Blind Identification of MIMO Channels Based on 2nd Order Statistics and Colored Inputs

The Riemannian Geometry of Algorithms, Performance Analysis and Bounds over the Quotient Manifold of Identifiable Channel Classes

João Xavier

Abstract

We study blind identification of multiple-input multiple-output (MIMO) systems based only on 2nd order statistics (SOS). This problem arises naturally in many applications, for example, SDMA (Space Division Multiple Access) networks for wireless multi-user communications. The problem of SOS-based blind MIMO channel identification is not strictly well-posed. At least, a phase ambiguity per input cannot be avoided. But other, more severe, channel ambiguities may also exist. We take the viewpoint of modeling the unavoidable phase ambiguities as an equivalence relation in the set of MIMO channels. We partition the set of MIMO channels in disjoint equivalence classes and work solely with this quotient space (the set of equivalence classes) throughout the thesis. We prove an identifiability theorem which shows that, under a certain spectral diversity condition on the input random signals, the MIMO channel equivalence class is uniquely determined by the output SOS. Although the proof of the identifiability theorem is not constructive, we develop a closed-form algorithm which achieves the predicted identifiability bound.

We show that the sufficient input spectral diversity condition can be easily induced by placing a coloring pre-filter at each transmitter. To achieve an optimal design for the pre-filters, we carry out an asymptotic performance analysis of our closed-form algorithm. Since we deal with an inference problem over a quotient space, our case is not covered by the standard theory used in Euclidean contexts. Instead, a Riemannian structure is induced in the quotient space and we set up some intrinsic theoretical tools to cope with this manifold setting. To place a fundamental limit on the MIMO channel equivalence estimate quality, we also present an extension of the Cramér-Rao bound to this Riemannian setup.

Keywords: Blind channel identification, Multiple-input multiple-output systems, Second-order statistics, Colored inputs, Performance Analysis, Riemannian geometry

Resumo

Esta tese aborda o problema da identificação cega de canais MIMO (multiple-input multiple-output) baseada apenas em estatísticas de 2ª ordem. Trata-se de um problema que encontra aplicação imediata em diversos cenários, como por exemplo, redes de acesso múltiplo por divisão espacial SDMA (Space Division Multiple Access) para comunicações móveis. O problema da identificação cega de canais MIMO a partir das estatísticas de 2ª ordem não está, de um ponto de vista puramente matemático, bem formulado. De facto, não pode ser resolvida pelo menos uma ambiguidade de fase por cada entrada do canal. Contudo, podem existir outras ambiguidades de carácter mais severo. Nesta tese, resolveu-se modelar as inevitáveis ambiguidades de fase como uma relação de equivalência no conjunto dos canais MIMO. Assim, particiona-se este conjunto em classes de equivalência disjuntas e analisam-se os diversos problemas a tratar nesta tese a partir deste espaço quociente. Para resolver o problema da identificação, prova-se um teorema que, sob determinada condição espectral nos processos estocásticos presentes à entrada, garante que a classe de equivalência do canal é univocamente determinada pelas estatísticas de 2ª ordem observadas à saída. Embora a prova deste teorema não seja construtiva, desenvolve-se um algoritmo que, num número finito de passos, realiza a identificação do canal.

Mostra-se que a condição espectral suficiente acima mencionada pode ser facilmente induzida colocando um filtro que correlaciona os símbolos de informação emitidos, em cada transmissor. De modo a conseguir um desenho óptimo para estes filtros, realiza-se um estudo de desempenho asimptótico do algoritmo de identificação proposto. Contudo, porque nos confrontamos com um problema de estimação formulado sobre um espaço quociente, o nosso caso não se encontra coberto pelas técnicas usuais de análise estatística que operam em espaços Euclideanos. Assim, induz-se uma estrutura Riemanniana sobre o espaço quociente e desenvolvem-se algumas ferramentas teóricas que possibilitam uma análise intrínsica neste contexto. Apresenta-se ainda uma extensão do conhecido limiar de Cramér-Rao para este contexto de variedades diferenciais Riemannianas.

Palavras-chave: Identificação cega de canal, Sistemas com múltiplas entradas e saídas, Estatísticas de segunda ordem, Entradas com espectro não-branco, Análise de desempenho, Geometria Riemanniana

Acknowledgements

My first words of gratitude go to my scientific advisor, Prof. Victor Barroso. Right from the start of this thesis, I felt his total and enthusiastic support for probing and researching outside the mainstream of statistical signal processing theory (namely, for attacking some standard problems from a differential-geometric viewpoint). Since, at that time, it was not at all clear whether this would be or not a valuable approach, I am sincerely grateful that he was willing to take the risk in pursuing this vision. At a more pragmatic level, I am in debt to him for countless stimulating and inspiring technical discussions held throughout these years. Only his technical advices permitted to reach the final form of this work. Finally, I know Prof. Victor Barroso from my last year as an undergraduate student. When this PhD project started, I conjectured to myself that, in addition to our scientific work, our friendship would also experience a major boost. It is, without any doubt, the happiest conclusion of this work that this initial conjecture has evolved into what is now a well-established theorem (which has a pleasant constructive proof).

At the beginning of this thesis, I was given the opportunity to visit Prof. José Moura at the Carnegie Mellon University. In several aspects, that event marked a turning point in my life. I have found that, each time I interact with him, new dimensions are added to my low-dimensional scientific world. Of course, I do not claim originality in this result: every person that I know that intersected his/her lifeline with his has reported the same phenomenon (and I know quite a few!). I am thus really deeply grateful for all the extremely lucid, vivid and fun (sometimes also highly provocative! – my favorites ones) scientific discussions, bright insights and many profound ideas that Prof. Moura shared with me.

Apart the brief appointments at the CMU, my physical coordinates throughout this thesis coincided with those of the Institute for Systems and Robotics (ISR). In my opinion, Prof. João Sentieiro (head of ISR) together with all other members of ISR, created a quite unique and perfect habitat for science development. This is not only an opinion: their remarkable work translates every year in high-quality (award winning) scientific theses and papers authored by ISR students and professors. It is simply a pleasure to feel part of this community. In particular, I want to single out my office mate, João Pedro Gomes, and thank him for the true friendship he provided throughout the years and for his never-ending patience with respect to: engaging in any technical discussion from which I always profit greatly at any time (usually, interrupting his own work), aiding me with many software/hardware problems (usually, interrupting his own work), and listening to my Monday comments about the week-end performance of the Sporting football team (usually, interrupting his own work). Many thanks are also due to Sébastien Bausson who read very carefully several versions of this manuscript (when it was only a mere mass of disconnected results). He acted as a genuine extra-reviewer and improved a lot the presentation of the material. I am also grateful to Etienne Grossmann for many clarifying discussions about math and Linux, and to Paulo Mónica and Jorge Barbosa for their genuine friendship.

I want also to acknowledge the support of Instituto Superior Técnico, namely, the Department of Electrical and Computer Engineering, for allowing me to pursue my PhD research in full time over three years. Moreover, I wish to thank the financial support provided by the Fundação para a Ciência e Tecnologia (FCT), grant PRAXIS XXI BD/9391/96, at the beginning of this thesis.

In the nontechnical world (the most important), I want to single out my friends Mr. Ricardo and his wife Mrs. Lena for their support over these years. I am also in debt to my uncles Nuno, Jorge and Fátima, together with my cousins Rosa, Cláudia and Bruno, for their constant encouragement. The same applies to my brother Nuno and, even more specially, to my sister Kátia whose friendship, support, and presence I always felt from day one. My last words go to my parents António e Maria Emília to whom I fully dedicate this thesis. The love and gratitude that I feel for them starts at the bottom of my heart and has no end.

Contents

List of Figures

Notation

In the following, we list the symbols and acronyms frequently used throughout this thesis. In each row, the first and second entries introduce the symbol or acronym and define its meaning, respectively. The third entry contains the number of the page where it appears for the first time.

List of Symbols

$\boldsymbol{A}^T$	transpose of $\boldsymbol{A}$	11
$\overline{\boldsymbol{A}}$	complex conjugate of $\boldsymbol{A}$	11
$\boldsymbol{A}^H$	Hermitean (conjugate transpose) of $\boldsymbol{A}$	11
$\boldsymbol{A}^+$	Moore-Penrose pseudo-inverse of $\boldsymbol{A}$	11
$\boldsymbol{A}^{-1}$	inverse of $\boldsymbol{A}$	11
$\boldsymbol{A}^{[n]}$	n-fold Kronecker product of $\boldsymbol{A}$	87
$\boldsymbol{A}_{n,m}$	diagonal concatenation of $\boldsymbol{I}_{nm}$ and $-\boldsymbol{I}_{nm}$	60
$\boldsymbol{B}_{n,m}$	diagonal concatenation of $\boldsymbol{K}_{n,m}$ and $\boldsymbol{K}_{n,m}$	60
$\boldsymbol{C}_{n,m}$	diagonal concatenation of $\boldsymbol{K}_{n,m}$ and $-\boldsymbol{K}_{n,m}$	60
$\mathbb{C}$	set of complex numbers	10
$\mathbb{C}^n$	set of n-dimensional column vectors with complex entries	11
$\mathbb{C}^{n\times m}$	set of $n \times m$ matrices with complex entries	11
$\mathbb{C}_d[z]$	set of polynomials in the indeterminate z^{-1} with degree d	12
$\operatorname{diag}(\boldsymbol{A}_1, \cdots, \boldsymbol{A}_n)$	diagonal concatenation of matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_n$	11
$\stackrel{d}{=}$	equality in distribution	59
$\xrightarrow{d}$	convergence in distribution	59
$\|\boldsymbol{A}\|$	Frobenius norm of $\boldsymbol{A}$	11
$\deg f(z)$	degree of the polynomial $f(z)$	12
$\operatorname{dom} F$	domain of the map F	10
$\mathbb{H}(n)$	set of $n \times n$ Hermitean matrices	12
$\mathbb{H}[z]$	set of $Q \times P$ polynomial matrices satisfying assumption A1)	20
$\mathbb{H}[z]/\sim$	set of equivalence classes of $\mathbb{H}[z]$ under the relation $\sim$	22
$\operatorname{image} F$	image of the map F	10
$\boldsymbol{i}_n$	n^2-dimensional column vector given by $\boldsymbol{i}_n = \operatorname{vec}(\boldsymbol{I}_n)$	86
$\boldsymbol{I}_n$	$n \times n$ identity matrix	11
$\boldsymbol{J}_n$	$n \times n$ Jordan block	11
$\mathbb{K}(n)$	set of $n \times n$ skew-Hermitean matrices	12
$\ker(\boldsymbol{A})$	kernel of $\boldsymbol{A}$	11
$\boldsymbol{K}_{n,m}$	commutation matrix of size $nm \times nm$	60
$\boldsymbol{K}_n$	commutation matrix of size $n^2 \times n^2$ ($\boldsymbol{K}_n = \boldsymbol{K}_{n,n}$)	60
$\boldsymbol{K}_n[m]$	mth shift matrix of size $n \times n$	11
$\delta[n]$	1D delta signal	13
$\delta[n,m]$	2D delta signal	13
$\boldsymbol{A} \boxtimes \boldsymbol{B}$	Khatri-Rao product of $\boldsymbol{A}$ and $\boldsymbol{B}$	150
$\boldsymbol{A} \otimes \boldsymbol{B}$	Kronecker product of $\boldsymbol{A}$ and $\boldsymbol{B}$	12
$\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$	Normal distribution with mean $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Sigma}$	59
$\mathbb{N}$	set of natural numbers $\{0, 1, 2, \ldots\}$	10
$\mathbf{1}_n$	n-dimensional column vector with all entries equal to 1	11

$\operatorname{ord} \boldsymbol{F}(z)$	order of the polynomial matrix $\boldsymbol{F}(z)$	12
$\xrightarrow{P}$	convergence in probability	59
$\mathbb{R}$	set of real numbers	10
$\mathbb{R}^n$	set of n-dimensional column vectors with real entries	11
$\mathbb{R}^{n\times m}$	set of $n \times m$ matrices with real entries	11
$\operatorname{rank}(\boldsymbol{A})$	rank of $\boldsymbol{A}$	11
$\circledast$	polynomial filter operator (convolution)	13
$\operatorname{tr}(\boldsymbol{A})$	trace of $\boldsymbol{A}$	11
$\mathbb{U}(n)$	set of $n \times n$ unitary matrices	12
$\operatorname{vec}(\boldsymbol{A})$	columns of $\boldsymbol{A}$ stacked from left to right	11
$\mathbb{Z}$	set of integer numbers $\{0, \pm 1, \pm 2, \ldots\}$	10
$\mathbf{0}_{n\times m}$	$n \times m$ zero matrix	11

Acronyms

ACMA	Analytical Constant Modulus Algorithm	5
BCIP	Blind Channel Identification Problem	4
BER	Bit Error Rate	45
BIDS	Blind Identification via Decorrelating Subchannels	5
CDMA	Code Division Multiple Access	1
CFIA	Closed-Form Identification Algorithm	16
CM	Constant Modulus	3
CRLB	Cramér-Rao Lower Bound	9
DOA	Direction of Arrival	3
EVD	Eigenvalue Decomposition	30
FA	Finite Alphabet	3
FIR	Finite-Impulse Response	6
FDMA	Frequency Division Multiple Access	1
HOS	Higher-Order Statistics	3
IIA	Iterative Identification Algorithm	35
IIR	Infinite-Impulse Response	5
ISI	Intersymbol Interference	6
IVLB	Intrinsic Variance Lower Bound	9
MIMO	Multiple-Input Multiple-Output	4
MSE	Mean-Square Error	38
QAM	Quadrature Amplitude Modulation	38
RF	Radio Frequency	1
SDMA	Space Division Multiple Access	2
SIMO	Single-Input Multiple-Output	4
SNR	Signal-to-Noise Ratio	39
SOS	Second-Order Statistics	3
SVD	Singular Value Decomposition	33
TDMA	Time Division Multiple Access	1
TICC	Transmitter Induced Conjugate Cyclostationarity	5
WSS	Wide-Sense Stationary	20

Chapter 1

Introduction

1.1 Problem formulation

During the past few years, we have witnessed a spectacular growth in the demand for mobile communication services, such as telephony, data, facsimile, etc. Indeed, whilst low data-rate digital voice was the main application of second-generation (2G) wireless systems (European GSM, North American IS-95, etc) with channels operating around 14.4kbps (kilobits per second), the new third-generation (3G) systems (Universal Mobile Telecommunications System (UMTS), International Mobile Telecommunications (IMT)-2000, etc) migrated towards high-rate data applications and aim to support Internet and multimedia applications of up to 144kbps in high mobility (vehicular) traffic and up to 2Mbps (megabits per second) in indoor (stationary) traffic [19]. To cope with these rates of demand, an increase of the current cellular networks' capacity is mandatory. Since the radio frequency (RF) spectrum is a scarce resource, this translates into the demand for new, highly efficient spectral bandwidth-saving multiple access techniques together with advanced signal processing methodologies which end up enabling the operators to reliably serve multiple high-rate users in parallel. Multiple-access schemes incorporated in some 1G and 2G standards include: code division multiple access (CDMA), time-division multiple access (TDMA) and frequency division multiple access (FDMA). In CDMA architectures, several users occupy simultaneously in time a spectral bandwidth that is substantially greater than their respective information-bearing signals [70]. Their baseband data signals are spread by means of a pseudo-random code (one per user) before transmission. Stretching the signal to the whole available bandwidth, provides protection against interference, noise, and permits to mitigate the negative effects of multipath propagation. CDMA reception consists in exploiting the near-orthogonality of the spreading codes to uniquely despread each intended user from the incoming wideband signal which contains the mixed co-channel transmissions. The CDMA concept is embodied in the IS-95 standard for cellular phone applications and is also envisaged for 3G systems. In TDMA architectures, each user has access to the whole available bandwidth, but only one communicator can be active at a time [46]. That is, the time axis is partitioned into disjoint intervals and each of these non-overlapping time slots is allocated to only one user. Since the users' transmissions do not intersect in time, there is no inherent crosstalk, which, apart some synchronization issues, simplifies TDMA reception. The TDMA concept is included in the GSM digital cellular system. In FDMA architectures, the available RF

bandwidth is split into disjoint frequency narrowband subchannels. Each user enjoys exclusive access to one frequency subchannel and all users transmit simultaneously in time. Due to the intrinsic separation of signals in the frequency domain, the FDMA receiver essentially picks up each user from the observed signal through a bandpass filter. The FDMA scheme was prominently employed in the first-generation (1G) cellular systems.

Recently, the space division multiple access (SDMA) concept has emerged as an attractive multiple-access technique which has the potential for supporting high-speed data transmission while maintaining an efficient signal spectral occupation [3, 45, 20]. In loose terms, SDMA networks take advantage of the spatial dimension of the radio resource for acommodating multiple users in parallel. In SDMA architectures, several users within the same geographical cell do coexist in the same time and/or frequency channel which may be either narrowband or wideband. SDMA reception relies on an multielement antenna deployed at the base station receiver together with sophisticated array signal processing techniques to discriminate the co-channel users based on their distinct spatial signatures. Combining the SDMA philosophy with other multiple-access techniques results in a more efficient spectral packing of users, which ends up boosting the overall system capacity without requiring further Hertzs. Figure 1.1 depicts the uplink (users to base station) of a standard SDMA network (all users are active at the same time). In addition to the net increase in cellular capacity, SDMA systems can provide better immunity to the multipath-induced fading phenomenon. Indeed, if the antenna elements are properly spaced so that signal decorrelation can be assumed, the SDMA receiver has access to several independent versions of each transmitted signal (spatial oversampling). The probability that a deep fade occurs simultaneously across all antenna elements becomes negligible (as the number of antennas grows), thus ensuring an adequate average power level for signal copy.

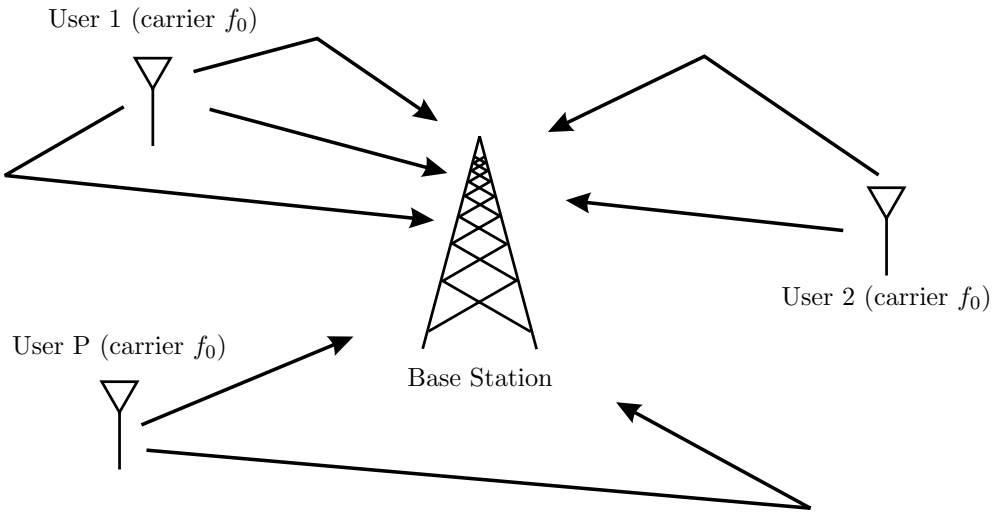

Figure 1.1: Space Division Multiple Acess (SDMA) wireless network: uplink

These performance gains are obtained at the cost of an increased complexity in the base station receiver, which must now demodulate several co-channel users' transmissions (without the help of orthogonal spreading codes). The first separation techniques in the spatial domain relied on the concept of spatial filtering or beamforming [69]. Essentially, this

consists in numerically weighting the outputs of the antenna array in order to create beam patterns (spatial filters) with the ability of illuminating desired users and nulling out the unintended ones. The implementation of beamformers at the receiver requires previous knowledge of the sources' geographical position, that is, the direction of arrival (DOA) of the wavefronts impinging on the array. Although, in principle, the receiver can learn these spatial channel characteristics through training sessions (supervised reception), this results in a waste of the available bandwidth. Thus, blind (unsupervised reception) methods is the preferred mode of operation. Furthermore, blind techniques can support automatic link re-establishment whenever data links are occasionally disruptured due to severe fading. A plethora of blind DOA finding methods have been proposed in the literature, see [52, 50, 21, 35] and the references therein. The DOA estimation techniques can resolve closely spaced sources but require highly accurate parametric channel models, calibrated arrays, and/or special array configurations. Their performance can drop significantly in the face of callibation errors, sensor position mismatches, etc. This motivated the search for new blind spatial separation techniques, free from the DOA/beamforming paradigm. Thus, to gain robustness, current approaches do not rely on known array responses, but instead, tend to interpret SDMA demodulation as an instance of a blind source separation (BSS) problem. The area of BSS is a very active, interesting research subject on its own which finds direct applications in many fields such as speech, digital communications, radar, etc. See [26] for a survey of theory, techniques and recent research findings. In the context of SDMA networks, the BSS techniques exploit special properties in the data signal model "outside" the realm of the array response matrix. Note that the structure of the channel matrix is basically controlled by the physical environment (position of scatters, etc) in which the data link is embedded. In contrast, the transmitted signals are man-made. Thus, any desired structure can be easily inserted in the transmitter side and can safely be assumed to be there (to be exploited) at the receiver end. In the context of wireless communications, current BSS approaches take advantage of several digital signal structures: cyclostationarity, constant-modulus (CM), finite-alphabet (FA), etc. In fact, a wide range of methodologies is now available, see [17, 18]. When BSS techniques exploit the information contained in the statistics of the observations, they can be classified as either higher-order statistics (HOS) methods or second-order statistics (SOS) methods. The SOS-based techniques are potentially more attractive since they tend to require less data samples to achieve channel identification when compared to cumulant-based (HOS) techniques. Furthermore, the majority of BSS techniques developed so far are iterative in nature and vulnerable to local minima convergence.

In this thesis, we address the problem of blindly separating the contribution of distinct digital sources, given a finite set of observation vectors from the noisy linear mixture. Although the primary motivation comes from the SDMA context, our results are presented in full generality and can be used in other relevant digital communication setups. The striking features of our solution are: it uses only 2-nd order statistics (SOS) and provides a closed-form (non-iterative) estimate of the mixing channel. This is obtained by conveniently structuring the transmitted signals (more precisely, by judiciously coloring their 2nd order statistics) in order to enable the receiver to recover the channel matrix from the correlation matrices of the observation vector. In addition to the proposed closed-form

channel identification algorithm, we study in this thesis some adjacent problems linked to this multiple users setup. A more detailed account of the topics covered in this work and the main contributions to be found herein is given in the sequel.

1.2 The thesis chapter by chapter: brief summary and contributions

In this section, we briefly review the contents of chapters 2, 3 and 4, which are the main chapters in this thesis. The goal is to highlight the more important contributions per chapter. A more exhaustive description (summary) of each one can be found in their respective introductory section. For each chapter, we also refer the publications (conference or journal papers) that it gave rise.

1.2.1 Chapter 2

Summary. In chapter 2, we concern ourselves with the blind channel identification problem (BCIP) mentioned earlier. More specifically, we aim at blindly identifying the convolutive (that is, with finite memory) mixing channel which underlies the available channel output observations. From this knowledge, a space-time linear equalizer retrieving all the the input users is easily synthesized at the receiver. Our solution identifies the channel analytically (non-iterative scheme) and is based only on the 2nd order statistics of the observations. To put in perspective our results, we now review the more relevant blind channel identification/equalization techniques but with a clear and strong emphasis on SOS-based only methods. We begin with the single-user context and then proceed to the multiple-users setup.

For single-input multiple-output (SIMO) systems, the work by Tong *et al.* [59, 60, 61] was a major achievement. By exploiting only the 2nd order statistics of the channel outputs, they derive an analytical solution for the unknown channel coefficients. Being a non-iterative identification scheme, it is not impaired by local (false) minima convergence issues which plague many iterative techniques (and oblige them to several time-consuming reinitializations). In their work, the input signal driving the SIMO channel is assumed to be white (uncorrelated in time), and this structure was shown to be sufficient for retrieving the channel by the method of moments. The subspace method developed in [43] exploits instead the Sylvester structure exhibited by the channel matrix in a stacked data model and yields another SOS-based closed-form channel identification technique. The subspace method can also be viewed as a deterministic method (no statistical description is assumed for the source signal), like the techniques in [80, 29]. See [62] for a survey on both these statistical and deterministic approaches.

The subspace-based methodology was generalized to the context of multiple-input multiple-output (MIMO) systems in [24, 1, 2]. It was shown that, by exploiting only the 2nd order statistics of the channel outputs, the MIMO channel can be recovered up to a block diagonal constant matrix. In equivalent terms, the convolutive mixture can be decoupled in several instantaneous (also called static or memoryless) mixtures of the input signals, with two sources being in the same group if and only if they share the same channel degree. In particular, the convolutive mixture is completely resolved if all

users are exposed to distinct system orders, that is, they exhibit memory diversity. The blind SOS-based whitening approach in [64, 65, 66] also converts a convolutive mixture into a static one, with a substantial weakening on the channel assumptions: infinite-impulse response (IIR) channels can be accommodated, as well as minimum-phase common zeros among the subchannels. Furthermore, the usual column-reduced condition can be dropped. Still, it is left with these approaches to resolve the residual static mixtures. To handle the latter, several BSS techniques can be invoked depending mainly on the characteristics of the sources, but also on the number of available samples, and the signal-to-noise ratio. Examples of these BSS techniques include: **i)** high-order statistics (HOS) approaches, e.g., the joint-diagonalization procedure in [9], which are feasible for non-Gaussian sources (although estimates of cumulants converge slower than SOS, [61]); **ii)** the analytical constant modulus algorithm (ACMA) [67], which separates constant modulus (CM) sources; **iii)** separation of finite-alphabet (FA) sources, which may be tackled by locally-convergent iterative algorithms [56, 57, 58, 25, 4, 73, 68].

Complete blind channel identification (that is, without requiring any BSS step) based only on 2nd order statistics can by achieved by the techniques in [5, 38, 30, 31, 12, 11]. The methods in [5, 38, 30, 31] rely on colored inputs while [12, 11] rely on conjugate cyclostationary inputs. The main drawback of the separation technique in [5] is that it only applies to static (memoryless) mixtures. The matrix-pencil (MP) method introduced in [38] can handle convolutive (with memory) mixtures and is formulated within the non-stationary scenario. The MP technique processes a pair of output correlation matrices in order to determine generalized eigenvectors. Each one of these eigenvectors can isolate a filtered version of one of the inputs from the given observations. Thus, by carefully grouping the extracted signals (that is, identifying which extracted signals correspond to the same source) the original multiple-user separation problem is reverted to several SIMO deconvolution subproblems which can then be solved by the aforementioned mono-user techniques. However, in the MP approach it is not clear which pair of output correlation matrices should be selected to support this identification scheme. Moreover, by processing only one pair of correlation matrices, the MP ignores important information conveyed by the remaining ones. The techniques in [30, 31] exploit the concept of blind identification by decorrelating subchannels (BIDS) introduced in [17, chapter 4]. The main advantage over the MP technique is that the BIDS algorithm does not require the channel to be simultaneously irreducible and column-reduced (as it is common in most multiple-user scenarios [1, 2, 24, 68]). However, the implementation of the BIDS technique is not in closed-form and global convergence is not proved for the iterative algorithm in [17, chapter 4]. At this point, we would like to stress that, common to all the works in [5, 38, 30, 31] is the fact that they do not assume the color of the input statistics to be known at the receiver. The closed-form transmitter induced conjugate cyclostationarity (TICC) technique introduced in [12, 11] does not require colored inputs. Rather, a distinct conjugate cyclic frequency per user is induced at each transmitter. This inserted structure in the emitted signals is then exploited at the receiver to reduce the problem to several SIMO channel estimation problems. The main drawback of TICC is its high sensitivity to carrier frequency misadjustments. In multi-user scenarios, this distortion may be expected to appear in the baseband demodulated signals, as the receiver has to synchronize its local

oscillator simultaneously with multiple independently generated carrier frequencies.

Contribution. In this chapter, we introduce a closed-form method for the blind identification of MIMO channels, based only on 2nd order statistics (SOS). However, in contrast with all aforementioned SOS-based techniques, we assume that the color of the 2nd order statistics of the channel inputs are known at the receiver (in addition to the common assumption that they are distinct). This is because, in our communication strategy, the emitted signals are colored at each transmitter. More precisely, we use correlative filters at the transmitters to assign distinct spectral patterns to the random messages emitted by the sources. Thus, in particular, their spectral colors are under our control. It is this extra piece of information available at the receiver (when compared with the remaining techniques) which simplifies and enables an analytical solution for the blind channel identification problem. We establish sufficient conditions on the correlative filters to ensure uniqueness of the MIMO channel matrix from the SOS of the channel outputs by proving an identifiability theorem. We exploit this theoretical result to derive an algorithm that blindly identifies the channel matrix by matching the theoretical and observed 2nd order statistics.

As in the MP approach [38] we handle convolutive (with memory) channels, but we take advantage of all correlation matrices of the channel outputs in an integrated identification scheme (not just a pair of them as in the MP approach). Compared with the BIDS [17, chapter 4] (see also [30, 31]) and the TICC [12, 11] methodologies, our channel assumptions are more restrictive because we assume the channel to be irreducible and column-reduced (as in the majority of multiple-user approaches [1, 2, 24, 68, 38]). However, since our solution can be implemented in closed-form, we do not have the local convergence issues of the BIDS algorithm [17, chapter 4]. When compared with the (non-iterative) TICC technique, we gain robustness to baseband carrier phase drifts (as we show through computer simulations).

Relation with previous work. We introduced the correlative framework in [71]. In that paper, the closed form solution relies on a certain quasi-diagonal structure of the sources's correlation matrices. However, to attain this analytical solution, we have to restrict the correlative filters to those that satisfy a minimal memory length condition. In loose terms, the channel order of each correlative filter must exceed the degree of intersymbol interference (ISI) experienced by each user. This condition imposes a significant lower bound on the computational complexity of the Viterbi decoding algorithm, as the number of states in the trellis diagram increases with the order of the correlative filters. Here, we drop the quasi-diagonal property, which makes feasible correlative filters with arbitrary non-zero degree. Thus, the computational complexity of the Viterbi decoding step is significantly reduced. In fact, we prove that minimum phase finite impulse response (FIR) filters of non-zero degree fulfill the requirements of the identifiability theorem. This allows for the direct inversion of the filters, and leads to a simpler scheme to decode the original data sequences that may have phase-tracking capability. Since the sources's autocorrelation matrices do not have the quasi-diagonal structure, the method in [71] no longer applies. We develop here a new consistent closed form estimator for the MIMO channel.

Publications. Parts of the material contained in chapter 2 have been published in the

conference papers [72, 74] and, in a more extended version, in the journal paper [75]. Although the main theoretical and simulation results included in chapter 2 can be found in [75], the material is presented here with a different language. More precisely, we alert the reader right from the start that, in fact, the MIMO channel cannot be fully identified with SOS-only: a phase offset per user cannot be avoided. Of course, this is a well-known result. However, instead of sweeping this fact under the carpet, we choose to construct an appropriate mathematical setup which takes into account this impossibility right from the beginning. This is accomplished by working with a quotient space of channel equivalence classes (in which all MIMO channels that are equal up to a phase offset per column are identified) rather than the set of MIMO channels directly. Obviously, this mathematical setup does not contribute by itself to the solution of the blind channel identification problem at hand (which explains why it is ignored in most published identification schemes). In fact, if blind channel identification is the only issue to be addressed, then we can even agree that the introduction of this mathematical structure into the problem is rather pedantic. But, the question here is that blind channel identification is not the only issue to be addressed in this thesis. For example, in chapter 3, we have to deal with the asymptotic performance of our proposed closed-form channel identification algorithm. How can we assess (in terms of a metric) the quality of the channel estimate with respect to the "true" channel, in the presence of the phase channel ambiguities ? In fact, what is the "true" channel, since all MIMO channels in the same equivalence class are equally valid ? The usual approach taken in most performance studies consists in arbitrarily normalizing certain channel entries and then take the Euclidean distance between the channel estimate and the true one (after both are normalized). In contrast, equipped with the quotient space concept, we can avoid this awkward technique and work with a truly intrinsic distance, that is, a metric on the quotient space (thus, no channel "normalization" is needed with the metric being invariant to phase modifications in channel columns). In fact, a substantially stronger result applies, as the quotient space can be given quite naturally a (Riemannian) geometry from which, in particular, the intrinsic distance pops out. Apart the elegance of this approach, structuring the quotient space as a smooth (geometrical) object in its own right, is a key theoretical step (almost mandatory) which enables a rigorous intrinsic performance analysis (chapter 3) and the determination of intrinsic variance bounds of given estimators (chapter 4) which take values in that manifold. In sum, within the geometrical framework of the (Riemannian) quotient space, we can address many interesting theoretical problems in an elegant, intrinsic and unified manner. In fact, the majority of our theoretical results are obtained in great generality, and might be of interest for other parametric estimation scenarios with certain structured ambiguities.

Exploitation of differential-geometric tools within the context of statistical signal processing is not new. The precursor paper by R. Bucy, J. M. F. Moura and S. Leung [8] examines a highly relevant signal processing problem (multiple source localization by a phased sonar or radar array) from a differential-geometric perspective. They showed that the standard DOA estimation problem can be recast as the problem of estimating a point in a certain submanifold of the Grassmannian manifold. This viewpoint not only brings geometrical insight into the problem, but also potentiates the development of efficient algorithms for the position estimation of multiple closely spaced targets. Another geometrical

re-interpretation of a typical signal processing problem occurs in M. Rendas and J. M. F. Moura [49]. By exploiting the geometrical structure behind ML parameter estimation in exponential families, they set up an ambiguity function – an analysis tool which can answer several (global) performance questions. Its application to passive and active sonar and radar location mechanisms is exemplified in [49].

1.2.2 Chapter 3

Summary. In chapter 3, we carry out a performance analysis of the closed-form identification algorithm proposed in chapter 2. The goal of this theoretical study is to assess the impact of the correlative filters (placed at the transmitters) on the quality of the channel estimate. More precisely, given a choice for the correlative filters and the MIMO channel, we obtain a closed-form expression which approximates the mean-square distance between the true channel equivalence class and the estimated one, when N channel output observations are available for processing. From this study, we can address the issue of optimal (off-line) pre-filter design in communications scenarios where the MIMO channel is a random object with a given a stochastic model. For obtaining our theoretical results, we generalize certain key results in classical large-sample analysis (derived within the context of Euclidean spaces) to the setting of Riemannian manifolds (see below). At the end of the chapter, the theoretical study is validated through numerical simulations. The computer experiments have shown a good agreement between the predicted (theoretical) and observed accuracy of the channel estimate, thus qualifying this study as an effective tool for attacking the problem of optimal pre-filter design.

Contribution. The main contribution and novelty of this chapter is contained in the differential-geometric framework that we introduce in order to theoretically support the aforementioned performance analysis. Indeed, notice that our estimator takes values in a non-Euclidean space: the quotient space of identifiable channel equivalence classes. Thus, our case is not covered by the standard theory of asymptotic analysis which is mainly concerned with Euclidean spaces, e.g., see [37, 53]. In this chapter, we generalize some classical concepts and results (such as asymptotic normal random sequences, the delta-method, etc) to the Riemannian manifold setting. To attain this level of abstraction, we were inspired by some definitions introduced in the work by Chang [10] which addresses asymptotic analysis on the Lie group of special orthogonal matrices. Our results are obtained in all generality (for example, they are applicable to either Euclidean submanifolds or quotient spaces). This Riemannian extension of classical large-sample theory is useful in our context because the quotient space of channel equivalence classes can easily (in fact, quite naturally) acquire a Riemannian structure. This structure follows directly (is uniquely defined) by requiring the projection map which sends each MIMO channel to its equivalence class to be a Riemannian submersion. In loose terms, we make our choices in order to have a nice interface between the geometry of the set of MIMO channels (a canonical Euclidean space) and the geometry of the quotient space of channel equivalence classes. This last property, together with the large-sample machinery developed for Riemannian settings, simplifies the analysis and leads to elegant (computable) results. As a last commentary, we would like to notice that as soon as a Riemannian structure is

introduced on the quotient space, the concept of distance (also called intrinsic or Riemannian distance in the sequel) becomes automatically available. In fact, this is the distance mentioned in the expression "mean-square distance" appearing in the previous paragraph. The point here is that a correct notion of distance between equivalence classes emerges spontaneously from this setup. This removes the need for guessing what should be the appropriate error measure for evaluating the performance of the estimator, as it is sometimes required in other parametric estimation problems affected by parameter ambiguities, see the discussion in [42].

Publications. Some techniques, tools and the spirit of the approach pursued in the performance study in chapter 3 have been published in our conference papers [76, 77]. However, these works are not concerned with Riemannian manifolds. This is because, in [76, 77], all data are assumed to take values in the set of real numbers $\mathbb{R}$ (for simplicity of the analysis). Thus, with that assumption, the phase offset ambiguity which affects each column of the MIMO channel transfer matrix boils down to a ± 1 ambiguity. As a consequence, each channel equivalence class contains only a finite number of channels, not a continuum of channels like it happens when complex-valued data is allowed. This discrete structure of the quotient space makes the formalism of Riemannian spaces perfectly avoidable, and that is why it was not invoked in [76, 77]. All said, the whole Riemannian viewpoint (together with all its related results) taken in chapter 3 is new and has not been published.

1.2.3 Chapter 4

Summary. In the previous chapters 3 and 4, we proposed a SOS-based solution to the blind MIMO channel identification problem and studied its performance, respectively. In chapter 4, we interest ourselves with the performance of any estimator (SOS-based or not) for this inference problem. More precisely, we aim at finding a lower bound on the mean-square (Riemannian) distance which is valid for any estimator of the MIMO channel equivalence class, for a given signal-to-noise ratio and number of channel observations. Notice that we face a parametric estimation problem where the parameter takes values in the Riemannian manifold of channel equivalence classes. Therefore, the familiar bounding tools such as the Cramér-Rao lower bound (CRLB), which were conceived for statistical families indexed by parameters in open Euclidean subsets, do not apply. In this chapter, we develop an extension of the Cramér-Rao bound to the Riemannian manifold setting, which we call the intrinsic variance lower bound (IVLB). The IVLB is developed in all generality and may be used in scenarios other than the blind MIMO channel identification problem. In fact, chapter 4 contains examples involving statistical families indexed over the Lie group of orthogonal matrices and the complex projective space. The IVLB limits the accuracy of estimators taking values in Riemannian manifolds, within the framework of parametric statistical models indexed over Riemannian manifolds. The accuracy mentioned here is the intrinsic (Riemannian) accuracy which is measured with respect to the intrinsic Riemannian distance carried by the Riemannian manifolds. The IVLB depends on the curvature of the Riemannian manifold where the estimators take values and on a

coordinate-free extension of the familiar Fisher information matrix.

Contribution. The theory behind the classical CRLB inequality can be found in many standard textbooks [63, 51, 34, 48]. It assumes that one is given a parametric statistical family with the parameter taking values in open subsets of Euclidean spaces. However, recently, several authors have achieved CRLB extensions to the context of parametric statistical families indexed over submanifolds of Euclidean spaces, see [22, 41, 55]. This is the natural setting for inference problems where the parameter indexing the family is known to satisfy certain deterministic (perhaps physically imposed) smooth constraints. The works in [22, 41, 55] do not deal with abstract Riemannian manifolds such as quotient spaces. This drawback is eliminated in the studies in [27, 44, 54]. They work, right from the start, with statistical families indexed over (abstract) Riemannian manifolds and express their results with intrinsic-only tools (that is, the ambient Euclidean spaces are discarded from the analysis or are even non-existent). Note that this level of abstraction automatically covers more statistical families, like ones indexed over quotient spaces (the problem treated in this thesis). The work in [27] extends the CRLB inequality in terms of tensor-like objects, but a lower bound on the intrinsic variance of estimators is not readily obtainable. The study in [44] expresses its results in terms of intrinsic distance. However, the Riemannian structure of the manifold where the estimator takes values is not arbitrary, but the one induced by its Fisher information metric (this choice of structure corresponds to the familiar Rao distance). The work in [54] improves this result, that is, it allows for arbitrary Riemannian metrics and applies to unbiased estimators. Our contribution in this chaper is thus the following one. We derive the IVLB maintaining the level of abstraction of the works in [27, 44]. That is, we express our results in the language of Riemannian differential geometry. However, contrary to the works in [27, 44], we arrive at a lower-bound which is formulated in terms of the intrinsic Riemannian distance (in contrast with [27] which provides tensor inequalities) and applies to arbitrary Riemannian structures (in contrast with [44] which assumes a specific one). Compared to [54], we allow for biased estimators.

Publications. The material presented in chapter 4 can be found in the conference papers [78, 79]. The work in [78] deals with unbiased estimation only, while [79] extends the IVLB to the general case of biased estimators. Here, we mainly refine the presentation (more details are discussed) and gather the examples contained in both papers.

1.3 Notation

In the following, we introduce the notation common to all chapters in this thesis. As the need arises, additional notation is defined within each chapter.

Mappings. Maps are usually denoted by capital letters in normal typeface ($F, G, \ldots$) or by greek letters ($\psi, \varphi, \ldots$). The symbols $\operatorname{dom} F$ and $\operatorname{image} F$ denote the domain and image of the map F, respectively. That is, if $F : \operatorname{dom} F \to S$ then $\operatorname{image} F = \{F(x) \in S : x \in \operatorname{dom} F\}$.

Matrix algebra. $\mathbb{N} = \{0, 1, 2, \ldots\}$, $\mathbb{Z} = \{0, \pm 1, \pm 2, \ldots\}$, $\mathbb{R}$, and $\mathbb{C}$ denote the set of

natural, integer, real, and complex numbers, respectively. Matrices (uppercase) and (column/row) vectors are in boldface type. Scalars are usually in lowercase and non-boldface type. $\mathbb{R}^{n\times m}$ and $\mathbb{R}^n$ denote the set of $n\times m$ matrices and the set of n-dimensional column vectors with real entries, respectively. A similar definition holds for $\mathbb{C}^n$ and $\mathbb{C}^{n\times m}$. The notations $(\cdot)^T$, $\overline{(\cdot)}$, $(\cdot)^H$, $(\cdot)^+$ and $(\cdot)^{-1}$ stand for the transpose, the complex conjugate, the Hermitean, the Moore-Penrose pseudo-inverse and the inverse operator, respectively. The symbols $\mathrm{tr}\,(\cdot)$ and $\mathrm{rank}\,(\cdot)$ represent the trace and the rank matrix functions. The Frobenius norm of $\boldsymbol{A}$ is denoted by

$$\|\boldsymbol{A}\| = \sqrt{\mathrm{tr}\left(\boldsymbol{A}^H\boldsymbol{A}\right)}.$$

The kernel (nullspace) of $\boldsymbol{A}\in\mathbb{C}^{n\times m}$ is represented by $\ker(\boldsymbol{A}) = \{\boldsymbol{x}\in\mathbb{C}^m : \boldsymbol{A}\boldsymbol{x}=\boldsymbol{0}\}$. The symbols $\boldsymbol{I}_n$, $\boldsymbol{0}_{n\times m}$ and

$$\boldsymbol{J}_n = \begin{bmatrix} 0 & & & & \\ 1 & 0 & & & \\ & 1 & 0 & & \\ & & \ddots & \ddots & \\ & & & 1 & 0 \end{bmatrix}$$

stand for the $n\times n$ identity, the $n\times m$ all-zero, and the $n\times n$ Jordan block (ones in the first lower diagonal) matrices, respectively. When the dimensions are clear from the context, the subscripts are dropped. Also, if either $n=0$ or $m=0$, we let $\boldsymbol{0}_{n\times m}$ be an empty matrix. The symbol $\boldsymbol{1}_n = (1,1,\ldots,1)^T$ denotes the n-dimensional vector with all entries equal to 1. For $m\in\mathbb{Z}$, we define the $n\times n$ shift matrices

$$\boldsymbol{K}_n[m] = \begin{cases} \boldsymbol{J}_n^m, & \text{if } m\geq 0 \\ \boldsymbol{J}_n^{-m}, & \text{if } m<0 \end{cases}.$$

Direct sum (diagonal concatenation) of matrices is represented by $\mathrm{diag}\,(\cdot)$,

$$\boldsymbol{A} = \mathrm{diag}\,(\boldsymbol{A}_1,\boldsymbol{A}_2,\ldots,\boldsymbol{A}_n) \quad\Leftrightarrow\quad \boldsymbol{A} = \begin{bmatrix} \boldsymbol{A}_1 & & & \\ & \boldsymbol{A}_2 & & \\ & & \ddots & \\ & & & \boldsymbol{A}_n \end{bmatrix}.$$

For a vector $\boldsymbol{a} = (a_1,a_2,\ldots,a_n)^T$, we define the corresponding diagonal matrix

$$\mathrm{Diag}\,(\boldsymbol{a}) = \begin{bmatrix} a_1 & & & \\ & a_2 & & \\ & & \ddots & \\ & & & a_n \end{bmatrix}.$$

The vectorization operator is denoted by $\mathrm{vec}\,(\cdot)$,

$$\boldsymbol{A} = \begin{bmatrix} \boldsymbol{a}_1 & \boldsymbol{a}_2 & \cdots & \boldsymbol{a}_m \end{bmatrix} : n\times m \quad\Rightarrow\quad \mathrm{vec}\,(\boldsymbol{A}) = \begin{bmatrix} \boldsymbol{a}_1 \\ \boldsymbol{a}_2 \\ \vdots \\ \boldsymbol{a}_m \end{bmatrix} : nm\times 1.$$

The Kronecker product of two matrices is represented by $\otimes$. If

$$\boldsymbol{A} = \begin{bmatrix} a_{11} & \cdots & a_{1m} \\ \vdots & \ldots & \vdots \\ a_{n1} & \cdots & a_{nm} \end{bmatrix} : n \times m$$

and $\boldsymbol{B} : k \times l$, then

$$\boldsymbol{A} \otimes \boldsymbol{B} = \begin{bmatrix} a_{11}\boldsymbol{B} & \cdots & a_{1m}\boldsymbol{B} \\ \vdots & \ldots & \vdots \\ a_{n1}\boldsymbol{B} & \cdots & a_{nm}\boldsymbol{B} \end{bmatrix} : nk \times ml.$$

We let $\mathbb{H}(n) = \left\{\boldsymbol{H} \in \mathbb{C}^{n\times n} : \boldsymbol{H}^H = \boldsymbol{H}\right\}$, $\mathbb{K}(n) = \left\{\boldsymbol{K} \in \mathbb{C}^{n\times n} : \boldsymbol{K}^H = -\boldsymbol{K}\right\}$, $\mathbb{U}(n) = \left\{\boldsymbol{U} \in \mathbb{C}^{n\times n} : \boldsymbol{U}^H\boldsymbol{U} = \boldsymbol{I}_n\right\}$ and $\mathbb{O}(n) = \left\{\boldsymbol{Q} : \mathbb{R}^{n\times n} : \boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}_n\right\}$ denote the set of Hermitean, skew-Hermitean, unitary and orthogonal matrices of size $n \times n$, respectively.

Polynomials and signals. The set of polynomials with coefficients in $\mathbb{C}$ and indeterminate z^{-1} is denoted by $\mathbb{C}[z]$. The polynomial

$$f(z) = \sum_{k=0}^{d} f[k]z^{-k}$$

is said to have degree d, written $\deg f(z) = d$, if $f[d] \neq 0$. The degree of the zero polynomial is defined to be $-\infty$. The set of polynomials with degree d is denoted by

$$\mathbb{C}_d[z] = \{f(z) \in \mathbb{C}[z] \,:\, \deg f(z) = d\}\,.$$

Similar definitions apply for $\mathbb{C}^n[z]$ and $\mathbb{C}^{n\times m}[z]$, the set of $n \times 1$ polynomial vectors and $n \times m$ polynomial matrices, respectively. The order of a polynomial matrix

$$\boldsymbol{F}(z) = \begin{bmatrix} \boldsymbol{f}_1(z) & \boldsymbol{f}_2(z) & \cdots & \boldsymbol{f}_m(z) \end{bmatrix} \in \mathbb{C}^{n\times m}[z]$$

is defined as $\operatorname{ord}\boldsymbol{F}(z) = \sum_{k=1}^{m} \deg \boldsymbol{f}_k(z)$. For the polynomial vector

$$\boldsymbol{f}(z) = \sum_{k=0}^{d} \boldsymbol{f}[k]z^{-k} \in \mathbb{C}_d^n[z],$$

and $l \in \mathbb{N}$, we define the $n(l+1) \times (d+1+l)$ block Sylvester matrix

$$\boldsymbol{\mathcal{T}}_l\left(\boldsymbol{f}(z)\right) = \begin{bmatrix} \boldsymbol{f}[0] & \cdots & \boldsymbol{f}[d] & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{f}[0] & \cdots & \boldsymbol{f}[d] & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \ddots & \boldsymbol{0} \\ \boldsymbol{0} & \cdots & \boldsymbol{0} & \boldsymbol{f}[0] & \cdots & \boldsymbol{f}[d] \end{bmatrix}.$$

In this thesis, we only consider discrete-time signals. Polynomial matrices

$$\boldsymbol{F}(z) = \sum_{k=0}^{d} \boldsymbol{F}[k]z^{-k} \in \mathbb{C}^{p\times q}[z]$$

act on (input) signals $\boldsymbol{x}[n] \in \mathbb{C}^q$ yielding (output) signals $\boldsymbol{y}[n] \in \mathbb{C}^p$:

$$\boldsymbol{y}[n] = \boldsymbol{F}(z) \circledast \boldsymbol{x}[n] \quad \Leftrightarrow \quad \boldsymbol{y}[n] = \sum_{k=0}^{d} \boldsymbol{F}[k]\boldsymbol{x}[n-k].$$

Let $\boldsymbol{x}[n] = (x_1[n], x_2[n], \ldots, x_m[n])^T$ denote a discrete-time signal. For an integer $l \geq 0$, the symbol $\boldsymbol{x}[n; l]$ denotes the signal

$$\boldsymbol{x}[n; l] = \begin{bmatrix} \boldsymbol{x}[n] \\ \boldsymbol{x}[n-1] \\ \vdots \\ \boldsymbol{x}[n-l] \end{bmatrix},$$

which is $(l+1)m$-dimensional. For an m-tuple of integers $\boldsymbol{l} = (l_1, l_2, \ldots, l_m)^T$, $l_m \geq 0$, the symbol $\boldsymbol{x}[n; \boldsymbol{l}]$ stands for the signal

$$\boldsymbol{x}[n; \boldsymbol{l}] = \begin{bmatrix} \boldsymbol{x}_1[n; l_1] \\ \boldsymbol{x}_2[n; l_2] \\ \vdots \\ \boldsymbol{x}_m[n; l_m] \end{bmatrix},$$

which is $(l_1 + l_2 + \cdots + l_m + m)$-dimensional. The 1D delta signal $\delta : \mathbb{Z} \to \mathbb{R}$ is defined as

$$\delta[n] = \begin{cases} 1, & \text{if } n = 0 \\ 0, & \text{if } n \neq 0 \end{cases}.$$

The 2D delta signal $\delta : \mathbb{Z} \times \mathbb{Z} \to \mathbb{R}$ is given by

$$\delta[n, m] = \begin{cases} 1, & \text{if } (n, m) = (0, 0) \\ 0, & \text{if } (n, m) \neq (0, 0) \end{cases}.$$

We use the same symbol δ for both the 1D and 2D delta signals, as the context easily resolves the ambiguity.

Chapter 2

Blind Channel Identification Based On 2nd Order Statistics

2.1 Chapter summary

Section 2.2 introduces the data model describing a discrete-time linear multiple-input multiple-output (MIMO) system with finite memory and noisy outputs. In section 2.3, we state the standard algebraic and statistical conditions which are assumed to hold with regard to the inputs, transfer matrix and observation noise of the MIMO system. Also, we formulate our blind channel identification problem (BCIP): we aim at identifying the transfer matrix of the MIMO system, from the 2nd order statistics (SOS) of its outputs. In section 2.4, we note that using only this statistical information, the MIMO channel is not fully identifiable. For example, at the very least, a phase ambiguity per input cannot be resolved. But other, much more severe ambiguities, can exist. We note that the input phase ambiguities can be modeled as an equivalence relation in the set of MIMO channels. Thus, after this set is partitioned into disjoint equivalence classes, we can, at most, identify each one of these equivalence classes. This obliges us to explicitly shift the target of our original BCIP: we abandon the identification of the MIMO channel (a meaningless goal), and now aim at estimating its equivalence class. Mathematically, the whole situation is described by introducing a mapping φ, which associates, to each MIMO channel equivalence class, the set of correlation matrices that it induces at the MIMO output. The mere fact that such map exists immediately asserts the unidentifiability of the MIMO channel itself, leaving however open the question of the identifiability of its equivalence class: is the mapping φ one-to-one ? In section 2.5, we settle this issue. We show that, with the standard assumption of white sources as channel inputs, the introduced map φ may fail to be injective, that is, two distinct channel equivalence classes may induce the same SOS at the channel output. This motivates us to pre-differentiate the inputs in the correlation domain, by assigning them distinct spectral colors. In the common multi-user digital communications context, this implies inserting a pre-filter in each transmitter (each user acts as an input in the MIMO system). Each pre-filter correlates the information symbols prior to transmission, thus coloring the previously flat spectral power density of the information-bearing random process. This pre-processing scheme requires no additional power or bandwidth. Furthermore, the original data rate is maintained and

no synchronization or cooperation between the sources is needed. We state an identifiability theorem (theorem 2.1). This theorem proves that, under a certain condition on the pre-filters, the map φ indeed becomes one-to-one, that is, the MIMO channel equivalence class becomes identifiable from the SOS of the channel output. We also state a feasibility theorem (theorem 2.2). This theorem proves that the set of pre-filters which meet the sufficient condition is dense in the set of all unit-norm, minimum-phase pre-filters. In other words, the sufficient condition in the identifiability theorem is easily fulfilled in practice. Having ensured the injectivity of φ and thereby the theoretical feasibility of the BCIP, we solve it in section 2.6. More precisely, we present an algorithm which reconstructs the MIMO channel equivalence class, given the correlation matrices observed at its output. This algorithm may also be interpreted as a computational scheme realizing the inverse map φ^{-1}. The algorithm is in closed-form (non-iterative), and leads itself to a natural implementation in parallel processors. We refer to this algorithm as the closed-form identification algorithm (CFIA). In section 2.7, we address the problem of decoding the sources' emitted information symbols, after the channel has been identified. We perform this under an additional channel impairment: baseband phase drifts due to carrier frequency misadjustements or Doppler effects. In multi-user setups, this distortion may be expected to appear in the baseband demodulated signals, as the receiver as to synchronize its local oscillator simultaneously with multiple independently generated carrier frequencies. We propose an iterative source separation and channel identification algorithm (IIA), which also permits to refine the closed-form channel estimated provided by the CFIA. Section 2.8 evaluates the performance of our proposed algorithms. We compare it with the transmitter induced conjugate cyclostationarity (TICC) approach in [12]. The simulation results show that our proposed technique yields symbol estimates with lower probability of error than TICC, in the presence of the aforementioned carrier frequency asynchronisms. Section 2.9 concludes this chaper.

2.2 Data model

MIMO channel. Consider a MIMO channel with P inputs $s_1[n], \ldots, s_P[n]$ and Q outputs $y_1[n], \ldots, y_Q[n]$ as depicted in figure 2.1. In figure 2.1, the signals $w_1[n], \ldots, w_Q[n]$ model observation noise. The pth input $s_p[n]$ appears at the qth output $y_q[n]$ through the sub-channel $h_{qp}(z)$. Each sub-channel $h_{qp}(z)$ is a finite impulse response (FIR) filter of degree D_{qp},

$$h_{qp}(z) = h_{qp}[0] + h_{qp}[1]\, z^{-1} + \cdots + h_{qp}[D_{qp}]\, z^{-D_{qp}}. \tag{2.1}$$

The qth output is given by

$$\begin{aligned} y_q[n] &= \sum_{p=1}^{P} h_{qp}(z) \circledast s_p[n] + w_q[n] \\ &= \sum_{p=1}^{P} \left(\sum_{d=0}^{D_{qp}} h_{qp}[d]\, s_p[n-d] \right) + w_q[n]. \end{aligned} \tag{2.2}$$

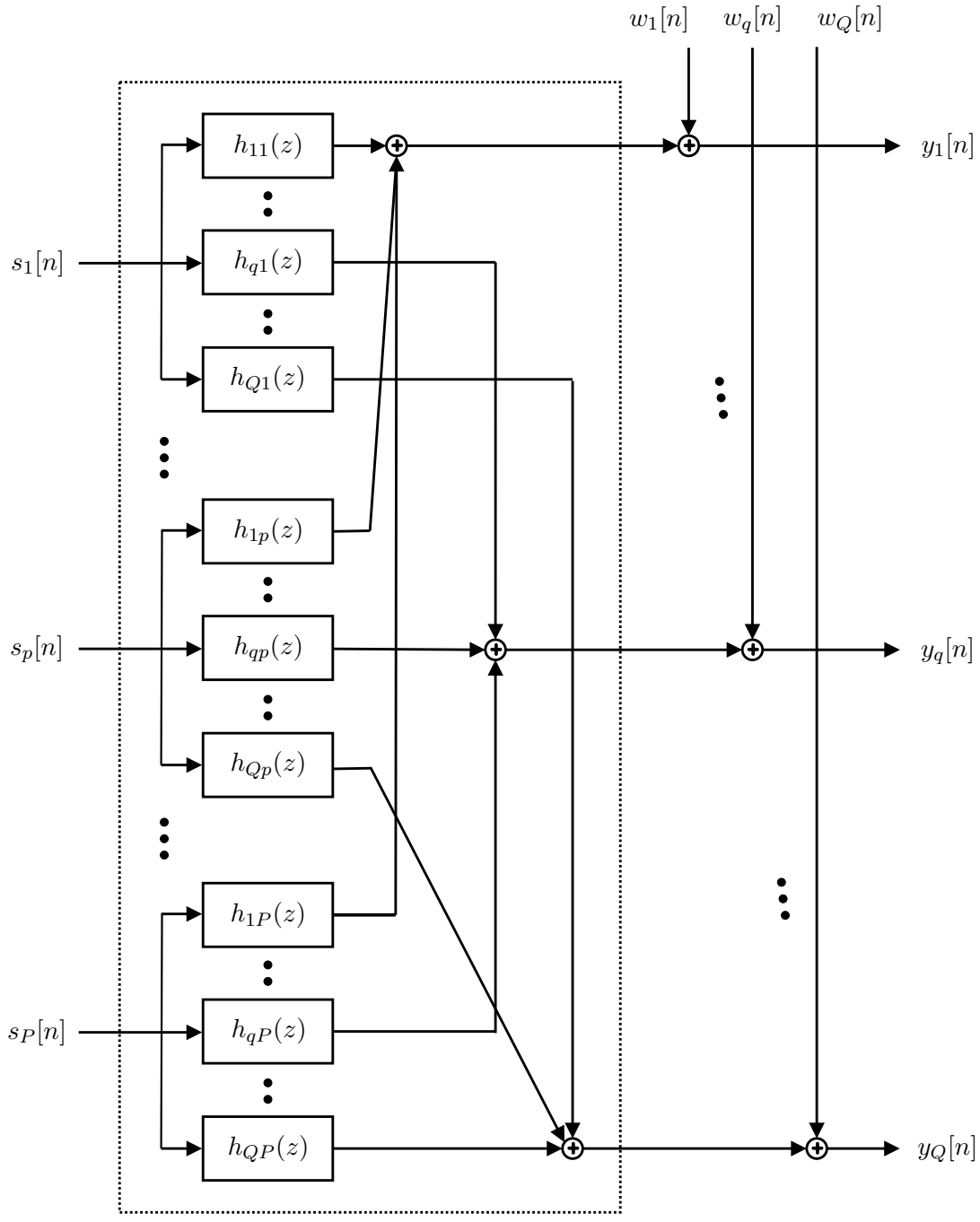

Figure 2.1: P-input/Q-output MIMO channel

Thus, each MIMO channel output is a noisy linear superposition of filtered versions of all inputs.

Matricial model. The MIMO channel in figure 2.1 is represented in more compact form in the block diagram of figure 2.2.

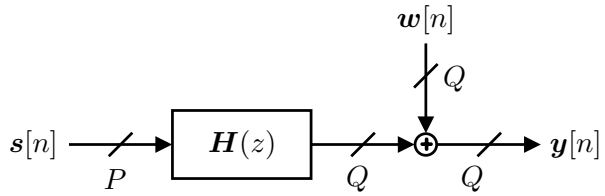

Figure 2.2: P-input/Q-output MIMO channel: block diagram

Here, $\boldsymbol{s}[n] = (s_1[n], \cdots, s_P[n])^T$, $\boldsymbol{y}[n] = (y_1[n], \cdots, y_Q[n])^T$, and $\boldsymbol{w}[n] = (w_1[n], \cdots, w_Q[n])^T$, denote the input, output and noise vectors, respectively. The $Q \times P$ polynomial matrix

$$\boldsymbol{H}(z) = \begin{bmatrix} h_{11}(z) & \cdots & h_{1p}(z) & \cdots & h_{1P}(z) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h_{q1}(z) & \cdots & h_{qp}(z) & \cdots & h_{qP}(z) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ h_{Q1}(z) & \cdots & h_{Qp}(z) & \cdots & h_{QP}(z) \end{bmatrix}$$

is termed the MIMO transfer matrix, or MIMO channel, and contains all subchannels. Its pth column is given by

$$\begin{aligned} \boldsymbol{h}_p(z) &= (h_{1p}(z), h_{2p}(z), \cdots, h_{Qp}(z))^T \\ &= \begin{bmatrix} h_{1p}[0] \\ h_{2p}[0] \\ \vdots \\ h_{Qp}[0] \end{bmatrix} + \begin{bmatrix} h_{1p}[1] \\ h_{2p}[1] \\ \vdots \\ h_{Qp}[1] \end{bmatrix} z^{-1} + \cdots + \begin{bmatrix} h_{1p}[D_p] \\ h_{2p}[D_p] \\ \vdots \\ h_{Qp}[D_p] \end{bmatrix} z^{-D_p} \\ &= \boldsymbol{h}_p[0] + \boldsymbol{h}_p[1]z^{-1} + \cdots + \boldsymbol{h}_p[D_p]z^{-D_p}, \end{aligned}$$

where $D_p = \max\{D_{1p}, D_{2p}, \ldots, D_{Qp}\}$. Thus, $\boldsymbol{h}_p(z)$, the pth column of $\boldsymbol{H}(z)$, contains the Q sub-channels activated by the pth input $s_p[n]$. With this notation, equation (2.2) gives rise to the matricial model

$$\boldsymbol{y}[n] = \sum_{p=1}^{P} \boldsymbol{h}_p(z) \circledast s_p[n] + \boldsymbol{w}[n] = \boldsymbol{H}(z) \circledast \boldsymbol{s}[n] + \boldsymbol{w}[n]. \tag{2.3}$$

The identification $\boldsymbol{H}(z) \simeq (\boldsymbol{d}; \boldsymbol{H})$. The polynomial channel matrix

$$\boldsymbol{H}(z) = \begin{bmatrix} \boldsymbol{h}_1(z) & \boldsymbol{h}_2(z) & \cdots & \boldsymbol{h}_P(z) \end{bmatrix}$$

is parameterized by **i)** discrete and **ii)** continuous variables. **i)** The discrete variables are the column degrees of $\boldsymbol{H}(z)$, $D_p = \deg \boldsymbol{h}_p(z)$, which we collect in the P-dimensional vector

$$\boldsymbol{d} = (D_1, D_2, \cdots, D_P)^T. \tag{2.4}$$

ii) The continuous variables are the coefficients of the column polynomial filters $\boldsymbol{h}_p(z)$, which we gather in the $Q \times (D+P)$ matrix

$$\boldsymbol{H} = \Big[\underbrace{\boldsymbol{h}_1[0]\,\boldsymbol{h}_1[1]\cdots\boldsymbol{h}_1[D_1]}_{\boldsymbol{H}_1} \;\; \underbrace{\boldsymbol{h}_2[0]\,\boldsymbol{h}_2[1]\cdots\boldsymbol{h}_2[D_2]}_{\boldsymbol{H}_2} \;\; \cdots \;\; \underbrace{\boldsymbol{h}_P[0]\,\boldsymbol{h}_P[1]\cdots\boldsymbol{h}_P[D_P]}_{\boldsymbol{H}_P} \Big]. \tag{2.5}$$

Here, $D = \operatorname{ord}\boldsymbol{H}(z) = \sum_{p=1}^{P} D_p$ denotes the order of the polynomial matrix $\boldsymbol{H}(z)$, that is, the sum of the degrees of its P column polynomial filters.

Equations (2.4) and (2.5) establish a one-to-one relationship between the object $\boldsymbol{H}(z)$ and the ordered pair $(\boldsymbol{d};\boldsymbol{H})$, denoted $\boldsymbol{H}(z) \simeq (\boldsymbol{d};\boldsymbol{H})$ in the sequel. This alternative representation for $\boldsymbol{H}(z)$ is more convenient for computations. Take the model (2.3), which is written in terms of the transfer matrix $\boldsymbol{H}(z)$. In terms of $(\boldsymbol{d};\boldsymbol{H})$, it reads as

$$\boldsymbol{y}[n] = \boldsymbol{H}\boldsymbol{s}[n;\boldsymbol{d}] + \boldsymbol{w}[n]. \tag{2.6}$$

Table 2.1 summarizes our MIMO channel model.

Data object			Dimension
$\boldsymbol{y}[n]$	$=$	$\boldsymbol{H}\boldsymbol{s}[n;\boldsymbol{d}] + \boldsymbol{w}[n]$	$Q \times 1$
$\boldsymbol{d}$	$=$	$(D_1, D_2, \cdots, D_P)^T$	$P \times 1$
$\boldsymbol{H}$	$=$	$\begin{bmatrix} \boldsymbol{H}_1 & \boldsymbol{H}_2 & \cdots & \boldsymbol{H}_P \end{bmatrix}$	$Q \times (D+P)$, $D = \sum_{p=1}^{P} D_p$
$\boldsymbol{H}_p$	$=$	$\begin{bmatrix} \boldsymbol{h}_p[0] & \boldsymbol{h}_p[1] & \cdots & \boldsymbol{h}_p[D_p] \end{bmatrix}$	$Q \times (D_p+1)$

Table 2.1: MIMO channel – data model

Stacked data model. For later use, it is also convenient to write down the equations of the stacked data model of order L, defined here as the one corresponding to the stacked observations $\boldsymbol{y}[n;L]$. The data model in table 2.1 corresponds to the special case $L = 0$. For the general situation, we have the stacked data model in table 2.2. Here, $\mathcal{D}_p = D_p + L$

Data object			Dimension
$\boldsymbol{y}[n;L]$	$=$	$\boldsymbol{\mathcal{H}}\,\boldsymbol{s}[n;\mathfrak{d}] + \boldsymbol{w}[n;L]$	$Q(L+1) \times 1$
$\mathfrak{d}$	$=$	$(\mathcal{D}_1, \mathcal{D}_2, \cdots, \mathcal{D}_P)^T$	$P \times 1$
$\boldsymbol{\mathcal{H}}$	$=$	$\begin{bmatrix} \boldsymbol{\mathcal{H}}_1 & \boldsymbol{\mathcal{H}}_2 & \cdots & \boldsymbol{\mathcal{H}}_P \end{bmatrix}$	$Q(L+1) \times (D+P+LP)$
$\boldsymbol{\mathcal{H}}_p$	$=$	$\boldsymbol{\mathcal{T}}_L\,(\boldsymbol{h}_p(z))$	$Q(L+1) \times (D_p+1+L)$

Table 2.2: MIMO channel – stacked data model of order L

and, recalling the definition of the Sylvester matrix operator $\boldsymbol{\mathcal{T}}$ in page 12, we have

$$\boldsymbol{\mathcal{H}}_p = \begin{bmatrix} \boldsymbol{h}_p[0] & \cdots & \boldsymbol{h}_p[D_p] & & \\ & \boldsymbol{h}_p[0] & \cdots & \boldsymbol{h}_p[D_p] & \\ & & \ddots & & \ddots \\ & & & \boldsymbol{h}_p[0] & \cdots \;\; \boldsymbol{h}_p[D_p] \end{bmatrix} : Q(L+1) \times (D_p+1+L).$$

2.3 The space $\mathbb{H}[z]$ and problem formulation

Assumptions. In the data model (2.3), we assume an algebraic property for the MIMO transfer matrix $\boldsymbol{H}(z)$ and a statistical condition on the random vectors $\boldsymbol{s}[n]$ and $\boldsymbol{w}[n]$:

A1. The $Q \times P$ polynomial matrix $\boldsymbol{H}(z)$ is: **i)** tall ($Q > P$), **ii)** irreducible ($\operatorname{rank} \boldsymbol{H}(z) = P$, for all $z \neq 0$ and $z = \infty$), **iii)** column-reduced ($\operatorname{rank} \begin{bmatrix} \boldsymbol{h}_1(D_1) & \cdots & \boldsymbol{h}_P(D_P) \end{bmatrix} = P$), and **iv)** memory limited by some known degree $D_{\max}$ ($\deg \boldsymbol{H}(z) \leq D_{\max}$). Hereafter, we let $\mathbb{H}[z]$ denote the set of all $Q \times P$ polynomial matrices satisfying condition A1 (P and Q are fixed);

A2. The inputs $s_p[n]$ and the noise $w_q[n]$ are zero-mean wide-sense stationary (WSS) processes. The inputs have unit-power, $\mathrm{E}\left\{|s_p[n]|^2\right\} = 1$, and are uncorrelated with each other, $\mathrm{E}\left\{s_p[n]\overline{s_q[m]}\right\} = 0$, for all $n, m \in \mathbb{Z}$ and $p \neq q$. The input and noise signals are uncorrelated with each other, $\mathrm{E}\left\{s_p[n]\overline{w_q[m]}\right\} = 0$, for all n, m, p, q, and the noise correlation matrices $\boldsymbol{R}_{\boldsymbol{w}}[m] = \mathrm{E}\left\{\boldsymbol{w}[n]\boldsymbol{w}[n-m]^H\right\}$ are known.

Assumption A1 requires the MIMO polynomial matrix $\boldsymbol{H}(z)$ to be irreducible and column-reduced. This is a common assumption in multi-user contexts, see [1, 2, 24, 68, 38], and believed to hold almost surely in realistic multipath models [2]. See [24] for a more detailed characterization and discussion on the generality of this channel assumption. However, other methodologies can tackle less restrictive channels [12, 31]. The main characteristic of irreducible and column-reduced channels is that they admit linear MIMO FIR inverses, that is, for each such channel $\boldsymbol{H}(z)$, there exists a polynomial matrix $\boldsymbol{E}(z)$ such that $\boldsymbol{E}(z)\boldsymbol{H}(z) = \boldsymbol{I}_P$, see [6]. The unit-power assumption in A2 entails no loss of generality, as $\boldsymbol{H}(z)$ may absorb any multiplicative constant. The noise correlation matrices may be assumed known, as they can be accurately estimated in the absence of input signals.

BCIP. In this chapter, we address the following blind channel identification problem (BCIP): given the 2nd order statistics of the MIMO channel output, that is, the set of output correlation matrices

$$\boldsymbol{R}_{\boldsymbol{y}} = \{\boldsymbol{R}_{\boldsymbol{y}}[m] \,:\, m \in \mathbb{Z}\}, \tag{2.7}$$

where $\boldsymbol{R}_{\boldsymbol{y}}[m] = \mathrm{E}\left\{\boldsymbol{y}[n]\boldsymbol{y}[n-m]^H\right\}$, find the MIMO transfer matrix $\boldsymbol{H}(z) \in \mathbb{H}[z]$.

In standard statistical language, BCIP asks for an estimator of $\boldsymbol{H}(z)$ based on the method of moments. We aim at extracting the transfer matrix $\boldsymbol{H}(z)$ from the 2nd-order moments of the observations $\boldsymbol{y}[n]$. To see the connection between the channel $\boldsymbol{H}(z)$ and the SOS of the MIMO channel output $\boldsymbol{R}_{\boldsymbol{y}}$ in (2.7), we express the output correlation matrices $\boldsymbol{R}_{\boldsymbol{y}}[m]$ in terms of the the transfer matrix $\boldsymbol{H}(z)$. More specifically, we use the identification $\boldsymbol{H}(z) \simeq (\boldsymbol{d}; \boldsymbol{H})$ to give $\boldsymbol{R}_{\boldsymbol{y}}[m]$ in terms of the vector of channel degrees $\boldsymbol{d}$ and the matrix of channel coefficients $\boldsymbol{H}$. For $\boldsymbol{e} = (E_1, E_2, \ldots, E_P)^T$, we use the notation $\boldsymbol{R}_{\boldsymbol{s}}[m; \boldsymbol{e}] = \mathrm{E}\left\{\boldsymbol{s}[n; \boldsymbol{e}]\boldsymbol{s}[n-m; \boldsymbol{e}]^H\right\}$ to designate the correlation matrix at lag m of $\boldsymbol{s}[n; \boldsymbol{e}]$. Note that, given the assumption A2 (uncorrelated inputs), this is a block diagonal matrix

$$\boldsymbol{R}_{\boldsymbol{s}}[m; \boldsymbol{e}] = \begin{bmatrix} \boldsymbol{R}_{\boldsymbol{s}_1}[m; E_1] & & & \\ & \boldsymbol{R}_{\boldsymbol{s}_2}[m; E_2] & & \\ & & \ddots & \\ & & & \boldsymbol{R}_{\boldsymbol{s}_P}[m; E_P] \end{bmatrix}, \tag{2.8}$$

with pth diagonal block

$$\boldsymbol{R}_{\boldsymbol{s}_p}[m; E_p] = \mathrm{E}\left\{\boldsymbol{s}_p[n; E_p]\boldsymbol{s}_p[n-m; E_p]^H\right\}. \tag{2.9}$$

Thus, with this notation, given the data model (2.6) and condition A2 (inputs and noise are uncorrelated), we have

$$\boldsymbol{R}_{\boldsymbol{y}}[m] = \boldsymbol{H}\boldsymbol{R}_{\boldsymbol{s}}[m; \boldsymbol{d}]\boldsymbol{H}^H + \boldsymbol{R}_{\boldsymbol{w}}[m]. \tag{2.10}$$

Equation (2.10) shows how the transfer matrix $\boldsymbol{H}(z) \simeq (\boldsymbol{d}; \boldsymbol{H})$ affects the 2nd order statistics of the observations.

2.4 The space $\mathbb{H}[z]/\sim$ and problem reformulation

In this section, we reformulate the BCIP and put it in firm mathematical ground. It will be clear from the forthcoming discussion that the original formulation is not well-posed. The correct formulation requires the introduction of a certain channel-to-statistics map φ. This viewpoint simplifies and unifies the analysis: channel identifiability corresponds to injectivity of φ (section 2.5) and solving the BCIP amounts to inverting φ (section 2.6).

We start by defining the set

$$\mathbb{C}_{\mathbb{Z}} = \prod_{m\in\mathbb{Z}} \boldsymbol{C}_m, \tag{2.11}$$

where $\boldsymbol{C}_m = \mathbb{C}^{Q\times Q}$. Thus, $\mathbb{C}_{\mathbb{Z}}$ is the Cartesian product of countably many copies of $\mathbb{C}^{Q\times Q}$. Each copy of $\mathbb{C}^{Q\times Q}$ plays the role of a coordinate axis within the space $\mathbb{C}_{\mathbb{Z}}$, as each factor $\mathbb{R}$ in the Cartesian product $\mathbb{R}^n = \mathbb{R}\times\cdots\times\mathbb{R}$. If $\boldsymbol{P}$ is a point in $\mathbb{C}_{\mathbb{Z}}$, we let $\boldsymbol{P}[m] \in \mathbb{C}^{Q\times Q}$ denote its mth coordinate ($m \in \mathbb{Z}$). A point $\boldsymbol{P}$ in $\mathbb{C}_{\mathbb{Z}}$ is specified by listing all its coordinates $\boldsymbol{P}[m]$, $m \in \mathbb{Z}$. Figure 2.3 illustrates these concepts (only three coordinate axis are shown).

