta)^H d\theta, \\ \tilde{\mathbf{R}}^u := \tilde{\mathbf{R}}^u[0] &= \int_{-\pi}^{\pi} \rho(\theta)\mathbf{a}^u(\theta)\mathbf{a}^u(\theta)^H d\theta,\end{aligned} \tag{2.18}$$

where the function $\rho : [-\pi, \pi] \longrightarrow \mathbb{R}^+$, here denominated again as "angular power spectrum" (APS), is defined as

$$\rho(\theta) := \sum_{c=1}^{N_c} f_c(\theta)\alpha_c.$$

Proof. Let us drop for simplicity the UL/DL superscripts. By recalling the channel model in (2.12), we obtain

$$\begin{aligned}\mathbf{R}[l, l'] :&= \mathbb{E}\left[\mathbf{h}[l]\mathbf{h}^H[l']\right] \\ &= \sum_{c=1}^{N_c}\sum_{c'=1}^{N_c} \mathbb{E}\left[\mathbf{h}_c\mathbf{h}_{c'}^H\right] \delta[l - l_c]\delta[l' - l_{c'}]\end{aligned}$$

This expression can be further simplified by observing that $\mathbf{h}_c$ has the same expression (and notation) of the quantity defined in (2.11) for the clustered model (2.7). Following

the same lines of the derivation in Section 2.2.1, it can be shown that

$$\begin{aligned}\mathbf{R}[l,l'] &= \sum_{c=1}^{N_c} \mathbb{E}\left[\mathbf{h}_c\mathbf{h}_c^H\right]\delta[l-l_c]\delta[l'-l_c] \\ &= \sum_{c=1}^{N_c} \mathbb{E}\left[\mathbf{h}_c\mathbf{h}_c^H\right]\delta[l-l_c]\delta[l-l'] \\ &= \left(\int_{-\pi}^{\pi} \alpha_c f_c(\theta)\mathbf{a}(\theta)\mathbf{a}^H(\theta)d\theta\right)\delta[l-l_c]\delta[l-l'].\end{aligned}$$

By considering now the channel in the sub-carrier domain defined in 2.15, it is possible to write the space-frequency correlation as

$$\begin{aligned}\mathbb{E}\left[\tilde{\mathbf{h}}[k]\tilde{\mathbf{h}}[k']^H\right] &= \sum_{l=0}^{L-1}\sum_{l'=0}^{L-1}\mathbf{R}[l,l']e^{-j\frac{2\pi}{N_s}(kl-k'l')} \\ &= \sum_{l=0}^{L-1}\mathbf{R}[l,l]e^{-j\frac{2\pi l}{N_s}(k-k')} =: \tilde{\mathbf{R}}[k-k']\end{aligned}$$

By considering now the spatial only covariance matrix in the sub-carrier domain we obtain

$$\begin{aligned}\tilde{\mathbf{R}}[0] &= \sum_{l=0}^{L-1}\mathbf{R}[l,l] \\ &= \sum_{l=0}^{L-1}\left(\int_{-\pi}^{\pi}\alpha_c f_c(\theta)\mathbf{a}(\theta)\mathbf{a}^H(\theta)d\theta\right)\delta[l-l_c] \\ &= \int_{-\pi}^{\pi}\left(\sum_{c=1}^{N_c}\alpha_c f_c(\theta)\right)\mathbf{a}(\theta)\mathbf{a}^H(\theta)d\theta\end{aligned}$$

The expressions (2.17) and (2.18) are then proved. □

2.3.2 Continuous Scattering Model

Let us now generalize the channel model described in Section 2.3.1 obtained by applying the continuous channel modeling philosophy as discussed in Section 2.2.2. A possible expression for the channel vector in the delay domain is given by

$$\begin{aligned}\mathbf{h}^{\mathbf{d}}(\tau) &= \int_{-\pi}^{\pi}\sqrt{\gamma(\theta,\tau)}z^d(\theta,\tau)\mathbf{a}^d(\theta)d\theta, \\ \mathbf{h}^{\mathbf{u}}(\tau) &= \int_{-\pi}^{\pi}\sqrt{\gamma(\theta,\tau)}z^u(\theta,\tau)\mathbf{a}^u(\theta)d\theta,\end{aligned} \tag{2.19}$$

where

- $\gamma : [-\pi, \pi] \times \mathbb{R}^+ \to \mathbb{R}^+$ describes the received or transmitted power per unit angle and unit delay. Angular reciprocity is here modeled by assuming equal γ for UL and DL.
- $\mathbf{a}^d, \mathbf{a}^u : [-\pi, \pi] \longrightarrow \mathbb{C}^{N \times 1}$ are the BTS array responses as in (2.6).
- $z^d(\theta, \tau), z^u(\theta, \tau)$ are the unitary power complex random processes in the angular and delay, satisfying

$$\mathbb{E}\left[z^d(\theta,\tau)(z^d(\theta',\tau'))^*\right] = \mathbb{E}\left[z^d(\theta,\tau)(z^d(\theta',\tau'))^*\right] = \delta(\theta-\theta')\delta_{\tau,\tau'},$$

  which models the small scale fading. In FDD systems, the UL and DL processes are uncorrelated. They are generally assumed to be circularly symmetric Gaussian. Note: $\delta(\cdot)$ denotes the Dirac delta, and $\delta_{\cdot,\cdot}$ the Kronecker delta.

To reflect the considerations given in Section 2.1.1, we model the time evolution of the channel as follows:

- The realizations of the random processes $z^d(\theta, \tau)$, $z^u(\theta, \tau)$ are drawn independently, and they are kept fixed at intervals corresponding to the coherence time T_c (block-fading assumption).
- A realization of the random process $\gamma(\theta, \tau)$ is kept fixed for a WSS window T_{WSS}, with $T_{WSS} \gg T_c$. Moreover, adjacent WSS windows must experience correlated realizations.

This model is very similar to the one proposed in [10], but it is here generalized so that γ is not assumed to be constant over the delays τ. This is justified intuitively by assuming that components with different delays (and, in particular, when the delay difference is large) are originated by different cluster of scatterers, and thus they may experience different average attenuation.

Let us now assume an OFDM system with bandwidth W and define the impulse response in the sub-carrier domain as follows:

$$\begin{aligned}
\tilde{\mathbf{h}}^d[k] &= \sum_{l=0}^{L-1} \mathbf{h}^d[l] e^{-j\frac{2\pi kl}{N_s}}, \quad \mathbf{h}^d[l] := \mathbf{h}^d\left(\frac{l}{W}\right), \\
\tilde{\mathbf{h}}^u[k] &= \sum_{l=0}^{L-1} \mathbf{h}^u[l] e^{-j\frac{2\pi kl}{N_s}}, \quad \mathbf{h}^u[l] := \mathbf{h}^u\left(\frac{l}{W}\right),
\end{aligned} \tag{2.20}$$

where L is the impulse response length (assumed finite), N_s is the chosen OFDM block length, and $k = 0 \dots (N_s - 1)$ is the sub-carrier index.

Proposition 2.4. *By assuming the continuous scattering model in (2.19), the space-frequency correlation matrices depend just on the sub-carrier indexes distance and not*

on the absolute sub-carrier indexes:

$$\begin{aligned}\mathbb{E}\left[\tilde{\mathbf{h}}^d[k]\left(\tilde{\mathbf{h}}^d[k']\right)^H\right] &= \tilde{\mathbf{R}}^d[k-k'],\\ \mathbb{E}\left[\tilde{\mathbf{h}}^u[k]\left(\tilde{\mathbf{h}}^u[k']\right)^H\right] &= \tilde{\mathbf{R}}^u[k-k'].\end{aligned} \tag{2.21}$$

Furthermore, the spatial covariance matrices in the sub-carrier domain have the following expression for every sub-carrier k:

$$\begin{aligned}\tilde{\mathbf{R}}^d := \tilde{\mathbf{R}}^d[0] &= \int_{-\pi}^{\pi}\rho(\theta)\mathbf{a}^d(\theta)\mathbf{a}^d(\theta)^H d\theta,\\ \tilde{\mathbf{R}}^u := \tilde{\mathbf{R}}^u[0] &= \int_{-\pi}^{\pi}\rho(\theta)\mathbf{a}^u(\theta)\mathbf{a}^u(\theta)^H d\theta,\end{aligned} \tag{2.22}$$

where the function $\rho : [-\pi,\pi] \longrightarrow \mathbb{R}^+$, here denominated again as APS, is defined as

$$\rho(\theta) := \sum_{l=0}^{L-1}\gamma\left(\theta, \frac{l}{W}\right),$$

which represent the total angular power density over all taps of the sampled impulse response.

Proof. (Informal) The proof is based on the heuristic characterization of the Dirac function. Let us drop for simplicity the UL/DL superscripts. Recalling the channel model in (2.19), we obtain

$$\begin{aligned}\mathbf{R}(\tau,\tau') :&= \mathbb{E}[\mathbf{h}(\tau)\mathbf{h}(\tau')^H]\\ &= \int_{-\pi}^{\pi}\int_{-\pi}^{\pi}\sqrt{\gamma(\theta,\tau)}\sqrt{\gamma(\theta',\tau')}\mathbb{E}[z(\theta,\tau)z(\theta',\tau')^*]\mathbf{a}(\theta)\mathbf{a}(\theta')^H d\theta d\theta'\\ &= \left(\int_{-\pi}^{\pi}\sqrt{\gamma(\theta,\tau)}\sqrt{\gamma(\theta,\tau')}\mathbf{a}(\theta)\mathbf{a}(\theta)^H d\theta\right)\delta_{\tau,\tau'}.\end{aligned}$$

By considering now the channel in the sub-carrier domain, it is possible to write the space-frequency correlation as

$$\begin{aligned}\mathbb{E}\left[\tilde{\mathbf{h}}[k]\tilde{\mathbf{h}}[k']^H\right] &= \sum_{l=0}^{L-1}\sum_{l'=0}^{L-1}\mathbf{R}\left(\frac{l}{W},\frac{l'}{W}\right)e^{-j\frac{2\pi}{N_s}(kl-k'l')}\\ &= \sum_{l=0}^{L-1}\mathbf{R}\left(\frac{l}{W},\frac{l}{W}\right)e^{-j\frac{2\pi l}{N_s}(k-k')} := \tilde{\mathbf{R}}[k-k'].\end{aligned}$$

The spatial only covariance matrix in the sub-carrier domain is

$$\begin{aligned}\tilde{\mathbf{R}}[0] &= \sum_{l=0}^{L-1} \mathbf{R}\left(\frac{l}{W}, \frac{l}{W}\right) \\ &= \sum_{l=0}^{L-1} \int_{-\pi}^{\pi} \gamma\left(\theta, \frac{l}{W}\right) \mathbf{a}(\theta)\mathbf{a}^H(\theta) d\theta \\ &= \int_{-\pi}^{\pi} \sum_{l=0}^{L-1} \gamma\left(\theta, \frac{l}{W}\right) \mathbf{a}(\theta)\mathbf{a}^H(\theta) d\theta.\end{aligned}$$

The expressions (2.21) and (2.22) are then proved. □

2.4 Realistic Propagation and Antennas

Although the channel models presented in Section 2.2 and Section 2.3, owing to their simplicity, are still very popular in the scientific literature (see for example [8, 10]), the increasing complexity of modern wireless communication systems (e.g., the 4G and 5G architectures) has demanded channel models able to describe real propagation phenomena more accurately (see, for example, the relatively recent 3GPP technical report [15]). In particular, this section focuses on the following aspects:

- In the models in Section 2.2 and Section 2.3, the UE antenna is assumed to be omni-directional, thus its angular response is ignored. In practice, omni-directional antennas do not exist, so a description of the UE antenna radiation pattern is needed.

- Polarization diversity, achieved for example by co-located cross-polarized antennas, has been shown to be a promising technique to increase the degrees of freedom (DoF) of MIMO channels at a relatively low cost for the operators [9]. For this reason, models for the propagation of dual-polarized radiation are required.

- For a better description of a real 3D environment, directional channel models should include a description of the multipath propagation in spherical coordinates; i.e., in both azimuth and zenith.

By considering the aspects outlined above, and by focusing, for simplicity, on a discrete narrow-band GSCM scattering model similar to the the one proposed in Section 2.2.1,

the channel vector $\mathbf{h} := \mathbf{h}[m_0]$ at an arbitrary time m_0 is given by

$$\mathbf{h}^d = \sum_{c=1}^{N_c} \mathbf{h}_c^d, \quad \mathbf{h}^u = \sum_{c=1}^{N_c} \mathbf{h}_c^u,$$

$$\mathbf{h}_c^d := \sqrt{\frac{\alpha_c}{N_p}} \sum_{i=1}^{N_p} \mathbf{A}^d(\boldsymbol{\theta}_{ic}) \begin{bmatrix} e^{j\varphi^d_{VV,ic}} & \frac{1}{\sqrt{K_{ic}}} e^{j\varphi^d_{VH,ic}} \\ \frac{1}{\sqrt{K_{ic}}} e^{j\varphi^d_{HV,ic}} & e^{j\varphi^d_{HH,ic}} \end{bmatrix} \mathbf{B}(\boldsymbol{\phi}_{ic})^H, \tag{2.23}$$

$$\mathbf{h}_c^u := \sqrt{\frac{\alpha_c}{N_p}} \sum_{i=1}^{N_p} \mathbf{A}^u(\boldsymbol{\theta}_{ic}) \begin{bmatrix} e^{j\varphi^u_{VV,ic}} & \frac{1}{\sqrt{K_{ic}}} e^{j\varphi^u_{VH,ic}} \\ \frac{1}{\sqrt{K_{ic}}} e^{j\varphi^u_{HV,ic}} & e^{j\varphi^u_{HH,ic}} \end{bmatrix} \mathbf{B}(\boldsymbol{\phi}_{ic})^H,$$

where

- N_c denotes the number of clusters and N_p the associated number of subpaths.
- $\boldsymbol{\theta}_{ic}$ and $\boldsymbol{\phi}_{ic}$ are either the DoD and DoA of subpath i for the DL case, or the DoA and DoD of subpath i for the UL case. The directions $\boldsymbol{\theta}_{ic}$ and $\boldsymbol{\phi}_{ic}$ are defined as tuples taking values in the set $\Omega := [-\pi, \pi] \times [0, \pi]$, representing the azimuth and the zenith of a spherical coordinate system. They are drawn independently from the joint distributions $f_c(\boldsymbol{\theta}, \boldsymbol{\phi})$, and they are assumed to be equal for UL and DL.
- $\mathbf{A}^d, \mathbf{A}^u : \Omega \to \mathbb{C}^{N\times 2}$ are the dual polarized antenna array responses of the BTS. In FDD systems, they are different from DL and UL. The columns of $\mathbf{A}^d, \mathbf{A}^u$ are denoted with $[\mathbf{a}_V^d, \mathbf{a}_H^d], [\mathbf{a}_V^u, \mathbf{a}_H^u]$, and they describe respectively the array responses for the vertical and for the horizontal polarization.
- $\alpha_c > 0$ is the average power of all the subpaths of cluster c, and it assumed to be equal for UL and DL.
- $\mathbf{B} : \Omega \to \mathbb{R}^{1\times 2}$ is the dual polarized antenna radiation pattern of the UE. It is assumed to be frequency independent. The columns of $\mathbf{B}$ are denoted with $[b_V, b_H]$, and they describe respectively the radiation patterns for the vertical and for the horizontal polarization.
- The random matrices

$$\mathbf{M}_{ic}^d := \begin{bmatrix} e^{j\varphi^d_{VV,ic}} & \frac{1}{\sqrt{K_i}} e^{j\varphi^d_{VH,ic}} \\ \frac{1}{\sqrt{K_i}} e^{j\varphi^d_{HV,ic}} & e^{j\varphi^d_{HH,ic}} \end{bmatrix},$$

$$\mathbf{M}_{ic}^u := \begin{bmatrix} e^{j\varphi^u_{VV,ic}} & \frac{1}{\sqrt{K_{ic}}} e^{j\varphi^u_{VH,ic}} \\ \frac{1}{\sqrt{K_{ic}}} e^{j\varphi^u_{HV,ic}} & e^{j\varphi^u_{HH,ic}} \end{bmatrix},$$

model the fading of the vertical and horizontal polarization as well as of the cross-polarization terms originated by the polarization changes that the electromagnetic waves undergo during the propagation. The random phases

$$\{\varphi^d_{VV,ic}, \varphi^d_{VH,ic}, \varphi^d_{HV,ic}, \varphi^d_{HH,ic}\},$$
$$\{\varphi^u_{VV,ic}, \varphi^u_{VH,ic}, \varphi^u_{HV,ic}, \varphi^u_{HH,ic}\},$$

are assumed i.i.d. uniformly distributed in $[-\pi, \pi]$. The parameters K_{ic} are the cross polarization power ratios (XPRs), and they are assumed to be i.i.d. random variables and to be equal for UL and DL. This polarization model is identical to the one suggested by the 3GPP technical document [15] and by [9, Chapter 7], where the two polarizations are assumed to experience independent fading.

To reflect the considerations given in section 2.1.1, we model the time evolution of the channel as follows:

- The fast time-varying parameters

$$\{\varphi^d_{VV,ic}, \varphi^d_{VH,ic}, \varphi^d_{HV,ic}, \varphi^d_{HH,ic}\},$$
$$\{\varphi^u_{VV,ic}, \varphi^u_{VH,ic}, \varphi^u_{HV,ic}, \varphi^u_{HH,ic}\},$$

  $\boldsymbol{\theta}_{ic}$, $\boldsymbol{\phi}_{ic}$, and K_{ic} are drawn independently and kept fixed at intervals corresponding to the coherence time T_c (block-fading assumption).

- The slow time-varying parameters α_c and f_c are assumed constant over a WSS window T_{WSS}, with $T_{WSS} \gg T_c$. Moreover, adjacent WSS windows must experience similar parameters.

Proposition 2.5. *With the model in (2.23), the spatial covariance matrices have the following expressions:*

$$\begin{aligned}\mathbf{R}^d &= \int_\Omega \rho_V(\boldsymbol{\theta})\mathbf{a}^d_V(\boldsymbol{\theta})\mathbf{a}^d_V(\boldsymbol{\theta})^H d^2\boldsymbol{\theta} + \int_\Omega \rho_H(\boldsymbol{\theta})\mathbf{a}^d_H(\boldsymbol{\theta})\mathbf{a}^d_H(\boldsymbol{\theta})^H d^2\boldsymbol{\theta},\\ \mathbf{R}^u &= \int_\Omega \rho_V(\boldsymbol{\theta})\mathbf{a}^u_V(\boldsymbol{\theta})\mathbf{a}^u_V(\boldsymbol{\theta})^H d^2\boldsymbol{\theta} + \int_\Omega \rho_H(\boldsymbol{\theta})\mathbf{a}^u_H(\boldsymbol{\theta})\mathbf{a}^u_H(\boldsymbol{\theta})^H d^2\boldsymbol{\theta},\end{aligned} \tag{2.24}$$

where the functions $\rho_V, \rho_H : \Omega \to \mathbb{R}^+$, here denominated respectively as "vertical polarization angular power spectrum" (V-APS) and "horizontal polarization angular power spectrum" (H-APS) are defined as

$$\rho_V(\boldsymbol{\theta}) := \sum_{c=1}^{N_c} \alpha_c \int_\Omega f_c(\boldsymbol{\theta}, \boldsymbol{\phi}) \left(b^2_V(\boldsymbol{\phi}) + \frac{1}{K} b^2_H(\boldsymbol{\phi}) \right) d^2\boldsymbol{\phi},$$
$$\rho_H(\boldsymbol{\theta}) := \sum_{c=1}^{N_c} \alpha_c \int_\Omega f_c(\boldsymbol{\theta}, \boldsymbol{\phi}) \left(b^2_H(\boldsymbol{\phi}) + \frac{1}{K} b^2_V(\boldsymbol{\phi}) \right) d^2\boldsymbol{\phi},$$

where $1/K := \mathbb{E}[1/K_{ic}]$ describes the average effect of the XPRs K_{ic}.

Proof. Let us drop for simplicity the UL/DL superscripts. By recalling the channel model in (2.23), and by computing the inter-cluster and intra-cluster expectations over the random phases $\boldsymbol{\varphi} := \{\varphi_{VV,ic}, \varphi_{VH,ic}, \varphi_{HV,ic}, \varphi_{HH,ic}\}$ conditioned on the random angles $\boldsymbol{\theta} := \{\boldsymbol{\theta}_{ic}\}$, $\boldsymbol{\phi} := \{\boldsymbol{\phi}_{ic}\}$, and on the random XPRs $\mathbf{K} := \{K_{ic}\}$, we obtain

$$\mathbf{R}'_{cc'} := \mathbb{E}_{\varphi|\boldsymbol{\theta},\boldsymbol{\phi},\mathbf{K}}[\mathbf{h}_c\mathbf{h}_{c'}^H] = \frac{\sqrt{\alpha_c\alpha_{c'}}}{N_p}\sum_{i=1}^{N_p}\sum_{l=1}^{N_p}\mathbf{A}(\boldsymbol{\theta}_{ic})\mathbf{X}_{ic,lc'}\mathbf{A}(\boldsymbol{\theta}_{lc'})^H,$$

where

$$\begin{aligned}\mathbf{X}_{ic,lc'} :&= \mathbb{E}_{\boldsymbol{\varphi}|\boldsymbol{\theta},\boldsymbol{\phi},\mathbf{K}}\left[\mathbf{M}_{ic}\mathbf{B}(\boldsymbol{\phi}_{ic})^H\mathbf{B}(\boldsymbol{\phi}_{lc'})\mathbf{M}_{lc'}^H\right]\\ &= \begin{cases}\begin{bmatrix} b_V(\boldsymbol{\phi}_{ic})^2 + \dfrac{b_H(\boldsymbol{\phi}_{ic})^2}{K_{ic}} & 0 \\ 0 & b_H(\boldsymbol{\phi}_{ic})^2 + \dfrac{b_V(\boldsymbol{\phi}_{ic})^2}{K_{ic}}\end{bmatrix}, & \text{if } i = l \cap c = c'\\ \mathbf{0}, & \text{otherwise}\end{cases},\end{aligned}$$

The expression for $\mathbf{X}_{ic,lc'}$ is due to the uncorrelated phases of different multi-path components, and its derivation is similar to that in Section 2.2.1. Since there is no inter-cluster correlation, it is possible to focus just on the intra-cluster correlation $\mathbf{R}'_{cc}$. By computing now the expectation of $\mathbf{X}_{ic,ic}$ over the XPRs $\mathbf{K}$ conditioned on the random angles $\boldsymbol{\theta}$, $\boldsymbol{\phi}$, we obtain

$$\mathbb{E}_{\mathbf{K}|\boldsymbol{\theta},\boldsymbol{\phi}}[\mathbf{X}_{ic,ic}] = \begin{bmatrix} b_V(\boldsymbol{\phi}_{ic})^2 + \dfrac{b_H(\boldsymbol{\phi}_{ic})^2}{K} & 0 \\ 0 & b_H(\boldsymbol{\phi}_{ic})^2 + \dfrac{b_V(\boldsymbol{\phi}_{ic})^2}{K}\end{bmatrix},$$

where we define $1/K := \mathbb{E}[1/K_{ic}]$. With this in hand, we obtain

$$\begin{aligned}\mathbf{R}''_c :=& \mathbb{E}_{\boldsymbol{\varphi},\mathbf{K}|\boldsymbol{\theta},\boldsymbol{\phi}}\left[\mathbf{h}_c\mathbf{h}_c^H\right] = \mathbb{E}_{\mathbf{K}|\boldsymbol{\theta},\boldsymbol{\phi}}[\mathbf{R}'_{cc}]\\ =& \frac{\alpha_c}{N_p}\sum_{i=1}^{N_p}\mathbf{A}(\boldsymbol{\theta}_{ic})\mathbb{E}_{\mathbf{K}|\boldsymbol{\theta},\boldsymbol{\phi}}[\mathbf{X}_{ic,ic}]\mathbf{A}(\boldsymbol{\theta}_{ic})^H\\ =& \frac{\alpha_c}{N_p}\sum_{i=1}^{N_p}\left(b_V(\boldsymbol{\phi}_{ic})^2 + \frac{b_H(\boldsymbol{\phi}_{ic})^2}{K}\right)\mathbf{a}_V(\boldsymbol{\theta}_{\mathbf{ic}})\mathbf{a}_V(\boldsymbol{\theta}_{ic})^H\\ &+ \frac{\alpha_c}{N_p}\sum_{i=1}^{N_p}\left(b_H(\boldsymbol{\phi}_{ic})^2 + \frac{b_V(\boldsymbol{\phi}_{ic})^2}{K}\right)\mathbf{a}_H(\boldsymbol{\theta}_{\mathbf{ic}})\mathbf{a}_H(\boldsymbol{\theta}_{ic})^H.\end{aligned}$$

By computing now the complete intra-cluster correlation we obtain

$$\begin{aligned}\mathbf{R}_c :=& \mathbb{E}\left[\mathbf{h}_c\mathbf{h}_c^H\right] = \mathbb{E}_{\boldsymbol{\theta},\boldsymbol{\phi}}[\mathbf{R}''_{cc}]\\ =& \int_\Omega\int_\Omega \alpha_c f_c(\boldsymbol{\theta},\boldsymbol{\phi})\left(b_V(\boldsymbol{\phi})^2 + \frac{b_H(\boldsymbol{\phi})^2}{K}\right)\mathbf{a}_V(\boldsymbol{\theta})\mathbf{a}_V(\boldsymbol{\theta})^H d^2\boldsymbol{\theta}d^2\boldsymbol{\phi}\\ &+ \int_\Omega\int_\Omega \alpha_c f_c(\boldsymbol{\theta},\boldsymbol{\phi})\left(b_H(\boldsymbol{\phi})^2 + \frac{b_V(\boldsymbol{\phi})^2}{K}\right)\mathbf{a}_H(\boldsymbol{\theta})\mathbf{a}_H(\boldsymbol{\theta})^H d^2\boldsymbol{\theta}d^2\boldsymbol{\phi}.\end{aligned}$$

Finally, we obtain the overall spatial covariance matrix as

$$\begin{aligned}\mathbf{R} :=& \mathbb{E}\left[\mathbf{h}\mathbf{h}^H\right] = \sum_{c=1}^{N_c} \mathbf{R_c} \\ =& \int_\Omega \left[\sum_{c=1}^{N_c} \alpha_c \int_\Omega f_c(\boldsymbol{\theta},\boldsymbol{\phi}) \left(b_V(\boldsymbol{\phi})^2 + \frac{b_H(\boldsymbol{\phi})^2}{K}\right) d^2\boldsymbol{\phi}\right] \mathbf{a}_V(\boldsymbol{\theta})\mathbf{a}_V(\boldsymbol{\theta})^H d^2\boldsymbol{\theta} \\ &+ \int_\Omega \left[\sum_{c=1}^{N_c} \alpha_c \int_\Omega f_c(\boldsymbol{\theta},\boldsymbol{\phi}) \left(b_H(\boldsymbol{\phi})^2 + \frac{b_V(\boldsymbol{\phi})^2}{K}\right) d^2\boldsymbol{\phi}\right] \mathbf{a}_H(\boldsymbol{\theta})\mathbf{a}_H(\boldsymbol{\theta})^H d^2\boldsymbol{\theta},\end{aligned}$$

and the expression (2.24) is proved. □

Remark 2.1. *(Non-independent polarization fading) In general, if the two polarizations cannot be assumed to fade independently, the resulting covariance model would have the form*

$$\begin{aligned}\mathbf{R} =& \int_\Omega \rho_{VV}(\boldsymbol{\theta})\mathbf{a}_V(\boldsymbol{\theta})\mathbf{a}_V(\boldsymbol{\theta})^H d^2\boldsymbol{\theta} + \int_\Omega \rho_{HH}(\boldsymbol{\theta})\mathbf{a}_H(\boldsymbol{\theta})\mathbf{a}_H(\boldsymbol{\theta})^H d^2\boldsymbol{\theta} \\ &+ \int_\Omega \rho_{VH}(\boldsymbol{\theta})\mathbf{a}_V(\boldsymbol{\theta})\mathbf{a}_H(\boldsymbol{\theta})^H d^2\boldsymbol{\theta} + \int_\Omega \rho_{HV}(\boldsymbol{\theta})\mathbf{a}_H(\boldsymbol{\theta})\mathbf{a}_V(\boldsymbol{\theta})^H d^2\boldsymbol{\theta}.\end{aligned} \tag{2.25}$$

2.5 Multi-antenna UE

We now generalize the system model presented in section 2.1 by considering a BTS with N antennas and a UE with M antennas. By assuming a flat-fading channel model, a typical discrete input-output relation describing the UL communication between the BTS and the UE is given by

$$\mathbf{y}[m] = \mathbf{H}^u[m]\mathbf{x}[m] + \bar{\mathbf{y}}[m] + \mathbf{w}[m], \tag{2.26}$$

where $m \in \mathbb{Z}$ is the discrete time index, $\mathbf{H}^u[m] \in \mathbb{C}^{N\times M}$ is the UL time-variant channel matrix between the UE and the BTS, $\mathbf{x}[m] \in \mathbb{C}^M$ is the UE transmitted sequence, $\mathbf{y}[m] \in \mathbb{C}^{N\times 1}$ is the received signal at the BTS, $\bar{\mathbf{y}}[m] \in \mathbb{C}^{N\times 1}$ is the MU interference, and $\mathbf{w}[m] \sim \mathcal{CN}(0, \sigma_W^2 I_N)$ is the received white noise process. Similarly, a typical discrete input-output relation describing the DL communication between the UE and the BTS is given by

$$\mathbf{y}[m] = \left(\mathbf{H}^d[m]\right)^T \mathbf{x}[m] + \mathbf{w}[m], \tag{2.27}$$

$\mathbf{H}^d[m] \in \mathbb{C}^{N\times M}$ is the DL time-variant channel matrix, $\mathbf{x}[m] \in \mathbb{C}^{N\times 1}$ is the transmitted sequence, $\mathbf{y}[m] \in \mathbb{C}^{M\times 1}$ is the received signal, and $\mathbf{w}[m] \sim \mathcal{CN}(0, \sigma_W^2 I_M)$ is the received white noise process.