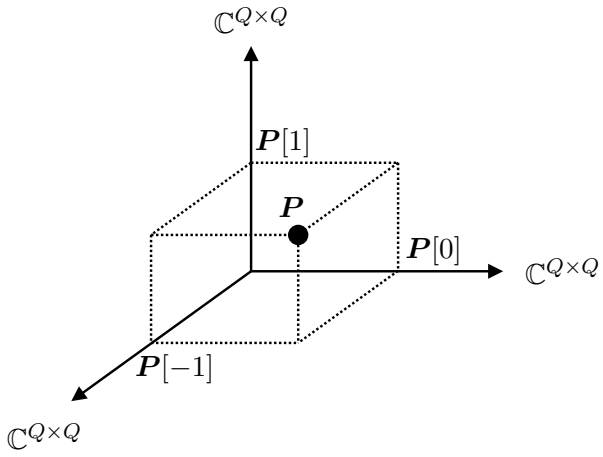

Figure 2.3: Sketch of space $\mathbb{C}_{\mathbb{Z}}$ and a point $\boldsymbol{P}$ in it

The set of all output correlation matrices $\boldsymbol{R}_{\boldsymbol{y}} = \{\boldsymbol{R}_{\boldsymbol{y}}[m] : m \in \mathbb{Z}\}$ in (2.7) is a point in the space $\mathbb{C}_{\mathbb{Z}}$. It depends on the underlying channel $\boldsymbol{H}(z) \simeq (\boldsymbol{d}; \boldsymbol{H})$ through (2.10). In

fact, equation (2.10) defines a map $\Phi : \mathbb{H}[z] \to \mathbb{C}_{\mathbb{Z}}$ which maps a transfer matrix $\boldsymbol{H}(z) \simeq (\boldsymbol{d}; \boldsymbol{H})$ to the point $\Phi(\boldsymbol{H}(z))$ with mth coordinate $\Phi(\boldsymbol{H}(z))[m] = \boldsymbol{H}\boldsymbol{R}_{\boldsymbol{s}}[m; \boldsymbol{d}]\boldsymbol{H}^H + \boldsymbol{R}_{\boldsymbol{w}}[m]$.

In the present context, it might be tempting to re-state BCIP as: given $\boldsymbol{R}_{\boldsymbol{y}} \in \operatorname{image}\Phi$, find $\Phi^{-1}(\boldsymbol{R}_{\boldsymbol{y}})$, that is, invert the map Φ. However, it turns out that Φ is not injective and, as such, its inverse does not exist. That is why the original formulation of BCIP is incorrect. The non-injectivity of Φ can be readily established as follows. Let $\boldsymbol{H}(z) \in \mathbb{H}[z]$ and consider transfer matrices $\boldsymbol{G}(z)$ given by

$$\boldsymbol{G}(z) = \boldsymbol{H}(z)\boldsymbol{\Theta}(\boldsymbol{\theta}), \tag{2.12}$$

where $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_P)^T \in \mathbb{R}^P$ and $\boldsymbol{\Theta}(\boldsymbol{\theta}) = \operatorname{diag}\left(e^{i\theta_1}, e^{i\theta_2}, \ldots, e^{i\theta_P}\right)$. Note that $\boldsymbol{G}(z) \in \mathbb{H}[z]$ and if some θ_p is not an integer multiple of 2π, then $\boldsymbol{G}(z) \neq \boldsymbol{H}(z)$. But, more importantly, the equality $\Phi(\boldsymbol{G}(z)) = \Phi(\boldsymbol{H}(z))$ holds, irrespective of the input autocorrelation sequences, $r_{s_p}[m] = \mathrm{E}\left\{s_p[n]\overline{s_p[n-m]}\right\}$. To check this, let $(\boldsymbol{e}; \boldsymbol{G})$ be defined by the identification $\boldsymbol{G}(z) \simeq (\boldsymbol{e}; \boldsymbol{G})$. Given (2.12), we have $\boldsymbol{e} = \boldsymbol{d}$ and $\boldsymbol{G} = \boldsymbol{H}\boldsymbol{\Lambda}$, where

$$\boldsymbol{\Lambda} = \operatorname{diag}\left(e^{i\theta_1}\boldsymbol{I}_{D_1+1},\, e^{i\theta_2}\boldsymbol{I}_{D_2+1},\, \cdots,\, e^{i\theta_P}\boldsymbol{I}_{D_P+1}\right). \tag{2.13}$$

Thus, the mth coordinate of the point $\Phi(\boldsymbol{G}(z))$ in the space $\mathbb{C}_{\mathbb{Z}}$ is given by

$$\begin{aligned}
\Phi(\boldsymbol{G}(z))[m] &= \boldsymbol{G}\boldsymbol{R}_{\boldsymbol{s}}[m; \boldsymbol{e}]\boldsymbol{G}^H + \boldsymbol{R}_{\boldsymbol{w}}[m] \\
&= \boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{R}_{\boldsymbol{s}}[m; \boldsymbol{d}]\overline{\boldsymbol{\Lambda}}\boldsymbol{H}^H + \boldsymbol{R}_{\boldsymbol{w}}[m] \\
&= \boldsymbol{H}\boldsymbol{R}_{\boldsymbol{s}}[m; \boldsymbol{d}]\boldsymbol{\Lambda}\overline{\boldsymbol{\Lambda}}\boldsymbol{H}^H + \boldsymbol{R}_{\boldsymbol{w}}[m] \\
&= \boldsymbol{H}\boldsymbol{R}_{\boldsymbol{s}}[m; \boldsymbol{d}]\boldsymbol{H}^H + \boldsymbol{R}_{\boldsymbol{w}}[m] \\
&= \Phi(\boldsymbol{H}(z))[m].
\end{aligned}$$

Notice that $\boldsymbol{R}_{\boldsymbol{s}}[m; \boldsymbol{d}]$ commutes with $\boldsymbol{\Lambda}$ because both share the same block diagonal structure, see (2.8) and (2.13), the pth block of $\boldsymbol{\Lambda}$ being the matrix $e^{i\theta_p}\boldsymbol{I}_{D_p+1}$ which commutes with all matrices of the same size. The non-injectivity of Φ, irrespective of the input autocorrelation sequences $r_{s_p}[m]$, corresponds to a well-known MIMO channel identifiability bound: with 2nd order statistics, the channel $\boldsymbol{H}(z)$ can only be identified up to a phase offset per user.

To bypass the non-injectivity of Φ, we must work with a version of Φ which acts on equivalence classes of channels in $\mathbb{H}[z]$, rather than on single points (channels) in $\mathbb{H}[z]$. We let φ denote this induced version. Formally, the construction is as follows. Introduce the relation $\sim$ on the set of MIMO channel matrices $\mathbb{H}[z]$ by declaring $\boldsymbol{G}(z) \sim \boldsymbol{H}(z)$ if (2.12) holds. Thus, $\sim$ denotes equality of polynomial matrices modulo a phase offset per column. It is easily checked that $\sim$ is an equivalence relation. Recall that a relation $\sim$ on a set X is said to be an equivalence relation if it is reflexive ($x \sim x$, for all $x \in X$), symmetric ($x \sim y$ implies $y \sim x$) and transitive ($x \sim y$ and $y \sim z$ imply $x \sim z$). We let $\mathbb{H}[z]/\sim$ denote the set of equivalence classes and $\pi : \mathbb{H}[z] \to \mathbb{H}[z]/\sim$ the map which projects each $\boldsymbol{H}(z)$ to its equivalence class $\pi(\boldsymbol{H}(z))$, also written $[\boldsymbol{H}(z)]$. Now, we make Φ descend to the quotient thereby inducing a quotient map φ such that the diagram in figure 2.4 commutes, that is, $\Phi = \varphi \circ \pi$.

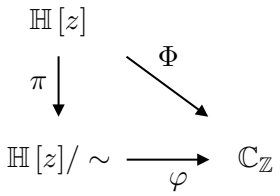

Figure 2.4: The map Φ descends to a quotient map φ

In equivalent terms, we are defining $\varphi([\boldsymbol{H}(z)]) = \Phi(\boldsymbol{H}(z))$, and φ is well-defined (the result does not depend on a particular representative of the equivalence class) because, as shown previously, $\Phi(\boldsymbol{H}(z)) = \Phi(\boldsymbol{G}(z))$, whenever $\boldsymbol{G}(z) \sim \boldsymbol{H}(z)$. In the next section, we prove that φ can be made injective by imposing a certain condition on the 2nd order statistics of the inputs. This will lead to a correct map formulation of the BCIP, anticipated here: given $\boldsymbol{R_y} \in \operatorname{image}\varphi$, find $\varphi^{-1}(\boldsymbol{R_y})$. In other words, we will shift our initial formulation of the BCIP. We will no longer aim for the identification of a channel (as seen, this is a meaningless goal). Rather, we will solve for its equivalence class.

2.5 Ensuring the identifiability of $\mathbb{H}[z]/\sim$

White inputs. Based only on assumptions A1 and A2, the map φ may also fail to be injective. That is, distinct channel equivalence classes may induce the same 2nd order statistics at the MIMO channel output. This also renders the BCIP meaningless. Consider the most common scenario: P spectrally white sources $a_1[n], \ldots, a_P[n]$ are plugged to the P inputs of the MIMO channel, see figure 2.5.

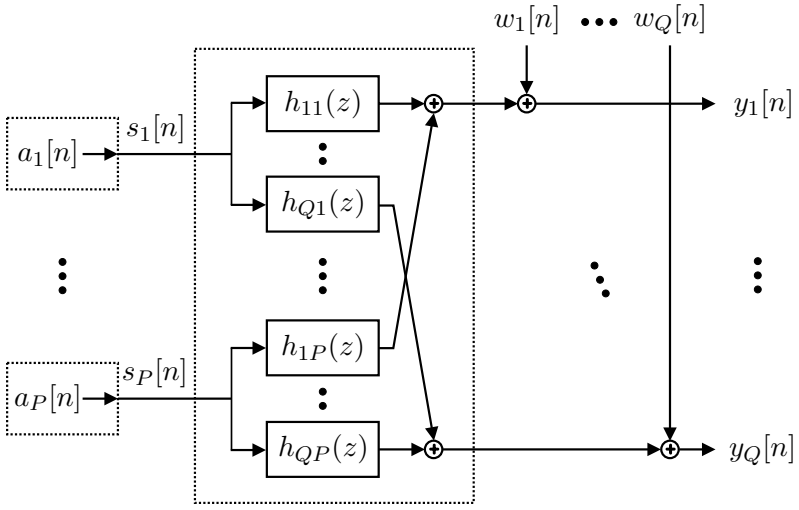

Figure 2.5: P-input/Q-output MIMO channel with P spectrally white sources as inputs

A zero-mean WSS random signal $x[n]$ is said to be spectrally white if its autocorrelation sequence is a delta signal, $r_x[m] = \mathrm{E}\left\{x[n]\overline{x[n-m]}\right\} = \delta[m]$. Equivalently, the power spectral density of $x[n]$ is flat. Notice that spectrally white sources arise naturally from

the common digital communication scenario. In that case, each $a_p[n]$ usually denotes an infinite string of independent and identically distributed (iid) information symbols drawn from a finite modulation alphabet $\mathbb{A} \subset \mathbb{C}$ such as $\mathbb{A} = \mathbb{A}_{\text{BSK}} = \{\pm 1\}$. The independence of the symbols implies, in particular, that they are uncorrelated.

The map φ can be seen to be non-injective with a simple example. Let $P = 2$, fix $Q > P$ and choose a channel with equal input degrees: $\boldsymbol{H}(z) \simeq (\boldsymbol{d}; \boldsymbol{H})$ with $\boldsymbol{d} = (D_1, D_2)^T$, $D_p = D_0$, for some $D_0 \in \mathbb{N}$. Since $s_p[n] = a_p[n]$ and each $a_p[n]$ is spectrally white, it can be checked that $\boldsymbol{R}_{\boldsymbol{s}_p}[m; D_p] = \boldsymbol{K}_{D_p+1}[m]$. Recall from page 11 that $\boldsymbol{K}_n[m]$ denotes the mth shift matrix of size $n \times n$. As a consequence,

$$\boldsymbol{R}_{\boldsymbol{s}}[m; \boldsymbol{d}] = \begin{bmatrix} \boldsymbol{R}_{\boldsymbol{s}_1}[m; D_1] & \boldsymbol{0} \\ \boldsymbol{0} & \boldsymbol{R}_{\boldsymbol{s}_2}[m; D_2] \end{bmatrix} = \boldsymbol{I}_2 \otimes \boldsymbol{K}_{D_0+1}[m]. \tag{2.14}$$

Therefore, the map φ sends $[\boldsymbol{H}(z)]$ to the point in $\mathbb{C}_{\mathbb{Z}}$ whose mth coordinate is

$$\begin{aligned} \varphi\left([\boldsymbol{H}(z)]\right)[m] &= \boldsymbol{H}\boldsymbol{R}_{\boldsymbol{s}}[m; \boldsymbol{D}]\boldsymbol{H}^H + \boldsymbol{R}_{\boldsymbol{w}}[m] \\ &= \boldsymbol{H}\left(\boldsymbol{I}_2 \otimes \boldsymbol{K}_{D_0+1}[m]\right)\boldsymbol{H}^H + \boldsymbol{R}_{\boldsymbol{w}}[m]. \end{aligned}$$

Define another channel $\boldsymbol{G}(z) = \boldsymbol{H}(z)\boldsymbol{Q}$, where

$$\boldsymbol{Q} = \frac{1}{\sqrt{2}} \begin{bmatrix} 1 & -1 \\ 1 & 1 \end{bmatrix}.$$

Then, $\boldsymbol{G}(z) \simeq (\boldsymbol{e}; \boldsymbol{G})$ with $\boldsymbol{e} = \boldsymbol{d}$ and

$$\boldsymbol{G} = \boldsymbol{H}\left(\boldsymbol{Q} \otimes \boldsymbol{I}_{D_0+1}\right). \tag{2.15}$$

Using (2.14) and (2.15), the image of $[\boldsymbol{G}(z)]$ under φ is the point in $\mathbb{C}_{\mathbb{Z}}$ with mth coordinate

$$\begin{aligned} \varphi\left([\boldsymbol{G}(z)]\right)[m] &= \boldsymbol{G}\boldsymbol{R}_{\boldsymbol{s}}[m; \boldsymbol{e}]\boldsymbol{G}^H + \boldsymbol{R}_{\boldsymbol{w}}[m] \\ &= \boldsymbol{H}\left(\boldsymbol{Q} \otimes \boldsymbol{I}_{D_0+1}\right)\left(\boldsymbol{I}_2 \otimes \boldsymbol{K}_{D_0+1}[m]\right)\left(\boldsymbol{Q}^H \otimes \boldsymbol{I}_{D_0+1}\right)\boldsymbol{H}^H + \boldsymbol{R}_{\boldsymbol{w}}[m] \\ &= \boldsymbol{H}\left(\boldsymbol{Q}\boldsymbol{Q}^H \otimes \boldsymbol{K}_{D_0+1}[m]\right)\boldsymbol{H}^H + \boldsymbol{R}_{\boldsymbol{w}}[m] \\ &= \boldsymbol{H}\left(\boldsymbol{I}_2 \otimes \boldsymbol{K}_{D_0+1}[m]\right)\boldsymbol{H}^H + \boldsymbol{R}_{\boldsymbol{w}}[m]. \end{aligned}$$

Here, we used the Kronecker product property $(\boldsymbol{A} \otimes \boldsymbol{B})(\boldsymbol{C} \otimes \boldsymbol{D}) = \boldsymbol{A}\boldsymbol{C} \otimes \boldsymbol{B}\boldsymbol{D}$, for conformable matrices $\boldsymbol{A}, \boldsymbol{B}, \boldsymbol{C}, \boldsymbol{D}$, see [39, page 28]. Since φ sends $[\boldsymbol{G}(z)] \neq [\boldsymbol{H}(z)]$ to the same image, it is not one-to-one.

Colored inputs. Thus, if φ is to be injective, the random signals seen at the P inputs of the MIMO channel cannot be all simultaneously white. The data sources must imprint their information through colored signals. This means that we need a mechanism that, within the pth source, transforms the uncorrelated signal $a_p[n]$ into a correlated one, say $s_p[n]$ for convenience, which becomes the new source output. This mechanism must be reversible, that is, $a_p[n]$ must be recoverable from $s_p[n]$, or else, information may be lost. We propose to implement such a mechanism through unit-power minimum phase filters. That is, the pth new information bearing signal $s_p[n]$ is a filtered version of the pth original white signal $a_p[n]$,

$$s_p[n] = c_p(z) \circledast a_p[n]. \tag{2.16}$$

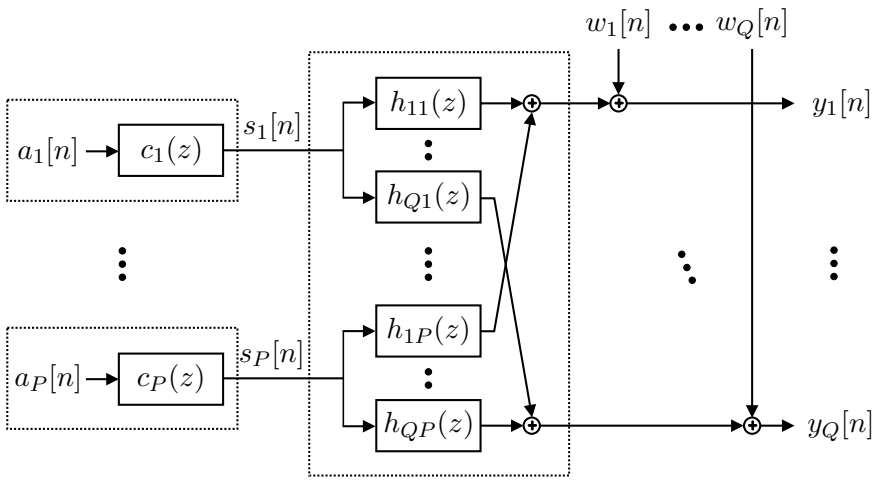

Figure 2.6: P-input/Q-output MIMO channel with P colored sources as inputs

Figure 2.6 depicts the new scenario discussed here. Compare with figure 2.5 which illustrates the original scenario.

For each $p = 1, 2, \ldots, P$,

$$c_p(z) = \sum_{d=0}^{C_p} c_p[d] z^{-d}$$

denotes an unit-power, that is,

$$\sum_{d=0}^{C_p} |c_p[d]|^2 = 1,$$

minimum phase filter of degree C_p. Moreover, without loss of generality, we assume $c_p[0] \neq 0$ (nonzero precursor). For future reference, we define

$$\mathbb{M}_C[z] = \left\{ c(z) = \sum_{d=0}^{C} c[d] z^{-d} \in \mathbb{C}_C[z] \,:\, c(z) \text{ is unit-power, minimum phase and } c[0] \neq 0 \right\}. \tag{2.17}$$

Thus, $c_p(z) \in \mathbb{M}_{C_p}[z]$, for all p. The unit-power property of each pre-filter guarantees that the transmission power is maintained relative to the original scenario, that is,

$$\mathrm{E}\left\{ |s_p[n]|^2 \right\} = \mathrm{E}\left\{ |a_p[n]|^2 \right\},$$

as can be easily seen. The minimum phase property ensures the existence of a stable inverse for the filter $c_p(z)$, hence, the reversibility of the transformation. Notice also that this filtering mechanism preserves the original data rate.

By passing the white signal $a_p[n]$ through the filter $c_p(z)$, we induce a spectral color in the random signal $s_p[n]$. Note that the 2nd order statistics of $s_p[n]$ are completely determined by $c_p(z)$. It is useful to expose this dependence in an explicit manner. We do so by writing the correlation matrices of the pth colored signal $\boldsymbol{s}_p[n; E_p]$, that is, the matrices $\boldsymbol{R}_{\boldsymbol{s}_p}[m; E_p]$ in (2.9), in terms of the coefficients of the correlative filter $c_p(z)$. Here, $E_p \in \mathbb{N}$ denotes a generic stacking parameter. Similarly, for $N \in \mathbb{N}$ and $m \in \mathbb{Z}$, we

define $\boldsymbol{R}_{\boldsymbol{a}_p}[m;N] = \mathrm{E}\left\{\boldsymbol{a}_p[n;N]\boldsymbol{a}_p[n-m;N]^H\right\}$. Since $a_p[n]$ is a spectrally white signal,

$$\boldsymbol{R}_{\boldsymbol{a}_p}[m;N] = \boldsymbol{K}_{N+1}[m]. \tag{2.18}$$

Given (2.16), we have

$$\boldsymbol{s}_p[n;E_p] = \boldsymbol{\mathcal{T}}_{E_p}\left(c_p(z)\right)\boldsymbol{a}_p[n;E_p+C_p]. \tag{2.19}$$

Note that

$$\boldsymbol{\mathcal{T}}_{E_p}\left(c_p(z)\right) = \begin{bmatrix} c_p[0] & \cdots & c_p[C_p] & & \\ & c_p[0] & \cdots & c_p[C_p] & \\ & & \ddots & & \ddots \\ & & & c_p[0] & \cdots & c_p[C_p] \end{bmatrix} : (E_p+1)\times(E_p+C_p+1)\,. \tag{2.20}$$

Finally, using (2.19) and then (2.18),

$$\begin{aligned} \boldsymbol{R}_{\boldsymbol{s}_p}[m;E_p] &= \boldsymbol{\mathcal{T}}_{E_p}\left(c_p(z)\right)\,\boldsymbol{R}_{\boldsymbol{a}_p}[m;E_p+C_p]\,\boldsymbol{\mathcal{T}}_{E_p}\left(c_p(z)\right)^H \\ &= \boldsymbol{\mathcal{T}}_{E_p}\left(c_p(z)\right)\,\boldsymbol{K}_{E_p+C_p+1}[m]\,\boldsymbol{\mathcal{T}}_{E_p}\left(c_p(z)\right)^H. \end{aligned} \tag{2.21}$$

Equation (2.21) shows how the pth input correlation matrices $\boldsymbol{R}_{\boldsymbol{s}_p}[m;E_p]$ depend on the coefficients of the pth correlative filter $c_p(z)$.

Spectral diversity assumption. It turns out that, if the pre-filters induce sufficiently distinct spectral colors to the random processes $s_p[n]$, then the map $\varphi : \mathbb{H}/\sim \longrightarrow \mathbb{C}_{\mathbb{Z}}$, which sends equivalence classes of channels to output correlation matrices, becomes injective, thereby turning the BCIP into a (at least, theoretically) feasible problem. In equivalent words, the channel $\boldsymbol{H}(z)$ becomes identifiable, up to a phase offset per column, from the 2nd order statistics of the MIMO channel output. Before stating this spectral diversity condition, we must introduce new notation. For a matrix $\boldsymbol{A} \in \mathbb{C}^{n\times n}$, we let $\sigma\left(\boldsymbol{A}\right) = \{\lambda_1, \lambda_2, \ldots, \lambda_n\}$ denote its spectrum, that is, the set of its eigenvalues (including multiplicities). The normalized correlation matrix at lag m of a zero-mean WSS random multivariate process $\boldsymbol{x}[n]$ is defined as

$$\boldsymbol{\Gamma}_{\boldsymbol{x}}[m] = \boldsymbol{R}_{\boldsymbol{x}}[0]^{-1/2}\boldsymbol{R}_{\boldsymbol{x}}[m]\boldsymbol{R}_{\boldsymbol{x}}[0]^{-1/2},$$

where $\boldsymbol{R}_{\boldsymbol{x}}[m] = \mathrm{E}\left\{\boldsymbol{x}[n]\boldsymbol{x}[n-m]^H\right\}$. This is the matricial counterpart of the usual correlation coefficient at lag m of a given zero-mean WSS scalar random process $x[n]$,

$$\begin{aligned} \gamma_x[m] &= \frac{\mathrm{E}\left\{x[n]\overline{x[n-m]}\right\}}{\sqrt{\mathrm{E}\left\{|x[n]|^2\right\}}\sqrt{\mathrm{E}\left\{|x[n-m]|^2\right\}}} \\ &= \frac{r_x[m]}{\sqrt{r_x[0]}\sqrt{r_x[0]}} \\ &= r_x[0]^{-1/2}r_x[m]r_x[0]^{-1/2}. \end{aligned}$$

We let $\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m;E_p]$ denote the normalized correlation matrix at lag m of $\boldsymbol{s}_p[n;E_p]$. Thus,

$$\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m;E_p] = \boldsymbol{R}_{\boldsymbol{s}_p}[0;E_p]^{-1/2}\boldsymbol{R}_{\boldsymbol{s}_p}[m;E_p]\boldsymbol{R}_{\boldsymbol{s}_p}[0;E_p]^{-1/2}. \tag{2.22}$$

Notice that $\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; E_p]$ in (2.22) is a function of $c_p(z)$ through the matrices $\boldsymbol{R}_{\boldsymbol{s}_p}[0; E_p]$ and $\boldsymbol{R}_{\boldsymbol{s}_p}[m; E_p]$. Similarly,

$$\boldsymbol{\Gamma}_{\boldsymbol{s}}[m; \boldsymbol{e}] = \boldsymbol{R}_{\boldsymbol{s}}[0; \boldsymbol{e}]^{-1/2} \boldsymbol{R}_{\boldsymbol{s}}[m; \boldsymbol{e}] \boldsymbol{R}_{\boldsymbol{s}}[0; \boldsymbol{e}]^{-1/2} \tag{2.23}$$

denotes the normalized correlation matrix at lag m of $\boldsymbol{s}[n; \boldsymbol{e}]$. Since the inputs are uncorrelated (condition A2 in page 20), this is a block diagonal matrix,

$$\boldsymbol{\Gamma}_{\boldsymbol{s}}[m; \boldsymbol{e}] = \begin{bmatrix} \boldsymbol{\Gamma}_{\boldsymbol{s}_1}[m; E_1] & & & \\ & \boldsymbol{\Gamma}_{\boldsymbol{s}_2}[m; E_2] & & \\ & & \ddots & \\ & & & \boldsymbol{\Gamma}_{\boldsymbol{s}_P}[m; E_P] \end{bmatrix}. \tag{2.24}$$

We are now in position to state our final assumption.

A3. The pre-filters are memory limited by some known degree $C_{\max}$ ($\deg c_p(z) = C_p \leq C_{\max}$). The P data sources correlate their white information sequences $a_p[n]$, that is, employ pre-filters $c_p(z)$, such that: for each $p = 1, 2, \ldots, P$, there exists a correlation lag $m(p) \in \mathbb{Z}$ satisfying

$$\sigma\left(\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m(p); E_p]\right) \cap \left[\bigcup_{q \neq p} \sigma\left(\boldsymbol{\Gamma}_{\boldsymbol{s}_q}[m(p); E_q]\right)\right] = \emptyset, \tag{2.25}$$

for all $0 \leq E_1, E_2, \ldots, E_P \leq E$. Here, E is some pre-chosen constant which verifies

$$E \geq (P+1) D_{\max} + C_{\max}.$$

Condition A3 means that, for each source $p = 1, 2, \ldots, P$, there must exist one correlation lag $m(p)$ which makes the spectra of $\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m(p); E_p]$ disjoint from those of the remaining sources, that is, from the spectra of $\boldsymbol{\Gamma}_{\boldsymbol{s}_q}[m(p); E_q]$, for $q \neq p$. Furthermore, this non-overlapping spectral condition must hold (with the same correlation lag $m(p)$) for any stacking parameters $E_1, E_2, \ldots, E_P$ taken in $\{0, 1, 2, \ldots, E\}$.

Identifiability theorem. As promised, the spectral richness introduced by the pre-filters, quantified in condition A3, guarantees the injectivity of φ. We formally state this result as theorem 2.1.

Theorem 2.1. *Consider the signal model in (2.3), and assume that conditions A1-A3 are fulfilled. Then, the map $\varphi : \mathbb{H}[z]/\sim \to \mathbb{C}_{\mathbb{Z}}$ is one-to-one, that is,*

$$\varphi([\boldsymbol{G}(z)]) = \varphi([\boldsymbol{H}(z)]) \quad \Rightarrow \quad [\boldsymbol{G}(z)] = [\boldsymbol{H}(z)].$$

Proof: See appendix A.

In terms of the non-injective map Φ (which acts on channels, not on equivalence classes, recall figure 2.4), the theorem reads: $\Phi(\boldsymbol{G}(z)) = \Phi(\boldsymbol{H}(z))$ if and only if $\boldsymbol{G}(z) \sim \boldsymbol{H}(z)$.

That is, two channels induce the same output 2nd order statistics if and only if they are equal modulo a phase ambiguity per column. Finally, this theorem permits us to formulate the BCIP as: given $\boldsymbol{R_y} \in \operatorname{image}\varphi$, find $\varphi^{-1}(\boldsymbol{R_y})$. In words, solving the BCIP consists in pinpointing the underlying equivalence class of the MIMO channel, given the 2nd order statistics of the output.

Connection with [5]. Assumption A3 admits a drastic simplification when the column degrees $D_p = \deg \boldsymbol{h}_p(z)$ of the channel matrix $\boldsymbol{H}(z) = [\,\boldsymbol{h}_1(z)\,\boldsymbol{h}_2(z)\,\cdots\,\boldsymbol{h}_P(z)\,]$ are known beforehand. In this very special case, it can be seen from the proof of theorem 2.1 that it suffices that, for each pair of sources (p, q), with $p \neq q$, there exists a correlation lag $m = m(p, q)$ such that

$$\sigma\left(\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; 1 + D_p + L)\right) \cap \sigma\left(\boldsymbol{\Gamma}_{\boldsymbol{s}_q}(m; 1 + D_q + L)\right) = \varnothing. \tag{2.26}$$

Here, L denotes a sufficiently large stacking parameter making the channel matrix $\boldsymbol{\mathcal{H}}$, which corresponds to the stacked observations $\boldsymbol{y}[n; L]$ (recall table 2.2), full column-rank. In particular, for the scenario of overdetermined static mixtures of sources (the case addressed in [5]), we have $D_p = 0$ and $L = 0$, and the condition in (2.26) recovers precisely the spectral conditions of the identifiability theorem 2 in [5].

Correlative filters. In the remaining of this section we investigate the feasibility of condition A3. This condition requires the pre-filters $c_1(z), \ldots, c_P(z)$ to insert a sufficiently diverse spectral structure in the random signals $s_1[n], \ldots, s_P[n]$, through the normalized input correlation matrices $\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; E_p]$. At first sight, this condition imposed on the pre-filters might appear too restrictive. In the following, we prove that it is not. In fact, we show that, in a certain sense to be explained shortly, almost every P-tuple $(c_1(z), \ldots, c_P(z))$ of pre-filters fulfills the requirements of condition A3 . This result is formally stated in theorem 2.2 below.

The first step towards theorem 2.2 is to endow each set $\mathbb{M}_C[z]$ in (2.17) with a metric space structure. This is accomplished by identifying $\mathbb{M}_C[z]$ with a subset of $\mathbb{C}^{C+1}$ through the map $\iota : \mathbb{M}_C[z] \to \mathbb{C}^{C+1}$ given by

$$c(z) = \sum_{d=0}^{C} c[d] z^{-d} \quad \overset{\iota}{\mapsto} \quad (c[0], c[1], \cdots, c[C])^T,$$

and letting this identification induce a distance function d on $\mathbb{M}_C[z]$. More precisely, $d : \mathbb{M}_C[z] \times \mathbb{M}_C[z] \to \mathbb{R}$ is defined by $d(c(z), d(z)) = \|\iota(c(z)) - \iota(d(z))\|$, where $\|\cdot\|$ denotes the usual Euclidean norm in $\mathbb{C}^{C+1}$. Hereafter, when we think of $\mathbb{M}_C[z]$ as a metric space, we are implicitly assuming it equipped with the metric d.

The next step is to define, for a given P-tuple of degrees, $\boldsymbol{c} = (C_1, C_2, \cdots, C_P)^T \in \mathbb{N}^P$, the Cartesian product $\mathbb{M}_{\boldsymbol{c}}[z] = \mathbb{M}_{C_1}[z] \times \mathbb{M}_{C_2}[z] \times \cdots \times \mathbb{M}_{C_P}[z]$. Thus, a point $\boldsymbol{c}(z)$ in $\mathbb{M}_{\boldsymbol{c}}[z]$ is a P-tuple of filters, the pth filter taken from the appropriate factor space $\mathbb{M}_{C_p}[z]$, that is, $\boldsymbol{c}(z) = (c_1(z), c_2(z), \cdots, c_P(z))^T \in \mathbb{M}_{\boldsymbol{c}}[z]$ if and only if $c_p(z) \in \mathbb{M}_{C_p}[z]$. As a Cartesian product of metric spaces, $\mathbb{M}_{\boldsymbol{c}}[z]$ is itself a metric space, with natural distance function given by

$$d(\boldsymbol{c}(z), \boldsymbol{d}(z)) = \sqrt{\sum_{p=1}^{P} d(c_p(z), d_p(z))^2}.$$

Notice that we are using the same symbol d to denote metrics on distinct spaces (in each factor and the Cartesian product itself). We rely on the context to resolve the ambiguity.

Now, think of each point in the space $\mathbb{M}_{\boldsymbol{c}}[z]$ as a P-tuple of candidate correlative filters. We say candidate because some points satisfy condition A3 and others do not. We let $\mathbb{F}_{\boldsymbol{c}}[z] \subset \mathbb{M}_{\boldsymbol{c}}[z]$ denote the subset of those fulfilling condition A3. Theorem 2.2 asserts that $\mathbb{F}_{\boldsymbol{c}}[z]$ occupies almost all the totality of the space $\mathbb{M}_{\boldsymbol{c}}[z]$, whenever $\boldsymbol{c} = (C_1, C_2, \cdots, C_P)^T$ is an all non-zero P-tuple of degrees ($C_p \neq 0$ for all p). Of course, we must also have $C_p \leq C_{\max}$, in order to comply with condition A3. We say that the P-tuple $\boldsymbol{c}$ is upper-bounded by $C_{\max}$.

Theorem 2.2. *Let $\boldsymbol{c}$ be an all non-zero P-tuple of degrees upper-bounded by $C_{\max}$. Then, $\mathbb{F}_{\boldsymbol{c}}[z]$ is dense in $\mathbb{M}_{\boldsymbol{c}}[z]$.*

Proof: See appendix A.

Recall that a subset F of a given metric space M is said to be dense in M if and only if every open ball in M intersects F. Thus, every open ball of radius $\epsilon > 0$ (no matter how small) and center $\boldsymbol{c}(z)$ (arbitrary) in $\mathbb{M}_{\boldsymbol{c}}[z]$,

$$B_\epsilon\left(\boldsymbol{c}(z)\right) = \left\{\boldsymbol{d}(z) \in \mathbb{M}_{\boldsymbol{c}}[z] \,:\, d\left(\boldsymbol{c}(z), \boldsymbol{d}(z)\right) < \epsilon\right\},$$

contains points of $\mathbb{F}_{\boldsymbol{c}}[z]$, that is, P-tuples of filters which satisfy condition A3. On the other hand, note that condition A3 assumes only the knowledge of an upper bound ($D_{\max}$) for the unknown column degrees $D_1, \ldots, D_P$. All said, the pre-filters can be easily (in the sense of theorem 2.2) selected off-line, without knowing $\boldsymbol{H}(z)$, to meet the spectral diversity assumption A3.

2.6 Closed-form identification algorithm (CFIA)

In this section, we solve the BCIP as formulated in section 2.5 (see page 28). More precisely, we present a closed-form (non-iterative) algorithm which, given the output correlation matrices $\boldsymbol{R}_{\boldsymbol{y}} = \{\boldsymbol{R}_{\boldsymbol{y}}[m] \,:\, m \in \mathbb{Z}\}$, determines the equivalence class $[\boldsymbol{H}(z)] = \varphi^{-1}(\boldsymbol{R}_{\boldsymbol{y}})$ of the underlying MIMO channel. Thus, the algorithm is a computational scheme which implements the inverse map $\varphi^{-1} : \operatorname{image}\varphi \subset \mathbb{C}_{\mathbb{Z}} \to \mathbb{H}[z]/\sim$, thereby solving the BCIP. Hereafter, we term this algorithm the closed-form identification algorithm (CFIA). The CFIA works in terms of the stacked observations $\boldsymbol{y}_L[n]$ in table 2.2, page 19. We assume that the choice for L makes the corresponding stacked channel coefficient matrix $\boldsymbol{\mathcal{H}}$ full column-rank. Note that, because we assumed the channel matrix $\boldsymbol{H}(z)$ to be tall, irreducible and column-reduced (condition A1), then $\boldsymbol{\mathcal{H}}$ is full column-rank for all L sufficiently large, say $L \geq L_0$. The value $L_0 = \operatorname{ord}\boldsymbol{H}(z)$ works, see [24], but this is unknown beforehand. Thus, a very conservative value is $L_0 = PD_{\max}$. The correlation matrices of $\boldsymbol{y}[n; L]$, that is, $\boldsymbol{R}_{\boldsymbol{y}}[m; L] = \mathrm{E}\left\{\boldsymbol{y}[n; L]\boldsymbol{y}[n-m; L]^H\right\}$, are completely determined by those of $\boldsymbol{y}[n]$. Indeed, we have

$$\boldsymbol{R}_{\boldsymbol{y}}[m; L] = \begin{bmatrix} \boldsymbol{R}_{\boldsymbol{y}}[m] & \boldsymbol{R}_{\boldsymbol{y}}[m+1] & \cdots & \boldsymbol{R}_{\boldsymbol{y}}[m+L] \\ \boldsymbol{R}_{\boldsymbol{y}}[m-1] & \boldsymbol{R}_{\boldsymbol{y}}[m] & \cdots & \boldsymbol{R}_{\boldsymbol{y}}[m+L-1] \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{R}_{\boldsymbol{y}}[m-L] & \boldsymbol{R}_{\boldsymbol{y}}[m-L+1] & \cdots & \boldsymbol{R}_{\boldsymbol{y}}[m] \end{bmatrix}. \tag{2.27}$$

The same applies to the random vector $\boldsymbol{w}[n\,;L]$: just replace $\boldsymbol{y}$ by $\boldsymbol{w}$ in (2.27). Since $\boldsymbol{R_y}[m]$ are given and $\boldsymbol{R_w}[m]$ are known (assumption A2), it follows that $\boldsymbol{R_y}[m;L]$ and $\boldsymbol{R_w}[m;L]$ are available, for any $m \in \mathbb{Z}$. We use them to define the (denoised) correlation matrices $\boldsymbol{R}[m] = \boldsymbol{R_y}[m;L] - \boldsymbol{R_w}[m;L]$. The latter satisfy the equality

$$\boldsymbol{R}[m] = \boldsymbol{\mathcal{H}}\, \boldsymbol{R_s}[m;\boldsymbol{\mathfrak{d}}]\, \boldsymbol{\mathcal{H}}^H. \tag{2.28}$$

Furthermore, the matrices $\boldsymbol{R_s}[m;\boldsymbol{e}]$ are also available, for arbitrary $\boldsymbol{e} = (E_1, E_2, \cdots, E_P)^T$. By equation (2.8), each of them is a block diagonal matrix with P blocks. The pth block, given in (2.21), depends only on the pth correlative filter. Since the pre-filters are chosen beforehand, the matrices $\boldsymbol{R_s}[m;\boldsymbol{e}]$ can thus be pre-stored (computed offline for arbitrary $\boldsymbol{e}$).

For later use, notice that $\boldsymbol{R_s}[0;\boldsymbol{e}]$ is a positive definite matrix, irrespective of $\boldsymbol{e}$. To verify this, note that the Sylvester matrix in (2.20) has always full row rank because $c_p[0] \neq 0$ (by definition). On the other hand, we have the identity $\boldsymbol{K}_n[0] = \boldsymbol{I}_n$, for arbitrary n. Using this identity in equation (2.21) gives

$$\boldsymbol{R_{s_p}}[0;E_p] = \boldsymbol{\mathcal{T}}_{E_p}\left(c_p(z)\right) \boldsymbol{\mathcal{T}}_{E_p}\left(c_p(z)\right)^H.$$

Since $\boldsymbol{\mathcal{T}}_{E_p}\left(c_p(z)\right)$ has full row rank, $\boldsymbol{R_s}[0;E_p]$ is positive definite (Gramian matrix). Finally, $\boldsymbol{R_s}[0;\boldsymbol{e}]$, being the diagonal concatenation of positive definite matrices, is itself a positive definite matrix.

CFIA: **step 1.** The CFIA involves three steps. The first step computes a matrix $\boldsymbol{\mathcal{G}}_0$ which will satisfy the equality

$$\boldsymbol{\mathcal{G}}_0 = \boldsymbol{\mathcal{H}}\, \boldsymbol{R_s}[0;\boldsymbol{\mathfrak{d}}]^{1/2}\, \boldsymbol{Q}^H, \tag{2.29}$$

where $\boldsymbol{Q}$ denotes a residual unknown unitary matrix. To accomplish this, we exploit the available denoised matrix $\boldsymbol{R}[0]$. Based on (2.28), we have $\boldsymbol{R}[0] = \boldsymbol{\mathcal{H}}\, \boldsymbol{R_s}[0;\boldsymbol{\mathfrak{d}}]\, \boldsymbol{\mathcal{H}}^H$. Defining $\boldsymbol{\mathcal{H}}_0 = \boldsymbol{\mathcal{H}}\, \boldsymbol{R_s}[0;\boldsymbol{\mathfrak{d}}]^{1/2}$, this can be rewritten as

$$\boldsymbol{R}[0] = \boldsymbol{\mathcal{H}}_0\, \boldsymbol{\mathcal{H}}_0^H. \tag{2.30}$$

By inspection, $\boldsymbol{R}[0]$ is positive semidefinite. Since $\boldsymbol{\mathcal{H}}$ is full column rank and $\boldsymbol{R_s}[0;\boldsymbol{\mathfrak{d}}]$ is positive definite (in particular, nonsingular), it follows that $\boldsymbol{\mathcal{H}}_0$ and $\boldsymbol{\mathcal{H}}$ have the same rank (the number of columns of $\boldsymbol{\mathcal{H}}$). Thus,

$$R = \operatorname{rank}\left(\boldsymbol{\mathcal{H}}\right) = \operatorname{rank}\left(\boldsymbol{\mathcal{H}}_0\right) = (L+1)P + \sum_{p=1}^{P} D_p. \tag{2.31}$$

To obtain $\boldsymbol{\mathcal{G}}_0$, we proceed as follows. Perform the eigenvalue decomposition (EVD)

$$\boldsymbol{R}[0] = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H, \tag{2.32}$$

where $\boldsymbol{U}$ denotes an unitary matrix, and $\boldsymbol{\Lambda} = \operatorname{diag}\left(\lambda_1, \ldots, \lambda_R, 0, \ldots, 0\right)$. Here, $\lambda_1 \geq \cdots \geq \lambda_R > 0$ (the nonzero eigenvalues are positive because $\boldsymbol{R}[0]$ is positive semidefinite). Thus, the EVD of $\boldsymbol{R}[0]$ reveals R, its rank. Once R is known, we use it to define the matrix $\boldsymbol{U}_1$ which contains the first R columns of $\boldsymbol{U}$, counting from the left. Moreover, define

$\boldsymbol{\Lambda}_1 = \mathrm{diag}(\lambda_1, \ldots, \lambda_R)$. Notice that both $\boldsymbol{U}_1$ and $\boldsymbol{\Lambda}_1$ are available from the EVD of $\boldsymbol{R}[0]$. With these definitions, equation (2.32) reads as

$$\boldsymbol{R}[0] = \boldsymbol{U}_1 \underbrace{\begin{bmatrix} \lambda_1 & & \\ & \ddots & \lambda_R \end{bmatrix}}_{\boldsymbol{\Lambda}_1} \boldsymbol{U}_1^H. \tag{2.33}$$

Thus, letting $\boldsymbol{\mathcal{G}}_0 = \boldsymbol{U}_1 \boldsymbol{\Lambda}_1^{1/2}$, we have

$$\boldsymbol{R}[0] = \boldsymbol{\mathcal{G}}_0 \, \boldsymbol{\mathcal{G}}_0^H. \tag{2.34}$$

Equations (2.30) and (2.34) assert that $\boldsymbol{\mathcal{H}}_0 \, \boldsymbol{\mathcal{H}}_0^H = \boldsymbol{\mathcal{G}}_0 \, \boldsymbol{\mathcal{G}}_0^H$, which, in turn, by trivial matrix algebra theory, implies that $\boldsymbol{\mathcal{G}}_0 = \boldsymbol{\mathcal{H}}_0 \, \boldsymbol{Q}^H$, for some unitary matrix $\boldsymbol{Q} : R \times R$.

CFIA: **step 2.** This step determines $\boldsymbol{\mathfrak{d}} = (\mathcal{D}_1, \mathcal{D}_2, \cdots, \mathcal{D}_P)^T$, $\mathcal{D}_p = D_p + L$, the vector of input degrees in the stacked data model (recall table 2.2), and the unknown unitary matrix $\boldsymbol{Q} \in \mathbb{U}(R)$ appearing in (2.29), modulo some phase ambiguities. Notice that although $\boldsymbol{\mathfrak{d}}$ is unknown, its entries must sum up to $R - P$. This follows from (2.31) and the fact $\mathcal{D}_p = D_p + L$. Thus, $\boldsymbol{\mathfrak{d}}$ belongs to the finite set

$$\mathbb{E} = \left\{ \boldsymbol{\mathfrak{e}} = (\mathcal{E}_1, \mathcal{E}_2, \cdots, \mathcal{E}_P)^T \in \mathbb{N}^P \, : \, \sum_{p=1}^{P} \mathcal{E}_p = R - P \right\}. \tag{2.35}$$

For future convenience, we partition $\boldsymbol{Q}$ into P submatrices,

$$\boldsymbol{Q} = \begin{bmatrix} \boldsymbol{Q}_1 & \boldsymbol{Q}_2 & \cdots & \boldsymbol{Q}_P \end{bmatrix}, \tag{2.36}$$

where $\boldsymbol{Q}_p : R \times (\mathcal{D}_p + 1)$. We start by using the available data, $\boldsymbol{\mathcal{G}}_0$ (from step 1) and $\boldsymbol{R}[m]$ (given), to define the matrices

$$\boldsymbol{\Upsilon}[m] = \boldsymbol{\mathcal{G}}_0^+ \, \boldsymbol{R}[m] \, \boldsymbol{\mathcal{G}}_0^{+H}. \tag{2.37}$$

We now relate the available matrices $\boldsymbol{\Upsilon}[m]$ with the unknowns $\boldsymbol{\mathfrak{d}}$ and $\boldsymbol{Q}$. By (2.29), we have $\boldsymbol{\mathcal{G}}_0^+ = \boldsymbol{Q} \, \boldsymbol{R}_{\boldsymbol{s}}[0; \boldsymbol{\mathfrak{d}}]^{-1/2} \, \boldsymbol{\mathcal{H}}^+$. Plugging this in (2.37) and recalling (2.28) gives

$$\boldsymbol{\Upsilon}[m] = \boldsymbol{Q} \, \boldsymbol{\Gamma}_{\boldsymbol{s}}[m; \boldsymbol{\mathfrak{d}}] \, \boldsymbol{Q}^H, \tag{2.38}$$

where $\boldsymbol{\Gamma}_{\boldsymbol{s}}[m; \boldsymbol{\mathfrak{d}}]$ follows the definition in (2.23). Equation (2.38) expresses the matrices $\boldsymbol{\Upsilon}[m]$, $m \in \mathbb{Z}$, as a function of both the vector of integers $\boldsymbol{\mathfrak{d}}$ and the unitary matrix $\boldsymbol{Q}$. The remarkable fact here is that this factorization is essentially unique in terms of $(\boldsymbol{\mathfrak{d}}, \boldsymbol{Q})$, as the second part of theorem 2.3 below shows. First, we need a definition. Let $\boldsymbol{\mathfrak{e}} = (\mathcal{E}_1, \mathcal{E}_2, \cdots, \mathcal{E}_p)^T$ denote a p-tuple of integers and $\boldsymbol{A}_1, \boldsymbol{A}_2$ two $n \times n$ matrices, where $n = p + \sum_{j=1}^{P} \mathcal{E}_j$. We say that $\boldsymbol{A}_1$ is similar to $\boldsymbol{A}_2$ with respect to $\boldsymbol{\mathfrak{e}}$, written $\boldsymbol{A}_1 \overset{\boldsymbol{\mathfrak{e}}}{\sim} \boldsymbol{A}_2$, if $\boldsymbol{A}_1 = \boldsymbol{A}_2 \, \boldsymbol{\Lambda}$ for some diagonal matrix $\boldsymbol{\Lambda} = \mathrm{diag}\left(\lambda_1 \boldsymbol{I}_{\mathcal{E}_1+1}, \lambda_2 \boldsymbol{I}_{\mathcal{E}_2+1}, \ldots, \lambda_p \boldsymbol{I}_{\mathcal{E}_p+1}\right)$, where $\lambda_j \in \mathbb{C}$, for $j = 1, 2, \ldots, p$. If, in addition, all λ_j are of the form $e^{i\,\theta_j}$, with $\theta_j \in \mathbb{R}$, then we write $\boldsymbol{A}_1 \overset{\boldsymbol{\mathfrak{e}}}{\approx} \boldsymbol{A}_2$.

Theorem 2.3. *Assume that conditions A1-A3 are fulfilled. Then,*

$$\boldsymbol{\Upsilon}[m]\,\boldsymbol{X} - \boldsymbol{X}\,\boldsymbol{\Gamma}_{\boldsymbol{s}}[m;\boldsymbol{\mathfrak{d}}] = \boldsymbol{0} \tag{2.39}$$

for all $m \in \mathbb{Z}$ *if and only if* $\boldsymbol{X} \overset{\boldsymbol{\mathfrak{d}}}{\sim} \boldsymbol{Q}$. *Moreover, let* $\boldsymbol{\mathfrak{e}} = (\mathcal{E}_1, \mathcal{E}_2, \cdots, \mathcal{E}_P)^T \in \mathbb{E}$ *and* $\boldsymbol{W} \in \mathbb{U}(R)$. *Then,*

$$\boldsymbol{\Upsilon}[m] = \boldsymbol{W}\,\boldsymbol{\Gamma}_{\boldsymbol{s}}[m;\boldsymbol{\mathfrak{e}}]\,\boldsymbol{W}^H, \quad \textit{for all } m \in \mathbb{Z}, \tag{2.40}$$

if and only if $\boldsymbol{\mathfrak{e}} = \boldsymbol{\mathfrak{d}}$ *and* $\boldsymbol{W} \overset{\boldsymbol{\mathfrak{d}}}{\approx} \boldsymbol{Q}$.

Proof: See appendix A.

We exploit both uniqueness results of theorem 2.3 as the basis of our strategy for determining $\boldsymbol{\mathfrak{d}}$ and $\boldsymbol{Q}$. First, notice that, since the pre-filters $c_p(z)$ and the MIMO subchannels $h_{pq}(z)$ have finite memory, the normalized correlation matrices $\boldsymbol{\Gamma}_{\boldsymbol{s}}[m;\boldsymbol{\mathfrak{e}}]$ are zero for $|m|$ sufficiently large. Thus, theorem 2.3 still holds if one replaces the condition $m \in \mathbb{Z}$ in (2.40) by, say, $m = \pm 1, \pm 2, \ldots, \pm M$, where M is chosen sufficiently large. A conservative value is, for example, $M = R - P + C_{\max} + 1$. With this data, we define the function $\chi : \mathbb{E} \times \mathbb{U}(R) \to \mathbb{R}$,

$$\chi\left(\boldsymbol{\mathfrak{e}}, \boldsymbol{W}\right) = \sum_{|m|=1}^{M} \left\| \boldsymbol{\Upsilon}[m] - \boldsymbol{W}\,\boldsymbol{\Gamma}_{\boldsymbol{s}}[m;\boldsymbol{\mathfrak{e}}]\,\boldsymbol{W}^H \right\|^2. \tag{2.41}$$

For a pair $(\boldsymbol{\mathfrak{e}}, \boldsymbol{W})$, the value $\chi\left(\boldsymbol{\mathfrak{e}}, \boldsymbol{W}\right)$ measures the mismatch of the factorization (2.40). Theorem 2.3 asserts that $\chi\left(\boldsymbol{\mathfrak{e}}, \boldsymbol{W}\right) = 0$ if and only if $\boldsymbol{\mathfrak{e}} = \boldsymbol{\mathfrak{d}}$ and $\boldsymbol{W} \overset{\boldsymbol{\mathfrak{d}}}{\approx} \boldsymbol{Q}$. Now, suppose $\boldsymbol{\mathcal{W}} : \mathbb{E} \to \mathbb{U}\left(R\right)$ is a map satisfying $\chi\left(\boldsymbol{\mathfrak{d}}, \boldsymbol{\mathcal{W}}(\boldsymbol{\mathfrak{d}})\right) = 0$, and define $\phi : \mathbb{E} \to \mathbb{R}$ by $\phi(\boldsymbol{\mathfrak{e}}) = \chi\left(\boldsymbol{\mathfrak{e}}, \boldsymbol{\mathcal{W}}(\boldsymbol{\mathfrak{e}})\right)$. Then, $\boldsymbol{\mathfrak{d}}$ is the unique minimizer of this non-negative function ϕ over $\mathbb{E}$, because $\phi(\boldsymbol{\mathfrak{d}}) = 0$ (by the hypothesis on the map $\boldsymbol{\mathcal{W}}$) and, as soon as $\boldsymbol{\mathfrak{e}} \neq \boldsymbol{\mathfrak{d}}$, $\phi\left(\boldsymbol{\mathfrak{e}}\right) > 0$ by theorem 2.3. Thus, $\boldsymbol{\mathfrak{d}}$ can be found as

$$\boldsymbol{\mathfrak{d}} = \arg \min_{\boldsymbol{\mathfrak{e}} \in \mathbb{E}} \phi\left(\boldsymbol{\mathfrak{e}}\right). \tag{2.42}$$

Moreover, $\boldsymbol{\mathcal{W}}(\boldsymbol{\mathfrak{d}}) \overset{\boldsymbol{\mathfrak{d}}}{\approx} \boldsymbol{Q}$, that is, $\boldsymbol{\mathcal{W}}(\boldsymbol{\mathfrak{d}})$ would reveal $\boldsymbol{Q}$ up to phase ambiguities.

It remains to exhibit the map $\boldsymbol{\mathcal{W}}$. It must satisfy the equation $\chi\left(\boldsymbol{\mathfrak{d}}, \boldsymbol{\mathcal{W}}(\boldsymbol{\mathfrak{d}})\right) = 0$, that is,

$$\boldsymbol{\Upsilon}[m] = \boldsymbol{\mathcal{W}}\left(\boldsymbol{\mathfrak{d}}\right)\,\boldsymbol{\Gamma}_{\boldsymbol{s}}[m;\boldsymbol{\mathfrak{d}}]\,\boldsymbol{\mathcal{W}}\left(\boldsymbol{\mathfrak{d}}\right)^H \quad \Leftrightarrow \quad \boldsymbol{\Upsilon}[m]\,\boldsymbol{\mathcal{W}}(\boldsymbol{\mathfrak{d}}) - \boldsymbol{\mathcal{W}}\left(\boldsymbol{\mathfrak{d}}\right)\boldsymbol{\Gamma}_{\boldsymbol{s}}[m;\boldsymbol{\mathfrak{d}}] = \boldsymbol{0}, \tag{2.43}$$

for $m = \pm 1, \pm 2, \ldots, \pm M$. Here, we used the fact that $\boldsymbol{\mathcal{W}}\left(\boldsymbol{\mathfrak{d}}\right)$ is unitary ($\boldsymbol{\mathcal{W}}\left(\boldsymbol{\mathfrak{d}}\right)^H \boldsymbol{\mathcal{W}}\left(\boldsymbol{\mathfrak{d}}\right) = \boldsymbol{I}_R$). Writing $\boldsymbol{\mathcal{W}}\left(\boldsymbol{\mathfrak{d}}\right) = [\,\boldsymbol{\mathcal{W}}_1\left(\boldsymbol{\mathfrak{d}}\right)\,\boldsymbol{\mathcal{W}}_2\left(\boldsymbol{\mathfrak{d}}\right)\,\cdots\,\boldsymbol{\mathcal{W}}_P\left(\boldsymbol{\mathfrak{d}}\right)\,]$, $\boldsymbol{\mathcal{W}}_p\left(\boldsymbol{\mathfrak{d}}\right) : R \times (\mathcal{D}_p + 1)$, and using the block diagonal structure of $\boldsymbol{\Gamma}_{\boldsymbol{s}}[m;\boldsymbol{\mathfrak{d}}]$ (see (2.24)), the equality in the right hand side of (2.43) decouples into P equalities,

$$\boldsymbol{\Upsilon}[m]\,\boldsymbol{\mathcal{W}}_p\left(\boldsymbol{\mathfrak{d}}\right) - \boldsymbol{\mathcal{W}}_p\left(\boldsymbol{\mathfrak{d}}\right)\,\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m;\mathcal{D}_p] = 0, \tag{2.44}$$

for $p = 1, 2, \ldots, P$. Thus, the pth submatrix $\boldsymbol{\mathcal{W}}_p\left(\boldsymbol{\mathfrak{d}}\right)$ of $\boldsymbol{\mathcal{W}}\left(\boldsymbol{\mathfrak{d}}\right)$ must satisfy (2.44) for $m = \pm 1, \pm 2, \ldots, \pm M$. Since $\boldsymbol{\Upsilon}[-m] = \boldsymbol{\Upsilon}[m]^H$ and $\boldsymbol{\Gamma}[-m;\boldsymbol{\mathfrak{e}}] = \boldsymbol{\Gamma}[m;\boldsymbol{\mathfrak{e}}]^H$, we conclude

that each $\boldsymbol{\mathcal{W}}_p$ must be a solution of the linear system

$$\left\{\begin{array}{lll} \boldsymbol{\Upsilon}[1] \quad \boldsymbol{X} - \boldsymbol{X}\,\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[1;\mathcal{D}_p] & = & \boldsymbol{0} \\ \boldsymbol{\Upsilon}[1]^H \; \boldsymbol{X} - \boldsymbol{X}\,\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[1;\mathcal{D}_p]^H & = & \boldsymbol{0} \\ \qquad\qquad \vdots & & \\ \boldsymbol{\Upsilon}[M] \quad \boldsymbol{X} - \boldsymbol{X}\,\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[M;\mathcal{D}_p] & = & \boldsymbol{0} \\ \boldsymbol{\Upsilon}[M]^H \; \boldsymbol{X} - \boldsymbol{X}\,\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[M;\mathcal{D}_p]^H & = & \boldsymbol{0} \end{array}\right. , \tag{2.45}$$

which containts $2M$ homogeneous matricial linear equations in $\boldsymbol{X} : R\times(\mathcal{D}_p+1)$. According to theorem 2.3, each solution $\boldsymbol{X}$ of (2.45) satisfies $\boldsymbol{X} = \lambda\boldsymbol{Q}_p$, for some $\lambda \in \mathbb{C}$. Here, $\boldsymbol{Q}_p$ denotes the pth submatrix of $\boldsymbol{Q}$, see (2.36). A nonzero solution of (2.45) can be computed as follows. Successively using the identity $\operatorname{vec}(\boldsymbol{ABC}) = \left(\boldsymbol{C}^T \otimes \boldsymbol{A}\right)\operatorname{vec}(\boldsymbol{B})$, for conformable matrices $\boldsymbol{A}, \boldsymbol{B}$ and $\boldsymbol{C}$ [39, page 30], on each matricial equation of (2.45) and stacking the results, we obtain $\boldsymbol{T}_p\,\boldsymbol{x} = \boldsymbol{0}$, where $\boldsymbol{x} = \operatorname{vec}(\boldsymbol{X})$ and

$$\boldsymbol{T}_p = \left[\begin{array}{c} \boldsymbol{T}_p[1] \\ \boldsymbol{T}_p[2] \\ \vdots \\ \boldsymbol{T}_p[M] \end{array}\right], \quad \boldsymbol{T}_p[m] = \left[\begin{array}{l} \boldsymbol{I}_{\mathcal{D}_p+1} \otimes \boldsymbol{\Upsilon}[m] - \boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m;\mathcal{D}_p]^T \otimes \boldsymbol{I}_R \\ \\ \boldsymbol{I}_{\mathcal{D}_p+1} \otimes \boldsymbol{\Upsilon}[m]^H - \overline{\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m;\mathcal{D}_p]} \otimes \boldsymbol{I}_R \end{array}\right]. \tag{2.46}$$

Thus, $\boldsymbol{x} \in \ker(\boldsymbol{T}_p) = \ker\left(\boldsymbol{T}_p^H\boldsymbol{T}_p\right)$. Let $\boldsymbol{S}_p = \boldsymbol{T}_p^H\boldsymbol{T}_p$ and $\boldsymbol{S}_p = \boldsymbol{U}\boldsymbol{\Lambda}\boldsymbol{U}^H$, be an EVD of $\boldsymbol{S}_p$, with the entries of the non-negative diagonal matrix $\boldsymbol{\Lambda} = \operatorname{diag}\left(\lambda_1, \ldots, \lambda_{R(\mathcal{D}_p+1)}\right)$ sorted in decreasing order. Then, $\lambda_{R(\mathcal{D}_p+1)} = 0$, and $\boldsymbol{x}$ can be set as the last column (counting from the left) of $\boldsymbol{U}$, say, $\boldsymbol{u}_{R(\mathcal{D}_p+1)}$. Reshaping this unit-norm vector into a $R \times (\mathcal{D}_p + 1)$ matrix, $\boldsymbol{U}_p = \operatorname{vec}^{-1}\left(\boldsymbol{u}_{R(\mathcal{D}_p+1)}\right)$, and scaling by $\left\|\boldsymbol{Q}_p\right\|$, that is, defining $\boldsymbol{\mathcal{X}}_p = \sqrt{\mathcal{D}_p+1}\,\boldsymbol{U}_p$, yields a solution of (2.45) with the same norm as $\boldsymbol{Q}_p$. Thus, this solution satisfies $\boldsymbol{\mathcal{X}}_p = e^{i\theta_p}\,\boldsymbol{Q}_p$, for some $\theta_p \in \mathbb{R}$, and, as a consequence, $\boldsymbol{\mathcal{X}} = [\,\boldsymbol{\mathcal{X}}_1\,\boldsymbol{\mathcal{X}}_2\,\cdots\,\boldsymbol{\mathcal{X}}_P\,] \overset{\mathfrak{d}}{\approx} \boldsymbol{Q}$.

In conclusion, if we knew $\mathfrak{d} = (\mathcal{D}_1, \mathcal{D}_2, \cdots, \mathcal{D}_P)^T$, then we could compute a matrix $\boldsymbol{\mathcal{X}} \overset{\mathfrak{d}}{\approx} \boldsymbol{Q}$ as described above: it suffices to take $\boldsymbol{\mathcal{X}}_p : R \times (\mathcal{D}_p + 1)$, the pth submatrix of $\boldsymbol{\mathcal{X}}$, to be $\sqrt{\mathcal{D}_p+1}$ times vec^{-1} of the unit-norm eigenvector associated with the smallest eigenvalue of $\boldsymbol{S}_p = \boldsymbol{T}_p^H\boldsymbol{T}_p$, with $\boldsymbol{T}_p$ as in (2.46). Now, viewing $\mathfrak{d}$ as a generic point in $\mathbb{E}$, this scheme gives a map from $\mathbb{E}$ to $\mathbb{C}^{R\times R}$. We let $\mathfrak{e} \in \mathbb{E} \mapsto \boldsymbol{\mathcal{X}}(\mathfrak{e}) \in \mathbb{C}^{R\times R}$ denote this map. As just seen, $\boldsymbol{\mathcal{X}}(\mathfrak{d})$ is unitary and is similar to $\boldsymbol{Q}$ with respect to $\mathfrak{d}$. For $\mathfrak{e} \neq \mathfrak{d}$, the matrix $\boldsymbol{\mathcal{X}}(\mathfrak{e})$ is not necessarily unitary. We define $\boldsymbol{\mathcal{W}}(\mathfrak{e})$ to be the projection of $\boldsymbol{\mathcal{X}}(\mathfrak{e})$ onto the unitary group $\mathbb{U}(R)$. That is, we define $\boldsymbol{\mathcal{W}}(\mathfrak{e}) = \boldsymbol{\Pi}(\boldsymbol{\mathcal{X}}(\mathfrak{e}))$, where $\boldsymbol{\Pi} : \mathbb{C}^{R\times R} \to \mathbb{U}(R)$ denotes the projection onto $\mathbb{U}(R)$. This nonlinear projector may be computed as follows. Let $\boldsymbol{X} \in \mathbb{C}^{R\times R}$ be given and perform the singular value decomposition (SVD) $\boldsymbol{X} = \boldsymbol{U}\,\boldsymbol{\Sigma}\,\boldsymbol{V}^H$, where $\boldsymbol{U}, \boldsymbol{V} \in \mathbb{U}(R)$ and $\boldsymbol{\Sigma}$ (diagonal) contains the singular values of $\boldsymbol{X}$. Then $\boldsymbol{\Pi}(\boldsymbol{X}) = \boldsymbol{U}\boldsymbol{V}^H$ is the unitary factor in the polar form of $\boldsymbol{X}$ [28, page 412], and minimizes the distance from $\boldsymbol{X}$ to $\mathbb{U}(R)$,

$$\|\boldsymbol{X} - \boldsymbol{\Pi}(\boldsymbol{X})\| = \min_{\boldsymbol{W}\,\in\,\mathbb{U}(R)} \|\boldsymbol{X} - \boldsymbol{W}\|.$$

That is, it is the solution of a special Procrustes problem [28, page 431]. Since $\boldsymbol{\mathcal{X}}(\mathfrak{d})$ is already unitary, $\boldsymbol{\mathcal{W}}(\mathfrak{d}) = \boldsymbol{\mathcal{X}}(\mathfrak{d}) \overset{\mathfrak{d}}{\approx} \boldsymbol{Q}$ and $\chi(\mathfrak{d}, \boldsymbol{\mathcal{W}}(\mathfrak{d})) = 0$ as required.

Some remarks are in order before proceeding to the next step of the CFIA. **i)** Even for moderate values of inputs P, the cardinality of the finite set $\mathbb{E}$ in (2.35) can be quite high. Thus, minimization of ϕ in (2.42) by enumeration may jeopardize this approach for real-time applications. Fortunately, in many communication scenarios, previous field studies are available which determine the typical profile of the propagation space-time channels, that is, good estimates of the column degrees D_p (hence, $\mathcal{D}_p$) of the channel matrix $\boldsymbol{H}(z)$ are available beforehand. This knowledge can be exploited to minimize ϕ only over a highly restricted domain (the vicinity of the typical channel orders). **ii)** As in most subspace-based approaches [61, 43, 24], the performance of our technique depends strongly on the accuracy of the estimated channel orders. In our case, this is connected with the ability of correctly detecting the zero of ϕ in (2.42) over $\mathbb{E}$, using only finite-length data packets. That is, when the CFIA operates on estimated correlation matrices, say,

$$\widehat{\boldsymbol{R}}_{\boldsymbol{y}}[m] = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{y}[n]\boldsymbol{y}[n-m]^H,$$

where $\{\boldsymbol{y}[n] : n = -M, \ldots, N\}$ denotes the available packet of observations, rather than on exact (theoretical) correlation matrices $\boldsymbol{R}_{\boldsymbol{y}}[m]$. In order to increase the robustness of this detection step, the matching cost proposed in (2.41) should be computed with an appropriate weighting matrix rather than the identity matrix. The optimal weighting matrix could be obtained on the basis of a more detailed theoretical study. Another possible approach is to assess theoretically the impact of the correlative filters on ϕ, in order to obtain an optimal design to minimize the probability of detection error. These issues are not pursued further in this thesis. See however the performance analysis in chapter 3, which may be used to design the correlative filters in order to minimize the estimation error, that is, the mean-square distance between true and estimated channel equivalence class, under the assumption that the column degrees D_p are known.

CFIA: **step 3.** This step determines $\boldsymbol{H}(z)$ up to a phase offset per column, that is, the equivalence class $[\boldsymbol{H}(z)] \in \mathbb{H}[z]/\sim$. We use the matrix $\boldsymbol{\mathcal{H}}_0$ in (2.29) and the pair $(\mathfrak{d}, \boldsymbol{\mathcal{W}}(\mathfrak{d}))$, which are available from the first and second steps, respectively. Recall that $\boldsymbol{\mathcal{W}}(\mathfrak{d}) = \boldsymbol{Q}\,\boldsymbol{\Lambda}$, where $\boldsymbol{\Lambda} = \operatorname{diag}\left(e^{i\,\theta_1}\boldsymbol{I}_{\mathcal{D}_1+1}, e^{i\,\theta_2}\boldsymbol{I}_{\mathcal{D}_2+1}, \ldots, e^{i\,\theta_P}\boldsymbol{I}_{\mathcal{D}_P+1}\right)$, for some $\theta_p \in \mathbb{R}$. We define

$$\boldsymbol{\mathcal{G}} = \boldsymbol{\mathcal{H}}_0\,\boldsymbol{\mathcal{W}}(\mathfrak{d})\;\boldsymbol{R}_{\boldsymbol{s}}[0;\mathfrak{d}]^{-1/2} = \boldsymbol{\mathcal{H}}\,\boldsymbol{\Lambda}. \tag{2.47}$$

Here, we used the fact that $\boldsymbol{R}_{\boldsymbol{s}}[0;\mathfrak{d}]^{1/2}$ commutes with $\boldsymbol{\Lambda}$, since they share the same block diagonal structure, the pth block of $\boldsymbol{\Lambda}$ being $e^{i\,\theta_p}\boldsymbol{I}_{\mathcal{D}_p+1}$. Equation (2.47) means that if one parses $\boldsymbol{\mathcal{G}} = [\,\boldsymbol{\mathcal{G}}_1\,\boldsymbol{\mathcal{G}}_2\,\cdots\,\boldsymbol{\mathcal{G}}_P\,]$ accordingly to $\boldsymbol{\mathcal{H}}$ in table 2.2, then $\boldsymbol{\mathcal{G}}_p = e^{i\,\theta_p}\,\boldsymbol{\mathcal{H}}_p$, or,

$$\boldsymbol{\mathcal{G}}_p = e^{i\,\theta_p}\begin{bmatrix} \boldsymbol{h}_p[0] & \cdots & \boldsymbol{h}_p[D_p] & & \\ & \boldsymbol{h}_p[0] & \cdots & \boldsymbol{h}_p[D_p] & \\ & & \ddots & & \ddots \\ & & & \boldsymbol{h}_p[0] & \cdots & \boldsymbol{h}_p[D_p] \end{bmatrix}.$$

Note that $\boldsymbol{d} = (D_1, D_2, \cdots, D_P)^T$ is obtained from $\mathfrak{d} = (\mathcal{D}_1, \mathcal{D}_2, \cdots, \mathcal{D}_P)^T$, as $D_p = \mathcal{D}_p - L$. Thus, the coefficients of the pth filter $\boldsymbol{h}_p(z) = \sum_{d=0}^{D_p}\boldsymbol{h}_p[d]z^{-d}$ may be read out directly (modulo a phase offset) from $\boldsymbol{\mathcal{G}}_p$, or, for improved estimation, by averaging

the $L+1$ copies available in $\boldsymbol{\mathcal{G}}_p$ yielding $\boldsymbol{G}_p = e^{i\theta_p}\boldsymbol{H}_p$. Defining $\boldsymbol{G} = [\,\boldsymbol{G}_1\,\boldsymbol{G}_2\,\cdots\,\boldsymbol{G}_P\,]$ and letting $\boldsymbol{G}(z)$ be the channel matrix which satisfies $\boldsymbol{G}(z) \simeq (\boldsymbol{d};\boldsymbol{G})$, it follows that $[\boldsymbol{G}(z)] = [\boldsymbol{H}(z)]$.

2.7 Iterative identification algorithm (IIA)

After the channel matrix $\boldsymbol{H}(z)$ is identified, we face the problem of detecting the unfiltered information-bearing sequences $a_p[n]$, recall figure 2.5, from the packet of available observations, say, $\{\boldsymbol{y}[n] : n = 1, 2, \ldots, N\}$. In the sequel, we assume the common digital communication scenario, that is, the pth data sequence $a_p[n]$ consists of iid symbols drawn from a given finite modulation alphabet $\mathbb{A}_p \subset \mathbb{C}$. Furthermore, for simplicity, we assume that $\boldsymbol{w}[n]$ denotes white spatio-temporal Gaussian noise,

$$\boldsymbol{R}_{\boldsymbol{w}}[m] = \mathrm{E}\left\{\boldsymbol{w}[n]\boldsymbol{w}[n-m]^H\right\} = \sigma^2\boldsymbol{I}_Q\delta[m]. \tag{2.48}$$

In the present context, the optimal maximum likelihood (ML) detector leads to a generalized (multi-user) maximum likelihood sequence estimation (MLSE) Viterbi algorithm [15]. However, the computational cost of this approach is very high, due to the trellis size involved. Here, we pursue a computationally simpler, yet sub-optimal technique, to detect the data symbols $a_p[n]$ from the samples $\boldsymbol{y}[n]$. The proposed technique exploits the fact that the correlative filters are minimum-phase and permits to handle carrier frequency asynchronisms. In the context of SDMA systems, these may be induced by, for example, Doppler effects due to the relative motion between the mobile sources and the base station antenna array. This distortion induces a baseband rotation in the received signals. We have a data model similar to (2.3), except for the inclusion of the residual phase drifts,

$$\boldsymbol{y}[n] = \sum_{p=1}^{P} \boldsymbol{h}_p(z) \circledast \widetilde{s}_p[n] + \boldsymbol{w}[n] \tag{2.49}$$

where

$$\widetilde{s}_p[n] = e^{i\omega_p n} s_p[n], \tag{2.50}$$

and ω_p denotes the baseband rotation frequency corresponding to the pth user. Although the filters $\boldsymbol{h}_p(z)$ in (2.49) are not exactly the same as in (2.3) because, in all rigor, some phase offset corrections are needed, we maintain the notation for the sake of clarity.

Each iteration of the proposed iterative procedure consists of two steps. In the first step, the data symbols $a_p[n]$ are detected, given the current estimate of the channel $\boldsymbol{H}(z)$. In the second step, the channel matrix $\boldsymbol{H}(z)$ is re-evaluated on the basis of the newly estimated data symbols $a_p[n]$. This resembles, in spirit, the methodology of the ILSP and ILSE approaches in [56, 57, 58]. The added difficulty here is that the data symbols are pre-filtered and distorted by baseband rotations. Hereafter, we refer to our proposed iterative algorithm as the iterative identification algorithm (IIA). We now discuss the two steps, in each iteration of the IIA, in more detail.

IIA**: step 1.** We are at the $(k+1)$th iteration cycle. Let

$$\boldsymbol{H}^{(k)}(z) = \left[\begin{array}{cccc} \boldsymbol{h}_1^{(k)}(z) & \boldsymbol{h}_2^{(k)}(z) & \cdots & \boldsymbol{h}_P^{(k)}(z) \end{array}\right]$$

denote the estimate of the MIMO channel matrix $\boldsymbol{H}(z)$ obtained from the previous iteration cycle. The algorithm is initialized with $\boldsymbol{H}^{(0)}(z)$, the closed-form solution furnished by the CFIA presented in section 2.6, and $k = 0$. We reason as if the current channel estimate is exact, that is, $\boldsymbol{H}^{(k)}(z) = \boldsymbol{H}(z)$. Focus on the pth user. First, we extract the baseband rotated sequence

$$\widetilde{s}_p[n] = e^{i\omega_p n} s_p[n]$$

from the observations $\boldsymbol{y}[n]$. To accomplish this, we employ a zero-forcing row polynomial filter

$$\boldsymbol{f}_p(z) = \sum_{d=0}^{F_p} \boldsymbol{f}_p[d] z^{-d}, \tag{2.51}$$

where $\boldsymbol{f}_p[d] \in \mathbb{C}^{1\times Q}$, satisfying

$$\boldsymbol{f}_p(z)\boldsymbol{H}^{(k)}(z) = \left[\begin{array}{ccccccc} 0 & \cdots & 0 & z^{-d_p} & 0 & \cdots & 0 \end{array}\right], \tag{2.52}$$

for some non-negative delay d_p in the pth entry. That is, $\boldsymbol{f}_p(z)$ exactly nulls the inter-symbol and co-channel interferences affecting the user p. Notice that the existence of such a zero-forcing filter is guaranteed by the irreducibility of the channel matrix in assumption A1, if the degree F_p is taken sufficiently high. The coefficients of the filter, arranged in the row vector,

$$\boldsymbol{f}_p = \left[\begin{array}{cccc} \boldsymbol{f}_p[0] & \boldsymbol{f}_p[1] & \cdots & \boldsymbol{f}_p[F_p] \end{array}\right],$$

correspond to the $((p-1)(1+F_p) + D_1 + \cdots + D_{p-1} + d_p + 1)$th row of the pseudo-inverse of the matrix

$$\boldsymbol{\mathcal{H}}^{(k)} = \left[\begin{array}{ccccc} \boldsymbol{\mathcal{T}}_{F_p}\left(\boldsymbol{h}_1^{(k)}\right) & \boldsymbol{\mathcal{T}}_{F_p}\left(\boldsymbol{h}_2^{(k)}\right) & \cdots & \boldsymbol{\mathcal{T}}_{F_p}\left(\boldsymbol{h}_P^{(k)}\right) \end{array}\right].$$

Apply the filter $f_p(z)$ to the observations $\boldsymbol{y}[n + d_p]$ (the delay d_p is introduced for notational convenience), and denote the resulting scalar sequence by $\alpha_p[n]$. That is,

$$\begin{aligned} \alpha_p[n] &= \boldsymbol{f}_p(z) \circledast \boldsymbol{y}[n + d_p] \\ &= e^{i\omega_p n} \left(\sum_{d=0}^{C_p} c_p[d] a_p[n-d] \right) + n_p[n], \end{aligned} \tag{2.53}$$

where

$$n_p[n] = \boldsymbol{f}_p(z) \circledast \boldsymbol{w}[n + d_p].$$

We have to detect the data symbols $a_p[n] \in \mathbb{A}$ from the samples $\alpha_p[n]$ in (2.53). The two main obstacles are: **i)** the presence of the correlative filter $c_p(z)$, and **ii)** the baseband rotation $e^{i\omega_p n}$. We address each distortion separately.

i) First we get rid of the correlative filter. Rewrite (2.53) as

$$\alpha_p[n] = \sum_{d=0}^{C_p} \widetilde{c}_p[d] \widetilde{a}_p[n-d] + n_p[n],$$

where $\widetilde{c}_p[d] = c_p[d]e^{i\omega_p d}$ and $\widetilde{a}_p[n] = a_p[n]e^{i\omega_p n}$ denote "rotated" versions of the filter $c_p(z)$ and the information-bearing signal $a_p[n]$, respectively. Now, for typical values of ω_p and

pre-filter degrees, say, $\omega_p = 2\pi/1000$ and $C_p = 5$, we have $\tilde{c}_p[d] \simeq c_p[d]$, since $e^{i\omega_p d} \simeq 1$ for small integers d. Therefore, within this approximation,

$$\alpha_p[n] = c_p(z) \circledast \widetilde{a}_p[n] + n_p[n].$$

Since the correlative filter $c_p(z)$ is minimum-phase, we use its stable inverse, denote $d_p(z)$, to recover the rotated signal $\widetilde{a}_p[n]$,

$$\beta_p[n] = d_p(z) \circledast \alpha_p[n] = \underbrace{e^{i\omega_p n} a_p[n]}_{\tilde{a}_p[n]} + u_p[n], \tag{2.54}$$

where $u_p[n] = d_p(z) \circledast n_p[n]$.

ii) Now, we handle the baseband rotation. Split the available samples $\beta_p[n]$ in B consecutive blocks of equal size T. Making T small enough, we have, within each block $b = 1, 2, \ldots, B$, the approximation

$$\beta_p[n] = e^{i\theta_p[b]} a_p[n] + u_p[n], \tag{2.55}$$

for some phase $\theta_p[b] \in \mathbb{R}$.

We process the samples $\beta_p[n]$ block by block. Within each block, we jointly detect the symbols $a_p[n]$ and the corresponding phase offset $\theta_p[b]$. Assume we are processing the bth block. This scheme starts with $\theta_p[0] = 0$ and $b = 1$. We use the estimate of the phase offset in the previous block, and the fact that the phase varies smoothly between adjacent blocks, to make the approximation $\theta_p[b] \approx \theta_p[b-1]$, from which we obtain almost phase-corrected symbols,

$$\gamma_p[n] = e^{-i\theta_p[b-1]} \beta_p[n] \approx a_p[n] + v_p[n],$$

where $v_p[n] = e^{-i\theta_p[b-1]} u_p[n]$ denotes circular complex Gaussian noise. The data symbol $a_p[n]$ is detected by projecting the soft estimate $\gamma_p[n]$ onto the finite modulation alphabet $\mathbb{A} \subset \mathbb{C}$. That is, a least-squares (LS) criterion is adopted, leading to

$$\widehat{a}_p[n] = \underset{a \in \mathbb{A}_p}{\arg\min} \; |\gamma_p[n] - a|^2 .$$

Now, we turn to the problem of refining the estimate of $\theta_p[b]$ in (2.55), using $a_p[n] = \widehat{a}_p[n]$. Again, we follow a LS model-fitting approach,

$$\widehat{\theta}_p[b] = \underset{\theta \in \mathbb{R}}{\arg\min} \; \sum_n \left| \beta_p[n] - e^{i\theta} a_p[n] \right|^2, \tag{2.56}$$

where it is understood that the summation in (2.56) only involves those time-indexes n contained in the bth block. It is easily seen that the solution of (2.56) is characterized by the necessary and sufficient optimality condition

$$e^{i\widehat{\theta}_p[b]} \left(\sum_n \overline{\beta}_p[n] a_p[n] \right) \geq 0,$$

which yields a simple computational scheme for retrieving $\widehat{\theta}_p[b]$.

It should be noticed that the estimation of ω_p in (2.54) may also be efficiently solved by exploiting the fact that ω_p is a conjugate cyclic frequency of the signal $\beta_p[n]$. The main advantage of the proposed methodology is that it permits to handle more generic phase drifts, that is, phase distortions of the form $e^{i\theta_p[n]}$ where the time-varying phase signal $\theta_p[n]$ does not necessarily follows a linear dynamic such as $\theta_p[n] = \omega_p n$.

IIA: step 2. Take the detected symbols $a_p[n]$ and phase offsets $\{\theta_p[b] : b = 1, 2, \ldots, B\}$ in step A and reconstruct the signal $\widetilde{s}_p[n]$ in (2.50) as

$$\widetilde{s}_p[n] = e^{i\theta_p[b]}\, c_p(z) \circledast a_p[n],$$

whenever the time index n falls in the bth block. Rewrite (2.49) in matrix form

$$\boldsymbol{Y} = \left[\begin{array}{cccc} \boldsymbol{y}[1] & \boldsymbol{y}[2] & \cdots & \boldsymbol{y}[N] \end{array}\right] = \boldsymbol{H}\,\widetilde{\boldsymbol{S}} + \boldsymbol{W}, \tag{2.57}$$

where

$$\widetilde{\boldsymbol{S}} = \left[\begin{array}{cccc} \widetilde{\boldsymbol{s}}[1] & \widetilde{\boldsymbol{s}}[2] & \cdots & \widetilde{\boldsymbol{s}}[N] \end{array}\right],$$

with $\widetilde{\boldsymbol{s}}[n] = \widetilde{\boldsymbol{s}}[n; \boldsymbol{d}]$, and

$$\boldsymbol{W} = \left[\begin{array}{cccc} \boldsymbol{w}[1] & \boldsymbol{w}[2] & \cdots & \boldsymbol{w}[N] \end{array}\right].$$

We now re-evaluate the coefficient channel matrix $\boldsymbol{H}$ in (2.57) by adopting a LS performance criterion, that is,

$$\boldsymbol{H}^{(k+1)} = \underset{\boldsymbol{G} \in \mathbb{C}^{Q \times D}}{\arg\min} \left\| \boldsymbol{Y} - \boldsymbol{G}\widetilde{\boldsymbol{S}} \right\|^2.$$

This yields

$$\boldsymbol{H}^{(k+1)} = \boldsymbol{Y}\,\widetilde{\boldsymbol{S}}^{+}.$$

Now, if $\boldsymbol{H}^{(k+1)} \simeq \boldsymbol{H}^{(k)}$, then stop the iterations. Else, set $k = k+1$ and return to step A.

2.8 Computer simulations

We present two sets of simulations. In the first set, we consider $P = 2$ users, without carrier misadjustments. The performance of the proposed blind channel identification technique is evaluated in terms of the mean-square error (MSE) of the MIMO channel estimate. For separation of the sources and the equalization step, the performance criterion is the symbol error rate (SER) of the estimated data symbols. In the second set of simulations, we consider $P = 3$ users with residual phase drifts. We compare our technique with the TICC approach [12], both in terms of the MSE of the channel estimate and the SER of the resulting symbol detection scheme.

Scenario with two users. We consider $P = 2$ users, with distinct digital modulation formats. User 1 employs the quaternary amplitude modulation (QAM) digital format normalized to unit-power,

$$\mathbb{A}_1 = \mathbb{A}_{\mathrm{QAM}} = \left\{ \pm\frac{1}{\sqrt{2}} \pm i\frac{1}{\sqrt{2}} \right\},$$

while user 2 employs the binary phase keying modulation format (BPSK),

$$\mathbb{A}_2 = \mathbb{A}_{\text{BSK}} = \{\pm 1\}\,.$$

Both users pass their iid symbol information sequences $a_p[n]$ through correlative filters $c_p(z)$ prior to transmission, as explained in section 2.5, recall also figure 2.6. We used correlative filters with minimal memory, that is, with just one zero,

$$c_p\,(z) = \kappa_p\left(1 - z_p z^{-1}\right).$$

The zeros of the correlative filters for users 1 and 2 are $z_1 = \frac{1}{4}e^{-i\pi/2}$ and $z_2 = \frac{1}{2}e^{i\pi/4}$, respectively. The constants κ_p are adjusted accordingly to ensure unit-power filters. For both users, the analog transmitter shaping filter $p(t)$ is a raised-cosine with $\alpha = 80\%$ excess-bandwidth. Each communication channel, activated between each user and one of the receiving antennas, is a random realization of the continuous-time multipath model

$$g(t) = g_0\delta(t) + g_{\max}\delta(t - \Delta_{\max}) + \sum_{k=1}^{K} g_k\delta\left(t - \Delta_k\right), \tag{2.58}$$

where $\delta(t)$ denotes the continuous-time Delta dirac signal. In (2.58), we have two fixed paths at $\Delta = 0$ and $\Delta_{\max} = 2.8T_s$, where T_s denotes the symbol period. The integer K is the random number of extra paths. We take K to be uniformly distributed in the finite set $\{5, 6, \ldots, 15\}$. For all paths $k = 1, \ldots, K$, the delays Δ_k are uniformly distributed in the interval $[0, \Delta_{\max}]$. The complex fading coefficients g_k, $q = k, \ldots, K$, as well as g_0 and $g_{\max}$ denote unit-power complex circular Gaussian random variables. Thus, each path experiences independent Rayleigh fading. Each composite continuous-time channel $h(t) = p(t) \circledast g(t)$, where $\circledast$ denotes here convolution in continuous-time, is then sampled at the baud rate and truncated at $4T_s$. This truncation, forcing the channel $h(t)$ to have compact support (hence the polynomial filters to have finite degree), is a reasonable approximation since it deletes, at most, 4% of the channel power. As a consequence, the channel matrix $\boldsymbol{H}(z) = [\,\boldsymbol{h}_1(z)\,\boldsymbol{h}_2(z)\,]$ has equal column degrees $D_p = \deg \boldsymbol{h}_p(z) = 3$, which are assumed known at the receiver. The receiver has $Q = 4$ antennas, and identifies the MIMO channel $\boldsymbol{H}(z)$ on the basis of one packet of $N = 350$ data samples, $\{\boldsymbol{y}[n] : n = 1, \ldots, N\}$.

The CFIA is runned with the stacking parameter $L = 3$, recall (2.27). The exact correlation matrices $\boldsymbol{R}_{\boldsymbol{y}}[m]$ are replaced everywhere by their corresponding finite-sample estimates,

$$\widehat{\boldsymbol{R}}_{\boldsymbol{y}}[m] = \frac{1}{N-m}\sum_{n=m+1}^{N} \boldsymbol{y}[n]\boldsymbol{y}[n-m]^H.$$

The choice $M = 3$ is adopted in step 2 of the CFIA, recall (2.41). After channel identification, the proposed IIA is runned. In the IIA, the sources are extracted from the observations $\boldsymbol{y}[n]$ by zero-forcing filters $\boldsymbol{f}_p(z)$ of degree $F_p = 2D_p = 6$ and delay $d_p = D_p + 1 = 4$, recall (2.51) and (2.52). In our simulations, the additive observation noise $\boldsymbol{w}[n]$ is taken as spatio-temporal white Gaussian noise with power σ^2, recall (2.48). The signal-to-noise ratio (SNR) is defined as

$$\text{SNR} = \frac{\sum_{p=1}^{P} \text{E}\left\{\|\boldsymbol{h}_p(z) \circledast s_p[n]\|^2\right\}}{\text{E}\left\{\|\boldsymbol{w}[n]\|^2\right\}} = \frac{\|\boldsymbol{H}\|^2}{Q\sigma^2},$$

where $\boldsymbol{H}$ came from the identification $\boldsymbol{H}(z) \simeq (\boldsymbol{d}; \boldsymbol{H})$.

We start by illustrating a typical run of our technique. Figure 2.7 plots in the complex plane $\mathbb{C}$ one of the Q observed signals at the MIMO channel output, that is, an entry $y_q[n]$ of the data vector $\boldsymbol{y}[n]$. The joint effect of the intersymbol and co-channel interference is

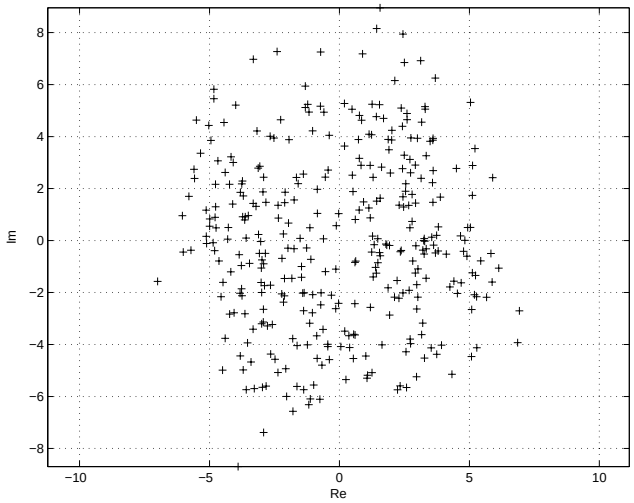

Figure 2.7: Output of the unequalized channel

clearly noticeable. Figures 2.8 and 2.9 (notice the difference in the vertical scale relative to figure 2.7) show the output of the equalized channel, that is, the signals $\beta_1[n]$ and $\beta_2[n]$ in (2.54). As seen, the algorithm recovers valid user signals from the observations.

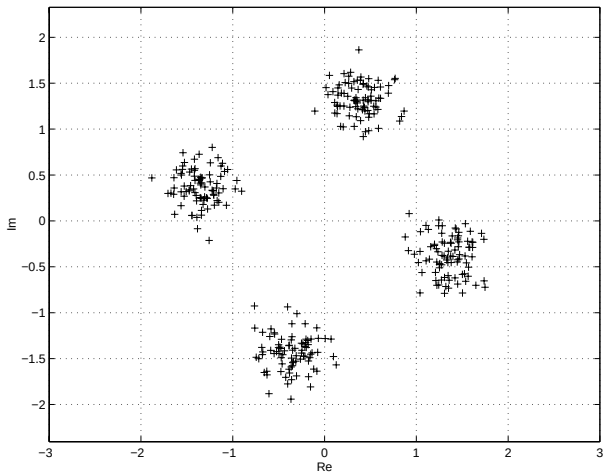

Figure 2.8: Signal estimate for user 1 ($\beta_1[n]$)

Note that the digital constellations are rotated in the signals $\beta_p[n]$ because the proposed derotating mechanism (recall step ii) in page 37) has not yet been runned. Moreover, note that this initial phase rotation cannot be avoided, since the CFIA can only estimate the channel modulo a phase ambiguity per user. The example pictured here corresponds to SNR $= 15\,\mathrm{dB}$.

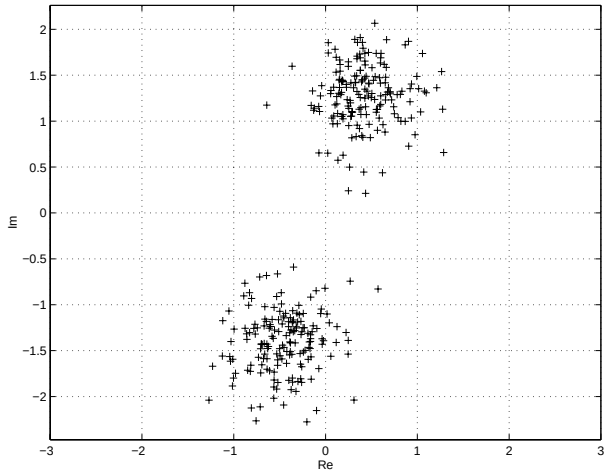

Figure 2.9: Signal estimate for user 2 ($\beta_2[n]$)

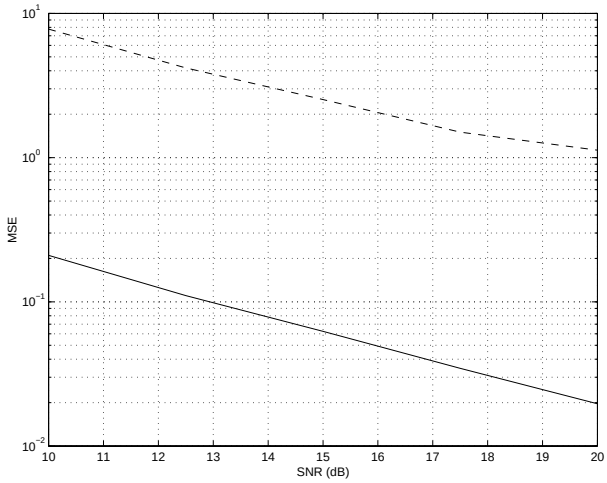

Figure 2.10: MSE of the CFIA (dashed) and the IIA (solid) channel estimate: SNR varies

We evaluated more extensively the performance of our proposed technique. We varied the SNR between $\mathrm{SNR}_{\min} = 10\,\mathrm{dB}$ and $\mathrm{SNR}_{\max} = 20\,\mathrm{dB}$, in steps of $\mathrm{SNR}_{\mathrm{step}} = 2.5\,\mathrm{dB}$. For each SNR, $J = 500$ statistically independent trials were considered. For each trial, we generated $N = 350$ data samples $\boldsymbol{y}[n]$, and ran the two proposed algorithms, CFIA and IIA. For both, we recorded the square-error (SE) of the channel estimate, that is,

$$\mathrm{SE} = \left\| \widehat{\boldsymbol{H}} - \boldsymbol{H} \right\|^2 .$$

The symbol error rates for both sources were obtained by error counting. Figure 2.10 displays the average results over the $J = 500$ trials, for the mean-square error (MSE) of the channel estimate. This is monotonically decreasing, as expected. The dashed and solid curves refer to the channel estimate provided by the CFIA and IIA, respectively. As seen, the iterative technique IIA improves significantly over the closed-form estimate CFIA. This is expected, since it exploits all the available statistical information, for example, the fact that the sources are discrete (digital), whereas the closed-form technique only exploits

SNR (dB)	User 1	User 2
10.0	0.0646	0.0276
12.5	0.0124	0.0032
15.0	0.0011	0.0001
17.5	0.0000	0.0000
20.0	0.0000	0.0000

Table 2.3: Symbol error rate (SER): CFIA ($N = 350$, SNR varies)

SNR (dB)	User 1	User 2
10.0	0.0014	0.0040
12.5	0.0002	0.0002
15.0	0.0000	0.0000
17.5	0.0000	0.0000
20.0	0.0000	0.0000

Table 2.4: Symbol error rate (SER): IIA ($N = 350$, SNR varies)

the 2nd order moments. However, the CFIA is needed to start the iterations in IIA, that is, to provide an accurate initial channel estimate, which is then subsequently refined by the iterations.

In tables 2.3 and 2.4, we show the symbol error rates (SER) associated to the two sources. These correspond to the symbol detectors implemented from the closed-form and iterative channel estimators, respectively. Notice that, as user 1 employs the QAM format, the SNR per symbol is lower and, as a consequence, the SER is higher. Moreover, as expected, the better accuracy of the iterative MIMO channel estimate results in a lower probability of error.

We also studied the performance of the proposed techniques with respect to the packet size N. We fixed SNR $= 10\,\mathrm{dB}$ and varied N between $N_{\min} = 200$ and $N_{\max} = 1000$ in steps of $N_{\text{step}} = 200$. Figure 2.11 shows the average results for the MSE, and tables 2.5 and 2.6 display the SER of both sources.

Scenario with three users. In this set of computer simulations, we consider $P = 3$ binary users, $\mathbb{A}_p = \mathbb{A}_{\text{BSK}}$, and compare our results with the TICC approach [12]. The TICC technique permits to identify the MIMO channel using only 2nd order statistics. It relies on a distinct data pre-processing scheme. This consists, at each transmitter, in multiplying the information sequence $a_p[n]$ by a cyclic frequency α_p, that is, $s_p[n] = a_p[n]\, e^{i\,\alpha_p n}$, in order to induce a conjugate cyclostationary frequency in the signal $s_p[n]$.

N	User 1	User 2
200	0.1527	0.0675
400	0.0461	0.0230
600	0.0218	0.0109
800	0.0168	0.0091
1000	0.0106	0.0075

Table 2.5: Symbol error rate (SER): CFIA (SNR $= 10\,\mathrm{dB}$, N varies)

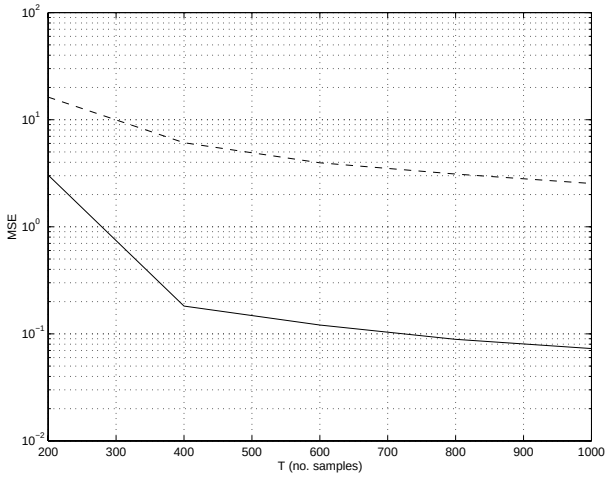

Figure 2.11: MSE of the CFIA (dashed) and the IIA (solid) channel estimate: N varies

N	User 1	User 2
200	0.0204	0.0156
400	0.0014	0.0041
600	0.0012	0.0041
800	0.0012	0.0041
1000	0.0012	0.0039

Table 2.6: Symbol error rate (SER): IIA (SNR = 10 dB, N varies)

See figure 2.12 and [12] for more details.

Each user employs a correlative filter with two zeros,

$$c_p(z) = \kappa_p \left(1 - z_{p,1}\, z^{-1}\right)\left(1 - z_{p,2}\, z^{-1}\right).$$

Table 2.7 discriminates the zeros of the correlative filters for each user. The constants κ_p are adjusted to guarantee unit-norm pre-filters. The multipath propagation model in (2.58) is maintained, but now $\Delta_{\max} = 2.5T_s$, and the composite channel is truncated at $3T_s$. This suppresses, at most, 4% of the channel power. Due to this truncation, the column degrees of the MIMO channel are given by $D_p = 3$. An $Q = 8$ antenna array sampled at the baud rate is assumed at the receiver. Also, the data packet size $N = 750$.

For the CFIA, we take the stacking parameter $L = 2$ in (2.27), and the value $M = 3$ is used in (2.41). For the IIA, we the degrees and delays of the zero-forcing filters $\boldsymbol{f}_p(z)$ are

User p	$z_{p,1}$	$z_{p,2}$
1	$\frac{1}{2}$	$-\frac{1}{2}$
2	$\frac{1}{3}e^{i\pi/3}$	$\frac{1}{4}e^{i\pi/2}$
3	$\frac{1}{2}e^{i3\pi/4}$	$-\frac{1}{4}e^{-i\pi/2}$

Table 2.7: Zeros of the correlative filters ($P = 3$ users)

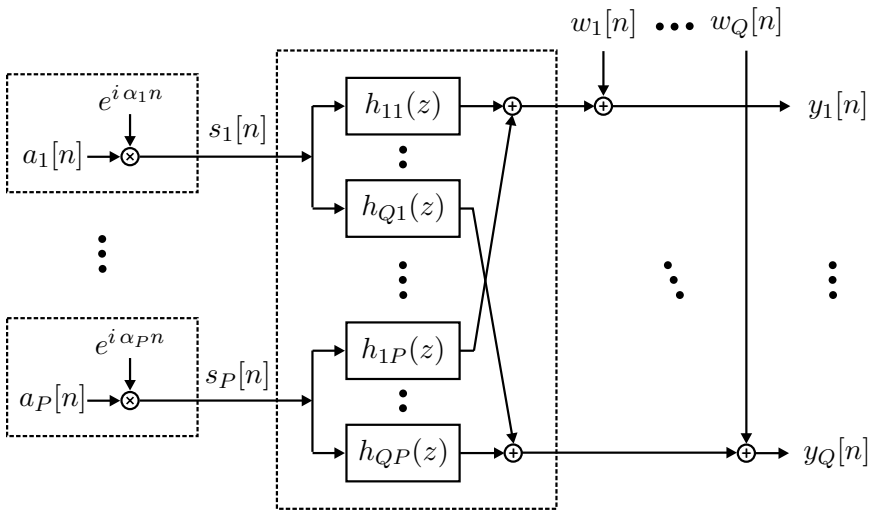

Figure 2.12: P-input/Q-output MIMO channel with P induced cyclic frequencies

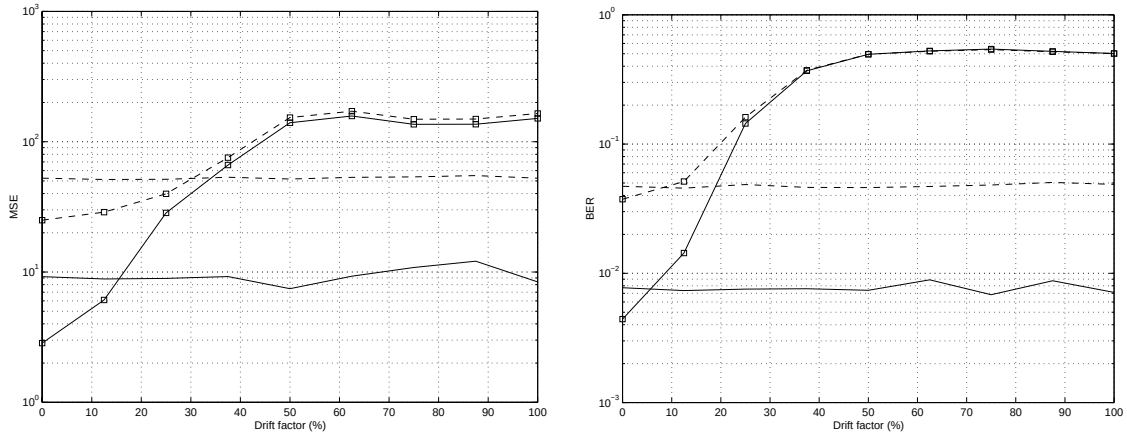

Figure 2.13: MSE (left) and BER of user 1 (right) for the proposed and TICC (with square marks) approaches : closed-form (dashed) and iterative (solid) algorithms (SNR = 5 dB)

$F_p = 2D_p = 6$ and $d_p = D_p + 1 = 4$, respectively. Also, we considered blocks of $T = 10$ samples in the derotating mechanism, see step ii) in page 37.

For the TICC approach, the three users employ cyclic frequencies given by $\alpha_1 = -0.35$, $\alpha_2 = 0$ and $\alpha_3 = 0.35$, respectively. Also, the Wiener filters in [12] are implemented with parameters $\delta = 4$ and $L = 7$. The nominal baseband rotations for the three users in (2.49) are given by $\omega_1 = \frac{2\pi}{750}$, $\omega_2 = -\frac{2\pi}{1000}$ and $\omega_3 = \frac{2\pi}{900}$, respectively. The channel degrees are assumed known for both approaches.

We performed computer simulations to compare the performance of our proposed technique and the TICC approach. We considered residual baseband rotations given by $\lambda\omega_p$, where the drift factor λ is varied between $\lambda_{\min} = 0\%$ (no baseband rotation) and $\lambda_{\max} = 100\%$ (moderate baseband rotation), in steps of $\lambda_{\text{step}} = 12.5\%$. For each λ, $J = 500$ statistically independent trials were performed. Each trial consisted in the generation of $N = 750$ data samples, and subsequent channel estimation and symbol detection as in the previous scenario with two users. The left plot in Figure 2.13 displays the average results, over the $J = 500$ trials considered. The SNR was fixed at 5 dB. For

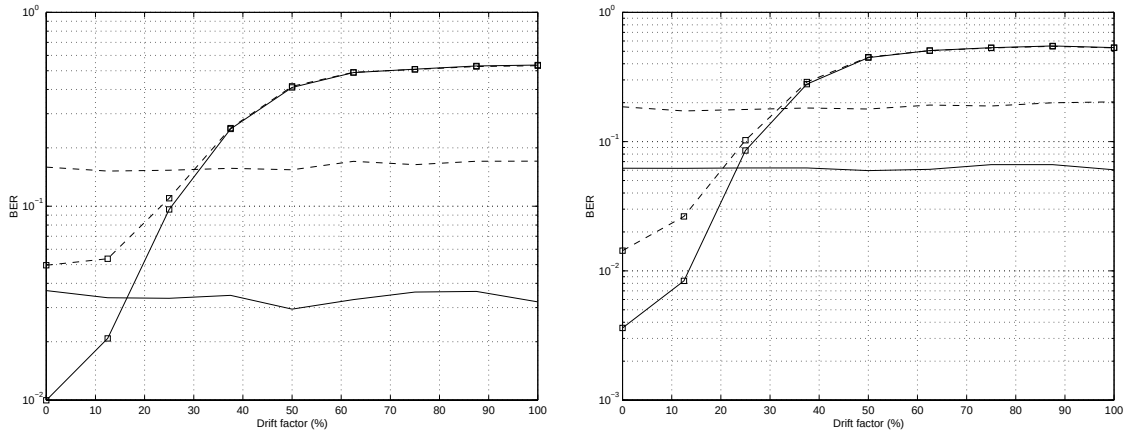

Figure 2.14: BER of user 2 (left) and user 3 (right) for the proposed and TICC (with square marks) approaches : closed-form (dashed) and iterative (solid) algorithms (SNR = 5 dB)

both approaches, the dashed and solid curves correspond to the closed-form and iterative channel estimates, respectively. Additionally, the curves associated with the TICC approach are labeled with a square mark. As seen, the accuracy of the channel estimate of our techniques (either closed-form or iterative) is almost insensitive to the drift baseband rotation factor λ. In contrast, the performance of the TICC estimators degrades as the carriers misadjustment gets worst. The right plot in Figure 2.13 and Figure 2.14 display the bit error rates (BER) associated with the two approaches, for the $P = 3$ users considered. As seen, the proposed technique outperforms TICC. A similar set of simulations was performed under SNR = 10 dB. The results are displayed in figures 2.15 and 2.16. We can infer the same conclusions as above. The fact that, in the face of uncontrollable baseband rotations, the TICC channel estimate performs worst than our channel estimate is hardly surprising. The vulnerability of the TICC technique to carrier frequency misadjustements is acknowledged already in [12], and can be explained quite easily: the presence of an unpredictable baseband rotation shifts all conjugate frequencies to unpredictable places, thereby destroying the structure previously inserted at each transmitter.

2.9 Conclusions

We started by formulating the blind channel identification problem (BCIP) over the set of MIMO channels $\mathbb{H}[z]$. The BCIP is an inverse problem. It asks for the point in $\mathbb{H}[z]$ which is consistent with the correlation matrices observed at the channel's output. However, irrespective of the spectral colors seen at the input of the MIMO system, two polynomial matrices in $\mathbb{H}[z]$ which are equal (modulo a phase offset per column) always induce the same 2nd order statistics at the output. Thus, infinitely many points in $\mathbb{H}[z]$ are consistent with the given correlation matrices. This implies that the original formulation of the BCIP is meaningless. We took the viewpoint of modelling the phase ambiguities as an equivalence relation $\sim$ in the set $\mathbb{H}[z]$, and reformulated the BCIP over the induced quotient space of channel equivalence classes $\mathbb{H}[z]/\sim$. However, under the standard assumption of each channel input being a time decorrelated sequence, the problem is still not well-posed:

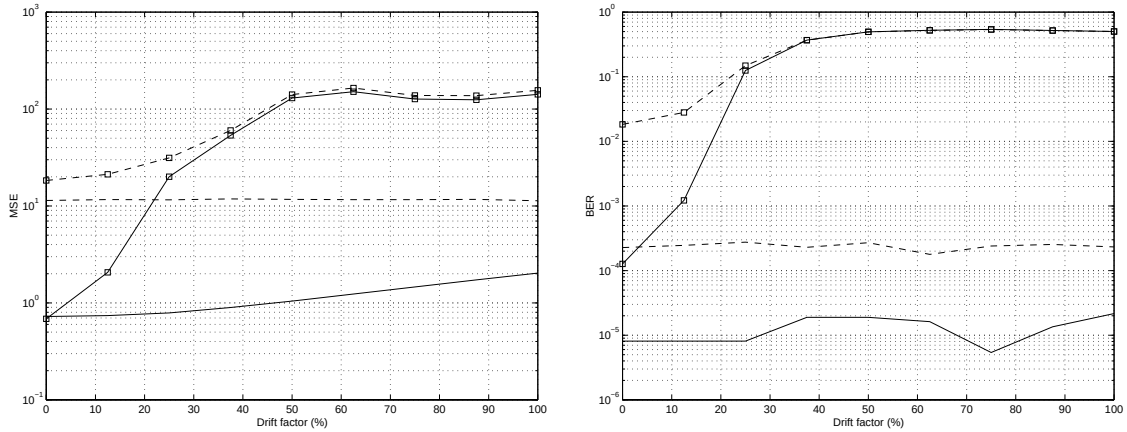

Figure 2.15: MSE (left) and BER of user 1 (right) for the proposed and TICC (with square marks) approaches : closed-form (dashed) and iterative (solid) algorithms (SNR = 10 dB)

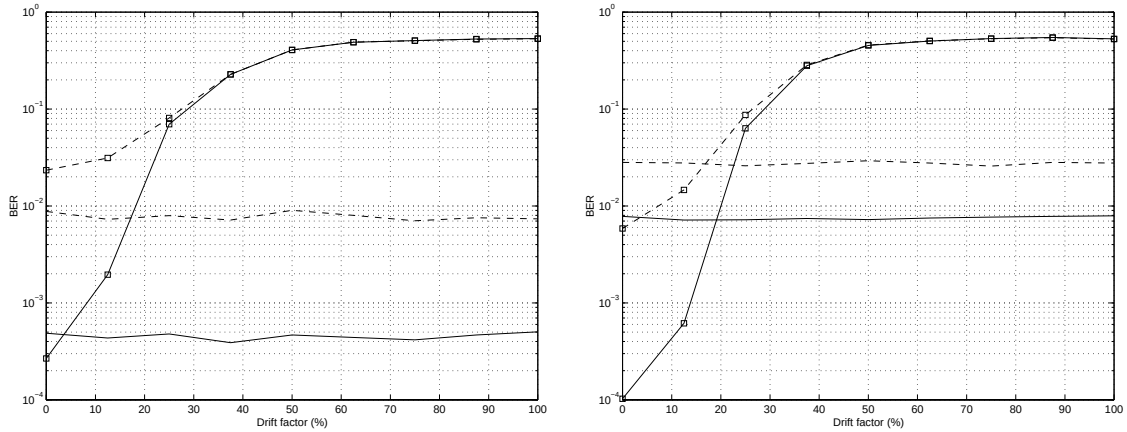

Figure 2.16: BER of user 2 (left) and user 3 (right) for the proposed and TICC (with square marks) approaches : closed-form (dashed) and iterative (solid) algorithms (SNR = 10 dB)

distinct channel equivalence classes, that is, points in $\mathbb{H}[z]/\sim$, may be consistent with the correlation matrices at the channel's output. That is, the map $\varphi : \mathbb{H}[z]/\sim \rightarrow \mathbb{C}_{\mathbb{Z}}$, which associates to each channel equivalence class the set of correlation matrices induced at the channel's output, may fail to be injective. We have shown how to circumvent this difficulty. More precisely, we proved an identifiability theorem (theorem 2.1) which asserts that, under a certain spectral diversity condition on the random signals seen at the input, there is one and only one point in $\mathbb{H}[z]/\sim$ which is consistent with the given 2nd order statistics of the channel's output. Next, we proved a feasibility theorem (theorem 2.2). It asserts that the sufficient spectral condition which ensures identifiability can be easily induced by unit-norm minimum-phase pre-filters located at the sources. The proof of the identifiability theorem is not constructive. We proceeded to develop a closed-form identification algorithm (CFIA) which achieves the predicted identifiability bound. The CFIA takes the 2nd order statistics of the channel output and reconstructs the underlying channel equivalence class. An iterative source separation and channel identification algorithm (IIA) was also presented. The IIA is an iterative scheme, initializated by the CFIA, which decodes the emitted sources' information symbols and refines the channel estimate, while tracking residual phase drifts in the baseband signals. In the final part of this chapter, we compared our approach, through computer simulations, with the transmitter induced conjugate cyclostationary (TICC) technique. The TICC approach consists in pre-processing the symbol information sequences in a distinct manner. It does not use correlative filters. Instead, a conjugate cyclostationary frequency is induced at each transmitter. Like our pre-processing, this inserted structure in the transmitted signals ensures channel identifiability (modulo a phase offset per column) from the 2nd order statistics of the channel's output. The simulation results have shown that, in contrast to the TICC approach, our pre-processing is resilient to baseband phase drifts induced by carrier frequency misadjustments.

Chapter 3

Performance Analysis

3.1 Chapter summary

In this chapter, we carry out a theoretical performance analysis of the closed-form identification algorithm (CFIA) introduced in chapter 2. More precisely, we assess the quality of the CFIA's estimate in the quotient space $\mathbb{H}[z]/\sim$ for a given (finite) number N of channel output observations. Section 3.2 starts by motivating this theoretical study. The main reason for embarking in such a study lies in the fact that this is a necessary step for the more ambitious project of optimally designing the pre-filters. Our analysis is developed under a strong simplification. We assume that the column degrees of the MIMO channel polynomial matrix $\boldsymbol{H}(z)$ are known. Thus, their detection is unnecessary and the CFIA focus only on the estimate of the sub-channels' coefficients. In equivalent terms, given the identification $\boldsymbol{H}(z) \simeq (\boldsymbol{d}; \boldsymbol{H})$, we assume that the discrete component $\boldsymbol{d}$ is known, and thereby, the CFIA only estimates the channel's continuous component $\boldsymbol{H}$. This simplification is invoked to maintain the analysis tractable. Therefore, we deal in this chapter with a somewhat "truncated" version of the space $\mathbb{H}[z]$ and its quotient $\mathbb{H}[z]/\sim$, in addition to a simplified CFIA. After these modifications are carefully explained, section 3.2 provides a macroscopic view of the strategy employed in the performance analysis. Four main phases are identified and briefly delineated. In a nutshell, we resort to the large-sample regime ($N \to \infty$) and make the distributions of all random objects of interest converge to corresponding normal distributions. These latter distributions can be characterized in closed-form and, for all practical N, approximate quite well the exact distributions. However, we face a rather non-standard performance study because our estimator takes values, not on an usual Euclidean space, but on an abstract space: the quotient space of identifiable channel classes $\mathbb{H}[z]/\sim$. Furthermore, because we want the concept of distance to be available in $\mathbb{H}[z]/\sim$ (in order to measure the quality of the estimate with respect to the true channel class), the quotient space (up to this point only a purely algebraic object) must experience a metamorphosis and became a Riemannian manifold. Some special machinery must be developed to tackle an asymptotic statistical analysis in this peculiar setting. This is done in section 3.3. We present the necessary theoretical framework to handle random objects on manifolds. More precisely, a proper definition of asymptotic normality of random sequences on Riemannian manifolds is needed, and some well known results for Euclidean spaces are then extended to this Riemannian context. Equipped with these theoretical tools, section 3.4 implements the four main phases previously enumerated

in the performance study. We devote a subsection to each phase. **i)** The first phase is implemented in subsection 3.4.1. It endows the quotient space $\mathbb{H}[z]/\sim$ with a smooth geometrical structure, making it a connected Riemannian manifold. Some pertinent geometrical features of $\mathbb{H}[z]/\sim$ are then investigated. **ii)** Subsection 3.4.2 implements the second phase of the performance analysis. Roughly, it re-interprets the CFIA as a smooth map between manifolds and computes its derivative at any given point in its domain (this data is needed in the sequel). **iii)** The third phase is implemented in subsection 3.4.3. It establishes the asymptotic normality (as $N \to \infty$) of the finite-sample estimates of the MIMO channel output correlation matrices. These estimated correlation matrices constitute the input of the CFIA. **iv)** The performance analysis is closed in subsection 3.4.4. It brings together all the pieces developed so far and establishes the asymptotic normality of the CFIA's estimate in the Riemannian manifold $\mathbb{H}[z]/\sim$. Moreover, a closed-form expression which approximates the mean-square error of the CFIA estimate is obtained. Section 3.5 validates the theoretical analysis. We compare the closed-form expression obtained for the mean-square error of the CFIA's estimate with Monte-Carlo computer simulations, as a function of the number of available channel observations. Section 3.6 contains the main conclusions of this chapter.

3.2 Performance analysis: macroscopic view

Motivation. In this chapter, we analyze the accuracy of the CFIA as a point estimator in $\mathbb{H}[z]/\sim$, the quotient space of identifiable channel equivalence classes. More precisely, let N designate the number of available measurements of the MIMO channel output, and $\left[\widehat{\boldsymbol{H}^N(z)}\right]$ the corresponding channel equivalence class estimate provided by the CFIA. Our goal is to derive a closed-form expression for

$$J\left[N\,;\,\boldsymbol{c}(z),[\boldsymbol{H}(z)]\right] = \mathrm{E}\left\{d\left(\left[\boldsymbol{H}(z)\right],\left[\widehat{\boldsymbol{H}^N(z)}\right]\right)^2\right\}, \tag{3.1}$$

the mean-square distance between the true and estimated channel equivalence classes, when N observations are available and given that $\boldsymbol{c}(z) = (c_1(z), c_2(z), \ldots, c_P(z))^T$ is the P-tuple of pre-filters coloring the 2nd order statistics of the underlying MIMO channel class $[\boldsymbol{H}(z)]$. In (3.1), $d : \mathbb{H}[z]/\sim \times \mathbb{H}[z]/\sim \to \mathbb{R}$ denotes a metric on the quotient space $\mathbb{H}[z]/\sim$ (to be discussed soon). The interest in obtaining an expression for (3.1) is twofold. **i)** At the very least, it permits to avoid unnecessary massive computer simulations (Monte-Carlo runs) to assess the quality (as measured by the distance d) of a N-sample based estimate of the channel class $[\boldsymbol{H}(z)]$ given the choice $\boldsymbol{c}(z)$ for the P-tuple of pre-filters. **ii)** More importantly, it permits to address the issue of optimum pre-filter design, when known random environments dictate the statistics of the MIMO channel $\boldsymbol{H}(z)$. That is, suppose one disposes of a probabilistic model for the entries (sub-channel coefficients) of $\boldsymbol{H}(z)$. Such statistical characterization of the channel could be based, for example, on field measurements. A sensible approach for designing the pre-filters, supposing data packets of length N, could be the constrained ($c_p(z)$: unit-power, minimum-phase, etc) minimization of

$$J\left(\boldsymbol{c}(z)\right) = \mathrm{E}\left\{d\left(\left[\boldsymbol{H}(z)\right],\left[\widehat{\boldsymbol{H}^N(z)}\right]\right)^2\right\}, \tag{3.2}$$

where the expectation is over the ensemble of transmitted information symbols, noise and channel realizations. Although not explicit in (3.2), $\big[\widehat{\boldsymbol{H}^N(z)}\big]$ denotes the estimate based on the choice $\boldsymbol{c}(z)$ for the pre-filters. A more exact but cumbersome notation would be $\big[\widehat{\boldsymbol{H}^N(z)}\big](\boldsymbol{c}(z))$. We employ the former symbol for brevity. Since (3.2) can be rewritten as

$$J\left(\boldsymbol{c}(z)\right) = \mathrm{E}_{\boldsymbol{H}(z)}\left\{\mathrm{E}\left\{d\left(\left[\boldsymbol{H}(z)\right], \big[\widehat{\boldsymbol{H}^N(z)}\big]\right)^2 \mid \boldsymbol{H}(z)\right\}\right\} \tag{3.3}$$

$$= \mathrm{E}_{\boldsymbol{H}(z)}\left\{J\left[N\,;\,\boldsymbol{c}(z), [\boldsymbol{H}(z)]\right]\right\}, \tag{3.4}$$

we see that knowledge of $J\left[N\,;\,\boldsymbol{c}(z), [\boldsymbol{H}(z)]\right]$ is necessary for implementing such an approach. Notice that, in the right-hand side of (3.3), the inner expectation is only over the information symbols and noise, whereas the outer expectation denotes an average over the ensemble of channel realizations. The identity in (3.4) is valid since we are assuming that the MIMO channel is a random object statistically independent of transmitted information symbols and noise (thus, it can be treated as a constant in the inner expectation).

Simplification. Guessing the channel matrix $\boldsymbol{H}(z) = [\,\boldsymbol{h}_1(z)\,\boldsymbol{h}_2(z)\,\cdots\,\boldsymbol{h}_P(z)\,]$ from the SOS of the observations is a joint detection-estimation problem. We must detect the degrees $D_p = \deg \boldsymbol{h}_p(z)$ of the P column polynomial filters

$$\boldsymbol{h}_p(z) = \sum_{d=0}^{D_p} \boldsymbol{h}_p[d] z^{-d}, \qquad\qquad p = 1, 2, \ldots, P,$$

which we collected in the vector of integers $\boldsymbol{d} = (D_1, D_2, \ldots, D_P)^T$, and estimate the corresponding filter coefficients $\boldsymbol{H}_p = \boldsymbol{\mathcal{T}}_0\left(\boldsymbol{h}_p(z)\right)$, which we gathered in the $Q \times (D+P)$ complex matrix

$$\boldsymbol{H} = \Big[\ \underbrace{\boldsymbol{h}_1[0]\,\boldsymbol{h}_1[1]\,\cdots\,\boldsymbol{h}_1[D_1]}_{\boldsymbol{H}_1}\ \ \underbrace{\boldsymbol{h}_2[0]\,\boldsymbol{h}_2[1]\,\cdots\,\boldsymbol{h}_2[D_2]}_{\boldsymbol{H}_2}\ \ \cdots\ \ \underbrace{\boldsymbol{h}_P[0]\,\boldsymbol{h}_P[1]\,\cdots\,\boldsymbol{h}_P[D_P]}_{\boldsymbol{H}_P}\ \Big], \tag{3.5}$$

where $D = \operatorname{ord} \boldsymbol{H}(z) = \sum_{p=1}^{P} D_p$. This is just the identification $\boldsymbol{H}(z) \simeq (\boldsymbol{d}; \boldsymbol{H})$ discussed in page 18. In order to obtain (3.1), we will make a major simplification: we assume that the vector of column degrees $\boldsymbol{d}$ is known, and focus only on the estimation of $\boldsymbol{H}$. We avoid the detection problem in order to keep the theoretical analysis tractable. This is an exact performance analysis of the proposed receiver in scenarios where the statistical characterization of the channel clearly indicates a strong (probability 1) mode for a certain delay-spread user configuration, that is, for a certain configuration of channel column degrees $\boldsymbol{d}$. In these cases, the receiver runs a simplified version of the CFIA which estimates $\boldsymbol{H}$ on the assumption that the discussed dominant mode $\boldsymbol{d}$ is the activated one. For the generic case where the receiver runs the full CFIA, the results we are about to obtain for (3.1) must be interpreted, in fact, only as a lower-bound, since they are derived on the optimistic assumption that the discrete part $\boldsymbol{d}$ of the MIMO channel is correctly detected.

Apart from the mentioned simplification we will work under some additional assumptions:

B1. The degrees $D_p = \deg \boldsymbol{h}_p(z)$ of the channel polynomial filters are equal, that is, $\boldsymbol{d} = (D_1, D_2, \ldots, D_P)^T = \boldsymbol{d}_0 = (D_0, D_0, \ldots, D_0)^T$, for a known $D_0 \in \mathbb{N}$. The same applies to the degrees $C_p = \deg \boldsymbol{c}_p(z)$ of the correlative pre-filters, that is, $\boldsymbol{c} = (C_1, C_2, \ldots, C_P)^T = (C_0, C_0, \ldots, C_0)^T$, for a certain $C_0 \in \mathbb{N}$;

B2. The number of MIMO channel outputs exceeds the order of the channel plus the number of its inputs, that is, $Q > D = \operatorname{ord} \boldsymbol{H}(z) + P = P(D_0 + 1)$;

B3. The unfiltered and uncorrelated information signal $a_p[n]$ (recall figure 2.6 in page 25) denotes a sequence of independent and identically symbols drawn from the QPSK alphabet (normalized to unit-power) $\mathbb{A}_{\text{QPSK}} = \left\{ \pm\frac{1}{\sqrt{2}} \pm i\frac{1}{\sqrt{2}} \right\}$. Moreover, the observation noise $\boldsymbol{w}[n]$ is taken to be white spatio-temporal Gaussian noise with power σ^2, that is, $\boldsymbol{R}_{\boldsymbol{w}}[m] = \mathrm{E}\left\{ \boldsymbol{w}[n]\boldsymbol{w}[n-m]^H \right\} = \sigma^2 \boldsymbol{I}_Q \delta[m]$.

These additional assumptions are only introduced for notational convenience. In loose terms, they symmetrize the problem and lead to more compact matrix formulas. This is clearly the justification for assumption B1. Assumption B2 permits us to run the CFIA with non-stacked data samples, that is, to take the value $L = 0$ as the choice for the stacking parameter in the CFIA. As a consequence, the algorithm manipulates smaller data objects. Finally, in assumption B3, we are only picking a choice for the information sources' and noise statistical models. This is necessary to make computations. We emphasize that the general case, that is, distinct channel degrees, $Q \leq D$, other digital sources $a_p[n]$, and so on, follows easily from the particular setup in B1-B3. The analysis of this particular case captures the flavor of the generic situation.

Redefinition of $\mathbb{H}[z]$ and $\mathbb{H}[z]/\sim$. Recall that the set $\mathbb{H}[z]$ was defined in page 20 as the set of tall, irreducible, column-reduced $Q \times P$ polynomial matrices that are memory-limited by $D_{\max}$. Since for any $\boldsymbol{H}(z) \in \mathbb{H}[z]$, we have the identification $\boldsymbol{H}(z) \simeq (\boldsymbol{d}; \boldsymbol{H})$, we may view $\mathbb{H}[z]$ as a finite stack of disjoint leaves indexed by the vector of integers

$$\boldsymbol{d} = (D_1, D_2, \ldots, D_P)^T, \qquad D_p \in \{0, 1, \ldots, D_{\max}\}, \tag{3.6}$$

see figure 3.1. The exact number of leaves is $(D_{\max}+1)^P$. The leaf corresponding to a con-

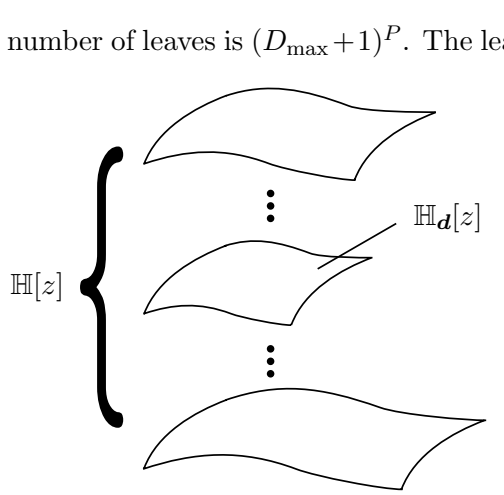

Figure 3.1: The set $\mathbb{H}[z]$ as a finite stack of leaves

figuration of column degrees $\boldsymbol{d} = (D_1, D_2, \ldots, D_P)^T$, called the $\boldsymbol{d}$-leaf and denoted $\mathbb{H}_{\boldsymbol{d}}[z]$, is

the subset of $\mathbb{H}[z]$ consisting of those polynomial matrices $\boldsymbol{H}(z) = [\,\boldsymbol{h}_1(z)\,\boldsymbol{h}_2(z)\,\cdots\boldsymbol{h}_P(z)\,]$ satisfying $\deg \boldsymbol{h}_p(z) = D_p$, for $p = 1, 2, \ldots, P$. That is,

$$\mathbb{H}[z] = \coprod_{\boldsymbol{d}} \mathbb{H}_{\boldsymbol{d}}[z],$$

where $\coprod$ stands for disjoint union and it is understood that the index $\boldsymbol{d}$ runs over all possible configurations satisfying (3.6). Furthermore, each single $\boldsymbol{d}$-leaf may be viewed as a subset of a complex space through the identification (3.5), that is, via the one-to-one mapping $\iota : \mathbb{H}_{\boldsymbol{d}}[z] \to \mathbb{C}^{*Q\times(D_1+1)} \times \mathbb{C}^{*Q\times(D_2+1)} \times \cdots \times \mathbb{C}^{*Q\times(D_P+1)}$, where $\mathbb{C}^{*Q\times(D_p+1)} = \mathbb{C}^{Q\times(D_p+1)} - \left\{\boldsymbol{0}_{Q\times(D_p+1)}\right\}$, given by

$$\boldsymbol{H}(z) = [\,\boldsymbol{h}_1(z)\,\boldsymbol{h}_2(z)\,\cdots\boldsymbol{h}_P(z)\,] \quad \overset{\iota}{\mapsto} \quad (\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_P)\,, \tag{3.7}$$

where $\boldsymbol{H}_p = \boldsymbol{\mathcal{T}}_0\,(\boldsymbol{h}_p(z))$. Note that the "dimensionality" of the leaves is not constant. Lemma 3.1 asserts that each leaf of $\mathbb{H}[z]$ occupies almost all the totality of its host complex space.

Lemma 3.1. *Consider the $\boldsymbol{d}$-leaf identification mapping ι defined in (3.7). Then, $\iota\,(\mathbb{H}_{\boldsymbol{d}}[z])$ is an open and dense subset of $\mathbb{C}^{*Q\times(D_1+1)} \times \mathbb{C}^{*Q\times(D_2+1)} \times \cdots \times \mathbb{C}^{*Q\times(D_P+1)}$.*

Proof: See appendix B.

Now, in this context, the aforementioned simplification means that both the MIMO channel $\boldsymbol{H}(z)$ and its estimate $\widehat{\boldsymbol{H}}^N(z)$ belong to the leaf $\mathbb{H}_{\boldsymbol{d}_0}[z]$. Since all the action takes place there we may cut the remaining leaves from the analysis. Furthermore, in light of lemma 3.1, we may view the $\boldsymbol{d}_0$-leaf as an open, dense subset of

$$\mathbb{C}^*_{\boldsymbol{d}_0} = \underbrace{\mathbb{C}^{*Q\times(D_0+1)} \times \mathbb{C}^{*Q\times(D_0+1)} \times \cdots \times \mathbb{C}^{*Q\times(D_0+1)}}_{P}.$$

We do not have the full identification $\mathbb{H}_{\boldsymbol{d}_0}[z] = \mathbb{C}^*_{\boldsymbol{d}_0}$ because the injective mapping $\iota : \mathbb{H}_{\boldsymbol{d}_0}[z] \to \mathbb{C}^*_{\boldsymbol{d}_0}$ defined in (3.7) is not onto. Its image misses the whole host complex space by a nowhere dense subset. However, the full identification is highly desirable since the set $\mathbb{C}^*_{\boldsymbol{d}_0}$ is much easier to work with than its subset $\iota\,(\mathbb{H}_{\boldsymbol{d}_0}[z]) \subset \mathbb{C}^*_{\boldsymbol{d}_0}$. Note that, although $\iota\,(\mathbb{H}_{\boldsymbol{d}_0}[z])$ is also open, it cannot be written as a Cartesian product of P copies of a single set. For example, the entries in distinct columns of $\boldsymbol{H}(z)$ must "cooperate" in order to ensure column-reducedness. But $\mathbb{C}^*_{\boldsymbol{d}_0}$ has this decoupling property. To achieve the full identification, we perform a slight enlargement of the domain of ι to

$$\mathbb{H}_{\leq \boldsymbol{d}_0}[z] = \left\{\boldsymbol{H}(z) = [\,\boldsymbol{h}_1(z)\,\boldsymbol{h}_2(z)\,\cdots\boldsymbol{h}_P(z)\,] \in \mathbb{C}^{Q\times P}[z] \,:\, \deg \boldsymbol{h}_p(z) \leq D_0\right\}.$$

Note that $\mathbb{H}_{\boldsymbol{d}_0}[z] \subset \mathbb{H}_{\leq \boldsymbol{d}_0}[z]$. Now, we redefine $\mathbb{H}[z]$ as the inverse image of $\mathbb{C}^*_{\boldsymbol{d}_0}$ under this (still one-to-one) extended mapping. That is, we set $\mathbb{H}[z] = \iota^{-1}\left(\mathbb{C}^*_{\boldsymbol{d}_0}\right)$, which gives

$$\mathbb{H}[z] = \left\{\boldsymbol{H}(z) = [\,\boldsymbol{h}_1(z)\,\boldsymbol{h}_2(z)\,\cdots\boldsymbol{h}_P(z)\,] \in \mathbb{C}^{Q\times P}[z] \,:\, \boldsymbol{h}_p(z) \neq 0 \text{ and } \deg \boldsymbol{h}_p(z) \leq D_0\right\}. \tag{3.8}$$

In sum, based on our previous assumptions, we started by simplifying the structure of $\mathbb{H}[z]$ by dropping the irrelevant leaves. Then, we enlarged a bit (in the topological sense of

lemma 3.1) the remaining leaf in order to attain a pleasant, full identification with an open "rectangular" subset of a Cartesian product of complex spaces. With respect to the quotient $\mathbb{H}[z]/\sim$ we maintain its definition, but now $\mathbb{H}[z]$ denotes the space in (3.8). That is, $\mathbb{H}[z]/\sim$ is the set of equivalence classes of $\mathbb{H}[z]$ under the equivalence relation $\sim$, where $\boldsymbol{G}(z) \sim \boldsymbol{H}(z)$ if and only if $\boldsymbol{G}(z) = \boldsymbol{H}(z)\boldsymbol{\Theta}(\boldsymbol{\theta})$ for some $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_P)^T \in \mathbb{R}^P$, where $\boldsymbol{\Theta}(\boldsymbol{\theta}) = \operatorname{diag}\left(e^{i\theta_1}, e^{i\theta_2}, \ldots, e^{i\theta_P}\right)$. Thus, $\sim$ continues to denote equality of polynomial matrices modulo a phase offset per column. Moreover, we keep the notation $\pi : \mathbb{H}[z] \to \mathbb{H}[z]/\sim$, $\pi(\boldsymbol{H}(z)) = [\boldsymbol{H}(z)]$ for the projection map merging all equivalent polynomial matrices into the corresponding equivalence class.

Redefinition of CFIA**.** Given that the column degrees of $\boldsymbol{H}(z)$ are known and assumptions B1-B3 are in force, the CFIA can be simplified to the algorithm shown in table 3.1, page 55. The CFIA takes an ordered $(M+1)$-tuple of $Q \times Q$ matrices, the finite-sample estimates of the MIMO output correlation matrices, and delivers a point in the quotient space $\mathbb{H}[z]/\sim$, which is the estimated channel equivalence class $\left[\widehat{\boldsymbol{H}^N(z)}\right]$. Here, we defined

$$\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^N[m] = \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{y}[n]\boldsymbol{y}[n-m]^H, \qquad m = 0, 1, \ldots, M,$$

and indexed the available $N+M$ data samples as $\{\boldsymbol{y}[-M+1], \ldots, \boldsymbol{y}[0], \boldsymbol{y}[1], \ldots, \boldsymbol{y}[N]\}$. Thus, in all rigor, we should have denoted the CFIA output by $\left[\widehat{\boldsymbol{H}^{N+M}(z)}\right]$, since $N+M$ is the number of observations. However, we always have $N \gg M$ in practice, so the approximation $N+M \approx N$ is valid. In fact, since we will perform an asymptotic ($N \to \infty$) analysis of the CFIA this distinction is irrelevant. Note that, at the exit of step 4), we are invoking the discussed enlarged identification mapping ι to view the matrix $\widehat{\boldsymbol{H}}^N \in \mathbb{C}^*_{\boldsymbol{d}_0}$ as a polynomial matrix $\widehat{\boldsymbol{H}}^N(z)$ in $\mathbb{H}[z]$.

In step 1), we assumed that the D largest eigenvalues of the Hermitean matrix $\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^N[0]$ are distinct. The subset of matrices satisfying this property is open and dense in the set of $Q \times Q$ Hermitean matrices, so the assumption that $\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^N[0]$, built on the basis of noisy observations, falls in this set is realistic. More importantly, note the presence of D reference vectors $\boldsymbol{r}_1, \ldots, \boldsymbol{r}_D$, which are supposed to be fixed beforehand. The dth vector $\boldsymbol{r}_d$ is introduced to desambiguate the choice of the eigenvector $\widehat{\boldsymbol{u}}_d$, which is only defined modulo a phase ambiguity. In other words, the reference vectors enable a smooth selection of the corresponding eigenvectors. This mechanism is based on the assumption that each $\widehat{\boldsymbol{u}}_d$ does not fall in the Lebesgue measure zero hyperplane orthogonal to the previously chosen reference vector $\boldsymbol{r}_d$. Again, because this property holds almost everywhere in the space where $\widehat{\boldsymbol{u}}_d$ (obtained from noisy observations) lives, we consider it to be realistic. Similar remarks apply to step 3), where the minimum eigenvalues of the P Hermitean matrices $\widehat{\boldsymbol{S}}_1, \ldots, \widehat{\boldsymbol{S}}_P$, are assumed to be simple (that is, with multiplicity 1) and the P reference vectors $\boldsymbol{s}_1, \ldots, \boldsymbol{s}_P$ are employed. It is important to show that the CFIA output is invariant with respect to the particular choice of reference vectors adopted in steps 1) and 3). To establish this, assume that one feeds the $(M+1)$-ordered tuple of estimated correlation matrices to another version of the CFIA, that is, one which is based on distinct reference vectors from those in table 3.1. We label the inner variables of this new version with a tilde rather than a hat. Thus, $\widetilde{\boldsymbol{\lambda}}$ is the counterpart of $\widehat{\boldsymbol{\lambda}}$ in table 3.1, $\widetilde{\boldsymbol{U}}$ is the

input: $\left(\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[0], \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[1], \ldots, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[M]\right)$

step 1) compute $\widehat{\boldsymbol{\lambda}} = \left(\widehat{\lambda}_1, \widehat{\lambda}_2, \ldots, \widehat{\lambda}_D\right)^T$ and $\widehat{\boldsymbol{U}} = [\,\widehat{\boldsymbol{u}}_1\, \widehat{\boldsymbol{u}}_2 \cdots \widehat{\boldsymbol{u}}_D\,]$ such that

$$\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[0]\,\widehat{\boldsymbol{u}}_d = \widehat{\lambda}_d \widehat{\boldsymbol{u}}_d, \quad \widehat{\boldsymbol{u}}_d^H \widehat{\boldsymbol{u}}_d = 1, \quad \mathrm{Re}\,\boldsymbol{r}_d^H \widehat{\boldsymbol{u}}_d > 0, \quad \mathrm{Im}\,\boldsymbol{r}_d^H \widehat{\boldsymbol{u}}_d = 0,$$

for $d = 1, 2, \ldots, D$. Moreover, the entries of $\widehat{\boldsymbol{\lambda}}$ are arranged in decreasing order, $\widehat{\lambda}_1 > \widehat{\lambda}_2 > \cdots > \widehat{\lambda}_D$, and denote the $D = (D_0 + 1)P = \mathrm{ord}\,\boldsymbol{H}(z)$ largest eigenvalues of $\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[0]$;

step 2) compute

$$\widehat{\boldsymbol{\Upsilon}}[m] = \widehat{\boldsymbol{\mathcal{G}}}_0^{+}\, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[m]\, \widehat{\boldsymbol{\mathcal{G}}}_0^{+H},$$

for $m = 1, 2, \ldots, M$, where

$$\widehat{\boldsymbol{\mathcal{G}}}_0 = \widehat{\boldsymbol{U}} \left(\widehat{\boldsymbol{\Lambda}} - \sigma^2 \boldsymbol{I}_D\right)^{1/2},$$

with

$$\widehat{\boldsymbol{\Lambda}} = \mathrm{diag}\left(\widehat{\lambda}_1, \ldots, \widehat{\lambda}_D\right);$$

step 3) compute $\widehat{\boldsymbol{w}}_p$, the eigenvector associated with the minimum eigenvalue of $\widehat{\boldsymbol{S}}_p$ and satisfying

$$\widehat{\boldsymbol{w}}_p^H \widehat{\boldsymbol{w}}_p = 1, \quad \mathrm{Re}\,\boldsymbol{s}_p^H \widehat{\boldsymbol{w}}_p > 0, \quad \mathrm{Im}\,\boldsymbol{s}_p^H \widehat{\boldsymbol{w}}_p = 0.$$

Here, $\widehat{\boldsymbol{S}}_p = \widehat{\boldsymbol{T}}_p{}^H \widehat{\boldsymbol{T}}_p$, where

$$\widehat{\boldsymbol{T}}_p = \begin{bmatrix} \widehat{\boldsymbol{T}}_p[1] \\ \widehat{\boldsymbol{T}}_p[2] \\ \vdots \\ \widehat{\boldsymbol{T}}_p[M] \end{bmatrix}, \; \widehat{\boldsymbol{T}}_p[m] = \begin{bmatrix} \boldsymbol{I}_{D_0+1} \otimes \widehat{\boldsymbol{\Upsilon}}[m] - \boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; D_0]^T \otimes \boldsymbol{I}_D \\ \\ \boldsymbol{I}_{D_0+1} \otimes \widehat{\boldsymbol{\Upsilon}}[m]^H - \overline{\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; D_0]} \otimes \boldsymbol{I}_D \end{bmatrix}.$$

Define $\widehat{\boldsymbol{\mathcal{W}}} = \left[\,\widehat{\boldsymbol{\mathcal{W}}}_1\, \widehat{\boldsymbol{\mathcal{W}}}_2 \cdots \widehat{\boldsymbol{\mathcal{W}}}_P\,\right]$, where

$$\widehat{\boldsymbol{\mathcal{W}}}_p = \sqrt{D_0 + 1}\; \mathrm{vec}^{-1}\left(\widehat{\boldsymbol{w}}_p\right) \,:\, D \times (D_0 + 1);$$

step 4) set $\widehat{\boldsymbol{H}}^N = \widehat{\boldsymbol{\mathcal{G}}}_0\, \widehat{\boldsymbol{\mathcal{W}}}\, \boldsymbol{R}_{\boldsymbol{s}}[0; \boldsymbol{d}_0]^{-1/2}$ and $\widehat{\boldsymbol{H}}^N(z) = \iota^{-1}\left(\widehat{\boldsymbol{H}}^N\right)$.

output: $\left[\widehat{\boldsymbol{H}^N(z)}\right] = \pi\left(\widehat{\boldsymbol{H}}^N(z)\right)$

Table 3.1: Simplified CFIA

counterpart of $\widehat{\boldsymbol{U}}$, and so on. At the end of step 1) in the new version, we have $\widetilde{\boldsymbol{U}} = \widehat{\boldsymbol{U}}\boldsymbol{\Omega}$, for some $\boldsymbol{\Omega} = \text{diag}\left(e^{i\omega_1}, \ldots, e^{i\omega_D}\right)$, because $\widetilde{\boldsymbol{u}}_d$ and $\widehat{\boldsymbol{u}}_d$ can only differ by a phase factor. Moreover, $\widetilde{\lambda}_d = \widehat{\lambda}_d$, since the eigenvalues cannot change. Thus, in step 2), we have

$$\begin{aligned}
\widetilde{\boldsymbol{\mathcal{G}}}_0 &= \widetilde{\boldsymbol{U}}\left(\widetilde{\boldsymbol{\Sigma}} - \sigma^2\boldsymbol{I}_D\right)^{1/2} \\
&= \widehat{\boldsymbol{U}}\boldsymbol{\Omega}\left(\widehat{\boldsymbol{\Sigma}} - \sigma^2\boldsymbol{I}_D\right)^{1/2} \\
&= \widehat{\boldsymbol{U}}\left(\widehat{\boldsymbol{\Sigma}} - \sigma^2\boldsymbol{I}_D\right)^{1/2}\boldsymbol{\Omega} \\
&= \widehat{\boldsymbol{\mathcal{G}}}_0\boldsymbol{\Omega}. \qquad (3.9)
\end{aligned}$$

Notice that $\boldsymbol{\Omega}$ commutes with $\left(\widehat{\boldsymbol{\Sigma}} - \sigma^2\boldsymbol{I}_D\right)^{1/2}$ because they are both $D \times D$ diagonal matrices. As a consequence of (3.9), $\widetilde{\boldsymbol{\mathcal{G}}}_0^+ = \overline{\boldsymbol{\Omega}}\,\widehat{\boldsymbol{\mathcal{G}}}_0^+$, and

$$\widetilde{\boldsymbol{\Upsilon}}[m] = \overline{\boldsymbol{\Omega}}\,\widehat{\boldsymbol{\Upsilon}}[m]\,\boldsymbol{\Omega}, \qquad (3.10)$$

for $m = 1, \ldots, M$. To attack step 3), we start by noting that, as can be easily seen, the matrix $\widehat{\boldsymbol{\mathcal{W}}}_p$ in table 3.1 can also be obtained as follows: compute

$$\widehat{\boldsymbol{X}}_p = \arg \min_{\|\boldsymbol{X}\| = 1} \widehat{f}\,(\boldsymbol{X})\,,$$

where

$$\widehat{f}\,(\boldsymbol{X}) = \sum_{m=1,\ldots,M} \left\|\widehat{\boldsymbol{\Upsilon}}[m]\boldsymbol{X} - \boldsymbol{X}\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; D_0]\right\|^2 + \left\|\widehat{\boldsymbol{\Upsilon}}[m]^H\boldsymbol{X} - \boldsymbol{X}\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; D_0]^H\right\|^2,$$

and find $\delta_p \in \mathbb{R}$ such that $\widehat{\boldsymbol{\mathcal{W}}}_p = \sqrt{D_0+1}\,\widehat{\boldsymbol{X}}_p\,e^{i\delta_p}$ satisfies

$$\text{Re}\left\{\boldsymbol{s}_p^H\text{vec}\left(\widehat{\boldsymbol{\mathcal{W}}}_p\right)\right\} > 0, \quad \text{Im}\left\{\boldsymbol{s}_p^H\text{vec}\left(\widehat{\boldsymbol{\mathcal{W}}}_p\right)\right\} = 0.$$

Thus, in what respects the new version of CFIA, we have

$$\widetilde{\boldsymbol{X}}_p = \arg \min_{\|\boldsymbol{X}\| = 1} \widetilde{f}\,(\boldsymbol{X})\,.$$

Since $\widetilde{f}\,(\boldsymbol{X}) = \widehat{f}\,(\boldsymbol{\Omega}\boldsymbol{X})$ (use (3.10) and the fact that $\boldsymbol{\Omega}$ is unitary), we must have

$$\boldsymbol{\Omega}\widetilde{\boldsymbol{X}}_p = \boldsymbol{X}_p\,e^{i\theta_p}, \qquad (3.11)$$

for some $\theta_p \in \mathbb{R}$, because the left-hand side of (3.11) is also an unit-norm global minimizer of the quadratic form $\widehat{f}$. This implies $\widetilde{\boldsymbol{\mathcal{W}}}_p = \overline{\boldsymbol{\Omega}}\,\widehat{\boldsymbol{\mathcal{W}}}_p\,e^{i\theta_p}$, where the θ_p's may have changed relative to (3.11) due to the new reference vectors. Thus, in a more compact form,

$$\widetilde{\boldsymbol{\mathcal{W}}} = \overline{\boldsymbol{\Omega}}\,\widehat{\boldsymbol{\mathcal{W}}}\,\boldsymbol{\Delta}, \qquad (3.12)$$

where $\boldsymbol{\Delta} = \boldsymbol{\Theta}(\boldsymbol{\theta}) \otimes \boldsymbol{I}_{D_0+1}$, with $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_P)^T$ and $\boldsymbol{\Theta}(\boldsymbol{\theta}) = \text{diag}\left(e^{i\theta_1}, \ldots, e^{i\theta_P}\right)$. This means that in step 4) of the new version

$$\begin{aligned}
\widetilde{\boldsymbol{H}}^N &= \widetilde{\boldsymbol{\mathcal{G}}}_0\,\widetilde{\boldsymbol{\mathcal{W}}}\,\boldsymbol{R}_{\boldsymbol{s}}[0; \boldsymbol{d}_0]^{-1/2} \\
&= \widehat{\boldsymbol{\mathcal{G}}}_0\,\widehat{\boldsymbol{\mathcal{W}}}\,\boldsymbol{\Delta}\,\boldsymbol{R}_{\boldsymbol{s}}[0; \boldsymbol{d}_0]^{-1/2} \qquad (3.13) \\
&= \widehat{\boldsymbol{\mathcal{G}}}_0\,\widehat{\boldsymbol{\mathcal{W}}}\,\boldsymbol{R}_{\boldsymbol{s}}[0; \boldsymbol{d}_0]^{-1/2}\,\boldsymbol{\Delta} \qquad (3.14) \\
&= \widehat{\boldsymbol{H}}^N\,\boldsymbol{\Delta}. \qquad (3.15)
\end{aligned}$$

To establish (3.13), we used (3.9) and (3.12). Equation (3.14) is valid because $\boldsymbol{\Delta}$ and $\boldsymbol{R_s}[0;\boldsymbol{d}_0]^{-1/2}$ commute: they share the same block diagonal structure, with the pth block of $\boldsymbol{\Delta}$ being $e^{i\theta_p}\boldsymbol{I}_{D_0+1}$ (which commutes with any square matrix of the same size). Now, the equality in (3.15) implies $\widetilde{\boldsymbol{H}}^N(z) = \widehat{\boldsymbol{H}}^N(z)\boldsymbol{\Theta}(\boldsymbol{\theta})$. Thus, $\widetilde{\boldsymbol{H}}^N(z) \sim \widehat{\boldsymbol{H}}^N(z)$ and both project to the same point in the quotient $\mathbb{H}[z]/\sim$. In other words, both versions of the CFIA deliver the same output.

Approach outline. We address the problem of finding an expression for (3.1) through asymptotic analysis. More precisely, we activate the large-sample regime, that is, we let $N \to \infty$, and find the asymptotic distribution of the statistic $\left[\widehat{\boldsymbol{H}}^N(z)\right]$ in the quotient space $\mathbb{H}[z]/\sim$. Truncating the asymptotic study at any given N yields a distribution which is only an approximation to the true (exact) distribution. However, the asymptotic distribution can be found in closed-form (in contrast to the exact one, which seems very difficult to tame analytically) and it is easy to evaluate the expectation operator appearing in (3.1) with respect to this asymptotic distribution. Moreover, it turns out that the approximation for $J\left[N;\boldsymbol{c}(z),[\boldsymbol{H}(z)]\right]$ thus obtained is already tight for small N, as shown through numerical computer simulations in section 3.5.