Similarly to what described in Section 2.1.1, a classic statistical channel model assume the coefficients of $\mathbf{H}$ (dropping UL and DL for simplicity) to be a zero-mean WSS process; i.e., with spatio-temporal covariance matrix given by [9]

$$\mathbb{E}\left[\mathrm{vec}(\mathbf{H}[m])\mathrm{vec}(\mathbf{H}[n])^H\right] = \mathbf{R}[m-n], \tag{2.28}$$

from where it is again possible to extract the channel coherence interval M_c and the spatial covariance matrix $\mathbf{R} := \mathbf{R}[0]$.

In many applications the second order statistics of the channel is often described (not completely) by using the "receive covariance matrix" $\mathbf{R}_{RX}$ and the "transmit covariance matrix" $\mathbf{R}_{TX}$, which are given by [9]

$$\begin{aligned} \mathbf{R}^u_{RX} &= \mathbb{E}\left[\mathbf{H}^u[m]\mathbf{H}^u[m]^H\right], \quad \mathbf{R}^u_{TX} = \mathbb{E}\left[\mathbf{H}^u[m]^H\mathbf{H}^u[m]\right], \\ \mathbf{R}^d_{RX} &= \mathbb{E}\left[\mathbf{H}^d[m]^H\mathbf{H}^d[m]\right], \quad \mathbf{R}^d_{TX} = \mathbb{E}\left[\mathbf{H}^d[m]\mathbf{H}^d[m]^H\right]. \end{aligned} \tag{2.29}$$

For example, under certain assumptions (e.g. independent statistics of DoAs and DoDs [9], which are not always realistic), the complete spatial covariance matrix can be expressed by using the following "Kronecker model:"

$$\mathbf{R}^u = \mathbf{R}^u_{RX} \otimes \mathbf{R^u}_{TX}, \quad \mathbf{R}^d = \mathbf{R}^d_{TX} \otimes \mathbf{R^d}_{RX}, \tag{2.30}$$

where $\otimes$ denotes the Kronecker product.

Let us consider the directional channel model (2.23), and let us extend it to the multi-antenna UE case as follows:

$$\begin{aligned} \mathbf{H}^d &= \sum_{c=1}^{N_c} \mathbf{H}^d_c, \quad \mathbf{H}^u = \sum_{c=1}^{N_c} \mathbf{H}^u_c, \\ \mathbf{H}^d_c &:= \sqrt{\frac{\alpha_c}{N_p}} \sum_{i=1}^{N_p} \mathbf{A}^d(\boldsymbol{\theta}_{ic}) \begin{bmatrix} e^{j\varphi^d_{VV,ic}} & \frac{1}{\sqrt{K_{ic}}} e^{j\varphi^d_{VH,ic}} \\ \frac{1}{\sqrt{K_{ic}}} e^{j\varphi^d_{HV,ic}} & e^{j\varphi^d_{HH,ic}} \end{bmatrix} \mathbf{B}^d(\boldsymbol{\phi}_{ic})^H, \\ \mathbf{H}^u_c &:= \sqrt{\frac{\alpha_c}{N_p}} \sum_{i=1}^{N_p} \mathbf{A}^u(\boldsymbol{\theta}_{ic}) \begin{bmatrix} e^{j\varphi^u_{VV,ic}} & \frac{1}{\sqrt{K_{ic}}} e^{j\varphi^u_{VH,ic}} \\ \frac{1}{\sqrt{K_{ic}}} e^{j\varphi^u_{HV,ic}} & e^{j\varphi^u_{HH,ic}} \end{bmatrix} \mathbf{B}^u(\boldsymbol{\phi}_{ic})^H, \end{aligned} \tag{2.31}$$

where all the quantities are identical to (2.23), except for the effect of the UE that is now modeled by using the frequency dependent dual polarized antenna array responses $\mathbf{B}^d, \mathbf{B}^u : \Omega \to \mathbb{C}^{M\times 2}$. The columns of $\mathbf{B}^d, \mathbf{B}^u$ are denoted with $[\mathbf{b}^d_V, \mathbf{b}^d_H], [\mathbf{b}^u_V, \mathbf{b}^u_H]$, and they describe respectively the array responses for the vertical and for the horizontal polarization.

Proposition 2.6. *By assuming model in (2.31), the receive and transmit covariance matrices have the following expressions:*

$$\begin{aligned}
\mathbf{R}_{RX}^{u} &= \int_{\Omega} \rho_V(\boldsymbol{\theta}) \mathbf{a}_V^u(\boldsymbol{\theta}) \mathbf{a}_V^u(\boldsymbol{\theta})^H d^2\boldsymbol{\theta} + \int_{\Omega} \rho_H(\boldsymbol{\theta}) \mathbf{a}_H^u(\boldsymbol{\theta}) \mathbf{a}_H^u(\boldsymbol{\theta})^H d^2\boldsymbol{\theta}, \\
\mathbf{R}_{TX}^{u} &= \int_{\Omega} \gamma_V(\boldsymbol{\phi}) \mathbf{b}_V^u(\boldsymbol{\phi}) \mathbf{b}_V^u(\boldsymbol{\phi})^H d^2\boldsymbol{\phi} + \int_{\Omega} \gamma_H(\boldsymbol{\phi}) \mathbf{b}_H^u(\boldsymbol{\phi}) \mathbf{b}_H^u(\boldsymbol{\phi})^H d^2\boldsymbol{\phi}, \\
\mathbf{R}_{RX}^{d} &= \int_{\Omega} \gamma_V(\boldsymbol{\phi}) \mathbf{b}_V^d(\boldsymbol{\phi}) \mathbf{b}_V^d(\boldsymbol{\phi})^H d^2\boldsymbol{\phi} + \int_{\Omega} \gamma_H(\boldsymbol{\phi}) \mathbf{b}_H^d(\boldsymbol{\phi}) \mathbf{b}_H^d(\boldsymbol{\phi})^H d^2\boldsymbol{\phi}, \\
\mathbf{R}_{TX}^{d} &= \int_{\Omega} \rho_V(\boldsymbol{\theta}) \mathbf{a}_V^d(\boldsymbol{\theta}) \mathbf{a}_V^d(\boldsymbol{\theta})^H d^2\boldsymbol{\theta} + \int_{\Omega} \rho_H(\boldsymbol{\theta}) \mathbf{a}_H^d(\boldsymbol{\theta}) \mathbf{a}_H^d(\boldsymbol{\theta})^H d^2\boldsymbol{\theta},
\end{aligned} \tag{2.32}$$

where the functions $\rho_V, \rho_H, \gamma_V, \gamma_H : \Omega \to \mathbb{R}^+$ are defined as

$$\begin{aligned}
\rho_V(\boldsymbol{\theta}) &:= \sum_{c=1}^{N_c} \alpha_c \int_{\Omega} f_c(\boldsymbol{\theta}, \boldsymbol{\phi}) \left(\|\mathbf{b}_V(\boldsymbol{\phi})\|^2 + \frac{1}{K} \|\mathbf{b}_H(\boldsymbol{\phi})\|^2 \right) d^2\boldsymbol{\phi}, \\
\rho_H(\boldsymbol{\theta}) &:= \sum_{c=1}^{N_c} \alpha_c \int_{\Omega} f_c(\boldsymbol{\theta}, \boldsymbol{\phi}) \left(\|\mathbf{b}_H(\boldsymbol{\phi})\|^2 + \frac{1}{K} \|\mathbf{b}_V(\boldsymbol{\phi})\|^2 \right) d^2\boldsymbol{\phi}, \\
\gamma_V(\boldsymbol{\phi}) &:= \sum_{c=1}^{N_c} \alpha_c \int_{\Omega} f_c(\boldsymbol{\theta}, \boldsymbol{\phi}) \left(\|\mathbf{a}_V(\boldsymbol{\theta})\|^2 + \frac{1}{K} \|\mathbf{a}_H(\boldsymbol{\theta})\|^2 \right) d^2\boldsymbol{\theta}, \\
\gamma_H(\boldsymbol{\phi}) &:= \sum_{c=1}^{N_c} \alpha_c \int_{\Omega} f_c(\boldsymbol{\theta}, \boldsymbol{\phi}) \left(\|\mathbf{a}_H(\boldsymbol{\theta})\|^2 + \frac{1}{K} \|\mathbf{a}_V(\boldsymbol{\theta})\|^2 \right) d^2\boldsymbol{\theta},
\end{aligned}$$

and where $1/K := \mathbb{E}[1/K_{ic}]$ describes the average effect of the XPRs K_{ic}, and we also define the following functions:

$$\begin{aligned}
\|\mathbf{b}_V(\boldsymbol{\phi})\|^2 &:= \|\mathbf{b}_V^u(\boldsymbol{\phi})\|^2 = \|\mathbf{b}_V^d(\boldsymbol{\phi})\|^2, \\
\|\mathbf{b}_H(\boldsymbol{\phi})\|^2 &:= \|\mathbf{b}_H^u(\boldsymbol{\phi})\|^2 = \|\mathbf{b}_H^d(\boldsymbol{\phi})\|^2, \\
\|\mathbf{a}_V(\boldsymbol{\phi})\|^2 &:= \|\mathbf{a}_V^u(\boldsymbol{\phi})\|^2 = \|\mathbf{a}_V^d(\boldsymbol{\phi})\|^2, \\
\|\mathbf{a}_H(\boldsymbol{\phi})\|^2 &:= \|\mathbf{a}_H^u(\boldsymbol{\phi})\|^2 = \|\mathbf{a}_H^d(\boldsymbol{\phi})\|^2,
\end{aligned}$$

which are assumed to be frequency invariant because they do not depend on the phase response of the array, but just on its magnitude (i.e. the radiation pattern), which is generally assumed to be frequency independent.

Proof. (Sketch) Let us drop the UL and DL superscripts for simplicity. By focusing on the expression for the matrix $\mathbb{E}[\mathbf{H}\mathbf{H}^H]$, the proof is identical to the one presented in Section 2.4, except we substitute $b_V(\cdot)^2$ and $b_H(\cdot)^2$ respectively with $\|\mathbf{b}_V(\boldsymbol{\phi})\|^2$ and $\|\mathbf{b}_H(\boldsymbol{\phi})\|^2$. Similarly, the expression for the matrix $\mathbb{E}[\mathbf{H}^H\mathbf{H}]$ is obtained by simply flipping the role of the UE and BTS array responses. By applying the definitions in (2.29), the expressions (2.32) are finally proved. □

2.6 Summary of Spatial Covariance Matrix Expressions

This chapter showed that, for many different channel models, the spatial covariance matrix $\mathbf{R}$ can be decomposed into the following two terms, combined via integration:

- A term that is frequency invariant and slowly time-varying, and it is denoted in general as APS, which describes the distribution of the average power in the angular domain.
- A term that is frequency dependent and constant in time, which describes the BTS antenna array response in the angular domain (both radiation pattern and phase response).

The frequency invariance property of the APS defines the angular reciprocity assumption outlined in Section 2.1.2. The slow time variance of the APS instead falls into the WSS assumption described in Section 2.1.1.

In Section 2.2 and 2.3 we showed that this decomposition is possible either by assuming a discrete or a continuous scattering modeling philosophy, and that it is not influenced by whether the flat-fading model comes from a narrow-band system or from a wide-band OFDM system. Furthermore, in Section 2.4 we showed that, with respect to a simplified 2D channel model, the extension to 3D environments is simply done by considering the double integration of bivariate (azimuth and zenith) functions instead of the single integration of azimuth-only dependent functions. Section 2.4 also shows that this decomposition is also possible with realistic antenna radiation patterns. Finally, the use of dual-polarized antennas is shown in Section 2.4 to introduce a finer decomposition in terms of the different antenna radiation patterns for the two orthogonal polarizations.

The expressions for $\mathbf{R}^d$ and $\mathbf{R}^u$, focusing on 3D modeling, are summarized in Table 2.1, and they hold for both narrow-band and wide-band OFDM systems as well as for both discrete and continuous scattering models. Table 2.1 also includes the matrices $\mathbf{R}^d_{TX}$ and $\mathbf{R}^u_{RX}$ defined in Section 2.5 for the multi-antenna UE scenario.

Table 2.1: Summary of covariance matrix expressions

	Unpolarized	Dual-polarized
$\mathbf{R}^d$, $\mathbf{R}^d_{TX}$	$\int_\Omega \rho(\boldsymbol{\theta})\mathbf{a}^d(\boldsymbol{\theta})\mathbf{a}^d(\boldsymbol{\theta})^H d^2\boldsymbol{\theta}$	$\int_\Omega \rho_V(\boldsymbol{\theta})\mathbf{a}^d_V(\boldsymbol{\theta})\mathbf{a}^d_V(\boldsymbol{\theta})^H d^2\boldsymbol{\theta}$ $+ \int_\Omega \rho_H(\boldsymbol{\theta})\mathbf{a}^d_H(\boldsymbol{\theta})\mathbf{a}^d_H(\boldsymbol{\theta})^H d^2\boldsymbol{\theta}$
$\mathbf{R}^u$, $\mathbf{R}^u_{RX}$	$\int_\Omega \rho(\boldsymbol{\theta})\mathbf{a}^u(\boldsymbol{\theta})\mathbf{a}^u(\boldsymbol{\theta})^H d^2\boldsymbol{\theta}$	$\int_\Omega \rho_V(\boldsymbol{\theta})\mathbf{a}^u_V(\boldsymbol{\theta})\mathbf{a}^u_V(\boldsymbol{\theta})^H d^2\boldsymbol{\theta}$ $+ \int_\Omega \rho_H(\boldsymbol{\theta})\mathbf{a}^u_H(\boldsymbol{\theta})\mathbf{a}^u_H(\boldsymbol{\theta})^H d^2\boldsymbol{\theta}$

Although the definitions of the APS (or V-APS, H-APS) varies according to the choice of the channel model, its physical and mathematical properties are shared (e.g., it is a power density, so it is positive and real valued).

Chapter 3

Channel Estimation: The Large-scale Array Regime

In conventional systems (e.g. 4G systems), the BTS acquires the channel state information (CSI) by using part of the available DL resources to transmit known reference signals (also called "pilots") to the UEs. These pilots are used by the UEs to estimate their channels, which are then fed back to the BTS on the reverse links. To reduce the feedback overhead, transceivers use a pre-defined codebook for CSI quantization [16]. The main assumption of these systems is that the channel remains constant for a time interval large enough to allow CSI acquisition and data transmission in the same coherence block (block-fading assumption, see Section 2.1.1). In practice, this assumption is crude, and the BTS receives an outdated version of the CSI. However, many techniques able to deal with this problem have been proposed, and the idea is to exploit prior information that is carried by the delayed CSI [17, 18].

In this chapter, we study the impact of the DL training overhead by ignoring the performance degradation caused by quantized and delayed CSIT. In particular, we show that, even under such ideal assumptions, the main problem of techniques used in current commercial systems is the use of pilot sequences with length T at least equal to the number of BTS antennas N. Briefly, as shown in the example below, in the large-scale system array regime (i.e., when N is comparable to the coherence block length M_c), the training overhead becomes so large that data transmission becomes unfeasible.

Example: Consider a typical 2 GHz LTE system with coherence bandwidth $B_c = 100$ kHz and coherence time $T_c = 1$ ms (UE speed of roughly 60 km/h). The corresponding coherence block length is $M_c = B_c T_c = 100$ time slots. Thus, a typical massive MIMO array equipped with $N = 64$ antennas would spend 64% of the coherence block just for channel estimation.

The goal of this chapter is to show how it is possible to exploit correlation properties of the channel to reduce the channel estimation effort.

This chapter is organized as follows:

- In Section 3.1 we start the analysis with a review of conventional channel estimation techniques
- In Section 3.2 we show that the main reason for using pilot sequences with length $T \geq N$ is justified by the i.i.d. fading assumption.
- However, in Sections 3.3 and 3.4 we show that, with correlated fading, and if the DL channel covariance matrix is ill-conditioned or rank deficient, the knowledge of this matrix at the BTS side can greatly reduce the overhead for CSI acquisition. We also briefly discuss how channel correlation plays a fundamental role when applying compressed sensing based channel estimation algorithms.
- Finally, in Section 3.5 we look at the problem from an Information Theory perspective. In particular, by using the concept of Degrees of Freedom (DoF), we provide another confirmation that the choice of using $T \geq N$ is deeply rooted in the i.i.d. fading assumption, and we give a preliminary outline of the potential benefits of releasing this assumption.

3.1 Review of Conventional Channel Estimation Techniques

3.1.1 Downlink Channel Estimation

Let us consider the DL MU-MIMO model in (2.2), and let us further focus on the part of the coherence block $m = 0, \ldots, T < M_c$, which is devoted to channel estimation. The received signal at a given UE for the entire estimation block, denoted by $\mathbf{y} \in \mathbb{C}^{T \times 1}$, can be rewritten in matrix form as

$$\mathbf{y} = \mathbf{X}\mathbf{h}^d + \mathbf{w}, \tag{3.1}$$

where $\mathbf{h}^d \in \mathbb{C}^{N \times 1}$ is the UE channel vector, $\mathbf{X} \in \mathbb{C}^{T \times N}$ is the transmitted pilot sequence at the BTS, and $\mathbf{w} \in \mathbb{C}^{T \times 1}$ is the noise. For simplicity, in the following we drop the DL superscript.

With a conventional linear estimator, the channel estimate is given by

$$\hat{\mathbf{h}} = \mathbf{F}\mathbf{y}, \tag{3.2}$$

where $\mathbf{F} \in \mathbb{C}^{N \times T}$ is the linear estimator to be designed. The most popular linear estimators are the least square (LS) estimator and the minimum mean square error (LMMSE) estimator, and they are given by the well known expressions

$$\begin{aligned} \mathbf{F}_{\text{LS}} &= \mathbf{X}^{\dagger}, \\ \mathbf{F}_{\text{LMMSE}} &= \mathbf{R}_{\mathbf{hy}}{\mathbf{R}_{\mathbf{yy}}}^{-1}, \end{aligned} \tag{3.3}$$

where $\mathbf{R}$ is the channel spatial covariance matrix as defined in section 2.1, $(\cdot)^{\dagger}$ denotes the Moore–Penrose pseudo-inverse, and where

$$\begin{aligned} \mathbf{R_{hy}} &:= \mathbb{E}[\mathbf{h}\mathbf{y}^H] = \mathbf{R}\mathbf{X}^H = \mathbf{R_{yh}}^H, \\ \mathbf{R_{yy}} &:= \mathbb{E}[\mathbf{y}\mathbf{y}^H] = \mathbf{X}\mathbf{R}\mathbf{X}^H + \sigma_w^2\mathbf{I}. \end{aligned} \tag{3.4}$$

Throughout the remaing parts of this chapter the noise variance σ_w^2 will be given in terms of a per antenna average SNR, defined as follows:

$$\text{SNR} := \frac{P_h P_x}{\sigma_w^2}, \tag{3.5}$$

where $P_x := \dfrac{\|\mathbf{X}\|_F}{NT}$ is the per antenna average transmit power, and $P_h := \dfrac{\text{tr}\{\mathbf{R}\}}{N}$ is the per antenna average channel power.

3.1.2 Uplink Channel Estimation

Let us consider the UL MU-MIMO model in (2.1), and let us further focus on the part of the coherence block $m = 0, \ldots, T < M_c$, which is devoted to channel estimation. The received signal at the BTS for the entire estimation block, denoted by $\mathbf{Y} \in \mathbb{C}^{N \times T}$, can be rewritten in matrix form as

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{W}, \tag{3.6}$$

where $\mathbf{H} := \begin{bmatrix} \mathbf{h}_1^u & \ldots & \mathbf{h}_K^u \end{bmatrix} \in \mathbb{C}^{N \times K}$ is the equivalent MIMO channel that groups all the UEs channels $\mathbf{h}_k^u \in \mathbb{C}^{N \times 1}$, $\mathbf{X} := \begin{bmatrix} \mathbf{x}_1 & \ldots & \mathbf{x}_K \end{bmatrix}^T \in \mathbb{C}^{K \times T}$ collects all the UEs pilot sequences $\mathbf{x}_k \in \mathbb{C}^{T \times 1}$, and $\mathbf{W} \in \mathbb{C}^{N \times T}$ is the noise.

To deal with multiple users, in the following we assume a typical solution based on orthogonal pilot sequences, with the assumption that the UEs are perfectly synchronized.

Let us define a set of K pilot sequences $\{\mathbf{x}_k, k = 1 \ldots K\}$ such that the following orthogonality property is satisfied:

$$\mathbf{x}_l^H \mathbf{x}_k = \begin{cases} \|\mathbf{x}_k\|^2, & \text{if } k = l \\ 0, & \text{otherwise} \end{cases}.$$

With this pilot design, the optimal channel estimator in terms of MSE and SNR (defined in the following) is the matched filter:

$$\hat{\mathbf{h}}_k^u = \mathbf{Y}\frac{\mathbf{x}_k^*}{\|\mathbf{x}_k\|^2}, \tag{3.7}$$

which results in a mean square error of

$$\begin{aligned} \mathrm{MSE}_k &:= \mathbb{E}\left[\|\mathbf{h}_k^u - \hat{\mathbf{h}}_k^u\|^2\right] \\ &= \frac{\mathbf{x}_k^T}{\|\mathbf{x}_k\|^2}\mathbb{E}\left[\mathbf{W}^H\mathbf{W}\right]\frac{\mathbf{x}_k^*}{\|\mathbf{x}_k\|^2} \\ &= \frac{\sigma_w^2}{\|\mathbf{x}_k\|^2} = \frac{\sigma_w^2}{P_k T} =: \frac{1}{\mathrm{SNR}_k} \end{aligned} \tag{3.8}$$

where we denote by P_k the average transmit power per time slot available at the kth UE.

Minimum Pilot Sequence Length By assuming a synchronous system with orthogonal pilots as described above, the pilot sequence length must satisfy $T \geq K$. The proof follows immediately from the definition of orthogonal vectors: a necessary condition for K vectors to be orthogonal is that they must be is a space of dimension of at least K.

3.2 Conventional Estimation of i.i.d Channels

3.2.1 Minimum Pilot Sequence Length for i.i.d. Channels

Conventional DL pilot based channel estimators are usually designed by imposing $T \geq N$, which can be intuitively explained by the common wisdom that, in order to meaningfully estimate N independent parameters, we need at least N observations. A more precise motivation for this design choice is given in the following two propositions:

Proposition 3.1. *In case of i.i.d. fading, i.e. with* $\mathbf{R} = \sigma_h^2\mathbf{I}$*, and given a pilot sequence* $\mathbf{X}$ *of length* T*, the mean square error* $\mathrm{MSE} := \mathbb{E}[\|\mathbf{h} - \hat{\mathbf{h}}\|^2]$ *of the LS estimator defined in section 3.1.1 is lower bounded by*

$$\mathrm{MSE} \geq \sigma_h^2 \max(N - T, 0) + \sum_{\lambda_i \neq 0} \frac{\sigma_w^2}{\lambda_i^2}, \tag{3.9}$$

where λ_i^2, $i = 1, \ldots N$, are the eigenvalues of the matrix $\mathbf{Q} = \mathbf{X}^H\mathbf{X}$. Furthermore, the equality in 3.9 is achieved if and only if $\mathrm{rank}\{\mathbf{X}\} = \min(N, T)$.

Proof. Let us consider the MSE expressed as $\mathbb{E}[\|\mathbf{h} - \hat{\mathbf{h}}\|^2] = \mathrm{tr}\{\mathbf{R}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}\}$, where $\mathbf{R}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}$ denotes the covariance matrix of the estimation error, given by

$$\begin{aligned} \mathbf{R}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}} &:= \mathbb{E}[(\mathbf{h} - \hat{\mathbf{h}})(\mathbf{h} - \hat{\mathbf{h}})^H] \\ &= (\mathbf{X}^\dagger\mathbf{X} - \mathbf{I})\mathbf{R}(\mathbf{X}^\dagger\mathbf{X} - \mathbf{I})^H + \sigma_w^2\mathbf{X}^\dagger(\mathbf{X}^\dagger)^H \\ &= (\mathbf{X}^\dagger\mathbf{X} - \mathbf{I})\sigma_h^2(\mathbf{X}^\dagger\mathbf{X} - \mathbf{I}) + \sigma_w^2(\mathbf{X}^H\mathbf{X})^\dagger, \end{aligned} \tag{3.10}$$

where we applied the well-known identities $\mathbf{X}^\dagger\mathbf{X} = (\mathbf{X}^\dagger\mathbf{X})^H$ and $\mathbf{X}^\dagger(\mathbf{X}^\dagger)^H = (\mathbf{X}^H\mathbf{X})^\dagger$. By defining the $N \times N$ PSD matrix $\mathbf{Q} := \mathbf{X}^H\mathbf{X}$ and its eigenvalues decomposition $\mathbf{Q} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, $\mathbf{\Lambda} = \text{diag}\{\lambda_1^2, \ldots, \lambda_N^2\}$, the MSE for the LS estimator can be expressed as

$$\begin{aligned}
\text{MSE} &= \text{tr}\left\{(\mathbf{X}^\dagger\mathbf{X} - \mathbf{I})\sigma_h^2(\mathbf{X}^\dagger\mathbf{X} - \mathbf{I}) + \sigma_w^2(\mathbf{X}^H\mathbf{X})^\dagger\right\} \\
&= \sigma_h^2\text{tr}\left\{\mathbf{X}^\dagger\mathbf{X}\mathbf{X}^\dagger\mathbf{X} - 2\mathbf{X}^\dagger\mathbf{X} + \mathbf{I}\right\} + \text{tr}\left\{\sigma_w^2\mathbf{Q}^\dagger\right\} \\
&= \sigma_h^2\text{tr}\left\{\mathbf{X}^\dagger\mathbf{X} - 2\mathbf{X}^\dagger\mathbf{X} + \mathbf{I}\right\} + \text{tr}\left\{\sigma_w^2(\mathbf{U}\mathbf{\Lambda}\mathbf{U}^H)^\dagger\right\} \\
&= \sigma_h^2\left(\text{tr}\left\{\mathbf{I}\right\} - \text{tr}\left\{\mathbf{P}_{\mathbf{X}^H}\right\}\right) + \sigma_w^2\text{tr}\left\{\mathbf{U}\Lambda^\dagger\mathbf{U}^H\right\} \\
&= \sigma_h^2\left(N - \text{rank}\{\mathbf{P}_{\mathbf{X}^H}\}\right) + \sigma_w^2\text{tr}\left\{\mathbf{U}^H\mathbf{U}\Lambda^\dagger\right\} \\
&= \sigma_h^2\left(N - \text{rank}\{\mathbf{P}_{\mathbf{X}^H}\}\right) + \sum_{\lambda_i \neq 0}\frac{\sigma_w^2}{\lambda_i^2},
\end{aligned}$$

where we applied the well-known identities $\mathbf{X}^\dagger\mathbf{X}\mathbf{X}^\dagger\mathbf{X} = \mathbf{X}^\dagger\mathbf{X}$, $\mathbf{X}^\dagger\mathbf{X} = \mathbf{P}_{\mathbf{X}^H}$, with $\mathbf{P}_{\mathbf{X}^H}$ denoting the projection matrix on the column space of $\mathbf{X}^H$, and where $\mathbf{\Lambda}^\dagger$ is a diagonal matrix with diagonal elements

$$\tilde{\lambda}_i^2 = \begin{cases} \dfrac{1}{\lambda_i^2} & \text{if } \lambda_i^2 > 0 \\ 0 & \text{otherwise} \end{cases}, \quad i = 1, \ldots, N.$$

Since $\text{rank}\{\mathbf{P}_{\mathbf{X}^H}\} = \text{rank}\{\mathbf{X}^H\} \leq \min(N, T)$, the MSE can be lower bounded as

$$\begin{aligned}
\text{MSE} &\geq \sigma_h^2\left(N - \min(N, T)\right) + \sum_{\lambda_i \neq 0}\frac{\sigma_w^2}{\lambda_i^2} \\
&= \sigma_h^2\max(N - T, 0) + \sum_{\lambda_i \neq 0}\frac{\sigma_w^2}{\lambda_i^2}
\end{aligned}$$

□

Proposition 3.2. *In case of i.i.d. fading, i.e. with* $\mathbf{R} = \sigma_h^2\mathbf{I}$*, and given a pilot sequence* $\mathbf{X}$ *of length* T*, the mean square error* $\text{MSE} := \mathbb{E}[\|\mathbf{h} - \hat{\mathbf{h}}\|^2]$ *of the LMMSE estimator defined in section 3.1.1 is lower bounded by*

$$\text{MSE} \geq \sigma_h^2\max(N - T, 0) + \sum_{\lambda_i \neq 0}\frac{1}{\sigma_h^{-2} + \dfrac{\lambda_i^2}{\sigma_w^2}}, \tag{3.11}$$

where λ_i^2*,* $i = 1, \ldots N$*, are the eigenvalues of the matrix* $\mathbf{Q} = \mathbf{X}^H\mathbf{X}$*. Furthermore, the equality in 3.11 is achieved if and only if* $\text{rank}\{\mathbf{X}\} = \min(N, T)$*.*