We now give a macroscopic view of our asymptotic analysis. Four main phases can be identified and we make the whole theory tick as follows. **i)** We start by turning the quotient space $\mathbb{H}[z]/\sim$ into a differentiable manifold. This differential structure is introduced by identifying $\mathbb{H}[z]/\sim$ with the space of orbits generated by a Lie group action on $\mathbb{H}[z]$. Then, we induce in a natural way a geometry on the quotient space by equipping it with a Riemannian tensor. With this added structure, $\mathbb{H}[z]/\sim$ is a connected Riemannian manifold and the concept of distance between any two of its points is available (the function d mentioned in (3.1)). Note that $\mathbb{H}[z]$ has a canonical Riemannian manifold structure which comes from its identification with $\mathbb{C}^*_{\boldsymbol{d}_0}$. The geometry on the quotient space is induced by requiring the projection map $\pi : \mathbb{H}[z] \to \mathbb{H}[z]/\sim$ to be a Riemannian submersion. In loose terms, we make our choices in order to have the two geometries interfacing nicely through their natural link π. **ii)** The next step is to view CFIA in table 3.1 as a mapping $\text{CFIA} : \mathcal{U} \subset \mathbb{C}^{Q\times Q} \times \cdots \times \mathbb{C}^{Q\times Q} \to \mathbb{H}[z]/\sim$, which sends an ordered $(M+1)$-tuple of $Q \times Q$ complex matrices lying in a certain open set $\mathcal{U}$ to a MIMO channel equivalence class. We proceed to show that, in fact, CFIA is smooth as a mapping between manifolds at the point $(\boldsymbol{R_y}[0], \boldsymbol{R_y}[1], \ldots, \boldsymbol{R_y}[M])$, where $\boldsymbol{R_y}[m] = \mathrm{E}\left\{\boldsymbol{y}[n]\boldsymbol{y}[n-m]^H\right\}$. That is, the mapping CFIA is smooth on an open neighborhood of $(\boldsymbol{R_y}[0], \boldsymbol{R_y}[1], \ldots, \boldsymbol{R_y}[M])$. Moreover, we obtain its derivative at this point. This is achieved by writing $\text{CFIA} = \pi \circ \psi$, where $\psi : \mathcal{U} \subset \mathbb{C}^{Q\times Q} \times \cdots \times \mathbb{C}^{Q\times Q} \to \mathbb{H}[z]$ is the mapping which, given a $(M+1)$-tuple of $Q \times Q$ matrices, performs steps 1) to 4) of the CFIA. Since π is smooth (essentially by construction) it suffices to prove that ψ is smooth at $(\boldsymbol{R_y}[0], \boldsymbol{R_y}[1], \ldots, \boldsymbol{R_y}[M])$, in order to establish the smoothness of their composition $\text{CFIA} = \pi \circ \psi$. We further decompose ψ as a concatenation of 4 mappings, $\psi = \psi_4 \circ \psi_3 \circ \psi_2 \circ \psi_1$, where the ith mapping ψ_i corresponds to the ith step of the CFIA in table 3.1. Figure 3.2 puts in perspective all the mappings discussed here. **iii)** The next ingredient in our analysis consists in establishing the asymptotic normality of the ordered $(M+1)$-tuple of random matrices $\left(\widehat{\boldsymbol{R}}^N_{\boldsymbol{y}}[0], \widehat{\boldsymbol{R}}^N_{\boldsymbol{y}}[1], \ldots, \widehat{\boldsymbol{R}}^N_{\boldsymbol{y}}[M]\right)$ appearing at the CFIA's input. **iv)** The analysis is then

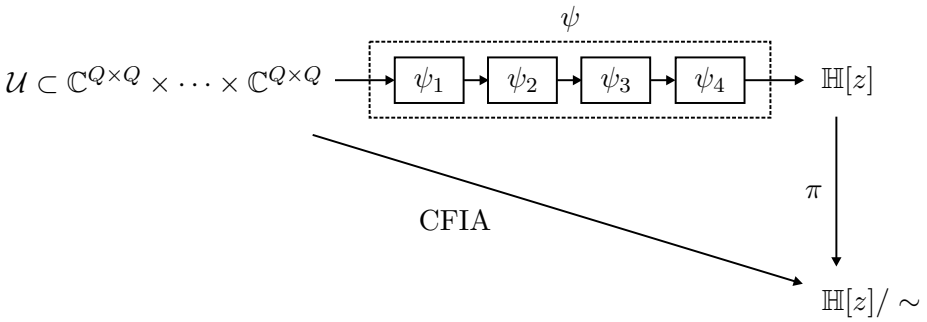

Figure 3.2: Mappings involved in the asymptotic analysis

closed by invoking a straightforward generalization of the delta-method to our manifold setting (the delta-method, for Euclidean spaces, is recalled later in more detail). Basically, this extension asserts that smooth maps between manifolds transform asymptotic normal inputs into asymptotic normal outputs. Thus, since $\left(\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[0], \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[1], \ldots, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[M]\right)$ is asymptotically normal with center $(\boldsymbol{R}_{\boldsymbol{y}}[0], \boldsymbol{R}_{\boldsymbol{y}}[1], \ldots, \boldsymbol{R}_{\boldsymbol{y}}[M])$ and the mapping CFIA is smooth at $(\boldsymbol{R}_{\boldsymbol{y}}[0], \boldsymbol{R}_{\boldsymbol{y}}[1], \ldots, \boldsymbol{R}_{\boldsymbol{y}}[M])$, the statistic

$$\widehat{[\boldsymbol{H}^{N}(z)]} = \mathrm{CFIA}\left(\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[0], \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[1], \ldots, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[M]\right),$$

behaves in the limit ($N \to \infty$) as a normal variable in the quotient $\mathbb{H}[z]/\sim$, with center $[\boldsymbol{H}(z)] = \mathrm{CFIA}\left(\boldsymbol{R}_{\boldsymbol{y}}[0], \boldsymbol{R}_{\boldsymbol{y}}[1], \ldots, \boldsymbol{R}_{\boldsymbol{y}}[M]\right)$. Moreover, its asymptotic covariance can be found in closed-form and it yields an immediate approximation for

$$\mathrm{E}\left\{d\left([\boldsymbol{H}(z)], \widehat{[\boldsymbol{H}^{N}(z)]}\right)^{2}\right\},$$

for any given N.

3.3 Differential-geometric framework

The purpose of this section is to define, in an intrinsic manner, asymptotic normality of random sequences on Riemannian manifolds and develop the corresponding extension of the delta-method to this setting. This is accomplished incrementally. First, coordinate-free generalizations of familiar statistical concepts in Euclidean spaces are obtained for finite-dimensional vector spaces. Then, the theoretical jump to Riemannian spaces is made by exploiting their natural local identifications with linear tangent spaces via the exponential mapping.

Euclidean spaces: random objects. Throughout this section, we assume fixed a probability space $(\Omega, \mathcal{A}, \mu)$, where Ω denotes the sample space, $\mathcal{A}$ is a σ-algebra, and μ stands for a probability measure. When we think of each Euclidean space $\mathbb{R}^m$ as a measure space, we let the Borel sets denote its σ-algebra and associate to it the Lesbesgue measure. A random vector in $\mathbb{R}^m$ is a measurable mapping $\boldsymbol{x} : \Omega \to \mathbb{R}^m$, $\omega \to \boldsymbol{x}(\omega)$. Let $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)^T$ denote a random vector in $\mathbb{R}^m$. We denote by $F_{\boldsymbol{x}} : \mathbb{R}^m \to \mathbb{R}$ its distribution function, that is, $F_{\boldsymbol{x}}(\boldsymbol{t}) = \mathrm{Prob}\{\boldsymbol{x} \preceq \boldsymbol{t}\}$. Here, and for further reference, the notation $\boldsymbol{a} \preceq \boldsymbol{b}$, where $\boldsymbol{a} = (a_1, \ldots, a_m)^T \in \mathbb{R}^m$ and $\boldsymbol{b} = (b_1, \ldots, b_m)^T \in \mathbb{R}^m$, means

that $a_i \leq b_i$ for all i. The notation $\boldsymbol{a} \prec \boldsymbol{b}$ means that we have strict inequality, that is, $a_i < b_i$ for all i. Similar definitions hold for $\succeq$ and $\succ$. Let $\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots$ denote a sequence of random vectors in $\mathbb{R}^m$. The sequence $\boldsymbol{x}_n$ is said to converge in distribution to the random vector $\boldsymbol{x}$, written $\boldsymbol{x}_n \xrightarrow{d} \boldsymbol{x}$, if $F_{\boldsymbol{x}_n}(\boldsymbol{t}) \to F_{\boldsymbol{x}}(\boldsymbol{t})$ as $n \to \infty$, for all $\boldsymbol{t}$ at which $F_{\boldsymbol{x}}$ is continuous. The sequence $\boldsymbol{x}_n$ is said to converge to $\boldsymbol{x}$ in probability, written $\boldsymbol{x}_n \xrightarrow{P} \boldsymbol{x}$, if for every $\epsilon > 0$, $\text{Prob}\{\|\boldsymbol{x}_n - \boldsymbol{x}\| > \epsilon\} \to 0$ as $n \to \infty$. The normal distribution in $\mathbb{R}^m$, with mean $\boldsymbol{\mu} \in \mathbb{R}^m$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times m}$ is denoted by $\mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$. Let $\boldsymbol{x} \in \mathbb{R}^m$ denote a random vector and let $\overset{d}{=}$ denote equality in distribution. We recall that $\boldsymbol{x} \overset{d}{=} \mathcal{N}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ if and only if $\boldsymbol{t}^T\boldsymbol{x} \overset{d}{=} \mathcal{N}(\boldsymbol{t}^T\boldsymbol{\mu}, \boldsymbol{t}^T\boldsymbol{\Sigma}\boldsymbol{t})$, for all $\boldsymbol{t} \in \mathbb{R}^m$. Let a_n denote a sequence of positive numbers converging to infinity. We say that the sequence of random vectors $\boldsymbol{x}_n$ is a_n-asymptotically normal with mean $\boldsymbol{\mu} \in \mathbb{R}^m$ and covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{m \times m}$, written $\boldsymbol{x}_n \sim a_n - \mathcal{AN}(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if $a_n(\boldsymbol{x}_n - \boldsymbol{\mu}) \xrightarrow{d} \mathcal{N}(\boldsymbol{0}, \boldsymbol{\Sigma})$. Random matrices are reverted to random vectors through the $\text{vec}(\cdot)$ operator. Thus, for example, $\boldsymbol{X}_n \xrightarrow{d} \boldsymbol{X}$, means that $\text{vec}(\boldsymbol{X}_n) \xrightarrow{d} \text{vec}(\boldsymbol{X})$. In terms of notation regarding matrix normal distributions, it is convenient to denote the mean in matrix form. More precisely, for a random matrix $\boldsymbol{X} \in \mathbb{R}^{n \times m}$, the notation $\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{\Upsilon}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Upsilon} \in \mathbb{R}^{n \times m}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{nm \times nm}$, means that $\text{vec}(\boldsymbol{X}) \sim \mathcal{N}(\text{vec}(\boldsymbol{\Upsilon}), \boldsymbol{\Sigma})$. The same notational principle applies to asymptotic normal distributions of random matrices.

Complex random objects are handled by embedding them in real Euclidean spaces and applying the previous definitions. We define three identifications which permit to embed complex vectors, ordered k-tuples of complex vectors, and ordered k-tuples of complex matrices. These are defined recursively as follows. We identify $\mathbb{C}^m$ with $\mathbb{R}^{2m}$ via the mapping $\imath : \mathbb{C}^m \to \mathbb{R}^{2m}$,

$$\imath(\boldsymbol{z}) = \begin{bmatrix} \text{Re}\, \boldsymbol{z} \\ \text{Im}\, \boldsymbol{z} \end{bmatrix}. \tag{3.16}$$

The remaining identifications to be used are: $\imath : \mathbb{C}^{m_1} \times \mathbb{C}^{m_2} \cdots \times \mathbb{C}^{m_k} \to \mathbb{R}^{2(m_1+m_2+\cdots+m_k)}$ given by

$$\imath(\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_k) = \begin{bmatrix} \imath(\boldsymbol{z}_1) \\ \imath(\boldsymbol{z}_2) \\ \vdots \\ \imath(\boldsymbol{z}_k) \end{bmatrix},$$

and $\imath : \mathbb{C}^{n_1 \times m_1} \times \mathbb{C}^{n_2 \times m_2} \cdots \times \mathbb{C}^{n_k \times m_k} \to \mathbb{R}^{2(n_1 m_1 + n_2 m_2 + \cdots + n_k m_k)}$ defined by

$$\imath(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_k) = \imath(\text{vec}(\boldsymbol{Z}_1), \text{vec}(\boldsymbol{Z}_2), \ldots, \text{vec}(\boldsymbol{Z}_k)). \tag{3.17}$$

Notice that the symbol $\imath$ designate all identifications. It suffices to look at the argument to sort out the ambiguity. In the sequel, these identifications are implicitly used whenever complex random objects are involved. As illustrative examples: **i)** saying that $\boldsymbol{z} \in \mathbb{C}^m$ is random means that $\imath(\boldsymbol{z})$ is random; **ii**) saying that

$$(\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_k) \sim \mathcal{N}((\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_k), \boldsymbol{\Sigma}),$$

where $\boldsymbol{z}_j, \boldsymbol{\mu}_j \in \mathbb{C}^{m_j}$ and $\boldsymbol{\Sigma} \in \mathbb{R}^{2(m_1+m_2+\cdots+m_k) \times 2(m_1+m_2+\cdots+m_k)}$, means that

$$\imath(\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_k) \sim \mathcal{N}(\imath(\boldsymbol{\mu}_1, \boldsymbol{\mu}_2, \ldots, \boldsymbol{\mu}_k), \boldsymbol{\Sigma});$$

iii) let a_n denote a sequence of positive numbers converging to infinity. Saying that a sequence (indexed by n) of ordered k-tuples of complex random matrices $(\boldsymbol{Z}_1^n, \boldsymbol{Z}_2^n, \ldots, \boldsymbol{Z}_k^n)$, where $\boldsymbol{Z}_j^n \in \mathbb{C}^{n_j \times m_j}$, is a_n-asymptotically normal with mean $(\boldsymbol{\Upsilon}_1, \boldsymbol{\Upsilon}_2, \ldots, \boldsymbol{\Upsilon}_k)$, where $\boldsymbol{\Upsilon}_j \in \mathbb{C}^{n_j \times m_j}$, and covariance $\boldsymbol{\Sigma} \in \mathbb{R}^{2(n_1m_1+n_2m_2+\cdots+n_km_k)\times 2(n_1m_1+n_2m_2+\cdots+n_km_k)}$, written

$$(\boldsymbol{Z}_1^n, \boldsymbol{Z}_2^n, \ldots, \boldsymbol{Z}_k^n) \sim a_n - \mathcal{AN}\left((\boldsymbol{\Upsilon}_1, \boldsymbol{\Upsilon}_2, \ldots, \boldsymbol{\Upsilon}_k), \boldsymbol{\Sigma}\right),$$

means that

$$\imath(\boldsymbol{Z}_1^n, \boldsymbol{Z}_2^n, \ldots, \boldsymbol{Z}_k^n) \sim a_n - \mathcal{AN}\left(\imath(\boldsymbol{\Upsilon}_1, \boldsymbol{\Upsilon}_2, \ldots, \boldsymbol{\Upsilon}_k), \boldsymbol{\Sigma}\right).$$

It is worth mentioning some basic properties of the embeddings, which will prove to be handy in the sequel. We state them without proof, since all of them are rather trivial to check. For $\boldsymbol{Z} \in \mathbb{C}^{n\times m}$, $\boldsymbol{z}_j \in \mathbb{C}^{m_j}$ and conformable matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, we have:

$$\begin{aligned}
\imath\left(\overline{\boldsymbol{Z}}\right) &= \boldsymbol{A}_{n,m}\,\imath(\boldsymbol{Z}) && (3.18)\\
\imath\left(\boldsymbol{Z}^T\right) &= \boldsymbol{B}_{n,m}\,\imath(\boldsymbol{Z}) && (3.19)\\
\imath\left(\boldsymbol{Z}^H\right) &= \boldsymbol{C}_{n,m}\,\imath(\boldsymbol{Z}) && (3.20)\\
\imath(\boldsymbol{A}\boldsymbol{z}) &= \jmath(\boldsymbol{A})\,\imath(\boldsymbol{z}) && (3.21)\\
\imath(\boldsymbol{A}\boldsymbol{Z}\boldsymbol{B}) &= \jmath\left(\boldsymbol{B}^T \otimes \boldsymbol{A}\right)\,\imath(\boldsymbol{Z}) && (3.22)\\
\imath\left(\begin{bmatrix} \boldsymbol{z}_1 \\ \vdots \\ \boldsymbol{z}_k \end{bmatrix}\right) &= \boldsymbol{\Pi}_{m_1,\ldots,m_k} \begin{bmatrix} \imath(\boldsymbol{z}_1) \\ \vdots \\ \imath(\boldsymbol{z}_k) \end{bmatrix}. && (3.23)
\end{aligned}$$

Here, and for further reference, we used the notation $\boldsymbol{A}_{n,m} = \operatorname{diag}(\boldsymbol{I}_{nm}, -\boldsymbol{I}_{nm})$ and $\boldsymbol{B}_{n,m} = \operatorname{diag}(\boldsymbol{K}_{n,m}, \boldsymbol{K}_{n,m})$, where the symbol $\boldsymbol{K}_{n,m}$ denotes the commutation matrix of size $nm \times nm$ [39]. It is a permutation matrix which acts on $n \times m$ matrices $\boldsymbol{A}$ as $\boldsymbol{K}_{n,m}\operatorname{vec}(\boldsymbol{A}) = \operatorname{vec}\left(\boldsymbol{A}^T\right)$. The notation $\boldsymbol{K}_n = \boldsymbol{K}_{n,n}$ is also used. Moreover, we define $\boldsymbol{C}_{n,m} = \operatorname{diag}(\boldsymbol{K}_{n,m}, -\boldsymbol{K}_{n,m})$, and

$$\jmath(\boldsymbol{A}) = \begin{bmatrix} \operatorname{Re}\boldsymbol{A} & -\operatorname{Im}\boldsymbol{A} \\ \operatorname{Im}\boldsymbol{A} & \operatorname{Re}\boldsymbol{A} \end{bmatrix}.$$

The matrix $\boldsymbol{\Pi}_{m_1,\ldots,m_k}$ denotes a permutation matrix which is implicitly (and uniquely) defined by (3.23).

Derivatives of complex mappings. Let $f : \mathbb{R}^m \to \mathbb{R}^p$, $\boldsymbol{x} = (x_1, x_2, \ldots, x_m)^T \mapsto f(\boldsymbol{x}) = (f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_p(\boldsymbol{x}))^T$ denote a real differentiable mapping. We recall that its derivative at $\boldsymbol{x}_0$ is given by the $p \times m$ matrix

$$Df(\boldsymbol{x}_0) = \begin{bmatrix} \frac{\partial f_1}{\partial x_1}(\boldsymbol{x}_0) & \cdots & \frac{\partial f_1}{\partial x_m}(\boldsymbol{x}_0) \\ \vdots & \ddots & \vdots \\ \frac{\partial f_p}{\partial x_1}(\boldsymbol{x}_0) & \cdots & \frac{\partial f_p}{\partial x_m}(\boldsymbol{x}_0) \end{bmatrix}.$$

The mapping f is said to be continuously differentiable at $\boldsymbol{x}_0$ if it is continuously differentiable in some open neighborhood of $\boldsymbol{x}_0$. Complex mappings are handled in a natural way through the identifications discussed above. For example, consider the mapping $f : \mathbb{C}^m \to \mathbb{C}^p$. Then, there is an unique mapping $\widehat{f} : \mathbb{R}^{2m} \to \mathbb{R}^{2p}$ such that the diagram

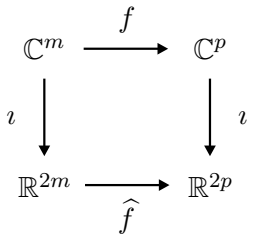

Figure 3.3: The complex mapping f induces the real mapping $\widehat{f}$

in figure 3.3 commutes, that is, $\widehat{f} \circ \imath = \imath \circ f$. The mapping f is said to be differentiable if $\widehat{f}$ is. In such case, its derivative at $\boldsymbol{z}_0 \in \mathbb{C}^m$ is defined as $D\widehat{f}(\imath(\boldsymbol{z}_0))$. The derivatives of mappings $f : \mathbb{C}^{m_1} \times \cdots \times \mathbb{C}^{m_k} \to \mathbb{C}^{p_1} \times \cdots \times \mathbb{C}^{p_l}$ and

$$F : \mathbb{C}^{n_1 \times m_1} \times \cdots \times \mathbb{C}^{n_k \times m_k} \to \mathbb{C}^{p_1 \times q_1} \times \cdots \times \mathbb{C}^{p_l \times q_l} \tag{3.24}$$

are defined similarly by working with the appropriate embeddings. With these definitions, the chain rule continues to hold. For example, suppose that $f : \mathbb{C}^m \to \mathbb{C}^p$ and $g : \mathbb{C}^p \to \mathbb{C}^q$, and let $h = g \circ f$. Then, for $\boldsymbol{z} \in \mathbb{C}^m$, we have $Dh(\boldsymbol{z}) = Dg\left(f(\boldsymbol{z})\right) Df(\boldsymbol{z})$. To establish this, simply notice that $\widehat{h} = \widehat{g} \circ \widehat{f}$. Furthermore, we can generalize the delta-method in a straightforward manner. We recall that, for Euclidean spaces, the delta-method asserts that if $f : \mathbb{R}^m \to \mathbb{R}^p$ is smooth on an open neighborhood of $\boldsymbol{\mu}$ and $\boldsymbol{x}_n \sim a_n - \mathcal{AN}\left(\boldsymbol{\mu}, \boldsymbol{\Sigma}\right)$, then $f(\boldsymbol{x}_n) \sim a_n - \mathcal{AN}\left(f(\boldsymbol{\mu}), Df(\boldsymbol{\mu})\boldsymbol{\Sigma}Df(\boldsymbol{\mu})^T\right)$, see [53, corollary 1.1, page 45]. Now, let F be as in (3.24), and suppose it is continuously differentiable at $(\boldsymbol{\Upsilon}_1, \ldots, \boldsymbol{\Upsilon}_k)$, where $\boldsymbol{\Upsilon}_j \in \mathbb{C}^{n_j \times m_j}$. Let a_n denote a sequence of positive numbers converging to infinity and suppose

$$\left(\boldsymbol{Z}_1^n, \ldots, \boldsymbol{Z}_k^n\right) \sim a_n - \mathcal{AN}\left(\left(\boldsymbol{\Upsilon}_1, \ldots, \boldsymbol{\Upsilon}_k\right), \boldsymbol{\Sigma}\right).$$

Then,

$$F\left(\boldsymbol{Z}_1^n, \ldots, \boldsymbol{Z}_k^n\right) \sim a_n - \mathcal{AN}\left(F\left(\boldsymbol{\Upsilon}_1, \ldots, \boldsymbol{\Upsilon}_k\right), DF\left(\boldsymbol{\Upsilon}_1, \ldots, \boldsymbol{\Upsilon}_k\right) \boldsymbol{\Sigma} DF\left(\boldsymbol{\Upsilon}_1, \ldots, \boldsymbol{\Upsilon}_k\right)^T\right).$$

Linear spaces: general definitions. Let V denote a finite-dimensional vector space over $\mathbb{R}$, with $\dim V = m$. We denote by V^* its dual space, that is, the set of linear functionals from V to $\mathbb{R}$. The elements of V^* are called covectors. To each basis $E_1, \ldots, E_m$ of V corresponds a unique dual basis $\omega_1, \ldots, \omega_m$ of V^* such that $\omega_i(E_j) = \delta[i - j]$. A tensor Φ on V of type (k, l) is a multilinear map

$$\Phi : \underbrace{V \times \cdots \times V}_{k \text{ copies}} \times \underbrace{V^* \times \cdots \times V^*}_{l \text{ copies}} \to \mathbb{R}.$$

Such a map is also called a k-covariant, l-contravariant tensor on V, and we define its rank as $k + l$. Therefore, the rank of a tensor is the total number of arguments that it takes. The space of all (k, l) tensors on V is denoted by $T_l^k(V)$. To simplify notation, we let $T^k(V) = T_0^k(V)$ and $T_l(V) = T_l^0(V)$. We have the identifications $T^1(V) = V^*$ and $T_1(V) = V^{**} = V$. Moreover, $\mathrm{End}(V)$, the set of linear mappings from V to V, can be identified with $T_1^1(V)$ by assigning to $A : V \to V$, the $(1, 1)$ tensor Φ_A acting on pairs $(X, \sigma) \in V \times V^*$ as $\Phi_A(X, \sigma) = \sigma(AX)$. This identification is used to define the trace of a

$(1,1)$ tensor Φ, written $\operatorname{tr}\Phi$, as the trace of its corresponding linear mapping in $\mathrm{End}(V)$. With these definitions, it is easily seen that if $E_1,\ldots,E_m$ and $\omega_1,\ldots,\omega_m$ denote dual basis in V and V^*, respectively, then

$$\operatorname{tr}\Phi = \sum_{i=1}^{m} \Phi(E_i,\omega_i).$$

The space $T_l^k(V)$ has a natural vector space structure. For tensors $\Phi,\Psi \in T_l^k(V)$, and scalars $\alpha,\beta\in\mathbb{R}$, the (k,l) tensor $\alpha\Phi+\beta\Psi$ is defined by

$$\begin{aligned}&(\alpha\Phi+\beta\Psi)\,(X_1,\ldots,X_k,\sigma_1,\ldots,\sigma_l) = \\ &\quad \alpha\Phi(X_1,\ldots,X_k,\sigma_1,\ldots,\sigma_l)+\beta\Psi(X_1,\ldots,X_k,\sigma_1,\ldots,\sigma_l),\end{aligned}$$

where $X_i\in V$ and $\sigma_j\in V^*$. Let V and W denote finite-dimensional vector spaces. A linear mapping $A: V\to W$ can be used to pull back covariant tensors on W and push forward contravariant tensors on V. More precisely, for given covariant order k and contravariant order l, we have the linear pull back mapping $A^*: T^k(W)\to T^k(V)$, $\Phi\mapsto A^*\Phi$, where

$$(A^*\Phi)\,(X_1,\ldots,X_k) = \Phi\,(AX_1,\ldots,AX_k)\,, \tag{3.25}$$

for $X_i\in V$, and the linear push forward mapping $A_*: T_l(V)\to T_l(W)$, $\Phi\mapsto A_*\Phi$, where

$$(A_*\Phi)\,(\sigma_1,\ldots,\sigma_l) = \Phi\,(A^*\sigma_1,\ldots,A^*\sigma_l)\,, \tag{3.26}$$

for $\sigma_j\in W^*$. An inner product on V is a bilinear form g on V (equivalently, a $(2,0)$ tensor on V), which is symmetric ($g(X,Y)=g(Y,X)$) and positive definite ($g(X,X)\geq 0$, with equality if and only if $X=0$). The existence of an inner-product g on V induces many constructions. The length of a vector $X\in V$, written $|X|$, is defined as $|X|=\sqrt{g(X,X)}$. The vectors $X_1,\ldots,X_k$ are said to be orthonormal if $g(X_i,X_j)=\delta[i-j]$. The inner product also provides a natural identification between V and its dual V^*, as follows. To each vector $X\in V$ we associate the covector $X^\flat\in V^*$ defined by

$$X^\flat(Y) = g(Y,X), \tag{3.27}$$

for $Y\in V$. The notation $X^\flat = g(\cdot,X)$ is also used. This identification $V\to V^*$ is one-to-one and onto, hence invertible. Therefore, to each covector $\sigma\in V^*$ corresponds a unique vector in V, which we denote $\sigma^\sharp$. In this manner, an inner product is introduced in the dual space V^*, denoted $g^\flat: V^*\times V^*\to\mathbb{R}$, by letting $g^\flat(\sigma,\omega)=g(\sigma^\sharp,\omega^\sharp)$. Also, we can use these identifications to convert a given tensor on V on any other of the same rank. An important example, that we shall use in the sequel, consists in converting $\Phi\in T_2(V)$ in $\Phi^\flat\in T_1^1(V)$ given by

$$\Phi^\flat(X,\sigma) = \Phi(X^\flat,\sigma),$$

for $(X,\sigma)\in V\times V^*$. Since $\Phi^\flat$ is a $(1,1)$ tensor its trace is defined, and we also call it the trace of Φ, that is, we define $\operatorname{tr}\Phi=\operatorname{tr}\Phi^\flat$. With these definitions, it easily seen that, if $\omega_1,\ldots,\omega_m$ denotes an orthonormal basis for V^* and $\Phi\in T_2(V)$, then

$$\operatorname{tr}\Phi = \sum_{i=1}^{m}\Phi\,(\omega_i,\omega_i)\,. \tag{3.28}$$

Linear spaces: random objects. Let V denote an m-dimensional vector space over $\mathbb{R}$. The vector space V acquires a topology by identifying it with $\mathbb{R}^m$ through a choice of basis. This topology is well-defined, that is, it does not depend on which particular basis of V is chosen. We think of V as a measurable space by equipping it with the σ-algebra generated by its class of open sets, that is, by its topology. A random vector in V is a measurable mapping $X : \Omega \to V$, $\omega \mapsto X(\omega)$. With this definition, it follows that for a given covector $\sigma \in V^*$, $\sigma(X)$ denotes a random variable in $\mathbb{R}$, that is, the mapping $\omega \in \Omega \mapsto \sigma(X(\omega)) \in \mathbb{R}$ is measurable, because it is a composition of measurable mappings. The random vector X is said to have mean $\mu \in V$ if $\mathrm{E}\{\sigma(X)\} = \sigma(\mu)$, for all $\sigma \in V^*$. Here, $\mathrm{E}\{\cdot\}$ denotes the expectation operator. The covariance of X is a symmetric 2-contravariant tensor on V, denoted $\mathrm{Cov}(X)$, and defined by $\mathrm{Cov}(X)(\sigma, \psi) = \mathrm{E}\{\sigma(X-\mu)\psi(X-\mu)\}$, for $(\sigma, \psi) \in V^* \times V^*$ and where μ denotes the mean value of X. A random vector X in V is said to have the normal distribution with mean $\mu \in V$ and covariance $\Sigma \in T_2(V)$, written $X \sim \mathcal{N}(\mu, \Sigma)$, if for any given $\sigma \in V^*$ the random variable $\sigma(X)$ in $\mathbb{R}$ has the distribution $\mathcal{N}(\sigma(\mu), \Sigma(\sigma, \sigma))$. Let $X_1, X_2, \ldots$ denote a sequence of random vectors in V. We say that the sequence X_n converges in distribution to the random vector X, written $X_n \xrightarrow{d} X$, if for any $\sigma \in V^*$, the sequence of random variables $\sigma(X_n)$ converges in distribution to the random variable $\sigma(X)$. Similarly, the sequence X_n is said to converge in probability to $\mu \in V$, written $X_n \xrightarrow{P} \mu$, if for any $\sigma \in V^*$, we have $\sigma(X_n) \xrightarrow{P} \sigma(\mu)$. The sequence X_n converges in probability to the random vector X, written $X_n \xrightarrow{P} X$, if $X_n - X \xrightarrow{P} 0$. Let a_n denote a sequence of positive numbers converging to infinity. We say that the sequence of random vectors X_n is a_n-asymptotically normal with mean $\mu \in V$ and covariance $\Sigma \in T_2(V)$, written $X_n \sim a_n - \mathcal{AN}(\mu, \Sigma)$, if $a_n(X_n - \mu) \xrightarrow{d} \mathcal{N}(0, \Sigma)$. The previous definitions did not require an inner product on the vector space V. If such an inner product exists, we define the variance of the random vector X in V as $\mathrm{var}(X) = \mathrm{E}\left\{|X-\mu|^2\right\}$, where μ denotes the mean value of X. Is is straightforward to check that $\mathrm{var}(X) = \mathrm{tr}\,\mathrm{Cov}(X)$, as it should be.

Manifolds: general definitions. We assume the reader to be familiar with basic concepts of differential Riemannian geometry, such as those at the level of [7, 16, 32]. Here, we mainly settle notation which, whenever possible, is compatible with the notation in [7]. For a smooth (C^∞) manifold M, the tangent space to M at p is denoted by T_pM and the respective cotangent space by $T_p^*M = (T_pM)^*$. The symbol $T_l^k(M)$ stands for the bundle of mixed tensors of type (k, l) on M, and $\mathcal{T}_l^k(M)$ denotes the vector space of smooth sections of $T_l^k(M)$. Thus, a section $\Phi \in \mathcal{T}_l^k(M)$ assigns a tensor of type (k, l) to each tangent space T_pM, $p \in M$, denoted

$$\Phi_p : \underbrace{T_pM \times \cdots \times T_pM}_{k \text{ copies}} \times \underbrace{T_p^*M \times \cdots \times T_p^*M}_{l \text{ copies}} \to \mathbb{R}.$$

To simplify notation, we let $T^k(M) = T_0^k(M)$, $T_l(M) = T_l^0(M)$, $\mathcal{T}^k(M) = \mathcal{T}_0^k(M)$ and $\mathcal{T}_l(M) = \mathcal{T}_l^0(M)$. The tangent bundle of M is denoted by $TM = T_1M$. We assume TM to be equipped with its canonical smooth structure, and we let $\mathcal{T}(M) = \mathcal{T}_1(M)$ stand for the space of smooth sections of TM, that is, the set of smooth vector fields on M.

The cotangent bundle of M is denoted by $T^*M = T^1M$, and it is assumed equipped with its canonical smooth structure. The set of smooth covector fields on M, that is, the space of smooth sections of T^*M, is denoted by $\mathcal{T}^*(M)$. The set of smooth real-valued functions defined on M is denoted by $\mathrm{C}^\infty(M)$. If M and N denote smooth manifolds, and $F : M \to N$ is a smooth mapping, $F_* : TM \to TN$ denotes its derivative. We also use the symbol F_* to denote the push forward operator $F_* : T_l(M) \to T_l(N)$ induced by the derivative of F. That is, at each $p \in M$, we let the linear (derivative) mapping $F_* : T_pM \to T_{F(p)}N$ between linear spaces induce the corresponding linear push forward operator $F_* : T_l(T_pM) \to T_l\left(T_{F(p)}N\right)$ which acts on contravariant tensors on T_pM. Exterior differentiation is represented by d.

Let M denote a Riemannian manifold with metric g, denoted sometimes by (M, g). This means that g is a tensor field $g \in \mathcal{T}^2(M)$ which is symmetric ($g(X,Y) = g(Y,X)$, for all $X, Y \in \mathcal{T}M$) and positive definite ($g(X,X) \geq 0$ for all $X \in \mathcal{T}(M)$, with equality if and only if $X = 0$). Thus, g is a smooth assignment of an inner-product $g_p : T_pM \times T_pM \to \mathbb{R}$ to each tangent space T_pM, $p \in M$. For $X_p \in T_pM$, we use the notation $|X_p| = \sqrt{g_p(X_p, X_p)}$ to designate the norm of the vector X_p with respect to the inner-product g_p. We let ∇ denote simultaneously the Levi-Civita connection on M, and the induced connection on all bundles $T_l^k(M)$. The gradient of $f \in \mathrm{C}^\infty(M)$, denoted $\operatorname{grad} f$, is the unique smooth section of TM satisfying $df(X) = g(\operatorname{grad} f, X)$, for all $X \in \mathcal{T}(M)$. Moreover, the Hessian of $f \in \mathrm{C}^\infty(M)$, denoted $\operatorname{Hess} f$, is defined as $\operatorname{Hess} f = \nabla df$, see [16, example 2.64, page 74]. It is a symmetric form and belongs to $\mathcal{T}_0^2(M)$. The exponential mapping is given by $\operatorname{Exp} : \mathcal{D} \subset TM \to M$, where the open set $\mathcal{D}$ denotes its domain. We recall that, for a given tangent vector $X_p \in \mathcal{D} \subset TM$, the exponential mapping is defined as $\operatorname{Exp} X_p = \gamma(1)$, where $\gamma(t)$ designates the unique geodesic which emanates from $p \in M$ along the tangent direction X_p, that is, $\gamma(0) = p$ and $\dot{\gamma}(0) = X_p$ [7, definition 6.3, page 337]. For $p \in M$, we let Exp_p denote the restriction of Exp to the tangent space T_pM. A geodesic ball centered at $p \in M$ and radius $\epsilon > 0$, denoted $B_\epsilon(p)$, is the diffeomorphic image under Exp_p of the tangent ball

$$TB_\epsilon(p) = \{X_p \in T_pM \,:\, |X_p| < \epsilon\},$$

that is, $B_\epsilon(p) = \operatorname{Exp}_p(TB_\epsilon(p))$. Recall that, in general, geodesic balls with center p are only defined for all positive ϵ below a certain threshold (which depends on p), see [16, corollary 2.89, page 85]. We say that a smooth map $F : M \to T_pM$ is a linearization of M at p if it agrees with the inverse mapping of Exp on some geodesic ball $B_\epsilon(p)$. The length of a smooth curve $\gamma : [a,b] \to M$ is defined as

$$\ell(\gamma) = \int_a^b |\dot{\gamma}(t)|\, dt.$$

The curve γ is said to be regular if it is an immersion, that is, $\dot{\gamma}(t) \neq 0$. A continuous map $\gamma : [a,b] \to M$ is said to be a piecewise regular curve if there is a finite partition $a = a_0 < a_1 < \cdots < a_n = b$ such that the restriction of γ to each subinterval $[a_{i-1}, a_i]$ is a regular curve. In such case, the length of γ is the sum of the lengths of the regular subsegments. If M is connected, the Riemannian distance between two points $p, q \in M$, denoted $d(p,q)$, is defined as the infimum of the lengths of all piecewise regular curves

from p to q. Note that we use the same symbol d for both the exterior derivative and the geodesic distance. The context easily resolves the ambiguity. Let $\gamma : [a,b] \to M$ denote a smooth curve, and V a vector field along γ, that is, a smooth map $V : [a,b] \to TM$. We let $D_tV : [a,b] \to TM$ denote the covariant derivative of the vector field V along the curve γ. The vector field V is said to be parallel along γ if $D_tV \equiv 0$. Recall that the curve γ is a geodesic if its velocity vector $\dot\gamma : [a,b] \to TM$ is parallel, that is, $D_t\dot\gamma \equiv 0$. The Riemannian curvature tensor of M is denoted by R. It is the $\mathrm{C}^\infty(M)$-multilinear map $R : \mathcal{T}(M) \times \mathcal{T}(M) \times \mathcal{T}(M) \times \mathcal{T}(M) \to \mathbb{R}$ defined by

$$R(X,Y,Z,W) = g(\nabla_X\nabla_Y Z - \nabla_Y\nabla_X Z - \nabla_{[X,Y]}Z, W), \tag{3.29}$$

where $X,Y,Z,W \in \mathcal{T}(M)$ and $[X,Y] = XY - YX \in \mathcal{T}(M)$ stands for the Lie bracket of the vectors fields X and Y. Let X_p, Y_p denote linearly independent vectors in T_pM. The sectional curvature of the plane $\Pi = \mathrm{span}\{X_p, Y_p\} \subset T_pM$ is denoted by $K(X_p, Y_p)$ or $K(\Pi)$. Recall that

$$K(X_p,Y_p) = \frac{R(X_p,Y_p,Y_p,X_p)}{g(X_p,X_p)g(Y_p,Y_p) - g(X_p,Y_p)^2}. \tag{3.30}$$

Manifolds: random objects. Let M denote a connected Riemannian manifold. It is also a measurable space by letting its topology generate a σ-algebra, called the Borel σ-algebra of M. A random point in M is a measurable mapping $x : \Omega \to M$, $\omega \mapsto x(\omega)$. Let $x_1, x_2, \ldots$ denote a sequence of random points in M. The sequence x_n is said to converge in probability to the random point x, written $x_n \xrightarrow{P} x$, if the sequence of random variables $d(x_n, x)$ converges to zero in probability. Recall that $d(p,q)$ denotes the Riemannian distance from p to q. Let a_n a sequence of positive numbers converging to infinity. We say that the sequence of random points x_n is a_n-asymptotically normal with mean $p \in M$ and covariance form $\Sigma \in T_2(T_pM)$, written $x_n \sim a_n - \mathcal{AN}(p,\Sigma)$, if $x_n \xrightarrow{P} p$ and $F(x_n) \sim a_n - \mathcal{AN}(0,\Sigma)$ for all linearizations $F : M \to T_pM$ of M at p. Note that $F(x_n)$ denotes a random vector in the linear space T_pM. Lemma 3.2 shows that, in fact, the analysis of one linearization is sufficient to establish asymptotic normality of random sequences.

Lemma 3.2. *Suppose $x_n \xrightarrow{P} p$ and let F and G denote linearizations of M at p. Then,*

$$F(x_n) \sim a_n - \mathcal{AN}(0,\Sigma) \quad \Leftrightarrow \quad G(x_n) \sim a_n - \mathcal{AN}(0,\Sigma).$$

Proof: See appendix B.

If x_n is a_n-asymptotically normal with mean p and covariance form Σ, we have lemma 3.3 which connects the asymptotic distribution of the random square-distance $d(x_n,p)^2$ to the covariance Σ.

Lemma 3.3. *Suppose $x_n \sim a_n - \mathcal{AN}(p,\Sigma)$. Then,*

$$a_n^2\, d(x_n,p)^2 \xrightarrow{d} z,$$

where $\mathrm{E}\{z\} = \mathrm{tr}\,\Sigma$.

Proof: See appendix B.

Thus, if $x_n \sim a_n - \mathcal{AN}(p, \Sigma)$, lemma 3.3 suggests the approximation $\mathrm{E}\left\{a_n^2\, d(x_n, p)^2\right\} \simeq \mathrm{E}\{z\}$, that is,

$$\mathrm{E}\left\{d(x_n, p)^2\right\} \simeq \frac{\operatorname{tr}\Sigma}{a_n^2}, \tag{3.31}$$

for large n. It is an important fact that asymptotic normality is preserved by smooth mappings of manifolds. Lemma 3.4 generalizes to the setting of Riemannian manifolds the well known delta-method for Euclidean spaces.

Lemma 3.4. *Let M, N denote Riemannian manifolds and $F : M \to N$ a smooth map. Suppose $x_n \sim a_n - \mathcal{AN}(p, \Sigma)$. Then, $F(x_n) \sim a_n - \mathcal{AN}(F(p), F_*\Sigma)$.*

Proof: See appendix B.

Note that the result still holds if F is only smooth on an open neighborhood U of p, because it suffices to replace U by M in the lemma (a look at the proof clarifies this point even better).

3.4 Performance analysis: microscopic view

In this section, we zoom in our performance analysis and work out the details of its four main phases. The four phases were defined in page 57. We devote a subsection to each phase. **i)** In subsection 3.4.1, we equip the quotient space $\mathbb{H}[z]/\sim$ with a Riemannian manifold structure and dissect a bit its geometry. **ii)** In subsection 3.4.2, we concern ourselves with the map ψ, which, essentially, performs steps 1) to 4) of the CFIA. We define it, show that it is a smooth map and compute its derivative at any given point. **iii)** In subsection 3.4.3, we establish the asymptotic ($N \to \infty$) normality of the random ordered P-tuple $\left(\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^N[0], \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^N[1], \ldots, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^N[M]\right)$ appearing at the input of the CFIA. A closed-form expression is obtained for its asymptotic covariance. **iv)** Subsection 3.4.4 is the last one. It assembles all the analytical pieces developed so far and, together with results from the differential-geometric framework in section 3.3, proves the asymptotic ($N \to \infty$) normality of the random point $[\widehat{\boldsymbol{H}^N(z)}]$ in the Riemannian manifold $\mathbb{H}[z]/\sim$. Furthermore, its asymptotic covariance is obtained in closed-form and an approximation to $J\left[N\,;\, \boldsymbol{c}(z), [\boldsymbol{H}(z)]\right] = \mathrm{E}\left\{d\left([\boldsymbol{H}(z)]\,,\, [\widehat{\boldsymbol{H}^N(z)}]\right)^2\right\}$ in (3.1) is deduced.

3.4.1 The geometry of $\mathbb{H}[z]/\sim$

In this subsection, we turn the quotient space $\mathbb{H}[z]/\sim$ into a Riemannian manifold. This smooth geometric structure is induced naturally by requiring the projection map $\pi : \mathbb{H}[z] \to \mathbb{H}[z]/\sim$ to be a Riemannian submersion. That is, we make the geometries on the two spaces $\mathbb{H}[z]$ and $\mathbb{H}[z]/\sim$, linked canonically by π, to interface nicely. As soon as the space $\mathbb{H}[z]/\sim$ acquires its Riemannian structure, it becames a geometrical object in its own right (for example, the concept of distance becomes available). We proceed to study its main geometric features. More precisely, we develop a closed-form expression for the

distance between any two of its points, and obtain its sectional curvatures at any given point. This data will be used in the sequel: the distance is needed later in this chapter, whereas an upper-bound on the sectional curvatures is required in chapter 4.

The Riemannian manifold $\mathbb{H}[z]$. We start by noticing that the space $\mathbb{H}[z]$ has a natural Riemannian manifold structure. It comes from the identification provided by the bijective map

$$\iota : \mathbb{H}[z] \to \mathbb{C}^*_{\boldsymbol{d}_0} = \underbrace{\mathbb{C}^{*Q\times(D_0+1)} \times \mathbb{C}^{*Q\times(D_0+1)} \times \cdots \times \mathbb{C}^{*Q\times(D_0+1)}}_{P}, \tag{3.32}$$

which operates as

$$\boldsymbol{H}(z) = [\,\boldsymbol{h}_1(z)\,\boldsymbol{h}_2(z)\cdots\boldsymbol{h}_P(z)\,] \quad \overset{\iota}{\mapsto} \quad (\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_P)\,, \tag{3.33}$$

where $\boldsymbol{H}_p = \boldsymbol{\mathcal{T}}_0\,(\boldsymbol{h}_p(z))$, recall the discussion in page 53. Moreover, because

$$\mathbb{C}^*_{\boldsymbol{d}_0} \subset \mathbb{C}_{\boldsymbol{d}_0} = \underbrace{\mathbb{C}^{Q\times(D_0+1)} \times \mathbb{C}^{Q\times(D_0+1)} \times \cdots \times \mathbb{C}^{Q\times(D_0+1)}}_{P},$$

we can use the embedding $\imath : \mathbb{C}_{\boldsymbol{d}_0} \to \mathbb{R}^{2PQ(D_0+1)}$ (as defined in (3.17)),

$$(\boldsymbol{Z}_1, \boldsymbol{Z}_2, \ldots, \boldsymbol{Z}_P) \quad \overset{\imath}{\mapsto} \quad \begin{bmatrix} \operatorname{Re}\operatorname{vec}(\boldsymbol{Z}_1) \\ \operatorname{Im}\operatorname{vec}(\boldsymbol{Z}_1) \\ \vdots \\ \operatorname{Re}\operatorname{vec}(\boldsymbol{Z}_P) \\ \operatorname{Im}\operatorname{vec}(\boldsymbol{Z}_P) \end{bmatrix}. \tag{3.34}$$

to further identify $\mathbb{C}_{\boldsymbol{d}_0}$ with the real space $\mathbb{R}^{2PQ(D_0+1)}$. In sum, through the one-to-one composition map $\imath\circ\iota : \mathbb{H}[z] \to \mathbb{R}^{2PQ(D_0+1)}$, we are identifying $\mathbb{H}[z]$ with an open subset of the real Euclidean space $\mathbb{R}^{2PQ(D_0+1)}$, which is assumed equipped with its usual Riemannian structure. With this identification, the space $\mathbb{H}[z]$ is a real smooth manifold with dimension $\dim \mathbb{H}[z] = 2PQ(D_0+1)$. It is a connected topological space, because each factor space $\mathbb{C}^{*Q\times(D_0+1)}$ is connected (it consists of $\mathbb{C}^{Q\times(D_0+1)}$ with the origin removed), and the Cartesian product of connected spaces is connected. Let $\boldsymbol{H}(z) = [\,\boldsymbol{h}_1(z)\,\boldsymbol{h}_2(z)\cdots\boldsymbol{h}_P(z)\,]$ denote a point in $\mathbb{H}[z]$. The tangent space to $\mathbb{H}[z]$ at $\boldsymbol{H}(z)$, written $T_{\boldsymbol{H}(z)}\mathbb{H}[z]$, is naturally identified with the vector space $\mathbb{C}_{\boldsymbol{d}_0}$ over the field $\mathbb{R}$. We denote by $\langle\cdot,\cdot\rangle_{\boldsymbol{H}(Z)}$ the inner product on the tangent space $T_{\boldsymbol{H}(z)}\mathbb{H}[z]$, or simply by $\langle\cdot,\cdot\rangle$ when the point $\boldsymbol{H}(z)$ is clear from the context. For any tangent vectors $\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2, \ldots, \boldsymbol{\Delta}_P)\,, \boldsymbol{\Gamma} = (\boldsymbol{\Gamma}_1, \boldsymbol{\Gamma}_2, \ldots, \boldsymbol{\Gamma}_P) \in T_{\boldsymbol{H}(z)}\mathbb{H}[z]$, where $\boldsymbol{\Delta}_p, \boldsymbol{\Gamma}_p \in \mathbb{C}^{Q\times(D_0+1)}$, it is given by

$$\langle\boldsymbol{\Delta}, \boldsymbol{\Gamma}\rangle_{\boldsymbol{H}(z)} = \sum_{p=1}^{P} \operatorname{Re}\operatorname{tr}\left(\boldsymbol{\Gamma}_p^H \boldsymbol{\Delta}_p\right). \tag{3.35}$$

Moreover, the geodesic $\gamma : I \subset \mathbb{R} \to \mathbb{H}[z]$ which emanates from $\boldsymbol{H}(z)$ in the direction $\boldsymbol{\Delta}$, that is, which satisfies $\gamma(0) = \boldsymbol{H}(z)$ and

$$\frac{d}{dt}\gamma(t)\bigg|_{t=0} = \boldsymbol{\Delta},$$

is given by

$$\gamma(t) = \iota^{-1}\left((\boldsymbol{H}_1 + t\boldsymbol{\Delta}_1, \boldsymbol{H}_2 + t\boldsymbol{\Delta}_2, \ldots, \boldsymbol{H}_P + t\boldsymbol{\Delta}_P)\right),$$

with $\boldsymbol{H}_p = \boldsymbol{\mathcal{T}}_0\left(\boldsymbol{h}_p(z)\right)$. Its domain is given by $I = (a, b)$, where

$$a = \inf\left\{t < 0 \,:\, \boldsymbol{H}_p + t\boldsymbol{\Delta}_p \neq \boldsymbol{0},\ \text{for all } p\right\}$$

and

$$b = \sup\left\{t > 0 \,:\, \boldsymbol{H}_p + t\boldsymbol{\Delta}_p \neq \boldsymbol{0},\ \text{for all } p\right\}.$$

The values $a = -\infty$ and $b = +\infty$ are possible.

The smooth manifold $\mathbb{H}[z]/\sim$. Recall that the space $\mathbb{H}/\sim$ is the set of equivalence classes induced by the equivalence relation $\sim$ on $\mathbb{H}[z]$, where $\boldsymbol{G}(z) \sim \boldsymbol{H}(z)$ if and only if $\boldsymbol{G}(z) = \boldsymbol{H}(z)\boldsymbol{\Theta}\left(\boldsymbol{\theta}\right)$ for some $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_P)^T \in \mathbb{R}^P$, where $\boldsymbol{\Theta}\left(\boldsymbol{\theta}\right) = \operatorname{diag}\left(e^{i\theta_1}, e^{i\theta_2}, \ldots, e^{i\theta_P}\right)$. Consider the P-dimensional torus Lie group

$$\mathbb{T}^P = \underbrace{\mathbb{S}^1 \times \mathbb{S}^1 \times \cdots \times \mathbb{S}^1}_{P},$$

where $\mathbb{S}^1 = \{u \in \mathbb{C} \,:\, |u| = 1\}$, with multiplication law

$$\boldsymbol{u} \cdot \boldsymbol{v} = (u_1, u_2, \ldots, u_P) \cdot (v_1, v_2, \ldots, v_P) = (u_1 v_1, u_2 v_2, \ldots, u_P v_P).$$

Define a right action of $\mathbb{T}^P$ on $\mathbb{H}[z]$ as $\lambda : \mathbb{H}[z] \times \mathbb{T}^P \to \mathbb{H}[z]$,

$$\lambda\left(\boldsymbol{H}(z), \boldsymbol{u}\right) = \left[\begin{array}{cccc} \boldsymbol{h}_1(z)u_1 & \boldsymbol{h}_2(z)u_2 & \cdots & \boldsymbol{h}_P(z)u_P \end{array}\right]. \tag{3.36}$$

In the sequel, we use the notation $\lambda\left(\boldsymbol{H}(z), \boldsymbol{u}\right) = \boldsymbol{H}(z) \cdot \boldsymbol{u}$. For given $\boldsymbol{H}(z) \in \mathbb{H}[z]$, the subset

$$\boldsymbol{H}(z)\mathbb{T}^P = \left\{\boldsymbol{H}(z) \cdot \boldsymbol{u} \,:\, \boldsymbol{u} \in \mathbb{T}^P\right\} \subset \mathbb{H}[z]$$

is termed the orbit of $\boldsymbol{H}(z)$ under the action of $\mathbb{T}^P$. It is obtained by acting on the point $\boldsymbol{H}(z)$ with all elements of the group $\mathbb{T}^P$. Note that the space $\mathbb{H}[z]$ is the disjoint union of all orbits generated by $\mathbb{T}^P$. The set of orbits is called the orbit space and is denoted by $\mathbb{H}[z]/\mathbb{T}^P$. We define the canonical projection map $\rho : \mathbb{H}[z] \to \mathbb{H}[z]/\mathbb{T}^P$, which associates to each point $\boldsymbol{H}(z)$ its corresponding orbit $\rho\left(\boldsymbol{H}(z)\right)$. Figure 3.4 illustrates the concepts introduced here. Now, it is clear that $\boldsymbol{H}(z) \sim \boldsymbol{G}(z)$ if and only if $\boldsymbol{H}(z)$ and $\boldsymbol{G}(z)$ are in the same orbit. Thus, the quotient space $\mathbb{H}[z]/\sim$ has a natural identification with the orbit space $\mathbb{H}[z]/\mathbb{T}^P$. Within this identification, $\pi = \rho$. It is readily seen that the action λ of the Lie group $\mathbb{T}^P$ on the manifold $\mathbb{H}[z]$ is smooth. Moreover, it is free and proper. We recall that the action of a topological group G on a topological space M, is said to be free if, for all $p \in M$, the identity $p \cdot g = p$ implies $g = e$, where e denotes the identity element of the group. That is, only $g = e$ can fix points. The action is said to be proper if, for each compact subset $K \subset M$ the set $G_K = \{g \in G \,:\, Kg \cap K \neq \emptyset\}$ is compact, see [16, page 32]. The action λ is free, because $\boldsymbol{H}(z) \cdot \boldsymbol{u} = \boldsymbol{H}(z)$ implies $\boldsymbol{h}_p(z)u_p = \boldsymbol{h}_p(z)$, for all p. Since $\boldsymbol{h}_p(z)$ is a nonzero polynomial, we conclude that $u_p = 1$, that is, $\boldsymbol{u} = (1, 1, \ldots, 1)$ is the identity element of the group $\mathbb{T}^P$. The action λ is proper because any smooth action by a compact Lie group on a smooth manifold is proper [16, page 32], and $\mathbb{T}^P$ is a compact

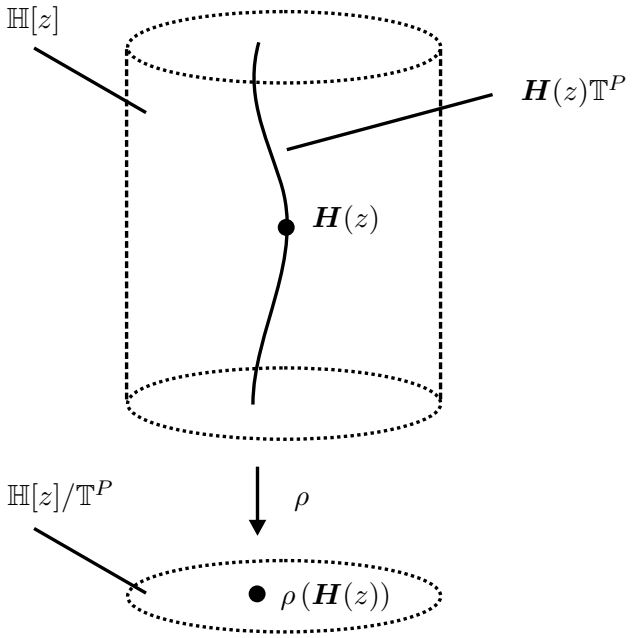

Figure 3.4: The space $\mathbb{H}[z]$, an orbit $\boldsymbol{H}(z)\mathbb{T}^P$ and the orbit space $\mathbb{H}[z]/\mathbb{T}^P$

space (Cartesian product of P compact circles). Thus, we are covered by theoreom 1.95 in [16, page 32]. It states that the orbit space $\mathbb{H}[z]/\mathbb{T}^P$ is a topological manifold and has a (in fact, unique) smooth structure making the canonical projection $\rho : \mathbb{H}[z] \to \mathbb{H}[z]/\mathbb{T}^P$ a smooth submersion. We recall that a smooth map $F : M \to N$ between smooth manifolds is said to be a submersion if, for any $p \in M$, the corresponding linear push forward map $F_{p*} : T_pM \to T_{F(p)}N$ is surjective, that is, $\operatorname{rank} F_{p*} = \dim F_{p*}(T_pM) = \dim T_pN = \dim N$. Hereafter, we consider the quotient space $\mathbb{H}[z]/\sim = \mathbb{H}[z]/\mathbb{T}^P$ equipped with such smooth structure. Its dimension is $\dim \mathbb{H}[z]/\mathbb{T}^P = \dim \mathbb{H}[z] - \dim \mathbb{T}^P = 2PQ(D_0+1) - P$.

The Riemannian manifold $\mathbb{H}[z]/\sim$. Up to this point, the quotient space is only a smooth manifold. We have not yet inserted in it a Riemannian structure. To induce such geometric structure, we exploit proposition 2.28 in [16, page 64]. This proposition asserts that if a Lie group G acts smoothly, freely and properly on a smooth Riemannian manifold (M, g) by isometries, say, $\lambda : M \times G \to M$, then there exits one and only one Riemannian metric on the orbit space M/G making the canonical projection $\rho : M \to M/G$ a Riemannian submersion. We recall that a map $\varrho : (M, g) \to (N, h)$ between Riemannian manifolds is said to be a Riemannian submersion if ϱ is a smooth submersion and, for each $p \in M$, the restriction of the push forward map $\varrho_* : T_pM \to T_{\varrho(p)}N$ to the horizontal linear subspace $H_pM \subset T_pM$ is a linear isometry, that is, $h_{\varrho(p)}(\varrho_*(X_p), \varrho_*(Y_p)) = g_p(X_p, Y_p)$, for all $X_p, Y_p \in H_pM$, see [16, definition 2.27, page 63]. Note that g_p and $h_{\varrho(p)}$ denote inner-products on the linear spaces T_pM and $T_{\varrho(p)}N$, respectively. The horizontal subspace H_pM is the orthogonal complement in T_pM (with respect to the inner product g_p) of the vertical linear subspace

$$V_pM = \ker \varrho_* = \{X_p \in T_pM : \varrho_*(X_p) = 0\} \subset T_pM.$$

Since ϱ is a submersion, it follows that $\dim \varrho_*(T_pM) = \dim T_{\varrho(p)}N = \dim N$, and thereby

$\dim V_pM = \dim M - \dim N$ and $\dim H_pM = \dim N$. Figure 3.5 illustrates these concepts. This figure must be interpreted with some care. We are not suggesting that the tangent

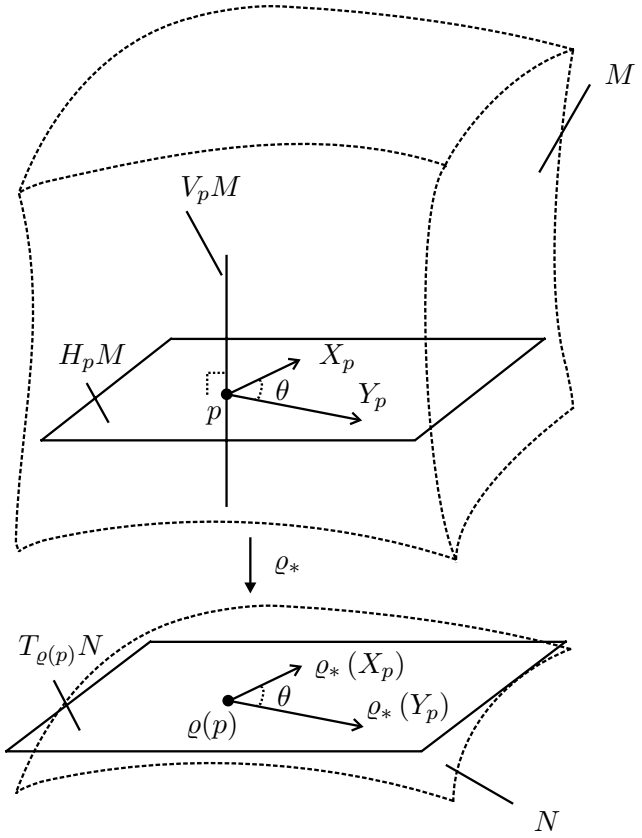

Figure 3.5: A Riemannian submersion $\varrho : M \to N$

space T_pM is a subset of the manifold M. We have chosen to draw it inside M (and not above it, as we did for N) to save space and keep the idea that this is a linear space "attached" to the point $p \in M$. We also recall the Lie group G acts on the Riemannian manifold M by isometries if, for all $g \in G$, the map $\lambda_g : M \to M$, $p \mapsto \lambda_g(p) = p \cdot g$ is a Riemannian isometry. That is, λ_g is a diffeomorphism (one-to-one, onto smooth map with a smooth inverse) and, for any $p \in M$, the push forward map $\lambda_{g*} : T_pM \to T_{p\cdot g}M$ is a linear isometry, see [16, definition 2.5, page 54].

To apply proposition 2.28 in [16, page 64], we take $M = \mathbb{H}[z]$, $G = \mathbb{T}^P$ and consider the action $\lambda : \mathbb{H}[z] \times \mathbb{T}^P \to \mathbb{H}[z]$ defined in (3.36). We have already seen that this action is smooth, free and proper. It remains only to show that, for any given $\boldsymbol{u} \in \mathbb{T}^P$, the map $\lambda_{\boldsymbol{u}} : \mathbb{H}(z) \to \mathbb{H}(z)$, $\lambda_{\boldsymbol{u}}(\boldsymbol{H}(z)) = \boldsymbol{H}(z)\cdot\boldsymbol{u}$ is a Riemannian isometry. It is clear that $\lambda_{\boldsymbol{u}}$ is a diffeomorphism since $\lambda_{\overline{\boldsymbol{u}}}$ is a smooth inverse. Let $\boldsymbol{H}(z) \in \mathbb{H}[z]$. It is easily seen that, under the identification $T_{\boldsymbol{H}(z)}\mathbb{H}[z] = \mathbb{C}_{\boldsymbol{d}_0}$, the push forward linear $\lambda_{\boldsymbol{u}*} : T_{\boldsymbol{H}(z)}\mathbb{H}[z] \to T_{\boldsymbol{H}(z)}\mathbb{H}[z]$ is given by

$$\boldsymbol{\Delta} = (\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2, \ldots, \boldsymbol{\Delta}_P) \in T_{\boldsymbol{H}(z)}\mathbb{H}[z] \quad \overset{\lambda_{\boldsymbol{u}*}}{\mapsto} \quad (\boldsymbol{\Delta}_1 u_1, \boldsymbol{\Delta}_2 u_2, \ldots, \boldsymbol{\Delta}_P u_P) \in T_{\boldsymbol{H}(z)\cdot\boldsymbol{u}}\mathbb{H}[z].$$

To prove that $\lambda_{\boldsymbol{u}*}$ is a linear isometry it suffices to show that it preserves norms, that is,

$\langle \lambda_{\boldsymbol{u}*}(\boldsymbol{\Delta}), \lambda_{\boldsymbol{u}*}(\boldsymbol{\Delta}) \rangle_{\boldsymbol{H}(z)\cdot\boldsymbol{u}} = \langle \boldsymbol{\Delta}, \boldsymbol{\Delta} \rangle_{\boldsymbol{H}(z)}$. But, according to (3.35),

$$\begin{aligned} \langle \lambda_{\boldsymbol{u}*}(\boldsymbol{\Delta}), \lambda_{\boldsymbol{u}*}(\boldsymbol{\Delta}) \rangle_{\boldsymbol{H}(z)\cdot\boldsymbol{u}} &= \sum_{p=1}^{P} \operatorname{tr}\left(\overline{u}_p u_p \boldsymbol{\Delta}_p^H \boldsymbol{\Delta}_p\right) \\ &= \sum_{p=1}^{P} \operatorname{tr}\left(\boldsymbol{\Delta}_p^H \boldsymbol{\Delta}_p\right) \\ &= \langle \boldsymbol{\Delta}, \boldsymbol{\Delta} \rangle_{\boldsymbol{H}(z)}, \end{aligned}$$

where we used the identity $\overline{u}_p u_p = |u_p| = 1$, which is due to the fact that $u_p \in \mathbb{S}^1$. Thus, proposition 2.28 in [16, page 64] can be applied in the present context. In the sequel, we consider the smooth manifold $\mathbb{H}[z]/\sim$ to be equipped with this unique Riemannian structure making the canonical projection $\pi : \mathbb{H}[z] \to \mathbb{H}[z]/\sim$, $\pi(\boldsymbol{H}(z)) = [\boldsymbol{H}(z)]$, a Riemannian submersion.

Distance between points in $\mathbb{H}[z]/\sim$. Note that $\mathbb{H}[z]/\sim$ is a connected topological space, since it is the image of the connected space $\mathbb{H}[z]$ under the continuous map π. Thus, the concept of distance between points in $\mathbb{H}[z]/\sim$ is well-defined (recall the definition of Riemannian distance in page 64). This is the distance function mentioned in (3.1). We now obtain a closed-form expression for $d([\boldsymbol{H}(z)], [\boldsymbol{G}(z)])$, the Riemannian distance between any two given points $[\boldsymbol{H}(z)]$ and $[\boldsymbol{G}(z)]$ in $\mathbb{H}[z]/\sim$.

Let $\boldsymbol{H} = (\boldsymbol{H}_1, \ldots, \boldsymbol{H}_P) = \iota(\boldsymbol{H}(z))$ and $\boldsymbol{G} = (\boldsymbol{G}_1, \ldots, \boldsymbol{G}_P) = \iota(\boldsymbol{G}(z))$. Recall that $\boldsymbol{H}_p = \boldsymbol{\mathcal{T}}_0(\boldsymbol{h}_p(z))$ and $\boldsymbol{G}_p = \boldsymbol{\mathcal{T}}_0(\boldsymbol{g}_p(z))$. We assume that

$$\operatorname{Re}\operatorname{tr}\left(\boldsymbol{G}_p^H \boldsymbol{H}_p\right) \geq 0, \quad \operatorname{Im}\operatorname{tr}\left(\boldsymbol{G}_p^H \boldsymbol{H}_p\right) = 0, \tag{3.37}$$

for all p. This entails no loss of generality. Suppose (3.37) does not hold for some p. Then, $\operatorname{tr}\left(\boldsymbol{G}_p^H \boldsymbol{H}_p\right)$ is a nonzero complex number. Define

$$\widetilde{\boldsymbol{g}}_p(z) = \boldsymbol{g}_p(z) \frac{\operatorname{tr}\left(\boldsymbol{G}_p^H \boldsymbol{H}_p\right)}{|\operatorname{tr}\left(\boldsymbol{G}_p^H \boldsymbol{H}_p\right)|}, \tag{3.38}$$

and $\widetilde{\boldsymbol{g}}_q(z) = \boldsymbol{g}_q(z)$, for all $q \neq p$. Let $\widetilde{\boldsymbol{G}}(z) = [\,\widetilde{\boldsymbol{g}}_1(z)\, \widetilde{\boldsymbol{g}}_2(z)\, \cdots \widetilde{\boldsymbol{g}}_P(z)\,]$. It is clear that $\boldsymbol{G}(z)$ and $\widetilde{\boldsymbol{G}}(z)$ are in the same orbit, that is, $[\boldsymbol{G}(z)] = \left[\widetilde{\boldsymbol{G}}(z)\right]$. Moreover, from (3.38), it follows that

$$\widetilde{\boldsymbol{G}}_p = \boldsymbol{\mathcal{T}}_0\left(\widetilde{\boldsymbol{g}}_p(z)\right) = \boldsymbol{G}_p \frac{\operatorname{tr}\left(\boldsymbol{G}_p^H \boldsymbol{H}_p\right)}{|\operatorname{tr}\left(\boldsymbol{G}_p^H \boldsymbol{H}_p\right)|},$$

and $\operatorname{tr}\left(\widetilde{\boldsymbol{G}}_p^H \boldsymbol{H}_p\right) = |\operatorname{tr}\left(\boldsymbol{G}_p^H \boldsymbol{H}_p\right)|$. Thus, $\operatorname{Re}\operatorname{tr}\left(\widetilde{\boldsymbol{G}}_p^H \boldsymbol{H}_p\right) \geq 0$ and $\operatorname{Im}\operatorname{tr}\left(\widetilde{\boldsymbol{G}}_p^H \boldsymbol{H}_p\right) = 0$. Redefine $\boldsymbol{G}(z) = \widetilde{\boldsymbol{G}}(z)$. If (3.37) does not hold for another p, we can repeat this procedure. In sum, for a given $\boldsymbol{G}(z)$, we see that, in at most P movements in its orbit, we can find an orbit representative satisfying (3.37).

Using the identification $\mathbb{H}[z] = \mathbb{C}^*_{\boldsymbol{d}_0}$, consider the geodesic in $\mathbb{H}[z]$ given by $\gamma : [0,1] \subset \mathbb{R} \to \mathbb{H}[z]$,

$$\gamma(t) = \left(\boldsymbol{H}_1 + t(\boldsymbol{G}_1 - \boldsymbol{H}_1), \boldsymbol{H}_2 + t(\boldsymbol{G}_2 - \boldsymbol{H}_2), \ldots, \boldsymbol{H}_P + t(\boldsymbol{G}_P - \boldsymbol{H}_P)\right), \tag{3.39}$$

which connects the point $\gamma(0) = \boldsymbol{H}(z)$ to the point $\gamma(1) = \boldsymbol{G}(z)$. Note that, indeed, $\gamma(t) \in \mathbb{C}^*_{\boldsymbol{d}_0}$ for all $t \in (0,1)$, that is,

$$\boldsymbol{H}_p + t\,(\boldsymbol{G}_p - \boldsymbol{H}_p) \neq \boldsymbol{0}, \tag{3.40}$$

for all $t \in (0,1)$ and any p. To check this, assume $\boldsymbol{H}_p + t_0\,(\boldsymbol{G}_p - \boldsymbol{H}_p) = \boldsymbol{0}$, for some $t_0 \in (0,1)$ and some p. Then,

$$(1-t_0)\,\boldsymbol{H}_p + t_0\,\boldsymbol{G}_p = \boldsymbol{0}. \tag{3.41}$$

Multiplying both sides of (3.41) on the left by $\boldsymbol{G}_p^H$, taking the trace and using (3.37), yields

$$(1-t_0)\,|\mathrm{tr}\left(\boldsymbol{G}_p^H\boldsymbol{H}_p\right)| + t_0\,\|\boldsymbol{G}_p\|^2 = 0.$$

This implies $\|\boldsymbol{G}_p\| = 0$, that is, $\boldsymbol{g}_p(z) = 0$, which is impossible. Thus, γ is a curve in $\mathbb{H}[z] = \mathbb{C}^*_{\boldsymbol{d}_0}$.

Under the natural identification $T_{\boldsymbol{H}(z)}\mathbb{H}[z]$ with the vector space $\mathbb{C}_{\boldsymbol{d}_0}$ over $\mathbb{R}$, the tangent vector $\dot\gamma(t) \in T_{\gamma(t)}\mathbb{H}[z]$ is given by

$$\dot\gamma(t) = (\boldsymbol{G}_1 - \boldsymbol{H}_1, \boldsymbol{G}_2 - \boldsymbol{H}_2, \ldots, \boldsymbol{G}_P - \boldsymbol{H}_P)\,. \tag{3.42}$$

Thus, using (3.35), the length of the curve γ is

$$\begin{aligned}
\ell\,(\gamma) &= \int_0^1 \langle\dot\gamma(t), \dot\gamma(t)\rangle_{\gamma(t)}^{1/2}\,dt \\
&= \sqrt{\sum_{p=1}^{P} \|\boldsymbol{G}_p - \boldsymbol{H}_p\|^2} \\
&= \sqrt{\sum_{p=1}^{P} \|\boldsymbol{G}_p\|^2 + \|\boldsymbol{H}_p\|^2 - 2|\mathrm{tr}\left(\boldsymbol{G}_p^H\boldsymbol{H}_p\right)|}.
\end{aligned} \tag{3.43}$$

It is important to note that $\ell\,(\gamma)$ in (3.43) is the distance from the point $\boldsymbol{H}(z)$ to the orbit $\boldsymbol{G}(z)\mathbb{T}^P$, that is,

$$\ell\,(\gamma) = \min_{\widetilde{\boldsymbol{G}}(z)\,\in\,\boldsymbol{G}(z)\mathbb{T}^P} d\left(\boldsymbol{H}(z), \widetilde{\boldsymbol{G}}(z)\right).$$

To see this, let $\widetilde{\boldsymbol{G}}(z) \in \boldsymbol{G}(z)\mathbb{T}^P$, that is, $\widetilde{\boldsymbol{G}}(z) = \boldsymbol{G}(z)\cdot\boldsymbol{u}$, for some $\boldsymbol{u}$ in the torus $\mathbb{T}^P$. Thus, $\widetilde{\boldsymbol{G}}_p = \boldsymbol{G}_p\,u_p$ with $|u_p| = 1$, and

$$\begin{aligned}
d\left(\boldsymbol{H}(z), \widetilde{\boldsymbol{G}}(z)\right)^2 &\geq \sum_{p=1}^{P} \left\|\boldsymbol{H}_p - \widetilde{\boldsymbol{G}}_p\right\|^2 && (3.44)\\
&= \sum_{p=1}^{P} \|\boldsymbol{H}_p\|^2 + \|\boldsymbol{G}_p\|^2 - 2\,\mathrm{Re}\left\{u_p\,\mathrm{tr}\left(\boldsymbol{G}_p^H\boldsymbol{H}_p\right)\right\} && (3.45)\\
&\geq \sum_{p=1}^{P} \|\boldsymbol{G}_p\|^2 + \|\boldsymbol{H}_p\|^2 - 2|\mathrm{tr}\left(\boldsymbol{G}_p^H\boldsymbol{H}_p\right)| && (3.46)\\
&= \ell(\gamma)^2.
\end{aligned}$$

The inequality (3.44) is valid because under the Riemannian identification $\mathbb{H}[z] = \mathbb{C}^*_{\boldsymbol{d}_0}$ the shortest distance between points is the usual Euclidean distance. In (3.45), we simply replaced $\widetilde{\boldsymbol{G}}_p$ by $\boldsymbol{G}_p u_p$. Inequality (3.46) follows from the fact that, for any $z \in \mathbb{C}$, we have $\operatorname{Re} z \leq |z|$.

Note that the tangent vector $\dot{\gamma}(0)$ is horizontal, that is, $\dot{\gamma}(0) \in H_{\gamma(0)}\mathbb{H}[z]$. In general, if $\varrho : M \to N$ is a Riemannian submersion, a smooth curve $\gamma : I \to M$ is said to be horizontal if the tangent vector $\dot{\gamma}(t)$ belongs to the horizontal subspace $H_{\gamma(t)}M$, for all $t \in I$. The tangent vector $\dot{\gamma}(0)$ can be seen to be horizontal as follows. The vertical space at any given point $\boldsymbol{H}(z) \in \mathbb{H}[z] = \mathbb{C}^*_{\boldsymbol{d}_0}$ is the P-dimensional subspace of $T_{\boldsymbol{H}(z)}\mathbb{H}[z] = \mathbb{C}_{\boldsymbol{d}_0}$ given by

$$V_{\boldsymbol{H}(z)}\mathbb{H}[z] = \operatorname{span}\left\{(i\boldsymbol{H}_1, \boldsymbol{0}, \ldots, \boldsymbol{0}), (\boldsymbol{0}, i\boldsymbol{H}_2, \ldots, \boldsymbol{0}), \ldots, (\boldsymbol{0}, \ldots, \boldsymbol{0}, i\boldsymbol{H}_P)\right\}. \tag{3.47}$$

This is precisely the tangent space to the orbit $\boldsymbol{H}(z)\mathbb{T}^P$ at the point $\boldsymbol{H}(z)$. See figure 3.6 for a sketch. From (3.47) and (3.35), we conclude that the horizontal subspace at $\boldsymbol{H}(z)$ is

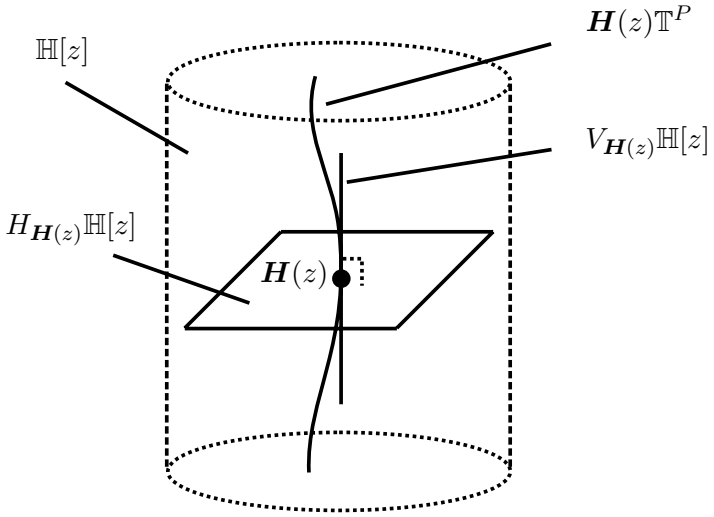

Figure 3.6: The horizontal $H_{\boldsymbol{H}(z)}\mathbb{H}[z]$ and vertical $V_{\boldsymbol{H}(z)}\mathbb{H}[z]$ subspaces of $T_{\boldsymbol{H}(z)}\mathbb{H}(z)$

the $2PQ(D_0+1) - P$-dimensional subspace of $T_{\boldsymbol{H}(z)}\mathbb{H}[z] = \mathbb{C}_{\boldsymbol{d}_0}$ given by

$$H_{\boldsymbol{H}(z)}\mathbb{H}[z] = \left\{(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2, \ldots, \boldsymbol{\Delta}_P) : \operatorname{Im}\operatorname{tr}\left(\boldsymbol{\Delta}_p^H \boldsymbol{H}_p\right) = 0, \text{ for all } p \right\}. \tag{3.48}$$

Since $\gamma(0) = \boldsymbol{H}(z)$, we have $H_{\gamma(0)}\mathbb{H}[z] = H_{\boldsymbol{H}(z)}\mathbb{H}[z]$. Let $(\boldsymbol{\Delta}_1, \boldsymbol{\Delta}_2, \ldots, \boldsymbol{\Delta}_P) = \dot{\gamma}(0)$. From (3.42), $\boldsymbol{\Delta}_p = \boldsymbol{G}_p - \boldsymbol{H}_p$, and, as a consequence,

$$\operatorname{tr}\left(\boldsymbol{\Delta}_p^H \boldsymbol{H}_p\right) = \operatorname{tr}\left(\boldsymbol{G}_p^H \boldsymbol{H}_p\right) - \|\boldsymbol{H}_p\|^2. \tag{3.49}$$

Taking the imaginary part of both sides of (3.49) and using (3.37) yields $\operatorname{Im}\operatorname{tr}\left(\boldsymbol{\Delta}_p^H \boldsymbol{H}_p\right) = 0$. That is, the tangent vector $\dot{\gamma}(0)$ is horizontal. Let the curve γ descend to the quotient space, that is, define the smooth curve $\eta : [0,1] \to \mathbb{H}[z]/\sim$ by $\eta(t) = \pi(\gamma(t))$. Note that η connects the point $\eta(0) = [\boldsymbol{H}(z)]$ to the point $\eta(1) = [\boldsymbol{G}(z)]$. Since γ is a geodesic of $\mathbb{H}[z]$ which starts with an horizontal velocity vector, proposition 2.109 [16, page 97] asserts that the whole curve γ is horizontal and, more importantly, that the curve η is a geodesic of $\mathbb{H}[z]/\sim$ with the same length of γ, that is, $\ell(\eta) = \ell(\gamma)$. Figure 3.7 provides an illustration.

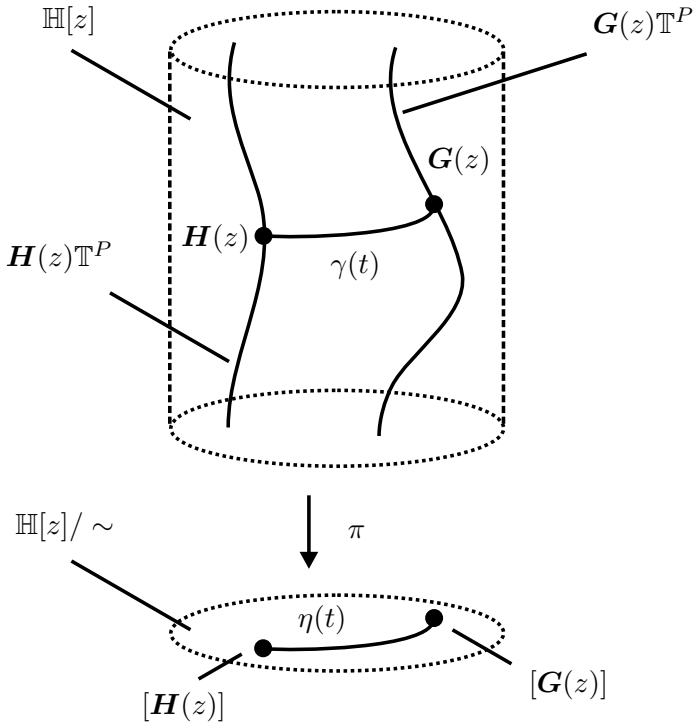

Figure 3.7: The geodesic $\gamma(t)$ in $\mathbb{H}[z]$ descends to the geodesic $\eta(t)$ in $\mathbb{H}[z]/\sim$

We now show that if $\widetilde{\eta} : [a,b] \to \mathbb{H}[z]/\sim$ denotes any piecewise regular curve connecting $\widetilde{\eta}(a) = [\boldsymbol{H}(z)]$ to $\widetilde{\eta}(b) = [\boldsymbol{G}(z)]$, then its length is greater or equal than that of η, that is, $\ell(\widetilde{\eta}) \geq \ell(\eta)$. This proves that the Riemannian distance $d([\boldsymbol{H}(z)], [\boldsymbol{G}(z)]) = \ell(\eta)$, that is, it is given by (3.43). We need lemma 3.5 below. It is an easy exercise in differential geometry and it is certainly a known result. However, because a specific reference to it could not be located in [16, 7], we include a proof for completness.

Lemma 3.5. *Let the Lie group G act smoothly, freely and properly on the smooth manifold M by isometries. Thus, M/G can be given a Riemannian structure making the canonical projection $\rho : M \to M/G$ a Riemannian submersion. Let $q : [a,b] \to M/G$ denote a piecewise regular curve and x an arbitrarily chosen point in the orbit $\rho^{-1}(q(a)) \subset M$, that is, $\rho(x) = q(a)$. Then, there is an unique horizontal piecewise regular curve $p : [a,b] \to M$ starting at x and projecting to the curve q under ρ. That is, $p(a) = x$, $\dot{p}(t) \in H_{p(t)}M$ and $\rho(p(t)) = q(t)$, for all t.*

Proof: See appendix B.

Now, let $\widetilde{\eta} : [a,b] \to \mathbb{H}[z]/\sim$ be any piecewise regular curve connecting $\widetilde{\eta}(a) = [\boldsymbol{H}(z)]$ to $\widetilde{\eta}(b) = [\boldsymbol{G}(z)]$. Let $\widetilde{\gamma} : [a,b] \to \mathbb{H}[z]$ denote the corresponding horizontal curve starting at $\boldsymbol{H}(z)$, whose existence is guaranteed by lemma 3.5. Note that $\widetilde{\gamma}$ connects $\boldsymbol{H}(z)$ to the orbit $\boldsymbol{G}(z)\mathbb{T}^P$, because $\pi(\widetilde{\gamma}(a)) = \widetilde{\eta}(a) = [\boldsymbol{H}(z)]$ and $\pi(\widetilde{\gamma}(b)) = \widetilde{\eta}(b) = [\boldsymbol{G}(z)]$. As shown previously, $\ell(\gamma)$ is the distance from the point $\boldsymbol{H}(z)$ to the orbit $\boldsymbol{G}(z)\mathbb{T}^P$. Thus, $\ell(\widetilde{\gamma}) \geq \ell(\gamma)$. But, since $\widetilde{\gamma}(t)$ is an horizontal curve and projects to $\widetilde{\eta}(t) = \pi(\widetilde{\gamma}(t))$, the curve $\widetilde{\eta}(t)$ has the same length, see [16, page 98]. That is, $\ell(\widetilde{\eta}) = \ell(\widetilde{\gamma}) \geq \ell(\gamma) = \ell(\eta)$. Thus, we conclude that the Riemannian distance between the points $[\boldsymbol{H}(z)]$ and $[\boldsymbol{G}(z)]$ is

given by

$$d\left([\boldsymbol{H}(z)],[\boldsymbol{G}(z)]\right)=\sqrt{\sum_{p=1}^{P}\|\boldsymbol{G}_p\|^2+\|\boldsymbol{H}_p\|^2-2|\mathrm{tr}\left(\boldsymbol{G}_p^H\boldsymbol{H}_p\right)|}. \tag{3.50}$$

Notice that the assumption (3.37), although used initially in the deduction, is not needed in (3.50). That is, $\boldsymbol{H}(z)$ and $\boldsymbol{G}(z)$ can be any orbit representatives. To see this, notice that the right-hand side of (3.50) remains the same, as it should, if one replaces $\boldsymbol{H}_p$ by $\boldsymbol{H}_p u_p$ (or $\boldsymbol{G}_p$ by $\boldsymbol{G}_p u_p$), with $|u_p|=1$, as this simply correspond to shifts of $\boldsymbol{H}(z)$ (or $\boldsymbol{G}(z)$) along the corresponding orbit, without changing the equivalence class $[\boldsymbol{H}(z)]$ (or $[\boldsymbol{G}(z)]$).

Sectional curvatures of $\mathbb{H}[z]/\sim$. We compute the sectional curvatures of the Riemannian manifold $\mathbb{H}[z]/\sim$ at a given point $[\boldsymbol{H}(z)]$ (recall the definition of sectional curvature in (3.30)). As mentioned previously, this data will be used in chapter 4. In loose terms, the sectional curvatures of a Riemannian manifold M encode the local geometry of M. Given a point $p\in M$, they probe the geometric structure of M around p by analysing the Gaussian curvature of the two-dimensional submanifolds $\mathrm{Exp}_p(\Pi)\subset M$, as Π varies over the set of two-dimensional planes in the tangent space T_pM. In fact, due to its intrinsic symmetries, the Riemannian curvature tensor at p can be fully recovered from the sectional curvatures at p, see [7, theorem 3.5, page 385].

Let $X_{[\boldsymbol{H}(z)]}$ and $Y_{[\boldsymbol{H}(z)]}$ denote two orthonormal vectors in $T_{[\boldsymbol{H}(z)]}\mathbb{H}[z]/\sim$. From (3.30), we must compute

$$K\left(X_{[\boldsymbol{H}(z)]},Y_{[\boldsymbol{H}(z)]}\right)=R\left(X_{[\boldsymbol{H}(z)]},Y_{[\boldsymbol{H}(z)]},Y_{[\boldsymbol{H}(z)]},X_{[\boldsymbol{H}(z)]}\right), \tag{3.51}$$

where R denotes the Riemannian curvature tensor of $\mathbb{H}[z]/\sim$. Recall (3.29) for the definition of the curvature tensor. To obtain (3.51), we exploit O'Neill's formula, see [16, theorem 3.61, page 127]. It asserts that, if $\varrho: M\to N$ is a Riemannian submersion, and X,Y denote smooth orthonormal vector fields on N, then

$$K(X_{\varrho(p)},Y_{\varrho(p)})=\widetilde{K}\left(\widetilde{X}_p,\widetilde{Y}_p\right)+\frac{3}{4}|[\widetilde{X},\widetilde{Y}]_p^V|^2,$$

for all $p\in M$, where $\widetilde{X},\widetilde{Y}\in\mathcal{T}(M)$ denote the horizontal lifts of $X,Y\in\mathcal{T}(N)$, $\widetilde{K}$ stands for the sectional curvature on M, and the smooth vector field $[\widetilde{X},\widetilde{Y}]^V$ denotes the vertical part of $[\widetilde{X},\widetilde{Y}]$. That is, $\widetilde{X}_p,\widetilde{Y}_p\in H_pM$ and $\varrho_{p*}\left(\widetilde{X}_p\right)=X_p$, $\varrho_{p*}\left(\widetilde{Y}_p\right)=Y_p$, for all $p\in M$. Moreover, $[\widetilde{X},\widetilde{Y}]_p^V\in V_pM$ denotes the orthogonal projection of the tangent vector $[\widetilde{X},\widetilde{Y}]_p\in T_pM$ onto the vertical subspace $V_pM\subset T_pM$. In our context, we may take $M=\mathbb{H}[z]$, $N=\mathbb{H}[z]/\sim$, $p=\boldsymbol{H}(z)$ and $\varrho=\pi$. Furthermore, we let $X,Y\in\mathcal{T}(U)$ denote smooth orthonormal vector fields defined in an open neighborhood $U\subset\mathbb{H}[z]/\sim$ of $\pi(\boldsymbol{H}(z))=[\boldsymbol{H}(z)]$, which extend (locally) the given tangent vectors $X_{[\boldsymbol{H}(z)]}$ and $Y_{[\boldsymbol{H}(z)]}$. Their corresponding horizontal lifts, defined in the open set $\widetilde{U}=\pi^{-1}(U)\subset\mathbb{H}[z]$, are denoted by $\widetilde{X}$ and $\widetilde{Y}$. Thus,

$$K(X_{[\boldsymbol{G}(z)]},Y_{[\boldsymbol{G}(z)]})=\widetilde{K}\left(\widetilde{X}_{\boldsymbol{G}(z)},\widetilde{Y}_{\boldsymbol{G}(z)}\right)+\frac{3}{4}|[\widetilde{X},\widetilde{Y}]^V_{\boldsymbol{G}(z)}|^2, \tag{3.52}$$

for all $\boldsymbol{G}(z)\in\widetilde{U}$ and where $\widetilde{K}$ denotes the sectional curvature of $\mathbb{H}[z]$. Recall that we are identifying $\mathbb{H}[z]$ with an open subset of $\mathbb{R}^{2PQ(D_0+1)}$ through the composition $\imath\circ\iota:\mathbb{H}[z]\to$

$\mathbb{R}^{2PQ(D_0+1)}$, see (3.32), (3.33) and (3.34). But, the real Euclidean spaces with their usual Riemannian structure are flat, that is, they are spaces of constant zero curvature. Thus,

$$\widetilde{K}\left(\widetilde{X}_{\boldsymbol{G}(z)}, \widetilde{Y}_{\boldsymbol{G}(z)}\right) = 0, \tag{3.53}$$

see [7, page 386]. Now, consider the P smooth vector fields $E_1, \ldots, E_P \in \mathcal{T}(\widetilde{U})$ defined, under the identification $T_{\boldsymbol{G}(z)}\mathbb{H}[z] = \mathbb{C}_{\boldsymbol{d}_0}$, by

$$E_{1\,\boldsymbol{G}(z)} = (i\boldsymbol{G}_1, \boldsymbol{0}, \ldots, \boldsymbol{0}), \ldots, E_{P\,\boldsymbol{G}(z)} = (\boldsymbol{0}, \ldots, i\boldsymbol{G}_P), \tag{3.54}$$

at each point $\boldsymbol{G}(z) \in \widetilde{U}$. As seen in (3.47), the frame $E_{1\,\boldsymbol{G}(z)}, \ldots, E_{P\,\boldsymbol{G}(z)}$ denotes a basis for the vertical subspace $V_{\boldsymbol{G}(z)}\mathbb{H}[z] \subset T_{\boldsymbol{G}(z)}\mathbb{H}[z]$. Furthermore, it is an orthogonal basis, because, from (3.35), we have

$$\langle E_{p\,\boldsymbol{G}(z)}, E_{q\,\boldsymbol{G}(z)} \rangle_{\boldsymbol{G}(z)} = r_p\left(\boldsymbol{G}(z)\right)\, \delta[p-q],$$

where $r_p : \widetilde{U} \to \mathbb{R}$ denotes the smooth function $r_p\left(\boldsymbol{G}(z)\right) = \|\boldsymbol{G}_p\|^2$. Define the P smooth vector fields $F_1, \ldots, F_P \in \mathcal{T}(\widetilde{U})$, by

$$F_p = 1/\sqrt{r_p}\, E_p. \tag{3.55}$$

It is clear that, when evaluated at each point $\boldsymbol{G}(z) \in \widetilde{U}$, they materialize into an orthonormal basis for $V_{\boldsymbol{G}(z)}\mathbb{H}[z]$. Let $\sigma_1, \ldots, \sigma_P \in \mathcal{T}^*(\widetilde{U})$ denote the dual covector fields. That is, $\sigma_{1\boldsymbol{G}(z)}, \ldots, \sigma_{P\boldsymbol{G}(z)}$ denotes the basis of $T^*_{\boldsymbol{G}(z)}\mathbb{H}[z]$ which is dual to the basis $F_{1\boldsymbol{G}(z)}, \ldots, F_{P\boldsymbol{G}(z)}$ of $T_{\boldsymbol{G}(z)}\mathbb{H}[z]$, at each point $\boldsymbol{G}(z) \in \widetilde{U}$. Thus,

$$\sigma_p\left(F_q\right) = \delta[p-q]. \tag{3.56}$$

From (3.56), and since $F_1, \ldots, F_P$ are orthonormal, it follows that $\sigma_{p\boldsymbol{G}(z)} = F^{\flat}_{p\,\boldsymbol{G}(z)}$, at each $\boldsymbol{G}(z) \in \widetilde{U}$. Recall from (3.27) that this means

$$\sigma_p\left(X_{\boldsymbol{G}(z)}\right) = \langle X_{\boldsymbol{G}(z)}, F_{p\,\boldsymbol{G}(z)} \rangle,$$

for any $X_{\boldsymbol{G}(z)} \in T_{\boldsymbol{G}(z)}\mathbb{H}[z]$. Thus, for an arbitrary tangent vector $X_{\boldsymbol{G}(z)}$, we have

$$|X^V_{\boldsymbol{G}(z)}|^2 = \sum_{p=1}^{P} \sigma_p\left(X_{\boldsymbol{G}(z)}\right)^2. \tag{3.57}$$

It is more convenient to work with the covector fields $\omega_1, \ldots, \omega_P \in \mathcal{T}^*(\widetilde{U})$, defined by

$$\omega_p = \sqrt{r_p}\, \sigma_p. \tag{3.58}$$

From (3.57) and (3.58), we have

$$|X^V_{\boldsymbol{G}(z)}|^2 = \sum_{p=1}^{P} \frac{1}{r_p(\boldsymbol{G}(z))}\, \omega_p\left(X_{\boldsymbol{G}(z)}\right)^2. \tag{3.59}$$

Using (3.53) and (3.59) in (3.52) yields

$$K(X_{[\boldsymbol{H}(z)]}, Y_{[\boldsymbol{H}(z)]}) = \frac{3}{4} \sum_{p=1}^{P} \frac{1}{\|\boldsymbol{H}_p\|^2}\, \omega_p\left([\widetilde{X}, \widetilde{Y}]_{\boldsymbol{H}(z)}\right)^2. \tag{3.60}$$

From (3.55), we see that ω_p is the dual covector field of E_p, that is, $\omega_p = E_p^\flat$. Furthermore, from [7, lemma 8.4, page 224] we have

$$\omega_p\left([\widetilde{X},\widetilde{Y}]\right) = \widetilde{X}\omega_p(\widetilde{Y}) - \widetilde{Y}\omega_p(\widetilde{X}) - d\omega_p(\widetilde{X},\widetilde{Y}), \tag{3.61}$$

where d denotes the exterior differentiation operator. But,

$$\omega_p(\widetilde{Y}) = \langle \widetilde{Y}, E_p\rangle = 0, \quad \omega_p(\widetilde{X}) = \langle \widetilde{X}, E_p\rangle = 0, \tag{3.62}$$

because $\widetilde{X}$ and $\widetilde{Y}$ denote horizontal vector fields, whereas E_p is a vertical vector field. Again, we recall that we are identifying $\mathbb{H}[z]$ with an open subset of $\mathbb{R}^{2PQ(D_0+1)}$ through the composition $\imath\circ\iota : \mathbb{H}[z] \to \mathbb{R}^{2PQ(D_0+1)}$. Let $n = Q(D_0+1)$ and label the coordinates of $\mathbb{R}^{2Pn}$ as $x_1^1,\ldots,x_1^n, y_1^1,\ldots,y_1^n,\ldots,x_P^1,\ldots,x_P^n,y_P^1,\ldots,y_P^n$. Within this identification the vertical vector fields E_p defined in (3.54) are given by

$$E_p = \sum_{k=1}^{n} x_p^k \frac{\partial}{\partial y_p^k} - y_p^k \frac{\partial}{\partial x_p^k}, \tag{3.63}$$

where, as usual, $\left\{\partial/\partial x_p^k, \partial/\partial y_p^k : k=1,\ldots,n, p=1,\ldots,P\right\}$ denote the tangent vector fields induced by the coordinates $\left\{x_p^k, y_p^k : k=1,\ldots,n, p=1,\ldots,P\right\}$. From (3.63), it follows that

$$\omega_p = E_p^\flat = \sum_{k=1}^{n} x_p^k\, dy_p^k - y_p^k\, dx_p^k.$$

Consequently,

$$d\omega_p = \sum_{k=1}^{n} dx_p^k \wedge dy_p^k - dy_p^k \wedge dx_p^k = -2\sum_{k=1}^{n} dy_p^k \wedge dx_p^k, \tag{3.64}$$

where $\wedge$ stands for the wedge product of differential forms. Plugging (3.62) and (3.64) in (3.61) yields

$$\begin{aligned}\omega_p\left([\widetilde{X},\widetilde{Y}]\right) &= 2\sum_{k=1}^{n} dy_p^k \wedge dx_p^k(\widetilde{X},\widetilde{Y}) \\ &= 2\sum_{k=1}^{n} dy_p^k(\widetilde{X})\, dx_p^k(\widetilde{Y}) - dy_p^k(\widetilde{Y})\, dx_p^k(\widetilde{X}).\end{aligned} \tag{3.65}$$

Under the identification $T_{\boldsymbol{H}(z)}\,\mathbb{H}[z] = \mathbb{C}_{\boldsymbol{d}_0}$, write $\widetilde{X}_{\boldsymbol{H}(z)} = (\widetilde{\boldsymbol{X}}_1,\ldots,\widetilde{\boldsymbol{X}}_P)$ and $\widetilde{Y}_{\boldsymbol{H}(z)} = (\widetilde{\boldsymbol{Y}}_1,\ldots,\widetilde{\boldsymbol{Y}}_P)$. Evaluating (3.65) at $\boldsymbol{H}(z)$ gives

$$\begin{aligned}\omega_p\left([\widetilde{X},\widetilde{Y}]_{\boldsymbol{H}(z)}\right) &= 2\,\mathrm{tr}\left\{\left(\mathrm{Re}\,\widetilde{\boldsymbol{Y}}_p\right)^T\left(\mathrm{Im}\,\widetilde{\boldsymbol{X}}_p\right)\right\} - 2\,\mathrm{tr}\left\{\left(\mathrm{Im}\,\widetilde{\boldsymbol{Y}}_p\right)^T\left(\mathrm{Re}\,\widetilde{\boldsymbol{X}}_p\right)\right\} \\ &= 2\,\mathrm{Re}\,\mathrm{tr}\left\{\left(i\widetilde{\boldsymbol{Y}}_p\right)^H\widetilde{\boldsymbol{X}}_p\right\}.\end{aligned} \tag{3.66}$$

Finally, inserting (3.66) in (3.60) yields

$$K(X_{[\boldsymbol{H}(z)]}, Y_{[\boldsymbol{H}(z)]}) = 3\sum_{p=1}^{P}\frac{1}{\|\boldsymbol{H}_p\|^2}\left[\mathrm{Re}\,\mathrm{tr}\left\{\left(i\widetilde{\boldsymbol{Y}}_p\right)^H\widetilde{\boldsymbol{X}}_p\right\}\right]^2. \tag{3.67}$$

In chapter 4, we will need an upper-bound on the sectional curvatures at a given point $[\boldsymbol{H}(z)]$, that is, an upper-bound on

$$C_{[\boldsymbol{H}(z)]} = \max_{X_{[\boldsymbol{H}(z)]}, Y_{[\boldsymbol{H}(z)]} \text{ : orthonormal}} K(X_{[\boldsymbol{H}(z)]}, Y_{[\boldsymbol{H}(z)]}).$$

From (3.67), we have

$$K(X_{[\boldsymbol{H}(z)]}, Y_{[\boldsymbol{H}(z)]}) \leq 3 \sum_{p=1}^{P} \frac{1}{\|\boldsymbol{H}_p\|^2} \left\|\widetilde{\boldsymbol{X}}_p\right\|^2 \left\|\widetilde{\boldsymbol{Y}}_p\right\|^2 \tag{3.68}$$

$$\leq 3 \max \left\{ \frac{1}{\|\boldsymbol{H}_p\|^2} \,:\, p = 1, 2, \ldots, P \right\}. \tag{3.69}$$

In (3.68) we used the inequality $|\mathrm{Re}\, z| \leq |z|$ for $z \in \mathbb{C}$ together with the Cauchy-Schwartz inequality $|\mathrm{tr}(\boldsymbol{A}^H \boldsymbol{B})| \leq \|\boldsymbol{A}\| \, \|\boldsymbol{B}\|$. In (3.69) we used the fact that the tangent vectors $\widetilde{\boldsymbol{X}}_{\boldsymbol{H}(z)} = (\widetilde{\boldsymbol{X}}_1, \ldots, \widetilde{\boldsymbol{X}}_P)$ and $\widetilde{\boldsymbol{Y}}_{\boldsymbol{H}(z)} = (\widetilde{\boldsymbol{Y}}_1, \ldots, \widetilde{\boldsymbol{Y}}_P)$ have unit-norm, that is,

$$\sum_{p=1}^{P} \left\|\widetilde{\boldsymbol{X}}_p\right\|^2 = \sum_{p=1}^{P} \left\|\widetilde{\boldsymbol{Y}}_p\right\|^2 = 1.$$

Thus, $\left\|\widetilde{\boldsymbol{Y}}_p\right\|^2 \leq 1$ and

$$\sum_{p=1}^{P} \left\|\widetilde{\boldsymbol{X}}_p\right\|^2 \left\|\widetilde{\boldsymbol{Y}}_p\right\|^2 \leq \sum_{p=1}^{P} \left\|\widetilde{\boldsymbol{X}}_p\right\|^2 = 1,$$

which, given (3.68), implies (3.69). In sum,

$$C_{[\boldsymbol{H}(z)]} \leq 3 \max \left\{ \frac{1}{\|\boldsymbol{H}_p\|^2} \,:\, p = 1, 2, \ldots, P \right\}. \tag{3.70}$$

3.4.2 The map ψ

In this subsection, we define the mapping ψ which, in brief, is the mapping performing steps 1) to 4) of the CFIA in table 3.1. Furthermore, we compute its derivative at the point $(\boldsymbol{R}_{\boldsymbol{y}}[0], \boldsymbol{R}_{\boldsymbol{y}}[1], \ldots, \boldsymbol{R}_{\boldsymbol{y}}[M])$. This data is needed to close the asymptotic analysis in section 3.4.4. The mapping ψ will be of the form

$$\psi : \mathcal{U} \subset \underbrace{\mathbb{C}^{Q \times Q} \times \cdots \times \mathbb{C}^{Q \times Q}}_{M+1} \to \mathbb{H}[z],$$

for a certain open set $\mathcal{U}$. We will write ψ as a composition of four mappings,

$$\psi = \psi_4 \circ \psi_3 \circ \psi_2 \circ \psi_1.$$

The mapping ψ_i essentially executes the ith step of the CFIA in table 3.1. More precisely, in terms of the inner variables of the CFIA, we will make our choices for the four mappings in order to have the input-output relationships expressed in figure 3.8.

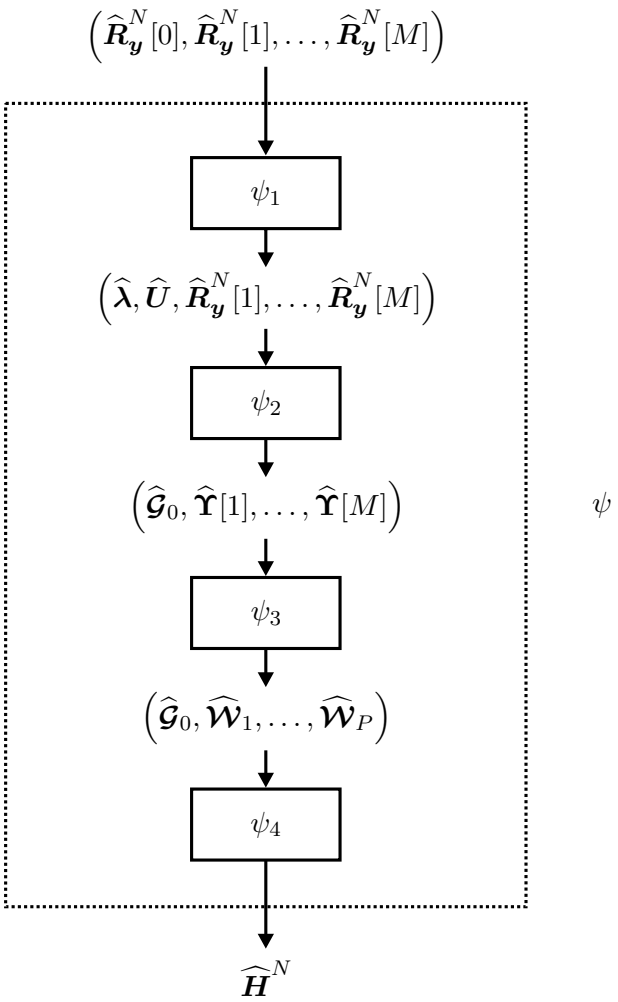

Figure 3.8: The map ψ as the composition $\psi = \psi_4 \circ \psi_3 \circ \psi_2 \circ \psi_1$

We start by defining the mappings ψ_i in the reversed order. That is, we begin with ψ_4 and finish with ψ_1. For each mapping ψ_i, we prove that it is smooth, and compute its derivative at an arbitrary point of its (open) domain. The actual computational details are left to the appendix C. The derivative of ψ at $(\boldsymbol{R_y}[0], \boldsymbol{R_y}[1], \ldots, \boldsymbol{R_y}[M]) \in \mathcal{U}$ is then obtained by the chain rule.

Mapping ψ_4: definition. This mapping performs step 4) of the CFIA in table 3.1, see also figure 3.8. In particular, we must have

$$\left(\widehat{\boldsymbol{\mathcal{G}}}_0, \widehat{\boldsymbol{\mathcal{W}}}_1, \ldots, \widehat{\boldsymbol{\mathcal{W}}}_P\right) \quad \stackrel{\psi_4}{\longmapsto} \quad \widehat{\boldsymbol{H}}^N.$$

We let it be of the form

$$\psi_4 : \mathcal{U}_4 \subset \underbrace{\mathbb{C}^{Q\times D} \times \underbrace{\mathbb{C}^{D\times(D_0+1)} \times \cdots \times \mathbb{C}^{D\times(D_0+1)}}_{P}}_{\mathcal{C}_4} \to \mathbb{C}^{Q\times D}$$

where

$$\mathcal{U}_4 = \{\, (\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_P) \in \mathcal{C}_4 \,:\, \operatorname{rank} \boldsymbol{Z}_0 = D,\, \operatorname{rank} [\, \boldsymbol{Z}_1 \,\cdots\, \boldsymbol{Z}_P \,] = P(D_0+1) = D \,\}\,.$$

We define

$$\psi_4 (\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_P) = \boldsymbol{Z}_0 \, [\, \boldsymbol{Z}_1 \,\cdots\, \boldsymbol{Z}_P \,] \, \boldsymbol{R_s}[0; \boldsymbol{d}_0]^{-1/2}. \tag{3.71}$$

Note that $\mathcal{U}_4$ is an open subset of $\mathcal{C}_4$ and $\operatorname{image} \psi_4 \subset \mathbb{C}^*_{\boldsymbol{d}_0} = \mathbb{H}[z] \subset \mathbb{C}_{\boldsymbol{d}_0} = \mathbb{C}^{Q\times D}$. Moreover, ψ_4 is smooth on all of its domain $\mathcal{U}_4$, since only elementary operations (matrix multiplications, and so forth) are involved. The derivative of ψ_4 at an arbitrary point $(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_P)$ of its domain $\mathcal{U}_4$ is computed in appendix C, section C.1.

Mapping ψ_3: definition. Before defining the mapping ψ_3, we state the technical lemma 3.6 which establishes the local differentiability of simple eigenvalues and associated unit-norm eigenvectors of Hermitean matrices, as functions of the matrix entries. This lemma will be invoked to compute the derivative of ψ_3 (and also ψ_1). The issue of differentiability of simple eigenvalues and associated eigenvectors of complex matrices is addressed in [39]. However, the result therein is inadequate for our purposes for two reasons. **i)** Let λ_0 denote a simple (that is, with multiplicity 1), possibly complex, eigenvalue of a $n \times n$ complex matrix $\boldsymbol{Z}_0$. Let $\boldsymbol{q}_0$ denote an associated unit-norm eigenvector, that is, $\boldsymbol{Z}_0 \boldsymbol{q}_0 = \lambda_0 \boldsymbol{q}_0$ with $\boldsymbol{q}_0^H \boldsymbol{q}_0 = 1$. The result in [39] states that there exists an open set $U \subset \mathbb{C}^{n\times n}$ containing $\boldsymbol{Z}_0$ and smooth mappings $\lambda : U \to \mathbb{C}$ and $\boldsymbol{q} : U \to \mathbb{C}^n$ such that $\lambda(\boldsymbol{Z}_0) = \lambda_0$, $\boldsymbol{q}(\boldsymbol{Z}_0) = \boldsymbol{q}_0$ and

$$\begin{aligned} \boldsymbol{Z}\, \boldsymbol{q}(\boldsymbol{Z}) &= \lambda(\boldsymbol{Z})\, \boldsymbol{q}(\boldsymbol{Z}) \\ \boldsymbol{q}_0^H \boldsymbol{q}(\boldsymbol{Z}) &= 1, \end{aligned} \tag{3.72}$$

for all $\boldsymbol{Z} \in U$. The first reason this result is inadequate for our purposes lies in the way the eigenvector $\boldsymbol{q}$ extending $\boldsymbol{q}_0$ is normalized in (3.72). It is not normalized to unit-power,

as we want it to be. Instead, it is scaled to satisfy a linear constraint. **ii)** The second, and more important, reason lies in the fact that the implementation of this version of the EVD requires the previous knowledge of an eigenvector $\boldsymbol{q}_0$ of $\boldsymbol{Z}_0$. As can be seen in (3.72), this knowledge is needed to scale eigenvectors of matrices $\boldsymbol{Z}$ in a vicinity of $\boldsymbol{Z}_0$. In our context, this means that, to code the CFIA with this version of the EVD, we should know beforehand, for example, the eigenvectors of $\boldsymbol{Z}_0 = \boldsymbol{R}_{\boldsymbol{y}}[0]$, because in step 1) of the CFIA we calculate eigenvalues and eigenvectors of matrices $\boldsymbol{Z} = \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[0]$ in the vicinity of $\boldsymbol{Z}_0 = \boldsymbol{R}_{\boldsymbol{y}}[0]$. Clearly, such knowledge is not available in practice. But, even so, this would make the CFIA unable to handle more than one MIMO channel (if the channel changes, the correlation matrix at the output may change, and, as a consequence, also its eigenstructure). Thus, what we need is a version of the EVD which allows for differentiable unit-norm eigenvectors along with a more realistic normalization procedure. This is proposed in lemma 3.6. Before stating lemma 3.6 we need some notation. Let $V \subset \mathbb{C}^n$ denote a linear subspace. We say that the vector $\boldsymbol{c} \in \mathbb{C}^n$ is oblique to V, if it does not lie in the orthogonal complement of V, that is,

$$\boldsymbol{c} \notin V^{\perp} = \left\{ \boldsymbol{z} \,:\, \boldsymbol{z}^H \boldsymbol{v} = 0, \text{ for all } \boldsymbol{v} \in V \right\}.$$

If λ denotes an eigenvalue of a $n \times n$ complex matrix $\boldsymbol{Z}$, we let $\mathrm{V}_{\lambda}(\boldsymbol{Z})$ denote its associated eigenspace. That is,

$$\mathrm{V}_{\lambda}(\boldsymbol{Z}) = \{ \boldsymbol{q} \in \mathbb{C}^n \,:\, \boldsymbol{Z}\boldsymbol{q} = \lambda \boldsymbol{q} \}.$$

Suppose λ is a simple eigenvalue of $\boldsymbol{Z}$ and $\boldsymbol{c}$ is oblique to $\mathrm{V}_{\lambda}(\boldsymbol{Z})$. It is clear that there is one and only one unit-norm eigenvector $\boldsymbol{q} \in \mathrm{V}_{\lambda}(\boldsymbol{Z})$ such that $\mathrm{Re}\, \boldsymbol{c}^H \boldsymbol{q} > 0$ and $\mathrm{Im}\, \boldsymbol{c}^H \boldsymbol{q} = 0$. We denote such eigenvector by $\boldsymbol{q}(\boldsymbol{Z}; \lambda; \boldsymbol{c})$.

Lemma 3.6. *Let $\lambda_0 \in \mathbb{R}$ denote a simple eigenvalue of a $n \times n$ Hermitean matrix $\boldsymbol{Z}_0$. Suppose $\boldsymbol{c}_0 \in \mathbb{C}^n$ is oblique to $\mathrm{V}_{\lambda_0}(\boldsymbol{Z}_0)$ and let $\boldsymbol{q}_0 = \boldsymbol{q}(\boldsymbol{Z}_0; \lambda_0; \boldsymbol{c}_0)$. Then, there exist open sets $U \subset \mathbb{C}^{n \times n}$ and $V \subset \mathbb{C} \times \mathbb{C}^n$ containing $\boldsymbol{Z}_0$ and $(\lambda_0, \boldsymbol{q}_0)$, respectively, such that, for each $\boldsymbol{Z} \in U$ there exists one and only one $(\lambda, \boldsymbol{q}) \in V$ satisfying $\boldsymbol{Z}\boldsymbol{q} = \lambda \boldsymbol{q}$, $\boldsymbol{q}^H \boldsymbol{q} = 1$, $\mathrm{Re}\, \boldsymbol{c}_0^H \boldsymbol{q} > 0$, $\mathrm{Im}\, \boldsymbol{c}_0^H \boldsymbol{q} = 0$ and λ is a simple eigenvalue of $\boldsymbol{Z}$. Denote this implicit mapping by $\boldsymbol{Z} \in U \mapsto (\lambda(\boldsymbol{Z}), \boldsymbol{q}(\boldsymbol{Z})) \in V$. Note that, for each $\boldsymbol{Z} \in U$, $\boldsymbol{c}_0$ is oblique to $\mathrm{V}_{\lambda(\boldsymbol{Z})}(\boldsymbol{Z})$ and $\boldsymbol{q}(\boldsymbol{Z}) = \boldsymbol{q}(\boldsymbol{Z}; \lambda(\boldsymbol{Z}); \boldsymbol{c}_0)$. The implicit mapping is smooth (infinitely differentiable) and the differentials $d\lambda$ and $d\boldsymbol{q}$, evaluated at $\boldsymbol{Z}_0$, are given by*

$$d\lambda = \boldsymbol{q}_0^H d\boldsymbol{Z} \boldsymbol{q}_0 \tag{3.73}$$

$$d\boldsymbol{q} = (\lambda_0 \boldsymbol{I}_n - \boldsymbol{Z}_0)^{+} d\boldsymbol{Z} \boldsymbol{q}_0 - i \frac{\mathrm{Im}\left\{ \boldsymbol{c}_0^H (\lambda_0 \boldsymbol{I}_n - \boldsymbol{Z}_0)^{+} d\boldsymbol{Z} \boldsymbol{q}_0 \right\}}{\boldsymbol{c}_0^H \boldsymbol{q}_0} \boldsymbol{q}_0. \tag{3.74}$$

Proof: See appendix B.

To implement this version of the EVD it is only required that $\boldsymbol{c}_0$ does not lie in the orthogonal complement of the one-dimensional eigenspace $\mathrm{V}_{\lambda_0}(\boldsymbol{Z}_0)$. Although, in practice, $\mathrm{V}_{\lambda_0}(\boldsymbol{Z}_0)$ is not known beforehand, almost all $\boldsymbol{c}_0$ do satisfy this condition. More precisely, those which do not verify it are confined to a Lebesgue measure zero set, the linear slice orthogonal to $\mathrm{V}_{\lambda_0}(\boldsymbol{Z}_0)$.

For our purposes, we need the derivatives, as defined in page 60, of the complex mappings $\lambda : U \subset \mathbb{C}^{n\times n} \to \mathbb{C}$ and $\boldsymbol{q} : U \subset \mathbb{C}^{n\times n} \to \mathbb{C}^n$ at $\boldsymbol{Z}_0$. These are straightforward to obtain from the differentials in (3.73) and (3.74). More precisely, $D\lambda\,(\boldsymbol{Z}_0)$ and $D\boldsymbol{q}\,(\boldsymbol{Z}_0)$ are uniquely defined by

$$\begin{aligned} \imath\,(d\lambda) &= D\lambda\,(\boldsymbol{Z}_0)\;\imath\,(d\boldsymbol{Z}) \\ \imath\,(d\boldsymbol{q}) &= D\boldsymbol{q}\,(\boldsymbol{Z}_0)\;\imath\,(d\boldsymbol{Z})\,. \end{aligned}$$

Plugging $d\boldsymbol{Z} = \mathrm{Re}\,\{d\boldsymbol{Z}\} + i\,\mathrm{Im}\,\{d\boldsymbol{Z}\}$ in (3.73) and (3.74), and working out the details yields

$$D\lambda\,(\boldsymbol{Z}_0) = \jmath\left(\boldsymbol{q}_0^T \otimes \boldsymbol{q}_0^H\right) \tag{3.75}$$

$$D\boldsymbol{q}\,(\boldsymbol{Z}_0) = \left(\boldsymbol{I}_{2n} - \frac{1}{\boldsymbol{c}_0^H\boldsymbol{q}_0}\begin{bmatrix} \mathrm{Im}\,\boldsymbol{q}_0 \\ -\mathrm{Re}\,\boldsymbol{q}_0 \end{bmatrix}\begin{bmatrix} \mathrm{Im}\,\boldsymbol{c}_0^T & -\mathrm{Re}\,\boldsymbol{c}_0^T \end{bmatrix}\right)\jmath\left(\boldsymbol{q}_0^T \otimes (\lambda_0\boldsymbol{I}_n - \boldsymbol{Z}_0)^+\right). \tag{3.76}$$

We use the notation $D\lambda\,(\boldsymbol{Z}_0;\lambda_0;\boldsymbol{c}_0)$ and $D\boldsymbol{q}\,(\boldsymbol{Z}_0;\lambda_0;\boldsymbol{c}_0)$ to designate the matrices on the right-hand sides of (3.75) and (3.76), respectively. The symbol $\boldsymbol{q}_0$ is not included in this notation, because $\boldsymbol{q}_0$ is not an independent entity. Recall that it is just a shorthand for $\boldsymbol{q}\,(\boldsymbol{Z}_0;\lambda_0;\boldsymbol{c}_0)$.

The mapping ψ_3 must perform step 3) of the CFIA in table 3.1. Recall also figure 3.8. It must satisfy

$$\left(\widehat{\boldsymbol{\mathcal{G}}}_0, \widehat{\boldsymbol{\Upsilon}}[1], \ldots, \widehat{\boldsymbol{\Upsilon}}[M]\right) \quad \overset{\psi_3}{\longmapsto} \quad \left(\widehat{\boldsymbol{\mathcal{G}}}_0, \widehat{\boldsymbol{\mathcal{W}}}_1, \ldots, \widehat{\boldsymbol{\mathcal{W}}}_P\right).$$

We will define it in terms of auxiliary mappings η_p and ϑ_p in order to have the input-output relationships described in figure 3.9.

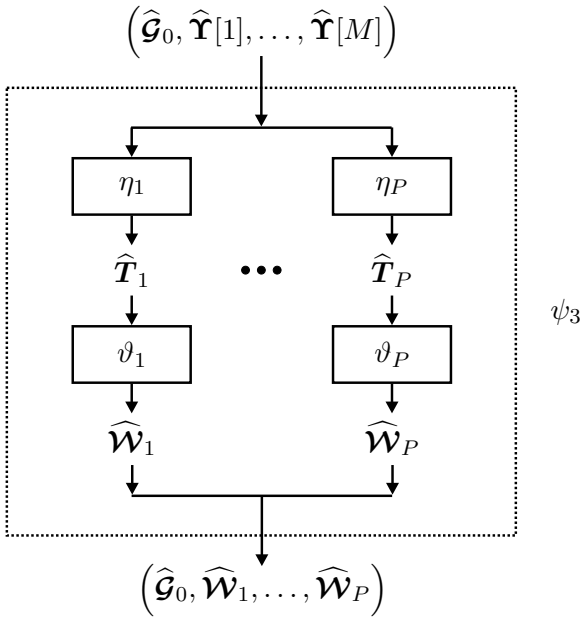

Figure 3.9: The internal structure of the map ψ_3

The pth mapping

$$\eta_p : \underbrace{\mathbb{C}^{D\times D} \times \cdots \times \mathbb{C}^{D\times D}}_{M} \to \mathbb{C}^{2MD(D_0+1)\times D(D_0+1)}$$

is given by

$$\eta_p\left(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M\right) = \begin{bmatrix} \boldsymbol{I}_{D_0+1} \otimes \boldsymbol{Z}_1 - \boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; D_0]^T \otimes \boldsymbol{I}_D \\ \boldsymbol{I}_{D_0+1} \otimes \boldsymbol{Z}_1^H - \overline{\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; D_0]} \otimes \boldsymbol{I}_D \\ \vdots \\ \boldsymbol{I}_{D_0+1} \otimes \boldsymbol{Z}_M - \boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; D_0]^T \otimes \boldsymbol{I}_D \\ \boldsymbol{I}_{D_0+1} \otimes \boldsymbol{Z}_M^H - \overline{\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; D_0]} \otimes \boldsymbol{I}_D \end{bmatrix}. \tag{3.77}$$

Note that, with this definition, $\widehat{\boldsymbol{T}}_p = \eta_p\left(\widehat{\boldsymbol{\Upsilon}}[1], \ldots, \widehat{\boldsymbol{\Upsilon}}[M]\right)$. To introduce the mappings ϑ_p, we need a bit of terminology. If $\boldsymbol{Z}$ is an Hermitean matrix, we denote its minimum eigenvalue by $\lambda_{\min}(\boldsymbol{Z})$. Also, to simplify notation, we let $\mathrm{V}_{\min}(\boldsymbol{Z}) = \mathrm{V}_{\lambda_{\min}(\boldsymbol{Z})}(\boldsymbol{Z})$. The pth mapping

$$\vartheta_p : \mathcal{V}_p \subset \mathbb{C}^{2MD(D_0+1)\times D(D_0+1)} \to \mathbb{C}^{D\times D(D_0+1)},$$

where

$$\mathcal{V}_p = \left\{ \boldsymbol{Z} : \dim \mathrm{V}_{\min}\left(\boldsymbol{Z}^H\boldsymbol{Z}\right) = 1,\ \boldsymbol{s}_p \notin \mathrm{V}_{\min}^{\perp}\left(\boldsymbol{Z}^H\boldsymbol{Z}\right) \right\},$$

is given by

$$\vartheta_p(\boldsymbol{Z}) = \sqrt{D_0+1}\,\mathrm{vec}^{-1}\left(\boldsymbol{q}\left(\boldsymbol{Z}^H\boldsymbol{Z}; \lambda_{\min}\left(\boldsymbol{Z}^H\boldsymbol{Z}\right); \boldsymbol{s}_p\right)\right).$$

Here, vec^{-1} denotes the inverse mapping of $\mathrm{vec} : \mathbb{C}^{D\times(D_0+1)} \to \mathbb{C}^{D(D_0+1)}$. Note that the domain of ϑ_p, denoted above by $\mathcal{V}_p$, consists of those $2MD(D_0+1)\times D(D_0+1)$ complex matrices $\boldsymbol{Z}$ for which the minimum eigenvalue of $\boldsymbol{Z}^H\boldsymbol{Z}$ is simple and the vector $\boldsymbol{s}_p$ is oblique to the corresponding eigenspace. Moreover, note that $\widehat{\boldsymbol{\mathcal{W}}}_p = \vartheta_p \circ \eta_p\left(\widehat{\boldsymbol{\Upsilon}}[1], \ldots, \widehat{\boldsymbol{\Upsilon}}[M]\right)$.

We define the mapping

$$\psi_3 : \mathcal{U}_3 \subset \underbrace{\mathbb{C}^{Q\times D} \times \underbrace{\mathbb{C}^{D\times D} \times \cdots \times \mathbb{C}^{D\times D}}_{M}}_{\mathcal{C}_3} \to \mathcal{C}_4,$$

where

$$\mathcal{U}_3 = \left\{ (\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M) \in \mathcal{C}_3 : (\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M) \in \eta_p^{-1}(\mathcal{V}_p),\ \text{for all } p = 1, \ldots, P \right\},$$

to be given by

$$\psi_3(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M) = \left(\boldsymbol{Z}_0, \vartheta_1 \circ \eta_1(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M), \ldots, \vartheta_P \circ \eta_P(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M)\right).$$

Note that $\mathcal{U}_3$ is an open subset of $\mathcal{C}_3$. Furthermore, the mapping ψ_3 is smooth on $\mathcal{U}_3$. This follows from the fact that both η_p (because it only involves elementary operations) and ϑ_p (by virtue of lemma 3.6) are smooth on their respective domains. Thus, in particular,

ψ_3 is continuous and, therefore, $\psi_3^{-1}(\mathcal{U}_4)$ is an open subset of $\mathcal{U}_3$. We now redefine the domain of ψ_3 as the open set $\mathcal{U}_3 \leftarrow \mathcal{U}_3 \cap \psi_3^{-1}(\mathcal{U}_4)$. This ensures that $\psi_3(\mathcal{U}_3) \subset \mathcal{U}_4$, that is, we may compose $\psi_4 \circ \psi_3$. The derivative of ψ_3 at an arbitrary point $(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M)$ of its domain $\mathcal{U}_3$, written $D\psi_3(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M)$, is computed in appendix C, section C.2.

Mapping ψ_2: definition. This mapping executes step 2) of the CFIA in table 3.1, see also figure 3.8, and must verify

$$\left(\widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{U}}, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^N[1], \ldots, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^N[M]\right) \quad \stackrel{\psi_2}{\longmapsto} \quad \left(\widehat{\boldsymbol{\mathcal{G}}}_0, \widehat{\boldsymbol{\Upsilon}}[1], \ldots, \widehat{\boldsymbol{\Upsilon}}[M]\right).$$

We introduce the auxiliary mappings

$$\tau : \mathcal{V} \subset \mathbb{C}^D \times \mathbb{C}^{Q\times D} \to \mathbb{C}^{Q\times D}, \quad \upsilon : \mathcal{V} \subset \mathbb{C}^D \times \mathbb{C}^{Q\times D} \to \mathbb{C}^{D\times Q},$$

where $\mathcal{V} = \left\{(\boldsymbol{z}, \boldsymbol{Z}) : \operatorname{Re}\boldsymbol{z} \succ \sigma^2 \mathbf{1}_D\right\}$, defined by

$$\tau(\boldsymbol{z}, \boldsymbol{Z}) = \boldsymbol{Z} \operatorname{Diag}\left(\operatorname{Re}\boldsymbol{z} - \sigma^2 \mathbf{1}_D\right)^{1/2} \tag{3.78}$$

$$\upsilon(\boldsymbol{z}, \boldsymbol{Z}) = \operatorname{Diag}\left(\operatorname{Re}\boldsymbol{z} - \sigma^2 \mathbf{1}_D\right)^{-1/2} \boldsymbol{Z}^H. \tag{3.79}$$

In terms of the inner variables of the CFIA, $\widehat{\boldsymbol{\mathcal{G}}}_0 = \tau\left(\widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{U}}\right)$ and $\widehat{\boldsymbol{\mathcal{G}}}_0^+ = \upsilon\left(\widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{U}}\right)$. We define

$$\psi_2 : \mathcal{U}_2 \subset \underbrace{\mathbb{C}^D \times \mathbb{C}^{Q\times D} \times \underbrace{\mathbb{C}^{Q\times Q} \times \cdots \times \mathbb{C}^{Q\times Q}}_{M}}_{\mathcal{C}_2} \to \mathcal{C}_3,$$

where

$$\mathcal{U}_2 = \left\{ (\boldsymbol{z}, \boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M) \in \mathcal{C}_2 : (\boldsymbol{z}, \boldsymbol{Z}_0) \in \mathcal{V} \right\},$$

to be given by

$$\psi_2(\boldsymbol{z}, \boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M) = \left(\tau(\boldsymbol{z}, \boldsymbol{Z}_0), \upsilon(\boldsymbol{z}, \boldsymbol{Z}_0) \boldsymbol{Z}_1 \upsilon(\boldsymbol{z}, \boldsymbol{Z}_0)^H, \ldots, \upsilon(\boldsymbol{z}, \boldsymbol{Z}_0) \boldsymbol{Z}_M \upsilon(\boldsymbol{z}, \boldsymbol{Z}_0)^H\right). \tag{3.80}$$

Note that $\mathcal{U}_2$ is an open subset of $\mathcal{C}_2$. Furthermore, because both τ and υ are smooth on their domain $\mathcal{V}$ the mapping ψ_2 is smooth on $\mathcal{U}_2$. We now redefine the domain of ψ_2 as the open set $\mathcal{U}_2 \leftarrow \mathcal{U}_2 \cap \psi_2^{-1}(\mathcal{U}_3)$, in order to satisfy $\psi_2(\mathcal{U}_2) \subset \mathcal{U}_3$. Thus, the composition $\psi_4 \circ \psi_3 \circ \psi_2$ is well-defined. The derivative of ψ_2 at an arbitrary point $(\boldsymbol{z}, \boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M)$ of its domain $\mathcal{U}_2$ is computed in appendix C, section C.3.

Mapping ψ_1: definition. This is the mapping performing step 1) of the CFIA in table 3.1. According to figure 3.8 it must operate as to satisfy

$$\left(\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^N[0], \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^N[1], \ldots, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^N[M]\right) \quad \stackrel{\psi_1}{\longmapsto} \quad \left(\widehat{\boldsymbol{\lambda}}, \widehat{\boldsymbol{U}}, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^N[1], \ldots, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^N[M]\right).$$

In order to describe it succinctly we introduce some notation. We denote by $\mathcal{S}_D$ the subset of $Q \times Q$ Hermitean matrices whose largest D eigenvalues are simple. For $\boldsymbol{Z} \in \mathcal{S}_D$, we let $\lambda_1(\boldsymbol{Z}) > \cdots > \lambda_D(\boldsymbol{Z})$ denote its largest D eigenvalues arranged in strictly decreasing order. We also define the vector $\boldsymbol{\lambda}(\boldsymbol{Z}) = (\lambda_1(\boldsymbol{Z}), \ldots, \lambda_D(\boldsymbol{Z}))^T$. Let $\boldsymbol{c}_d$ denote a vector oblique to

the one-dimensional eigenspace $\mathrm{V}_{\lambda_d(\boldsymbol{Z})}(\boldsymbol{Z})$. Recall that $\boldsymbol{q}(\boldsymbol{Z};\lambda_d(\boldsymbol{Z});\boldsymbol{c}_d)$ designates the unique unit-norm eigenvector $\boldsymbol{q} \in \mathrm{V}_{\lambda_d(\boldsymbol{Z})}(\boldsymbol{Z})$ which satisfies $\mathrm{Re}\,\boldsymbol{c}_d^H\boldsymbol{q} > 0$ and $\mathrm{Im}\,\boldsymbol{c}_d^H\boldsymbol{q} = 0$. These eigenvectors can be arranged in the matrix

$$\boldsymbol{Q}(\boldsymbol{Z};\boldsymbol{c}_1,\ldots,\boldsymbol{c}_D) = \left[\begin{array}{cccc} \boldsymbol{q}(\boldsymbol{Z};\lambda_1(\boldsymbol{Z});\boldsymbol{c}_1) & \boldsymbol{q}(\boldsymbol{Z};\lambda_2(\boldsymbol{Z});\boldsymbol{c}_2) & \cdots & \boldsymbol{q}(\boldsymbol{Z};\lambda_D(\boldsymbol{Z});\boldsymbol{c}_D) \end{array}\right]. \tag{3.81}$$

Note that each $Q \times Q$ complex matrix can be uniquely decomposed as $\boldsymbol{Z} = \boldsymbol{Z}_h + \boldsymbol{Z}_s$, where $\boldsymbol{Z}_h = (\boldsymbol{Z} + \boldsymbol{Z}^H)/2$ is Hermitean and $\boldsymbol{Z}_s = (\boldsymbol{Z} - \boldsymbol{Z}^H)/2$ is skew-Hermitean. We denote by $\rho : \mathbb{C}^{Q\times Q} \to \mathbb{C}^{Q\times Q}$ the projector

$$\rho(\boldsymbol{Z}) = \frac{\boldsymbol{Z} + \boldsymbol{Z}^H}{2}, \tag{3.82}$$

retrieving the Hermitean part of its matrix argument.

We define the mapping

$$\psi_1 : \mathcal{U}_1 \subset \underbrace{\underbrace{\mathbb{C}^{Q\times Q} \times \cdots \times \mathbb{C}^{Q\times Q}}_{M+1}}_{\mathcal{C}_1} \to \mathcal{C}_2,$$

where

$$\mathcal{U}_1 = \left\{(\boldsymbol{Z}_0,\boldsymbol{Z}_1,\ldots,\boldsymbol{Z}_M) \in \mathcal{C}_1 \,:\, \rho(\boldsymbol{Z}_0) \in \mathcal{S}_D,\, \boldsymbol{r}_d \notin \mathrm{V}_d^{\perp}(\rho(\boldsymbol{Z}_0)), d = 1,\ldots,D\right\},$$

to be given by

$$\psi_1(\boldsymbol{Z}_0,\boldsymbol{Z}_1,\ldots,\boldsymbol{Z}_M) = (\boldsymbol{\lambda}(\rho(\boldsymbol{Z}_0)), \boldsymbol{Q}(\rho(\boldsymbol{Z}_0);\boldsymbol{r}_1,\ldots,\boldsymbol{r}_D), \boldsymbol{Z}_1,\ldots,\boldsymbol{Z}_M). \tag{3.83}$$

Note that $\mathcal{U}_1$ is an open subset of $\mathcal{C}_1$ and ψ_1 is smooth on it, because simple eigenvalues and associated eigenvectors of Hermitean matrices are smooth functions of its entries, as established in lemma 3.6. As expected, we now redefine the domain of ψ_1 to be $\mathcal{U}_1 \leftarrow \mathcal{U}_1 \cap \psi_1^{-1}(\mathcal{U}_2)$. This makes the composition $\psi_4 \circ \psi_3 \circ \psi_2 \circ \psi_1$ well-defined. The derivative of ψ_1 at an arbitrary point $(\boldsymbol{Z}_0,\boldsymbol{Z}_1,\ldots,\boldsymbol{Z}_M)$ of its domain $\mathcal{U}_1$, denote by the symbol $D\psi_1(\boldsymbol{Z}_0,\boldsymbol{Z}_1,\ldots,\boldsymbol{Z}_M)$, is computed in appendix C, section C.4.

Derivative of ψ. The mapping $\psi : \mathcal{U} \subset \mathbb{C}^{Q\times Q} \times \cdots \times \mathbb{C}^{Q\times Q} \to \mathbb{C}_{\boldsymbol{d}_0} = \mathbb{C}^{Q\times D}$ is defined by putting $\mathcal{U} = \mathcal{U}_1$ and letting $\psi = \psi_4 \circ \psi_3 \circ \psi_2 \circ \psi_1$. By the above considerations, $\mathcal{U}$ is open, ψ is smooth on it, $\mathrm{image}\,\psi \subset \mathbb{C}^*_{\boldsymbol{d}_0} = \mathbb{H}[z]$, and we have

$$\widehat{\boldsymbol{H}}^N = \psi\left(\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^N[0], \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^N[1], \ldots, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^N[M]\right).$$

That is, ψ is a smooth mapping performing steps 1) to 4) of the CFIA in table 3.1. The derivative of ψ at the point $(\boldsymbol{R}_{\boldsymbol{y}}[0], \boldsymbol{R}_{\boldsymbol{y}}[1], \ldots, \boldsymbol{R}_{\boldsymbol{y}}[M])$ is given by the chain rule. That is,

$$D\psi(\boldsymbol{R}_{\boldsymbol{y}}[0], \boldsymbol{R}_{\boldsymbol{y}}[1], \ldots, \boldsymbol{R}_{\boldsymbol{y}}[M]) = D\psi_4 \circ D\psi_3 \circ D\psi_2 \circ D\psi_1(\boldsymbol{R}_{\boldsymbol{y}}[0], \boldsymbol{R}_{\boldsymbol{y}}[1], \ldots, \boldsymbol{R}_{\boldsymbol{y}}[M]).$$

3.4.3 Asymptotic normality of $\left(\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[0], \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[1], \ldots, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[M]\right)$

Before proving the asymptotic normality of the ordered $(M+1)$-tuple of random matrices $\left(\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[0], \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[1], \ldots, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[M]\right)$, we establish some lemmas regarding the asymptotic normality of certain random processes. This will shorten the proof substantially. Let $\boldsymbol{x}, \boldsymbol{y}$ denote random vectors, not necessarily of the same size. The notation $\boldsymbol{x} \perp \boldsymbol{y}$ means that $\boldsymbol{x}$ and $\boldsymbol{y}$ are statistically independent. Let $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \ldots, \kappa_P)^T$ denote a P-dimensional vector with non-negative entries. We say that the real P-dimensional random vector $\boldsymbol{x} = (x_1, x_2, \ldots, x_P)^T$ belongs to the class $\mathcal{R}(\boldsymbol{\kappa})$, written $\boldsymbol{x} \in \mathcal{R}(\boldsymbol{\kappa})$, if its entries are mutually independent ($x_p \perp x_q$ for $p \neq q$) and their first four moments satisfy $\mathrm{E}\{x_p\} = 0$, $\mathrm{E}\{x_p^2\} = 1$, $\mathrm{E}\{x_p^3\} = 0$, and $\mathrm{E}\{x_p^4\} = \kappa_p$, for $p = 1, 2, \ldots, P$. Lemma 3.7 provides a formula for the fourth order moments of random vectors $\boldsymbol{x}$ in $\mathcal{R}(\boldsymbol{\kappa})$, or equivalently, for the correlation matrix of the random vector $\boldsymbol{x} \otimes \boldsymbol{x}$.

Lemma 3.7. *Suppose that the P-dimensional random vector $\boldsymbol{x}$ belongs to the class $\mathcal{R}(\boldsymbol{\kappa})$. Then,*

$$\operatorname{corr}\{\boldsymbol{x} \otimes \boldsymbol{x}\} = \boldsymbol{I}_{P^2} + \boldsymbol{K}_P + \boldsymbol{i}_P \boldsymbol{i}_P^T + \operatorname{diag}\left((\kappa_1 - 3)\,\boldsymbol{e}_1 \boldsymbol{e}_1^T, \ldots, (\kappa_P - 3)\,\boldsymbol{e}_P \boldsymbol{e}_P^T\right), \tag{3.84}$$

where $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \ldots, \kappa_P)^T$ and $\boldsymbol{e}_p$ denotes the pth column of the identity matrix $\boldsymbol{I}_P$.

Proof: See appendix B.

In lemma 3.7, and for further reference, we define $\boldsymbol{i}_n = \operatorname{vec}(\boldsymbol{I}_n)$, the n^2-dimensional column vector obtained by vectorizing the $n \times n$ identity matrix $\boldsymbol{I}_n$. For $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \ldots, \kappa_P)^T$, we define $\boldsymbol{\Sigma}(\boldsymbol{\kappa})$ as the matrix in (3.84).

Let $x_p[n]$, $p = 1, 2, \ldots, P$, denote independent random signals with each $x_p[n]$ being a sequence of independent realizations of $\mathcal{R}(\kappa_p)$. That is, each pth signal is statistically white ($x_p[n] \perp x_p[m]$ for $n \neq m$) and $x_p[n] \in \mathcal{R}(\kappa_p)$ for all $n \in \mathbb{Z}$. Moreover, $x_p[n] \perp x_q[m]$ for all n, m and $p \neq q$. In that case, we say that the discrete-time random signal $\boldsymbol{x}[n] = (x_1[n], x_2[n], \ldots, x_P[n])^T$ belongs to the class $\mathcal{R}_{\mathbb{Z}}(\boldsymbol{\kappa})$, where $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \ldots, \kappa_P)^T$. Lemma 3.8 establishes the asymptotic normality of the finite-sample estimate of the correlation matrix of the random signal $\boldsymbol{x}[n; \boldsymbol{l}]$, for any given choice of stacking parameters in $\boldsymbol{l} = (L_1, L_2, \ldots, L_P)^T$.

Lemma 3.8. *Let $\boldsymbol{x}[n] = (x_1[n], x_2[n], \ldots, x_P[n])^T$ belong to $\mathcal{R}_{\mathbb{Z}}(\boldsymbol{\kappa})$. Then, for given $\boldsymbol{l} = (L_1, L_2, \ldots, L_P)^T$, the sequence of random matrices (indexed by N)*

$$\boldsymbol{R}^N = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}[n; \boldsymbol{l}] \boldsymbol{x}[n; \boldsymbol{l}]^T$$

is asymptotically normal. More precisely, we have

$$\boldsymbol{R}^N \sim \sqrt{N} - \mathcal{AN}\left(\boldsymbol{I}_L, \boldsymbol{R}(\boldsymbol{\kappa}; \boldsymbol{l})\right),$$

where $L = L_1 + \cdots + L_P + P$ and

$$\boldsymbol{R}(\boldsymbol{\kappa}; \boldsymbol{l}) = \boldsymbol{D}(\boldsymbol{l})^{[2]} \left(\boldsymbol{R}_0 + 2 \sum_{l=1}^{L_0} \boldsymbol{R}_l - (2L_0 + 1) \boldsymbol{i}_{P(L_0+1)} \boldsymbol{i}_{P(L_0+1)}^T \right) \boldsymbol{D}(\boldsymbol{l})^{[2]^T},$$

with $L_0 = \max\{L_1, \ldots, L_P\}$, $\boldsymbol{D}(\boldsymbol{l}) = \operatorname{diag}(\boldsymbol{D}_1, \ldots, \boldsymbol{D}_P)$, $\boldsymbol{D}_p = \left[\, \boldsymbol{I}_{L_p+1}\, \boldsymbol{0}_{(L_p+1)\times(L_0-L_p)} \,\right]$, *and* $\boldsymbol{R}_l = \boldsymbol{S}^{[2]}_{P,L_0,l}\, \boldsymbol{\Sigma}(\boldsymbol{\kappa} \otimes \boldsymbol{1}_{L_0+l+1})\, \boldsymbol{T}^{[2]}_{P,l,L_0}{}^{T}$.

Proof: See appendix B.

In lemma 3.8, and in the sequel, $\boldsymbol{A}^{[n]} = \boldsymbol{A} \otimes \cdots \otimes \boldsymbol{A}$ denotes the n-fold Kronecker product of $\boldsymbol{A}$. Also,

$$\begin{aligned} \boldsymbol{S}_{p,n,i} &= \boldsymbol{I}_p \otimes \left[\, \boldsymbol{I}_{n+1} \quad \boldsymbol{0}_{(n+1)\times i} \,\right] \\ \boldsymbol{T}_{p,i,n} &= \boldsymbol{I}_p \otimes \left[\, \boldsymbol{0}_{(n+1)\times i} \quad \boldsymbol{I}_{n+1} \,\right], \end{aligned}$$

for $p \geq 1$ and $n, i \geq 0$. The matrix $\boldsymbol{S}_{p,n,i}$ consists of p copies, concatenated along the diagonal, of a $(n+1) \times (n+i+1)$ block. The same applies to $\boldsymbol{T}_{p,i,n}$. The generalization of lemma 3.8 to complex random signals is more useful to our purposes. We say that the discrete-time complex random signal $\boldsymbol{x}[n] = (x_1[n], x_2[n], \ldots, x_P[n])^T$ belongs to the class $\mathcal{C}_{\mathbb{Z}}(\boldsymbol{\kappa})$, where $\boldsymbol{\kappa} = (\kappa_1, \kappa_2, \ldots, \kappa_P)^T$, if the discrete-time real random signal $\imath(\boldsymbol{x}[n])$ belongs to $\mathcal{R}_{\mathbb{Z}}\left(\boldsymbol{\kappa}^{(2)}\right)$. Here, and for further reference, we use the notation $\boldsymbol{a}^{(n)} = \boldsymbol{1}_n \otimes \boldsymbol{a}$ to designate the vector consisting of n concatenated copies of the vector $\boldsymbol{a}$.

Lemma 3.9. *Let* $\boldsymbol{x}[n] = (x_1[n], x_2[n], \ldots, x_P[n])^T$ *belong to* $\mathcal{C}_{\mathbb{Z}}(\boldsymbol{\kappa})$. *Then, for given* $\boldsymbol{l} = (L_1, L_2, \ldots, L_P)^T$, *the sequence of random matrices (indexed by* N*)*

$$\boldsymbol{R}^N = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{x}[n; \boldsymbol{l}] \boldsymbol{x}[n; \boldsymbol{l}]^H$$

is asymptotically normal. More precisely, we have

$$\boldsymbol{R}^N \sim \sqrt{N} - \mathcal{AN}\left(2\boldsymbol{I}_L, \boldsymbol{C}(\boldsymbol{\kappa}; \boldsymbol{l})\right),$$

where $L = L_1 + \cdots + L_P + P$ *and* $\boldsymbol{C}(\boldsymbol{\kappa}; \boldsymbol{l}) = \boldsymbol{E}[L] \boldsymbol{R}(\boldsymbol{\kappa}^{(2)}; \boldsymbol{l}^{(2)}) \boldsymbol{E}[L]^T$.

Proof: See appendix B.

In lemma 3.9, we use the notation

$$\boldsymbol{E}[m] = \left[\, \begin{array}{c} \boldsymbol{E}_R[m] \otimes \boldsymbol{E}_R[m] + \boldsymbol{E}_I[m] \otimes \boldsymbol{E}_I[m] \\ \boldsymbol{E}_R[m] \otimes \boldsymbol{E}_I[m] - \boldsymbol{E}_I[m] \otimes \boldsymbol{E}_R[m] \end{array} \,\right],$$

where $\boldsymbol{E}_R[m] = [\, \boldsymbol{I}_m\, \boldsymbol{0}_{m\times m} \,]$ and $\boldsymbol{E}_I[m] = [\, \boldsymbol{0}_{m\times m}\, \boldsymbol{I}_m \,]$, for any non-negative integer m.

In this section, we establish the asymptotic normality of the ordered $(M+1)$-tuple of random matrices $\left(\widehat{\boldsymbol{R}}^N_{\boldsymbol{y}}[0], \widehat{\boldsymbol{R}}^N_{\boldsymbol{y}}[1], \ldots, \widehat{\boldsymbol{R}}^N_{\boldsymbol{y}}[M]\right)$, where

$$\widehat{\boldsymbol{R}}^N_{\boldsymbol{y}}[m] = \frac{1}{N} \sum_{n=1}^{N} \boldsymbol{y}[n] \boldsymbol{y}[n-m]^H. \tag{3.85}$$

More precisely, we show that

$$\left(\widehat{\boldsymbol{R}}^N_{\boldsymbol{y}}[0], \widehat{\boldsymbol{R}}^N_{\boldsymbol{y}}[1], \ldots, \widehat{\boldsymbol{R}}^N_{\boldsymbol{y}}[M]\right) \sim \sqrt{N} - \mathcal{AN}\left(\left(\boldsymbol{R}_{\boldsymbol{y}}[0], \boldsymbol{R}_{\boldsymbol{y}}[1], \ldots, \boldsymbol{R}_{\boldsymbol{y}}[M]\right), \boldsymbol{\Sigma}\right), \tag{3.86}$$

for a certain covariance matrix $\boldsymbol{\Sigma}$, to be explicited in the sequel. We start by performing a conceptual twist on the MIMO channel model, which, roughly, consists in interpreting the Q random noise processes $w_q[n]$ as extra (virtual) sources. That is, we rewrite the observations in the condensed (noiseless) form

$$\boldsymbol{y}[n] = \boldsymbol{\mathcal{G}}\boldsymbol{\gamma}[n] \tag{3.87}$$

as follows. Plug the pre-filtering identities

$$s_p[n] = c_p(z) \circledast a_p[n]$$

into the convolution-and-add equation

$$\boldsymbol{y}[n] = \sum_{p=1}^{P} \boldsymbol{h}_p(z) \circledast s_p[n] + \boldsymbol{w}[n].$$

We have

$$\begin{aligned}
\boldsymbol{y}[n] &= \sum_{p=1}^{P} (\boldsymbol{h}_p(z)c_p(z)) \circledast a_p[n] + \boldsymbol{w}[n] \\
&= \sum_{p=1}^{P} \left(\frac{1}{\sqrt{2}}\boldsymbol{h}_p(z)c_p(z) \right) \circledast \underbrace{\left(\sqrt{2}a_p[n]\right)}_{\alpha_p[n]} + \frac{\sigma}{\sqrt{2}} \underbrace{\left(\frac{\sqrt{2}}{\sigma}\boldsymbol{w}[n] \right)}_{\boldsymbol{\beta}[n]} \\
&= \sum_{p=1}^{P} \boldsymbol{\mathcal{G}}_p \, \alpha_p[n; E_0] + \frac{\sigma}{\sqrt{2}}\boldsymbol{\beta}[n] \\
&= \underbrace{[\ \boldsymbol{\mathcal{G}}_1 \quad \boldsymbol{\mathcal{G}}_2 \quad \cdots \quad \boldsymbol{\mathcal{G}}_P \quad \tfrac{\sigma}{\sqrt{2}}\boldsymbol{I}_Q\ ]}_{\boldsymbol{\mathcal{G}}} \underbrace{\begin{bmatrix} \alpha_1[n; E_0] \\ \vdots \\ \alpha_P[n; E_0] \\ \boldsymbol{\beta}[n] \end{bmatrix}}_{\boldsymbol{\gamma}[n]},
\end{aligned}$$

where we defined $E_0 = C_0 + D_0$, and $\boldsymbol{\mathcal{G}}_p = \boldsymbol{\mathcal{T}}_0\left(\boldsymbol{h}_p(z)c_p(z)/\sqrt{2}\right)$. Notice the appearance of $\alpha_p[n] = \sqrt{2}a_p[n]$ and $\beta_q[n] = \sigma/\sqrt{2}w_q[n]$ which denote power-scaled versions of the information $a_p[n]$ and noise $w_q[n]$ random sequences, respectively. Due to assumption B3 in page 52, we have

$$\operatorname{Re}\alpha_p[n] \overset{d}{=} \operatorname{Im}\alpha_p[n] \overset{d}{=} \mathcal{U}\left(\mathbb{A}_{\text{BSK}}\right), \quad \operatorname{Re}\beta_q[n] \overset{d}{=} \operatorname{Im}\beta_q[n] \overset{d}{=} \mathcal{N}\left(0,1\right), \tag{3.88}$$

where $\overset{d}{=}$ means equality in distribution and $\mathcal{U}\left(\mathbb{A}_{\text{BSK}}\right)$ denotes the uniform distribution over the binary alphabet $\mathbb{A}_{\text{BSK}} = \{\pm 1\}$. In fact, letting $\boldsymbol{\alpha}[n] = (\alpha_1[n], \alpha_2[n], \ldots, \alpha_P[n])^T$ and recalling that $\boldsymbol{\beta}[n] = (\beta_1[n], \beta_2[n], \ldots, \beta_Q[n])^T$, define

$$\boldsymbol{x}[n] = \begin{bmatrix} \boldsymbol{\alpha}[n] \\ \boldsymbol{\beta}[n] \end{bmatrix}.$$

It is easily seen that $\boldsymbol{x}[n]$ is a random process belonging to the class $\mathcal{C}_{\mathbb{Z}}(\boldsymbol{\kappa})$, where

$$\boldsymbol{\kappa} = (\underbrace{1,\ldots,1}_{P},\underbrace{3,\ldots,3}_{Q})^T.$$

From (3.87), it follows that

$$\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[m] = \boldsymbol{\mathcal{G}}\widehat{\boldsymbol{R}}_{\boldsymbol{\gamma}}^{N}[m]\boldsymbol{\mathcal{G}}^H, \tag{3.89}$$

where

$$\widehat{\boldsymbol{R}}_{\boldsymbol{\gamma}}^{N}[m] = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{\gamma}[n]\boldsymbol{\gamma}[n-m]^H. \tag{3.90}$$

Now, each vector in the set $\{\boldsymbol{\gamma}[n],\boldsymbol{\gamma}[n-1],\ldots,\boldsymbol{\gamma}[n-M]\}$ is a sub-vector of $\boldsymbol{x}[n;\boldsymbol{l}]$, where

$$\boldsymbol{l} = (\underbrace{E_0+M,\ldots,E_0+M}_{P},\underbrace{M,\ldots,M}_{Q})^T.$$

More precisely, we have the relationship

$$\boldsymbol{\gamma}[n-m] = \boldsymbol{I}[m]\,\boldsymbol{x}[n;\boldsymbol{l}], \tag{3.91}$$

where $\boldsymbol{I}[m]$ is a selection matrix given by

$$\boldsymbol{I}[m] = \begin{bmatrix} \boldsymbol{I}_{\alpha}[m] & \\ & \boldsymbol{I}_{\beta}[m] \end{bmatrix},$$

with

$$\begin{aligned} \boldsymbol{I}_{\alpha}[m] &= \boldsymbol{I}_P \otimes \begin{bmatrix} \boldsymbol{0}_{(E_0+1)\times m} & \boldsymbol{I}_{E_0+1} & \boldsymbol{0}_{(E_0+1)\times(M-m)} \end{bmatrix} \\ \boldsymbol{I}_{\beta}[m] &= \boldsymbol{I}_Q \otimes \begin{bmatrix} \boldsymbol{0}_{1\times m} & 1 & \boldsymbol{0}_{1\times(M-m)} \end{bmatrix}, \end{aligned}$$

for $m = 0,1,\ldots,M$. Thus, using (3.91) in (3.90), and the result in (3.89) yields

$$\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[m] = (\boldsymbol{\mathcal{G}}\boldsymbol{I}[0])\,\boldsymbol{R}^N\,(\boldsymbol{\mathcal{G}}\boldsymbol{I}[m])^H, \tag{3.92}$$

where

$$\boldsymbol{R}^N = \frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}[n;\boldsymbol{l}]\boldsymbol{x}[n;\boldsymbol{l}]^H. \tag{3.93}$$

This means that

$$\left(\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[0],\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[1],\ldots,\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[M]\right) = F\left(\boldsymbol{R}^N\right), \tag{3.94}$$

where $F : \mathbb{C}^{[PE_0+(P+Q)(M+1)]\times[PE_0+(P+Q)(M+1)]} \rightarrow \mathbb{C}^{Q\times Q}\times\cdots\times\mathbb{C}^{Q\times Q}$ denotes the linear mapping given by

$$F(\boldsymbol{Z}) = (\boldsymbol{A}_0\boldsymbol{Z}\boldsymbol{B}_0,\boldsymbol{A}_1\boldsymbol{Z}\boldsymbol{B}_1,\ldots,\boldsymbol{A}_M\boldsymbol{Z}\boldsymbol{B}_M),$$

where $\boldsymbol{A}_m = \boldsymbol{\mathcal{G}}\boldsymbol{I}[0]$ and $\boldsymbol{B}_m = (\boldsymbol{\mathcal{G}}\boldsymbol{I}[m])^H$ for $m = 0,1,\ldots,M$. Applying lemma 3.9 to the sequence (indexed by N) of random matrices in (3.93) yields

$$\boldsymbol{R}^N \sim \sqrt{N} - \mathcal{AN}\left(2\boldsymbol{I}_{PE_0+(P+Q)(M+1)},\boldsymbol{C}(\boldsymbol{\kappa};\boldsymbol{l})\right).$$

Finally, due to (3.94), we may invoke the delta-method (see page 61) and conclude that

$$\left(\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[0], \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[1], \ldots, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[M]\right) \sim \sqrt{N} - \mathcal{AN}\left(\left(\boldsymbol{R}_{\boldsymbol{y}}[0], \boldsymbol{R}_{\boldsymbol{y}}[1], \ldots, \boldsymbol{R}_{\boldsymbol{y}}[M]\right), \boldsymbol{\Sigma}\right), \tag{3.95}$$

where

$$\boldsymbol{\Sigma} = \boldsymbol{D}\, \boldsymbol{C}(\boldsymbol{\kappa}; \boldsymbol{l})\, \boldsymbol{D}^T$$

with

$$\boldsymbol{D} = \begin{bmatrix} \jmath\left(\boldsymbol{B}_0^T \otimes \boldsymbol{A}_0\right) \\ \jmath\left(\boldsymbol{B}_1^T \otimes \boldsymbol{A}_1\right) \\ \vdots \\ \jmath\left(\boldsymbol{B}_M^T \otimes \boldsymbol{A}_M\right) \end{bmatrix}.$$

The fact that the asymptotic mean is $(\boldsymbol{R}_{\boldsymbol{y}}[0], \boldsymbol{R}_{\boldsymbol{y}}[1], \ldots, \boldsymbol{R}_{\boldsymbol{y}}[M])$ may be established by computing $F\left(2\boldsymbol{I}_{PE_0+(P+Q)(M+1)}\right)$ or, more easily, by noting that

$$\mathrm{E}\left\{\left(\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[0], \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[1], \ldots, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[M]\right)\right\} = \left(\boldsymbol{R}_{\boldsymbol{y}}[0], \boldsymbol{R}_{\boldsymbol{y}}[1], \ldots, \boldsymbol{R}_{\boldsymbol{y}}[M]\right),$$

irrespective of N, as it is clear from (3.85).