Proof. Let us consider the MSE expressed as $\mathbb{E}[\|\mathbf{h}-\hat{\mathbf{h}}\|^2] = \mathrm{tr}\{\mathbf{R}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}\}$, where $\mathbf{R}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}$ denotes the covariance matrix of the estimation error, given by

$$\begin{aligned}\mathbf{R}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}} &:= \mathbb{E}[(\mathbf{h}-\hat{\mathbf{h}})(\mathbf{h}-\hat{\mathbf{h}})^H] \\ &= \mathbf{R} - \mathbf{R}_{\mathbf{hy}}{\mathbf{R}_{\mathbf{yy}}}^{-1}\mathbf{R}_{\mathbf{yh}} \\ &= \mathbf{R} - \mathbf{R}\mathbf{X}^H\left(\mathbf{X}\mathbf{R}\mathbf{X}^H + \sigma_w^2\mathbf{I}\right)^{-1}\mathbf{X}\mathbf{R} \\ &= \left(\mathbf{R}^{-1} + \sigma_w^{-2}\mathbf{X}^H\mathbf{X}\right)^{-1},\end{aligned}$$

where for the last equality we used the matrix inversion lemma. In case of i.i.d. fading, $\mathbf{R}$ is a scaled identity. By defining the $N \times N$ PSD matrix $\mathbf{Q} := \mathbf{X}^H\mathbf{X}$ and its eigenvalues decomposition $\mathbf{Q} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, $\mathbf{\Lambda} = \mathrm{diag}\{\lambda_1^2, \ldots, \lambda_N^2\}$, the MSE for the LMMSE estimator can be expressed as

$$\begin{aligned}\mathrm{MSE} &= \mathrm{tr}\left\{\left(\sigma_h^{-2}\mathbf{I} + \sigma_w^{-2}\mathbf{Q}\right)^{-1}\right\} \\ &= \mathrm{tr}\left\{\left(\mathbf{U}(\sigma_h^{-2}\mathbf{I} + \sigma_w^{-2}\mathbf{\Lambda})\mathbf{U}^H\right)^{-1}\right\} \\ &= \mathrm{tr}\left\{\mathbf{U}(\sigma_h^{-2}\mathbf{I} + \sigma_w^{-2}\mathbf{\Lambda})^{-1}\mathbf{U}^H\right\} \\ &= \mathrm{tr}\left\{\mathbf{U}^H\mathbf{U}(\sigma_h^{-2}\mathbf{I} + \sigma_w^{-2}\mathbf{\Lambda})^{-1}\right\} \\ &= \sum_{i=1}^{N}\frac{1}{\sigma_h^{-2} + \dfrac{\lambda_i^2}{\sigma_w^2}},\end{aligned} \tag{3.12}$$

By definition, $\mathbf{Q}$ has N' zero eigenvalues, with $N' = N - \mathrm{rank}\{\mathbf{X}\} \geq N - \min(N,T) = \max(N-T,0)$, thus the MSE can be further expressed as

$$\mathrm{MSE} = \sigma_h^2 N' + \sum_{\lambda_i \neq 0}\frac{1}{\sigma_h^{-2} + \dfrac{\lambda_i^2}{\sigma_w^2}} \geq \sigma_h^2 \max(N-T,0) + \sum_{\lambda_i \neq 0}\frac{1}{\sigma_h^{-2} + \dfrac{\lambda_i^2}{\sigma_w^2}}.$$

□

Remark 3.1. *In the high SNR regime, the LMMSE estimator reduces to a projector onto the row space of* $\mathbf{X}$ *of dimension* $\dim(\mathbf{X}) \leq \min(N-T, N)$*. More precisely, by letting* $\sigma_w^2 \to \infty$*, we have*

$$\begin{aligned}\hat{\mathbf{h}} &= \mathbf{F}_{\mathrm{LMMSE}}\mathbf{y} \\ &= \mathbf{R}_{\mathbf{hy}}{\mathbf{R}_{\mathbf{yy}}}^{-1}\mathbf{X}\mathbf{h} \\ &= \mathbf{X}^H\left(\mathbf{X}\mathbf{X}^H\right)^{-1}\mathbf{X}\mathbf{h} \\ &= \mathbf{P}_{\mathbf{X^H}}\mathbf{h}\end{aligned} \tag{3.13}$$

where $\mathbf{P}_{\mathbf{X^H}}$ *denotes the projection matrix onto the column space of* $\mathbf{X}^H$*.*

As an immediate consequence of 3.1 and 3.2, we have the following result:

Corollary 3.1. *A necessary and sufficient condition for the MSE of both the LS and LMMSE estimator to vanish as the SNR increases is to let* $\mathbf{X}$ *to be a full-rank pilot sequence of length at least* N*. More precisely:*

$$\lim_{\sigma_w^2 \to 0} \text{MSE} = 0 \iff \text{rank}\{\mathbf{X}\} = N.$$

3.2.2 Pilot Sequence Design for i.i.d. Channels

In this section we derive the optimal structure of the pilot sequences in case of i.i.d. channels. Interestingly, the optimal pilot design is shown in practice to satisfy the minimum pilot sequence criteria derived in section 3.2.1 (given an additional mild condition on the SNR level for the LS estimator only).

Let us constrain the overall energy of the training sequence as

$$\|\mathbf{X}\|_F^2 \leq NT, \tag{3.14}$$

so that the average power per antenna and per time slot is at most unitary. We adopt the definition of SNR (3.5) by setting $P_x = 1$. We further define the parameter $\nu := T/N$, and the matrix $\mathbf{Q} := \mathbf{X}^H\mathbf{X}$ as well as its eigenvalues decomposition $\mathbf{Q} = \mathbf{U}\mathbf{\Lambda}\mathbf{U}^H$, $\mathbf{\Lambda} = \text{diag}\{\lambda_1^2, \ldots, \lambda_N^2\}$. Let us optimize the training sequence for both LMMSE and LS estimators by considering the MSE as cost function.

Proposition 3.3. *In case of i.i.d. fading, i.e. with* $\mathbf{R} = \sigma_h^2\mathbf{I}$*, the optimal training sequence in terms of MSE for LMMSE channel estimation is any sequence satisfying*

$$\lambda_i^2 = T, \quad \forall i = 1 \ldots N,$$

resulting in

$$\text{MSE} = \frac{\sigma_h^2}{\dfrac{1}{N} + \nu\text{SNR}}.$$

Proof. (sketch) The proof follows by applying standard Lagrangian minimization to the convex cost function (3.12) of variables $c_i := \lambda_i^2$, under the convex constraint (3.14) expressed as $0 \leq \sum_{i=1}^N c_i \leq NT$. The existence of the minimum is guaranteed by the Weierstrass theorem. □

Proposition 3.4. *In case of i.i.d. fading, i.e. with* $\mathbf{R} = \sigma_h^2\mathbf{I}$*, and by assuming* $\text{SNR} \geq \frac{2}{T}$*, the optimal training sequence in terms of MSE for the LS channel estimator is any sequence satisfying*

$$\lambda_i^2 = T, \quad \forall i = 1 \ldots N$$

resulting in

$$\text{MSE} = \frac{\sigma_h^2}{\nu\text{SNR}}.$$

Proof. (sketch) Let us focus on the term

$$\sum_{\lambda_i \neq 0} \frac{\sigma_w^2}{\lambda_i^2}$$

of the bound (3.9), which is a convex cost function of the $r := \text{rank}(\mathbf{X})$ variables $\{c_i = \lambda_i^2 : \lambda_i^2 \geq \epsilon > 0\}$, with ϵ arbitrarily small. By applying standard Lagrangian minimization to this convex cost function, under the convex constraint (3.14) expressed as $r\epsilon \leq \sum_{i:\lambda_i \neq 0} c_i \leq NT$, we obtain the optimal MSE

$$\begin{aligned}
\text{MSE} &= \sigma_h^2 (N - r) + \sum_{\lambda_i \neq 0} \frac{\sigma_w^2 r}{NT} \\
&= \sigma_h^2 (N - r) + \frac{\sigma_w^2 r^2}{NT} \\
&= \text{tr}\{\mathbf{R}\} \left[1 - \frac{r}{N} + \frac{\sigma_w^2 r^2}{\text{tr}\{\mathbf{R}\} NT} \right] \\
&= \text{tr}\{\mathbf{R}\} \left[1 - \frac{r}{N} + \frac{r^2}{\text{SNR} N^2 T} \right].
\end{aligned}$$

The existence of the minimum is guaranteed by the Weierstrass theorem, and it is attained by setting c_i (i.e. the non-zero eigenvalues λ_i^2) to $\frac{NT}{r}$.

Let us now analyze the behaviour of the MSE as the rank r of the training sequence varies. The function

$$f(t) = 1 - \frac{t}{N} + \frac{t^2}{\text{SNR} N^2 T}, \quad t \in \mathbf{R},$$

is a parabola, monotonically decreasing only for $t \in [-\infty, t_0]$, where t_0 is the unique critical point computed as

$$\frac{d}{dt} f(t_0) = 0 \iff t_0 = \frac{NT\text{SNR}}{2}.$$

Therefore, by considering the maximum possible rank $r_{\max} = N$, and by assuming the following condition on the SNR level

$$N \leq \frac{NT\text{SNR}}{2} \implies \text{SNR} \geq \frac{2}{T},$$

the MSE always decreases as the rank r increases. Hence, under this assumption on the SNR, the MSE is minimized by letting $\mathbf{X}$ to be a full rank pilot sequence of length $T \geq N$, with eigenvalues λ_i such that $\lambda_i^2 = T$, $\forall i = 1, \ldots, N$. □

Remark 3.2. *The additional assumption on the SNR level required by proposition 3.4 with respect to proposition 3.3 can be explained by the well-known fact that the LS estimator, as opposed to the LMMSE estimator, is not designed to take into account noise. We further point out that this assumption on the SNR level is typically met in practical applications.*

3.3 Conventional Estimation of Correlated Channels

In this section, we extend the analysis of the conventional channel estimation techniques described in Section 3.1.1 to the case of correlated MIMO channels.

3.3.1 Minimum Pilot Sequence Length (SU-MIMO)

In the following, we show that, for a channel with covariance matrix $\mathbf{R}$, it is possible to reduce the channel estimation overhead down to $T = \text{rank}\{\mathbf{R}\} \leq N$ while guaranteeing the estimation error to vanish as the SNR increases. More precisely, we have the following propositions:

Proposition 3.5. *Given the model in (3.1), a channel with spatial covariance matrix* $\mathbf{R}$*, and the eigen-decomposition* $\mathbf{R} = \mathbf{V}\mathbf{\Sigma}\mathbf{V}^H$*,* $\mathbf{\Sigma} = \text{diag}\{\sigma_1^2, \ldots, \sigma_N^2\}$*, the mean square error* $\text{MSE} := \mathbb{E}[\|\mathbf{h} - \hat{\mathbf{h}}\|^2]$ *of a LS channel estimate* $\hat{\mathbf{h}}$ *defined in Section 3.1.1 is given by*

$$\text{MSE} = \frac{c_1}{\text{SNR}} + c_2, \tag{3.15}$$

where $\text{SNR} := P_h P_x / \sigma_w^2$ *as defined in (3.5), and where the constants* c_1 *and* f_2 *are defined as follows:*

$$c_1 := \sum_{\lambda_i \neq 0} \frac{P_h P_x}{\lambda_i^2} \geq 0, \tag{3.16}$$

where $\lambda_1^2 \geq \ldots \geq \lambda_N^2$ *are the eigenvalues of* $\mathbf{Q} := \mathbf{X}^H\mathbf{X}$*, and*

$$c_2 := \left\| \mathbf{P}^{\perp}_{\mathbf{X}^H} \mathbf{V}\mathbf{\Sigma}^{\frac{1}{2}} \right\|_F^2 \geq 0, \tag{3.17}$$

where $\mathbf{P}^{\perp}_{\mathbf{X}^H}$ *denotes the projection matrix onto the subspace orthogonal to the column space of* $\mathbf{X}^H$*. Furthermore, we have that*

$$c_2 = 0 \iff \text{span}\{\mathbf{X}^H\} \supseteq \text{span}\{\mathbf{V}\mathbf{\Sigma}\}.$$

(We use the convention that $\text{span}\{\cdot\}$ *denotes the column span of a matrix.) Thus, a necessary condition for having* $c_2 = 0$ *is* $\text{rank}\{\mathbf{X}^H\} \geq \text{rank}\{\mathbf{R}\}$*, which implies* $T \geq \text{rank}\{\mathbf{R}\}$.

Proof. By (3.10), we obtain

$$\begin{aligned}
\mathbf{R}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}} :&= \mathbb{E}[(\mathbf{h} - \hat{\mathbf{h}})(\mathbf{h} - \hat{\mathbf{h}})^H] \\
&= (\mathbf{X}^\dagger\mathbf{X} - \mathbf{I})\mathbf{R}(\mathbf{X}^\dagger\mathbf{X} - \mathbf{I}) + \sigma_w^2(\mathbf{X}^H\mathbf{X})^\dagger \\
&= \mathbf{X}^\dagger\mathbf{X}\mathbf{R}\mathbf{X}^\dagger\mathbf{X} - \mathbf{X}^\dagger\mathbf{X}\mathbf{R} - \mathbf{R}\mathbf{X}^\dagger\mathbf{X} + \mathbf{R} + \sigma_w^2(\mathbf{X}^H\mathbf{X})^\dagger,
\end{aligned}$$

thus, by following the same lines as in Section 3.1.1, the MSE can be expressed as

$$\begin{aligned}
\text{MSE} &= \text{tr}\left\{\mathbf{R}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}\right\} \\
&= \text{tr}\{\mathbf{R}\} - \text{tr}\left\{\mathbf{X}^{\dagger}\mathbf{X}\mathbf{R}\right\} + \sigma_w^2 \text{tr}\left\{\mathbf{U}\Lambda^{\dagger}\mathbf{U}^H\right\} \\
&= \text{tr}\left\{(\mathbf{I} - \mathbf{X}^{\dagger}\mathbf{X})\mathbf{R}\right\} + \sum_{\lambda_i \neq 0} \frac{\sigma_w^2}{\lambda_i^2} \\
&= \text{tr}\left\{\mathbf{P}^{\perp}_{\mathbf{X}^H}\mathbf{R}\right\} + \sum_{\lambda_i \neq 0} \frac{\sigma_w^2}{\lambda_i^2} \\
&= \text{tr}\left\{\mathbf{P}^{\perp}_{\mathbf{X}^H}\mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{V}^H\mathbf{P}^{\perp}_{\mathbf{X}^H}\right\} + \sum_{\lambda_i \neq 0} \frac{\sigma_w^2}{\lambda_i^2} \\
&= \left\|\mathbf{P}^{\perp}_{\mathbf{X}^H}\mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}}\right\|_F^2 + \sum_{\lambda_i \neq 0} \frac{\sigma_w^2}{\lambda_i^2},
\end{aligned} \tag{3.18}$$

where $\mathbf{P}^{\perp}_{\mathbf{X}^H}$ denotes the projection matrix[1] onto the subspace orthogonal to the column space space of $\mathbf{X}^H$, and where we defined the eigenvalue decomposition $\mathbf{X}^H\mathbf{X} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$, $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1^2, \ldots, \lambda_N^2\}$. By defining $\text{SNR} := P_h P_x/\sigma_w^2$ as in (3.5) and

$$c_1 := \sum_{\lambda_i \neq 0} \frac{P_h P_x}{\lambda_i^2},$$

and

$$c_2 := \left\|\mathbf{P}^{\perp}_{\mathbf{X}^H}\mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}}\right\|_F^2,$$

the expression in 3.15 is proved.

We further have that

$$\begin{aligned}
c_2 = \left\|\mathbf{P}^{\perp}_{\mathbf{X}^H}\mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}}\right\|_F^2 = 0 \iff & \mathbf{P}^{\perp}_{\mathbf{X}^H}\mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}} = \mathbf{0} \\
\iff & (\mathbf{I} - \mathbf{P}_{\mathbf{X}^H})\mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}} = \mathbf{0} \\
\iff & \mathbf{P}_{\mathbf{X}^H}\mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}} = \mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}} \\
\iff & \text{span}\{\mathbf{X}^H\} \supseteq \text{span}\{\mathbf{V}\boldsymbol{\Sigma}\},
\end{aligned} \tag{3.19}$$

which proves the second part of the proposition. □

Proposition 3.6. *Given the model (3.1), a channel with spatial covariance matrix* $\mathbf{R}$*, and the eigen-decomposition* $\mathbf{R} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^H$*,* $\boldsymbol{\Sigma} = \text{diag}\{\sigma_1^2, \ldots, \sigma_N^2\}$*,* $\sigma_1^2 \geq \ldots \geq \sigma_N^2$*, the mean square error* $\text{MSE} := \mathbb{E}[\|\mathbf{h} - \hat{\mathbf{h}}\|^2]$ *of a LMMSE channel estimate* $\hat{\mathbf{h}}$ *defined in Section 3.1.1 is given by*

$$\text{MSE} = \sum_{c_i \neq 0} \frac{f_i}{1 + \dfrac{c_i}{\text{SNR}}} + \sum_{c_i = 0} f_i. \tag{3.20}$$

[1] We recall that every projection matrix $\mathbf{P}$ is idempotent, i.e. it satisfies $\mathbf{P} = \mathbf{P}\mathbf{P}$

where $\text{SNR} := P_h P_x/\sigma_w^2$ *as defined in (3.5), and where the constants* c_i *and* f_i *are defined as follows:*

$$c_i := \frac{l_i^2}{P_h P_x} \geq 0,$$

where $l_1^2 \geq \ldots \geq l_N^2$ *are the eigenvalues of* $\boldsymbol{\Xi} := \left(\mathbf{X}\mathbf{R}^{\frac{1}{2}}\right)^H \mathbf{X}\mathbf{R}^{\frac{1}{2}}$, $\mathbf{R}^{\frac{1}{2}} := \mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{V}^H$, *and*

$$f_i := f(\tilde{\mathbf{u}}_i) \geq 0,$$

where $f(\mathbf{x}) = \mathbf{x}^H\mathbf{R}\mathbf{x}$ *is the quadratic form associated to* $\mathbf{R}$, *and* $\tilde{\mathbf{u}}_i$ *denotes the eigenvector of* $\boldsymbol{\Xi}$ *associated to* l_i^2. *Furthermore, we have that*

$$\sum_{c_i=0} f_i = 0 \iff \text{span}\{\mathbf{X}^H\} \supseteq \text{span}\{\mathbf{V}\boldsymbol{\Sigma}\},$$

thus a necessary condition for having $\sum_{c_i=0} f_i = 0$ *is* $\text{rank}\{\mathbf{X}^H\} \geq \text{rank}\{\mathbf{R}\}$, *which implies* $T \geq \text{rank}\{\mathbf{R}\}$.

Proof. Let us consider the MSE expressed as $\mathbb{E}[\|\mathbf{h}-\hat{\mathbf{h}}\|^2] = \text{tr}\{\mathbf{R}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}\}$, where $\mathbf{R}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}$ denotes the covariance matrix of the estimation error, given by

$$\begin{aligned}\mathbf{R}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}} :&= \mathbb{E}[(\mathbf{h}-\hat{\mathbf{h}})(\mathbf{h}-\hat{\mathbf{h}})^H] \\ &= \mathbf{R} - \mathbf{R}_{\mathbf{hy}}{\mathbf{R}_{\mathbf{yy}}}^{-1}\mathbf{R}_{\mathbf{yh}} \\ &= \mathbf{R} - \mathbf{R}\mathbf{X}^H\left(\mathbf{X}\mathbf{R}\mathbf{X}^H + \sigma_w^2\mathbf{I}\right)^{-1}\mathbf{X}\mathbf{R}.\end{aligned}$$

By defining the square root matrix decomposition $\mathbf{R} = \mathbf{R}^{\frac{1}{2}}\mathbf{R}^{\frac{H}{2}}$,[2] with $\mathbf{R}^{\frac{1}{2}} := \mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{V}^H$, we can express the MSE as follows:

$$\begin{aligned}\text{MSE} &= \text{tr}\{\mathbf{R}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}\} \\ &= \text{tr}\left\{\mathbf{R}^{\frac{1}{2}}\left[\mathbf{I} - \mathbf{R}^{\frac{H}{2}}\mathbf{X}^H\left(\mathbf{X}\mathbf{R}\mathbf{X}^H + \sigma_w^2\mathbf{I}\right)^{-1}\mathbf{X}\mathbf{R}^{\frac{1}{2}}\right]\mathbf{R}^{\frac{H}{2}}\right\} \\ &= \text{tr}\left\{\mathbf{R}^{\frac{H}{2}}\mathbf{R}^{\frac{1}{2}}\left[\mathbf{I} - \mathbf{R}^{\frac{H}{2}}\mathbf{X}^H\left(\mathbf{X}\mathbf{R}^{\frac{1}{2}}\mathbf{R}^{\frac{H}{2}}\mathbf{X}^H + \sigma_w^2\mathbf{I}\right)^{-1}\mathbf{X}\mathbf{R}^{\frac{1}{2}}\right]\right\} \\ &= \text{tr}\left\{\mathbf{R}\left(\mathbf{I} + \sigma_w^{-2}\mathbf{R}^{\frac{H}{2}}\mathbf{X}^H\mathbf{X}\mathbf{R}^{\frac{1}{2}}\right)^{-1}\right\},\end{aligned}$$

where the last equality comes from the matrix inversion lemma. Let us now consider the eigenvalues $l_1^2 \geq \ldots \geq l_N^2$ of the matrix $\boldsymbol{\Xi} := \mathbf{R}^{\frac{H}{2}}\mathbf{X}^H\mathbf{X}\mathbf{R}^{\frac{1}{2}}$. Since $\text{Ker}\{\boldsymbol{\Xi}\} \supseteq \text{Ker}\{\mathbf{R}^{\frac{1}{2}}\} = \text{Ker}\{\mathbf{R}\}$, we have that $l_i^2 = 0$, $\forall i > p := \text{rank}\{\mathbf{R}\}$. By denoting the eigenvalues decomposition of $\boldsymbol{\Xi}$ with $\boldsymbol{\Xi} = \tilde{\mathbf{U}}\boldsymbol{\Gamma}\tilde{\mathbf{U}}^H$, $\boldsymbol{\Gamma} := \text{diag}(l_1^2, \ldots, l_N^2)$, $\tilde{\mathbf{U}} := \begin{bmatrix}\tilde{\mathbf{u}}_1 & \ldots & \tilde{\mathbf{u}}_N\end{bmatrix}$, we can

[2] $\mathbf{R}^{\frac{H}{2}}$ is a shorthand for $\left(\mathbf{R}^{\frac{1}{2}}\right)^H$

further express the MSE as

$$\begin{aligned}
\text{MSE} &= \text{tr}\left\{\mathbf{R}\left(\mathbf{I}+\sigma_w^{-2}\mathbf{R}^{\frac{H}{2}}\mathbf{X}^H\mathbf{X}\mathbf{R}^{\frac{1}{2}}\right)^{-1}\right\} \\
&= \text{tr}\left\{\mathbf{R}\left(\mathbf{I}+\sigma_w^{-2}\tilde{\mathbf{U}}\boldsymbol{\Gamma}\tilde{\mathbf{U}}^H\right)^{-1}\right\} \\
&= \text{tr}\left\{\mathbf{R}\tilde{\mathbf{U}}\left(\mathbf{I}+\sigma_w^{-2}\boldsymbol{\Gamma}\right)^{-1}\tilde{\mathbf{U}}^H\right\} \\
&= \text{tr}\left\{\tilde{\mathbf{U}}^H\mathbf{R}\tilde{\mathbf{U}}\left(\mathbf{I}+\sigma_w^{-2}\boldsymbol{\Gamma}\right)^{-1}\right\} \\
&= \sum_{i=1}^{p}\frac{1}{1+\dfrac{l_i^2}{\sigma_w^2}}\tilde{\mathbf{u}}_i^H\mathbf{R}\tilde{\mathbf{u}}_i \\
&= \sum_{l_i\neq 0}\frac{1}{1+\dfrac{l_i^2}{\sigma_w^2}}\tilde{\mathbf{u}}_i^H\mathbf{R}\tilde{\mathbf{u}}_i + \sum_{l_i=0}\tilde{\mathbf{u}}_i^H\mathbf{R}\tilde{\mathbf{u}}_i.
\end{aligned}$$

Let us define $f_i := \tilde{\mathbf{u}}_i^H\mathbf{R}\tilde{\mathbf{u}}_i$. By recalling that the positive semi-definite quadratic form $f(\mathbf{x}) = \mathbf{x}^H\mathbf{R}\mathbf{x}$ satisfies $f(\mathbf{x}) = 0 \iff \mathbf{x} \in \text{Ker}\{\mathbf{R}\}$, we can conclude that $c_i > 0\ \forall i \leq p$, because by construction we have that $\tilde{\mathbf{u}}_i \notin \text{Ker}\{\mathbf{R}\}\ \forall i \leq p$. Hence, we have that

$$\begin{aligned}
&\sum_{l_i=0} f_i = 0 \\
\iff & l_i^2 \neq 0 \quad \forall i \leq p \\
\overset{(a)}{\iff} & \text{Ker}\{\boldsymbol{\Xi}\} = \text{Ker}\{\mathbf{R}\} \\
\iff & \text{span}\{\boldsymbol{\Xi}^H\} = \text{span}\{\mathbf{R}^H\} \\
\iff & \text{span}\{\mathbf{R}^{\frac{H}{2}}\mathbf{X}^H\} = \text{span}\{\mathbf{R}^H\} \\
\iff & \text{span}\{\mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{V}^H\mathbf{X}^H\} = \text{span}\{\mathbf{V}\boldsymbol{\Sigma}\} \\
\iff & \text{span}\{\mathbf{X}^H\} \supseteq \text{span}\{\mathbf{V}\boldsymbol{\Sigma}\},
\end{aligned}$$

where for condition (a) we recall that $\text{Ker}\{\boldsymbol{\Xi}\} \supseteq \text{Ker}\{\mathbf{R}\}$. □

Corollary 3.2. *A corollary of proposition 3.6 is that, when the condition* $\text{span}\{\mathbf{X}^H\} \supseteq \text{span}\{\mathbf{V}\boldsymbol{\Sigma}\}$ *is satisfied, the expression for the MSE boils down to*

$$\text{MSE} = \sum_{i=1}^{p}\frac{1}{\dfrac{1}{\sigma_i^2}+\dfrac{\lambda_i^2}{\text{SNR}}},$$

where $\lambda_1^2 \geq \ldots \geq \lambda_N^2$ *are the eigenvalues of* $\mathbf{Q} := \mathbf{X}^H\mathbf{X}$.

Proof. By assuming $\text{span}\{\mathbf{X}^H\} \supseteq \text{span}\{\mathbf{V}\boldsymbol{\Sigma}\}$, we can write the eigenvalue decomposition $\mathbf{Q} = \mathbf{V}\text{diag}(\lambda_1, \ldots, \lambda_N)\mathbf{V}^H$, with $\lambda_1 \geq \ldots \geq \lambda_N$, $\lambda_i \neq 0\ \forall i \leq \text{rank}(\mathbf{R})$, and $\mathbf{V} =$

$\begin{bmatrix}\mathbf{v}_1 & \dots & \mathbf{v}_N\end{bmatrix}$. Thus, we have that

$$\begin{aligned}\boldsymbol{\Xi} &= \mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{V}^H\mathbf{Q}\mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{V}^H \\ &= \mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{V}^H\mathbf{V}\mathrm{diag}(\lambda_1^2,\dots,\lambda_N^2)\mathbf{V}^H\mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{V}^H \\ &= \mathbf{V}\boldsymbol{\Sigma}^{\frac{1}{2}}\mathrm{diag}(\lambda_1^2,\dots,\lambda_N^2)\boldsymbol{\Sigma}^{\frac{1}{2}}\mathbf{V}^H,\end{aligned}$$

which means that $l_i^2 = \sigma_i\lambda_i^2\sigma_i = \sigma_i^2\lambda_i^2$. Furthermore, we have that $\tilde{\mathbf{u}}_i = \mathbf{v}_i$, which implies $f_i = \mathbf{v}_i^H\mathbf{R}\mathbf{v}_i = \mathbf{v}_i^H\mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^H\mathbf{v}_i = \sigma_i^2$. □

Remark 3.3. *The analysis in proposition 3.6 is similar to the analysis of [19, Theorem 1], where the MSE given by a random pilot sequence is taken into account. However, here the focus is on the MSE given by a deterministic pilot sequence (or, equivalently, the conditional MSE given by a random pilot sequence, where the conditioning is on an arbitrary pilot realization).*

As an immediate consequence of 3.5 and 3.6, we have the following result:
Corollary 3.3. *By considering the the MSE of both the LS and LMMSE estimator, the following property holds:*

$$\lim_{\sigma_w^2\to 0}\mathrm{MSE} = 0 \iff \mathrm{span}\{\mathbf{X}^H\} \supseteq \mathrm{span}\{\mathbf{V}\boldsymbol{\Sigma}\}.$$

Remark 3.4. *The meaning of the minimum pilot sequence length given by the propositions in this section must be intended only as a necessary condition for the MSE to vanish as the SNR increase.*

3.3.2 Pilot Sequence Design for Correlated Channels (SU-MIMO)

In section 3.2.2 we have seen as the optimal pilot sequence satisfies the minimum pilot length criteria described in section 3.2.1 (given an additional mild condition on the SNR level for the LS estimator). In this section we repeat the analysis for generic correlated channels, highlighting the differences from the i.i.d. case. In particular, in contrast with the i.i.d. fading case, we show that the optimal pilot sequence design does not necessarily satisfies the minimum pilot sequence length criteria described in section 3.3.1. In fact, the optimal design is shown to be related to the joint effect of the eigenvalue distribution of the channel covariance matrix, and of the SNR level. The intuition is that, for sufficiently small SNR, the signal in the subspace corresponding to the smallest eigenvalues of $\mathbf{R}$ contains mostly noise, so signals lying in this subspace should be filtered out.