3.4.4 Asymptotic normality of $[\widehat{\boldsymbol{H}^N(z)}]$

By using the results of subsection 3.4.3, we have

$$\left(\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[0], \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[1], \ldots, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[M]\right) \sim \sqrt{N} - \mathcal{AN}\left(\left(\boldsymbol{R}_{\boldsymbol{y}}[0], \boldsymbol{R}_{\boldsymbol{y}}[1], \ldots, \boldsymbol{R}_{\boldsymbol{y}}[M]\right), \boldsymbol{\Sigma}\right),$$

see (3.95). The $M+1$-tuple $\left(\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[0], \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[1], \ldots, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[M]\right)$ is the input of the smooth map $\mathrm{CFIA} : \mathcal{U} \to \mathbb{H}[z]/\sim$, where

$$\mathrm{CFIA} = \pi \circ \psi,$$

recall figure 3.2. Thus, by invoking lemma 3.4, we conclude that the random sequence (indexed by N)

$$\left[\widehat{\boldsymbol{H}^N(z)}\right] = \mathrm{CFIA}\left(\widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[0], \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[1], \ldots, \widehat{\boldsymbol{R}}_{\boldsymbol{y}}^{N}[M]\right)$$

is asymptoticaly normal in the quotient space $\mathbb{H}[z]/\sim$. That is,

$$\left[\widehat{\boldsymbol{H}^N(z)}\right] \sim \sqrt{N} - \mathcal{AN}\left(\left[\boldsymbol{H}(z)\right], \Upsilon\right),$$

where $\Upsilon = \pi_*\left(\boldsymbol{\Lambda}\right)$ and

$$\boldsymbol{\Lambda} = D\psi\left(\left(\boldsymbol{R}_{\boldsymbol{y}}[0], \boldsymbol{R}_{\boldsymbol{y}}[1], \ldots, \boldsymbol{R}_{\boldsymbol{y}}[M]\right)\right) \boldsymbol{\Sigma} D\psi\left(\left(\boldsymbol{R}_{\boldsymbol{y}}[0], \boldsymbol{R}_{\boldsymbol{y}}[1], \ldots, \boldsymbol{R}_{\boldsymbol{y}}[M]\right)\right)^T.$$

Thus, we have the approximation

$$\mathrm{E}\left\{d\left(\left[\boldsymbol{H}(z)\right], \left[\widehat{\boldsymbol{H}^N(z)}\right]\right)^2\right\} \simeq \frac{\mathrm{tr}\,\Upsilon}{N}, \tag{3.96}$$

recall lemma 3.3 and (3.31). It remains to determine $\mathrm{tr}\,\Upsilon$. For this, we exploit lemma 3.10.

Lemma 3.10. *Let $\varrho : M \to N$ denote a Riemannian submersion. Let $\Sigma \in T_2(T_pM)$ for some $p \in M$, and define $\Upsilon = \varrho_*(\Sigma) \in T_2\left(T_{\varrho(p)}N\right)$. Then,*

$$\operatorname{tr}(\Upsilon) = \sum_{i=1}^{n} \Sigma(\omega_{ip}, \omega_{ip}),$$

where $\omega_{ip} = E_{ip}^{\flat} \in T_p^*M$ and $E_{1p}, \ldots, E_{np}$ denotes any orthonormal basis for the horizontal subspace $H_pM \subset T_pM$ with $n = \dim N$.

Proof: See appendix B.

To apply lemma 3.10, we proceed as follows. Let $\boldsymbol{H}_0 = \psi((\boldsymbol{R}_{\boldsymbol{y}}[0], \boldsymbol{R}_{\boldsymbol{y}}[1], \ldots, \boldsymbol{R}_{\boldsymbol{y}}[M]))$. Note that $\boldsymbol{H}_0$ depends on the choice of reference vectors $\boldsymbol{r}_i, \boldsymbol{s}_j$ in the implementation of the CFIA, see table 3.1. Let

$$\left\{ \left(\boldsymbol{\Delta}_1^{(q)}, \boldsymbol{\Delta}_2^{(q)}, \ldots, \boldsymbol{\Delta}_P^{(q)} \right) : q = 1, 2, \ldots, 2PQ(D_0+1) - P \right\}$$

designate an orthonormal basis for the horizontal subspace $H_{\boldsymbol{H}_0}\mathbb{H}[z]$, see (3.48). Let

$$\boldsymbol{\Delta}^{(q)} = \left[\begin{array}{cccc} \boldsymbol{\Delta}_1^{(q)} & \boldsymbol{\Delta}_2^{(q)} & \cdots & \boldsymbol{\Delta}_P^{(q)} \end{array} \right]$$

and define the matrix

$$\boldsymbol{\Delta} = \left[\begin{array}{cccc} \imath\left(\boldsymbol{\Delta}^{(1)}\right) & \imath\left(\boldsymbol{\Delta}^{(2)}\right) & \cdots & \imath\left(\boldsymbol{\Delta}^{(2PQ(D_0+1)-P)}\right) \end{array} \right].$$

Then, using the definition for the inner product in $T_{\boldsymbol{H}_0}\mathbb{H}[z]$, see (3.35), it is easily seen that

$$\operatorname{tr}\Upsilon = \operatorname{tr}\pi_*(\boldsymbol{\Lambda}) = \operatorname{tr}\left(\boldsymbol{\Delta}^T \boldsymbol{\Lambda} \boldsymbol{\Delta}\right).$$

3.5 Computer simulations

We conducted some numerical experiments to validate the theoretical analysis developed throughout this chapter. We considered a scenario with $P = 2$ users. Each user employs the QPSK digital modulation format, that is, $a_p[n] \in \mathbb{A}_{\text{QPSK}} = \left\{ \pm\frac{1}{\sqrt{2}} \pm i\frac{1}{\sqrt{2}} \right\}$. The pth symbol information sequence is pre-filtered as $s_p[n] = c_p(z) \circledast a_p[n]$, where the pre-filters $c_p(z)$ satisfy $C_0 = \deg c_p(z) = 1$. That is, $c_p(z) = \kappa_p\left(1 - z_p z^{-1}\right)$ where κ_p is an unit-power normalizing constant and z_p denotes the zero of the pth filter. For this set of simulations, the correlative filters have the same zeros as in the two-users scenario of section 2.8 (page 39), that is, $z_1 = \frac{1}{4}e^{-i\pi/2}$ and $z_2 = \frac{1}{2}e^{i\pi/4}$. For simplicity, we considered a memoryless mixing MIMO channel ($D_0 = 0$), with $Q = 3$ outputs. The channel was randomly generated and kept fixed throughout the simulations. The MIMO channel obtained was

$$\boldsymbol{H} = \left[\begin{array}{cc} -1.1972 + 0.4322i & -1.0286 - 0.4584i \\ 0.4435 - 0.7287i & -0.5949 + 0.5841i \\ -0.5140 + 0.2351i & -0.4865 - 0.3461i \end{array} \right].$$

The reference vectors $\boldsymbol{r}_i, \boldsymbol{s}_j$ appearing in table 3.1 were also randomly generated (independently from the channel). The observation noise $\boldsymbol{w}[n]$ is taken to be zero-mean spatio-temporal white Gaussian noise with power σ^2. The signal-to-noise ratio (SNR) is defined

as

$$\mathrm{SNR} = \frac{\mathrm{E}\left\{\|\boldsymbol{H}\boldsymbol{s}[n]\|^2\right\}}{\mathrm{E}\left\{\|\boldsymbol{w}[n]\|^2\right\}} = \frac{\|\boldsymbol{H}\|^2}{Q\sigma^2},$$

and was kept fixed at SNR = 15 dB. The goal of our simulations is to compare both sides of the approximation in (3.96), for several values of N, the number of available MIMO output observations. To accomplish this, the number of samples N was varied between $N_{\min} = 100$ and $N_{\max} = 500$ in steps of $N_{\text{step}} = 50$ samples. For each N, $K = 1000$ independent Monte-Carlo runs were simulated. For the kth run, the CFIA produced the estimate $\left[\widehat{\boldsymbol{H}^N(z)}\right]$ in the quotient space $\mathbb{H}[z]/\sim$, and the intrinsic (Riemannian) squared distance

$$d\left([\boldsymbol{H}(z)], \left[\widehat{\boldsymbol{H}^N(z)}\right]\right)^2$$

was recorded. We recall that the distance between two points in the quotient space was obtained in (3.50). The average of these squared-distances, over the K Monte-Carlos, constitute the estimate for the left-hand side of (3.96), for a given N. The right-hand side of (3.96) was computed using the theoretical expressions obtained throught the asymptotic analysis. Figure 3.10 shows the results thus obtained. The dashed and solid lines refer to

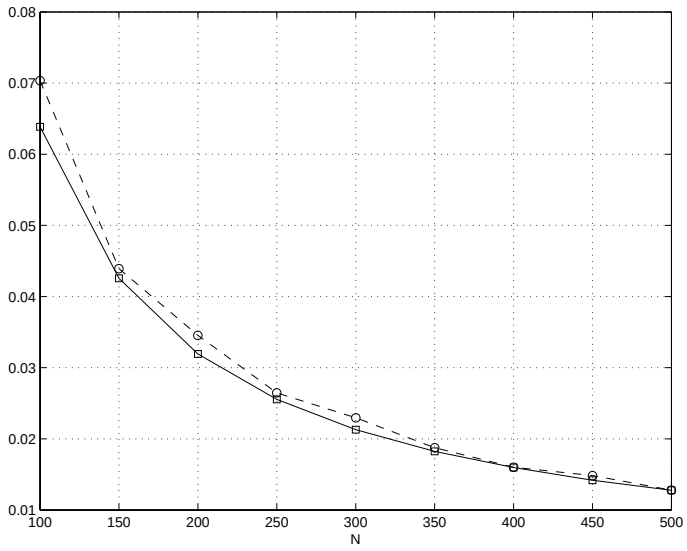

Figure 3.10: Mean squared (Riemannian) distance of the channel class estimate: theoretical (solid) and observed (dashed) (SNR = 15 dB)

the left and right hand sides of (3.96), respectively. As can be observed, the two curves show a good agreement even for modest values of the data packet length N, say $N \simeq 200$.

We performed a similar set of simulations, but now we decreased the signal-to-noise ratio to SNR = 5 dB. The results relative to this new scenario are presented in figure 3.11. We see that the quality of the channel class estimate degrades relative to the previous situation. This is expected and it is due to the loss of SNR. However, more importantly, the theoretically predicted and experimentally observed averaged squared distance of the channel class estimated continue to match each other.

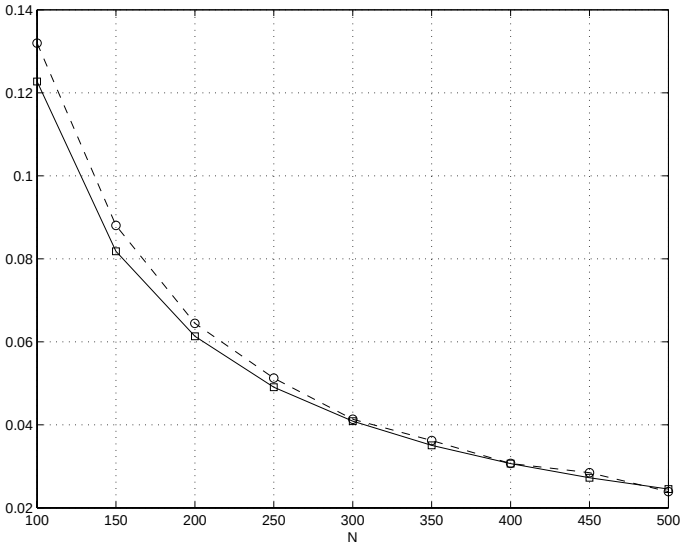

Figure 3.11: Mean squared (Riemannian) distance of the channel class estimate: theoretical (solid) and observed (dashed) (SNR = 5 dB)

3.6 Conclusions

In this chapter, we developed a performance analysis study of the closed-form identification algorithm (CFIA) proposed in chapter 2. More precisely, we obtained a closed-form expression which approximates

$$J\left[N\,;\,\boldsymbol{c}(z),[\boldsymbol{H}(z)]\right]=\mathrm{E}\left\{d\left(\left[\boldsymbol{H}(z)\right],\left[\widehat{\boldsymbol{H}^{N}(z)}\right]\right)^{2}\right\},\tag{3.97}$$

the mean-square distance between the true and the CFIA's channel estimate equivalence classes, given N channel observations and that $\boldsymbol{c}(z)=(c_1(z),c_2(z),\ldots,c_P(z))^T$ is the P-tuple of pre-filters placed at the transmitters. This study can be exploited to attain an optimal design for the pre-filter bank $\boldsymbol{c}(z)$, given a statistical model for the space-time channel $\boldsymbol{H}(z)$.

The theoretical study developed in this chapter was carried out in the setting of Riemannian manifolds. Indeed, we started by inducing a Riemannian geometry on the quotient space $\mathbb{H}[z]/\sim$. We made our choices for the geometry of the quotient in order to have it interfacing nicely with the geometry of $\mathbb{H}[z]$ (we required the canonical projection $\pi:\mathbb{H}[z]\rightarrow\mathbb{H}[z]/\sim$ to be a Riemannian submersion). Note that, previously, the quotient space was only an algebraic object. By endowing it with a geometric structure, we made available, in particular, the concept of distance. This is the distance function d appearing in (3.97) and measuring the quality of the channel estimate. We proceeded by re-interpreting the CFIA as a smooth map which sends a $M+1$-tuple of matrices (the estimated channel output correlation matrices) to a point in the smooth manifold $\mathbb{H}[z]/\sim$ (the estimated MIMO channel equivalence class). We proved that such estimate was asymptotically normal in the quotient manifold as $N\rightarrow\infty$, in a sense that generalizes the usual definition of asymptotic normality in Euclidean spaces. To prove this, we relied on a generalization of the well-known delta-method to this manifold setting. Fi-

nally, from the characterization of the asymptotic distribution of the CFIA's estimate, the mean-square distance in (3.97) is obtained in a straightforward manner. At this point, we would like to mention that, of course, a Riemannian structure is not strictly needed for introducing a distance function in the set $\mathbb{H}[z]/\sim$. However, by proceeding was we proposed, the map $\pi : \mathbb{H}[z] \to \mathbb{H}[z]/\sim$ is both smooth and a Riemannian submersion. It is this happy intersection of events which enables all the subsequent theoretical analysis. By imposing another, say ad-hoc, distance function on the quotient space, it would be more difficult to theoretically characterize and predict the asymptotic mean-square distance of the CFIA's estimate. The performance study was validated against computer simulations. The numerical experiments have shown a good agreement between the theoretical and the observed asymptotic behavior of the CFIA's estimate in the quotient space.

Chapter 4

Performance Bounds

4.1 Chapter summary

In chapter 2, we proposed a closed-form solution to the blind multiple-input multiple-output (MIMO) channel identification problem. Our analytical solution processes the observations at the channel's output through their 2nd order statistics (SOS) and not the observed data directly. That is, a kind of data-reduction occurs before the actual channel identification starts. Our channel estimator was structured like this because, as it was proved in chapter 2, the correlation matrices of the channel outputs contain sufficient information to reveal the underlying MIMO channel equivalence class (under a certain spectral diversity at the channel inputs). Furthermore, the identification method can be put in a non-iterative form, thus avoiding time-consuming re-initializations of typical iterative approaches. This proposed identification scheme is practical when enough observations are available to reliably estimate the output correlation matrices. However, for the purpose of identifying the MIMO channel equivalence class, there is no reason at all to restrict ourselves to the SOS of the channel output. For example, when only a scarce number of observations are available, all channel identification strategies based only on 2nd order statistics (or beyond, that is, higher-order statistics) are expected to experience a significant drop in performance, since the output correlation matrices' estimates will exhibit high variance about their nominal value (the true output correlation matrices). This leads us to the following question. What is the fundamental limit in estimating the channel equivalence class given a number N of observations ? By fundamental, we mean "estimator-independent" (based on SOS or not, etc). Note that our estimation problem is not standard in the sense that the parameter space is now $\mathbb{H}[z]/\sim$, that is, a connected Riemannian manifold. Usually, the parameter space is an open subset of a Euclidean space and, for these cases, we can resort to well-known results such as the Cramér-Rao lower bound (CRLB) to place a limit on the accuracy of any estimator for a given parametric statistical model. In this chapter, motivated by our inference problem on $\mathbb{H}[z]/\sim$, we develop a lower-bound on the intrinsic variance of estimators (with arbitrary bias) which take values in connected Riemannian manifolds. Thus, our results are presented in all generality. The solution to the inference problem concerning the quotient space $\mathbb{H}[z]/\sim$ will be obtained as a special case. This chapter is organized as follows. In section 4.2, we start by reviewing extensions of the CRLB inequality to the context of parametric statistical models indexed over smooth manifolds. We then define, in a differential-geometric

language, familiar statistical concepts such as bias, variance, etc. This sets the stage for stating precisely the bounding problem addressed in this chapter. In section 4.3, we give a solution to the stated problem. We present the intrinsic-variance lower bound (IVLB). The IVLB places a limit on the intrinsic accuracy of estimators taking values in Riemannian manifolds, in the context of parametric statistical models also indexed by Riemannian manifolds. The accuracy is measured with respect to the intrinsic Riemannian distance carried by the Riemannian manifolds. The derived bound depends on the curvature of the Riemannian manifold where the estimators take values and on a coordinate-free extension of the familiar Fisher information matrix. In section 4.4, we assess the tightness of the IVLB in some inference problems involving curved manifolds. We discuss inference problems on the unit-sphere $\mathbb{S}^{n-1}$, on the complex projective space $\mathbb{CP}^n$ and on the quotient space $\mathbb{H}[z]/\sim$, in subsections 4.4.1, 4.4.2 and 4.4.3, respectively. Note that in the last two problems we are dealing with coset spaces, that is, manifolds not immersed in Euclidean spaces (at least, not directly). Section 4.5 concludes this chapter and provides directions for future research.

4.2 Problem formulation

CRLB inequality. Consider a parametric family $\mathcal{F}$ of positive probability density functions over $\Omega = \mathbb{R}^n$ with parameter $\boldsymbol{p}$ taking values on an open subset P of some Euclidean space. That is, $\mathcal{F} = \{f_{\boldsymbol{p}} : \boldsymbol{p} \in P\}$, where $f_{\boldsymbol{p}} : \Omega \to \mathbb{R}$, $f_{\boldsymbol{p}}(\boldsymbol{\omega}) > 0$ for all $\boldsymbol{\omega} \in \Omega$, and $\int_\Omega f_{\boldsymbol{p}}(\boldsymbol{\omega})\, d\nu = 1$, for all $\boldsymbol{p} \in P$. The symbol ν denotes the Lebesgue measure on $\mathbb{R}^n$. We are interested in getting a lower-bound for the variance of a given unbiased estimator $\boldsymbol{\vartheta} : \Omega \to M$ of some map $\boldsymbol{b} : P \to M$. In equivalent terms, the estimator $\boldsymbol{\vartheta}$ satisfies $\mathrm{E}_{\boldsymbol{p}}\{\boldsymbol{\vartheta}\} = \boldsymbol{b}(\boldsymbol{p})$, for all $\boldsymbol{p} \in P$, that is, the map $\boldsymbol{b}$ denotes the bias of $\boldsymbol{\vartheta}$, and we want a lower bound for $\mathrm{var}_{\boldsymbol{p}}(\boldsymbol{\vartheta}) = \mathrm{E}_{\boldsymbol{p}}\left\{d\left(\boldsymbol{\vartheta}, \boldsymbol{b}(\boldsymbol{p})\right)^2\right\}$. Here, $M \subset \mathbb{R}^m$ denotes an open subset and, for $\boldsymbol{x}, \boldsymbol{y} \in M$,

$$d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\| \tag{4.1}$$

denotes the usual Euclidean distance. Several statistical signal processing problems can be cast in this canonical format, see [63, 51, 34]. Under suitable regularity conditions on $\mathcal{F}$, the Cramér-Rao lower bound (CRLB) [63, 51, 34, 48] provides the inequality

$$\mathrm{Cov}_{\boldsymbol{p}}\left(\boldsymbol{\vartheta}\right) \succeq D\boldsymbol{b}(\boldsymbol{p})\, \boldsymbol{I}_{\boldsymbol{p}}^{-1}\, D\boldsymbol{b}(\boldsymbol{p})^T, \tag{4.2}$$

where

$$\mathrm{Cov}_{\boldsymbol{p}}\left(\boldsymbol{\vartheta}\right) = \mathrm{E}_{\boldsymbol{p}}\left\{\left(\boldsymbol{\vartheta} - \boldsymbol{b}(\boldsymbol{p})\right)\left(\boldsymbol{\vartheta} - \boldsymbol{b}(\boldsymbol{p})\right)^T\right\}$$

denotes the covariance matrix of $\boldsymbol{\vartheta}$ with respect to $f_{\boldsymbol{p}}$, and the symbol

$$\boldsymbol{I}_{\boldsymbol{p}} = \mathrm{E}_{\boldsymbol{p}}\left\{\left(\boldsymbol{\nabla} \log\, f_{\boldsymbol{p}}\left(\boldsymbol{\omega}\right)\right)\, \left(\boldsymbol{\nabla} \log\, f_{\boldsymbol{p}}\left(\boldsymbol{\omega}\right)\right)^T\right\}$$

stands for the Fisher information matrix. For a differentiable function $f : \mathbb{R}^n \to \mathbb{R}$, the notation

$$\boldsymbol{\nabla} f\left(\boldsymbol{x}\right) = \left(\frac{\partial}{\partial x_1} f\left(\boldsymbol{x}\right), \frac{\partial}{\partial x_2} f\left(\boldsymbol{x}\right), \ldots, \frac{\partial}{\partial x_1} f\left(\boldsymbol{x}\right)\right)^T$$

denotes the gradient of f at $\boldsymbol{x}$. Furthermore, in (4.2), $D\boldsymbol{b}(\boldsymbol{p})$ denotes the derivative of the map $\boldsymbol{b}$ at the point $\boldsymbol{p}$, and, for symmetric matrices $\boldsymbol{A}$ and $\boldsymbol{B}$, the notation $\boldsymbol{A} \succeq \boldsymbol{B}$ means that $\boldsymbol{A} - \boldsymbol{B}$ is positive semidefinite. Note that

$$\mathrm{Cov}_{\boldsymbol{p}}(\boldsymbol{\vartheta}) = \int_{\Omega} (\boldsymbol{\vartheta}(\boldsymbol{\omega}) - \boldsymbol{b}(\boldsymbol{p})) \, (\boldsymbol{\vartheta}(\boldsymbol{\omega}) - \boldsymbol{b}(\boldsymbol{p}))^T \, f_{\boldsymbol{p}}(\boldsymbol{\omega}) \, d\nu.$$

Also,

$$\boldsymbol{I}_{\boldsymbol{p}} = \int_{\Omega} (\boldsymbol{\nabla} \log \, f_{\boldsymbol{p}}(\boldsymbol{\omega})) \, (\boldsymbol{\nabla} \log \, f_{\boldsymbol{p}}(\boldsymbol{\omega}))^T \, f_{\boldsymbol{p}}(\boldsymbol{\omega}) \, d\nu.$$

From (4.2), we can readily deduce the bound

$$\begin{aligned} \mathrm{var}_{\boldsymbol{p}}(\boldsymbol{\vartheta}) &= \mathrm{E}_{\boldsymbol{p}} \left\{ d(\boldsymbol{\vartheta}, \boldsymbol{b}(\boldsymbol{p}))^2 \right\} \\ &= \mathrm{tr}\left(\mathrm{Cov}_{\boldsymbol{p}}(\boldsymbol{\vartheta})\right) \\ &\geq \mathrm{tr}\left(D\boldsymbol{b}(\boldsymbol{p}) \, \boldsymbol{I}_{\boldsymbol{p}}^{-1} \, D\boldsymbol{b}(\boldsymbol{p})^T \right). \end{aligned} \tag{4.3}$$

Figure 4.1 illustrates the main points discussed so far.

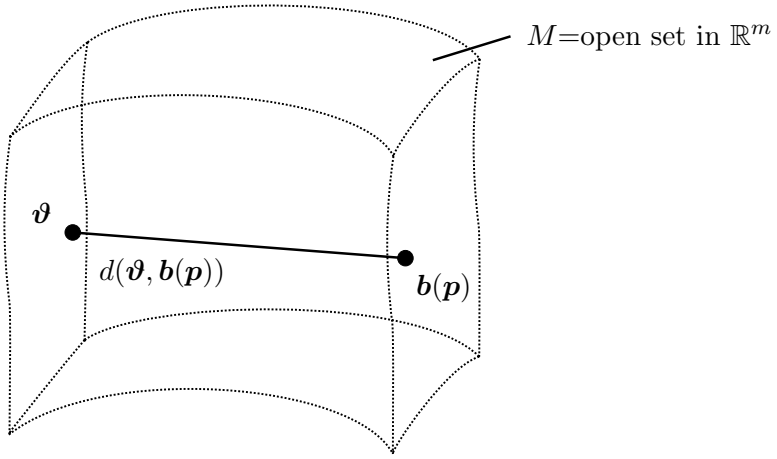

Figure 4.1: The CRLB places a limit to $\mathrm{var}_{\boldsymbol{p}}(\boldsymbol{\vartheta}) = \mathrm{E}_{\boldsymbol{p}} \left\{ d(\boldsymbol{\vartheta}, \boldsymbol{b}(\boldsymbol{p}))^2 \right\}$

CRLB Extensions. In the past few years, the classical CRLB inequalities (4.2) and (4.3) have been extended in several directions. The studies in [22, 41, 55] address unbiased estimation in scenarios where $M = P$. Thus, $\boldsymbol{b} : P \to P$ is given by $\boldsymbol{b}(\boldsymbol{p}) = \boldsymbol{p}$. The main novelty common to these works is that $M = P$ is no longer full-dimensional, that is, an open subset in $\mathbb{R}^m$, but instead contracts to a lower-dimensional immersed submanifold of $\mathbb{R}^m$, see figure 4.2 and compare with figure 4.1. This situation arises naturally in estimation problems where the parameter $\boldsymbol{p}$ to be estimated looses degrees of freedom due to *a priori* smooth deterministic constraints. For example, for physical reasons, the parameter $\boldsymbol{p}$ may be restricted to the surface of a unit-sphere (power constraint). Referring, for example, to the work in [55] the CRLB inequality (4.2) is now extended to

$$\mathrm{Cov}_{\boldsymbol{p}}(\boldsymbol{\vartheta}) \succeq \boldsymbol{U}_{\boldsymbol{p}} \left(\boldsymbol{U}_{\boldsymbol{p}}^T \, \boldsymbol{I}_{\boldsymbol{p}} \, \boldsymbol{U}_{\boldsymbol{p}} \right)^{-1} \boldsymbol{U}_{\boldsymbol{p}}^T, \tag{4.4}$$

where the columns of $\boldsymbol{U}_{\boldsymbol{p}}$ denote an orthonormal basis for $T_{\boldsymbol{p}}M$, the tangent space to M at the point $\boldsymbol{p}$, see figure 4.2. Thus, the full-rank Fisher information matrix $\boldsymbol{I}_{\boldsymbol{p}}$ is

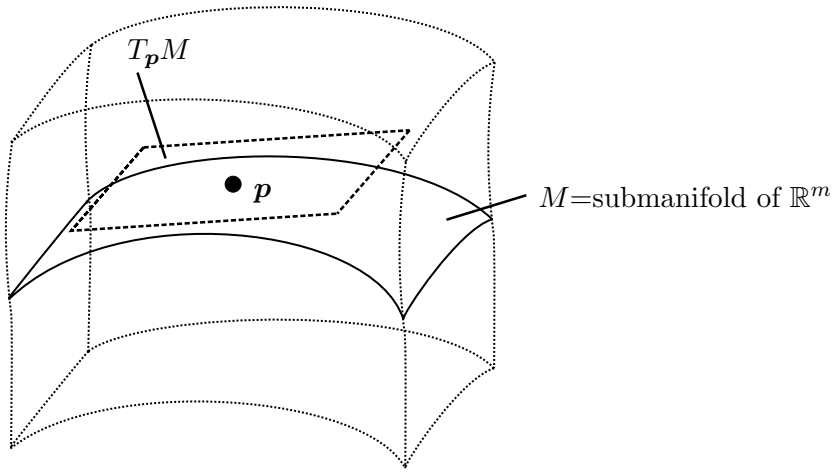

Figure 4.2: M contracts to a lower-dimensional submanifold of $\mathbb{R}^m$

now "compressed" (or projected) to the tangent plane of the constraint surface at $\boldsymbol{p}$. From (4.4), we have the bound

$$\begin{aligned} \mathrm{var}_{\boldsymbol{p}}(\boldsymbol{\vartheta}) &= \mathrm{E}_{\boldsymbol{p}}\left\{ d(\boldsymbol{\vartheta}, \boldsymbol{b}(\boldsymbol{p}))^2 \right\} \\ &= \mathrm{tr}\left(\mathrm{Cov}_{\boldsymbol{p}}(\boldsymbol{\vartheta})\right) \\ &\geq \mathrm{tr}\left(\boldsymbol{U}_{\boldsymbol{p}} \left(\boldsymbol{U}_{\boldsymbol{p}}^T \boldsymbol{I}_{\boldsymbol{p}} \boldsymbol{U}_{\boldsymbol{p}} \right)^{-1} \boldsymbol{U}_{\boldsymbol{p}}^T \right). \end{aligned} \tag{4.5}$$

The bound in (4.5) is still expressed in terms of the extrinsic Euclidean distance (4.1), not the intrinsic Riemannian distance, which measures the distance between points in M by taking only into account curve segments entirely contained in M (recall the definition of Riemannian distance in page 64). In figure 4.3, we sketch this discrepancy. Of course,

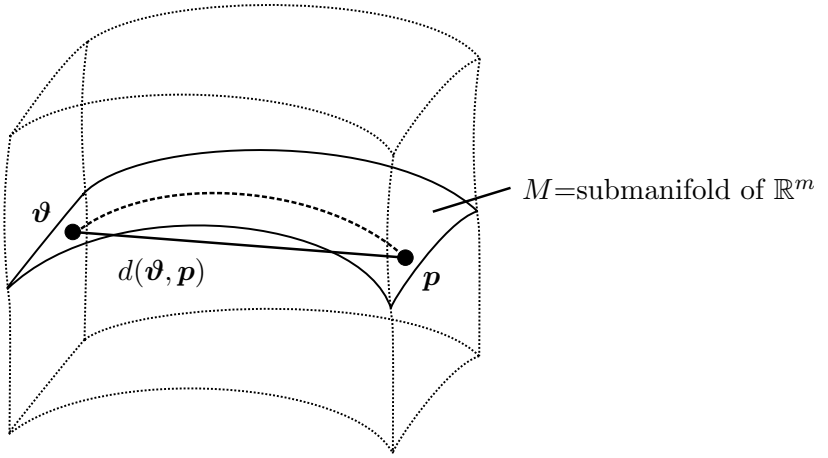

Figure 4.3: $d(\boldsymbol{\vartheta}, \boldsymbol{p})$ denotes the Euclidean distance not the Riemannian distance

since the Riemannian distance between two points in M (assumed to inherit its metric structure from the ambient space $\mathbb{R}^m$) is greater than or equal to the Euclidean distance, the lower bound in (4.5) still applies if d denotes the Riemannian distance. However, this bound clearly ignores the curvature of M and, as a consequence, might result too optimistic.

The studies in [27, 44, 54] avoid this discrepancy right from the start. They handle parametric statistical models, in the context of Riemannian manifolds, with intrinsic-only tools. That is, the ambient Euclidean spaces are simply ignored or even non-existent, and each manifold is treated as a geometrical object in its own right. Figure 4.4 illustrates the conceptual shift involved (the ambient space "disappears" from the analysis). The analysis

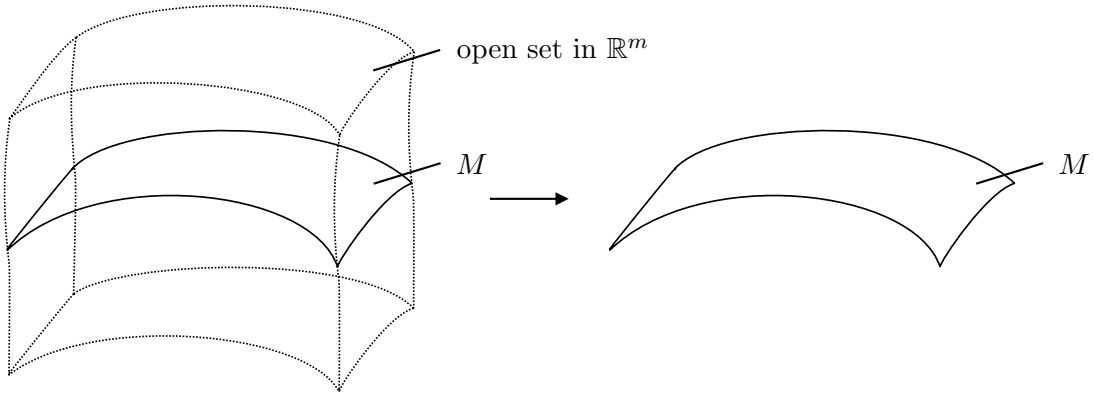

Figure 4.4: The Euclidean ambient space is discarded

in [27, 44, 54] is developed in the setting of Riemannian differential geometry. This level of abstraction automatically encompasses more general and interesting statistical families, like ones indexed over quotient spaces such as, for example, Grassmann manifolds [7, 32, 16]). Recall also that this is the context in blind MIMO channel identification, where we aim at estimating a point in the quotient space $\mathbb{H}[z]/\sim$. In [27], an intrinsic extension of the information inequality (4.2) is achieved in terms of tensor-like objects, but a lower bound on the variance of estimators is not readily available. The study in [44] provides such a bound. However, in [44], the Riemannian structure carried by M is not arbitrary, but the one induced by its Fisher information metric. The distance between points in M is thus the familiar information or Rao distance [33]. The work in [54] circumvents this difficulty: it is applicable to unbiased estimators and allows for arbitrary Riemannian metrics.

Problem Formulation. In this chapter, we adopt the differential-geometric viewpoint of the works in [27, 44, 54]. More specifically, we assume fixed a probability space $(\Omega, \mathcal{A}, \mu)$ and we are given a parametric statistical model $\mathcal{F} = \{f_p : \Omega \to \mathbb{R}\}$ with parameter p lying in a Riemannian manifold P. Our goal is to derive a lower bound for

$$\mathrm{var}_p(\vartheta) = E_p\left\{d(b(p), \vartheta)^2\right\}, \tag{4.6}$$

where $\vartheta : \Omega \to M$ denotes an estimator with mean value $\mathrm{E}_p\{\vartheta\} = b(p)$, where $b : P \to M$ is a smooth mapping, and M denotes a connected Riemannian manifold. The definition of mean-value of an estimator taking values in a Riemannian manifold is given in the next section, thus explaining in a precise manner what $\mathrm{E}_p\{\vartheta\} = b(p)$ means. In (4.6), $d : M \times M \to \mathbb{R}$ stands for the Riemannian distance function on M. Figure 4.5 illustrates the problem addressed in this chapter. Thus, the novel point here is: **i)** relative to [27], we derive a bound for the intrinsic variance $\mathrm{var}_p(\vartheta)$. This is not readily available from [27] which works in terms of tensor objects; **ii)** relative to [44], we allow $M \neq P$ and,

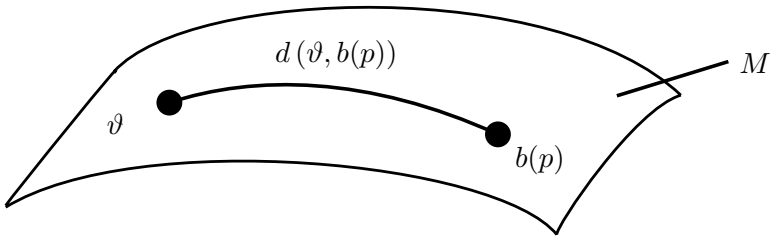

Figure 4.5: We aim at finding a tight lower bound for $\mathrm{var}_p(\vartheta) = \mathrm{E}_p\left\{d(\vartheta, b(p))^2\right\}$

furthermore, even if $M = P$ we allow M to be equipped with an arbitrary Riemannian structure. This is important for situations in which the Riemannian distance on M has a clear physical meaning whereas a real-world interpretation of the information metric is not so apparent (think of estimation problems where M stands for the Earth's surface equipped with our familiar concept of geodesic distance). In these cases, the intrinsic distance on M is usually the preferred figure of merit to evaluate the accuracy of estimators; **iii)** relative to [54], we allow for biased estimators.

4.3 Intrinsic variance lower bound (IVLB)

The reader is urged to recall the definitions and the differential-geometric concepts discussed in section 3.3. Here, we extend that framework in order to handle parametric statistical families in the setting of Riemannian manifolds. Let M denote a connected Riemannian manifold. For $m \in M$, we define the dispersion function about m as $k_m : M \to \mathbb{R}$,

$$k_m(n) = \frac{1}{2} d(m, n)^2,$$

where d stands for the Riemannian distance on M. We say that an open set $U \subset M$ is ϵ-uniformly normal if $U \subset B(m; \epsilon)$, for all $m \in U$. Recall that $B(m; \epsilon)$ denotes the geodesic ball centered at m with radius ϵ. See [7, theorem 6.9, page 340] for the existence of uniformly normal sets. Riemannian tools are used to formalize objects in statistical models indexed over Riemannian manifolds. Mostly, we keep the framework in [27, 44]. As mentioned before, a probability space $(\Omega, \mathcal{A}, \mu)$ is assumed fixed, where Ω stands for the sample space, $\mathcal{A}$ is a σ-algebra of subsets of Ω, and μ denotes the probability measure on $\mathcal{A}$. We are given a parametric family of positive densities $\mathcal{F} = \{f(\omega; p) : \omega \in \Omega\}$ with the parameter p taking values in a Riemannian manifold P. We define the log-likelihood function $l : \Omega \times P \to \mathbb{R}$ as $l(\omega, p) = \log f(\omega; p)$. We also write $f(\omega; p) = f_\omega(p)$, for $\omega \in \Omega$ and $p \in P$. Similarly, $l(\omega, p) = l_\omega(p)$. We assume $f_\omega \in \mathrm{C}^\infty(P)$ for all $\omega \in \Omega$. The Fisher information form, denoted I, is a section of $T_0^2(P)$. It is defined as

$$I(X_p, Y_p) = \mathrm{E}_p\left\{X_p l_\omega Y_p l_\omega\right\}, \tag{4.7}$$

where $X_p, Y_p \in T_pP$, see [27]. That is,

$$I(X_p, Y_p) = \int_\Omega (X_p l_\omega)(Y_p l_\omega) f(\omega; p)\, d\mu. \tag{4.8}$$

We define

$$\lambda_p = \max_{|X_p|=1} I(X_p, X_p). \tag{4.9}$$

In this context, an estimator ϑ is a random point in M, that is, a measurable mapping $\vartheta : \Omega \to M$. The estimator ϑ is said to have bias $b : P \to M$, if, for each $p \in P$, the function $\rho_p : M \to \mathbb{R}$, $\rho_p(n) = \mathrm{E}_p\{k_\vartheta(n)\}$, is globally minimized by $b(p)$. Notice that

$$\rho_p(n) = \int_\Omega k_{\vartheta(\omega)}(n)\, f(\omega; p)\, d\mu.$$

This notion of mean-value is also called the Riemannian center of mass [44]. The variance of an estimator $\vartheta : \Omega \to M$ with bias b is defined as

$$\mathrm{var}_p(\vartheta) = \mathrm{E}_p\left\{d(\vartheta, b(p))^2\right\}, \tag{4.10}$$

where d denotes the Riemannian distance on M.

The IVLB. We have now all the ingredients to state the main result of this chapter, which we call the intrinsic variance lower bound (IVLB). This is established in theorem 4.1.

Theorem 4.1. *Let the sectional curvature of M be bounded above by the constant $C \geq 0$. Let $\vartheta : \Omega \to M$ denote an estimator with bias $b : P \to M$. Assume that, for each $p \in P$, there exists $\epsilon > 0$, such that $\sqrt{C}\epsilon < T = \sqrt{3/2}$ and $\mathrm{Prob}_p\left\{\vartheta \in U_{b(p)}\right\} = 1$, where $U_{b(p)}$ is a $\epsilon-$uniformly normal neighborhood of $b(p) \in M$. Define*

$$\sigma_p = \max_{|X_p|=1} (b_*(X_p), b_*(X_p)). \tag{4.11}$$

Then, $\mathrm{var}_p(\vartheta) \geq 1/\eta_p$, if $C = 0$, and

$$\mathrm{var}_p(\vartheta) \geq \frac{4C + 3\eta_p - \sqrt{\eta_p(9\eta_p + 24C)}}{\frac{8}{3}C^2}, \tag{4.12}$$

if $C > 0$, where $\eta_p = \lambda_p/\sigma_p$ and λ_p is given by (4.9).

Proof: See appendix D.

The IVLB takes into account the geometry of M through the upper-bound C on its sectional curvatures. More precisely, it is required that $K(\Pi) \leq C$ for all two-dimensional planes $\Pi \subset T_pP$ and $p \in P$, where $K(\Pi)$ denotes the sectional curvature at p, recall (3.30). The Fisher information form and the estimator bias enter into the bound through

$$\eta_p = \frac{\lambda_p}{\sigma_p} = \frac{\max_{|X_p|=1} (b_*(X_p), b_*(X_p))}{\max_{|X_p|=1} I(X_p, X_p)}.$$

Finally, note that the theorem places an accuracy condition on the estimators ϑ, by requiring $d(\vartheta, b(p)) < T/\sqrt{C}$, where $T = \sqrt{3/2}$ is an universal (problem-independent) constant. Thus, the more curved is the manifold M, the more restricted is the class of estimators ϑ to which the IVLB applies.

Link with CRLB. It is possible to recover the IVLB as a weaker version of the classical Cramér-Rao inequality (4.3), by specializing P and M to open subsets of Euclidean spaces equipped with the usual metric. First notice that the inequality (4.3) imply

$$\begin{aligned} \mathrm{var}_p(\vartheta) &\geq \lambda_{\min}(\boldsymbol{I}_{\boldsymbol{p}}^{-1})\, \mathrm{tr}\left(D\boldsymbol{b}(\boldsymbol{p})\, D\boldsymbol{b}(\boldsymbol{p})^T\right) \\ &= \frac{1}{\lambda_{\max}(\boldsymbol{I}_{\boldsymbol{p}})}\, \mathrm{tr}\left(D\boldsymbol{b}(\boldsymbol{p})\, D\boldsymbol{b}(\boldsymbol{p})^T\right) \\ &\geq \frac{\lambda_{\max}\left(D\boldsymbol{b}(\boldsymbol{p})\, D\boldsymbol{b}(\boldsymbol{p})^T\right)}{\lambda_{\max}(\boldsymbol{I}_{\boldsymbol{p}})}. \end{aligned} \tag{4.13}$$

Here, for a symmetric matrix $\boldsymbol{A}$, we used the notation $\lambda_{\min}(\boldsymbol{A})$ (respectively, $\lambda_{\max}(\boldsymbol{A})$) to denote the minimum (respectively, maximum) eigenvalue of $\boldsymbol{A}$. Inequality (4.13) is precisely the IVLB inequality $\mathrm{var}_p(\vartheta) \geq 1/\eta_p$ for the case $C = 0$. Thus, for flat (Euclidean) spaces, the CRLB inequality (4.3) provides, in general, a better bound. The interest of the IVLB lies thus in the curved cases.

4.4 Examples

We examine three examples to assess the tightness of the IVLB. We study inference problems on the unit-sphere $\mathbb{S}^{n-1}$, on the complex projective space $\mathbb{CP}^n$ and on the space of identifiable MIMO channel equivalence classes $\mathbb{H}[z]/\sim$. Note that the last two examples require the formalism of Riemannian manifolds. The first example is dissected in all rigor in order to illustrate the main computations involved in applying the IVLB. The level of exposition for the remaining examples is not as stringent.

4.4.1 Inference on the unit-sphere $\mathbb{S}^{n-1}$

For illustrative purposes only, we consider a simple estimation problem over the unit-sphere in $\mathbb{R}^n$, denoted by

$$\mathbb{S}^{n-1} = \{\boldsymbol{x} \in \mathbb{R}^n \,:\, \|\boldsymbol{x}\| = 1\}\,.$$

The observation model is

$$\boldsymbol{Y} = \boldsymbol{Q} + \boldsymbol{W}, \tag{4.14}$$

where $\boldsymbol{Y} \in \mathbb{R}^{n\times n}$ denotes the observation, $\boldsymbol{Q} \in \mathbb{O}(n)$ denotes an (unknown) orthogonal matrix, and $\boldsymbol{W}$ stands for a random matrix whose entries are independent and identically distributed as zero-mean Gaussian random variables with variance σ^2. Thus, in this case, the sample space is $\Omega = \mathbb{R}^{n\times n}$, $\mathcal{A}$ denotes the σ-algebra of Lebesgue measurable sets and μ stands for Lebesgue measure. Furthermore, our statistical family is indexed, through the parameter $\boldsymbol{Q}$, over the Riemannian manifold $P = \mathbb{O}(n)$ of $n \times n$ real orthogonal matrices. In the sequel, we assume that $n \geq 3$. Consider the estimator $\boldsymbol{\vartheta} : \mathbb{R}^{n\times n} \to M = \mathbb{S}^{n-1}$,

$$\boldsymbol{\vartheta}(\boldsymbol{Y}) = \frac{\boldsymbol{y}}{\|\boldsymbol{y}\|},$$

where $\boldsymbol{y}$ denotes the first column of $\boldsymbol{Y}$. The map $\boldsymbol{\vartheta} : \mathbb{R}^{n\times n} \to \mathbb{S}^{n-1}$ can be interpreted as an estimator for the first column of the (unknown) orthogonal matrix $\boldsymbol{Q}$.

Bias. We start by establishing the bias of $\boldsymbol{\vartheta}$. We will show that the estimator $\boldsymbol{\vartheta}$ has bias given by the map $\boldsymbol{b} : \mathbb{O}(n) \to \mathbb{S}^{n-1}$, $\boldsymbol{b}(\boldsymbol{Q}) = \boldsymbol{q}$, where $\boldsymbol{q}$ denotes the first column of $\boldsymbol{Q}$. To establish this, we must (by definition) prove that $\boldsymbol{q}$ is a global minimum of the function $\rho_{\boldsymbol{Q}} : \mathbb{S}^{n-1} \to \mathbb{R}$,

$$\rho_{\boldsymbol{Q}}(\boldsymbol{u}) = \mathrm{E}_{\boldsymbol{Q}}\left\{d\left(\boldsymbol{\vartheta}(\boldsymbol{Y}), \boldsymbol{u}\right)^2\right\},$$

where $d(\boldsymbol{v}, \boldsymbol{u})$ denote the Riemannian distance between the points $\boldsymbol{v}, \boldsymbol{u} \in \mathbb{S}^{n-1}$. By letting $\mathbb{S}^{n-1}$ inherit the ambient Euclidean metric, we have $d(\boldsymbol{v}, \boldsymbol{u}) = \arccos(\boldsymbol{v}^{\mathrm{T}}\boldsymbol{u})$ because the geodesics of $\mathbb{S}^{n-1}$ are great circles [16, example 2.80.c, page 81]. See figure 4.6 for an illustration. In equivalent terms, we are letting the canonical embedding $\iota : \mathbb{S}^{n-1} \to \mathbb{R}^n$,

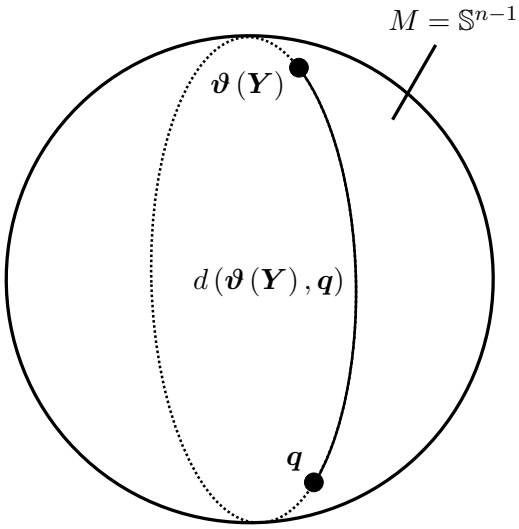

Figure 4.6: A great circle on the unit-sphere

$\iota(\boldsymbol{x}) = \boldsymbol{x}$, induce the Riemannian structure on $\mathbb{S}^{n-1}$. By analysing the proof of theorem 4.1 (see also [27]), it suffices, in fact, to prove that $\boldsymbol{q}$ is a stationary point of $\rho_{\boldsymbol{Q}}$. That is, we have to show that $X_{\boldsymbol{q}}\, \rho_{\boldsymbol{Q}} = 0$, for every tangent vector $X_{\boldsymbol{q}} \in T_{\boldsymbol{q}}\mathbb{S}^{n-1}$. By using the embedding ι to identify

$$T_{\boldsymbol{q}}\mathbb{S}^{n-1} \equiv \iota_* \left(T_{\boldsymbol{q}}\mathbb{S}^{n-1}\right) = \left\{\boldsymbol{d} \in \mathbb{R}^n \,:\, \boldsymbol{d}^T\boldsymbol{q} = 0\right\}, \tag{4.15}$$

this is equivalent to prove that $\boldsymbol{d}^T\, \nabla\rho_{\boldsymbol{Q}}(\boldsymbol{q}) = 0$, for every $\boldsymbol{d} \in T_{\boldsymbol{q}}\mathbb{S}^{n-1}$ and where $\nabla\rho_{\boldsymbol{Q}}(\boldsymbol{q})$ stands for the gradient of $\rho_{\boldsymbol{Q}}$ (viewed as a function on $\mathbb{R}^n$) evaluated at the point $\boldsymbol{q}$.

Since

$$\rho_{\boldsymbol{Q}}(\boldsymbol{u}) \propto \int_{\mathbb{R}^n} \arccos\left(\frac{\boldsymbol{x}^T\boldsymbol{u}}{\|\boldsymbol{x}\|}\right)^2 e^{-\frac{1}{2\sigma^2}\,\|\boldsymbol{x}-\boldsymbol{q}\|^2}\, d\boldsymbol{x},$$

where $\propto$ means equality up to a constant, we have

$$\nabla\rho_{\boldsymbol{Q}}(\boldsymbol{q}) \propto \int_{\mathbb{R}^n} h\left(\frac{\boldsymbol{x}^T\boldsymbol{q}}{\|\boldsymbol{x}\|}\right) \frac{\boldsymbol{x}}{\|\boldsymbol{x}\|}\, e^{-\frac{1}{2\sigma^2}\,\|\boldsymbol{x}-\boldsymbol{q}\|^2}\, d\boldsymbol{x}, \tag{4.16}$$

with

$$h(t) = \frac{\arccos(t)}{\sqrt{1-t^2}}.$$

Let $\boldsymbol{d} \in T_{\boldsymbol{q}}\mathbb{S}^{n-1}$. Using (4.16) and the change of variables $\boldsymbol{x} \mapsto (R, \boldsymbol{u}) \equiv (\|\boldsymbol{x}\|, \boldsymbol{x}/\|\boldsymbol{x}\|)$, it follows that

$$\boldsymbol{d}^T\nabla\rho_{\boldsymbol{Q}}(\boldsymbol{q}) \propto \int_0^{+\infty} R^{n-1}\, e^{-\frac{R^2+1}{2\sigma^2}}\, g(R)\, dR, \tag{4.17}$$

where

$$g(R) = \int_{\mathbb{S}^{n-1}} \boldsymbol{u}^T\boldsymbol{d}\, h(\boldsymbol{u}^T\boldsymbol{q})\, e^{\frac{R}{\sigma^2}\boldsymbol{u}^T\boldsymbol{q}}\, d\boldsymbol{u}. \tag{4.18}$$

Here, we used the "change-of-variables" formula which asserts that, for a smooth function $f : \mathbb{R}^n \to \mathbb{R}$, we have

$$\int_{\mathbb{R}^n} f(\boldsymbol{x})\, d\boldsymbol{x} = \int_0^{+\infty} \int_{\mathbb{S}^{n-1}} f(R\boldsymbol{u})\, R^{n-1}\, dR d\boldsymbol{u}.$$

This is a trivial application of the theory of integration based on differential forms [7, 16]. Moreover, basic properties of this theory are implicitly used in the forthcoming manipulations. Let $\boldsymbol{W} = [\,\boldsymbol{q}\,\boldsymbol{w}_2\,\cdots\,\boldsymbol{w}_n\,]$ denote a special orthogonal matrix, that is, $\boldsymbol{W}$ is orthogonal and $\det(\boldsymbol{W}) = 1$. Making the change of variables $\boldsymbol{u} \mapsto \boldsymbol{v} = (v_1, v_2, \ldots, v_n)^T = \boldsymbol{W}^T\boldsymbol{u}$ in (4.18), and recalling that $\boldsymbol{d}^T\boldsymbol{q} = 0$, leads to

$$g(R) = \int_{\mathbb{S}^{n-1}} h(v_1)\, e^{\frac{R}{\sigma^2}v_1} \sum_{i=2}^{n} a_i\, v_i\, d\boldsymbol{v},$$

where $a_i = \boldsymbol{d}^T\boldsymbol{w}_i$. Thus,

$$g(R) = \sum_{i=2}^{n} a_i\, g_i(R)$$

where

$$g_i(R) = \int_{\mathbb{S}^{n-1}} h(v_1)\, e^{\frac{R}{\sigma^2}v_1}\, v_i\, d\boldsymbol{v}. \tag{4.19}$$

Let $i \in \{2, \ldots, n\}$. Let $\boldsymbol{D}$ denote a diagonal matrix with diagonal entries equal to 1, except the (i,i)th and the (j,j)th entries which are -1. Here, $j \in \{2, \ldots, i-1, i+1, \ldots, n\}$ is arbitrarily chosen (note that this can always be done, because it was previously assumed that $n \geq 3$). This guarantees that $\det(\boldsymbol{D}) = 1$. Using the change of variables $\boldsymbol{w} \mapsto \boldsymbol{D}\boldsymbol{w}$ in (4.19), leads to $g_i(R) = -g_i(R)$, that is, $g_i(R) = 0$. Thus, $g(R) = 0$, and from (4.17) we have $\boldsymbol{d}^T\nabla\rho_{\boldsymbol{Q}}(\boldsymbol{q}) = 0$. We conclude that the estimator $\boldsymbol{\vartheta}$ has bias $\boldsymbol{b}(\boldsymbol{Q}) = \boldsymbol{q}$.

Bias derivative. The embedding $\iota : \mathbb{O}(n) \to \mathbb{R}^{n\times n}$, $\iota(\boldsymbol{X}) = \boldsymbol{X}$, provides the identification

$$T_{\boldsymbol{Q}}\mathbb{O}(n) \equiv \iota_*\left(T_{\boldsymbol{Q}}\mathbb{O}(n)\right) = \left\{\boldsymbol{Q}\boldsymbol{K} : \boldsymbol{K} = -\boldsymbol{K}^T\right\}. \tag{4.20}$$

With the identifications (4.20) and (4.15) for the tangent spaces of $\mathbb{O}(n)$ and $\mathbb{S}^{n-1}$, respectively, the push-forward map $\boldsymbol{b}_* : T_{\boldsymbol{Q}}\mathbb{O}(n) \to T_{\boldsymbol{q}}\mathbb{S}^{n-1}$ is given by $\boldsymbol{b}_*(\boldsymbol{Q}\boldsymbol{K}) = \boldsymbol{Q}\boldsymbol{k}$, where $\boldsymbol{k}$ denotes the first column of the $n \times n$ skew-symmetric matrix $\boldsymbol{K}$. We assume that $\mathbb{O}(n)$ inherits its metric structure from the ambient Euclidean space $\mathbb{R}^{n\times n}$. Let $\langle , \rangle$ denote the induced inner-product on the tangent space $T_{\boldsymbol{Q}}\mathbb{O}(n)$ and $|\cdot|$ stand for the respective norm. Notice that, with these choices, this norm coincides with the usual Frobenius norm on $\mathbb{R}^{n\times n}$ under the identification (4.20). We have

$$\begin{aligned}
\sigma_{\boldsymbol{Q}} &= \max_{\substack{\|\boldsymbol{Q}\boldsymbol{K}\| = 1 \\ \boldsymbol{K} = -\boldsymbol{K}^T}} \langle \boldsymbol{b}_*(\boldsymbol{Q}\boldsymbol{K}), \boldsymbol{b}_*(\boldsymbol{Q}\boldsymbol{K})\rangle \\
&= \max_{\substack{\|\boldsymbol{k}\| = 1 \\ \boldsymbol{K} = -\boldsymbol{K}^T}} \|\boldsymbol{k}\|^2 \\
&= 1.
\end{aligned}$$

Fisher-information form. To obtain the Fisher-information form associated with our parametric family, we start by noticing that

$$f(\boldsymbol{Y}; \boldsymbol{Q}) \propto e^{-\frac{1}{2\sigma^2}\|\boldsymbol{Y} - \boldsymbol{Q}\|^2}.$$

Thus,

$$l(\boldsymbol{Y},\boldsymbol{Q}) = \log f(\boldsymbol{Y},\boldsymbol{Q}) \propto -\frac{1}{2\sigma^2}\left\|\boldsymbol{Y}-\boldsymbol{Q}\right\|^2.$$

The directional derivative of $l_{\boldsymbol{Y}}(\cdot) = l(\boldsymbol{Y},\cdot)$ at $\boldsymbol{Q}$ in the tangent direction $\boldsymbol{QK}$ ($\boldsymbol{K} = -\boldsymbol{K}^T$), written $\nabla_{\boldsymbol{Q}} l(\boldsymbol{Y},\boldsymbol{Q};\boldsymbol{QK})$, is given by

$$\nabla_{\boldsymbol{Q}} l(\boldsymbol{Y},\boldsymbol{Q};\boldsymbol{QK}) = \frac{1}{\sigma^2}\mathrm{tr}\left(\boldsymbol{Y}^T\boldsymbol{QK}\right).$$

It follows that the Fisher-information form acts on tangent vectors $\boldsymbol{QK} \in T_{\boldsymbol{Q}}\mathbb{O}(n)$ at $\boldsymbol{Q}$ as

$$\begin{aligned} I_{\boldsymbol{Q}}(\boldsymbol{QK},\boldsymbol{QK}) &= \frac{1}{\sigma^4}\mathrm{E}\left\{\left(\mathrm{tr}\left(\boldsymbol{Y}^T\boldsymbol{QK}\right)\right)^2\right\} \\ &= \frac{1}{\sigma^4}\mathrm{E}\left\{\left(\mathrm{tr}\left(\boldsymbol{W}^T\boldsymbol{QK}\right)\right)^2\right\} \qquad (4.21) \\ &= \frac{1}{\sigma^4}\mathrm{E}\left\{\left(\mathrm{tr}\left(\boldsymbol{Z}^T\boldsymbol{K}\right)\right)^2\right\}. \qquad (4.22) \end{aligned}$$

In (4.21), we used the identities $\boldsymbol{Y} = \boldsymbol{Q}+\boldsymbol{W}$, $\boldsymbol{Q}^T\boldsymbol{Q} = \boldsymbol{I}_n$ and $\mathrm{tr}\left(\boldsymbol{K}\right) = 0$. In (4.22), $\boldsymbol{Z}$ denotes a $n \times n$ random matrix whose entries are independent and identically distributed as zero-mean Gaussian random variables with variance σ^2. Note that (4.22) holds because $\boldsymbol{Z}$ and $\boldsymbol{W}^T\boldsymbol{Q}$ are equal in distribution. Now, noting that $\mathrm{tr}\left(\boldsymbol{A}^T\boldsymbol{B}\right) = \mathrm{vec}(\boldsymbol{A})^T\mathrm{vec}(\boldsymbol{B})$, equation (4.22) reads as

$$\begin{aligned} I_{\boldsymbol{Q}}(\boldsymbol{QK},\boldsymbol{QK}) &= \frac{1}{\sigma^4}\mathrm{E}\left\{\mathrm{vec}(\boldsymbol{K})^T\,\mathrm{vec}(\boldsymbol{Z})\mathrm{vec}(\boldsymbol{Z})^T\,\mathrm{vec}(\boldsymbol{K})\right\} \\ &= \frac{1}{\sigma^4}\mathrm{E}\left\{\mathrm{vec}(\boldsymbol{K})^T\,\left(\sigma^2 I_{n^2}\right)\,\mathrm{vec}(\boldsymbol{K})\right\} \\ &= \frac{1}{\sigma^2}\left\|\boldsymbol{K}\right\|^2. \end{aligned}$$

Thus, we have

$$\begin{aligned} \lambda_{\boldsymbol{Q}} &= \max_{\substack{\|\boldsymbol{QK}\|=1 \\ \boldsymbol{K}=-\boldsymbol{K}^T}} I_{\boldsymbol{Q}}(\boldsymbol{QK},\boldsymbol{QK}) \\ &= \max_{\substack{\|\boldsymbol{K}\|=1 \\ \boldsymbol{K}=-\boldsymbol{K}^T}} \frac{1}{\sigma^2}\left\|\boldsymbol{K}\right\|^2 \\ &= \frac{1}{\sigma^2}. \end{aligned}$$

Computer simulations. It is well known that the sectional curvature of $\mathbb{S}^{n-1}$ is constant and equal to 1, see [36, page 148]. Thus, we can take the upper bound $C = 1$ in theorem 4.1. Inserting $C = 1$, $\sigma_{\boldsymbol{Q}} = 1$ and $\lambda_{\boldsymbol{Q}} = 1/\sigma^2$ in (4.12) yields

$$\mathrm{var}_{\boldsymbol{Q}}\left(\vartheta\right) \geq \frac{4 + 3/\sigma^2 - \sqrt{\frac{1}{\sigma^2}\left(9/\sigma^2+24\right)}}{\frac{8}{3}}. \qquad (4.23)$$

We performed computer simulations to compare both sides of the inequality (4.23). We considered the case $n = 3$. We randomly generated an orthogonal matrix $\boldsymbol{Q} \in \mathbb{O}(n)$. This matrix was kept fixed during all Monte-Carlo experiments. We ran experiments from $\mathrm{SNR}_{\min} = -5\,\mathrm{dB}$ to $\mathrm{SNR}_{\max} = 30\,\mathrm{dB}$, in steps of $\Delta = 5\,\mathrm{dB}$. Here, SNR stands for the signal-to-noise ratio in the data model (4.14), that is,

$$\mathrm{SNR} = \frac{\|\boldsymbol{Q}\|^2}{\mathrm{E}\left\{\|\boldsymbol{W}\|^2\right\}} = \frac{1}{n\sigma^2}.$$

For each SNR, we considered $L = 1000$ statistically independent experiments. For each SNR, the variance of ϑ was taken as the mean value of $d(\boldsymbol{q}, \vartheta(\boldsymbol{Y}_l))^2$, $l = 1, 2, \ldots, L$, where $\boldsymbol{q}$ denotes the first column of $\boldsymbol{Q}$ and $\boldsymbol{Y}_l$ stands for the lth realization of the observation $\boldsymbol{Y}$ in (4.14). Figure 4.7 plots the result of the experiments. The dashed and solid line refer

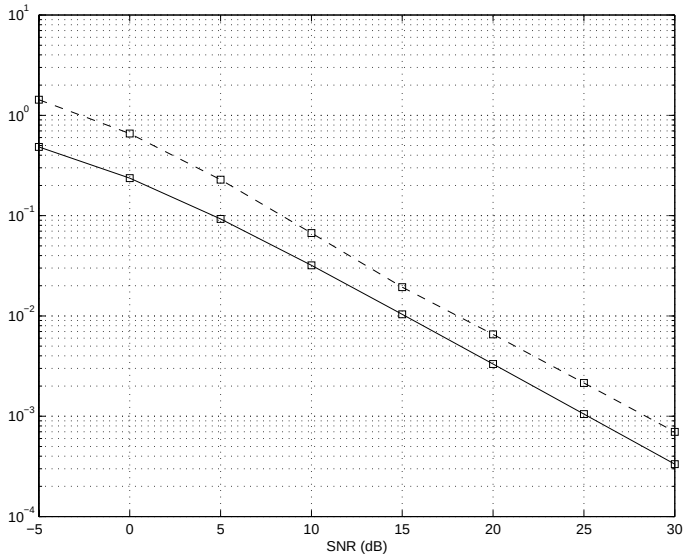

Figure 4.7: Estimated $\mathrm{var}_{\boldsymbol{Q}}(\vartheta)$ (dashed) and IVLB (solid)

to the estimated $\mathrm{var}_{\boldsymbol{Q}}(\vartheta)$ and the intrinsic variance lower bound in (4.23), respectively. We see that, at least for the example considered herein, the IVLB is reasonably tight on the whole range of simulated SNRs.

4.4.2 Inference on the complex projective space $\mathbb{CP}^n$

We consider a statistical model indexed over $P = \mathbb{CP}^n$, the complex projective space of dimension n [32, 16]. This is the set of 1-dimensional complex subspaces (lines) of $\mathbb{C}^{n+1}$. It is a real connected manifold with dimension $2n$. Hereafter, identify $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ using the embedding $\imath : \mathbb{C}^n \to \mathbb{R}^{2n}$,

$$\imath(\boldsymbol{z}) = \begin{bmatrix} \mathrm{Re}\,\boldsymbol{z} \\ \mathrm{Im}\,\boldsymbol{z} \end{bmatrix}, \tag{4.24}$$

recall (3.16). Also, recall that, for complex matrices, we have the identification $\imath : \mathbb{C}^{n\times m} \to \mathbb{R}^{2nm}$,

$$\imath(\boldsymbol{Z}) = \begin{bmatrix} \mathrm{Re}\,\mathrm{vec}\,(\boldsymbol{Z}) \\ \mathrm{Im}\,\mathrm{vec}\,(\boldsymbol{Z}) \end{bmatrix},$$

see (3.17). Using the identification (4.24), we can view the unit-sphere in $\mathbb{C}^n$,

$$\{\boldsymbol{u} \in \mathbb{C}^n \,:\, \|\boldsymbol{u}\| = 1\},$$

as the unit-sphere in $\mathbb{R}^{2n}$,

$$\mathbb{S}^{2n-1} = \left\{\boldsymbol{x} \in \mathbb{R}^{2n} \,:\, \|\boldsymbol{x}\| = 1\right\},$$

for all n.

The space $\mathbb{CP}^n$ can be realized as the orbit space of a Lie group action on the unit-sphere of $\mathbb{C}^{n+1}$. More precisely, we have $\mathbb{CP}^n = \mathbb{S}^{2n+1}/\mathbb{S}^1$ by considering the action $\varphi : \mathbb{S}^{2n+1} \times \mathbb{S}^1 \to \mathbb{S}^{2n+1}$, $\varphi(\boldsymbol{u}, c) = \boldsymbol{u}c$. Here, $\mathbb{S}^1 = \left\{e^{it} : t \in \mathbb{R}\right\} \subset \mathbb{C}$ is seen as a real $1-$dimensional Lie group, with complex multiplication as group operation. We let $\pi : \mathbb{S}^{2n+1} \to \mathbb{CP}^n$ denote the canonical submersion. See figure 4.8 for a sketch. Then,

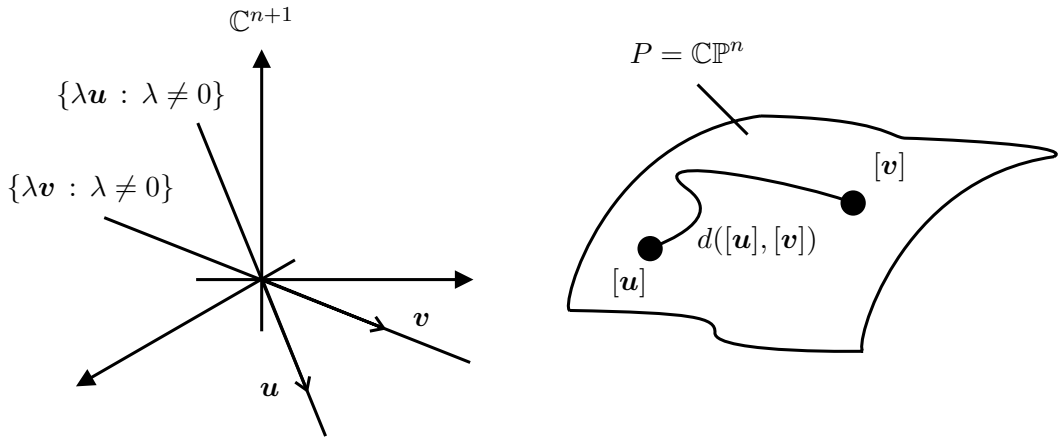

Figure 4.8: Complex projective space as a quotient space

$\mathbb{CP}^n$ has a natural Riemannian metric, that is, the only one making the projection π a Riemannian submersion [16]. With this induced metric, the intrinsic distance between two points $\pi(\boldsymbol{u}) = [\boldsymbol{u}]$ and $\pi(\boldsymbol{v}) = [\boldsymbol{v}]$ in $\mathbb{CP}^n$, with $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{S}^{2n+1} \subset \mathbb{C}^{n+1}$, is given by $d([\boldsymbol{u}], [\boldsymbol{v}]) = \arccos\left(\left|\boldsymbol{u}^H\boldsymbol{v}\right|\right)$, also called the Fubini-Study metric [36, 13], and the sectional curvature of $\mathbb{CP}^n$ obeys $1 \leq K \leq 4$ for $n \geq 2$, see [36, page 153]. Thus, we have the upper bound $C = 4$. We consider the parametric family

$$\mathcal{F} = \left\{f(\boldsymbol{X}; p) \,:\, \boldsymbol{X} \in \mathbb{C}^{(n+1)\times(n+1)}\right\}$$

corresponding to the observation data model

$$\boldsymbol{X} = \boldsymbol{u}\boldsymbol{u}^H + \boldsymbol{W}, \tag{4.25}$$

where $\boldsymbol{W}$ denotes a $(n+1) \times (n+1)$ random matrix where each entry is identically and independently distributed as a zero-mean complex circular Gaussian random variable with variance $\sigma^2 > 0$. Moreover, $\boldsymbol{u} \in \pi^{-1}(p) \subset \mathbb{C}^{n+1}$ denotes a representative (chosen in the unit-sphere) of the subspace $p \in \mathbb{CP}^n$, that is, $p = [\boldsymbol{u}]$. Notice that the particular choice of $\boldsymbol{u}$ in the fiber $\pi^{-1}(p)$ is irrelevant (as it should be). In words, given the data model (4.25), we are trying to estimate p, the complex line in $\mathbb{C}^{n+1}$ spanned by $\boldsymbol{u}$. Note that $\boldsymbol{u}$ itself is

not identifiable from $\boldsymbol{X}$ as $\boldsymbol{u}e^{i\phi}$ induces the same data matrix $\boldsymbol{X}$. Thus, in loose terms, the line spanned by $\boldsymbol{u}$ is the "true" parameter.

We now evaluate the Fisher information form associated with our parametric statistical family. Given the data model in (4.25), we have

$$\begin{aligned} \log p\left(\boldsymbol{X};[\boldsymbol{u}]\right) &\propto -\frac{1}{\sigma^2}\left\|\boldsymbol{X}-\boldsymbol{u}\boldsymbol{u}^H\right\|^2 \\ &\propto -\frac{2}{\sigma^2}\operatorname{Re}\left(\boldsymbol{u}^H\boldsymbol{X}\boldsymbol{u}\right), \end{aligned} \tag{4.26}$$

where $\propto$ stands for equality up to an additive constant. Thus, for any given tangent vector $X_{[\boldsymbol{u}]} \in T_{[\boldsymbol{u}]}\mathbb{CP}^n$, we have

$$\begin{aligned} I_{[\boldsymbol{u}]}\left(X_{[\boldsymbol{u}]},X_{[\boldsymbol{u}]}\right) &= \mathrm{E}\left\{X_{[\boldsymbol{u}]}\log p\left(\boldsymbol{X};\cdot\right)\, X_{[\boldsymbol{u}]}\log p\left(\boldsymbol{X};\cdot\right)\right\} \\ &= \frac{4}{\sigma^4}\mathrm{E}\left\{\left(X_{[\boldsymbol{u}]}f_{\boldsymbol{X}}\right)^2\right\}, \end{aligned} \tag{4.27}$$

where $f_{\boldsymbol{X}} : \mathbb{CP}^n \to \mathbb{R}$ is defined as $f_{\boldsymbol{X}}\left([\boldsymbol{u}]\right) = \operatorname{Re}\left(\boldsymbol{u}^H\boldsymbol{X}\boldsymbol{u}\right)$. Note that we have $g_{\boldsymbol{X}} = f_{\boldsymbol{X}}\circ\pi$, where $g_{\boldsymbol{X}} : \mathbb{S}^{2n+1} \to \mathbb{R}$ is given by $g_{\boldsymbol{X}}(\boldsymbol{v}) = \operatorname{Re}\left(\boldsymbol{v}^H\boldsymbol{X}\boldsymbol{v}\right)$. Since π is a Riemannian submersion we have the equality $X_{[\boldsymbol{u}]} = \pi_*\left(X_{\boldsymbol{u}}\right)$ for some horizontal vector $X_{\boldsymbol{u}} \in H_{\boldsymbol{u}}\mathbb{S}^{2n+1}$. For further reference, we notice that, within the identification $T_{\boldsymbol{u}}\mathbb{S}^{2n+1} \equiv \mathbb{C}^{n+1}$, we have

$$T_{\boldsymbol{u}}\mathbb{S}^{2n+1} = \left\{\boldsymbol{\delta} \in \mathbb{C}^{n+1} \,:\, \operatorname{Re}\boldsymbol{u}^H\boldsymbol{\delta} = 0\right\},$$

and

$$H_{\boldsymbol{u}}\mathbb{S}^{2n+1} = \left\{\boldsymbol{\delta} \in \mathbb{C}^{n+1} \,:\, \boldsymbol{u}^H\boldsymbol{\delta} = 0\right\}.$$

Now,

$$\begin{aligned} X_{[\boldsymbol{u}]}f_{\boldsymbol{X}} &= \pi_*\left(X_{\boldsymbol{u}}\right)f_{\boldsymbol{X}} \\ &= X_{\boldsymbol{u}}\left(\pi^* f_{\boldsymbol{X}}\right) \\ &= X_{\boldsymbol{u}}\left(f_{\boldsymbol{X}}\circ\pi\right) \\ &= X_{\boldsymbol{u}}\, g_{\boldsymbol{X}}. \end{aligned}$$

Writing $X_{\boldsymbol{u}} = \boldsymbol{\delta} \in \mathbb{C}^{2n+1}$ and noting that

$$g_{\boldsymbol{X}}\left(\boldsymbol{v}\right) = \imath\left(\overline{\boldsymbol{v}}\otimes\boldsymbol{v}\right)^T\imath\left(\boldsymbol{X}\right),$$

we have (after some trivial computations)

$$X_{\boldsymbol{u}}g_{\boldsymbol{X}} = \imath\left(\overline{\boldsymbol{\delta}}\otimes\boldsymbol{u}+\overline{\boldsymbol{u}}\otimes\boldsymbol{\delta}\right)^T\imath\left(\boldsymbol{X}\right). \tag{4.28}$$

Plugging (4.28) in (4.27) yields

$$I_{[\boldsymbol{u}]}\left(X_{[\boldsymbol{u}]},X_{[\boldsymbol{u}]}\right) = \frac{4}{\sigma^4}\imath\left(\overline{\boldsymbol{\delta}}\otimes\boldsymbol{u}+\overline{\boldsymbol{u}}\otimes\boldsymbol{\delta}\right)^T\mathrm{E}\left\{\imath\left(\boldsymbol{X}\right)\imath\left(\boldsymbol{X}\right)^T\right\}\imath\left(\overline{\boldsymbol{\delta}}\otimes\boldsymbol{u}+\overline{\boldsymbol{u}}\otimes\boldsymbol{\delta}\right). \tag{4.29}$$

Using the facts: **i)** the random vector $\imath\left(\boldsymbol{X}\right)$ is distributed as

$$\imath\left(\boldsymbol{X}\right) \sim \mathcal{N}\left(\imath\left(\overline{\boldsymbol{u}}\otimes\boldsymbol{u}\right), \frac{\sigma^2}{2}\boldsymbol{I}_{(n+1)^2}\right),$$

ii) for complex vectors $\boldsymbol{z}_1$ and $\boldsymbol{z}_2$ of the same size, the equality $\imath(\boldsymbol{z}_1)^T \imath(\boldsymbol{z}_2) = \mathrm{Re}\left(\boldsymbol{z}_1^H \boldsymbol{z}_2\right)$ holds, and **iii)** the horizontality of $\boldsymbol{\delta}$ and the unit-norm of $\boldsymbol{u}$ imply

$$\imath\left(\overline{\boldsymbol{\delta}} \otimes \boldsymbol{u} + \overline{\boldsymbol{u}} \otimes \boldsymbol{\delta}\right)^T \imath\left(\overline{\boldsymbol{u}} \otimes \boldsymbol{u}\right) = 0,$$

and

$$\left\|\imath\left(\overline{\boldsymbol{\delta}} \otimes \boldsymbol{u} + \overline{\boldsymbol{u}} \otimes \boldsymbol{\delta}\right)\right\|^2 = 2\left\|\boldsymbol{\delta}\right\|^2,$$

the equality in (4.29) becomes

$$I_{[\boldsymbol{u}]}\left(X_{[\boldsymbol{u}]}, X_{[\boldsymbol{u}]}\right) = \frac{4}{\sigma^2}\left\|\boldsymbol{\delta}\right\|^2.$$

Thus,

$$\begin{aligned} \lambda_{[\boldsymbol{u}]} &= \max_{|X_{[\boldsymbol{u}]}| = 1} I_{[\boldsymbol{u}]}\left(X_{[\boldsymbol{u}]}, X_{[\boldsymbol{u}]}\right) \\ &= \max_{\|\boldsymbol{\delta}\| = 1} \frac{4}{\sigma^2}\left\|\boldsymbol{\delta}\right\|^2 \\ &= 4/\sigma^2. \end{aligned} \tag{4.30}$$

In (4.30), we used the fact that π is a Riemannian submersion. Thus, $|X_{[\boldsymbol{u}]}| = |X_{\boldsymbol{u}}| = \|\boldsymbol{\delta}\|$. As an estimator for $[\boldsymbol{u}]$, we take the maximum likelihood (ML) estimator. Based on the log-likelihood function in (4.26), this is given by $\vartheta(\boldsymbol{X}) = [\boldsymbol{v}]$, where $\boldsymbol{v}$ denotes the unit-norm eigenvector corresponding to the maximum eigenvalue of $\boldsymbol{X} + \boldsymbol{X}^H$. It is difficult to establish analytically the bias of this proposed estimator. However, since it is consistent in SNR (that is, it converges to $[\boldsymbol{u}]$ as the noise power vanishes) we will assume that it is unbiased. That is, we consider that $\sigma_{[\boldsymbol{u}]} = 1$. Inserting $C = 4$, $\lambda_{[\boldsymbol{u}]} = 4/\sigma^2$ and $\sigma_{[\boldsymbol{u}]} = 1$ in (4.12) yields the bound

$$\mathrm{var}_{[\boldsymbol{u}]}(\vartheta) \geq \frac{4 + 3/\sigma^2 - \sqrt{\frac{1}{\sigma^2}\left(9/\sigma^2 + 24\right)}}{\frac{32}{3}}. \tag{4.31}$$

We conducted a set of computer simulations (similar to the one described in subsection 4.4.1) to evaluate the gap between both sides of (4.31). We considered $n = 2$ and let $p = [\boldsymbol{u}]$ denote the line spanned by $\boldsymbol{u} = (1, 0, 0)^T$. The SNR for the data model (4.25) is defined as

$$\mathrm{SNR} = \frac{\left\|\boldsymbol{u}\boldsymbol{u}^H\right\|^2}{\mathrm{E}\left\{\left\|\boldsymbol{W}\right\|^2\right\}} = \frac{1}{\sigma^2(n+1)^2}.$$

Figure 4.9 plots the result obtained. As seen, the IVLB underestimates the variance of ϑ, within a reasonable margin of error.

4.4.3 Inference on the quotient space $\mathbb{H}[z]/\sim$

As our last example, we consider an inference problem involving the quotient space of identifiable MIMO channel equivalence classes $\mathbb{H}[z]/\sim$. Consider the data model in (2.3) reproduced here for convenience,

$$\boldsymbol{y}[n] = \sum_{p=1}^{P} \boldsymbol{h}_p(z) \circledast s_p[n] + \boldsymbol{w}[n] = \boldsymbol{H}(z) \circledast \boldsymbol{s}[n] + \boldsymbol{w}[n]. \tag{4.32}$$

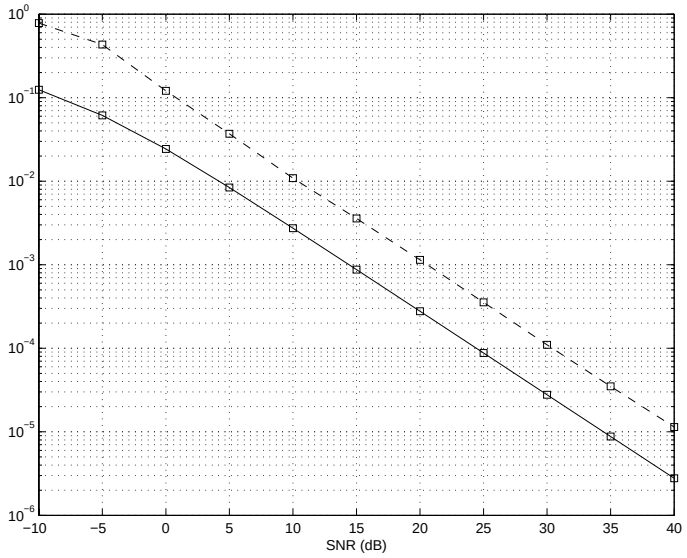

Figure 4.9: Estimated $\mathrm{var}_p(\vartheta)$ (dashed) and IVLB (solid)

For simplicity, we consider that $\deg \boldsymbol{H}(z) = 0$ (memoryless channel). That is, $\boldsymbol{h}_p(z) = \boldsymbol{h}_p[0] = \boldsymbol{h}_p$, for all p. Thus, in the identification $\boldsymbol{H}(z) \simeq (\boldsymbol{d}; \boldsymbol{H})$, we have $\boldsymbol{d} = (0, 0, \ldots, 0)^T$ and $\boldsymbol{H} = [\, \boldsymbol{h}_1 \, \boldsymbol{h}_2 \, \cdots \boldsymbol{h}_P \,]$. Equation (4.32) becomes

$$\boldsymbol{y}[n] = \boldsymbol{H}\boldsymbol{s}[n] + \boldsymbol{w}[n], \tag{4.33}$$

where $\boldsymbol{s}[n] = (s_1[n], s_2[n], \ldots, s_P[n])^T$. Arranging the N available channel observations in the data matrix $\boldsymbol{Y} = [\, \boldsymbol{y}[1] \, \boldsymbol{y}[2] \, \cdots \, \boldsymbol{y}[N] \,]$ leads to

$$\boldsymbol{Y} = \boldsymbol{H}\boldsymbol{S} + \boldsymbol{W}, \tag{4.34}$$

where $\boldsymbol{S} = [\, \boldsymbol{s}[1] \, \boldsymbol{s}[2] \, \cdots \, \boldsymbol{s}[N] \,]$ and $\boldsymbol{W} = [\, \boldsymbol{w}[1] \, \boldsymbol{w}[2] \, \cdots \, \boldsymbol{w}[N] \,]$. Equivalently,

$$\boldsymbol{y} = (\boldsymbol{H} \otimes \boldsymbol{I}_N) \; \boldsymbol{s} + \boldsymbol{w}, \tag{4.35}$$

where $\boldsymbol{y} = \mathrm{vec}\left(\boldsymbol{Y}^T\right)$, $\boldsymbol{s} = \mathrm{vec}\left(\boldsymbol{S}^T\right)$ and $\boldsymbol{w} = \mathrm{vec}\left(\boldsymbol{W}^T\right)$. Note that $\boldsymbol{s} = \left(\boldsymbol{s}_1^T, \boldsymbol{s}_2^T, \ldots, \boldsymbol{s}_P^T\right)^T$, where $\boldsymbol{s}_p = (s_p[1], s_p[2], \ldots, s_p[N])^T$. In (4.32), we assume that the sources $s_p[n]$ denote mutually independent zero-mean WSS Gaussian processes with known autocorrelation function $r_{s_p}[m] = \mathrm{E}\left\{ s_p[n] \overline{s_p[n-m]} \right\}$. The observation noise $\boldsymbol{w}[n]$ is taken to be spatio-temporal white Gaussian noise with variance σ^2 and statistically independent from the sources. This means that the random vector $\boldsymbol{y}$ in (4.35) is normally distributed as

$$\boldsymbol{y} \sim \mathcal{N}\left(\boldsymbol{0}, \boldsymbol{C}\left(\boldsymbol{H}\right)\right), \tag{4.36}$$

where the correlation matrix $\boldsymbol{C}(\boldsymbol{H}) = \mathrm{E}\left\{\boldsymbol{y}\boldsymbol{y}^H\right\}$ can be written as

$$\boldsymbol{C}\left(\boldsymbol{H}\right) = \sum_{p=1}^{P} \left(\boldsymbol{h}_p \boldsymbol{h}_p^H\right) \otimes \boldsymbol{R}_{\boldsymbol{s}_p} + \sigma^2 \boldsymbol{I}_{QN}, \tag{4.37}$$

with

$$\boldsymbol{R}_{\boldsymbol{s}_p} = \mathrm{E}\left\{\boldsymbol{s}_p \boldsymbol{s}_p^H\right\} = \begin{bmatrix} r_{s_p}[0] & \overline{r_{s_p}[1]} & \cdots & \cdots & \overline{r_{s_p}[N-1]} \\ r_{s_p}[1] & r_{s_p}[0] & \overline{r_{s_p}[1]} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & r_{s_p}[1] & r_{s_p}[0] & \overline{r_{s_p}[1]} \\ r_{s_p}[N-1] & \cdots & \cdots & r_{s_p}[1] & r_{s_p}[0] \end{bmatrix}.$$

Given the observation $\boldsymbol{y}$, we aim at estimating the MIMO channel $\boldsymbol{H}$. From (4.36) we see that all information about the channel is contained in the covariance matrix of $\boldsymbol{y}$. But, more importantly, we see that $\boldsymbol{H}$ is not identifiable: it is clear from (4.37) that the distribution of $\boldsymbol{y}$ is invariant to phase offsets in the columns of $\boldsymbol{H}$. That is, $\boldsymbol{C}(\boldsymbol{H}) = \boldsymbol{C}(\boldsymbol{H}\boldsymbol{\Theta})$ for any $\boldsymbol{\Theta} = \mathrm{diag}\left(e^{i\theta_1}, e^{i\theta_2}, \ldots, e^{i\theta_P}\right)$. Thus, only the MIMO channel equivalence classes are identifiable (we are assuming that the correlation matrices $\boldsymbol{R}_{\boldsymbol{s}_p}$ are sufficiently diverse to ensure uniqueness, see the discussion in section 2.5). Thus, our parametric statistical family is given by $\mathcal{F} = \{p(\boldsymbol{y}; [\boldsymbol{H}(z)]) : [\boldsymbol{H}(z)] \in \mathbb{H}[z]/\sim\}$, where

$$p(\boldsymbol{y}; [\boldsymbol{H}(z)]) = \frac{1}{\pi^{QN} \det(\boldsymbol{C}(\boldsymbol{H}))} \exp\left(-\boldsymbol{y}^H \boldsymbol{C}(\boldsymbol{H})^{-1} \boldsymbol{y}\right). \tag{4.38}$$

We recall that the main geometrical features of the Riemannian manifold $\mathbb{H}[z]/\sim$ were studied in subsection 3.4.1. In particular, the intrinsic distance between the points $[\boldsymbol{H}(z)]$ and $[\boldsymbol{G}(z)]$ is given by

$$d([\boldsymbol{H}(z)], [\boldsymbol{G}(z)]) = \sqrt{\sum_{p=1}^{P} \|\boldsymbol{g}_p\|^2 + \|\boldsymbol{h}_p\|^2 - 2|\boldsymbol{g}_p^H \boldsymbol{h}_p|},$$

where $\boldsymbol{H} = [\,\boldsymbol{h}_1\, \boldsymbol{h}_2\, \cdots\, \boldsymbol{h}_P\,]$ and $\boldsymbol{G} = [\,\boldsymbol{g}_1\, \boldsymbol{g}_2\, \cdots\, \boldsymbol{g}_P\,]$, recall (3.50). Furthermore, is was seen that the sectional curvature at each point $[\boldsymbol{H}(z)]$ can be upper-bounded as

$$C_{[\boldsymbol{H}(z)]} \leq 3 \max\left\{\frac{1}{\|\boldsymbol{h}_p\|^2} : p = 1, 2, \ldots, P\right\}, \tag{4.39}$$

recall (3.70).

We now derive the Fisher information form associated with our parametric statistical family. Given the distribution in (4.38), we have

$$\log p(\boldsymbol{y}; [\boldsymbol{H}(z)]) \propto -\log\det(\boldsymbol{C}(\boldsymbol{H})) - \boldsymbol{y}^H \boldsymbol{C}(\boldsymbol{H})^{-1} \boldsymbol{y}.$$

Letting $X_{[\boldsymbol{H}(z)]}$ denote a tangent vector in $T_{[\boldsymbol{H}(z)]}\mathbb{H}[z]/\sim$, we have

$$\begin{aligned} I_{[\boldsymbol{H}(z)]}\left(X_{[\boldsymbol{H}(z)]}, X_{[\boldsymbol{H}(z)]}\right) &= \mathrm{E}\left\{X_{[\boldsymbol{H}(z)]} \log p(\boldsymbol{y}; \cdot)\; X_{[\boldsymbol{H}(z)]} \log p(\boldsymbol{y}; \cdot)\right\} \\ &= \mathrm{E}\left\{\left(X_{[\boldsymbol{H}(z)]} f_{\boldsymbol{y}}\right)^2\right\}, \end{aligned} \tag{4.40}$$

where $f_{\boldsymbol{y}} : \mathbb{H}[z]/\sim \to \mathbb{R}$ is given by

$$f_{\boldsymbol{y}}([\boldsymbol{H}(z)]) = \log\det(\boldsymbol{C}(\boldsymbol{H})) + \boldsymbol{y}^H \boldsymbol{C}(\boldsymbol{H})^{-1} \boldsymbol{y}.$$

We have the composition $g_{\boldsymbol{y}} = f_{\boldsymbol{y}} \circ \pi$, where $g_{\boldsymbol{y}} : \mathbb{H}[z] \to \mathbb{R}$ is defined as

$$g_{\boldsymbol{y}}(\boldsymbol{H}(z)) = \log\det(\boldsymbol{C}(\boldsymbol{H})) + \boldsymbol{y}^H \boldsymbol{C}(\boldsymbol{H})^{-1} \boldsymbol{y}.$$

Since the projection onto the quotient $\pi : \mathbb{H}[z] \to \mathbb{H}[z]/\sim$ is a Riemannian submersion, there exists a horizontal tangent vector $X_{\boldsymbol{H}(z)} \in H_{\boldsymbol{H}(z)}\mathbb{H}[z]$ such that $X_{[\boldsymbol{H}(z)]} = \pi_* \left(X_{\boldsymbol{H}(z)}\right)$. Thus,

$$\begin{aligned} X_{[\boldsymbol{H}(z)]} f_{\boldsymbol{y}} &= \pi_* \left(X_{\boldsymbol{H}(z)}\right) f_{\boldsymbol{y}} \\ &= X_{\boldsymbol{H}(z)} \left(\pi^* f_{\boldsymbol{y}}\right) \\ &= X_{\boldsymbol{H}(z)} \left(f_{\boldsymbol{y}} \circ \pi\right) \\ &= X_{\boldsymbol{H}(z)}\, g_{\boldsymbol{y}}. \end{aligned}$$

Recall that the horizontal space at $\boldsymbol{H}(z)$ is given by

$$H_{\boldsymbol{H}(z)}\mathbb{H}[z] = \left\{(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_P) \in \mathbb{C}^Q \times \mathbb{C}^Q \times \cdots \times \mathbb{C}^Q : \operatorname{Im}\left(\boldsymbol{\delta}_p^H \boldsymbol{h}_p\right) = 0\right\},$$

see (3.48). Using this identification, let $X_{\boldsymbol{H}(z)} = (\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_P)$. It can be seen (after some straightforward computations) that we have

$$X_{\boldsymbol{H}(z)} g_{\boldsymbol{y}} = \operatorname{tr}\left(\boldsymbol{C}\left(\boldsymbol{H}\right)^{-1} \boldsymbol{\Delta}\right) - \boldsymbol{y}^H \boldsymbol{C}\left(\boldsymbol{H}\right) \boldsymbol{\Delta} \boldsymbol{C}\left(\boldsymbol{H}\right)^{-1} \boldsymbol{y}, \tag{4.41}$$

where

$$\boldsymbol{\Delta} = \sum_{p=1}^{P} \left(\boldsymbol{\delta}_p \boldsymbol{h}_p^H + \boldsymbol{h}_p \boldsymbol{\delta}_p^H\right) \otimes \boldsymbol{R}_{\boldsymbol{s}_p}.$$

Plugging (4.41) in (4.40) and simplifying yields

$$I_{[\boldsymbol{H}(z)]}\left(X_{[\boldsymbol{H}(z)]}, X_{[\boldsymbol{H}(z)]}\right) = \operatorname{tr}\left(\boldsymbol{C}\left(\boldsymbol{H}\right)^{-1} \boldsymbol{\Delta} \boldsymbol{C}\left(\boldsymbol{H}\right)^{-1} \boldsymbol{\Delta}\right).$$

Thus,

$$\begin{aligned} \lambda_{[\boldsymbol{H}(z)]} &= \max_{\left|X_{[\boldsymbol{H}(z)]}\right| = 1} I_{[\boldsymbol{H}(z)]}\left(X_{[\boldsymbol{H}(z)]}, X_{[\boldsymbol{H}(z)]}\right) \\ &= \max_{\substack{\|(\boldsymbol{\delta}_1, \boldsymbol{\delta}_2, \ldots, \boldsymbol{\delta}_P)\| = 1 \\ \operatorname{Im}\left(\boldsymbol{\delta}_p^H \boldsymbol{h}_p\right) = 0}} \operatorname{tr}\left(\boldsymbol{C}\left(\boldsymbol{H}\right)^{-1} \boldsymbol{\Delta} \boldsymbol{C}\left(\boldsymbol{H}\right)^{-1} \boldsymbol{\Delta}\right). \end{aligned} \tag{4.42}$$

The optimization problem expressed in (4.42) requires the maximization of a quadratic function subject to linear and quadratic constraints. Upon a choice of basis for the horizontal space $H_{\boldsymbol{H}(z)}\mathbb{H}[z]$ the linear constraints can be dropped and (4.42) boils down to the computation of the maximum eigenvalue of an Hermitean matrix. We do not find an explicit formula for this Hermitean matrix, because this leads to a rather complicated formula and it is not strictly needed for our purposes. In the computer simulations to be presented, we evaluated $\lambda_{[\boldsymbol{H}(z)]}$ by solving the optimization problem in (4.42) through a general-purpose optimization software package.

We carried out some simulations to compare the IVLB with the (intrinsic) variance of the ML estimator of the channel class $[\boldsymbol{H}(z)]$ in the data model (4.34). Given the distribution in (4.38) the ML estimator is given by $\vartheta\left(\boldsymbol{y}\right) = [\widehat{\boldsymbol{H}}(z)]$ where

$$\widehat{\boldsymbol{H}} = \operatorname*{arg\,min}_{\boldsymbol{G} = \left[\begin{smallmatrix} \boldsymbol{g}_1 \boldsymbol{g}_2 \cdots \boldsymbol{g}_p \end{smallmatrix}\right] \in \mathbb{H}[z]} \log\det\left(\boldsymbol{C}\left(\boldsymbol{G}\right)\right) + \boldsymbol{y}^H \boldsymbol{C}\left(\boldsymbol{G}\right)^{-1} \boldsymbol{y}.$$

In the simulations, the ML estimate was found through an optimization software package. We considered a MIMO channel (randomly generated and kept fixed throughout the simulations) given by

$$\boldsymbol{H} = \begin{bmatrix} 0.0860 - 0.6313i & 0.4620 + 1.0556i \\ -2.0046 - 2.3252i & -0.3210 - 0.1132i \\ -0.4931 - 1.2316i & 1.2366 + 0.3792i \end{bmatrix}.$$

The pth source signal $s_p[n]$ is a WSS Gaussian process obtained by passing an unit-power white Gaussian process $a_p[n]$ through a correlative filter $c_p(z)$, that is, $s_p[n] = c_p(z) \circledast a_p[n]$. The pre-filters are given by $c_p(z) = \kappa_p \left(1 - z_p z^{-1}\right)$ where κ_p is an unit-power normalizing constant and z_p denotes the zero of the pth filter. We used the same zeros as in the two-users scenario of section 2.8 (page 39), that is, $z_1 = \frac{1}{4}e^{-i\pi/2}$ and $z_2 = \frac{1}{2}e^{i\pi/4}$. The ML estimator is assumed to be unbiased that is, $\sigma_{[\boldsymbol{H}(z)]} = 1$. Furthermore, we suppose that the receiver knows the power of each user's channel vector, within a 50% relative error. That is, we take as an upper-bound for the sectional curvature, the value

$$C = 3 \max \left\{ \frac{1}{\left(0.5 \left\|\boldsymbol{h}_p\right\|\right)^2} \, : \, p = 1, 2 \right\} = 3.8503.$$

The signal-to-noise ratio for the data model (4.33) is defined by

$$\mathrm{SNR} = \frac{\mathrm{E}\left\{ \left\|\boldsymbol{H}\boldsymbol{s}[n]\right\|^2 \right\}}{\mathrm{E}\left\{ \left\|\boldsymbol{w}[n]\right\|^2 \right\}} = \frac{\|\boldsymbol{H}\|^2}{Q\sigma^2},$$

and is varied between $\mathrm{SNR}_{\min} = -5$ dB and $\mathrm{SNR}_{\max} = 10$ dB, in steps of $\mathrm{SNR}_{\mathrm{step}} = 5$ dB. Figure 4.10 shows the results obtained through computer simulations. We see that,

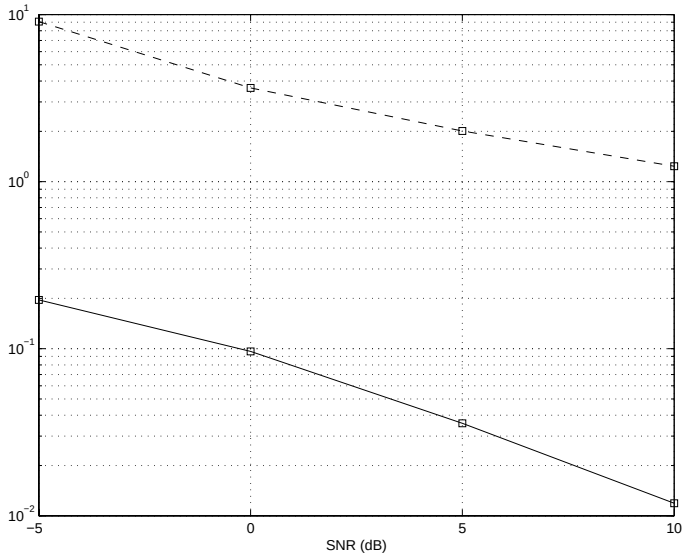

Figure 4.10: Estimated $\mathrm{var}_{[\boldsymbol{H}(z)]}(\vartheta)$ (dashed) and IVLB (solid)

although the IVLB lower bounds the observed intrinsic variance of the ML estimator, a more significative gap is noticeable. This might be due to the fact that, for this inference problem, the IVLB is not attainable (by any estimator), in the same sense that the CRLB is not attainable in certain estimation scenarios.

4.5 Conclusions

In this chapter, we considered parametric statistical families $\mathcal{F} = \{f_p : \Omega \to \mathbb{R}\}$ (Ω=sample space) where the parameter p (indexing the family) lies in a Riemannian manifold P. This is the setup emerging spontaneously from several applications, either due to parameter restrictions (P is a submanifold of some Euclidean space) or to intrinsic ambiguities in the observation model (P is a coset space, that is, the set of identifiable parameter equivalence classes). The latter situation arises, for example, in the problem of identifying a MIMO channel excited by complex circular Gaussian inputs and given a finite number of observations (see subsection 4.4.3 for a more detailed discussion). In either case, the language of differential-geometry permits to unify their treatment. Let M denote a connected Riemannian manifold, possibly distinct from P, and let $b : P \to M$ denote a smooth mapping. We proposed the intrinsic variance lower bound (IVLB) which, in loose terms (see theorem 4.1 for the precise statement), places a lower limit on the intrinsic variance $\mathrm{var}_p(\vartheta) = \mathrm{E}_p\left\{d(\vartheta, b(p))^2\right\}$ of any estimator ϑ with bias b. Here, d denotes the intrinsic (Riemannian) distance on M. The IVLB takes into account: **i)** the geometry of the manifold M through its sectional curvatures, **ii)** the given statistical model $\mathcal{F}$ through its associated Fisher information form and **iii)** the bias b through its "derivative", that is, the push-forward map b_*. Some inference problems were analyzed to examine the utility (tightness) of the bound. Although the preliminary results obtained were satisfactory, a more in-depth study is needed to fully characterize the capabilities of the proposed bound.

Chapter 5

Conclusions

5.1 Open issues and future work

The main problem addressed in this thesis was blind identification of multiple-input multiple-output (MIMO) channels based only on the 2nd order statistics of the channel observations. Since a phase ambiguity per column of the MIMO transfer matrix cannot be avoided, this consists in an inference problem on the quotient space of channel equivalence classes (where two channels are identified if they are equal modulo a phase offset per column). As we have seen, an in-depth treatment of this problem under this viewpoint has launched some new interesting challenges such as asymptotic analysis and performance bounds within the setting of Riemannian manifolds. These latter problems have found elegant solutions in the language of Riemannian differential geometry. Furthermore, the theory was developed in all generality and can thus be applied to other parametric estimation problems. In this last chapter, we complement the discussions provided at the end of chapters 2–4. However, the goal here is to identify some open issues and point directions for future work and research.

5.1.1 Chapter 2

Our analytical solution for the blind channel identification problem (BCIP) exploits the fact that the inputs of the MIMO channel have distinct 2nd order spectra and that their 2nd order statistical characterization is known by the receiver. This can be assumed since the information symbols are colored prior to transmission by correlative filters which are known at the receiver. The proposed closed-form identification algorithm (CFIA) finds the MIMO channel coefficients by matching the theoretical and observed correlation matrices of the channel observations. The main weakness of this algorithm is its vulnerability with respect to estimation errors in the channel orders. This is a common drawback of several blind channel identification approaches [61, 43, 24]. Therefore, an interesting possibility to explore in the future consists in circumventing this weakness by estimating directly the linear space-time equalizer, instead of first identifying the MIMO channel and then design the corresponding equalizer (as we do here). Remark that the former approach does not need, in principle, the knowledge of the exact channel orders. To achieve this extension, the identifiability theorem 2.1 should be rephrased in terms of the equalizer coefficients. That is, it should be re-formulated in order to guarantee channel

inversion, for example, as soon as the equalizer is able to reproduce signals with the same 2nd order statistics of the inputs. At this moment, it is not clear that this alternative approach would be also implementable in closed-form but this topic certainly deserves further attention. Another line for future research consists in investigating the possibility of dropping certain channel assumptions, like the technique of blind identification by decorrelating subchannels (BIDS), see [30, 31], which does not require the channel to be simultaneously irreducible and column-reduced as we do (and also most other multiple-user approaches do, see [1, 2, 24, 68, 38]). Furthermore, even if the identifiability theorem 2.1 can be extended to such relaxed channel assumptions, it seems that the CFIA must also suffer major modifications in order to cope with that new scenario.

5.1.2 Chapter 3

We recall that the asymptotic performance analysis presented in chapter 3 was carried out under a major assumption: the channel orders are known and the CFIA only estimates the MIMO channel coefficients. We made such a simplifying assumption in order to keep the theoretical study tractable. We also recall that the main utility of the performance analysis consists in providing guidelines for the (off-line) optimum design of the correlative filters (in terms of the mean-square distance of the channel class estimate) when one is given a stochastic model for the MIMO channel. For future work, we should develop a broader theoretical study which would also permit to design the correlative filters in order to minimize the probability of detection errors in the MIMO channel orders (as mentioned above, this is the main drawback of the proposed CFIA). As we pointed out earlier in page 34, correct detection of the channel orders boils down to the correct detection of the (unique) zero of the function ϕ in (2.42) over the discrete set $\mathbb{E}$, using only finite-length data packets. This is in turn linked with the matching cost functional proposed in (2.41) which should perhaps be computed with respect to a weighting matrix (to be determined) rather than simply the identity matrix. Since the new performance study should be conducted without the assumption of known channel orders, this means that the "full" geometrical model of $\mathbb{H}[z]$ as a finite set of leaves, recall figure 3.1 in page 52, should now be taken into account. That is, the simplification taken in chapter 3 which consists in redefining the set of MIMO channels $\mathbb{H}[z]$ as a single leaf (corresponding to the known channel orders) is no longer valid. This complicates matters greatly since a Riemannian model for this set is not clearly available (distinct leaves can have distinct dimensionality). Nevertheless, it seems an interesting problem, both from the theoretical and practical viewpoints, to work on.

5.1.3 Chapter 4

In this chapter, we introduced the intrinsic variance lower bound (IVLB). The IVLB essentially extends the Cramér-Rao lower bound (CRLB) to inference problems involving statistical families indexed over Riemannian manifolds and where the intrinsic mean-square (Riemannian) distance is the figure of merit used to evaluate the accuracy of estimators. This chapter is perhaps the most fertile in open issues and contains several exciting directions to explore in the future. A basic question to be answered is, for example, what are the sufficient and necessary conditions on the statistical family and on the geometries

of the involved Riemannian manifolds for the IVLB to be achievable ? Another theoretical point to be further investigated consists in the improvement of the proposed IVLB for certain Riemannian structures, for example spaces of constant sectional curvature, for which detailed characterizations and specific results are available in the mathematical literature. Also, parametric inference problems over quotient spaces usually arise from parametric estimation problems formulated over Euclidean spaces but with ambiguities in the parameters. A nice result to be obtained here would consist in showing that the IVLB can be obtained from the (necessarily singular) Fisher-information matrix (FIM) of the parameter plus some modifications arising from the problem-dependent geometry. In that line, the link between the IVLB and the works in [22, 41, 55] which address the special case of submanifolds embedded in Euclidean spaces deserves also more investigation. Finally, another direction for future research is the extension of the IVLB to the context of Bayesian estimation problems, where one disposes of a probabilistic prior for the parameter of interest (which lives in a Riemannian manifold now).

Appendix A

Proofs for Chapter 2

A.1 Proof of Theorem 2.1

First, we need some technical lemmas.

Lemma A.1. *Let $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{C}^n$, and $\boldsymbol{y} \neq \boldsymbol{0}$. If $\boldsymbol{x}^H \boldsymbol{K}_n[m]\boldsymbol{y} = 0$, for all $m \in \mathbb{Z}$, then $\boldsymbol{x} = \boldsymbol{0}$.*

Proof. Write $\boldsymbol{y} = (0, \ldots, 0, y_l, \ldots, y_n)^T$, where $y_l \neq 0$. Define the $n \times n$ matrix

$$\boldsymbol{Y} = \left[\begin{array}{cccc} \boldsymbol{K}_n[-l+1]\boldsymbol{y} & \boldsymbol{K}_n[-l+2]\boldsymbol{y} & \cdots & \boldsymbol{K}_n[n-l]\boldsymbol{y} \end{array} \right].$$

Note that $\boldsymbol{Y}$ is a Toeplitz lower triangular matrix of the form

$$\boldsymbol{Y} = \left[\begin{array}{cccc} y_l & 0 & \cdots & 0 \\ * & y_l & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ * & \cdots & * & y_l \end{array} \right].$$

Since $y_l \neq 0$, the matrix $\boldsymbol{Y}$ is non-singular. By hypothesis, $\boldsymbol{x}^H \boldsymbol{Y} = \boldsymbol{0}$. Thus, $\boldsymbol{x} = \boldsymbol{0}$ □

Lemma A.2. *Let $\boldsymbol{A} = diag(\boldsymbol{A}_1, \boldsymbol{A}_2, \ldots, \boldsymbol{A}_n)$, where $\boldsymbol{A}_i \in \mathbb{C}^{n_i \times n_i}$. Assume that $\sigma(\boldsymbol{A}_i) \cap \sigma(\boldsymbol{A}_j) = \varnothing$, for $i \neq j$. If $\boldsymbol{B}$ commutes with $\boldsymbol{A}$, that is, $\boldsymbol{AB} = \boldsymbol{BA}$, then $\boldsymbol{B} = diag(\boldsymbol{B}_1, \boldsymbol{B}_2, \ldots, \boldsymbol{B}_n)$, for some $\boldsymbol{B}_i \in \mathbb{C}^{n_i \times n_i}$.*

Proof. We use the fact that, if $\boldsymbol{XY} - \boldsymbol{YZ} = \boldsymbol{0}$ and $\sigma(\boldsymbol{X}) \cap \sigma(\boldsymbol{Z}) = \emptyset$, then $\boldsymbol{Y} = \boldsymbol{0}$ [23, lemma 7.1.5, page 336]. Write

$$\boldsymbol{B} = \left[\begin{array}{cccc} \boldsymbol{B}_{11} & \boldsymbol{B}_{12} & \cdots & \boldsymbol{B}_{1n} \\ \boldsymbol{B}_{21} & \boldsymbol{B}_{22} & \cdots & \boldsymbol{B}_{2n} \\ \vdots & \vdots & & \vdots \\ \boldsymbol{B}_{n1} & \boldsymbol{B}_{n2} & \cdots & \boldsymbol{B}_{nn} \end{array} \right],$$

where $\boldsymbol{B}_{ij} \in \mathbb{C}^{n_i \times n_j}$. From $\boldsymbol{AB} = \boldsymbol{BA}$, it follows that $\boldsymbol{A}_i \boldsymbol{B}_{ij} = \boldsymbol{B}_{ij} \boldsymbol{A}_j$. Since $\sigma(\boldsymbol{A}_i) \cap \sigma(\boldsymbol{A}_j) = \emptyset$ whenever $i \neq j$, we conclude that $\boldsymbol{B}_{ij} = \boldsymbol{0}$ whenever $i \neq j$ □

Lemma A.3. *Let $\boldsymbol{V} \in \mathbb{C}^{m \times n} (m \geq n)$ denote an isometry, that is, $\boldsymbol{V}^H \boldsymbol{V} = \boldsymbol{I}_n$, and let $\boldsymbol{F}[m] = \boldsymbol{V}^H \boldsymbol{K}_n[m] \boldsymbol{V}$, for $m \in \mathbb{Z}$. If $\boldsymbol{X} \in \mathbb{C}^{n \times n}$ commutes with $\boldsymbol{F}[m]$, for all $m \in \mathbb{Z}$, then $\boldsymbol{X} = \lambda \boldsymbol{I}_n$, for some $\lambda \in \mathbb{C}$.*

Proof. First, consider that $\boldsymbol{X}$ is a normal matrix, that is,

$$\boldsymbol{X} = \boldsymbol{Q}\boldsymbol{\Lambda}\boldsymbol{Q}^H, \tag{A.1}$$

where $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_1 \boldsymbol{I}_{n_1}, \lambda_2 \boldsymbol{I}_{n_2}, \ldots, \lambda_l \boldsymbol{I}_{n_l})$, with $\lambda_i \neq \lambda_j$, for $i \neq j$, and $\boldsymbol{Q}$ is unitary, see [23, corollary 7.1.4, page 336]. Consider the hypothesis that $l \geq 2$. Using (A.1) in

$$\boldsymbol{F}[m]\boldsymbol{X} = \boldsymbol{X}\boldsymbol{F}[m], \tag{A.2}$$

yields $\boldsymbol{G}[m]\boldsymbol{\Lambda} = \boldsymbol{\Lambda}\boldsymbol{G}[m]$, where $\boldsymbol{G}[m] = \boldsymbol{W}^H \boldsymbol{K}_n[m]\boldsymbol{W}$ and $\boldsymbol{W} = \boldsymbol{V}\boldsymbol{Q}$ is an isometry. Since $\boldsymbol{G}[m]$ commutes with $\boldsymbol{\Lambda}$, lemma A.2 asserts that $\boldsymbol{G}[m] = \operatorname{diag}(\boldsymbol{G}_1[m], \boldsymbol{G}_2[m], \ldots, \boldsymbol{G}_l[m])$, where $\boldsymbol{G}_i[m] \in \mathbb{C}^{n_i \times n_i}$. Thus, all matrices $\boldsymbol{G}[m]$ share at least one common zero entry. For example, any entry off the block diagonal. That is, there exists a pair of indices (i, j) such that $\boldsymbol{G}_{ij}[m] = \boldsymbol{w}_i^H \boldsymbol{K}_n[m]\boldsymbol{w}_j = 0$, for all $m \in \mathbb{Z}$, where $\boldsymbol{w}_i$ denotes the ith column of $\boldsymbol{W}$. From lemma A.1, we conclude that $\boldsymbol{w}_i = \boldsymbol{0}$ (contradiction). Thus, we must have $l = 1$ and $\boldsymbol{X} = \lambda \boldsymbol{I}_n$. We now turn to the general case. Given that $\boldsymbol{K}_n[m] = \boldsymbol{K}_n[-m]^H$, we have

$$\boldsymbol{F}[m]\boldsymbol{X}^H = \boldsymbol{X}^H \boldsymbol{F}[m] \tag{A.3}$$

Combining (A.2) and (A.3) gives $\boldsymbol{F}[m]\boldsymbol{A} = \boldsymbol{A}\boldsymbol{F}[m]$ and $\boldsymbol{F}[m]\boldsymbol{B} = \boldsymbol{B}\boldsymbol{F}[m]$, for all $m \in \mathbb{Z}$, where $\boldsymbol{A}$ and $\boldsymbol{B}$ came from the Cartesian decomposition of $\boldsymbol{X}$. That is $\boldsymbol{X} = \boldsymbol{A} + i\boldsymbol{B}$ and

$$\boldsymbol{A} = \frac{\boldsymbol{X} + \boldsymbol{X}^H}{2}, \qquad \boldsymbol{B} = \frac{\boldsymbol{X} - \boldsymbol{X}^H}{2i}.$$

Since both $\boldsymbol{A}$ and $\boldsymbol{B}$ are normal matrices (in fact, Hermitean), it follows from the first part of the proof that $\boldsymbol{A} = \lambda_r \boldsymbol{I}_n$ and $\boldsymbol{B} = \lambda_i \boldsymbol{I}_n$, for some $\lambda_r, \lambda_i \in \mathbb{R}$. Thus, $\boldsymbol{X} = (\lambda_r + i\lambda_i)\boldsymbol{I}_n$ □

Proof of Theorem 2.1. To settle notation $\boldsymbol{H}(z) = [\,\boldsymbol{h}_1(z)\,\boldsymbol{h}_2(z)\,\cdots\,\boldsymbol{h}_P(z)\,]$, $D_p = \deg \boldsymbol{h}_p(z)$ and $\boldsymbol{H}_p = \boldsymbol{\mathcal{T}}_0(\boldsymbol{h}_p(z)) = [\,\boldsymbol{h}_p[0]\,\boldsymbol{h}_p[1]\,\cdots\,\boldsymbol{h}_p[D_p]\,]$. Similarly, we let $\boldsymbol{G}(z) = [\,\boldsymbol{g}_1(z)\,\boldsymbol{g}_2(z)\,\cdots\,\boldsymbol{g}_P(z)\,]$, $E_p = \deg \boldsymbol{g}_p(z)$ and $\boldsymbol{G}_p = \boldsymbol{\mathcal{T}}_0\left(\boldsymbol{g}_p(z)\right) = \left[\,\boldsymbol{g}_p[0]\,\boldsymbol{g}_p[1]\,\cdots\,\boldsymbol{g}_p[E_p]\,\right]$. The proof that $\varphi([\boldsymbol{G}(z)]) = \varphi([\boldsymbol{H}(z)])$ implies $[\boldsymbol{G}(z)] = [\boldsymbol{H}(z)]$ is carried out incrementally in 3 steps, in which we establish: **(1)** $\operatorname{ord}\boldsymbol{G}(z) = \operatorname{ord}\boldsymbol{H}(z)$, that is, $\sum_{p=1}^{P} E_p = \sum_{p=1}^{P} D_p$, **(2)** $E_p = D_p$, for all p, and, finally, **(3)** $\boldsymbol{G}_p = \boldsymbol{H}_p e^{i\theta_p}$ for some $\theta_p \in \mathbb{R}$. Keep in mind that both $\boldsymbol{H}(z)$ and $\boldsymbol{G}(z)$ are taken from the "admissible" set $\mathbb{H}[z]$ (assumption A1 in page 20). For simplicity, throughout the proof, we consider noiseless samples $\boldsymbol{y}[n]$, that is, $\boldsymbol{w}[n] = \boldsymbol{0}$ in (2.3). This entails no loss of generality, because our proof only relies on the 2nd order statistics of $\boldsymbol{y}[n]$.

Step 1: $\operatorname{ord}\boldsymbol{G}(z) = \operatorname{ord}\boldsymbol{H}(z)$**.** Let $\mathbb{C}(z)$ denote the vector space of Q-tuples of rational functions, in the indeterminate z^{-1}, over the field of rational functions [14, 1]. Let $\mathcal{S}_{\boldsymbol{H}}(z) \subset \mathbb{C}(z)$ denote the P-dimensional subspace spanned by $\boldsymbol{H}(z)$, and $\mathcal{S}_{\boldsymbol{H}}^{\perp}(z) \subset \mathbb{C}(z)$ its $(Q-P)$-dimensional dual subspace [14, 1]. Similar definitions hold for $\mathcal{S}_{\boldsymbol{G}}(z)$ and $\mathcal{S}_{\boldsymbol{G}}^{\perp}(z)$. As shown in [1], if L (a stacking parameter) is high enough, then $\mathcal{S}_{\boldsymbol{H}}^{\perp}(z)$ is uniquely determined from the correlation matrix $\boldsymbol{R}_{\boldsymbol{y}}[0; L]$. Thus, $\mathcal{S}_{\boldsymbol{H}}^{\perp}(z) = \mathcal{S}_{\boldsymbol{G}}^{\perp}(z)$ and, as a consequence, $\mathcal{S}_{\boldsymbol{H}}(z) = \mathcal{S}_{\boldsymbol{G}}(z)$. Because both $\boldsymbol{H}(z)$ and $\boldsymbol{G}(z)$ are irreducible and column-reduced, they are minimal polynomial basis for $\mathcal{S}_{\boldsymbol{H}}(z) = \mathcal{S}_{\boldsymbol{G}}(z)$, see [14]. Thus, they have the same order [14]. That is, $\sum_{p=1}^{P} E_p = \sum_{p=1}^{P} D_p$.

Step 2: $E_p = D_p$**.** Recall the notation for the stacked data model of order L given in table 2.2, page 19. Furthermore, recall assumptions A1 and A3 in pages 20 and 27 for the

definitions of the constants $D_{\max}$ and E, respectively. Choose a stacking parameter

$$L \leq E - D_{\max}, \tag{A.4}$$

such that the stacked channel matrix

$$\boldsymbol{\mathcal{H}} = \left[\begin{array}{cccc} \boldsymbol{\mathcal{H}}_1 & \boldsymbol{\mathcal{H}}_2 & \cdots & \boldsymbol{\mathcal{H}}_P \end{array} \right],$$

$\boldsymbol{\mathcal{H}}_p = \boldsymbol{\mathcal{T}}_L(\boldsymbol{h}_p(z))$, is full column-rank. Such a L exists because the channel matrix $\boldsymbol{H}(z)$ was assumed to be tall, irreducible and column-reduced (see assumption A1 in page 20). As a consequence of this assumption, the channel matrices $\boldsymbol{\mathcal{H}}$ corresponding to stacked data models of order L are full column-rank for all L sufficiently large, say $L \geq L_0$. The value $L_0 = \operatorname{ord} \boldsymbol{H}(z)$ works, see [24]. Thus, a valid lower bound covering all possible $\boldsymbol{H}(z) \in \mathbb{H}[z]$ is $L_0 = PD_{\max}$. Note that $L_0 \leq E - D_{\max}$.

Let

$$R = \operatorname{rank}(\boldsymbol{\mathcal{H}}) = P(L+1) + \sum_{p=1}^{P} D_p$$

and let $\boldsymbol{R}_{\boldsymbol{y}}[m; L] = \mathrm{E}\left\{\boldsymbol{y}[n; L]\boldsymbol{y}[n-m; L]^H\right\}$, $m \in \mathbb{Z}$, denote the correlation matrices of $\boldsymbol{y}[n; L]$. Notice that, from $\boldsymbol{y}[n; L] = \boldsymbol{\mathcal{H}}\, \boldsymbol{s}[n; \boldsymbol{\mathfrak{d}}]$, where $\boldsymbol{\mathfrak{d}} = (\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_P)^T$, $\mathcal{D}_p = D_p + L$, we have

$$\begin{aligned} \boldsymbol{R}_{\boldsymbol{y}}[0; L] &= \boldsymbol{\mathcal{H}}\, \boldsymbol{R}_{\boldsymbol{s}}[0; \boldsymbol{\mathfrak{d}}]\, \boldsymbol{\mathcal{H}}^H & \text{(A.5)} \\ &= \boldsymbol{\mathcal{H}}_0\, \boldsymbol{\mathcal{H}}_0^H, & \text{(A.6)} \end{aligned}$$

where $\boldsymbol{\mathcal{H}}_0 = \boldsymbol{\mathcal{H}}\, \boldsymbol{R}_{\boldsymbol{s}}[0; \boldsymbol{\mathfrak{d}}]^{1/2}$. Furthermore, note that $\boldsymbol{R}_{\boldsymbol{s}}[0; \boldsymbol{\mathfrak{d}}]$ is a $R \times R$ positive definite (in particular, nonsingular) matrix. This results from the fact that, as seen in (2.8), the matrix $\boldsymbol{R}_{\boldsymbol{s}}[0; \boldsymbol{\mathfrak{d}}]$ is a block diagonal concatenation of matrices of the form $\boldsymbol{R}_{\boldsymbol{s}_p}[0; \mathcal{D}_p]$, which in turn, from (2.21), are Gramian matrices given by

$$\boldsymbol{R}_{\boldsymbol{s}_p}[0; \mathcal{D}_p] = \boldsymbol{\mathcal{T}}_{\mathcal{D}_p}(\boldsymbol{c}_p(z))\, \boldsymbol{\mathcal{T}}_{\mathcal{D}_p}(\boldsymbol{c}_p(z))^H.$$

Since $\boldsymbol{\mathcal{T}}_{\mathcal{D}_p}(\boldsymbol{c}_p(z))$ is full row rank (see (2.20) and use the fact that $c_p[0] \neq 0$ by definition), we conclude that each $\boldsymbol{R}_{\boldsymbol{s}_p}[0; \mathcal{D}_p]$ is positive-definite. Finally, $\boldsymbol{R}_{\boldsymbol{s}}[0; \boldsymbol{\mathfrak{d}}]$, being the diagonal concatenation of positive-definite matrices, is itself positive-definite.

From (A.5) and the above considerations, we have $\operatorname{rank}(\boldsymbol{R}_{\boldsymbol{y}}[0; L]) = R$. Let

$$\boldsymbol{R}_{\boldsymbol{y}}[0; L] = \boldsymbol{V}\, \boldsymbol{\Sigma}^2\, \boldsymbol{V}^H \tag{A.7}$$

denote a R-truncated EVD of the matrix $\boldsymbol{R}_{\boldsymbol{y}}[0; L]$. That is, $\boldsymbol{V}^H\boldsymbol{V} = \boldsymbol{I}_R$ and $\boldsymbol{\Sigma} = \operatorname{diag}(\sigma_1, \ldots, \sigma_R)$ with $\sigma_i > 0$. Define $\boldsymbol{P} = \boldsymbol{V}\boldsymbol{\Sigma}$. Thus,

$$\boldsymbol{R}_{\boldsymbol{y}}[0; L] = \boldsymbol{P}\boldsymbol{P}^H. \tag{A.8}$$

From (A.6) and (A.8), we conclude that

$$\boldsymbol{P} = \boldsymbol{\mathcal{H}}_0\, \boldsymbol{Q}^H \tag{A.9}$$

for some $R \times R$ unitary matrix $\boldsymbol{Q}$. Thus, defining $\boldsymbol{\Upsilon}[m] = \boldsymbol{P}^+\, \boldsymbol{R}_{\boldsymbol{y}}[m; L]\, \boldsymbol{P}^{+H}$ yields

$$\boldsymbol{\Upsilon}[m] = \boldsymbol{Q}\, \boldsymbol{\Gamma}_{\boldsymbol{s}}[m; \boldsymbol{\mathfrak{d}}]\, \boldsymbol{Q}^H, \tag{A.10}$$

where

$$\boldsymbol{\Gamma}_{\boldsymbol{s}}[m;\mathfrak{d}] = \boldsymbol{R}_{\boldsymbol{s}}[0;\mathfrak{d}]^{-1/2}\,\boldsymbol{R}_{\boldsymbol{s}}[m;\mathfrak{d}]\,\boldsymbol{R}_{\boldsymbol{s}}[0;\mathfrak{d}]^{-1/2},$$

recall (2.23).

But, the same reasoning in terms of the polynomial matrix $\boldsymbol{G}(z)$, leads to $\boldsymbol{R}_{\boldsymbol{y}}[0;L] = \boldsymbol{\mathcal{G}}_0\,\boldsymbol{\mathcal{G}}_0^H$, where $\boldsymbol{\mathcal{G}}_0 = \boldsymbol{\mathcal{G}}\,\boldsymbol{R}_{\boldsymbol{s}}[0;\mathfrak{e}]^{1/2}$,

$$\boldsymbol{\mathcal{G}} = \left[\begin{array}{cccc} \boldsymbol{\mathcal{G}}_1 & \boldsymbol{\mathcal{G}}_2 & \cdots & \boldsymbol{\mathcal{G}}_P \end{array}\right],$$

with $\boldsymbol{\mathcal{G}}_p = \boldsymbol{\mathcal{T}}_L\left(\boldsymbol{g}_p(z)\right)$, $\mathfrak{e} = (\mathcal{E}_1, \mathcal{E}_2, \ldots, \mathcal{E}_P)^T$ and $\mathcal{E}_p = E_p + L$. Thus, we also have

$$\boldsymbol{P} = \boldsymbol{\mathcal{G}}_0\,\boldsymbol{W}^H, \tag{A.11}$$

for some $R \times R$ unitary matrix $\boldsymbol{W}$, and, consequently,

$$\boldsymbol{\Upsilon}[m] = \boldsymbol{W}\,\boldsymbol{\Gamma}_{\boldsymbol{s}}[m;\mathfrak{e}]\,\boldsymbol{W}^H. \tag{A.12}$$

We want to prove that $E_p = D_p$, or, equivalently, $\mathcal{E}_p = \mathcal{D}_p$, for all p. Assume the opposite. Since, from step 1, we have

$$\sum_{p=1}^{P} D_p = \sum_{p=1}^{P} E_p,$$

then we must have $\mathcal{D}_p > \mathcal{E}_p$ for some p. In the sequel, whenever $\boldsymbol{X}$ denotes a $n \times n$ matrix, the notation $p(t) \sim \boldsymbol{X}$ means that $p(t)$ is the characteristic polynomial of $\boldsymbol{X}$. That is, $p(t)$ is the polynomial in the indeterminate t and complex coefficients given by $p(t) = \det(t\boldsymbol{I}_n - \boldsymbol{X})$. Note that such a polynomial $p(t)$ has always degree n. Furthermore, for a given polynomial $p(t)$ of degree n and indeterminate t, we let $\sigma(p(t))$ denote its set of n roots in $\mathbb{C}$, including multiplicities. Thus, trivially, $p(t) \sim \boldsymbol{X}$ implies $\sigma(\boldsymbol{X}) = \sigma(p(t))$.

Notice that

$$\begin{aligned} \mathcal{D}_p &= D_p + L \\ &\leq D_{\max} + L \\ &\leq E, \end{aligned}$$

where the last inequality follows from (A.4). The same reasoning yields $\mathcal{E}_p \leq E$. Thus, by assumption A3 in page 27, there exists a correlation lag $m(p)$ satisfying (2.25). Let $p(t) \sim \boldsymbol{\Upsilon}[m(p)]$. From (A.10), we have

$$p(t) = h_p(t) \prod_{q \neq p} h_q(t),$$

where $h_i(t) \sim \boldsymbol{\Gamma}_{\boldsymbol{s}_i}[m(p);\mathcal{D}_i]$. But, from (A.12), we must also have

$$p(t) = g_p(t) \prod_{q \neq p} g_q(t),$$

where $g_i(t) \sim \boldsymbol{\Gamma}_{\boldsymbol{s}_i}[m(p);\mathcal{E}_i]$. The property of the lag $m(p)$ expressed in (2.25) implies that

$$\sigma\left(h_p(t)\right) \cap \sigma\left(g_q(t)\right) = \emptyset,$$

for $q \neq p$. Thus, necessarily, $\sigma\left(h_p(t)\right) \subset \sigma\left(g_p(t)\right)$. But, this is a contradiction, since the cardinality of $\sigma\left(h_p(t)\right\}$ (that is, $\mathcal{D}_p + 1$) is greater than the cardinality of $\sigma\left(g_p(t)\right)$ (that is, $\mathcal{E}_p + 1$). Thus, $D_p = E_p$, for $p = 1, \ldots, P$.

Step 3: $\boldsymbol{G}_p = \boldsymbol{H}_p\, e^{i\theta_p}$. From step 2, we know that $\mathcal{D}_p = \mathcal{E}_p$, that is, $\mathfrak{d} = \mathfrak{e}$. Note that

$$\boldsymbol{\Gamma}_{\boldsymbol{s}}[m; \mathfrak{d}] = \begin{bmatrix} \boldsymbol{\Gamma}_{\boldsymbol{s}_1}[m; \mathcal{D}_1] & & & \\ & \boldsymbol{\Gamma}_{\boldsymbol{s}_2}[m; \mathcal{D}_2] & & \\ & & \ddots & \\ & & & \boldsymbol{\Gamma}_{\boldsymbol{s}_P}[m; \mathcal{D}_P] \end{bmatrix}. \tag{A.13}$$

Let $\boldsymbol{\Theta}$ denote the unitary matrix $\boldsymbol{\Theta} = \boldsymbol{Q}^H \boldsymbol{W}$, where $\boldsymbol{Q}$ and $\boldsymbol{W}$ are defined in (A.9) and (A.11), respectively. From (A.10) and (A.12), we have

$$\boldsymbol{\Gamma}_{\boldsymbol{s}}[m; \mathfrak{d}]\, \boldsymbol{\Theta} = \boldsymbol{\Theta}\, \boldsymbol{\Gamma}_{\boldsymbol{s}}[m; \mathfrak{d}], \tag{A.14}$$

for all $m \in \mathbb{Z}$. Partition the matrix $\boldsymbol{\Theta}$ in P^2 submatrices as in

$$\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\Theta}_{11} & \boldsymbol{\Theta}_{12} & \cdots & \boldsymbol{\Theta}_{1P} \\ \boldsymbol{\Theta}_{21} & \boldsymbol{\Theta}_{22} & \cdots & \boldsymbol{\Theta}_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Theta}_{P1} & \boldsymbol{\Theta}_{P2} & \cdots & \boldsymbol{\Theta}_{PP} \end{bmatrix},$$

where $\boldsymbol{\Theta}_{pq} : (\mathcal{D}_p + 1) \times (\mathcal{D}_q + 1)$. From (A.13) and (A.14), we have

$$\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; \mathcal{D}_p] \boldsymbol{\Theta}_{pq} = \boldsymbol{\Theta}_{pq} \boldsymbol{\Gamma}_{\boldsymbol{s}_q}[m; \mathcal{D}_q], \tag{A.15}$$

for all $m \in \mathbb{Z}$ and $p, q = 1, 2, \ldots, P$. Consider $p \neq q$. By assumption A3 in page 27, there is a correlation lag $m(p)$ which, in particular, satisfies $\sigma\left(\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m(p); \mathcal{D}_p]\right) \cap \sigma\left(\boldsymbol{\Gamma}_{\boldsymbol{s}_q}[m(p); \mathcal{D}_q]\right) = \emptyset$. Thus, the identity $\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m(p); \mathcal{D}_p] \boldsymbol{\Theta}_{pq} = \boldsymbol{\Theta}_{pq} \boldsymbol{\Gamma}_{\boldsymbol{s}_q}[m(p); \mathcal{D}_q]$ imply $\boldsymbol{\Theta}_{pq} = \boldsymbol{0}$, see [23, lemma 7.1.5, page 336]. In sum,

$$\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\Theta}_{11} & & & \\ & \boldsymbol{\Theta}_{22} & & \\ & & \ddots & \\ & & & \boldsymbol{\Theta}_{PP} \end{bmatrix},$$

with

$$\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; \mathcal{D}_p] \boldsymbol{\Theta}_{pp} = \boldsymbol{\Theta}_{pp} \boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; \mathcal{D}_p], \quad \text{for all } m \in \mathbb{Z}. \tag{A.16}$$

Notice that each $\boldsymbol{\Theta}_{pp}$ is an unitary matrix.

As mentioned above (see page 121) the matrix $\boldsymbol{\mathcal{T}}_{\mathcal{D}_p}\left(c_p(z)\right)$ is full row rank. Let

$$\boldsymbol{\mathcal{T}}_{\mathcal{D}_p}\left(\boldsymbol{c}_p(z)\right) = \boldsymbol{U}_p \boldsymbol{\Sigma}_p \boldsymbol{V}_p^H \tag{A.17}$$

denote a SVD, where $\boldsymbol{U}_p$ is unitary, $\boldsymbol{\Sigma}_p$ is non-singular (because $\boldsymbol{\mathcal{T}}_{M_p}\left(\boldsymbol{c}_p\right)$ has full row rank), and $\boldsymbol{V}_p$ is an isometry, that is, $\boldsymbol{V}_p^H \boldsymbol{V}_p$ is the identity matrix. Using (A.17) in (2.21) yields

$$\boldsymbol{R}_{\boldsymbol{s}_p}[0; \mathcal{D}_p] = \boldsymbol{U}_p \boldsymbol{\Sigma}_p^2 \boldsymbol{U}_p^H,$$

and, consequently

$$\boldsymbol{R}_{\boldsymbol{s}_p}[0;\mathcal{D}_p]^{-1/2} = \boldsymbol{U}_p\boldsymbol{\Sigma}_p^{-1}\boldsymbol{U}_p^H. \tag{A.18}$$

Using both (A.17) and (A.18) in (2.22) gives

$$\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m;\mathcal{D}_p] = \left(\boldsymbol{V}_p\boldsymbol{U}_p^H\right)^H \boldsymbol{K}_{\mathcal{D}_p+C_p+1}[m]\left(\boldsymbol{V}_p\boldsymbol{U}_p^H\right). \tag{A.19}$$

Since $\boldsymbol{V}_p\boldsymbol{U}_p^H$ is an isometry, equations (A.16) and (A.19) and lemma A.3 imply that

$$\boldsymbol{\Theta}_{pp} = \lambda_p\boldsymbol{I}_{\mathcal{D}_p+1}, \tag{A.20}$$

for some $\lambda_p \in \mathbb{C}$. But, because each $\boldsymbol{\Theta}_{pp}$ is an unitary matrix, we have $\lambda_p = e^{i\theta_p}$, for some $\theta_p \in \mathbb{R}$.

From (A.9) and (A.11), we have

$$\begin{aligned}\boldsymbol{\mathcal{G}}_0 &= \boldsymbol{P}\boldsymbol{W}\\ &= \boldsymbol{\mathcal{H}}_0\boldsymbol{Q}^H\boldsymbol{W}\\ &= \boldsymbol{\mathcal{H}}_0\boldsymbol{\Theta}.\end{aligned}$$

Thus,

$$\begin{aligned}\boldsymbol{\mathcal{G}} &= \boldsymbol{\mathcal{H}}\boldsymbol{R}_{\boldsymbol{s}}[0;\mathfrak{d}]^{1/2}\boldsymbol{\Theta}\boldsymbol{R}_{\boldsymbol{s}}[0;\mathfrak{e}]^{-1/2}\\ &= \boldsymbol{\mathcal{H}}\boldsymbol{\Theta}.\end{aligned}$$

We conclude that $\boldsymbol{\mathcal{G}}_p = \boldsymbol{\mathcal{H}}_p e^{i\theta_p}$, that is, $\boldsymbol{G}_p = \boldsymbol{H}_p e^{i\theta_p}$ □

A.2 Proof of Theorem 2.2

Throughout this proof we let $\boldsymbol{c} = (C_1, C_2, \ldots, C_P)^T$. Recall that $1 \le C_p \le C_{\max}$ for all p. Let $\boldsymbol{c}(z) = (c_1(z), c_2(z), \ldots, c_P(z))^T$ denote a point in the set $\mathbb{M}_{\boldsymbol{c}}[z]$ and let $\epsilon > 0$ be given. We must produce a point $\boldsymbol{d}(z) = (d_1(z), d_2(z), \ldots, d_P(z))^T$ in the set $\mathbb{F}_{\boldsymbol{c}}[z]$ such that $d(\boldsymbol{c}(z), \boldsymbol{d}(z)) < \epsilon$. Before proceeding, we need some definitions. For $\theta \in \mathbb{R}$ and a polynomial $p(z) = \sum_{k=0}^{d} p[k]z^{-k}$, we let $p(z;\theta)$ denote the "rotated" polynomial

$$p(z;\theta) = \sum_{k=0}^{d}\left(p[k]e^{ik\theta}\right)z^{-k}. \tag{A.21}$$

For $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_n)^T \in \mathbb{R}^n$ and a polynomial vector $\boldsymbol{p}(z) = (p_1(z), p_2(z), \ldots, p_n(z))^T$, we let

$$\boldsymbol{p}(z;\boldsymbol{\theta}) = (p_1(z;\theta_1), p_2(z;\theta_2), \ldots, p_n(z;\theta_n))^T.$$

Note that if $p(z)$ denotes a polynomial in $\mathbb{M}_C[z]$ then $p(z;\theta)$ is also a polynomial in $\mathbb{M}_C[z]$ for any θ. Indeed, fix a θ and recall the definition of $\mathbb{M}_C[z]$ in (2.17). It is clear that the polynomial $p(z;\theta)$ is unit-power and has nonzero precursor. To check that it is minimum phase, notice that, from (A.21), we have $p(z;\theta) = p(e^{-i\theta}z)$. Thus, the roots of $p(z;\theta)$ are rotated versions in the complex plane of those of $p(z)$. Since the latter are strictly included in the open disk with radius one and centered at the origin, so are the former. It is clear that this closure property generalizes to polynomial vectors, that is, if $\boldsymbol{p}(z) \in \mathbb{M}_{\boldsymbol{c}}[z]$ then

$\boldsymbol{p}(z;\boldsymbol{\theta}) \in \mathbb{M}_{\boldsymbol{c}}[z]$ for any $\boldsymbol{\theta}$. Finally, for $E_p \in \mathbb{N}$ and $c_p(z) \in \mathbb{M}_{C_p}[z]$, we let $\boldsymbol{R}_{\boldsymbol{s}_p}[m; E_p, c_p(z)]$ denote the correlation matrix of $\boldsymbol{s}_p[n; E_p]$ at lag m, induced by the correlative filter $c_p(z)$. That is,

$$\boldsymbol{R}_{\boldsymbol{s}_p}[m; E_p, c_p(z)] = \boldsymbol{\mathcal{T}}_{E_p}(c_p(z))\ \boldsymbol{K}_{E_p+C_p+1}[m]\, \boldsymbol{\mathcal{T}}_{E_p}(c_p(z))^H, \tag{A.22}$$

see also (2.21). Similarly, we let $\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; E_p, c_p(z)]$ denote the corresponding normalized correlation matrix, that is,

$$\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; E_p, c_p(z)] = \boldsymbol{R}_{\boldsymbol{s}_p}[0; E_p, c_p(z)]^{-1/2} \boldsymbol{R}_{\boldsymbol{s}_p}[m; E_p, c_p(z)] \boldsymbol{R}_{\boldsymbol{s}_p}[0; E_p, c_p(z)]^{-1/2}, \tag{A.23}$$

see (2.22).

Recall that $\boldsymbol{c}(z) = (c_1(z), c_2(z), \ldots, c_P(z))^T \in \mathbb{M}_{\boldsymbol{c}}[z]$. Let $\theta_p \in \mathbb{R}$ and $E_p \in \mathbb{N}$. We will need two fundamental properties in the sequel. The first one is given by

$$\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; E_p, c_p(z; \theta_p)] \sim e^{im\theta_p} \boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; E_p, c_p(z)]. \tag{A.24}$$

For two square matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ of the same size, the notation $\boldsymbol{A} \sim \boldsymbol{B}$ means that they are similar, that is, there exists a nonsingular matrix $\boldsymbol{S}$ such that $\boldsymbol{A} = \boldsymbol{S}\boldsymbol{B}\boldsymbol{S}^{-1}$. Recall that two similar matrices have the same eigenvalues including multiplicities, that is, the same spectrum $\sigma(\boldsymbol{A}) = \sigma(\boldsymbol{B})$. Thus, equation (A.24) asserts that "rotating" the correlative filter $c_p(z)$ by θ_p, makes the spectrum of corresponding normalized correlation matrix of $\boldsymbol{s}_p[n; E_p]$ at lag m rotate by $m\theta_p$. To establish (A.24), we start by noticing that, from (2.20), we have

$$\boldsymbol{\mathcal{T}}_{E_p}(c_p(z; \theta_p)) = \boldsymbol{\Theta}_{E_p+1}(\theta_p)^H \boldsymbol{\mathcal{T}}_{E_p}(c_p(z))\, \boldsymbol{\Theta}_{E_p+C_p+1}(\theta_p), \tag{A.25}$$

where $\boldsymbol{\Theta}_k(\theta) = \operatorname{diag}\left(1, e^{i\theta}, e^{i2\theta}, \ldots, e^{i(k-1)\theta}\right)$. Also, is is easily seen that

$$\boldsymbol{\Theta}_{E_p+C_p+1}(\theta_p) \boldsymbol{K}_{E_p+C_p+1}[m] \boldsymbol{\Theta}_{E_p+C_p+1}(\theta)^H = e^{im\theta} \boldsymbol{K}_{E_p+C_p+1}[m]. \tag{A.26}$$

Plugging (A.25) and (A.26) in (A.22) yields

$$\boldsymbol{R}_{\boldsymbol{s}_p}[m; E_p, c_p(z; \theta)] = e^{im\theta_p}\, \boldsymbol{\Theta}_{E_p+1}(\theta)^H \boldsymbol{R}_{\boldsymbol{s}_p}[m; E_p, c_p(z)] \boldsymbol{\Theta}_{E_p+1}(\theta). \tag{A.27}$$

From (A.27) it follows that

$$\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; E_p, c_p(z; \theta_p)] = \boldsymbol{\Theta}_{E_p+1}(\theta)^H \left(e^{im\theta_p} \boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; E_p, c_p(z)]\right) \boldsymbol{\Theta}_{E_p+1}(\theta),$$

and since $\boldsymbol{\Theta}_{E_p+1}(\theta)$ is the inverse of $\boldsymbol{\Theta}_{E_p+1}(\theta)^H$ we see that (A.24) holds. The other property that we will need is

$$\det\left(\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[C_p; E_p, c_p(z)]\right) \neq 0, \tag{A.28}$$

that is, all eigenvalues of $\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; E_p, c_p(z)]$ are nonzero at the special correlation lag $m = m(p) = C_p$. To establish this, it suffices to prove that $\boldsymbol{R}_{\boldsymbol{s}_p}[C_p; E_p, c_p(z)]$ is nonsingular, see (A.23). Now, we have

$$\boldsymbol{K}_{E_p+C_p+1}[C_p] = \begin{bmatrix} \boldsymbol{0} & \boldsymbol{0} \\ \boldsymbol{I}_{E_p+1} & \boldsymbol{0} \end{bmatrix}. \tag{A.29}$$

Using (A.29) in (A.22) yields

$$\boldsymbol{R}_{\boldsymbol{s}_p}[C_p; E_p, c_p(z)] = \begin{bmatrix} c_p[C_p] & 0 & \cdots & 0 \\ * & c_p[C_p] & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ * & \cdots & * & c_p[C_p] \end{bmatrix} \begin{bmatrix} \overline{c_p[0]} & 0 & \cdots & 0 \\ * & \overline{c_p[0]} & \ddots & \vdots \\ \vdots & \ddots & \ddots & 0 \\ * & \cdots & * & \overline{c_p[0]} \end{bmatrix}. \tag{A.30}$$

Both square lower-triangular matrices in the right-hand side of (A.30) are nonsingular due to the fact that $c_p[C_p] \neq 0$ (because $c_p(z) \in \mathbb{M}_{C_p}(z)$ implies that $\deg c_p(z) = C_p$) and $c_p[0] \neq 0$ (because $c_p(z) \in \mathbb{M}_{C_p}(z)$ implies that $c_p(z)$ has nonzero precursor). Thus, we conclude that (A.28) holds.

We conclude our proof as follows. Notice that the point $\boldsymbol{c}(z;\boldsymbol{\theta})$ varies continuously with $\boldsymbol{\theta}$, that is, "small" changes in $\boldsymbol{\theta}$ induce "small" distances $d(\boldsymbol{c}(z), \boldsymbol{c}(z;\boldsymbol{\theta}))$. Now, as θ_p starts to depart from 0 and does not leave the interval $(-\pi, \pi)$, the spectrum of $\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[C_p; E_p, c_p(z;\theta_p)]$ is rotated in the complex plane due to (A.24). Note that no eigenvalue is kept fixed because all of them are nonzero due to (A.28). Thus, clearly, we can choose $\theta_p \in (-\pi, \pi)$, with $|\theta_p|$ as small as we want, such that the spectrum of $\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[C_p; E_p, c_p(z;\theta_p)]$ does not intersect a given finite set of points in $\mathbb{C}$. Apply this property for all possible E_p in order to satisfy (2.25) and set $\boldsymbol{d}(z) = \boldsymbol{c}(z;\boldsymbol{\theta})$, where it assumed that $\boldsymbol{\theta}$ has been chosen so small that $d(\boldsymbol{c}(z), \boldsymbol{d}(z)) < \epsilon$ holds $\square$

A.3 Proof of Theorem 2.3

Inserting (2.38) in (2.39) yields

$$\boldsymbol{\Gamma}_{\boldsymbol{s}}[m; \mathfrak{d}]\, \boldsymbol{Q}^H \boldsymbol{X} = \boldsymbol{Q}^H \boldsymbol{X}\, \boldsymbol{\Gamma}_{\boldsymbol{s}}[m; \mathfrak{d}], \tag{A.31}$$

where we also used the fact that $\boldsymbol{Q}$ is unitary. Define

$$\boldsymbol{\Theta} = \boldsymbol{Q}^H \boldsymbol{X}, \tag{A.32}$$

and partition the matrix $\boldsymbol{\Theta}$ in P^2 submatrices as in

$$\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\Theta}_{11} & \boldsymbol{\Theta}_{12} & \cdots & \boldsymbol{\Theta}_{1P} \\ \boldsymbol{\Theta}_{21} & \boldsymbol{\Theta}_{22} & \cdots & \boldsymbol{\Theta}_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{\Theta}_{P1} & \boldsymbol{\Theta}_{P2} & \cdots & \boldsymbol{\Theta}_{PP} \end{bmatrix},$$

where $\boldsymbol{\Theta}_{pq} : (\mathcal{D}_p + 1) \times (\mathcal{D}_q + 1)$. Using the fact that

$$\boldsymbol{\Gamma}_{\boldsymbol{s}}[m; \mathfrak{d}] = \begin{bmatrix} \boldsymbol{\Gamma}_{\boldsymbol{s}_1}[m; \mathcal{D}_1] & & & \\ & \boldsymbol{\Gamma}_{\boldsymbol{s}_2}[m; \mathcal{D}_2] & & \\ & & \ddots & \\ & & & \boldsymbol{\Gamma}_{\boldsymbol{s}_P}[m; \mathcal{D}_P] \end{bmatrix}$$

and (A.32), we see that (A.31) is equivalent to

$$\boldsymbol{\Gamma}_{\boldsymbol{s}_p}[m; \mathcal{D}_p]\boldsymbol{\Theta}_{pq} = \boldsymbol{\Theta}_{pq}\boldsymbol{\Gamma}_{\boldsymbol{s}_q}[m; \mathcal{D}_q] \tag{A.33}$$

for all p, q and $m \in \mathbb{Z}$. The identity in (A.33) is precisely the identity in (A.15). Thus, recalling the arguments in step 3 of the proof of theorem 2.1 between (A.15) and (A.20), we have $\boldsymbol{\Theta}_{pq} = \mathbf{0}$ whenever $p \neq q$ and $\boldsymbol{\Theta}_{pp} = \lambda_p \boldsymbol{I}_{\mathcal{D}_p+1}$. Thus, from (A.32), we conclude that $\boldsymbol{X} \overset{\mathfrak{d}}{\sim} \boldsymbol{Q}$.