Similarly to section 3.2.2, we constrain the overall energy of the training sequence with (3.14), so that the average power per antenna and per time slot is at most unitary. We adopt the definition of SNR (3.5) by setting $P_x = 1$, and we define the the matrix $\mathbf{Q} := \mathbf{X}^H\mathbf{X}$ as well as its eigenvalues decomposition $\mathbf{Q} = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H$, $\boldsymbol{\Lambda} = \mathrm{diag}\{\lambda_1^2,\dots,\lambda_N^2\}$.

In the following, we focus on the LS estimator only. A similar analysis for the LMMSE estimator is left as future work. We approach the design of the optimal (in terms of

MSE) pilot sequence by firstly analyzing in a disjoint way the two terms c_1 and c_2 defined in lemma 3.5 and in its proof. Then, we discuss the joint optimization of these two parameters.

Proposition 3.7. *Among all the pilot sequences* $\mathbf{X}$ *of length* T *and rank* r*, the term* c_1 *defined in proposition 3.5 is minimized by letting*

$$\lambda_i^2 = \begin{cases} \dfrac{NT}{r} & \text{if} i \in \mathcal{I}_r \\ 0 & \text{otherwise} \end{cases},$$

where $\mathcal{I}_r$ *is a generic* r*-combination of the set of indexes* $\mathcal{I} := 1, \ldots, N$*, resulting in*

$$c_1 = \frac{\text{tr}\{\mathbf{R}\}}{N^2 T} r^2.$$

Proof. Let us consider the expression for c_1 given by 3.16. The proof is identical to the first part of the proof of proposition 3.4, and it is here omitted. □

Proposition 3.8. *Among all the pilot sequences* $\mathbf{X}$ *of length* T *and rank* r*, the term* c_2 *defined in proposition 3.5 is minimized by letting*

$$\text{span}\{\mathbf{X}^H\} = \text{span}\{\mathbf{V}_r\}, \tag{3.21}$$

where $\mathbf{V}_r$ *is the matrix which columns are the eigenvectors of* $\mathbf{R}$ *corresponding to the* r *strongest eigenvalues.*

Proof. From equations 3.18 and 3.17, let us rewrite c_2 as

$$c_2 = \text{tr}\left\{\mathbf{P}^{\perp}_{\mathbf{X}^H}\mathbf{R}\right\} = \text{tr}\left\{\mathbf{R}\right\} - \text{tr}\left\{\mathbf{P}_{\mathbf{X}^H}\mathbf{R}\right\}.$$

Let $\sigma_1^2 \geq \ldots \geq \sigma_N^2$ be the eigenvalues of the covariance matrix $\mathbf{R}$, $\mathbf{V}\mathbf{\Sigma}\mathbf{V}^H = \mathbf{R}$, $\mathbf{\Sigma} = \text{diag}(\sigma_1^2, \ldots, \sigma_N^2)$, be its eigenvalue decomposition, and $\mathbf{V}_n$ be the tall matrix constructed with the first $n \leq N$ columns of $\mathbf{V}$. By focusing on the term $\text{tr}\left\{\mathbf{P}_{\mathbf{X}^H}\mathbf{R}\right\}$ we further obtain:

$$\begin{aligned}
\text{tr}\left\{\mathbf{P}_{\mathbf{X}^H}\mathbf{R}\right\} &= \text{tr}\left\{\mathbf{X}^{\dagger}\mathbf{X}\mathbf{V}\mathbf{\Sigma}\mathbf{V}^H\right\} \\
&= \text{tr}\left\{\mathbf{V}^H\mathbf{X}^{\dagger}\mathbf{X}\mathbf{V}\mathbf{\Sigma}\right\} \\
&= \text{tr}\left\{\mathbf{V}^{\dagger}\mathbf{X}^{\dagger}\mathbf{X}\mathbf{V}\mathbf{\Sigma}\right\} \\
&\overset{(a)}{=} \text{tr}\left\{(\mathbf{X}\mathbf{V})^{\dagger}\mathbf{X}\mathbf{V}\mathbf{\Sigma}\right\} \\
&= \text{tr}\left\{\mathbf{P}_{(\mathbf{X}\mathbf{V})^H}\mathbf{\Sigma}\right\} \\
&= \sum_{i=1}^{N} \sigma_i^2 p_i,
\end{aligned}$$

where equality (a) holds because $\mathbf{V}$ has orthogonal columns, and p_i denotes the ith diagonal element of the projection matrix $\mathbf{P}_{(\mathbf{XV})^H}$. By recalling that the diagonal elements of a generic projection matrix are bounded in the compact interval $[0, 1]$, and by noticing that

$$\sum_{i=1}^{N} p_i = \mathrm{tr}\{\mathbf{P}_{(\mathbf{XV})^H}\} = \mathrm{tr}\{\mathbf{V}^H\mathbf{X}^\dagger\mathbf{XV}^H\} = \mathrm{tr}\{\mathbf{X}^\dagger\mathbf{XVV}^H\} = \mathrm{tr}\{\mathbf{P}_{\mathbf{X}^H}\} = r,$$

we can minimize c_2 by solving the following convex (LP) optimization problem:

$$\begin{aligned} \underset{\mathbf{x}}{\text{minimize}} \quad & -\mathbf{b}^T\mathbf{x} \\ \text{subject to} \quad & \mathbf{0} \preceq \mathbf{x} \preceq 1, \\ & \mathbf{1}^T\mathbf{x} = r \end{aligned}$$

where $\mathbf{b}^T := [\sigma_1^2, \ldots, \sigma_N^2]$, $\mathbf{x} := [p_1, \ldots, p_N]^T$, and $\mathbf{1}$ denotes the all-ones column vector. The optimal solution, obtained by solving the KKT conditions[3], is given by

$$p_i = \begin{cases} 1 & \text{if } i = 1, \ldots, r \\ 0 & \text{otherwise} \end{cases}.$$

The interpretation of this solution is immediate by looking at the optimal c_2:

$$c_2 = \mathrm{tr}\{\mathbf{R}\} - \sum_{i=1}^{r} \sigma_i^2 = \sum_{i=r+1}^{N} \sigma_i^2, \tag{3.22}$$

which corresponds to a residual channel estimation error that includes just the $N - r$ weaker dimensions of the subspace on which $\mathbf{h}$ lies.

Let us now analyze the impact of this optimality condition on the structure of the pilot sequence $\mathbf{X}$. Firstly we notice that, under the given optimal diagonal elements p_i of $\mathbf{P}_{(\mathbf{XV})^H}$, and by recalling that the eigenvalues of a generic projection matrix assume values only in $\{0, 1\}$, we have that $\mathbf{P}_{(\mathbf{XV})^H}$ must be already in diagonal form (it is diagonalized by the standard basis). Thus, we can write

$$\mathbf{P}_{(\mathbf{XV})^H} = (\mathbf{XV})^\dagger(\mathbf{XV}) = \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix}.$$

By pre-multiplying all members by $(\mathbf{XV})$, and by applying the definition of the pseudo-inverse, we obtain

$$\begin{aligned} (\mathbf{XV})(\mathbf{XV})^\dagger(\mathbf{XV}) &= (\mathbf{XV}) \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \\ (\mathbf{XV}) &= (\mathbf{XV}) \begin{bmatrix} \mathbf{I}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{bmatrix} \\ \begin{bmatrix} \mathbf{XV}_r & \mathbf{X}\bar{\mathbf{V}}_r \end{bmatrix} &= \begin{bmatrix} \mathbf{XV}_r & \mathbf{0} \end{bmatrix}, \end{aligned}$$

[3]One can check that this solution satisfies the KKT conditions given that $\sigma_i^2 \geq \sigma_j^2$ for $i < j$, which is always true by construction.

where $\bar{\mathbf{V}}_r$ denote the matrix which columns are the $N-r$ eigenvectors not included in $\mathbf{V}_r$. From the last equality we observe that

$$\mathbf{X}\bar{\mathbf{V}}_r = \mathbf{0} \iff \text{span}\{\mathbf{X}^H\} \perp \text{span}\{\bar{\mathbf{V}}_r\},$$

which by construction[4] of $\mathbf{X}^H$ and $\bar{\mathbf{V}}_r$ implies

$$\mathbf{X}\bar{\mathbf{V}}_r = \mathbf{0} \iff \text{span}\{\mathbf{X}^H\} = \text{span}\{\mathbf{V}_r\}.$$

□

We can now state the following corollary:

Corollary 3.4. *Among all the pilot sequences* $\mathbf{X}$ *of length* T *and rank* r*, the mean square error* $\text{MSE} := \mathbb{E}[\|\mathbf{h}-\hat{\mathbf{h}}\|^2]$ *of a LS channel estimate* $\hat{\mathbf{h}}$ *defined in Section 3.1.1 is minimized by letting*

$$\text{span}\{\mathbf{X}^H\} = \text{span}\{\mathbf{V}_r\},$$

where $\mathbf{V}_r$ *is the matrix which columns are the eigenvectors of* $\mathbf{R}$ *corresponding to the* r *strongest eigenvalues, and*

$$\lambda_i^2 = \begin{cases} \dfrac{NT}{r} & \text{if} i \in \mathcal{I}_r^* \\ 0 & \text{otherwise} \end{cases},$$

where $\mathcal{I}_r^*$ *is the set of indexes corresponding to the* r *eigenvectors* $\mathbf{U}_r$ *such that* $\text{span}\{\mathbf{U}_r\} = \text{span}\{\mathbf{V}_r\}$*. The resulting MSE is given by*

$$\text{MSE} = \sum_{i=r+1}^{N} \sigma_i^2 + \frac{\text{tr}\{\mathbf{R}\}r^2}{N^2T}\frac{1}{\text{SNR}},$$

where $\sigma_1^2 \geq \ldots \geq \sigma_N^2$ *are the eigenvalues of* $\mathbf{R}$*.*

Proof. The proof follows directly by jointly imposing the optimality conditions on the terms c_1 and c_2 as described in proposition 3.7 and proposition 3.8. □

An example of a training sequence satisfying the optimal pilot design given by proposition 3.4 is given in section 3.4.2.

Remark 3.5. *Corollary 3.4 gives the optimal pilot design, in terms of MSE, by fixing the pilot length* T *and its rank* r*. An important open question is how to design* T *and* r*. In general, we should look at a more complicated optimization problem, for example in terms of achievable rate, over the tuple* (r,T)*, which takes into account the joint effect of the channel estimation error and the penalty due to the resources allocated for pilot transmission (*T *slots), similarly to the analysis in section 4.8.3. However, for simplicity, in the following we focus on the optimization of the MSE by assuming a*

[4] $\mathbf{X}^H$ has rank r, and $\bar{\mathbf{V}}_r$ spans a subspace of dimension $N-r$

system that allows a maximum pilot length of $T = T_{\max} \ll M_c$*, where we recall that* M_c *is the coherence block length. We introduce the following notation*

$$\eta(r) := \frac{\sum_{i=1}^{r} \sigma_i^2}{\sum_{i=1}^{N} \sigma_i^2}, \quad \bar{\eta}(r) := 1 - \eta(r), \tag{3.23}$$

where $\sigma_1^2 \geq \ldots \geq \sigma_N^2$ *are the eigenvalues of* $\mathbf{R}$*. We can then optimize the MSE, expressed as*

$$\text{MSE} = \text{tr}\{\mathbf{R}\} \left[\bar{\eta}(r) + \frac{r^2}{N^2 T} \frac{1}{\text{SNR}} \right],$$

by letting $T = T_{\max}$ *and* $r = r^*$*, where* r^* *is the solution of*

$$\begin{aligned} \underset{r}{\text{minimize}} \quad & \bar{\eta}(r) + \frac{r^2}{N^2 T_{\max} \text{SNR}}, \\ \textit{subject to} \quad & r \leq \min(N, T_{\max}) \end{aligned}$$

which shows a clear dependence of the optimal value on the SNR level and on the distribution $\eta(r)$ of the eigenvalues of $\mathbf{R}$.

3.4 Training Overhead Reduction in Correlated Channels

3.4.1 Gains of Correlation-aware LS Channel Estimation (SU-MIMO)

In this section we compare the performance of the correlation-aware LS pilot design with the performance of the optimal LS pilot design for i.i.d. fading channels. We show that, if the channel covariance matrix is low rank, the training overhead can be reduced significantly. More precisely, we have the following proposition:

Proposition 3.9. *Let us consider a channel* $\mathbf{h}_{\text{corr}}$ *with covariance matrix* $\mathbf{R}$*, with* $p := \text{rank}\{\mathbf{R}\}$*, and a LS channel estimate* $\hat{\mathbf{h}}_{\text{corr}}$ *given by 3.1.1 and obtained with an optimal pilot sequence of length* $T_{\text{corr}} \geq p$ *satisfying proposition 3.5 and corollary 3.4. Let us further consider another channel* $\mathbf{h}_{\text{iid}}$ *with covariance matrix* $\sigma_h^2 \mathbf{I}$*, and a LS estimate* $\hat{\mathbf{h}}_{\text{iid}}$ *obtained with an optimal pilot sequence of length* $T_{\text{iid}} \geq N$ *satisfying proposition 3.4. By defining* $\text{MSE}_{\text{corr}} := \mathbb{E}[\|\mathbf{h}_{\text{corr}} - \hat{\mathbf{h}}_{\text{corr}}\|^2]$ *and* $\text{MSE}_{\text{iid}} := \mathbb{E}[\|\mathbf{h}_{\text{iid}} - \hat{\mathbf{h}}_{\text{iid}}\|^2]$*, and by using for fairness the same SNR and channel power* $N\sigma_h^2 = \text{tr}\{\mathbf{R}\}$*, we have that*

$$\text{MSE}_{\text{corr}} \leq \text{MSE}_{\text{iid}} \iff T_{\text{corr}} \geq \max\left[p, \left(\frac{p}{N}\right)^2 T_{\text{iid}} \right].$$

Proof. By recalling the expressions for the MSEs given by corollary 3.4 and proposition

3.4, and that proposition 3.5 is satisfied iff $r \geq p$, we have

$$\begin{aligned} &\mathrm{MSE}_{\mathrm{corr}} \leq \mathrm{MSE}_{\mathrm{iid}} \\ \iff &\frac{\mathrm{tr}\{\mathbf{R}\}r^2}{N^2 T_{\mathrm{corr}}\mathrm{SNR}} \leq \frac{N\sigma_g^2}{T_{\mathrm{iid}}\mathrm{SNR}} \\ \iff &T_{\mathrm{corr}} \geq \frac{r^2}{N^2} T_{\mathrm{iid}} \\ \iff &T_{\mathrm{corr}} \geq \frac{p^2}{N^2} T_{\mathrm{iid}} \end{aligned}$$

The proof is completed by recalling that $T_{\mathrm{corr}} \geq p$. □

Remark 3.6. *In Massive MIMO systems, because of the high impact of the training overhead, we typically have $T_{\mathrm{iid}} = N$. Thus, proposition 3.9 reduces to*

$$\mathrm{MSE}_{\mathrm{corr}} \leq \mathrm{MSE}_{\mathrm{iid}} \iff T_{\mathrm{corr}} \geq \max\left[p, p\left(\frac{p}{N}\right)\right] = p.$$

3.4.2 Example: Subspace-based SU-MIMO DL Channel Estimation

Let us focus on the SU-MIMO channel estimation problem. A popular approach for this channel estimation task in the case of correlated fading is based on the following Karhuen-Loeve transform:

$$\mathbf{h} = \mathbf{V}_r \boldsymbol{\Sigma}_r^{\frac{1}{2}} \boldsymbol{\beta}, \tag{3.24}$$

where $\boldsymbol{\beta} \in \mathbb{C}^{r\times 1}$ is an equivalent channel vector with covariance matrix $\mathbb{E}[\boldsymbol{\beta}\boldsymbol{\beta}^H] = \mathbf{I}$, and where $\mathbf{V}_r \boldsymbol{\Sigma}_r^{\frac{1}{2}} \in \mathbb{C}^{N\times r}$ corresponds to the $r = \mathrm{rank}\{\mathbf{R}\}$ non-zero eigenvalues and associated eigenvectors of the eigen-decomposition $\mathbf{R} = \mathbf{V}\boldsymbol{\Sigma}\mathbf{V}^H$, $\boldsymbol{\Sigma} = \mathrm{diag}\{\sigma_1^2, \ldots, \sigma_N^2\}$. This representation shows how the knowledge of $\mathbf{R}$ can be exploited to improve channel estimation, as it gives important information about the subspace where most of the energy of $\mathbf{h}$ lies. In particular, if the covariance matrix has low rank, and if it is known at the BTS, this representation can be effectively exploited to reduce the training overhead. To make this statement concrete, let us rewrite 3.1 as

$$\mathbf{y} = \mathbf{X}\mathbf{h} + \mathbf{w} = \mathbf{X}\mathbf{V}_r \boldsymbol{\Sigma}_r^{\frac{1}{2}} \boldsymbol{\beta} + \mathbf{w}, \tag{3.25}$$

By setting for example a training sequence $\mathbf{X} = \sqrt{P}\mathbf{V}_r^H$ of length $T = r \leq N$ and with a per-slot power constraint P, we obtain

$$\mathbf{y} = \sqrt{P}\mathbf{V}_r^H \mathbf{V}_r \boldsymbol{\Sigma}_r^{\frac{1}{2}} \boldsymbol{\beta} + \mathbf{w} = \sqrt{P}\boldsymbol{\Sigma}_r^{\frac{1}{2}} \boldsymbol{\beta} + \mathbf{w}.$$

An estimate of $\mathbf{h}$ can be obtained by computing

$$\hat{\mathbf{h}} = \frac{1}{\sqrt{P}}\mathbf{V}_r \mathbf{y} = \mathbf{h} + \frac{1}{\sqrt{P}}\mathbf{V}_r \mathbf{w}. \tag{3.26}$$

We point out that this estimation scheme corresponds to an LS channel estimation scheme as described in section 3.1.1, with pilot sequence satisfying the conditions of proposition 3.5 and corollary 3.4, and by setting $P = N$.

Channel Estimation based on Principal Subspaces

In many applications, the covariance matrix $\mathbf{R}$ is not necessarily low-rank, but it experiences a low-rank behavior, in the sense that only few eigen-modes carry most of the signal energy. In this case, a significant reduction in training overhead can be obtained with a variant of the subspace estimation scheme analyzed previously.

In more detail, let $\sigma_1^2 \geq \ldots \geq \sigma_N^2$ be the eigenvalues of the covariance matrix $\mathbf{R}$. Denote by $\mathbf{V}\text{diag}(\sigma_1^2, \ldots, \sigma_N^2)\mathbf{V}^H = \mathbf{R}$ the eigenvalue decomposition of the covariance matrix $\mathbf{R}$, and by $\mathbf{V}_n$ the tall matrix constructed with the first $n \leq N$ columns of $\mathbf{V}$. For a given design parameter $\eta \in [0, 1]$, let p be the smallest integer satisfying

$$\frac{\sum_{i=1}^{p} \sigma_i^2}{\sum_{i=1}^{N} \sigma_i^2} \geq \eta.$$

The principal subspace scheme for channel estimation is obtained by using the pilot sequence $\mathbf{X} = \sqrt{P}\mathbf{V}_p^H$ and the LS estimator defined in Section 3.1.1; i.e., $\mathbf{f}_{\text{LS}} = \mathbf{X}^\dagger = \frac{1}{\sqrt{P}}\mathbf{V}_p$.

This scheme satisfies the optimal LS pilot sequence design described in section 3.3.2, by letting $r = T \leq N$. According to the analysis in section 3.3.2, the signal in the subspace corresponding to the smallest eigenvalues of the covariance matrix contains mostly noise, so we should filter out signals in this subspace, especially when the SNR is the main bottleneck. As outlined in remark 3.5, a proper performance evaluation of this channel estimation scheme should include the the penalty due to the training overhead T. In Massive MIMO systems, and when T is high, the training overhead penalty may weight more than the penalty due to the estimation error. Hence, intuitively, in the large-scale array regime and for relatively low SNR, this estimation scheme may greatly improve the system performance. This intuition is experimentally confirmed by the simulation in section 4.8.3, which uses the realistic channel model in section 2.4, the proposed estimation scheme for different choices of η, and a transmission scheme based on the conventional MRC precoder.

3.4.3 A Link with Compressed Sensing Based Channel Estimation

In the traditional compressed sensing framework a high N-dimensional signal, e.g $\mathbf{h}\mathbb{C}^N$, is said to be *sparsely representable* if there exist an orthonormal basis $\mathbf{B} \in \mathbb{C}^{N \times N}$ such that

$$\mathbf{h} = \mathbf{B}\boldsymbol{\beta}, \qquad \|\boldsymbol{\beta}\|_0 \ll N$$

where $\|\boldsymbol{\beta}\|_0$ indicates the number of non-zero entries of $\boldsymbol{\beta}$.

In recent years lots of attention has been given also to an extension of the traditional compressed sensing framework: it has been shown that instead of an orthonormal basis

it is possible to use as a sparsifying matrix also an *overcomplete dictionary* $\mathbf{D} \in \mathbb{C}^{N\times K}$ (with $K > N$) that allows for even sparser representations [20]. However, even though this will limit the validity of our analysis, in the following we focus for simplicity on the traditional framework.

Compressed Sensing (CS) theory suggests that it is possible to measure a signal that is *sparse* in some basis just focusing on the few non-zero elements, saving a lot of estimation effort, by solving the *sparse signal recovery problem* [20]. This approach has been successfully applied to reduce training overhead in massive MIMO channel estimation, for example in [5].

In the following we show that, in case of i.i.d. fading, it is not possible to reduce the channel estimation effort by means of CS based algorithms, at least for the traditional framework. It turns out that a necessary condition for a channel vector to be sparsely representable is that it is correlated. This results strengthen the analysis given by this chapter, which shows that channel correlation is a fundamental requirement for reducing the channel estimation overhead.

More precisely, we have the following proposition:
Proposition 3.10. *Let us consider a random vector* $\mathbf{h} \in \mathbb{C}^N$ *and let us assume that it is sparsely representable as* $\mathbf{h} = \mathbf{B}\boldsymbol{\beta}$, $\|\boldsymbol{\beta}\|_0 \ll N$ *for a given matrix* $\mathbf{B} \in \mathbb{C}^{N\times N}$. *Then,*

$$\nexists k \in \mathbb{R} \quad \text{s.t.} \quad \mathbf{R} := \mathbb{E}[\mathbf{h}\mathbf{h}^H] = k\mathbf{I}.$$

Proof. By definition

$$\begin{aligned} \mathbf{R} = \mathbb{E}[\mathbf{h}\mathbf{h}^H] = \mathbf{B}\mathbb{E}[\boldsymbol{\beta}\boldsymbol{\beta}^H]\mathbf{B}^H = k\mathbf{I} \\ \iff \mathbb{E}[\boldsymbol{\beta}\boldsymbol{\beta}^H] = \mathbf{B}^H k\mathbf{B} = k\mathbf{I} \end{aligned}$$

which means that the vector $\boldsymbol{\beta}$ must be uncorrelated. However this is in not possible, because the assumption $L_0 := \|\boldsymbol{\beta}\|_0 < N$ implies that $\boldsymbol{\beta}$ is indeed correlated. In fact, by denoting by $\boldsymbol{\beta}_{L_0} \in \mathbb{C}^{L_0}$ an arbitrary subset of the elements of $\boldsymbol{\beta}$ of length L_0, the random vector $\bar{\boldsymbol{\beta}}_{L_0} \in \mathbb{C}^{N-L_0}$ collecting all the other elements of $\boldsymbol{\beta}$ [5] satisfies

$$\bar{\boldsymbol{\beta}}_{L_0} | (\boldsymbol{\beta}_{L_0} \neq \mathbf{0}) = \mathbf{0},$$

which clearly shows that the entries of $\boldsymbol{\beta}$ are statistically dependent. □

[5] By assumption $\bar{\boldsymbol{\beta}}_{L_0}$ has at least one element

3.5 Fundamental Limits of Massive MIMO: A DoF Perspective

3.5.1 A DoF Bound based on Non-coherent Capacity

Let us consider the DL MU-MIMO model (2.2), and let us further focus on a transmission block $m = 0, \ldots, M_c - 1$ such that all the UEs channel vectors can be assumed constant. The received signal to all UEs for the entire transmission block, denoted as $\mathbf{Y} \in \mathbb{C}^{K \times M_c}$, can be rewritten in matrix form as

$$\mathbf{Y} = \mathbf{H}\mathbf{X} + \mathbf{W}, \tag{3.27}$$

where $\mathbf{H} := \begin{bmatrix} \mathbf{h}_1^d & \ldots & \mathbf{h}_K^d \end{bmatrix}^T \in \mathbb{C}^{K \times N}$ is the equivalent MIMO channel obtained by grouping all the UEs channel vectors $\mathbf{h}_k^d$, $\mathbf{X} \in \mathbb{C}^{N \times M_c}$ is the transmitted sequence at the BTS, and $\mathbf{W} \in \mathbb{C}^{K \times M_c}$ is the noise. We further normalize $\mathbf{X}$ to that the average transmit power at each transmit antenna in one time slot is unitary, i.e. we impose the following power constraint:

$$\mathbb{E}\left[\|\mathbf{X}\|_F^2\right] = NM_c. \tag{3.28}$$

Under this normalization, we define the average SNR per receive antenna as $\text{SNR} := N/\sigma_w^2$.

A popular performance metric for coherent block fading channels is the notion of ergodic capacity [1], defined in this case as

$$C := \mathbb{E}\left[\mathcal{I}(\mathbf{H})\right] \tag{3.29}$$

where $\mathcal{I}(\mathbf{H})$ is a random variable that denotes the maximum mutual information per time slot between the input sequence $\mathbf{X}$ and the output sequence $\mathbf{Y}$ conditioned on a given channel realization $\mathbf{H}_0$, i.e.

$$\mathcal{I}(\mathbf{H} = \mathbf{H}_0) := \frac{1}{M_c} \max_{f_X(\cdot)} I(\mathbf{X}; \mathbf{Y} | \mathbf{H} = \mathbf{H}_0),$$

where the optimization is carried over all input sequence distributions f_X satisfying the power constraint (3.28). Theoretically, the notion of ergodic capacity (from now on denoted simply as capacity) describes the maximum rate (in bits per channel use) that can be achieved via channel coding over infinite coherence blocks. This implies that its operational meaning is justified for applications that have delay constraints sufficiently loose to allow coding over multiple coherence intervals (fast-fading assumption).

By assuming perfect CSI at the transmitter and at the receivers, and with full cooperation among UEs, the ergodic capacity of the considered DL MU-MIMO system boils down to the well-known Gaussian parallel channels expression given by [1, Chapter 7]:

$$C_{\text{coh}} = \mathbb{E}\left[\sum_{i=1}^{\min(N,K)} \log_2\left(1 + \frac{P_i^* \lambda_i^2}{\sigma_w^2}\right)\right], \tag{3.30}$$

where $P_i^* \geq 0$, $\sum_{i=1}^{\min(N,K)} P_i^* = N$ are the power allocations obtained with the water-filling algorithm, and λ_i are the random eigenvalues of $\mathbf{H}$. In the high SNR regime, and assuming i.i.d. Rayleigh fading (i.e. elements of $\mathbf{H}$ are $\mathcal{CN}(0,1)$), the capacity is given by

$$C_{\text{coh}} = N^* \log_2(\text{SNR}) + o(1), \tag{3.31}$$

where $N^* = \min(N, K)$. The quantity N^* is often referred to as the multiplexing gain, or the number of degrees of freedom (DoF) of the channel, and it describes the maximum number of independent streams per channel use that can be multiplexed in a MIMO channel, compared to an AWGN channel with the same (high) quality. Formally, we define the DoF as

$$\text{DoF} := \lim_{\text{SNR}\to\infty} \frac{C(\text{SNR})}{\log_2(\text{SNR})}.$$

In the low SNR regime C is proportional to the capacity of the AWGN channel, and the main bottleneck is given by the power constraint rather than the spatial DoF offered by multiple antennas. Thus, in the following, the focus is on the high SNR regime only.

The main limitation of using (3.31) as a performance metric for MU-MIMO systems is that it does not takes into account the resource consumed by real systems to become coherent, i.e. to acquire the CSI. Indeed, real systems are not coherent, thus a more theoretically appropriate capacity expression is the one derived in [21] for non-coherent i.i.d. Rayleigh block fading point-to-point MIMO capacity:

$$C = N^* \left(1 - \frac{N^*}{M_c}\right) \log_2(\text{SNR}) + o(1), \quad \text{SNR} \to \infty, \tag{3.32}$$

where

$$N^* = \min\left(N, K, \left\lfloor \frac{M_c}{2} \right\rfloor\right)$$

describes the optimal number of transmit antennas to be used at the BTS in the high SNR regime, among the N available, which saturates at $\left\lfloor \frac{M_c}{2} \right\rfloor$. It is important to underline that this result represents a very optimistic performance bound, since it assume perfect cooperation among the UE. Nevertheless, even in this extremely ideal case, it is evident how for a massive MIMO system, i.e. when both N and K are large, in case of i.i.d. Rayleigh fading, the limited coherence block length represents a major performance bottleneck: the available (optimistic) number of DoF saturate to $\frac{M_c}{4}$.

Example: Consider again the example given in the introduction of this chapter, which is a typical 2 GHz LTE system with coherence bandwidth $B_c = 100$ KHz and coherence time $T_c = 1$ ms (UE speed of roughly 60 Km/h). The corresponding coherence block length is $M_c = B_c T_c = 100$ time slots. Thus, the maximum DoF is $M_c/4 = 25$, no matter how large the system is (i.e no matter how large N and K are).

An interesting insight that links the CSI acquisition overhead to the non-coherent capacity formula (3.32) is given in [21, Section V], where a scheme based on pilot based channel estimation is shown to achieve the envisioned DoF. We point out that the adopted DoF

achieving channel estimation scheme assumes a training sequence length at least equal to the number of transmit antennas. This important assumption gives an operational meaning to the choice of the optimal number of transmit antennas N^*: there is no point in using more transmit antennas than $\left\lfloor \frac{M_c}{2} \right\rfloor$, as the system will spend more time for channel estimation than for data transmission.

3.5.2 Impact of Training Overhead Reduction on the DoF Bound

Let us consider the block fading DL MU-MIMO model (3.27), with coherence block length M_c. As discussed in [21, Section V], by using pilot based channel estimation schemes that require a training sequence of length N, massive MIMO achievable DoF are upper bounded by $M_c/4$.

However, let us know assume that is is possible to reduce the training overhead by a factor $G := N/T$, where T is the training sequence length. Under this assumption, by following the same lines as in [21, Section V], the analysis in [22] proposes a new upper bound on the DoF, given by

$$\text{DoF} = N^* \left(1 - \frac{N^*}{GM_c}\right), \tag{3.33}$$

where

$$N^* = \min\left(N, K, \left\lfloor \frac{GM_c}{2} \right\rfloor\right).$$

As a consequence, because of the resources released by the reduction of the training overhead, Massive MIMO systems DoF are now upper bounded by $GM_c/4$, i.e. with a gain G with respect to the bound given by (3.32). The main challenge is to understand how and in which scenario it is possible to achieve this bound.

The main intuition provided by [22] is that, since the bound given by (3.32) is derived by assuming i.i.d. fading, it is possible to break this performance bottleneck, and possibly achieve the new bound (3.33), by considering instead correlated channels.

As we have seen in Section 3.4, for a SU-MIMO system, if the channel is correlated, it is possible to obtain the same CSI estimation performance of an equivalent channel estimation scheme for i.i.d. fading but with a training overhead gain of $G = N/p \geq 1$. A similar analysis for MU-MIMO system has not been given in this work. Nevertheless advanced MU-MIMO channel estimation techniques (e.g. [7]) are available in the literature, and they show that by exploiting correlation such a training overhead reduction is possible also for MU-MIMO. However, a rigorous analysis of the achievable DoF of correlated channels and its connection with channel estimation is still an ongoing line of research.

3.5.3 Reducing the Training in i.i.d. Rayleigh Channels

Clearly, for $G > 1$, the bound (3.33) cannot be achieved in the case of i.i.d. Rayleigh channels, as it violates the capacity formula (3.32). This directly implies that the reduction of the training overhead in i.i.d. Rayleigh channels does not give any improvement in terms of DoF upper bound. Of course this does not imply that a reduction in channel estimation overhead cannot help in how fast the system can reach the high SNR regime. However, at least for conventional channel estimation techniques, it is possible to show that reducing the training overhead is actually detrimental, in the high SNR regime.

Let us consider, for example, the following conventional pilot-based scheme:

- The coherence block M_c of model (3.27) is split into a training phase of length T and into a data transmission phase of length $M_c - T$.
- In the training phase, the received signal can be written as

  $$\mathbf{Y}_T = \mathbf{H}\mathbf{X}_T + \mathbf{W}_T,$$

  where $\mathbf{X}_T$ is a full rank pilot sequence such that $\|\mathbf{X}_T\|_F^2 = NT$. Every user estimates is channel vector using the LMMSE estimator described in section 3.1.1. We assume that that the entire system instantaneously acquire the complete channel estimate $\hat{\mathbf{H}}$. We denote further define $\tilde{\mathbf{H}} := \mathbf{H} - \hat{\mathbf{H}}$. Because of the properties of the LMMSE estimator, $\tilde{\mathbf{H}}$ and $\hat{\mathbf{H}}$ are uncorrelated. Moreover, since $\mathbf{H}$ is zero mean complex Gaussian distributed, also $\hat{\mathbf{H}}$ and $\tilde{\mathbf{H}}$ are zero mean Gaussian distributed.
- In the data transmission part, the transmitted signal in one given time slot can be written as

  $$\mathbf{y} = \mathbf{H}\mathbf{x} + \mathbf{w} = \hat{\mathbf{H}}\mathbf{x} + \tilde{\mathbf{H}}\mathbf{x} + \mathbf{w}.$$

  The signal $\mathbf{x}$ is assumed i.i.d. and satisfying the power constraint $\mathbb{E}[\|\mathbf{x}\|^2] = N$. One can check that the overall power of the scheme satisfies (3.28). We use the same definition SNR $:= N/\sigma_w^2$ as in section 3.5. For reasons that will be clarified later, we further assume $\mathbb{E}\left[\log_2 \|\mathbf{x}\|_{\mathbf{P}}^2\right]$ to exist and to be finite for every projection matrix $\mathbf{P}$ with at least one non-zero eigenvalue. This assumption holds for typical communication signals (e.g. QAM or Gaussian signals).

This scheme is suboptimal, but still interesting to understand the performance limits of typical communication systems, which are often designed with a similar separation of channel estimation and data transmission phases. Interestingly, assuming $T = N$, this scheme is shown in [21] to achieve full non-coherent DoF. This happens under the assumption that the channel estimation error vanishes for high SNR. However, in section 3.1.1 we have seen that this condition is not met for $T < N$. The intuition is that in this case the term $\tilde{\mathbf{H}}\mathbf{x}$ acts as an equivalent noise due to channel estimation that limits the system performance, as it do not vanishes for high SNR. More formally, we have the following proposition:

Proposition 3.11. *Let us consider the MU-MIMO block-fading i.i.d. Rayleigh channel model (3.27). By applying the communication strategy with pilot-based channel estimation described above, with training length $1 \leq T < N$, the mutual information of this system has the following asymptotic behaviour:*

$$\mathrm{DoF}_{\mathrm{LMMSE},G>1} := \lim_{\mathrm{SNR}\to\infty} \frac{I(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}})}{\log_2(\mathrm{SNR})} = 0 \tag{3.34}$$

Proof. Let us define the shorthand $I(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}} = \hat{\mathbf{H}}_0) := I(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}}_0)$ By following similar arguments as in [23], we have

$$\begin{aligned} I(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}}_0) =& h(\mathbf{y}|\hat{\mathbf{H}}_0) - h(\mathbf{y}|\mathbf{x},\hat{\mathbf{H}}_0) \\ =& h(\mathbf{y}|\hat{\mathbf{H}}_0) - \mathbb{E}\left[\log_2\left|\pi e\left(\mathbf{R}_{\tilde{\mathbf{H}}\mathbf{x}|\mathbf{x}} + \sigma_w^2\mathbf{I}\right)\right|\right] \\ \leq& \log_2\left|\pi e\left(\hat{\mathbf{H}}_0\mathbf{Q}\hat{\mathbf{H}}_0^H + \mathbf{R}_{\tilde{\mathbf{H}}\mathbf{x}} + \sigma_w^2\mathbf{I}\right)\right| \\ & - \mathbb{E}\left[\log_2\left|\pi e\left(\mathbf{R}_{\tilde{\mathbf{H}}\mathbf{x}|\mathbf{x}} + \sigma_w^2\mathbf{I}\right)\right|\right] \end{aligned}$$

where $\mathbf{Q} := \mathbb{E}[\mathbf{x}\mathbf{x}^H]$ is the signal covariance matrix, satisfying $\mathrm{tr}\{\mathbf{Q}\} = N$. The second equality comes form the fact that $(\mathbf{y}|\mathbf{x},\hat{\mathbf{H}})$ is complex Gaussian distributed with mean $\hat{\mathbf{H}}\mathbf{x}$ and covariance matrix $\mathbf{R}_{\tilde{\mathbf{H}}\mathbf{x}|\mathbf{x}} + \sigma_w^2\mathbf{I}$, and the last inequality is obtained by bounding $(\mathbf{y}|\hat{\mathbf{H}})$ with the differential entropy of a complex Gaussian vector with the same covariance matrix $\mathbf{R}_{\mathbf{y}|\hat{\mathbf{H}}} = \hat{\mathbf{H}}\mathbf{Q}\hat{\mathbf{H}}^H + \mathbf{R}_{\tilde{\mathbf{H}}\mathbf{x}} + \sigma_w^2\mathbf{I}$. We further have

$$\mathbf{R}_{\tilde{\mathbf{H}}\mathbf{x}} := \mathbf{E}\left[\tilde{\mathbf{H}}\mathbf{x}\mathbf{x}^H\tilde{\mathbf{H}}^H\right] = \mathbf{E}\left[\mathbf{x}^H\tilde{\mathbf{R}}\mathbf{x}\right]\mathbf{I} = \mathbf{E}\left[\|\mathbf{x}\|^2_{\tilde{\mathbf{R}}}\right]\mathbf{I},$$

$$\mathbf{R}_{\tilde{\mathbf{H}}\mathbf{x}|\mathbf{x}} := \mathbf{E}\left[\tilde{\mathbf{H}}\mathbf{x}\mathbf{x}^H\tilde{\mathbf{H}}^H|\mathbf{x}\right] = \|\mathbf{x}\|^2_{\tilde{\mathbf{R}}}\mathbf{I},$$

where $\tilde{\mathbf{R}}$ is a shorthand for the estimation error covariance matrix $\mathbf{R}_{\tilde{\mathbf{h}}\tilde{\mathbf{h}}}$ defined in section 3.1.1, and the uncorrelation property results from the assumption that the users channel vectors are independent. Thus we can further develop the mutual information as

$$\begin{aligned} I(\mathbf{x};\mathbf{y}|\hat{\mathbf{H}}_0) \leq& \log_2\left|\frac{\hat{\mathbf{H}}_0\mathbf{Q}\hat{\mathbf{H}}_0^H}{\mathbf{E}\left[\|\mathbf{x}\|^2_{\tilde{\mathbf{R}}}\right] + \sigma_w^2} + \mathbf{I}\right| \\ & + K\log_2\left(\mathbf{E}\left[\|\mathbf{x}\|^2_{\tilde{\mathbf{R}}}\right] + \sigma_w^2\right) \\ & - K\mathbf{E}\left[\log_2\left(\|\mathbf{x}\|^2_{\tilde{\mathbf{R}}} + \sigma_w^2\right)\right]. \end{aligned}$$

Let us now analyze the behaviour of these terms for $\mathrm{SNR}\to\infty$, i.e. for $\sigma_w^2 \to 0$. In the high SNR regime, the LMMSE filter is just a projector $\mathbf{P}_{\mathbf{X}_T}$ on the column space of $\mathbf{X}_T$

(note that, with respect to section 3.1.1, $\mathbf{X}_T$ is here defined transposed), hence

$$\lim_{\sigma_w^2 \to 0} \tilde{\mathbf{R}} = \mathbf{E}\left[(\mathbf{h} - \mathbf{P}_{\mathbf{X}_T}\mathbf{h})(\mathbf{h} - \mathbf{P}_{\mathbf{X}_T}\mathbf{h})^H\right] = \mathbf{I} - \mathbf{P}_{\mathbf{X}_T} =: \mathbf{P}^{\perp}_{\mathbf{X}_T},$$
$$\lim_{\sigma_w^2 \to 0} \|\mathbf{x}\|^2_{\tilde{\mathbf{R}}} = \|\mathbf{x}\|^2_{\mathbf{P}^{\perp}_{\mathbf{X}_T}},$$
$$\lim_{\sigma_w^2 \to 0} \mathbf{E}\left[\|\mathbf{x}\|^2_{\tilde{\mathbf{R}}}\right] = \mathbf{E}\left[\|\mathbf{x}\|^2_{\mathbf{P}^{\perp}_{\mathbf{X}_T}}\right] = \mathrm{tr}\left\{\mathbf{P}^{\perp}_{\mathbf{X}_T}\mathbb{E}[\mathbf{x}\mathbf{x}^H]\mathbf{P}^{\perp}_{\mathbf{X}_T}\right\} = \mathrm{tr}\left\{\mathbf{P}^{\perp}_{\mathbf{X}_T}\right\} =: \sigma_e^2,$$

which leads to

$$\begin{aligned}\lim_{\sigma_w^2 \to 0} I(\mathbf{x}; \mathbf{y}|\hat{\mathbf{H}}_0) &\leq \log_2 \left| \frac{\hat{\mathbf{H}}_0 \mathbf{Q} \hat{\mathbf{H}}_0^H}{\sigma_e^2} + \mathbf{I} \right| + K \log_2 \sigma_e^2 - K\mathbf{E}\left[\log_2 \|\mathbf{x}\|^2_{\mathbf{P}^{\perp}_{\mathbf{X}_T}}\right] \\ &= \log_2 \left| \frac{\hat{\mathbf{H}}_0 \mathbf{Q} \hat{\mathbf{H}}_0^H}{\sigma_e^2} + \mathbf{I} \right| + c,\end{aligned}$$

where c is a non-negative finite constant that does not depend on the channel realization (it depends on K, and on the signal and pilot design). The constant c is non-negative because, by applying the Jensen's inequality:

$$\mathbf{E}\left[\log_2 \|\mathbf{x}\|^2_{\mathbf{P}^{\perp}_{\mathbf{X}_T}}\right] \leq \log_2 \mathbf{E}\left[\|\mathbf{x}\|^2_{\mathbf{P}^{\perp}_{\mathbf{X}_T}}\right] = \log_2 \sigma_e^2.$$

Moreover, it is finite because of the assumption on the distribution of $\mathbf{x}$.

By considering now

$$\lim_{\mathrm{SNR} \to \infty} I(\mathbf{x}; \mathbf{y}|\hat{\mathbf{H}}) \leq \mathbb{E}\left[\log_2 \left| \frac{\hat{\mathbf{H}} \mathbf{Q} \hat{\mathbf{H}}^H}{\sigma_e^2} + \mathbf{I} \right|\right] + c,$$

the theorem follows immediately.

It is interesting to underline that

$$\sigma_e^2 = \mathrm{tr}\left\{\mathbf{P}^{\perp}_{\mathbf{X}_T}\right\} = \mathrm{rank}\left\{\mathbf{P}^{\perp}_{\mathbf{X}_T}\right\} = N - T > 0,$$

thus, it is evident that the system is bottle necked by a non-vanishing residual noise that corresponds exactly to the number of channel dimensions that are not correctly estimated. □

Chapter 4

Spatial Covariance Matrix Conversion via Projection Methods

4.1 Overview

4.1.1 Problem Description

Knowledge of second-order statistics of channels, in particular in the form of spatial covariance matrices, is crucial for many algorithms for Massive MIMO systems operating in the FDD mode, as it provides useful long-term information that can be exploited for beamforming and for CSI acquisition.

Current MIMO systems usually obtain DL covariance information via feedback of the estimated covariance matrix from the UE, but in the massive MIMO regime operating in FDD mode this approach can be problematic because of the large overhead. More precisely, the high overhead of this scheme applied to FDD massive MIMO systems is mainly due the following two problems:

- Direct estimation of $\mathbf{R}^d$ from DL channel realizations obtained via conventional pilot-based DL channel estimation schemes suffers from unacceptably high training overhead, as described in Chapter 3.
- The large dimensionality of $\mathbf{R}^d$ imposes a high covariance feedback overhead. The estimation of the covariance matrix is done at the UE because, if available, the CSIT at the BTS is usually heavily quantized to reduce the overhead.

Although in chapter 2 we have seen that it is actually possible to reduce the impact of the training overhead, the conventional scheme based on DL covariance estimation

cannot avoid the feedback overhead.

Against this background, in this chapter we propose a novel technique to infer $\mathbf{R}^d$ from the observed UL covariance $\mathbf{R}^u$. This approach has many benefits compared to traditional mechanisms for covariance estimation. Among them, the most relevant are:

- Continuous covariance feedback from the UE is eliminated, thus the proposed scheme is an efficient enabling technique for channel estimation algorithms based on the $\mathbf{R}^d$, such as those described in Section 3.3.
- If long-term beamforming techniques based on $\mathbf{R}^d$ are used (e.g., for mm-Waves systems [24]), then the DL training could in principle be completely eliminated from the system.
- Operators can immediately apply the proposed scheme to boost the already implemented beamforming and CSI acquisition algorithms in perfect compliance with current standards, because the proposed mechanism for DL covariance estimation is completely transparent to the UEs.

4.1.2 State-of-the-art Solutions

Related state-of-the-art solutions in literature include:

- [25] Resampling of $\mathbf{R}^u$ for a uniform linear array (ULA) at a different wavelength by using cubic splines.
- [26] (and the follow-up study [27]) Interpolation of $\mathbf{R}^d$ from $\mathbf{R}^u$ and a dictionary of stored $(\mathbf{R}^d, \mathbf{R}^u)$ pairs measured at different UE locations.
- [28] Definition of a frequency calibration matrix obtained via a truncated Fourier series representation of the so called *angular power spectrum* (APS).

The main underlying assumption of the state-of-the-art techniques and of this work is the channel reciprocity in the angular domain, which is here modeled with the frequency invariance property of the APS (see Sect. 2.1). Because of its importance in building the connection between $\mathbf{R}^d$ and $\mathbf{R}^u$, the core part of this work is devoted to the development of an accurate technique for APS estimation given $\mathbf{R}^u$. Unlike related studies, we formalize the problem as a *convex feasibility* problem, so that we can apply very effective solutions based on projection methods on an infinite-dimensional Hilbert space. The resulting scheme is shown to outperform existing solutions under multiple point of views (see Sect. 4.5.2). In fact, it achieves estimation accuracy and flexibility comparable to [26] (the most accurate and robust algorithm considered here) but with complexity comparable to [25] and [28] (simple dictionary-less approaches).

4.1.3 Proposed Technique: Overview and Main Assumptions

Let us recall the channel spatial covariance models introduced in Chapter 2. As described in Section 4.1.1, the goal of the proposed algorithm is to estimate $\mathbf{R}^d$ from the observed UL covariance $\mathbf{R}^u$. By focusing for simplicity on the unpolarized antennas case, the main idea can be summarized into two steps as follows:

1. We obtain an estimate $\hat{\rho}$ of the APS ρ based on the knowledge of $\mathbf{R^u}$ (see its expression in Table 2.1) and known properties of ρ.
2. We compute an estimate of $\mathbf{R^d}$ by using its expression given in Table 2.1, and by substituting ρ with its estimate $\hat{\rho}$.

In particular, the APS estimation problem in the first step is addressed by formalizing it as a convex feasibility problem. We propose two versions of a set-theoretic approach differing in the definition of the solution space, leading to two variants of the proposed algorithm with different accuracy-complexity trade-offs.

In this work, analytic or experimental knowledge of the array responses $\mathbf{a}^u$ and $\mathbf{a}^d$ is assumed; this knowledge is cell-independent and it holds for the entire lifetime of the antenna array. In Section 4.2, to simplify the description of the proposed solutions, we focus on a simple 2D channel model. This part follows closely our preliminary study [29]. Later, in Section 4.6, we extend these results to 3D environments with dual-polarized antennas, by using the approach given by our follow-up study [30].

4.2 Two Algorithms for Covariance Conversion

4.2.1 Algorithm 1: APS Estimation via Projection onto a Linear Variety

Let us consider for simplicity a 2D scenario with unpolarized antennas, so that the spatial covariance matrix expressions are given in the one-variable integral form (azimuth only) as in equation (2.8) and (2.9). Let us further rewrite expression (2.8) as a system of equations of the form

$$r^u_m = \int_{-\pi}^{\pi} \rho(\theta) g^u_m(\theta) d\theta \quad m = 1 \dots M, \quad M = 2N^2, \tag{4.1}$$

where $r^u_m \in \mathbb{R}$ is the mth element of $\mathbf{r}^u := \text{vec}\left(\left[\Re\{\mathbf{R}^u\}\ \Im\{\mathbf{R}^u\}\right]\right)$, and $g^u_m : [-\pi, \pi] \longrightarrow \mathbb{R}$ is the mth coordinate function of the corresponding vectorization of the matrix $\mathbf{a}^u(\theta)\mathbf{a}^u(\theta)^H$. In general, since covariance matrices are Hermitian, the number of different (real) equations is at most $N(N-1)$. For complexity reasons, it is possible to modify the definition of the vec$(\cdot)$ operator such that all the duplicated equations of (4.1) are removed, but for notation simplicity in this section this trivial operation is omitted.

Now let $\mathcal{H}$ be the Hilbert space of real functions in $L^2[-\pi,\pi]$ equipped with the inner product

$$\langle f, g\rangle := \int_{-\pi}^{\pi} f(\theta)g(\theta)d\theta.$$

By assumption ρ and g_m^u are members of $\mathcal{H}$, thus (4.1) can can be written as

$$r_m^u = \langle \rho, g_m^u\rangle \quad m = 1\dots M. \tag{4.2}$$

The inverse problem of finding ρ given g_m^u and r_m^u, $m = 1\dots M$, is obviously ill-posed. Nevertheless, we can use a set-theoretic paradigm [31–34] to obtain an estimate of ρ based on the available information. More precisely, in this paradigm, we estimate ρ by solving

$$\text{find } \rho^* \in V := \cap_{m=1}^{M} V_m \neq \emptyset, \tag{4.3}$$

where $V_m := \{\rho \in \mathcal{H} : \langle \rho, g_m^u\rangle = r_m^u\}$ for $m = 1\dots M$.

The above problem is a feasibility problem involving simple hyperplanes V_m, $m = 1,\dots,M$, so we can easily restrict the candidate solution to keep the resulting algorithm simple and to avoid solutions corresponding to high-energy signals, which are unlikely to be sent in practical systems. In particular, among all the possible solutions of (4.3) (all equivalent based on only the information given by (2.8)), we choose the minimum norm solution

$$\hat{\rho} = \arg\min_{\rho^* \in V} \|\rho^*\|,$$

which corresponds to the orthogonal projection $P_V(0)$ of the zero vector onto the linear variety V. This projection has the following well-known closed-form expression [35, Chapter 3]:

$$\hat{\rho}(\theta) = \sum_{m=1}^{M} \alpha_m g_m^u(\theta), \tag{4.4}$$

where $\boldsymbol{\alpha} := [\alpha_1 \dots \alpha_M]$ is a solution to the linear system

$$\mathbf{r}^u = \mathbf{G}^u\boldsymbol{\alpha}, \tag{4.5}$$

$$\mathbf{G}^u = \begin{bmatrix} \langle g_1^u, g_1^u\rangle & \langle g_1^u, g_2^u\rangle & \dots & \langle g_1^u, g_M^u\rangle \\ \langle g_2^u, g_1^u\rangle & \langle g_2^u, g_2^u\rangle & \dots & \langle g_2^u, g_M^u\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle g_M^u, g_1^u\rangle & \langle g_M^u, g_2^u\rangle & \dots & \langle g_M^u, g_M^u\rangle \end{bmatrix},$$

which is guaranteed to have at least one solution. Moreover, from the projection theorem, all solutions give the unique projection $\hat{\rho}$.

We obtain an estimate of $\mathbf{R}^d$ by replacing ρ in the DL equivalent of (4.2) with its estimate $\hat{\rho}$ obtained in (4.4):

$$\hat{r}_m^d = \langle \hat{\rho}, g_m^d\rangle = \sum_{l=1}^{M} \alpha_l \langle g_l^u, g_m^d\rangle \quad m = 1\dots M, \tag{4.6}$$

which can be rewritten in matrix form as

$$\hat{\mathbf{r}}^d = \mathbf{Q}\boldsymbol{\alpha},$$

where $\hat{\mathbf{r}}^d$ is an estimate of the vector $\mathbf{r}^d := \text{vec}(\left[\Re\{\mathbf{R}^d\}\ \Im\{\mathbf{R}^d\}\right])$, $\boldsymbol{\alpha}$ is a solution to the linear system (4.5) given the UL measurements $\mathbf{r}^u = \mathbf{G}^u\boldsymbol{\alpha}$ as mentioned above, and

$$\mathbf{Q} = \begin{bmatrix} \langle g_1^d, g_1^u\rangle & \langle g_1^d, g_2^u\rangle & \dots & \langle g_1^d, g_M^u\rangle \\ \langle g_2^d, g_1^u\rangle & \langle g_2^d, g_2^u\rangle & \dots & \langle g_2^d, g_M^u\rangle \\ \vdots & \vdots & \ddots & \vdots \\ \langle g_M^d, g_1^u\rangle & \langle g_M^d, g_2^u\rangle & \dots & \langle g_M^d, g_M^u\rangle \end{bmatrix}.$$

It is important to underline that both $\mathbf{G}^u$ and $\mathbf{Q}$ depend only on the array geometry, and they can thus be computed or measured only once for the entire system lifetime.

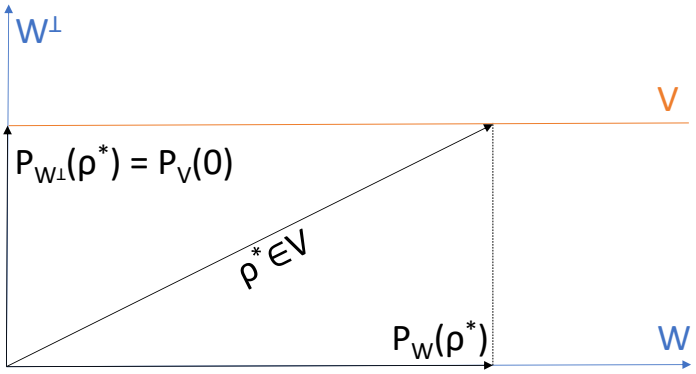


Figure 4.1: Geometrical representation of the solution to (4.3)

Geometrical Representation

A geometrical representation of the solution to (4.3) is given in Figure 4.1, and it can be be explained by noticing that[1]

$$V = \rho + \text{span}(g_1^u, \dots, g_M^u)^\perp,$$

since V is a linear variety defined by the intersection of M infinite-dimensional hyperplanes $V_m = \rho + \text{span}(g_m^u)^\perp$, and thus it can be expressed as a translation of the subspace

$$W = \cap_{m=1}^M \text{span}(g_m^u)^\perp = \text{span}(g_1^u, \dots, g_M^u)^\perp.$$

[1]Here the notation $\text{span}(x_1, \dots, x_k)$ must be intended as the subspace of $\mathcal{H}$ spanned by $x_1, \dots, x_k \in \mathcal{H}$

4.2.2 Algorithm 2: Exploiting Further Properties of the APS

In many applications, additional knowledge about the APS ρ is available (for example, support information). Given that this prior knowledge can be expressed in terms of closed convex sets, it is then possible to narrow the set of candidate solutions of (4.3) to obtain more accurate APS estimates. By separating the real and imaginary part of (2.8), and by working in the space of real functions, the previous algorithm in Sect. 4.2.1 already implicitly takes into account the knowledge that ρ is real valued. In the following, we propose an extension of the previous algorithm by considering that, being a power spectrum, ρ is always non-negative. More precisely, we look at the problem

$$\text{find } \rho^* \in C := V \cap Z, \tag{4.7}$$

where V is the linear variety defined in (4.3) and $Z = \{\rho \in \mathcal{H} : (\forall \theta \in [-\pi, \pi]) \quad \rho(\theta) \geq 0\}$ is the closed convex set of non-negative functions in $\mathcal{H}$. A solution to (4.7) can be found by applying one of the many existing iterative projection methods for convex feasibility problems available in literature. These methods typically produce a sequence $(\rho^{(i)})_{i\in\mathbb{N}} \subset \mathcal{H}$ such that $\rho^{(i)} \rightharpoonup \rho^* \in C$. In particular, we use the following fast iterative method called *extrapolated alternating projection method (EAPM)* (which can also be interpreted as a particular case of an adaptive projected subgradient method [33, Example 5]), which is given by [36]

$$\rho^{(i+1)} = \rho^{(i)} + \nu K_i \left[P_V(P_Z(\rho^{(i)})) - \rho^{(i)} \right] \quad (\forall i \in \mathbb{N}), \tag{4.8}$$

where $\nu \in (0, 2)$ is a step size, and K_i is the extrapolation parameter defined as

$$K_i = \begin{cases} \dfrac{\|P_Z(\rho^{(i)}) - \rho^{(i)}\|^2}{\|P_V(P_Z(\rho^{(i)})) - \rho^{(i)}\|^2}, & \text{if } \rho^{(i)} \notin Z \\ 1, & \text{if } \rho^{(i)} \in Z \end{cases}.$$

The initial condition $\rho^{(0)} \in V$ can be arbitrary, but here it is set to $\rho^{(0)} = P_V(0)$, the solution proposed in section 4.2.1. The projection $P_V : \mathcal{H} \to \mathcal{H}$ onto the set V is given by [35, Chapter 3]

$$P_V(x) = x - \sum_{m=1}^{M} \beta_m g_m^u + P_V(0),$$

with $\boldsymbol{\beta} := [\beta_1 \dots \beta_M]$ being a solution to the linear system $\mathbf{b} = \mathbf{G}^u\boldsymbol{\beta}$ where the mth element of $\mathbf{b}$ is given by $b_m = \langle x, g_m^u \rangle$ and g_m^u, G^u are defined in section (4.2.1). The projection $P_Z : \mathcal{H} \to \mathcal{H}$ is given by [32, p. 284]

$$P_Z(x) = \begin{cases} x(\theta), & \text{if } x(\theta) \geq 0 \\ 0, & \text{otherwise} \end{cases}.$$

Now, by proceeding along the same lines as in section 4.2.1, an estimate of $\mathbf{R^d}$ can be obtained by

$$\hat{r}_m^d = \langle \hat{\rho}, g_m^d \rangle \quad m = 1 \dots M.$$

which has the same form as (4.6), except that $\hat{\rho}$ results from (4.8).

4.3 Imperfect UL Covariance Knowledge

The algorithms presented in Section 4.2 are designed by assuming perfect knowledge of $\mathbf{R}^u$. In this section instead we consider a practical scenario in which the BTS has access only to the UL sample covariance

$$\bar{\mathbf{C}}^u := \frac{1}{N_s}\sum_{n=1}^{N_s} \hat{\mathbf{h}}^u[n](\hat{\mathbf{h}}^u[n])^H$$

computed from a limited number N_s of channel estimates defined as

$$\hat{\mathbf{h}}^u[n] = \mathbf{h}^u[nM_s] + \mathbf{z}[n], \quad \mathbf{z}[n] \sim \mathcal{CN}(\mathbf{0}, \sigma_z^2\mathbf{I}),$$

obtained by using the UL estimator described in Section 3.1.2. It is important to underline that the samples are taken with a spacing equal to the coherence time M_c, so that they can be considered independent. This imperfect knowledge leads to a performance degradation of the proposed algorithms. However, in the following, we discuss a correction procedure that can be applied to $\bar{\mathbf{C}}^u$ in order to mitigate these effects.

Let $\mathcal{H}_M$ be the Hilbert space of all $N \times N$ Hermitian matrices whose inner product is defined by $\langle A, B\rangle = \text{trace}(B^H A)$, and let $\mathcal{C}$,$\mathcal{T}$ be the subsets of $\mathcal{H}_M$ composed respectively by positive semi-definite (PSD) and Toeplitz matrices.

In case of Gaussian channels, the matrix $\bar{\mathbf{C}}^u$ is a sufficient statistic for

$$\mathbf{C}^u := \mathbf{E}[\hat{\mathbf{h}}^u[n](\hat{\mathbf{h}}^u[n])^H] = \mathbf{R}^u + \sigma_z^2\mathbf{I},$$

and in [37] the matrix

$$\bar{\mathbf{R}}^u := \bar{\mathbf{C}}^u - \sigma_z^2\mathbf{I}$$

is used to obtain the maximum-likelihood (ML)-PSD estimate of $\mathbf{R}^u$ by projecting it onto $\mathcal{C}$ as follows:

$$P_{\mathcal{C}}(\bar{\mathbf{R}}^u) = \mathbf{U}\boldsymbol{\Delta}_+\mathbf{U}^H,$$

with $\mathbf{U}$ and $\boldsymbol{\Delta}_+$ obtained form the eigen-decomposition $\bar{\mathbf{R}}^u = \mathbf{U}\boldsymbol{\Delta}\mathbf{U}^H$, and by defining $\boldsymbol{\Delta}_+ := \max(\boldsymbol{\Delta}, \mathbf{0})$, which is a short-hand for an element-wise standard $\max(\cdot,\cdot)$ operator over real numbers.

In this work, this operation is applied also in the case of non-Gaussian channels (e.g. the channel models presented in Chapter 2) as a correction algorithm able to extract a PSD estimate of $\mathbf{R}^u$ from the noisy sample covariance matrix $\bar{\mathbf{C}}^u$, even tough it does not necessarily correspond to the ML-PSD estimate of $\mathbf{R}^u$.

Moreover, according to the specific array geometry, the covariance matrix often shows an additional structure on top of the positive semi-definiteness. Thus, it is reasonable

to further process $P_{\mathcal{C}}(\bar{\mathbf{R}}^u)$ to restore the envisioned structure. The description of this procedure is left for the following sections, where the specific implementation of the algorithms for ULA and UPA are discussed.

4.4 Implementation for Uniform Linear Array

In this section we discuss an implementation of the proposed schemes to a uniform linear array (ULA) with N antennas at the BTS. Recall that the array response of a ULA is given by

$$\mathbf{a}(\theta) = \frac{1}{\sqrt{N}} \left[1 \;\; e^{j2\pi \frac{d}{\lambda} \sin\theta} \;\; \dots \;\; e^{j2\pi \frac{d}{\lambda}(N-1)\sin\theta}\right]^T,$$

where $d \in \mathbb{R}$ and $\lambda \in \mathbb{R}$ denote, respectively, the inter-antenna spacing and the carrier wavelength.

4.4.1 Analytical expressions for $\mathbf{G}^u$ and $\mathbf{Q}$

ULAs are not able to distinguish among a DoA/DoD θ and its reciprocal $\theta + \pi$, so we assume that the multipath components are confined to the interval $[-\pi/2, \pi/2]$, and we modify the definition of the of the scalar product for $\mathcal{H}$ accordingly, such that $\langle f, g\rangle = \int_{-\pi/2}^{\pi/2} f(\theta)g(\theta)d\theta$. This assumption is supported by the fact that real systems often work with a similar or even narrower cell sectorization.

For ULAs, the covariance matrices are Hermitian Toeplitz, so they can be completely represented by their first columns.

Proposition 4.1. *By redefining vec*$(\mathbf{A}) := \mathbf{a}_1$*, where* $\mathbf{a}_1$ *indicates the first column of* $\mathbf{A}$*, we can prove that the matrices* $\mathbf{G}^u$ *and* $\mathbf{Q}$ *defined in Sect. 4.2.1 have the following analytical form expressed in terms of the Bessel function of the first kind, zero order* $J_0 : \mathbb{R} \to \mathbb{R}$*:*

$$\mathbf{G}^u = \frac{\pi}{2N^2}\begin{bmatrix}\mathbf{G}_{\Re} & \mathbf{0}\\ \mathbf{0} & \mathbf{G}_{\Im}\end{bmatrix} \quad \mathbf{Q} = \frac{\pi}{2N^2}\begin{bmatrix}\mathbf{Q}_{\Re} & \mathbf{0}\\ \mathbf{0} & \mathbf{Q}_{\Im}\end{bmatrix},$$

where the elements corresponding to the (n, m)*-entries of* $\mathbf{G}_{\Re}$*,* $\mathbf{G}_{\Im}$*,* $\mathbf{Q}_{\Re}$*,*$\mathbf{Q}_{\Im} \in \mathbb{R}^{N\times N}$ *are given by*

$$\mathbf{G}_{\Re,nm} = J_0(x_{nm}) + J_0(y_{nm}),\ \mathbf{Q}_{\Re,nm} = J_0(p_{nm}) + J_0(q_{nm}),$$
$$\mathbf{G}_{\Im,nm} = J_0(x_{nm}) - J_0(y_{nm}),\ \mathbf{Q}_{\Im,nm} = J_0(p_{nm}) - J_0(q_{nm}),$$

and where

$$x_{nm} = 2\pi\frac{d}{\lambda^u}(n-m), \quad p_{nm} = 2\pi d\left(\frac{n-1}{\lambda^d} - \frac{m-1}{\lambda^u}\right),$$
$$y_{nm} = 2\pi\frac{d}{\lambda^u}(n+m-2), \quad q_{nm} = 2\pi d\left(\frac{n-1}{\lambda^d} + \frac{m-1}{\lambda^u}\right).$$

Proof. (sketch) By dropping UL/DL superscripts for simplicity, the chosen vectorization operation gives:

$$g_m(\theta) = \Re[v_m(\theta)] \quad m = 1 \dots N,$$
$$g_m(\theta) = \Im[v_{m-N}(\theta)] \quad m = N+1 \dots 2N,$$

where

$$v_n(\theta) := [\mathbf{a}(\theta)\mathbf{a}(\theta)^H]_{n,1} = \frac{1}{N} e^{j2\pi \frac{d}{\lambda}(n-1)\sin\theta},$$

with $n = 1 \dots N$. To compute $\mathbf{G}^u$, we evaluate the following four inner products:

$$\langle \Re[v_n], \Re[v_m] \rangle = \frac{\pi}{2N^2} \left[J_0(x) + J_0(y) \right],$$
$$\langle \Im[v_n], \Im[v_m] \rangle = \frac{\pi}{2N^2} \left[J_0(x) - J_0(y) \right],$$
$$\langle \Re[v_n], \Im[v_m] \rangle = 0,$$
$$\langle \Im[v_n], \Re[v_m] \rangle = 0,$$

evaluated at

$$x_{nm} = 2\pi \frac{d}{\lambda}(n-m),$$
$$y_{nm} = 2\pi \frac{d}{\lambda}(n+m-2),$$

with $n, m = 1 \dots N$.

The expression for $\mathbf{Q}$ is obtained similarly. □

4.4.2 Improving the Estimation of the UL Covariance Matrix

The direct feeding of either $\bar{\mathbf{R}}^u$ or the ML-PSD estimate defined in Section 4.3 as input to the proposed algorithms may result in poor performance because the Toeplitz assumption imposed by the ULA is not satisfied.

To overcome this problem, we propose to feed as input the projection of $\bar{\mathbf{R}}^u$ onto the set $\mathcal{T}_+ := \mathcal{C} \cap \mathcal{T}$. More precisely, we choose as input a solution of

$$\hat{\mathbf{R}}^u = \arg \min_{\mathbf{X} \in \mathcal{T}_+} \|\mathbf{X} - \bar{\mathbf{R}}^u\|_F.$$

Since the projections on $\mathcal{C}$ and $\mathcal{T}$ are known [38] and easy to compute, it is possible to compute $\hat{\mathbf{R}}^u$ by applying standard methods such as the Dykstra's or Haugazeau's algorithm [39, Chapter 20]. In this work, we use the approach described in [38], which solves

$$\text{find } \mathbf{X}^* \in \mathcal{T}_+ \cap C_3, \quad C_3 = \{\mathbf{X} \in \mathcal{H}_\mathcal{M} : \|\mathbf{X} - \bar{\mathbf{R}}^u\|_F \leq \delta\},$$

where δ is a tunable error tolerance, by using an alternating projection method producing a sequence convergent to a point in $\mathcal{T}_+ \cap C_3$.

4.5 Comparison with State-of-the-art Techniques

4.5.1 Simulated Scenario

This section presents numerical results of the proposed algorithms. For simplicity, in this first numerical evaluation, we assume the following correlated Rayleigh channel model:

$$\mathbf{h}^u[k] \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}^u), \quad \mathbf{h}^d[k] \sim \mathcal{CN}(\mathbf{0}, \mathbf{R}^d),$$

with spatial covariance matrices $\mathbf{R}^u$ and $\mathbf{R}^d$ given by (2.8) and (2.9). We simulate a typical model for the APS in cellular environments inspired by the GSCM channel model described in Section 2.2.1, where ρ is assumed to be composed by a weighted superposition of probability density functions as follows:

$$\rho(\theta) = \sum_{q=1}^{Q} f_q(\theta)\alpha_q.$$

As an example, in the following we assume Gaussian distributions $f_q \sim \mathcal{N}\left(\phi_q, \Delta_q^2\right)$ with ϕ_q uniformly drawn from $[-\pi/3, \pi/3]$ and standard deviation (also called *angular spread*) Δ_q uniformly drawn from $[3°, 8°]$, weights α_q uniformly drawn from $[0, 1]$ and further normalized such that $\sum_{q=1}^{Q} \alpha_q = 1$, and Q uniformly drawn from $\{1, 2, 3, 4, 5\}$. These statistical quantities are introduced to emulate the effect of different scattering patterns corresponding to random user locations.

A ULA is assumed for the BTS operating at UL/DL carrier wavelengths of $\lambda = 3 \cdot 10^8/f$ with $f = 1.8$ Ghz and 1.9 Ghz respectively. The antenna spacing d is set to half UL wavelength.

Channel realizations are given by $\mathbf{h} = \mathbf{R}^{\frac{1}{2}}\mathbf{w}$, with $\mathbf{w} \sim \mathcal{CN}(\mathbf{0}, \mathbf{I})$. The BTS is assumed to have access only to a UL sample covariance matrix computed from $N_s = 1000$ noisy channel estimates as described in Sections 4.3, with estimation noise power computed from a given per-antenna $\text{SNR}_{est} := \frac{\mathbf{E}[|h_n|^2]}{\sigma_z^2} = \frac{1}{N\sigma_z^2}$.

4.5.2 Numerical Results

The performance of the two algorithms defined in Sect. 4.2.1 and 4.2.2 are compared with the algorithms proposed in [25], [27], and [28], referred, respectively, to *splines*-based, *dictionary*-based, and *Fourier*-based. The DL sample covariance, obtained with the same number of samples and SNR as for the UL, is used as a baseline. For fairness, all the sample covariances used in this comparison are corrected with the Toeplitzation procedure outlined in Sect. 4.4.2.

The accuracy of an estimate $\hat{\mathbf{R}}$ of $\mathbf{R}$ is evaluated in terms of the mean square error $\text{MSE} := \mathbf{E}[e^2(\mathbf{R}, \hat{\mathbf{R}})]$, where $e(\cdot, \cdot)$ is a given error metric. In particular, we consider:

- The normalized Euclidean distance

$$e(\mathbf{R}, \hat{\mathbf{R}}) := \frac{\|\mathbf{R} - \hat{\mathbf{R}}\|_F}{\|\mathbf{R}\|_F}.$$

- [26, 40] The affine invariant distance in the Riemannian space of PSD matrices

$$e(\mathbf{R}, \hat{\mathbf{R}}) := \| \log(\mathbf{R}^{\frac{1}{2}} \hat{\mathbf{R}}^{-1} \mathbf{R}^{\frac{1}{2}}) \|_F.$$

- [40] The Grassmanian distance between the principal subspaces $\mathbf{V}_p$,$\hat{\mathbf{V}}_p$ defined from $\mathbf{R}$,$\hat{\mathbf{R}}$ by considering their eigenvectors corresponding to the minimum number p of largest eigenvalues λ_n satisfying

$$\frac{\sum_{n=1}^{p} \lambda_n}{\sum_{n=1}^{N} \lambda_n} \geq 95\%.$$

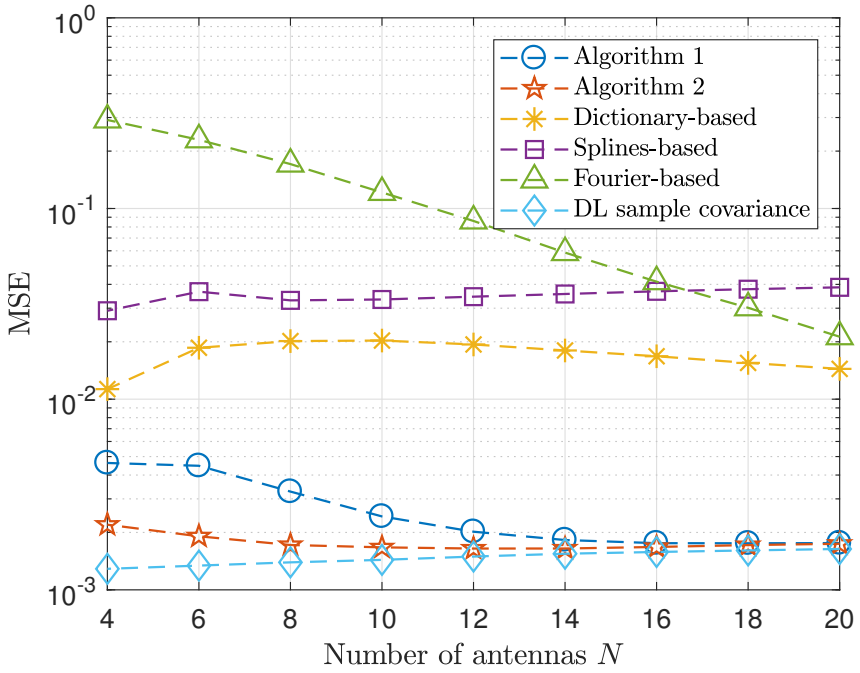
  The metric is then

$$e(\mathbf{R}, \hat{\mathbf{R}}) := \sqrt{\sum_{n=1}^{p} \gamma_n^2},$$

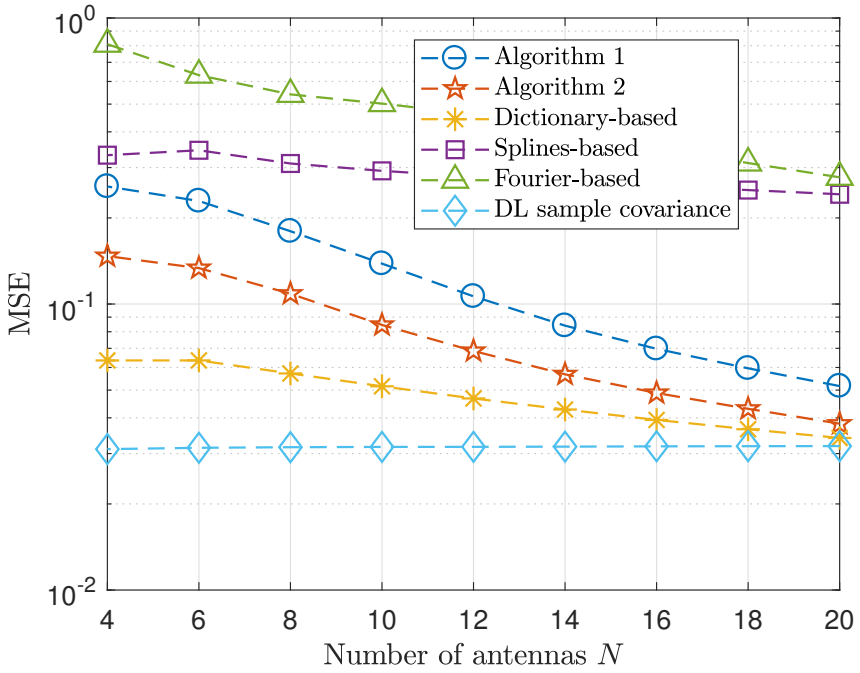
  where $\cos(\gamma_n)$ are the eigenvalues of $\mathbf{V}_p^H \hat{\mathbf{V}}_p$. This metric is particularly meaningful for the massive MIMO channel estimation problem, where a reliable signal subspace knowledge plays a crucial role.

The statistical mean is then obtained by Monte-Carlo simulations. For every Monte-Carlo run, a new APS and SNR_{est} level $\in [10, 30]$ (dB) are drawn.

Figure 4.2 compares the algorithms for different numbers of BTS antennas N. The performance of both proposed algorithms approach that of the DL sample covariance estimator as the number of constraints in the convex feasibility problem grows with N. The performance of both algorithms are comparable or better (depending on the metric and on the number of antennas) than the *dictionary*-based method, which in principle can achieve extremely high accuracy given that the dictionary is sufficiently large (here we used 1000 training samples). However, the proposed algorithms are dictionary-less, thus not requiting any overhead for dictionary acquisition. *Algorithm 1* has the same very low complexity as the *Fourier*-based method, but it achieves a much better accuracy. Compared to *Algorithm 1*, *Algorithm 2* shows better performance, especially in the low N region, where the prior information about the positivity of the APS becomes important. However, the performance gains are achieved at a cost of a higher complexity, which is due to the fact that the algorithm requires the numerical evaluation of integrals of the form $\int_{-\pi}^{\pi} x(\theta) d\theta$.

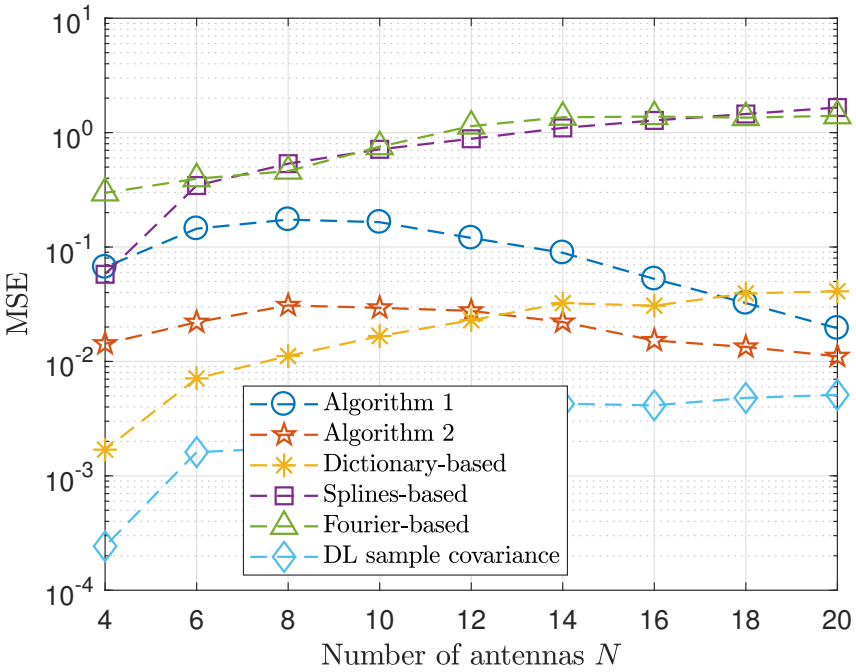


(a) Normalized Euclidean distance.



(b) Affine invariant distance.

Figure 4.2: Comparison of different DL covariance estimators vs number of BTS antennas N.



(c) Principal subspaces distance.

Figure 4.2: Comparison of different DL covariance estimators vs number of BTS antennas N. (cont.)

4.6 Extension to 3D Environments and Dual-Polarized Antennas

In this section we show that the proposed algorithms can be naturally extended to channel models and antenna arrays more complex than those described above by simply selecting an appropriate inner product. The analysis in chapter 2 will play a key role in the derivation of the extensions.

4.6.1 3D Environment, Unpolarized Antennas

Let us consider the expressions for the channel covariance matrices given in Table 2.1, by first focusing on the case of unpolarized antennas. Similarly to (4.1), let us further rewrite the expression of $\mathbf{R}^u$ as a system of equations of the form

$$r_m^u = \int_\Omega \rho(\boldsymbol{\theta}) g_m^u(\boldsymbol{\theta}) d^2\boldsymbol{\theta} \quad m = 1 \dots M, \quad M = 2N^2, \tag{4.9}$$

where $r_m^u \in \mathbb{R}$ is the mth element of $\mathbf{r}^u := \mathrm{vec}(\left[\Re\{\mathbf{R}^u\}\, \Im\{\mathbf{R}^u\}\right])$, $g_m^u : \Omega \longrightarrow \mathbb{R}$ is the mth coordinate function of the corresponding vectorization of the matrix $\mathbf{a}^u(\boldsymbol{\theta})\mathbf{a}^u(\boldsymbol{\theta})^H$, and where we recall that $\Omega := [-\pi, \pi] \times [0, \pi]$.

Let us now consider the Hilbert space $\mathcal{H}'$ of bivariate real functions in $L^2[\Omega]$ equipped with the inner product

$$\langle f, g \rangle := \int_\Omega f(\boldsymbol{\theta}) g(\boldsymbol{\theta}) d^2\boldsymbol{\theta}.$$

By assumption, ρ and g_m^u are members of $\mathcal{H}'$, thus (4.9) can can be written as

$$r_m^u = \langle \rho, g_m^u \rangle \quad m = 1 \dots M,$$

which is the analog of the expression in (4.2) to 3D environments with unpolarized antennas. A similar expression obviously holds for the DL. Hence, the proposed algorithms can be directly applied to 3D scenarios, by just changing the evaluation of the inner products, which now involves a double integral instead of an integral involving univariate functions.

4.6.2 3D Environment, Dual-Polarized Antennas

To consider the case of dual-polarized antennas, similarly to (4.1), we rewrite the expression of $\mathbf{R}^u$ as a system of equations of the form

$$r_m^u = \int_\Omega \rho_V(\boldsymbol{\theta}) g_{V,m}^u(\boldsymbol{\theta}) d^2\boldsymbol{\theta} + \int_\Omega \rho_H(\boldsymbol{\theta}) g_{H,m}^u(\boldsymbol{\theta}) d^2\boldsymbol{\theta}, \quad m = 1 \dots M, \quad M = 2N^2, \tag{4.10}$$

where $r^u_m \in \mathbb{R}$ is the mth element of $\mathbf{r}^u := \text{vec}(\left[\Re\{\mathbf{R}^u\}\ \Im\{\mathbf{R}^u\}\right])$, $g^u_{(\cdot),m} : \Omega \longrightarrow \mathbb{R}$ is the mth coordinate function of the corresponding vectorization of the matrix $\mathbf{a}^u_{(\cdot)}(\boldsymbol{\theta})\mathbf{a}^u_{(\cdot)}(\boldsymbol{\theta})^H$, and where we recall that $\Omega := [-\pi, \pi] \times [0, \pi]$.

Let us consider now the Hilbert space $\mathcal{H}'' := L^2[\Omega] \times L^2[\Omega]$ of tuples of bivariate real functions equipped with the inner product

$$\langle (f_V, f_H), (g_V, g_H) \rangle := \int_\Omega f_V(\boldsymbol{\theta}) g_V(\boldsymbol{\theta}) d^2\boldsymbol{\theta} + \int_\Omega f_H(\boldsymbol{\theta}) g_H(\boldsymbol{\theta}) d^2\boldsymbol{\theta}.$$

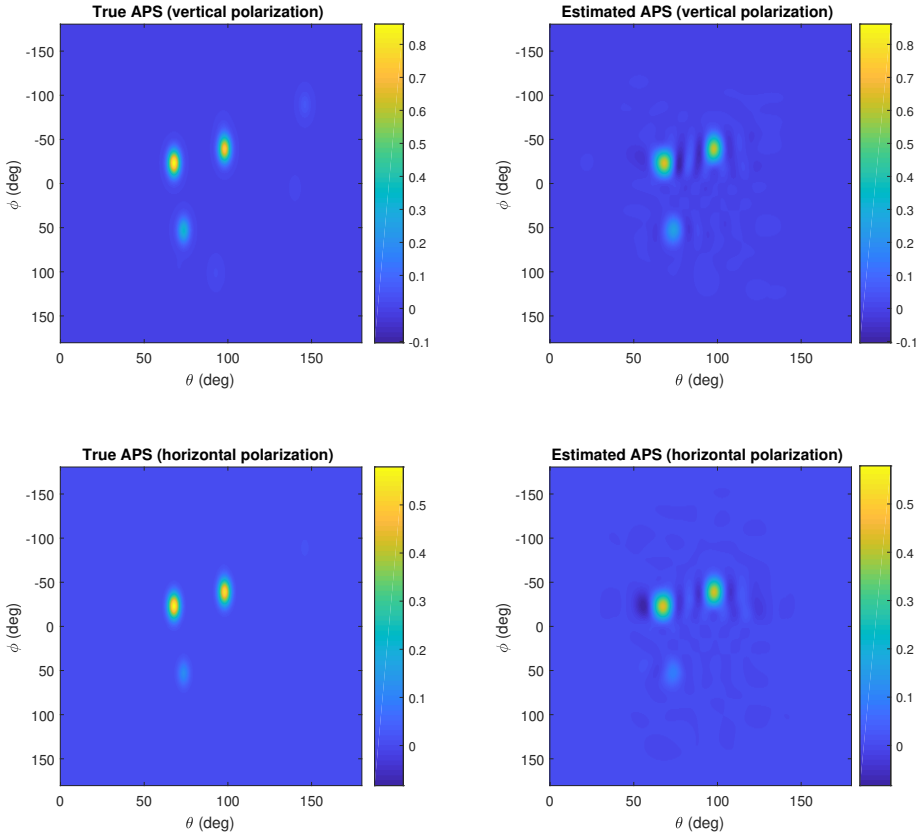
By assumption (ρ_V, ρ_H) and $(g^u_{V,m}, g^u_{H,m})$ are members of $\mathcal{H}''$, thus (4.10) can can be written as

$$r^u_m = \langle (\rho_V, \rho_H), (g^u_{V,m}, g^u_{H,m}) \rangle \quad m = 1 \dots M,$$

which is once again equivalent to the expression (4.2). As a result, by simply redefining the inner products as shown above, we can also apply the proposed algorithms in 3D scenarios with dual-polarized antennas.

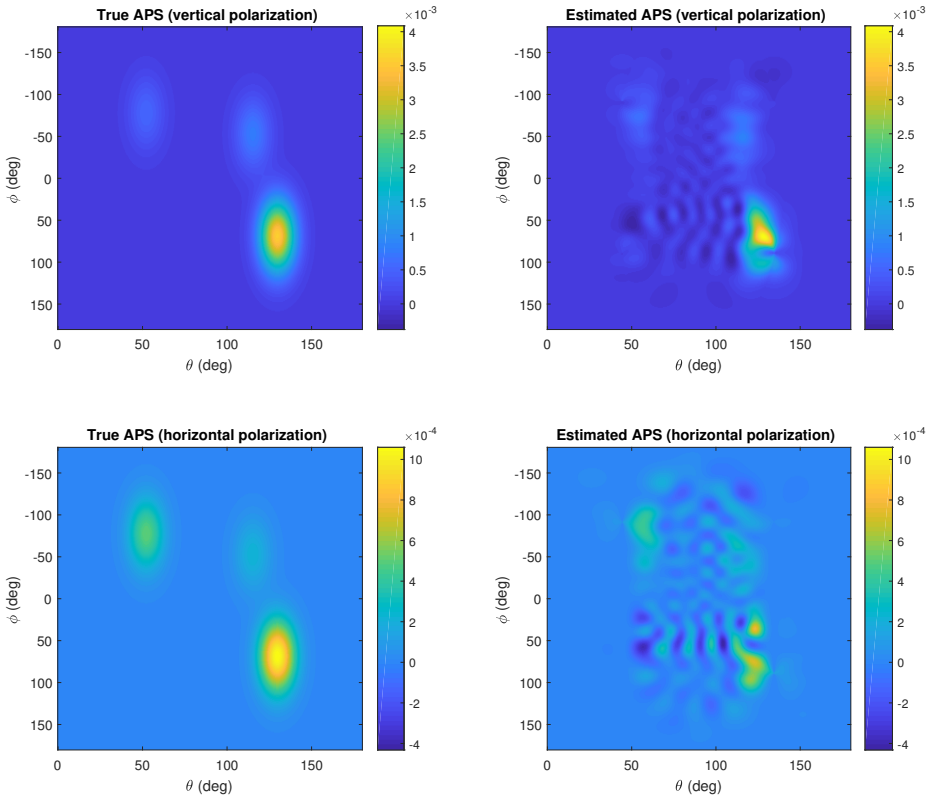
Example: Figure 4.3 shows 2 examples of APS estimation in a 3D environment with a dual polarized antenna array. The plots are obtained by applying the algorithm in Section 4.2.1 to a simplified version of the simulation scenario detailed in Section 4.8.

- Example 4.3a shows that the developed algorithm is able to reconstruct the given vertical and horizontal APS from the covariance matrix $\mathbf{R}_u$.
- Example 4.3b illustrates that the reconstruction may show artifacts because the true APS is just one of the possible solutions of the convex feasibility problem (recall that the inverse problem is ill-posed). Nevertheless, in our simulations, the APSs reconstructed with the proposed algorithms have typically been accurate enough so that the error of the covariance conversion is small by considering different metrics. More importantly, we will show soon that the conversion error has had little influence on the achievable rate of the communication links in the simulated scenarios.



(a) Good APS estimation.

Figure 4.3: Examples of V-APS and H-APS estimation



(b) APS estimation with artifacts.

Figure 4.3: Examples of V-APS and H-APS estimation. (cont.)

4.7 Implementation for Uniform Planar Array with Pairs of Cross-Polarized Antennas

A popular BTS array structure adopted in practical massive MIMO applications is the uniform planar array (UPA) with cross-polarized antennas, defined as a rectangular grid of identical and equispaced antenna elements, each of them composed of a pair of two vertically polarized antennas with a polarization slant of $\pm 45°$.

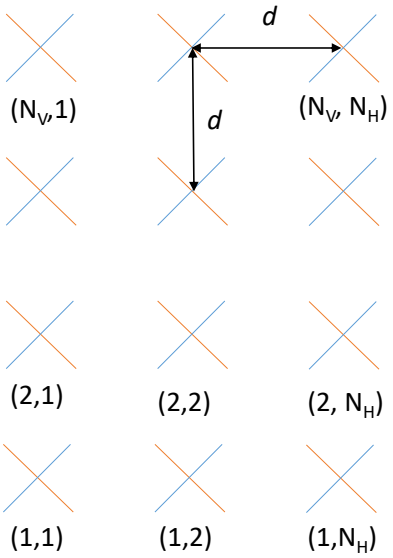


Figure 4.4: Array structure of UPA with pairs of X-pol. antennas.

The antenna array structure can be visualized in Figure 4.4, where N_V and N_H denotes respectively the number of vertical and horizontal elements, and where d denotes the horizontal and vertical inter-antenna spacing. Let us denote by $x(u, v, 1)$ the antenna in position (u, v), $u = 1, \ldots, N_V$ and $v = 1, \ldots, N_H$, with $+45°$ polarization slant, and with $u(u, v, 2)$ the co-located antenna with $-45°$ polarization slant. Let us further denote with $a_{V,1}(\boldsymbol{\theta})$ and $a_{H,1}(\boldsymbol{\theta})$ the vertical and horizontal radiation pattern of the $+45°$ polarized antennas, and with $a_{V,2}(\boldsymbol{\theta})$, $a_{H,2}(\boldsymbol{\theta})$ the vertical and horizontal radiation patters for the $-45°$ polarized antennas.

4.7.1 Structure of the Array Response

By omitting for simplicity the time index and the UL/DL superscripts, let us define the channel vector $\mathbf{h} \in \mathbf{C}^{2N_V N_H \times 1}$ as follows:

$$\mathbf{h} := \begin{bmatrix} \mathbf{h}_1^T & \mathbf{h}_2^T \end{bmatrix}^T,$$

where the channel coefficient for antenna $x(u, v, k)$ corresponds to the nth element of the vector $\mathbf{h}_k \in \mathbf{C}^{N_V N_H \times 1}$, with $n = (u-1)N_H + v$. Under this definition, the array responses for the vertical and horizontal polarization, including the effect of the radiation patterns, are given by

$$\begin{aligned} \mathbf{a}_V(\boldsymbol{\theta}) &:= \begin{bmatrix} \mathbf{a}_{V,1}^T(\boldsymbol{\theta}) & \mathbf{a}_{V,2}^T(\boldsymbol{\theta}) \end{bmatrix}^T \in \mathbf{C}^{2N_V N_H \times 1}, \\ \mathbf{a}_H(\boldsymbol{\theta}) &:= \begin{bmatrix} \mathbf{a}_{H,1}^T(\boldsymbol{\theta}) & \mathbf{a}_{H,2}^T(\boldsymbol{\theta}) \end{bmatrix}^T \in \mathbf{C}^{2N_V N_H \times 1}, \end{aligned} \tag{4.11}$$

$$\begin{aligned} \mathbf{a}_{V,1}(\boldsymbol{\theta}) &:= a_{V,1}(\boldsymbol{\theta}) e^{j\boldsymbol{\Psi}(\boldsymbol{\theta})} \in \mathbf{C}^{N_V N_H \times 1}, \\ \mathbf{a}_{V,2}(\boldsymbol{\theta}) &:= a_{V,2}(\boldsymbol{\theta}) e^{j\boldsymbol{\Psi}(\boldsymbol{\theta})} \in \mathbf{C}^{N_V N_H \times 1}, \\ \mathbf{a}_{H,2}(\boldsymbol{\theta}) &:= a_{H,2}(\boldsymbol{\theta}) e^{j\boldsymbol{\Psi}(\boldsymbol{\theta})} \in \mathbf{C}^{N_V N_H \times 1}, \\ \mathbf{a}_{H,1}(\boldsymbol{\theta}) &:= a_{H,1}(\boldsymbol{\theta}) e^{j\boldsymbol{\Psi}(\boldsymbol{\theta})} \in \mathbf{C}^{N_V N_H \times 1}, \end{aligned}$$

where we used the shorthand

$$e^{j\boldsymbol{\Psi}(\boldsymbol{\theta})} := \begin{bmatrix} e^{j\Psi_1(\boldsymbol{\theta})} & e^{j\Psi_2(\boldsymbol{\theta})} & \dots & e^{j\Psi_{N_V N_H}(\boldsymbol{\theta})} \end{bmatrix},$$

and where $\Psi_n(\boldsymbol{\theta}) = \Psi_n(\theta_1, \theta_2)$ is the geometry-only dependent phase term of an antenna in position (u, v), given by:

$$\Psi_{(u-1)N_H+v}(\boldsymbol{\theta}) = 2\pi \frac{d}{\lambda} \left[(u-1)\cos(\theta_1) + (v-1)\sin(\theta_1)\sin(\theta_2) \right].$$

4.7.2 Structure of the Covariance Matrix

With the structure of the array response defined in Section 4.7.1, by considering the expressions given in Table 2.1, and by assuming without loss of generality that $N_V \geq N_H$, the dual-polarized covariance matrix has the following structure:

$$\mathbf{R} = \begin{bmatrix} \mathbf{B}_1 & \mathbf{B}_2^H \\ \mathbf{B}_2 & \mathbf{B}_3 \end{bmatrix} \in \mathbf{C}^{2N_V N_H \times 2N_V N_H}, \tag{4.12}$$

where every macro-block $\mathbf{B}_l \in \mathbf{C}^{N_V N_H \times N_V N_H}$, $l = 1, 2, 3$, is Hermitian and it has the following block structure:

$$\mathbf{B}_l = \begin{bmatrix} \mathbf{B}_{l,1} & & & & \\ \mathbf{B}_{l,2} & \mathbf{B}_{l,1} & & & \\ \mathbf{B}_{l,3} & \mathbf{B}_{l,2} & \mathbf{B}_{l,1} & & \\ \vdots & \vdots & \vdots & \ddots & \\ \mathbf{B}_{l,N_V} & \dots & \mathbf{B}_{l,3} & \mathbf{B}_{l,2} & \mathbf{B}_{l,1} \end{bmatrix}, \tag{4.13}$$

where every block $\mathbf{B}_{l,i} \in \mathbf{C}^{N_H \times N_H}$, $i = 1, \dots N_V$ has identical diagonal entries b_{li}, and every block $\mathbf{B}_{l,1}$ is Hermitian Toeplitz.

4.7.3 Efficient Covariance Vectorization

Because of its particular structure described in Section 4.7.2, it is possible to describe completely the covariance matrix $\mathbf{R}$ by using the blocks $\mathbf{B}_{l,i} \in \mathbf{C}^{N_H \times N_H}$, $i = 1, \dots N^V$, $l = 1, 2, 3$. Furthermore, the blocks $\mathbf{B}_{l,1}$ can be represented by only N_H complex numbers because of the Hermitian Toeplitz structure, while the blocks $\mathbf{B}_{l,i}$, $i \neq 1$, can be represented by $N_H^2 - (N_H - 1)$ complex numbers because only one diagonal entry is sufficient. Overall, this means that it is possible to completely represent $\mathbf{R}$ with

$$M' = 3(N_H + (N_V - 1)(N_H^2 - N_H + 1)$$

complex numbers, or equivalently with

$$M = 2M' = 6(N_H + (N_V - 1)(N_H^2 - N_H + 1))$$

real numbers.

By using an appropriate vectorization operation using only the minimum number of reals that are sufficient to reconstruct the full matrix $\mathbf{R}$, we can reduce the size of the vectorized covariance $\mathbf{r}$ from $M = \mathcal{O}(N_V^2 N_H^2)$ down to $M = \mathcal{O}(N_V N_H^2)$. As a result, the algorithm using only the vectorization with minimum dimension has substantially lower complexity compared to the structure-unaware algorithm presented in Section 4.2.1. We note that the above representation holds for $N_V \geq N_H$. If $N_V \leq N_H$, it is possible to find a similar representation that requires $M = \mathcal{O}(N_H N_V^2)$ real numbers by simply exchanging the role of N_H and N_V in Section 4.7.2.

4.7.4 Improving the Estimation of the UL Covariance Matrix

Similarly to Section 4.4.2, the estimation $\hat{\mathbf{R}}^u$ of the covariance matrix $\mathbf{R}^u$ obtained from the sample covariance as described in Section 4.3 can be further improved by exploiting its particular structure. In particular, in this work, we propose to substitute an element (i, j) of the matrix $\hat{\mathbf{R}}^u$ with the arithmetic average of all its elements that are assumed to be identical according to the structure of $\mathbf{R}^u$ described in Section 4.7.2. It can be shown that this operation can be seen as a projection of $\hat{\mathbf{R}}^u$ onto the subspace of matrices with the structure described above.

4.8 Simulations with Realistic Channel Model

In this section we evaluate the proposed algorithms by simulating a realistic communication scenario between a BTS equipped with an 8x4 cross-polarized antenna array and single antenna UEs in a typical macro-cell environment. We adopt the multipath channel model described in Section 2.23, thus considering propagation in 3D environments and the effects of polarized antennas. The results shown here are valid for both

narrow-band systems and for wide-band OFDM systems. The content of this section can be summarized as follows:

- First, we describe the simulation scenario, highlighting the connections between the proposed simulation environment and the standard for MIMO simulations in [15].
- Second, we evaluate the UL/DL spatial covariance conversion error of the proposed algorithms by considering two common metrics.
- Third, we evaluate the algorithms in terms of the achievable rate when applied to a simple pilot-based communication scheme with MRC beamforming in the data transmission phase and subspace-based DL channel estimation in the CSI acquisition phase. The adopted rate metric takes into account the joint effect of channel estimation error and DL training overhead. Our results show that the proposed algorithms can greatly reduce the overheard in the DL.
- Finally, we compare the set-theoretic methods with machine learning approaches based on neural networks.

4.8.1 Simulated Scenario

To verify the performance of the algorithms, we consider the following simulation scenario, which follows closely the procedure outlined in [15, Section 7.3]:

1. General parameters:

Carrier frequency (f_c)	1.8 GHz for UL, 1.9 GHz for DL
System type	Narrow-band or wide-band OFDM
BTS	8x4 cross-polarized UPA $d_V = d_H = \lambda_u/2$
UEs	Single antenna, vertically polarized
Antennas radiation pattern	3GPP [15, Section 7.1], 3D-UMa

2. Channel coefficients are given by the narrow-band multipath model (2.23). This multipath model corresponds exactly to the 3GPP model given by [15, Eq. 7.3-22] computed at time $t = 0$, and by considering the first sub-carrier of an OFDM modulation. We recall that every sub-carrier experience the same spatial covariance matrix (Section 2.15), thus, for the purposes of this simulation, considering just the first sub-carrier is sufficient.
3. Cluster powers α_c are drawn uniformly from $[0, 1]$ and further normalized such that $\sum_{c=1}^{N_c} \alpha_c = 1$.
4. The XPRs values K_{ic} are drawn from a log-Normal distribution with parameters $(\mu_{\text{XPR}}, \sigma_{\text{XPR}}) = (7, 3)$[dB]. This is identical to the 3GPP model [15, Sect. 7.3, Step 9], with parameters for 3D-UMa, NLOS propagation.

5. The angles $\boldsymbol{\theta}_{ic} \in [-\pi, \pi] \times [0, \pi]$, $\boldsymbol{\phi}_{ic} \in [-\pi, \pi] \times [0, \pi]$ are generated from the following jointly Gaussian distribution:

$$\begin{aligned} f_c(\boldsymbol{\theta}, \boldsymbol{\phi}) &= f_{\mathrm{BTS},c}(\boldsymbol{\theta}) f_{\mathrm{UE},c}(\boldsymbol{\phi}), \\ f_{\mathrm{BTS},c}(\boldsymbol{\theta}) &\sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{BTS}}, \boldsymbol{\sigma}^2_{\mathrm{BTS}} \mathbf{I}), \\ f_{\mathrm{UE},c}(\boldsymbol{\theta}) &\sim \mathcal{N}(\boldsymbol{\mu}_{\mathrm{UE}}, \boldsymbol{\sigma}^2_{\mathrm{UE}} \mathbf{I}), \end{aligned}$$

where the clusters means and angular spreads

$$\begin{aligned} \boldsymbol{\mu}_{\mathrm{BTS}} &:= [\mu_{\mathrm{BTS},a} \quad \mu_{\mathrm{BTS},z}], \\ \boldsymbol{\sigma}^2_{\mathrm{BTS}} &:= [\sigma^2_{\mathrm{BTS},a} \quad \sigma^2_{\mathrm{BTS},z}], \\ \boldsymbol{\mu}_{\mathrm{UE}} &:= [\mu_{\mathrm{UE},a} \quad \mu_{\mathrm{UE},z}], \\ \boldsymbol{\sigma}^2_{\mathrm{UE}} &:= [\sigma^2_{\mathrm{UE},a} \quad \sigma^2_{\mathrm{UE},z}], \end{aligned}$$

are drawn as follows:

$$\begin{aligned} &\mu_{\mathrm{BTS},a} \sim \mathcal{U}\left[-\frac{2}{3}\pi, \frac{2}{3}\pi\right], \quad \mu_{\mathrm{UE},a} \sim \mathcal{U}\left[-\frac{2}{3}\pi, \frac{2}{3}\pi\right], \\ &\mu_{\mathrm{BTS},z} \sim \mathcal{U}\left[\frac{\pi}{4}, \frac{3}{4}\pi\right], \quad \mu_{\mathrm{UE},z} \sim \mathcal{U}\left[\frac{\pi}{4}, \frac{3}{4}\pi\right], \\ &\sigma_{\mathrm{BTS},a} \sim \mathcal{U}\left[3^\circ, 5^\circ\right], \quad \sigma_{\mathrm{UE},a} \sim \mathcal{U}\left[5^\circ, 10^\circ\right], \\ &\sigma_{\mathrm{BTS},z} \sim \mathcal{U}\left[1^\circ, 3^\circ\right], \quad \sigma_{\mathrm{UE},z} \sim \mathcal{U}\left[3^\circ, 5^\circ\right]. \end{aligned}$$

Note: the subscripts $(\cdot)_a$ and $(\cdot)_z$ denote the quantities referring to, respectively, the azimuth and the zenith of the chosen spherical coordinate system. The quantities defined in degrees $(.)^\circ$ should be converted into radians.

We highlight that the proposed algorithms do not require any specific choice of the distribution f_c, given that it is continuous. For this simulation, the above distributions have been motivated by the following considerations:

- Gaussian or Laplacian distributions are often suggested in the literature [12, 15] to model the angular distribution of NLOS clusters. A possible variation of the simulated environment would be to include a LOS cluster (with higher power) with uniform angular distribution.
- The distribution of the cluster spread is chosen to reflect the fact that the zenith spread is typically smaller than the azimuth spread, and that the BTS spread is typically smaller than the UE spread. See, for example, the 3GPP document [15] and the study in [12].
- The independence of $\boldsymbol{\theta}_{ic}$ from $\boldsymbol{\phi}_{ic}$ is introduced with the sole purpose of simplifying the analytic evaluation of the true covariance matrix.

- The cluster means are bounded to avoid dealing with the problem of angular wrapping in case the distributions have a significant tail outside the range $[-\pi, \pi] \times [0, \pi]$, which makes the analytic evaluation of the true covariance matrix more difficult.

The adopted procedure has some differences with that suggested by the 3GPP document [15, Sect. 7.3, Step 8]. In particular, here all angles are independently drawn, so we do not use the 3GPP simplification in which only the cluster mean is random, while the subpaths angles are defined from a predefined table (which has been possibly suggested in [15, Sect. 7.3, Step 8] to decrease the computational complexity of simulations).

6. The angles $\boldsymbol{\theta}$ and $\boldsymbol{\phi}$ are given in the global coordinate system (GCS). The BTS array response is given assuming that its local coordinate system (LCS) is aligned with the GCS. To simulate different UE antenna orientation, the UE antenna array response is given by assuming that its LCS is a 3D rotation of the GCS parametrized by (α, β, γ) as described in 3GPP [15, Sect. 5.1.3], where $\alpha, \beta, \gamma \sim \mathcal{U}\left[0, \frac{\pi}{6}\right]$.

7. For the time evolution of the channel we refer to the statistical model outlined in section 2.1.1, which is further detailed specifically for the adopted channel model in section 2.4. This is different from that suggested by the 3GPP document [15], where the time evolution is modeled deterministically as a phase term depending on the user trajectory and speed. Note: we further assume the random UE antenna orientation to be a slowly-varying parameter.

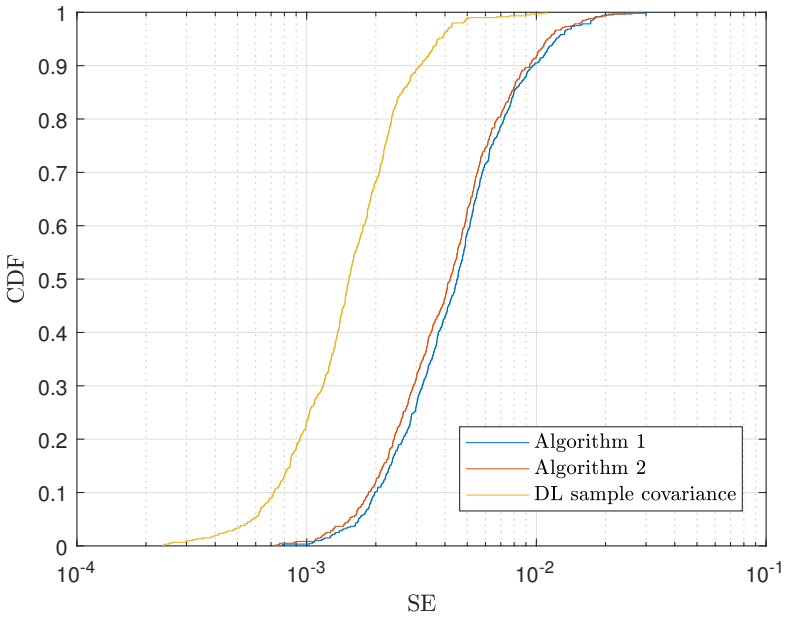
8. The pathloss is indirectly modeled by assuming a given system SNR level.

4.8.2 Performance of the Proposed Covariance Conversion Algorithms

The BTS is assumed to have access to estimated UL covariance matrices $\hat{\mathbf{R}}^u$ obtained as described in Section 4.3 from $N_s = 1000$ channel estimates, with a noise level defined by setting an average per-antenna SNR_{est} to $\text{SNR}_{est} := \dfrac{\text{tr}\{\mathbf{R}^u\}}{N\sigma_z^2} = 10$ [dB], where $N = 2N_V N_H$ denotes the number of BTS antennas.

The proposed algorithms are implemented in their variant for 3D environments with dual-polarized antennas as described in section 4.6.2, and by exploiting the efficient implementation for UPA described in section 4.7.

The accuracy of an estimate $\hat{\mathbf{R}}^d$ of $\mathbf{R}^d$ is evaluated in terms of the square error $\text{SE} := e^2(\mathbf{R}^d, \hat{\mathbf{R}}^d)$, where $e(\cdot, \cdot)$ is a given error metric. In particular, we consider as error metrics the normalized Frobenius norm and the 90% principal subspace distance defined in section 4.5.2. The true value of $\mathbf{R}^d$ is computed from the parameters defined in section 4.8.1 by using 2.24.



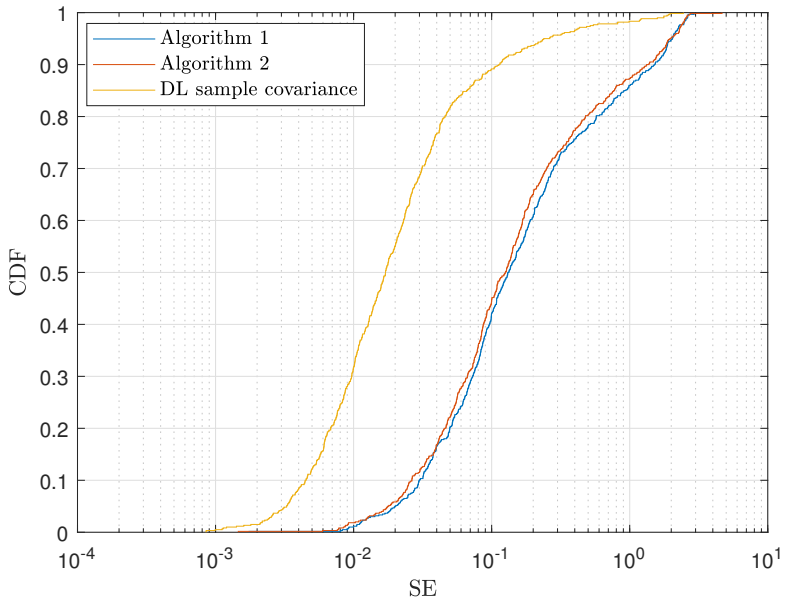
(a) Normalized Frobenius norm

Figure 4.5: Empirical CDF of the covariance conversion error.

We compare the proposed algorithms with an approach that directly estimates the DL covariance matrix based on downlink pilots. More specifically, in this last approach we used the same procedure that is used for the estimation of $\hat{\mathbf{R}}^u$, including the correction steps that take into account the structure of the covariance matrix for UPA. The results are shown in Figure 4.5, which shows the empirical cumulative distribution function (CDF) of the SE for the two chosen metrics, obtained by drawing independent realizations of the quantities that are assumed to stay fixed for a WSS window (i.e. by drawing a new V-APS and H-APS). The simulation confirms that the proposed algorithms are able to provide an accurate DL estimate by using only UL training, thus it can be used as an effective solution to the DL channel covariance acquisition problem. Furthermore, as already shown for the preliminary results for 2D systems given in Section 4.5.2, when the number of BTS antennas is large, the performance gap between Algorithm 1 and Algorithm 2 is small.

4.8.3 Application to SU-MIMO pilot-based systems with MRC beamforming

In this section we evaluate the performance of the proposed algorithms in terms of the achievable rate when they are used in following DL SU-MIMO transmission scheme:



(b) 90% subspace distance

Figure 4.5: Empirical CDF of the covariance conversion error. (cont.)

- The first $T < M_c$ slots of the coherence block length M_c are devoted to CSI acquisition by applying the conventional LS estimation scheme described in Section 3.1.1, with pilot sequence $\mathbf{X} = \sqrt{N}\mathbf{V}_T^H$, where $\mathbf{V}_T$ is the matrix of the T strongest eigenvectors of the estimated DL covariance $\hat{\mathbf{R}}^d$, and $N = 2N_V N_H$ is the number of BTS antennas.

- The remaining $M_c - T$ slots are then used for data transmission by using a maximum ratio combining (MRC) precoder $\mathbf{f}_{\text{MRC}} := \dfrac{\hat{\mathbf{h}}^H}{\|\hat{\mathbf{h}}\|}$ based on the estimated DL channel $\hat{\mathbf{h}}$, assumed for simplicity to be fed back from the UE instantaneously and without quantization. The received signal $y \in \mathbb{C}$ for a given time slot in the data transmission scheme can be expressed as follows:

$$y = (\mathbf{f}_{\text{MRC}} d)\mathbf{h} + w = \frac{\hat{\mathbf{h}}^H}{\|\hat{\mathbf{h}}\|}\mathbf{h}d + w,$$

where $d \in \mathbb{C}$ is the input symbol with zero mean and variance $\sigma_d^2 = N$, $\mathbf{x} := \mathbf{f}_{\text{MRC}} d$ is the transmitted signal with power $\mathbb{E}[\mathbf{x}^H\mathbf{x}] = N$, and $w \sim \mathcal{CN}(0, \sigma_w^2)$ is the receiver noise.

It can be verified that this system satisfies a unitary power constraint per transmitted antenna. More precisely, it satisfies

$$\mathbb{E}[\|\mathbf{X}\|_F^2] = NM_c,$$

where $\mathbf{X} \in \mathbb{C}^{T \times M_c}$ denotes the overall transmitted signal in the entire coherence block.

To evaluate the performance of the above transmission scheme, we consider the concept of ergodic achievable rate (similar to the notion of channel ergodic capacity [1], but where the mutual information is constrained by the specific transmission scheme) defined as

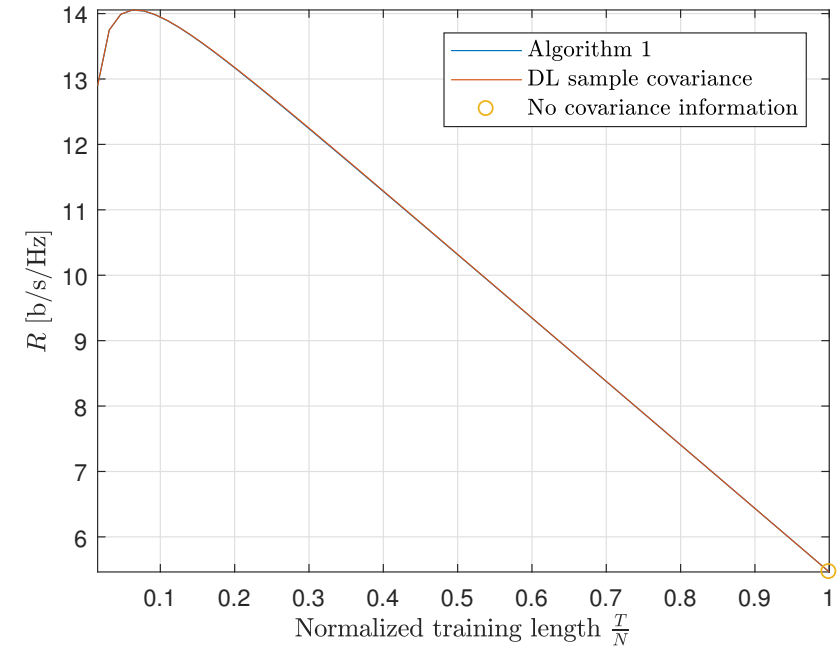
$$R' := \mathbb{E}\left[\log_2\left(1 + \frac{N}{\sigma_w^2}\frac{\hat{\mathbf{h}}^H \mathbf{h}\mathbf{h}^H \hat{\mathbf{h}}}{\|\hat{\mathbf{h}}\|^2}\right)\right] = \mathbb{E}\left[\log_2\left(1 + \frac{N}{\sigma_w^2}\|\mathbf{h}\|^2 \cos^2(\xi)\right)\right],$$

where the term $\cos(\xi) = \dfrac{|\hat{\mathbf{h}}^H \mathbf{h}|}{\|\hat{\mathbf{h}}\|\|\mathbf{h}\|}$ takes into account the channel estimation error. To include the reduction in rate owing to training, we multiply R' by the fraction of M_c that is used for data transmission, obtaining the following rate metric (in [bit/s/Hz]):

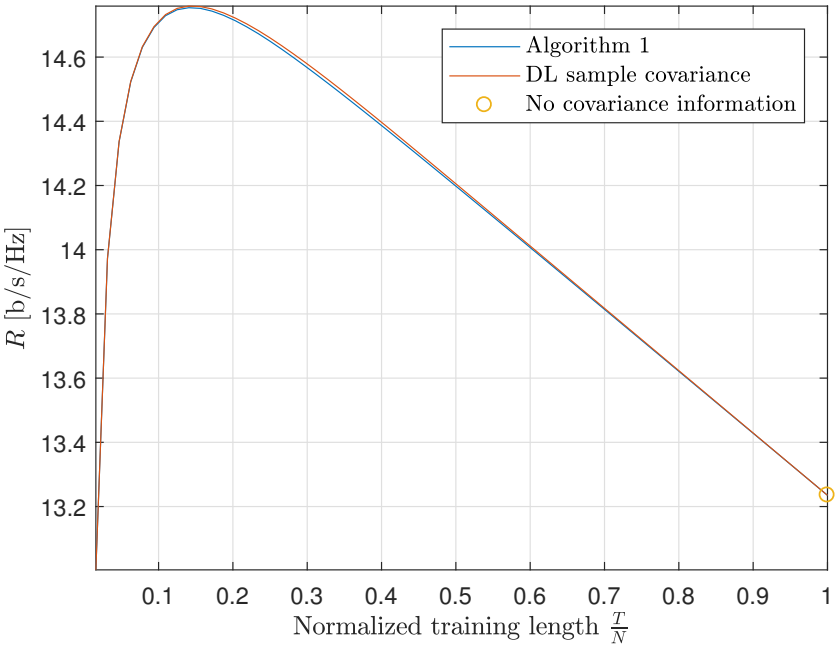
$$R := \left(\frac{M_c - T}{M_c}\right)\mathbb{E}\left[\log_2\left(1 + \frac{N}{\sigma_w^2}\|\mathbf{h}\|^2 \cos^2(\xi)\right)\right]. \tag{4.14}$$

In the following, we simulate realizations of the V-APS and H-APS as described in Section 4.8.2. For every realization of the V-APS and H-APS, we estimate the ergodic achievable rate 4.14 by considering Montecarlo averages of 1,000 samples of channels and their corresponding estimates. We further average the results over the multiple realizations of the V-APSs and H-APSs. The noise level is defined by setting an average per-antenna SNR to $\text{SNR} := \dfrac{\text{tr}\{\mathbf{R}^d\}}{N\sigma_w^2} = 10$ [dB]. Since the performance of Algorithm 1 and 2 are similar owing to the large number of antennas, and the former algorithm has much lower complexity, we omit the performance of the latter.

The results are shown in Figure 4.6, where the rate metric 4.14 is computed for a coherence block length of $M_c = 100, 500, 1000$ resource blocks, and by letting the training length T to vary in the interval $[1, N]$. We clearly see that it is crucial to decrease the training overhead by exploiting the information obtained from the estimation of the DL covariance matrix, especially if the coherence time becomes comparable to the number of BTS antennas (we recall that for this simulation we have $N = 64$). In particular, the results show that the proposed method for UL/DL covariance conversion have no or negligible performance loss with respect to the use of the DL sample covariance if applied with the practical transmission scheme analyzed in this section, but without the additional covariance feedback overhead.

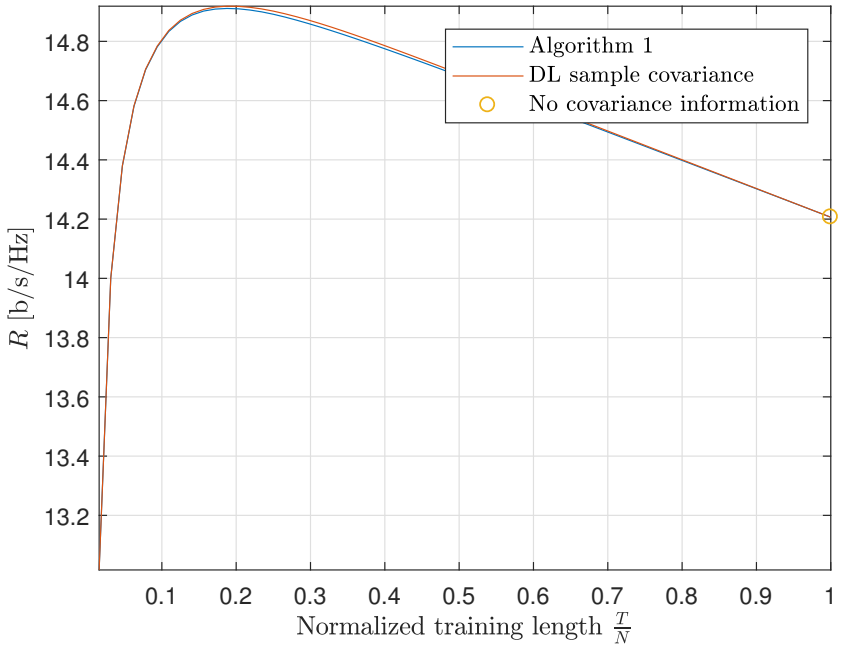


(a) Coherence block length: $M_c = 100$.



(b) Coherence block length: $M_c = 500$.

Figure 4.6: Ergodic achievable rate vs training sequence length.



(c) Coherence block length: $M_c = 1000$.

Figure 4.6: Ergodic achievable rate vs training sequence length. (cont.)

4.8.4 Comparison with Machine Learning Solutions

In this section we repeat the experiments in Section 4.8.2 and Section 4.8.3 to compare the performance of the proposed covariance conversion scheme with solutions that rely on a pre-stored dictionary of covariance matrices $(\hat{\mathbf{R}}^u, \hat{\mathbf{R}}^d)$, similar to the approach proposed in [26, 27], already analyzed with the preliminary results in Section 4.5.2.

In particular, we consider two approaches that used $L = 300$ samples of uplink and downlink covariance matrices:

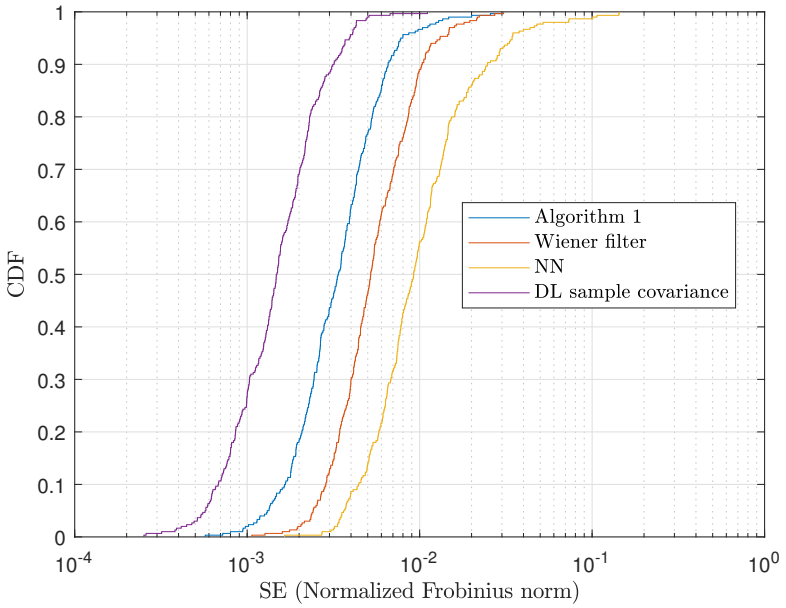
- A neural network with one dense hidden layer containing 181 neurons and one dense output layer containing 8,192 neurons. The standard rectified linear unit (ReLU) functions are used as the activation functions of the hidden layer, and linear activation functions are used in the output layer. The network is trained by concatenating the real and imaginary parts of vectorized versions of the uplink and downlink covariance matrices. The mean square error is used as the cost function, and the network is trained with the Adam algorithm in batches of size 100 in 1,000 passes (epochs). In our experiments, changing the depth of the network or the number of layers have either reduced the performance of the network or resulted in negligible performance gains.

- A Wiener filter approach. More precisely, denote the training set by $\{(\mathbf{x}_i = \mathrm{vec}(\hat{\mathbf{R}}_i^u), \mathbf{y}_i^d = \hat{\mathbf{R}}_i^d)\}_{i=1,\ldots,L}$. Given $\mathbf{x} = \mathrm{vec}(\hat{\mathbf{R}}^u)$, we use $\mathbf{y} = \mathrm{vec}(\hat{\mathbf{R}}^d) = \mathbf{M}\mathrm{vec}(\hat{\mathbf{R}}^u)$ as the estimate of the DL covariance, where $\mathbf{M}$ is a square matrix solving the following optimization problem:

$$\mathbf{M} \in \arg\min \sum_{i=1}^{L} \|\mathbf{M}\mathbf{x}_i - \mathbf{y}_i\|_2^2.$$

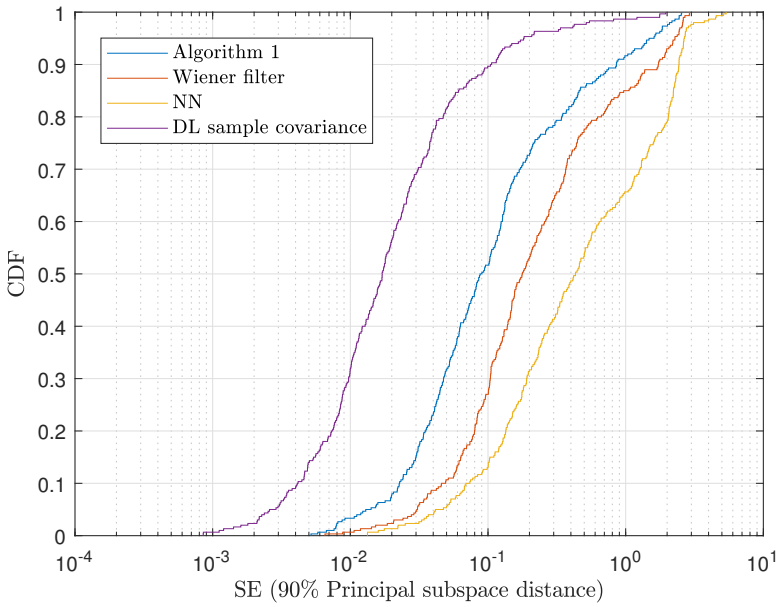
For fairness, the covariance matrices obtained with both approaches have been corrected by projecting those matrices onto the subspace of matrices with the structure induced by the array geometry (see Section 4.7.4, and note that the the uplink and downlink covariance matrices have the same structure).

The results for the covariance conversion error are shown in Figure 4.7. The results for the ergodic achievable rate in case of the SU-MIMO transmission scheme described in Section 4.8.3 are shown in Figure 4.8, and note that we focus on the interval around the optimal achievable rate. As we can see from the figures, the set-theoretic method outperforms the traditional learning schemes without requiring any training sets, which can be difficult to obtain in real systems.



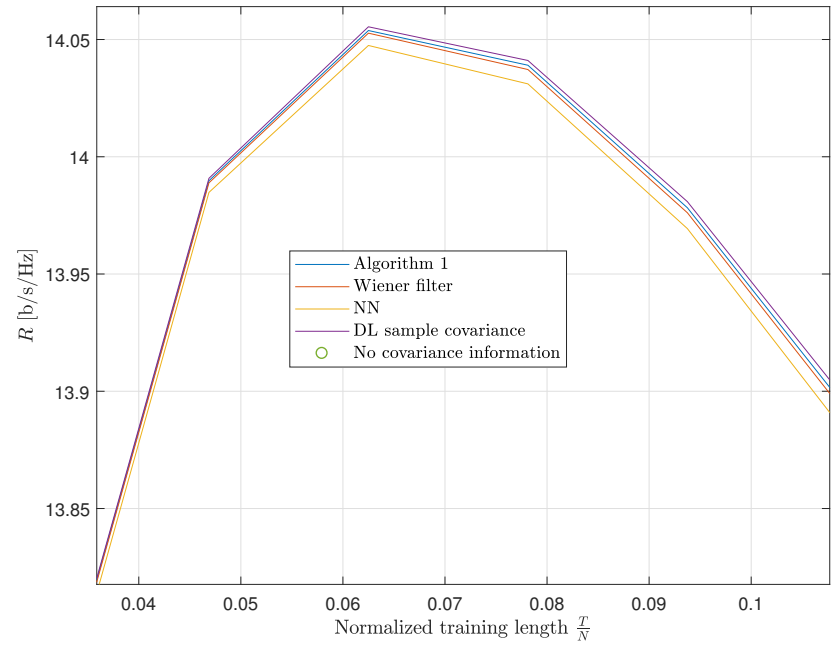
(a) Normalized Frobenius norm

Figure 4.7: Empirical CDF of the covariance conversion error.

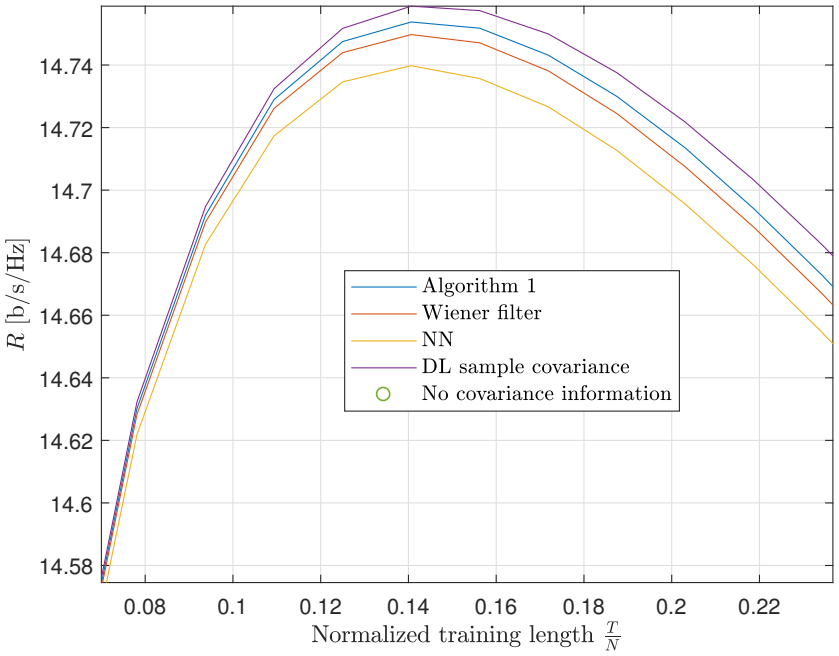


(b) 90% subspace distance

Figure 4.7: Empirical CDF of the covariance conversion error. (cont.)

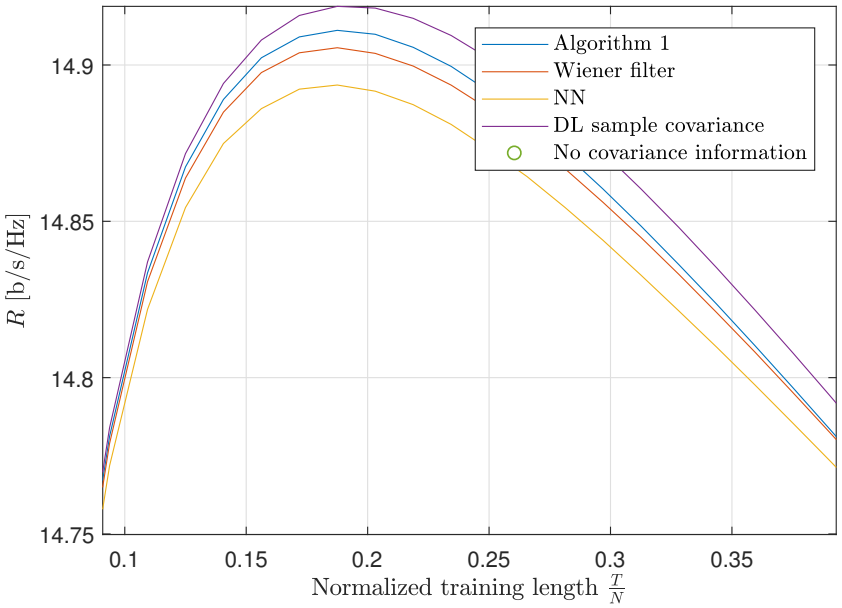


(a) Coherence block length: $M_c = 100$.



(b) Coherence block length: $M_c = 500$.

Figure 4.8: Ergodic achievable rate vs training sequence length.



(c) Coherence block length: $M_c = 1000$.

Figure 4.8: Ergodic achievable rate vs training sequence length. (cont.)

Chapter 5

Conclusion

5.1 Discussion of the Results

This thesis gives an in-dept overview of the role of channel correlation in FDD massive MIMO systems, particularly in relation to the DL CSI acquisition problem, and it proposes practical techniques for its exploitation.

The analysis in Chapter 3 provides a comprehensive analysis of lest-square (LS) and minimum mean square error (MMSE) estimation performance in case of correlated channels, and derives optimal conditions on the pilot sequence design for SU-MIMO systems. In particular, the importance of the knowledge of the DL spatial covariance matrix is stressed, as it provides information about the subspace on which the channel vector lies, which is fundamental for effectively reducing the channel estimation overhead.

Because of its fundamental importance, the structure of the spatial covariance matrix is analyzed in Chapter 2. Interesting expressions linking the UL and DL covariance matrices by means of a frequency independent function called *angular power spectrum* (APS) are derived, by considering several different channel models that cover narrowband and wide-band OFDM systems, realistic propagation with polarization effects, and different modeling philosophies considering either discrete or continuous scattering environments. We point out that the expression derived for polarized antenna arrays does not appear in the literature. Related expressions are instead available for the unpolarized antennas case, which is however an unrealistic assumption for many practical systems.

In Chapter 4, as an enabling technique for correlation-based CSI estimation techniques, a novel DL spatial covariance matrix estimation scheme is proposed, and its advantages with respect to competing schemes is discussed. We believe that the proposed invention makes a relevant contribution to the general optimization effort that massive MIMO systems are currently undergoing, in the process towards commercialization.

The concept of Degrees of Freedom (DoF) is also adopted in Chapter 3, to justify from

an Information Theoretical (IT) point of view the benefits of channel correlation in massive MIMO systems. However, the analysis is incomplete, as it provides an overview of the fundamental limits of conventional massive MIMO systems just for channels with independent fading. The performance bound for correlated channels are just briefly outlined, but a discussion about their achievability from an IT point of view is completely missing. An interesting future line of research would be to link the effect of channel estimation to the notion of non-coherent capacity also in case of correlated channels.

A complete performance analysis of the training overhead reduction has been given just for the LS estimator. An equivalent analysis for the MMSE estimator is missing. Another important point that has not be considered in this work is the extension of the results for SU-MIMO to MU-MIMO systems. However, we believe that the methodology proposed for SU-MIMO in Chapter 3 represents a useful starting point in that direction.

Finally, another interesting research line, that here has just been briefly introduced, is to further explore the connection between compressed sensing based CSI acquisition techniques and the spatial correlation of the channel.

5.2 Achievements

The work for this thesis lead to the following achievements:

- A first conference paper [29], accepted at IEEE ICASSP, April 2018.
- A patent, currently undergoing the application process, requested by Huawei China.
- A second conference paper [30], submitted.

5.3 Acknowledgements

This thesis has been carried out in the department of Wireless Communications and Networks, Fraunhofer Heinrich Hertz Institute, Berlin, during the period July 2017 to December 2017.

A preliminary version of this thesis has been successfully defended in January 2018, under the supervision of Prof. David Gesbert, as a conclusion of a one year study period at Eurecom, Sophia Antipolis.

A special thank goes to Dr. Renato L.G. Cavalcante, who illuminated me on some dark sides of math, engineering, and scientific writing. His supervision activity has been outstanding.

I also want to thank Prof. Slawomir Stańczak and Prof. Giuseppe Caire for the useful discussions and for the great opportunity that they offered me with this project at HHI.

Going a bit more back in time, my gratitude extends also to Prof. Roberto Garello and Prof. Monica Visintin, whom I thank for having encouraged me to follow my ambition for a career in research, and of course for the supervision work on this thesis.

Last but not least, I thank the colorful community of Berliners for having welcomed me in the best of the possible ways, and of course my family for they precious and fundamental support.

Bibliography

[1] D. Tse and P. Viswanath, *Fundamentals of wireless communication*. Cambridge university press, 2005.

[2] F. Rusek, D. Persson, B. K. Lau, E. G. Larsson, T. L. Marzetta, O. Edfors, and F. Tufvesson, "Scaling up MIMO: Opportunities and challenges with very large arrays," *IEEE Signal Processing Magazine*, vol. 30, no. 1, pp. 40–60, 2013.

[3] H. Huh, A. M. Tulino, and G. Caire, "Network mimo with linear zero-forcing beamforming: Large system analysis, impact of channel estimation, and reduced-complexity scheduling," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 2911–2934, 2012.

[4] H. Yin, D. Gesbert, M. Filippou, and Y. Liu, "A coordinated approach to channel estimation in large-scale multiple-antenna systems," *IEEE Journal on Selected Areas in Communications*, vol. 31, no. 2, pp. 264–273, 2013.

[5] Y. Ding and B. D. Rao, "Channel estimation using joint dictionary learning in FDD massive mimo systems," in *IEEE Global Conference on Signal and Information Processing (GlobalSIP)*, 2015, pp. 185–189.

[6] J. H. Kotecha and A. M. Sayeed, "Transmit signal design for optimal estimation of correlated MIMO channels," *IEEE Transactions on Signal Processing*, vol. 52, no. 2, pp. 546–557, 2004.

[7] A. Adhikary, J. Nam, J.-Y. Ahn, and G. Caire, "Joint spatial division and multiplexing–The large-scale array regime," *IEEE transactions on information theory*, vol. 59, no. 10, pp. 6441–6463, 2013.

[8] H. Yin, D. Gesbert, and L. Cottatellucci, "Dealing with interference in distributed large-scale MIMO systems: A statistical approach," *IEEE Journal of Selected Topics in Signal Processing*, vol. 8, no. 5, pp. 942–953, 2014.

[9] A. Molish, *Wireless Communications*, 2nd ed. Wiley, 2010.

[10] S. Haghighatshoar and G. Caire, "Massive MIMO channel subspace estimation from low-dimensional projections," *IEEE Transactions on Signal Processing*, vol. 65, no. 2, pp. 303–318, 2017.

[11] K. Hugl, K. Kalliola, and J. Laurila, "Spatial reciprocity of uplink and downlink radio channels in FDD systems," *Proc. COST 273 Technical Document TD (02)*, vol. 66, p. 7, 2002.

[12] H. Asplund, A. Glazunov, A. Molisch, K. Pedersen, and M. Steinbauer, "The cost 259 directional channel model-part ii: Macrocells," *IEEE Transactions on Wireless Communications*, vol. 5, no. 12, 2006.

[13] 3GPP, "Spatial channel model for multiple input multiple output MIMO simulations (release 14)," 3rd Generation Partnership Project (3GPP), Tech. Rep. TR 25.996 V14.0.0, 2017.

[14] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE transactions on antennas and propagation*, vol. 34, no. 3, pp. 276–280, 1986.

[15] 3GPP, "Study on 3D channel model for LTE (release 12)," 3rd Generation Partnership Project (3GPP), Tech. Rep. TR 36.873 V12.6.0, 2017.

[16] D. J. Love and R. W. Heath, "Limited feedback unitary precoding for spatial multiplexing systems," *IEEE Transactions on Information Theory*, vol. 51, no. 8, pp. 2967–2976, 2005.

[17] P. de Kerret, X. Yi, and D. Gesbert, "On the degrees of freedom of the k-user time correlated broadcast channel with delayed CSIT," in *2013 IEEE International Symposium on Information Theory*, 2013, pp. 624–628.

[18] M. A. Maddah-Ali and D. Tse, "Completely stale transmitter channel state information is still very useful," *IEEE Transactions on Information Theory*, vol. 58, no. 7, pp. 4418–4431, 2012.

[19] J. Fang, X. Li, H. Li, and F. Gao, "Low-rank covariance-assisted downlink training and channel estimation for FDD massive MIMO systems," *IEEE Transactions on Wireless Communications*, vol. 16, no. 3, pp. 1935–1947, 2017.

[20] H. Rauhut, K. Schnass, and P. Vandergheynst, "Compressed sensing and redundant dictionaries," *IEEE Transactions on Information Theory*, vol. 54, no. 5, pp. 2210–2219, 2008.

[21] L. Zheng and D. Tse, "Communication on the Grassmann manifold: A geometric approach to the noncoherent multiple-antenna channel," *IEEE Transactions on Information Theory*, vol. 48, no. 2, pp. 359–383, 2002.

[22] J. Nam, "Fundamental limits in correlated fading MIMO broadcast channels: Benefits of transmit correlation diversity," in *Information Theory (ISIT), 2014 IEEE International Symposium on*. IEEE, 2014, pp. 2889–2893.

[23] T. Yoo and A. Goldsmith, "Capacity and power allocation for fading MIMO channels with channel estimation error," *IEEE Transactions on Information Theory*, vol. 52, no. 5, pp. 2203–2214, 2006.

[24] M. R. Akdeniz, Y. Liu, M. K. Samimi, S. Sun, S. Rangan, T. S. Rappaport, and E. Erkip, "Millimeter wave channel modeling and cellular capacity evaluation," *IEEE journal on selected areas in communications*, vol. 32, no. 6, pp. 1164–1179, 2014.

[25] M. Jordan, A. Dimofte, X. Gong, and G. Ascheid, "Conversion from uplink to downlink spatio-temporal correlation with cubic splines," in *IEEE 69th Vehicular Technology Conference*, 2009, pp. 1–5.

[26] A. Decurninge, M. Guillaud, and D. Slock, "Channel covariance estimation in massive MIMO frequency division duplex systems," in *IEEE Globecom*, 2015, pp. 1–6.

[27] ——, "Riemannian coding for covariance interpolation in massive MIMO frequency division duplex systems," in *IEEE Sensor Array and Multichannel Signal Processing Workshop*, 2016, pp. 1–5.

[28] Y.-C. Liang and F. P. S. Chin, "Downlink channel covariance matrix (DCCM) estimation and its applications in wireless DS-CDMA systems," *IEEE Journal on Selected Areas in Communications*, vol. 19, no. 2, pp. 222–232, 2001.

[29] L. Miretti, R. Cavalcante, and S. Stanczak, "FDD Massive MIMO channel spatial covariance conversion using projection methods," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, April 2018.

[30] ——, ""downlink channel spatial covariance estimation in realistic FDD massive MIMO systems," *submitted to IEEE SPAWC 2018*, 2018.

[31] P. L. Combettes, "The foundations of set theoretic estimation," *Proceedings of the IEEE*, vol. 81, no. 2, pp. 182–208, 1993.

[32] H. Stark, Y. Yang, and Y. Yang, *Vector space projections: a numerical approach to signal and image processing, neural nets, and optics.* John Wiley & Sons, Inc., 1998.

[33] I. Yamada and N. Ogura, "Adaptive projected subgradient method for asymptotic minimization of sequence of nonnegative convex functions," *Numerical functional analysis and optimization*, vol. 25, pp. 593–617, 2005.

[34] S. Theodoridis, K. Slavakis, and I. Yamada, "Adaptive learning in a world of projections," *IEEE Signal Processing Magazine*, vol. 28, no. 1, pp. 97–123, 2011.

[35] D. G. Luenberger, *Optimization by Vector Space Methods.* Wiley, jan 1998.

[36] H. Bauschke, P. Combettes, and S. Kruk, "Extrapolation algorithm for affine-convex feasibility problems," *Numerical Algorithms*, vol. 41, no. 3, pp. 239–274, 2006.

[37] D. Neumann, M. Joham, L. Weiland, and W. Utschick, "Low-complexity computation of LMMSE channel estimates in massive MIMO," in *Proceedings of the 19th International ITG Workshop on Smart Antennas (WSA 2015).* VDE, 2015, pp. 1–6.

[38] K. Grigoriadis, A. Frazho, and R. Skelton, "Application of alternating convex projection methods for computation of positive toeplitz matrices," *IEEE transactions on signal processing*, vol. 42, no. 7, pp. 1873–1875, 1994.

[39] H. H. Bauschke and P. L. Combettes, *Convex analysis and monotone operator theory in Hilbert spaces*. Springer, 2011, vol. 408.

[40] S. T. Smith, "Covariance, subspace, and intrinsic Cramér-Rao bounds," *IEEE Transactions on Signal Processing*, vol. 53, no. 5, pp. 1610–1630, 2005.