To establish the second part of theorem 2.3, we start by noticing that (2.40) imply that $\mathfrak{e} = \mathfrak{d}$. To see this, just examine step 2 of the proof of theorem 2.1. Now, using the fact that $\boldsymbol{W}$ is unitary, we have

$$\boldsymbol{\Upsilon}[m]\, \boldsymbol{W} - \boldsymbol{W}\, \boldsymbol{\Gamma}_{\boldsymbol{s}}[m; \mathfrak{d}] = \mathbf{0}$$

for all $m \in \mathbb{Z}$. But, as shown above, this implies that $\boldsymbol{W} \overset{\mathfrak{d}}{\sim} \boldsymbol{Q}$. Since both $\boldsymbol{Q}$ and $\boldsymbol{W}$ are unitary, we necessarily have $\boldsymbol{W} \overset{\mathfrak{d}}{\approx} \boldsymbol{Q}$ □

Appendix B

Proofs for Chapter 3

B.1 Proof of Lemma 3.1

First, some definitions and two auxiliary lemmas. For $r \in \mathbb{C}$ and $\boldsymbol{a} = (a_0, a_1, \ldots, a_n)^T \in \mathbb{C}^{n+1}$, we let $p_{\boldsymbol{a}}(r) = a_0 + a_1 r + a_2 r^2 + \cdots + a_n r^n$. For $\boldsymbol{A} \in \mathbb{C}^{m\times(n+1)}$, we let

$$p_{\boldsymbol{A}}(r) = \begin{bmatrix} p_{\boldsymbol{a}_1}(r) \\ p_{\boldsymbol{a}_2}(r) \\ \vdots \\ p_{\boldsymbol{a}_m}(r) \end{bmatrix},$$

where $\boldsymbol{a}_i^T$ designates the ith row of $\boldsymbol{A}$. Finally, for $\boldsymbol{r} = (r_1, r_2, \ldots, r_d)^T$ and $\boldsymbol{A} \in \mathbb{C}^{m\times(n+1)}$, we write

$$p_{\boldsymbol{A}}(\boldsymbol{r}) = (p_{\boldsymbol{A}}(r_1), p_{\boldsymbol{A}}(r_2), \ldots, p_{\boldsymbol{A}}(r_d)) \in \mathbb{C}^m \times \mathbb{C}^m \times \cdots \times \mathbb{C}^m.$$

Let $n \geq 2$ and $d \geq 0$ denote integers. Recall that $\mathbb{C}_d^n[z] = \{\boldsymbol{f}(z) \in \mathbb{C}^n[z] \,:\, \deg \boldsymbol{f}(z) = d\}$. We identify $\mathbb{C}_d^n[z]$ with a subset of $\mathbb{C}^{n\times(d+1)}$ through the map $\iota : \mathbb{C}_d^n[z] \to \mathbb{C}^{n\times(d+1)}$,

$$\boldsymbol{f}(z) = \boldsymbol{f}[0] + \boldsymbol{f}[1]z^{-1} + \cdots + \boldsymbol{f}[d]z^{-d} \quad \xrightarrow{\iota} \quad \begin{bmatrix} \boldsymbol{f}[0] & \boldsymbol{f}[1] & \cdots & \boldsymbol{f}[d] \end{bmatrix}. \tag{B.1}$$

Within this identification, $\mathbb{C}_d^n[z] \equiv \iota\left(\mathbb{C}_d^n[z]\right)$ consists of those $n \times (d+1)$ complex matrices whose last column is nonzero. Therefore, it is an open subset of $\mathbb{C}^{n\times(d+1)}$. In the sequel, we consider $\mathbb{C}_d^n[z]$ a topological space by equipping it with the subspace topology. Consider the subset

$$\mathbb{I}_d^n[z] = \{\boldsymbol{f}(z) \in \mathbb{C}_d^n[z] \,:\, \boldsymbol{f}(z) \text{ is irreducible }\}.$$

Thus, $\boldsymbol{f}(z) \in \mathbb{I}_n^d[z]$ if and only if $\boldsymbol{f}(z) \neq \boldsymbol{0}$ for all $z \in (\mathbb{C} \setminus \{0\}) \cup \infty$. We have lemma B.1.

Lemma B.1. *The subset $\mathbb{I}_d^n[z]$ is an open subset of $\mathbb{C}_d^n[z]$.*

Proof. Let

$$\boldsymbol{f}_0(z) = (f_1(z), f_2(z), \ldots, f_n(z))^T = \sum_{i=0}^{d} \boldsymbol{f}[i]z^{-i}$$

be a point in $\mathbb{I}_d^n[z]$. We must produce an open subset $\mathbb{O}[z] \subset \mathbb{I}_d^n[z]$ containing $\boldsymbol{f}_0(z)$. Without loss of generality, we assume that $\deg f_1(z) = d$ (some polynomial $f_j(z)$, $1 \leq j \leq d$, must have degree d). Let $\boldsymbol{r}_0 = (r_1, r_2, \ldots, r_d)^T \in \mathbb{C}^d$ denote the d roots of $f_1(z)$, viewed as a polynomial in the indeterminate z^{-1} (that is, $f_1(z) = 0$ if and only if $z^{-1} = r_j$ for

some j) including multiplicities and arranged in some order, and let $\boldsymbol{A}_0 \in \mathbb{C}^{(n-1)\times(d+1)}$ contain the last $n-1$ rows of

$$\boldsymbol{F}_0 = \iota\left(\boldsymbol{f}_0(z)\right) = \left[\begin{array}{cccc} \boldsymbol{f}[0] & \boldsymbol{f}[1] & \cdots & \boldsymbol{f}[d] \end{array}\right] \in \mathbb{C}^{n\times(d+1)}.$$

Let $\mathcal{F} = \mathbb{C}^{n-1} \times \mathbb{C}^{n-1} \times \cdots \times \mathbb{C}^{n-1}$ (d times). Because $\boldsymbol{f}_0(z) \in \mathbb{I}_d^n[z]$, we have

$$p_{\boldsymbol{A}_0}\left(\boldsymbol{r}_0\right) \in \mathcal{S} = \left\{(\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_d) \in \mathcal{F} \,:\, \boldsymbol{z}_1 \neq \boldsymbol{0}, \boldsymbol{z}_2 \neq \boldsymbol{0}, \ldots, \boldsymbol{z}_d \neq \boldsymbol{0}\right\}.$$

Since $\mathcal{S}$ is an open subset of $\mathcal{F}$, and the map $f : \mathbb{C}^d \times \mathbb{C}^{(n-1)\times(d+1)} \to \mathcal{F}$, $f\left(\boldsymbol{r}, \boldsymbol{A}\right) = p_{\boldsymbol{A}}(\boldsymbol{r})$ is continuous, there exist open subsets $W \subset \mathbb{C}^d$ and $V \subset \mathbb{C}^{(n-1)\times(d+1)}$ containing $\boldsymbol{r}_0$ and $\boldsymbol{A}_0$, respectively, such that $f\left(W \times V\right) \subset \mathcal{S}$. Furthermore, let $\boldsymbol{a}_0 = (a_0, a_1, \ldots, a_d)^T \in \mathbb{C}^d$ denote the coefficients of the polynomial $f_1(z)$, that is, $f_1(z) = a_0 + a_1 z^{-1} + \cdots + a_d z^{-d}$. Note that the roots of $f_1(z)$ (the vector $\boldsymbol{r}_0$) are included in W. Since $a_d \neq 0$, by the proposition in [40, page 429] there exists an open subset $U \subset \mathbb{C}^d$ containing $\boldsymbol{a}_0$ such that for all $\boldsymbol{a} \in U$, the polynomial $p_{\boldsymbol{a}}(z)$ has degree d and all roots contained in W (within a suitable arrangement of the roots). Define the open set

$$\mathbb{O} = \left\{\, \boldsymbol{F} \in \mathbb{C}^{n\times(d+1)} \,:\, \boldsymbol{F} = \left[\begin{array}{c} \boldsymbol{a}^T \\ \boldsymbol{A} \end{array}\right],\, \boldsymbol{a} \in U,\, \boldsymbol{A} \in V \,\right\},$$

and let $\mathbb{O}[z] = \iota^{-1}\left(\mathbb{O}\right)$, where the injective map ι is defined in (B.1). Then, $\boldsymbol{f}_0(z) \in \mathbb{O}[z]$, $\mathbb{O}[z]$ is open and $\mathbb{O}[z] \subset \mathbb{I}_d^n[z]$ □

Lemma B.2. *Let $1 \leq m \leq n$ denote integers. Consider the linear map $\Phi : \mathbb{C}^n \to \mathbb{C}$ given by*

$$\Phi\left(\boldsymbol{c}\right) = \det\left[\begin{array}{c} \boldsymbol{E} \\ \boldsymbol{c}^T\boldsymbol{F} \end{array}\right],$$

where $\boldsymbol{E} : (m-1)\times m$ has full row rank and $\boldsymbol{F} : n \times m$ has full column rank. Then, the map Φ is not trivial, that is, the dimension of the linear subspace $\ker\Phi = \{\boldsymbol{c} : \Phi(\boldsymbol{c}) = 0\}$ is strictly less than n.

Proof. Let $\boldsymbol{E}^T = \boldsymbol{Q}\boldsymbol{R}$ denote a truncated QR-decomposition. That is, $\boldsymbol{Q} : m \times (m-1)$ satisfies $\boldsymbol{Q}^H\boldsymbol{Q} = \boldsymbol{I}_{m-1}$ and $\boldsymbol{R} : (m-1)\times(m-1)$ is an upper-triangular matrix with nonzero diagonal entries. Let $\boldsymbol{U} = \left[\,\boldsymbol{Q}\,\boldsymbol{u}\,\right]$ denote an unitary matrix. Then,

$$\begin{aligned}
\Phi(\boldsymbol{c}) = 0 \quad &\Leftrightarrow \quad \det\left[\begin{array}{c} \boldsymbol{E} \\ \boldsymbol{c}^T\boldsymbol{F} \end{array}\right] = 0 \\
&\Leftrightarrow \quad \det\left[\begin{array}{c} \boldsymbol{E} \\ \boldsymbol{c}^T\boldsymbol{F} \end{array}\right]\boldsymbol{U} = 0 \\
&\Leftrightarrow \quad \det\left[\begin{array}{cc} \boldsymbol{R}^T & \boldsymbol{0} \\ * & \boldsymbol{c}^T\boldsymbol{F}\boldsymbol{u} \end{array}\right] = 0 \\
&\Leftrightarrow \quad \boldsymbol{c}^T\boldsymbol{F}\boldsymbol{u} = 0.
\end{aligned}$$

Since $\boldsymbol{u} \neq \boldsymbol{0}$ the vector $\boldsymbol{v} = \boldsymbol{F}\boldsymbol{u}$ is nonzero. Thus, $\ker\Phi = \left\{\boldsymbol{c} : \boldsymbol{c}^T\boldsymbol{v} = 0\right\}$ is a proper subspace of $\mathbb{C}^n$ □

Proof of Lemma 3.1. Throughout the proof, we let $\mathbb{C}^*_{\boldsymbol{d}} = \mathbb{C}^{*Q\times(D_1+1)} \times \mathbb{C}^{*Q\times(D_2+1)} \times \cdots \times \mathbb{C}^{*Q\times(D_P+1)}$, where $\boldsymbol{d} = (D_1, D_2, \ldots, D_P)^T$. We start by proving that $\iota\left(\mathbb{H}_{\boldsymbol{d}}[z]\right)$ is an open

subset of $\mathbb{C}^*_{\boldsymbol{d}}$. Consider the map $\jmath : \mathbb{C}^*_{\boldsymbol{d}} \to \mathbb{C}^{Q\times P}[z]$ which sends $(\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_P) \in \mathbb{C}^*_{\boldsymbol{d}}$ to the polynomial matrix $\boldsymbol{H}(z) = [\,\boldsymbol{h}_1(z)\,\boldsymbol{h}_2(z)\,\cdots\,\boldsymbol{h}_P(z)\,]$ satisfying $\boldsymbol{H}_p = \boldsymbol{\mathcal{T}}_0\,(\boldsymbol{h}_p(z))$. Let $\boldsymbol{H}(z) = [\,\boldsymbol{h}_1(z)\,\boldsymbol{h}_2(z)\,\cdots\,\boldsymbol{h}_P(z)\,]$ denote a point in $\mathbb{H}_{\boldsymbol{d}}[z]$ and let $\boldsymbol{H} = \iota\,(\boldsymbol{H}(z))$. We must show that there exists an open subset $U \subset \mathbb{C}^*_{\boldsymbol{d}}$ containing $\boldsymbol{H}$ such that, for all $\boldsymbol{G} \in U$, the polynomial matrix $\jmath(\boldsymbol{G})$ is column-reduced and irreducible. Since $\jmath(\boldsymbol{H})$ is column-reduced, we have rank $([\,\boldsymbol{h}_1[D_1]\,\boldsymbol{h}_2[D_2]\,\cdots\,\boldsymbol{h}_P[D_P]\,]) = P$. Because $\det(\boldsymbol{A})$ is a continuous function of $\boldsymbol{A}$ and, for $\boldsymbol{B} : Q \times P$, we have $\operatorname{rank}(\boldsymbol{B}) = P$ if and only if $\det\left(\boldsymbol{B}^H\boldsymbol{B}\right) > 0$, we conclude that there exists an open subset $V \subset \mathbb{C}^*_{\boldsymbol{d}}$ containing $\boldsymbol{H}$ such that, for all $\boldsymbol{G} \in V$, $\jmath(\boldsymbol{G})$ is a column-reduced polynomial matrix. Now, consider the map $f : V \to \mathbb{C}^n_d[z]$, where $d = \operatorname{ord}\boldsymbol{H}(z) = D_1 + \cdots + D_P$ and

$$n = \left(\begin{array}{c} Q \\ P \end{array}\right) = \frac{Q!}{P!(Q-P)!},$$

given by

$$f(\boldsymbol{G}) = \left[\begin{array}{c} \det\left([\jmath(\boldsymbol{G})]\,(1,2,\ldots,P)\right) \\ \det\left([\jmath(\boldsymbol{G})]\,(1,2,\ldots,P+1)\right) \\ \vdots \\ \det\left([\jmath(\boldsymbol{G})]\,(Q-P+1,\ldots,Q)\right) \end{array}\right].$$

Here, for a P-tuple of indexes $1 \le i_1 < i_2 < \cdots < i_P \le Q$, the symbol $[\jmath(\boldsymbol{G})](i_1, i_2, \ldots, i_P)$ denotes the $P \times P$ polynomial submatrix of $\jmath(\boldsymbol{G})$ lying in the rows $i_1, i_2, \ldots, i_P$. In words, $f(\boldsymbol{G})$ consists of all $P \times P$ minors of the polynomial matrix $\jmath(\boldsymbol{G})$. Notice that, indeed, $f(V) \subset \mathbb{C}^n_d[z]$, that is, $\deg\jmath(\boldsymbol{G}) = d$ for all $\boldsymbol{G} \in V$, because for each $\boldsymbol{G} \in V$ the polynomial matrix $\jmath(\boldsymbol{G})$ is column-reduced. Furthermore, $\jmath(\boldsymbol{G})$ is irreducible if and only if $f(\boldsymbol{G}) \in \mathbb{I}^n_d[z]$. Since f is clearly continuous, $f(\boldsymbol{H}) \in \mathbb{I}^n_d[z]$, and by lemma B.1 the set $\mathbb{I}^n_d[z]$ is open in $\mathbb{C}^n_d[z]$, we conclude that there exists an open subset $U \subset V$ containing $\boldsymbol{H}$ such that $f(U) \subset \mathbb{I}^n_d[z]$. That is, for all $\boldsymbol{G} \in U$, the polynomial matrix $\jmath(\boldsymbol{G})$ is column-reduced and irreducible. This shows that $\iota\,(\mathbb{H}_{\boldsymbol{d}}[z])$ is an open subset of $\mathbb{C}^*_{\boldsymbol{d}}$.

We now prove that $\iota\,(\mathbb{H}_{\boldsymbol{d}}[z])$ is a dense subset of $\mathbb{C}^*_{\boldsymbol{d}}$. We start by claiming that for given $\boldsymbol{H} \in \mathbb{C}^*_{\boldsymbol{d}}$ and $\epsilon > 0$ there exists a $\boldsymbol{G} \in \mathbb{C}^*_{\boldsymbol{d}}$ such that $\|\boldsymbol{H} - \boldsymbol{G}\| < \epsilon$, $\boldsymbol{G}(z) = \jmath\,(\boldsymbol{G})$ is irreducible, and $\deg g_{qp}(z) = \deg h_{qp}(z)$ if $h_{qp}(z) \neq 0$ and $\deg g_{qp}(z) \le 0$ otherwise. Here, $h_{qp}(z)$ and $g_{qp}(z)$ denote the (q,p)th polynomial entry of $\boldsymbol{H}(z)$ and $\boldsymbol{G}(z)$, respectively. We prove this by induction on P. We start with $P = 1$. Let $\boldsymbol{h}(z) = \jmath\,(\boldsymbol{H})$ and write $\boldsymbol{h}(z) = (h_1(z), h_2(z), \ldots, h_Q(z))^T$. Recall that $Q > P = 1$. We assume that none of the entries of $\boldsymbol{h}(z)$ are zero (otherwise, add a positive constant less than ϵ to a zero entry and we are done because this change makes the perturbed polynomial vector irreducible). Furthermore, we can assume that all polynomial entries of $\boldsymbol{h}(z)$ have degree greater than or equal to 1 (otherwise, $\boldsymbol{h}(z)$ is already irreducible). Now, it is clear that by perturbing the roots of the polynomial $h_2(z)$ (thereby its coefficients) the polynomials $h_1(z)$ and $h_2(z)$ can be made coprime, that is, without any common root. Furthermore, this coprimeness can be achieved with perturbations (of the roots of $h_2(z)$) as small as wanted. Also, changing the roots of $\deg h_2(z)$ does not change its degree. Let $h_2'(z)$ denote such a perturbation of $h_2(z)$ making $\boldsymbol{g}(z) = (h_1(z), h_2'(z), h_3(z), \ldots, h_Q(z))^T$ satisfy $\|\boldsymbol{H} - \boldsymbol{G}\| < \epsilon$, where $\boldsymbol{G} = \boldsymbol{\mathcal{T}}_0\,(\boldsymbol{g}(z))$. Since $h_1(z)$ and $h_2'(z)$ are coprime, $\boldsymbol{g}(z) \neq \boldsymbol{0}$ for all $z \in (\mathbb{C} \setminus \{0\}) \cup \infty$, because $h_1(z)$ and $h_2'(z)$ cannot vanish simultaneously. Thus $\boldsymbol{g}(z)$ is irreducible. Assume

now the hypothesis holds for $P = P'$ and consider the case $P = P' + 1$. Let $\boldsymbol{H} = (\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_P) \in \mathbb{C}^*_{\boldsymbol{d}}$ and $\epsilon > 0$ be given. Let $\boldsymbol{H}(z) = \jmath(\boldsymbol{H})$ and write

$$\boldsymbol{H}(z) = \begin{bmatrix} \boldsymbol{A}(z) \\ \boldsymbol{b}(z)^T \\ \boldsymbol{c}(z)^T \\ \boldsymbol{D}(z) \end{bmatrix},$$

where $\boldsymbol{A}(z)$ has dimensions $(P-1) \times P$, $\boldsymbol{b}(z)^T$ and $\boldsymbol{c}(z)^T$ denote row polynomial vectors and $\boldsymbol{D}(z)$ is a polynomial matrix which is empty if $Q = P+1$. By adding a small positive constant in every zero entry of $\boldsymbol{A}(z)$, if any, we may assume that each entry in $\boldsymbol{A}(z)$ is nonzero. Thus, none of the rows of $\boldsymbol{A}(z)$ is zero and we may apply the hypothesis to the transpose of $\boldsymbol{A}(z)$, that is, $\boldsymbol{A}(z)^T$ (because $\boldsymbol{A}(z)^T$ is a tall matrix with nonzero columns). Thus, by hypothesis, there exists a perturbation $\boldsymbol{A}'(z)^T$ of $\boldsymbol{A}(z)^T$ such that $\boldsymbol{A}'(z)^T$ is irreducible and

$$\boldsymbol{H}'(z) = \begin{bmatrix} \boldsymbol{A}'(z) \\ \boldsymbol{b}^T(z) \\ \boldsymbol{c}^T(z) \\ \boldsymbol{D}(z) \end{bmatrix} = \begin{bmatrix} \boldsymbol{h}'_1(z) & \boldsymbol{h}'_2(z) & \cdots & \boldsymbol{h}'_P(z) \end{bmatrix}$$

satisfies $\|\boldsymbol{H} - \boldsymbol{H}'\| < \epsilon/2$, where $\boldsymbol{H}' = (\boldsymbol{H}'_1, \boldsymbol{H}'_2, \ldots, \boldsymbol{H}'_P)$ with $\boldsymbol{H}'_p = \boldsymbol{\mathcal{T}}_0\left(\boldsymbol{h}'_p(z)\right)$. Also, $\deg h'_{qp}(z) = \deg h_{qp}(z)$ if $h_{qp}(z) \neq 0$ and $\deg h'_{qp}(z) \leq 0$ otherwise. Define two $P \times P$ polynomial submatrices of $\boldsymbol{H}'(z)$ as

$$\boldsymbol{\Delta}_1(z) = \begin{bmatrix} \boldsymbol{A}'(z) \\ \boldsymbol{b}(z)^T \end{bmatrix} \qquad \boldsymbol{\Delta}_2(z) = \begin{bmatrix} \boldsymbol{A}'(z) \\ \boldsymbol{c}(z)^T \end{bmatrix}.$$

If the polynomial $\Delta_1(z) = \det\left(\boldsymbol{\Delta}_1(z)\right)$ does not have roots, then $\boldsymbol{\Delta}_1(z)$ (hence, $\boldsymbol{H}'(z)$) is irreducible and we can take $\boldsymbol{G} = \boldsymbol{H}'$. Otherwise, let $\{z_1, z_2, \ldots, z_r\}$ denote the distinct roots of the polynomial $\Delta_1(z)$, that is, $\Delta_1(z) = 0$ if and only if $z = z_j$ for some j. Without loss of generality, we assume that all entries of $\boldsymbol{c}(z) = (c_1(z), c_2(z), \ldots, c_P)^T$ are nonzero (otherwise, add a small positive constant to each zero entry). Thus, $0 \leq d_p = \deg c_p(z) \leq D_p$. Consider the map $\Phi : \mathbb{C}^d \to \mathbb{C}^r$, $d = (d_1 + 1) + \cdots + (d_P + 1)$, given by

$$\boldsymbol{c} = (c_1[0], \ldots, c_1[d_1]; \ldots; c_P[0], \ldots, c_P[d_P])^T \quad \mapsto \quad \Phi(\boldsymbol{c}) = \begin{bmatrix} \Phi_1(\boldsymbol{c}) \\ \Phi_2(\boldsymbol{c}) \\ \vdots \\ \Phi_d(\boldsymbol{c}) \end{bmatrix},$$

where

$$\Phi_i(\boldsymbol{c}) = \det \begin{bmatrix} & \boldsymbol{A}'(z_i) & \\ c_1[0] + \cdots + c_1[d_1] z_i^{-d_1} & \cdots & c_P[0] + \cdots + c_P[d_P] z_i^{-d_P} \end{bmatrix}.$$

Rewrite the linear map Φ_i as

$$\Phi_i(\boldsymbol{c}) = \det \begin{bmatrix} \boldsymbol{A}'(z_i) \\ \boldsymbol{c}^T \boldsymbol{F}_i \end{bmatrix},$$

where

$$\boldsymbol{F}_i = \begin{bmatrix} 1 & z_i^{-1} & \cdots & z_i^{-d_1} & & & & \\ & & & & \ddots & & & \\ & & & & & 1 & z_i^{-1} & \cdots & z_i^{-d_P} \end{bmatrix}^T.$$

Notice that $\boldsymbol{A}'(z_i)$ has full row rank (because $\boldsymbol{A}'(z)^T$ is irreducible) and $\boldsymbol{F}_i$ has full column rank (by simple inspection). Thus, we find ourselves in the conditions of lemma B.2. We conclude that $\ker \Phi_i$ is a linear subspace of dimension strictly less than d. Thus, $\mathcal{K} = \cup_{i=1}^{r} \ker \Phi_i$ is a nowhere dense subset of $\mathbb{C}^d$. This means that given any $\boldsymbol{c} \in \mathbb{C}^d$ there is another point $\boldsymbol{c}'$ arbitrarily close to $\boldsymbol{c}$ such that $\boldsymbol{c}' \notin \mathcal{K}$. Recall that $\boldsymbol{c}(z) = (c_1(z), c_2(z), \ldots, c_P(z))^T$ and write $c_p(z) = c_p[0] + c_p[1]z^{-1} + \cdots + c_p[d_p]z^{-d_p}$. Define the vector $\boldsymbol{c} = (c_1[0], \ldots, c_1[d_1]; \ldots; c_P[0], \ldots, c_P[d_P])^T$. Choose

$$\boldsymbol{c}' = \big(c_1'[0], \ldots, c_1'[d_1]; \ldots; c_P'[0], \ldots, c_P'[d_P]\big)^T$$

such that $\|\boldsymbol{c} - \boldsymbol{c}'\| < \epsilon/2$, $c_p'[d_p] \neq 0$ and $\boldsymbol{c}' \notin \mathcal{K}$. Let $c_p'(z) = c_p'[0] + c_p'[1]z^{-1} + \cdots + c_p'[d_p]z^{-d_p}$ and $\boldsymbol{c}'(z) = (c_1'(z), c_2'(z), \ldots, c_P'(z))^T$. Notice that we have $\deg c_p'(z) = \deg c_p(z)$. Let

$$\boldsymbol{G}(z) = \begin{bmatrix} \boldsymbol{A}'(z) \\ \boldsymbol{b}(z)^T \\ \boldsymbol{c}'(z)^T \\ \boldsymbol{D}(z) \end{bmatrix} = \begin{bmatrix} \boldsymbol{g}_1(z) & \boldsymbol{g}_2(z) & \cdots & \boldsymbol{g}_P(z) \end{bmatrix}.$$

Then, $\boldsymbol{G}(z)$ is irreducible because the minors $\Delta_1'(z) = \det \boldsymbol{\Delta}_1'(z)$ and $\Delta_2'(z) = \det \boldsymbol{\Delta}_2'(z)$, where

$$\boldsymbol{\Delta}_1'(z) = \begin{bmatrix} \boldsymbol{A}'(z) \\ \boldsymbol{b}(z)^T \end{bmatrix} \qquad \boldsymbol{\Delta}_2'(z) = \begin{bmatrix} \boldsymbol{A}'(z) \\ \boldsymbol{c}'(z)^T \end{bmatrix},$$

do not vanish simultaneously (we made our choices in order to guarantee that whenever $\Delta_1'(z)$ vanishes, $\Delta_2'(z)$ does not). Moreover, let $\boldsymbol{G} = (\boldsymbol{G}_1, \boldsymbol{G}_2, \ldots, \boldsymbol{G}_P)$ with $\boldsymbol{G}_p = \boldsymbol{\mathcal{T}}_0\big(\boldsymbol{g}_p(z)\big)$. Then,

$$\|\boldsymbol{H} - \boldsymbol{G}\| \leq \big\|\boldsymbol{H} - \boldsymbol{H}'\big\| + \big\|\boldsymbol{H}' - \boldsymbol{G}\big\| < \epsilon/2 + \epsilon/2 = \epsilon.$$

This proves the claim. We resume the proof that $\imath\left(\mathbb{H}_{\boldsymbol{d}}[z]\right)$ is a dense subset of $\mathbb{C}_{\boldsymbol{d}}^*$. Let $\boldsymbol{H} = (\boldsymbol{H}_1, \boldsymbol{H}_2, \ldots, \boldsymbol{H}_P) \in \mathbb{C}_{\boldsymbol{d}}^*$ and $\epsilon > 0$ be given. We must show that there exists $\boldsymbol{G}(z) \in \mathbb{H}_{\boldsymbol{d}}[z]$ such that $\|\boldsymbol{H} - \imath\left(\boldsymbol{G}(z)\right)\| < \epsilon$. The polynomial matrix $\boldsymbol{H}(z)$ is column-reduced if

$$\operatorname{rank}\left[\, \boldsymbol{h}_1[D_1]\, \boldsymbol{h}_2[D_2] \cdots \boldsymbol{h}_P[D_P] \,\right] = P, \tag{B.2}$$

where $\boldsymbol{h}_p[D_p]$ stands for the last column of the matrix $\boldsymbol{H}_p$. Thus, if $\boldsymbol{H}(z)$ is not already column-reduced we can perturb the last columns of $\boldsymbol{H}_p$ in order to make (B.2) hold. Furthermore, this perturbation can be made as small as wanted. All said, we can find $\boldsymbol{H}' \in \mathbb{C}_{\boldsymbol{d}}^*$ such that $\|\boldsymbol{H} - \boldsymbol{H}'\| < \epsilon/2$ and $\jmath\left(\boldsymbol{H}'\right)$ is column-reduced. In the first part of this proof, it was shown that the subset of those matrices $\boldsymbol{G} \in \mathbb{C}_{\boldsymbol{d}}^*$ such that $\jmath(\boldsymbol{G})$ is column-reduced, is an open subset of $\in \mathbb{C}_{\boldsymbol{d}}^*$. Thus, there exists an open subset $V \subset \mathbb{C}_{\boldsymbol{d}}^*$ containing $\boldsymbol{H}'$ such that $\jmath(\boldsymbol{G})$ is column-reduced for all $\boldsymbol{G} \in V$. Shrinking V (if necessary) we can further

assume that $\|\boldsymbol{H}' - \boldsymbol{G}\| < \epsilon/2$ for all $\boldsymbol{G} \in V$. By our claim, we can find $\boldsymbol{G} \in V$ such that $\jmath(\boldsymbol{G})$ is irreducible. Since

$$\|\boldsymbol{H} - \boldsymbol{G}\| \leq \left\|\boldsymbol{H} - \boldsymbol{H}'\right\| + \left\|\boldsymbol{H}' - \boldsymbol{G}\right\| < \epsilon/2 + \epsilon/2 = \epsilon,$$

the proof is finished □

B.2 Proof of Lemma 3.2

We will need the auxiliary lemmas B.3 and B.4. They are easy coordinate-free extensions of known results in Euclidean spaces to the context of linear spaces.

Lemma B.3. *Let V be a finite-dimensional vector space equipped with an inner-product denoted $\langle , \rangle$. For $X \in V$, we let $|X| = \sqrt{\langle X, X \rangle}$ denote its norm. Let X_n denote a sequence of random vectors in V. Then $X_n \xrightarrow{P} 0$ if and only if $|X_n| \xrightarrow{P} 0$.*

Proof. Assume $X_n \xrightarrow{P} 0$. Let $F_1, \ldots, F_m$ denote an orthonormal basis for V. Let $\omega_1, \ldots, \omega_m$ denote the corresponding dual basis in V^*, that is, $\omega_i(F_j) = \delta[i - j]$. Thus, $\omega_i = F_i^\flat$, or, equivalently, $\omega_i(X) = \langle X, F_i \rangle$ for any $X \in V$. Notice that

$$|X| = \sqrt{\sum_{i=1}^{m} \omega_i(X)^2}, \tag{B.3}$$

for all $X \in V$. Now, $X_n \xrightarrow{P} 0$ implies, by definition, that $\omega_i(X_n) \xrightarrow{P} 0$ for each $i = 1, \ldots, n$. Since $f(x) = x^2$ is continuous, theorem 2.1.4 in [37, page 51] asserts that $\omega_i(X_n)^2 = f(\omega_i(X_n)) \xrightarrow{P} f(0) = 0$ for each i. By repeated use of theorem 2.1.3 [37, page 50], we have $a_n = \sum_{i=1}^{m} \omega_i(X_n)^2 \xrightarrow{P} 0$. Finally, because $g(x) = \sqrt{x}$ is continuous, theorem 2.1.4 in [37, page 51] shows that $|X_n| = g(a_n) \xrightarrow{P} g(0) = 0$.

Now, for the reverse part, assume that $|X_n| \xrightarrow{P} 0$. Let $\sigma \in V^*$. By definition, we must show that $\sigma(X_n) \xrightarrow{P} 0$. Let $\omega_1, \ldots, \omega_m$ be as above and write $\sigma = c_1\omega_1 + \cdots + c_m\omega_m$ for some constants $c_i \in \mathbb{R}$. From theorem 2.1.3 [37, page 50], it suffices to show that $\omega_i(X_n) \xrightarrow{P} 0$, for each i, that is, for every $\epsilon > 0$, we have $\text{Prob}\{|\omega_i(X_n)| > \epsilon\} \to 0$, as $n \to \infty$. But, from (B.3), we have $|\omega_i(X_n)| \leq |X_n|$. Thus,

$$\text{Prob}\{|\omega_i(X_n)| > \epsilon\} \leq \text{Prob}\{|X_n| > \epsilon\} \to 0 \,\square$$

Lemma B.4. *Let X_n and Y_n denote sequences of random vectors in the finite-dimensional vector space V. Let X denote a random vector in V. If $X_n \xrightarrow{d} X$ and $Y_n - X_n \xrightarrow{P} 0$, then $Y_n \xrightarrow{d} X$.*

Proof. Let $\sigma \in V^*$. By definition, we must show that $\sigma(Y_n) \xrightarrow{d} \sigma(X)$. By hypothesis, we have $\sigma(X_n) \xrightarrow{d} \sigma(X)$ and $\sigma(Y_n) - \sigma(X_n) \xrightarrow{P} 0$. Notice that $\sigma(X_n), \sigma(Y_n)$ and $\sigma(X)$ denote real random variables. Apply [37, corollary 2.3.1, page 70] □

Proof of Lemma 3.2. By symmetry, it suffices to prove either the sufficiency of the necessity part of the lemma. Assume that $F(x_n) \sim a_n - \mathcal{AN}(0, \Sigma)$. We show that

$G(x_n) \sim a_n - \mathcal{AN}(0,\Sigma)$, that is, $a_n G(x_n) \xrightarrow{d} \mathcal{N}(0,\Sigma)$. In view of lemma B.4, it suffices to prove that $a_n\, G(x_n) - a_n\, F(x_n) \xrightarrow{P} 0$. From lemma B.3, this is equivalent to showing that, for given $\epsilon > 0$, we have $\text{Prob}\{|a_n G(x_n) - a_n F(x_n)| > \epsilon\} \to 0$, as $n \to \infty$. Let $\epsilon > 0$ be given. Let $U = B_\lambda(p)$, where $0 < \lambda < \epsilon$, denote a geodesic ball such that $F|_U = G|_U = \text{Exp}_p^{-1}|_U$. We recall that we have $d(p,x) = |\text{Exp}_p^{-1}(x)|$ for all $x \in U$, see [16, theorem 2.92, page 89]. Since $x_n \xrightarrow{P} p$, we have $\text{Prob}\{x_n \notin U\} = \text{Prob}\{d(x_n,p) \geq \lambda\} \to 0$, as $n \to \infty$. Thus, we have $\text{Prob}\{|a_n G(x_n) - a_n F(x_n)| > \epsilon\} \leq \text{Prob}\{x_n \notin U\} \to 0\,\square$

B.3 Proof of Lemma 3.3

We start by establishing the auxiliary lemma B.5.

Lemma B.5. *Let $x_n \sim a_n - \mathcal{AN}(p,\Sigma)$ and let $F : M \to T_pM$ denote a linearization of M at p. Then, $a_n^2\, d(x_n,p)^2 - a_n^2\, |F(x_n)|^2 \xrightarrow{P} 0$.*

Proof. Let $\epsilon > 0$ be given. We must show that $\text{Prob}\left\{|a_n^2|F(x_n)|^2 - a_n^2 d(x_n,p)^2| > \epsilon\right\} \to 0$, as $n \to \infty$. Let $U = B_\lambda(p)$, where $0 < \lambda < \epsilon$, denote a geodesic ball such that $F|_U = \text{Exp}_p^{-1}|_U$. For $x \in U$, we have the equality $d(x,p) = |\text{Exp}_p^{-1}(x)|$, see [16, theorem 2.92, page 89]. Thus,

$$\text{Prob}\left\{|a_n^2|F(x_n)|^2 - a_n^2 d(x_n,p)^2| > \epsilon\right\} \leq \text{Prob}\left\{d(x_n,p) \geq \lambda\right\} \to 0\,\square$$

Proof of Lemma 3.3. Let the inner-product $g_p : T_pM \times T_pM \to \mathbb{R}$ induce the inner-product $g_p^\flat : T_p^*M \times T_p^*M \to \mathbb{R}$, by letting $g_p^\flat(\omega,\sigma) = g_p(\omega^\sharp,\sigma^\sharp)$, recall the discussion in page 62. In the remaning of this proof, we denote $g_p^\flat$ by $\langle\,,\,\rangle$. Since Σ is a symmetric bilinear form on the finite-dimensional vector space T_p^*M, there exists an orthonormal basis $\omega_1,\ldots,\omega_m$ of T_p^*M that diagonalizes it, that is, $\Sigma(\omega_i,\omega_j) = \lambda_i^2\delta[i-j]$ and $\langle\omega_i,\omega_j\rangle = \delta[i-j]$. Notice that for any $X \in T_pM$, we have

$$|X|^2 = \sum_{i=1}^m \omega_i(X)^2. \tag{B.4}$$

Moreover,

$$\text{tr}\,\Sigma = \sum_{i=1}^m \Sigma(\omega_i,\omega_i) = \sum_{i=1}^m \lambda_i^2. \tag{B.5}$$

Define the random vector $\boldsymbol{x}_n \in \mathbb{R}^m$ by $\boldsymbol{x}_n = a_n\left(\omega_1\left(F(x_n)\right),\omega_2\left(F(X_n)\right),\ldots,\omega_m\left(F(x_n)\right)\right)^T$. We claim that $\boldsymbol{x}_n \xrightarrow{d} \mathcal{N}(\boldsymbol{0},\boldsymbol{\Lambda})$, where $\boldsymbol{\Lambda} = \text{diag}\left(\lambda_1^2,\lambda_2^2,\ldots,\lambda_m^2\right)$, that is, $\boldsymbol{t}^T\boldsymbol{x} \xrightarrow{d} \mathcal{N}(0,\boldsymbol{t}^T\boldsymbol{\Lambda}\boldsymbol{t})$, for all $\boldsymbol{t} \in \mathbb{R}^m$, see [37, theorem 5.1.8, page 284]. Let $\boldsymbol{t} = (t_1,t_2,\ldots,t_m)^T \in \mathbb{R}^m$ be given and define the covector $\omega = t_1\omega_1 + t_2\omega_2 + \cdots + t_m\omega_m$. Then,

$$\boldsymbol{t}^T\boldsymbol{x}_n = \omega\left(a_n F(x_n)\right) \xrightarrow{d} \mathcal{N}\left(0,\Sigma(\omega,\omega)\right) = \mathcal{N}\left(0,\boldsymbol{t}^T\boldsymbol{\Lambda}\boldsymbol{t}\right).$$

In the sequel, we let $\boldsymbol{z} \overset{d}{=} \mathcal{N}(\boldsymbol{0},\boldsymbol{\Lambda})$. Note that, from (B.4),

$$\|\boldsymbol{x}_n\|^2 = a_n^2 \sum_{i=1}^m \omega_i\left(F(x_n)\right)^2 = a_n^2|F(x_n)|^2.$$

Since the function $f : \mathbb{R}^m \to \mathbb{R}$, $f(\boldsymbol{x}) = \|\boldsymbol{x}\|^2$, is continuous, and $\boldsymbol{x}_n \xrightarrow{d} \boldsymbol{z}$, we have

$$a_n^2 \, |F(x_n)|^2 = \|\boldsymbol{x}_n\|^2 = f(\boldsymbol{x}_n) \xrightarrow{d} f(\boldsymbol{z}) = \|\boldsymbol{z}\|^2 , \tag{B.6}$$

by [37, theorem 5.1.5, page 281]. Letting $z \stackrel{d}{=} \|\boldsymbol{z}\|^2$, we see, from (B.5), that $\mathrm{E}\{z\} = \mathrm{tr}(\boldsymbol{\Lambda}) = \mathrm{tr}\,\Sigma$. Now, from lemma B.5, we have $a_n^2 \, d(x_n, p)^2 - a_n^2 \, |F(x_n)|^2 \xrightarrow{P} 0$. Thus, from (B.6) and [37, corollary 2.3.1, page 70], we conclude that $a_n^2 \, d(x_n, p)^2 \xrightarrow{d} z \, \square$

B.4 Proof of Lemma 3.4

Lemma B.6. *Let V denote a finite-dimensional vector space equipped with an inner-product $\langle , \rangle$. As usual, we let $|Y| = \sqrt{\langle Y, Y \rangle}$. If $X_n \xrightarrow{d} X$, where X denotes some random vector, then the sequence X_n is bounded in probability. That is, for every $\epsilon > 0$, there exists a constant $C > 0$ and an integer N such that $n \geq N$ implies $\mathrm{Prob}\{|X_n| < C\} > 1 - \epsilon$.*

Proof. Let $F_1, \ldots, F_m$ denote an orthonormal basis for V. Let $\omega_1, \ldots, \omega_m$ denote the corresponding dual basis in V^*. Thus, for any $Y \in V$, we have $\langle Y, F_i \rangle = \omega_i(Y)$, and, as a consequence,

$$|Y|^2 = \sum_{i=1}^{m} \omega_i(Y)^2.$$

Let $\epsilon > 0$ be given. By hypothesis, $\omega_i(X_n) \xrightarrow{d} \omega_i(X)$, for each i. By invoking [37, theorem 2.3.2, page 67] on the sequence of real random variables $\omega_i(X_n)$, we see that there exists an integer N_i and a constant $C_i > 0$ such that $n \geq N_i$ implies $\mathrm{Prob}\,\{|\omega_i(X_n)| < C_i\} > 1 - \epsilon/m$. Let $N = \max\{N_1, \ldots, N_m\}$ and $C = \sqrt{C_1^2 + \cdots + C_m^2}$. Thus, $n \geq N$ implies

$$\begin{aligned} \mathrm{Prob}\,\{|X_n| < C\} &= \mathrm{Prob}\left\{|X_n|^2 < C^2\right\} \\ &\geq \mathrm{Prob}\left\{|\omega_1(X_n)|^2 < C_1^2, \ldots, |\omega_m(X_n)|^2 < C_m^2\right\} \\ &\geq 1 - \epsilon \, \square \end{aligned}$$

Lemma B.7. *Let V denote a finite-dimensional vector space. Let X_n denote a sequence of random vectors in V. Assume that $X_n \sim a_n - \mathcal{AN}(0, \Sigma)$, where $a_n \to +\infty$ and $\Sigma \in T_2(V)$. Then, $X_n \xrightarrow{P} 0$.*

Proof. Let $\sigma \in V^*$. By definition, we must show that $\sigma(X_n) \xrightarrow{P} 0$. Notice that $\sigma(X_n)$ denotes a sequence of real random variables. By hypothesis, $a_n X_n \xrightarrow{d} \mathcal{N}(0, \Sigma)$. Thus, $a_n \sigma(X_n) \xrightarrow{d} \mathcal{N}(0, \Sigma(\sigma, \sigma))$. Now, apply [37, theorem 2.3.4, page 70] $\square$

Proof of Lemma 3.4. By definition, we must show that **i)** the sequence $F(x_n)$ converges in probability to $F(p)$, and **ii)** for any linearization $H : N \to T_{F(p)}N$ we have $H(F(x_n)) \sim a_n - \mathcal{AN}(0, F_* \Sigma)$.

i) To prove that $F(x_n) \xrightarrow{P} F(p)$ we must show that, for any given $\lambda > 0$, we have $\mathrm{Prob}\,\{d(F(x_n), F(p)) > \lambda\} \to 0$ as $n \to \infty$. Let $\lambda > 0$ be given and choose geodesic balls $U = B_\delta(p) \subset M$ and $V = B_\epsilon(F(p)) \subset N$ centered at p and $F(p)$, respectively, such that

$0 < \epsilon < \lambda$ and $F(U) \subset V$. Since, by hypothesis, $x_n \xrightarrow{P} 0$, we have $\text{Prob}\{d(x_n, p) \geq \delta\} \to 0$. Thus,

$$\text{Prob}\{d(F(x_n), F(p)) > \lambda\} \leq \text{Prob}\{x_n \notin U\} = \text{Prob}\{d(x_n, p) \geq \delta\} \to 0.$$

ii) Let $G : M \to T_pM$ and $H : N \to T_{F(p)}N$ denote linearizations of M (at p) and N (at $F(p)$), respectively. We start by showing that

$$a_n F_*(G(x_n)) \xrightarrow{d} \mathcal{N}(0, F_*\Sigma), \tag{B.7}$$

where $F_* : T_pM \to T_{F(p)}N$ denote the push forward linear (derivative) mapping induced by the smooth map $F : M \to N$. To prove this, we must show that $a_n\omega(F_*(G(x_n))) \xrightarrow{d} \mathcal{N}(0, (F_*\Sigma)(\omega, \omega))$, for any given covector ω in the dual space $T^*_{F(p)}N$. By hypothesis, for any $\sigma \in T^*_pM$, we have

$$a_n\sigma(G(x_n)) \xrightarrow{d} \mathcal{N}(0, \Sigma(\sigma, \sigma)). \tag{B.8}$$

Recall that $F^* : T^*_{F(p)}N \to T^*_pM$ denotes the pull back map induced by the linear map $F_* : T_pM \to T_{F(p)}N$, see (3.25) in page 62. Taking $\sigma = F^*\omega$ in (B.8) we conclude that

$$a_n\omega(F_*(G(x_n))) = a_n(F^*\omega)(G(x_n)) \xrightarrow{d} \mathcal{N}(0, \Sigma(F^*\omega, F^*\omega)) = \mathcal{N}(0, (F_*\Sigma)(\omega, \omega)),$$

where the equality $\Sigma(F^*\omega, F^*\omega) = (F_*\Sigma)(\omega, \omega)$ follows from the definition of the push forward map $F_* : T_2(T_pM) \to T_2\left(T_{F(p)}N\right)$, see (3.26) in page 62. Thus, (B.7) holds.

Our next step consists in proving that

$$a_n H(F(x_n)) - a_n F_*(G(x_n)) \xrightarrow{P} 0. \tag{B.9}$$

According to lemma B.3, we must show that, for any given $\alpha, \beta > 0$, there exists an integer N such that

$$n \geq N \quad \Rightarrow \quad \text{Prob}\{|a_n H(F(x_n)) - a_n F_*(G(x_n))| < \alpha\} > 1 - \beta. \tag{B.10}$$

Let $\alpha, \beta > 0$ be given. Since $a_n G(x_n) \xrightarrow{d} \mathcal{N}(0, \Sigma)$, lemma B.6 asserts the existence of an integer N_1 and a constant $C > 0$ such that

$$n \geq N_1 \quad \Rightarrow \quad \text{Prob}\{|a_n G(x_n)| < C\} > 1 - \frac{\beta}{2}. \tag{B.11}$$

Choose geodesic balls $U = B_\delta(p) \subset M$ and $V = B_\epsilon(F(p)) \subset N$ such that

$$F(U) \subset V, \quad \text{Exp}_p^{-1}|_U = G|_U, \quad \text{Exp}_{F(p)}^{-1}|_V = H|_V. \tag{B.12}$$

Expressing the smooth map $F : M \to N$ with respect to normal coordinates [7, theorem 6.6, page 339] centered at p and $F(p)$ and supported on U and V, respectively, it is easily seen that

$$\text{Exp}_{F(p)}^{-1}\left(F\left(\text{Exp}_p(X_p)\right)\right) - F_*(X_p) = o(|X_p|), \tag{B.13}$$

for $X_p \in TB_\delta(F(p)) = \{X_p \in T_pM : |X_p| < \delta\}$. Thus, there exists $0 < \lambda < \delta$ such that

$$|X_p| < \lambda \quad \Rightarrow \quad |\text{Exp}_{F(p)}^{-1}\left(F\left(\text{Exp}_p(X_p)\right)\right) - F_*(X_p)| < \frac{\alpha}{C}|X_p|. \tag{B.14}$$

By hypothesis, $G(x_n) \sim a_n - \mathcal{AN}(0, \Sigma)$. Thus, from lemmas B.7 and B.3, we see that $|G(x_n)| \xrightarrow{P} 0$. Choose N_2 such that

$$n \geq N_2 \quad \Rightarrow \quad \text{Prob}\{|G(x_n)| < \lambda\} > 1 - \frac{\beta}{2}. \tag{B.15}$$

From (B.11) and (B.15) we conclude that

$$n \geq N = \max\{N_1, N_2\} \quad \Rightarrow \quad \text{Prob}\{|a_n G(x_n)| < C, |G(x_n)| < \lambda\} > 1 - \beta.$$

Moreover, if $|a_n G(x_n)| < C$ and $|G(x_n)| < \lambda$ we have

$$\begin{aligned} |a_n H(F(x_n)) - a_n F_*(G(x_n))| &= a_n |H(F(x_n)) - F_*(G(x_n))| \\ &< a_n \frac{\alpha}{C} |G(x_n)| \\ &< \alpha. \end{aligned} \tag{B.16}$$

In (B.16), we used the properties in (B.14) and (B.12). Thus, (B.10) holds.

From (B.7), (B.9), and lemma B.4, we conclude that $H(F(x_n)) \sim a_n - \mathcal{AN}(0, F_*\Sigma)$ □

B.5 Proof of Lemma 3.5

Since a piecewise regular curve is a finite concatenation of smooth segments, we may assume without loss of generality that the curve $q : I \to M/G$, where $I = [a, b]$ is smooth.

We start by establishing the uniqueness of the curve $p(t)$. That is, assume the existence of two smooth curves $p_1 : I \to M$ and $p_2 : I \to M$ such that $p_i(a) = x$, $\dot{p}_i(t) \in H_{p_i(t)}M$ and $\rho(p_i(t)) = q(t)$ for all $t \in I$, and $i = 1, 2$. Let $A = \{t \in I : p_1(t) = p_2(t)\}$. Our goal is to prove that $A = I$. We show this by proving that A is open and closed in I. This imply that A is a connected component of I (note that A is non-empty because $a \in A$). Thus, $A = I$. We start by showing that A is open. Let $m = \dim M$ and $n = \dim N$. Let $t_0 \in A$. Choose (cubical) coordinates U, φ of $p_1(t_0) = p_2(t_0)$ and V, ψ of $q(t_0) = \rho(p_1(t_0)) = \rho(p_2(t_0))$ such that $\varphi(U) = C_\epsilon^m(0) = \{(x_1, \ldots, x_m) : |x_i| < \epsilon\}$, $V = C_\epsilon^n(0)$, $\varphi(p_i(t_0)) = (0, 0, \ldots, 0)$, $\psi(q(t_0)) = (0, 0, \ldots, 0)$, and the map $\rho|_U$ is given in these local coordinates by $\widehat{\rho}(x_1, x_2, \ldots, x_m) = (x_1, x_2, \ldots, x_n)$.

Let $\widehat{q}(t) = \psi(q(t)) = (\widehat{q}_1(t), \ldots, \widehat{q}_n(t))$ denote the curve $q(t)$ with respect to the coordinates V, ψ. Similarly, let $\widehat{p}_i(t) = \varphi(p_i(t))$ and write $\widehat{p}_i(t) = (a_i(t), b_i(t))$, where $a_i(t) = (a_{i1}(t), \ldots, a_{in}(t))$ and $b_i(t) = \big(b_{i1}(t), \ldots, b_{i(m-n)}(t)\big)$. Note that $\widehat{p}_i(t)$ and $\widehat{q}(t)$ are defined on an open subset of I containing t_0. Let

$$g(y) = g(y_1, \ldots, y_n) = \begin{bmatrix} g_{11}(y) & g_{12}(y) \\ g_{21}(y) & g_{22}(y) \end{bmatrix},$$

where $y \in \varphi(U)$, denote the Riemannian tensor in the coordinates U, φ. Here, $g_{11} : n \times n$ and $g_{22} : (m-n) \times (m-n)$. Thus, we are abandoning temporarily (within this proof) our notation style. We do not write matrices or vectors in boldface type and adhere to the notation in [7]. By hypothesis, we have $\rho(p_i(t)) = q(t)$. Thus, $a_{ij}(t) = \widehat{q}_j(t)$, for $j = 1, \ldots, n$. Also, by hypothesis, $\dot{p}_i(t) \in H_{p_i(t)}M$. Thus, $g_{21}(\widehat{p}(t))\dot{a}_i(t) + g_{22}(\widehat{p}_i(t))\dot{b}_i(t) = 0$, that is, $\dot{b}_i(t) = -g_{22}(a_i(t), b_i(t))^{-1} g_{21}(a_i(t), b_i(t))\dot{a}_i(t)$. Using $a_i(t) = \widehat{q}(t)$, leads to

$$\dot{b}_i(t) = -g_{22}(\widehat{q}(t), b_i(t))^{-1} g_{21}(\widehat{q}(t), b_i(t))\dot{\widehat{q}}(t). \tag{B.17}$$

Thus, $b_1(t)$ and $b_2(t)$ must agree within an subset of I containing t_0, because they are both determined by the solution of the same system of ordinary differential equations (B.17) and $b_1(t_0) = b_2(t_0)$ by hypothesis (see the invoked uniqueness result for ordinary differential equations in [7, theorem 4.1, page 131], for exampe). Thus, A is open. Since the defining conditions of A are continuous, A is also closed.

We now establish the existence of $p(t)$. First, we see that two horizontal curves in M which overlap in M/G can be "glued" together to produce an extended horizontal curve. More precisely, let $p_1 : [a_1, b_1] \to M$ and $p_2 : [a_2, b_2] \to M$ denote two smooth curves such that $a_1 < a_2 < b_1 < b_2$ and $\rho(p_1(a_2)) = \rho(p_2(a_2))$, that is, $p_1(a_2)$ and $p_2(a_2)$ are in the same orbit. Let g be the unique element in the group G satisfying $p_2(a_2) \cdot g = p_1(a_2)$. Since the map $\theta_g : M \to M$, $x \mapsto x \cdot g$, is an isometry, the curve $p_2' : [a_2, b_2] \to M$ given by $p_2'(t) = \theta_g(p_2(t))$ is also horizontal. By the uniquess result above, we see that $p_2'(t) = p_1(t)$ for all $t \in [a_2, b_1]$. All said, the curves $p_1(t)$ and $p_2(t)$ can be glued to obtain the "bigger" horizontal curve $p : [a_1, b_2] \to M$ where $p(t) = p_1(t)$ if $t \in [a_1, b_1]$ and $p(t) = p_2'(t)$ if $t \in [a_2, b_2]$. Now, for each $s \in I$, let $p_s : I_s = (a_s, b_s) \to M$ denote any horizontal curve satisfying $\rho(p_s(t)) = q(t)$ for $t \in I_s$. Notice that, for each s, the curve p_s exists thanks to the the existence of local solutions to the system of ordinary differential equations in (B.17). The collection of open intervals $\{I_s : s \in I\}$ cover the compact set I. Thus, there exists a finite subcover, say, $\{I_{s_1}, \ldots, I_{s_k}\}$. Now, glue the corresponding curves $p_{s_1}, \ldots, p_{s_k}$ $\square$

B.6 Proof of Lemma 3.6

We prove the existence of U, V, the implicit mapping $\boldsymbol{Z} \in U \mapsto (\lambda(\boldsymbol{Z}), \boldsymbol{q}(\boldsymbol{Z})) \in V$ and its differentiability, by using a re-statement of the well-known Implicit Function Theorem in the context of differentiable manifolds.

Fact. *Let A, B and C denote smooth manifolds, such that $\dim B = \dim C$. Let $F : A \times B \to C$ denote a smooth mapping. For $a \in A$, we define $F_a : B \to C$ as $F_a(b) = F(a, b)$. Let $(a_0, b_0) \in A \times B$ and define $c_0 = F(a_0, b_0)$. If $F_{a_0 *} : T_{b_0}(B) \to T_{c_0}(C)$ is a linear isomorphism then there exist open sets $A_0 \subset A$ and $B_0 \subset B$ containing a_0 and b_0, respectively, such that, for each $a \in A_0$ there exists one and only one $b \in B_0$ satisfying $F(a, b) = c_0$. Moreover, the mapping $a \in A_0 \mapsto b(a) \in B_0$ is smooth.*

The proof of this fact is omitted since it is trivial (invoke local coordinates and use the classical version of the Implicit Function Theorem). In our case, we let $A = \mathbb{C}^{n \times n}$, $B = \mathbb{C} \times \mathbb{S}_{\mathbb{C}}^{n-1}$, $C = \mathbb{C}^n \times \mathbb{R}$, and define the smooth mapping $F : A \times B \to C$, $F(\boldsymbol{Z}; \lambda, \boldsymbol{q}) = \left(\boldsymbol{Z}\boldsymbol{q} - \lambda \boldsymbol{q}, \operatorname{Im} \boldsymbol{c}_0^H \boldsymbol{q}\right)$. Note that $\dim B = \dim \mathbb{C} + \dim \mathbb{S}_{\mathbb{C}}^{n-1} = 2 + (2n-1) = 2n+1$ equals $\dim C = \dim \mathbb{C}^n + \dim \mathbb{R} = 2n + 1$, as required. Moreover, $a_0 = \boldsymbol{Z}_0$, $b_0 = (\lambda_0, \boldsymbol{q}_0)$, and $c_0 = F(a_0, b_0) = \boldsymbol{0}$. For the tangent space to B at b_0, we have the identification

$$T_{(\lambda_0, \boldsymbol{q}_0)}\left(\mathbb{C} \times \mathbb{S}_{\mathbb{C}}^{n-1}\right) \simeq \left\{(\Delta, \boldsymbol{\delta}) \in \mathbb{C} \times \mathbb{C}^n : \operatorname{Re} \boldsymbol{q}_0^H \boldsymbol{\delta} = 0\right\}. \tag{B.18}$$

With this identification, the linear mapping $F_{a_0 *} : T_{b_0}(B) \to T_{c_0}(C)$ is given by

$$F_{\boldsymbol{Z}_0 *}(\Delta, \boldsymbol{\delta}) = \left(\boldsymbol{Z}_0 \boldsymbol{\delta} - \lambda_0 \boldsymbol{\delta} - \Delta \boldsymbol{q}_0, \operatorname{Im} \boldsymbol{c}_0^H \boldsymbol{\delta}\right).$$

We must establish the injectivity of $F_{\boldsymbol{Z}_0*}$. Suppose $F_{\boldsymbol{Z}_0*}(\Delta, \boldsymbol{\delta}) = (\boldsymbol{0}, 0)$, that is,

$$(\boldsymbol{Z}_0 - \lambda_0 \boldsymbol{I}_n)\,\boldsymbol{\delta} = \Delta \boldsymbol{q}_0 \tag{B.19}$$

$$\operatorname{Im} \boldsymbol{c}_0^H \boldsymbol{\delta} = 0. \tag{B.20}$$

Multiplying both sides of (B.19) by $\boldsymbol{q}_0^H$ (on the left), and using the facts $\boldsymbol{q}_0^H \boldsymbol{Z}_0 = \lambda_0 \boldsymbol{q}_0^H$ and $\boldsymbol{q}_0^H \boldsymbol{q}_0 = 1$, yields $\Delta = 0$. Thus,

$$(\boldsymbol{Z}_0 - \lambda_0 \boldsymbol{I}_n)\,\boldsymbol{\delta} = \boldsymbol{0}. \tag{B.21}$$

Now, the rows of the Hermitean matrix $\boldsymbol{Z}_0 - \lambda_0 \boldsymbol{I}_n$ span the orthogonal complement of $\boldsymbol{q}_0$ in $\mathbb{C}^n$. To see this, let $\boldsymbol{Z}_0 = \boldsymbol{Q}\,\boldsymbol{\Lambda}\,\boldsymbol{Q}^H$ denote an EVD of $\boldsymbol{Z}_0$, where $\boldsymbol{Q} = \left[\, \boldsymbol{q}_0\, \boldsymbol{q}_1\, \cdots\, \boldsymbol{q}_{n-1} \,\right]$ is unitary, and $\boldsymbol{\Lambda} = \operatorname{diag}(\lambda_0, \lambda_1, \ldots, \lambda_{n-1})$ contains the eigenvalues. Then,

$$\boldsymbol{Z}_0 - \lambda_0 \boldsymbol{I}_n = \boldsymbol{Q}_1\,(\boldsymbol{\Lambda}_1 - \lambda_0 \boldsymbol{I}_{n-1})\,\boldsymbol{Q}_1^H, \tag{B.22}$$

where

$$\boldsymbol{Q}_1 = \left[\; \boldsymbol{q}_1 \quad \cdots \quad \boldsymbol{q}_{n-1} \;\right] \tag{B.23}$$

spans the orthogonal complement of $\boldsymbol{q}_0$ and $\boldsymbol{\Lambda}_1 = \operatorname{diag}(\lambda_1, \ldots, \lambda_{n-1})$. Since λ_0 is a simple eigenvalue, $\boldsymbol{\Lambda}_1 - \lambda_0 \boldsymbol{I}_{n-1}$ is non-singular, and the row space of $\boldsymbol{Z}_0 - \lambda_0 \boldsymbol{I}_n$ is identical to the column space of $\boldsymbol{Q}_1$. With this fact in mind, (B.21) implies that $\boldsymbol{\delta}$ is colinear with $\boldsymbol{q}_0$, that is, there exists $\alpha \in \mathbb{C}$ such that

$$\boldsymbol{\delta} = \boldsymbol{q}_0 \alpha. \tag{B.24}$$

Multiplying both sides of (B.24) by $\boldsymbol{c}_0^H$ (on the left) and recalling that $\boldsymbol{c}_0^H \boldsymbol{q}_0 \in \mathbb{R} - \{0\}$ (by hypothesis) and $\boldsymbol{c}_0^H \boldsymbol{\delta} \in \mathbb{R}$ (by (B.20)), we have $\alpha = \boldsymbol{c}_0^H \boldsymbol{\delta} / \boldsymbol{c}_0^H \boldsymbol{q}_0 \in \mathbb{R}$. Multiplying both sides of (B.24) by $\boldsymbol{q}_0^H$ (on the left) yields

$$\alpha = \operatorname{Re} \alpha = \operatorname{Re} \boldsymbol{q}_0^H \boldsymbol{\delta} = 0,$$

where the last equality follows from (B.18). Thus, $\boldsymbol{\delta} = \boldsymbol{0}$, and the mapping $F_{\boldsymbol{Z}_0*}$ is injective. This establishes the existence and differentiability of the mapping $\boldsymbol{Z} \in U \mapsto (\lambda(\boldsymbol{Z}), \boldsymbol{q}(\boldsymbol{Z})) \in V$. Notice that since, in particular, $\boldsymbol{q}(\boldsymbol{Z})$ is continuous and $\boldsymbol{q}(\boldsymbol{Z}_0) = \boldsymbol{q}_0$, we may now restrict U (if necessary) in order to satisfy $\operatorname{Re} \boldsymbol{c}_0^H \boldsymbol{q}(\boldsymbol{Z}) > 0$, for all $\boldsymbol{Z} \in U$. Also, since the eigenvalues of a matrix are continuous functions of its entries, we may restrict U further (if necessary) to guarantee that $\lambda(\boldsymbol{Z})$ is a simple eigenvalue of $\boldsymbol{Z}$, for all $\boldsymbol{Z} \in U$.

The differentiabiliy of the mappings being proved, we now compute the differentials $d\lambda$ and $d\boldsymbol{q}$ at $\boldsymbol{Z}_0$. From $\boldsymbol{Z}\boldsymbol{q} = \lambda \boldsymbol{q}$, we have (evaluating at $\boldsymbol{Z}_0$)

$$d\boldsymbol{Z}\,\boldsymbol{q}_0 + \boldsymbol{Z}_0 d\boldsymbol{q} = d\lambda\,\boldsymbol{q}_0 + \lambda_0 d\boldsymbol{q}. \tag{B.25}$$

Multiplying both sides of (B.25) by $\boldsymbol{q}_0^H$ (on the left), and using the facts $\boldsymbol{q}_0^H \boldsymbol{Z}_0 = \lambda_0 \boldsymbol{q}_0^H$ and $\boldsymbol{q}_0^H \boldsymbol{q}_0 = 1$, yields

$$d\lambda = \boldsymbol{q}_0^H d\boldsymbol{Z}\,\boldsymbol{q}_0. \tag{B.26}$$

Since $d\boldsymbol{q} \in \mathbb{C}^n$, we can write (uniquely)

$$d\boldsymbol{q} = \boldsymbol{q}_0 \alpha + \boldsymbol{v}, \tag{B.27}$$

for some $\alpha \in \mathbb{C}$ and vector $\boldsymbol{v} \in \mathbb{C}^n$ orthogonal to $\boldsymbol{q}_0$ ($\boldsymbol{q}_0^H \boldsymbol{v} = 0$). Note that $\operatorname{Re} \boldsymbol{q}_0^H d\boldsymbol{q} = 0$, because

$$d\boldsymbol{q} \in T_{\boldsymbol{q}_0}\left(\mathbb{S}_{\mathbb{C}}^{n-1}\right) \simeq \left\{\boldsymbol{\delta} \in \mathbb{C}^n \,:\, \operatorname{Re} \boldsymbol{q}_0^H \boldsymbol{\delta} = 0\right\}.$$

Since, from (B.27), we have $\alpha = \boldsymbol{q}_0^H d\boldsymbol{q}$, it follows that α is a pure imaginary number. From $\operatorname{Im} \boldsymbol{c}_0^H \boldsymbol{q} = 0$, we have

$$\operatorname{Im} \boldsymbol{c}_0^H d\boldsymbol{q} = 0, \tag{B.28}$$

and using (B.27) in (B.28) yields

$$\operatorname{Im} \boldsymbol{c}_0^H d\boldsymbol{q} = \operatorname{Im}\left\{\boldsymbol{c}_0^H \boldsymbol{q}_0\, \alpha + \boldsymbol{c}_0^H \boldsymbol{v}\right\} = -i\,(\boldsymbol{c}_0^H \boldsymbol{q}_0)\alpha + \operatorname{Im} \boldsymbol{c}_0^H \boldsymbol{v} = 0,$$

where we used the fact that $\boldsymbol{c}_0^H \boldsymbol{q}_0$ denotes a (nonzero) real number. Thus,

$$\alpha = -i\,\frac{\operatorname{Im} \boldsymbol{c}_0^H \boldsymbol{v}}{\boldsymbol{c}_0^H \boldsymbol{q}_0}. \tag{B.29}$$

We now find a suitable expression for $\boldsymbol{v}$. Plugging (B.26) in (B.25) and rearranging yields

$$(\lambda_0 \boldsymbol{I}_n - \boldsymbol{Z}_0)\, d\boldsymbol{q} + \left(\boldsymbol{q}_0^H d\boldsymbol{Z} \boldsymbol{q}_0\right) \boldsymbol{q}_0 = d\boldsymbol{Z} \boldsymbol{q}_0. \tag{B.30}$$

Using (B.27) in (B.30) provides

$$(\lambda_0 \boldsymbol{I}_n - \boldsymbol{Z}_0)\, \boldsymbol{v} + \left(\boldsymbol{q}_0^H d\boldsymbol{Z} \boldsymbol{q}_0\right) \boldsymbol{q}_0 = d\boldsymbol{Z} \boldsymbol{q}_0. \tag{B.31}$$

Now, since $\boldsymbol{v}^H \boldsymbol{q}_0 = 0$, the vector $\boldsymbol{v}$ must lie in the subspace spanned by the columns of $\boldsymbol{Q}_1$ in (B.23), that is, $\boldsymbol{v} = \boldsymbol{Q}_1 \boldsymbol{Q}_1^H \boldsymbol{v}$. On the other hand, we have

$$(\lambda_0 \boldsymbol{I}_n - \boldsymbol{Z}_0)^+ (\lambda_0 \boldsymbol{I}_n - \boldsymbol{Z}_0) = \boldsymbol{Q}_1 \boldsymbol{Q}_1^H,$$

due to (B.22). Thus, multiplying both sides of (B.31) by $(\lambda_0 \boldsymbol{I}_n - \boldsymbol{Z}_0)^+$ (on the left) and recalling that $\boldsymbol{Q}_1^H \boldsymbol{q}_0 = \boldsymbol{0}$ yields

$$\boldsymbol{v} = (\lambda_0 \boldsymbol{I}_n - \boldsymbol{Z}_0)^+\, d\boldsymbol{Z} \boldsymbol{q}_0. \tag{B.32}$$

Finally, using (B.29) and (B.32) in (B.27) gives

$$d\boldsymbol{q} = (\lambda_0 \boldsymbol{I}_n - \boldsymbol{Z}_0)^+\, d\boldsymbol{Z} \boldsymbol{q}_0 - i\,\frac{\operatorname{Im}\left\{\boldsymbol{c}_0^H\, (\lambda_0 \boldsymbol{I}_n - \boldsymbol{Z}_0)^+\, d\boldsymbol{Z} \boldsymbol{q}_0\right\}}{\boldsymbol{c}_0^H \boldsymbol{q}_0}\, \boldsymbol{q}_0 \,\square$$

B.7 Proof of Lemma 3.7

Note that two P^2-dimensional column vectors $\boldsymbol{a}$ and $\boldsymbol{b}$ are identical if and only if $(\boldsymbol{e}_p^T \otimes \boldsymbol{e}_q^T)\boldsymbol{a} = (\boldsymbol{e}_p^T \otimes \boldsymbol{e}_q^T)\boldsymbol{b}$ for all $1 \le p, q \le P$ (recall that $\boldsymbol{e}_p$ denotes the pth column of the identiy matrix $\boldsymbol{I}_P$). We use this to show $\mathrm{E}\{\boldsymbol{x} \otimes \boldsymbol{x}\} = \boldsymbol{i}_P$. On one hand,

$$(\boldsymbol{e}_p^T \otimes \boldsymbol{e}_q^T)\mathrm{E}\{\boldsymbol{x} \otimes \boldsymbol{x}\} = \mathrm{E}\left\{\boldsymbol{e}_p^T \boldsymbol{x} \cdot \boldsymbol{e}_q^T \boldsymbol{x}\right\} = \mathrm{E}\{x_p\, x_q\} = \delta[p-q].$$

On the other hand,

$$(\boldsymbol{e}_p^T \otimes \boldsymbol{e}_q^T)\boldsymbol{i}_P = (\boldsymbol{e}_p^T \otimes \boldsymbol{e}_q^T)\mathrm{vec}\,(\boldsymbol{I}_P) = \mathrm{vec}\left(\boldsymbol{e}_p^T\, \boldsymbol{I}_P\, \boldsymbol{e}_q\right) = \boldsymbol{e}_p^T \boldsymbol{e}_q = \delta[p-q].$$

We now turn our attention to corr $\{\boldsymbol{x} \otimes \boldsymbol{x}\}$. We have

$$\begin{aligned}
\text{corr}\,\{\boldsymbol{x} \otimes \boldsymbol{x}\} &= \text{E}\left\{(\boldsymbol{x} \otimes \boldsymbol{x})\,(\boldsymbol{x} \otimes \boldsymbol{x})^T\right\} \\
&= \text{E}\left\{\boldsymbol{x}\boldsymbol{x}^T \otimes \boldsymbol{x}\boldsymbol{x}^T\right\} \\
&= \begin{bmatrix} \text{E}\left\{x_1x_1\boldsymbol{x}\boldsymbol{x}^T\right\} & \text{E}\left\{x_1x_2\boldsymbol{x}\boldsymbol{x}^T\right\} & \cdots & \text{E}\left\{x_1x_P\boldsymbol{x}\boldsymbol{x}^T\right\} \\ \text{E}\left\{x_2x_1\boldsymbol{x}\boldsymbol{x}^T\right\} & \text{E}\left\{x_2x_2\boldsymbol{x}\boldsymbol{x}^T\right\} & \cdots & \text{E}\left\{x_2x_P\boldsymbol{x}\boldsymbol{x}^T\right\} \\ \vdots & \vdots & \ddots & \vdots \\ \text{E}\left\{x_Px_1\boldsymbol{x}\boldsymbol{x}^T\right\} & \text{E}\left\{x_Px_2\boldsymbol{x}\boldsymbol{x}^T\right\} & \cdots & \text{E}\left\{x_Px_P\boldsymbol{x}\boldsymbol{x}^T\right\} \end{bmatrix}.
\end{aligned}$$

Thus, corr $\{\boldsymbol{x} \otimes \boldsymbol{x}\}$ has P^2 blocks. Each block is $P \times P$, and the (p, q) block is given by $\text{E}\left\{x_px_q\boldsymbol{x}\boldsymbol{x}^T\right\}$, for $1 \leq p, q \leq P$. We now find a formula for each block. We begin with those lying in the diagonal, that is, we focus on $\text{E}\left\{x_px_p\boldsymbol{x}\boldsymbol{x}^T\right\}$, for given p. This is a $P \times P$ matrix with (k, l) entry given by $\text{E}\left\{x_p^2x_kx_l\right\}$, where $1 \leq k, l \leq P$. If $k = l$ and $k = p$, then $\text{E}\left\{x_p^2x_kx_l\right\} = \text{E}\left\{x_p^4\right\} = \kappa_p$. If $k = l$ but $k \neq p$, then $\text{E}\left\{x_p^2x_kx_l\right\} = \text{E}\left\{x_p^2x_k^2\right\} = \text{E}\left\{x_p^2\right\}\text{E}\left\{x_k^2\right\} = 1$. The remaning entries (off-diagonal) can be easily seen to be zero, by similar reasoning. Thus, the (p, p) block of corr $\{\boldsymbol{x} \otimes \boldsymbol{x}\}$ can be written as

$$\boldsymbol{I}_P + (\kappa_p - 1)\boldsymbol{e}_p\boldsymbol{e}_p^T. \tag{B.33}$$

We now focus on a block $\text{E}\left\{x_px_q\boldsymbol{x}\boldsymbol{x}^T\right\}$ with $p \neq q$. This is a $P \times P$ matrix with (k, l) entry given by $\text{E}\{x_px_qx_kx_l\}$. Since $p \neq q$, $\text{E}\{x_px_qx_kx_l\}$ is non-zero if and only if $(k, l) = (p, q)$ or $(k, l) = (q, p)$. In both cases, the result is 1. Thus, the (p, q) $(p \neq q)$ block of corr $\{\boldsymbol{x} \otimes \boldsymbol{x}\}$ can be written as

$$\boldsymbol{e}_p\boldsymbol{e}_q^T + \boldsymbol{e}_q\boldsymbol{e}_p^T. \tag{B.34}$$

It is now a simple matter to check that the (p, p) and (p, q) $(p \neq q)$ blocks of $\boldsymbol{I}_{P^2} + \boldsymbol{K}_P + \boldsymbol{i}_P\boldsymbol{i}_P^T + \text{diag}\left((\kappa_1 - 3)\,\boldsymbol{e}_1\boldsymbol{e}_1^T, \ldots, (\kappa_P - 3)\,\boldsymbol{e}_P\boldsymbol{e}_P^T\right)$ coincide with those in (B.33) and (B.34), respectively, once we notice that $\boldsymbol{I}_{P^2} = \text{diag}\,(\boldsymbol{I}_P, \boldsymbol{I}_P, \ldots, \boldsymbol{I}_P)$ (P copies of $\boldsymbol{I}_P$),

$$\boldsymbol{K}_P = \begin{bmatrix} \boldsymbol{e}_1\boldsymbol{e}_1^T & \boldsymbol{e}_2\boldsymbol{e}_1^T & \cdots & \boldsymbol{e}_P\boldsymbol{e}_1^T \\ \boldsymbol{e}_1\boldsymbol{e}_2^T & \boldsymbol{e}_2\boldsymbol{e}_2^T & \cdots & \boldsymbol{e}_P\boldsymbol{e}_2^T \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{e}_1\boldsymbol{e}_P^T & \boldsymbol{e}_2\boldsymbol{e}_P^T & \cdots & \boldsymbol{e}_P\boldsymbol{e}_P^T \end{bmatrix},$$

and

$$\boldsymbol{i}_P\boldsymbol{i}_P^T = \begin{bmatrix} \boldsymbol{e}_1\boldsymbol{e}_1^T & \boldsymbol{e}_1\boldsymbol{e}_2^T & \cdots & \boldsymbol{e}_1\boldsymbol{e}_P^T \\ \boldsymbol{e}_2\boldsymbol{e}_1^T & \boldsymbol{e}_2\boldsymbol{e}_2^T & \cdots & \boldsymbol{e}_2\boldsymbol{e}_P^T \\ \vdots & \vdots & \ddots & \vdots \\ \boldsymbol{e}_P\boldsymbol{e}_1^T & \boldsymbol{e}_P\boldsymbol{e}_2^T & \cdots & \boldsymbol{e}_P\boldsymbol{e}_P^T \end{bmatrix} \square$$

B.8 Proof of Lemma 3.8

Within the scope of this proof only, the notation $\boldsymbol{x}[n; K]$ means $\boldsymbol{x}[n; (K, K, \ldots, K)^T]$ (P copies of K) . Define the scalar random sequence $\lambda[n] = \boldsymbol{\theta}^T(\boldsymbol{x}[n; L] \otimes \boldsymbol{x}[n; L])$, where $\boldsymbol{\theta}$ denotes a previously chosen deterministic vector. It is easily seen that $\lambda[n]$ is a stationary

L-dependent sequence. See [37, page 62] for the concept of m-dependent sequences. Thus,

$$\sqrt{N}\left(\frac{1}{N}\sum_{n=1}^{N}\lambda[n]-\mu\right)\xrightarrow{d}\mathcal{N}(0,\nu^2),$$

as $N\to\infty$, where $\mu=\mathrm{E}\{\lambda[n]\}$, $\nu^2=\nu_0+2\sum_{l=1}^{L}\nu_l$ and $\nu_l=\mathrm{cov}\{\lambda[n],\lambda[n-l]\}$, for $l=0,1,\ldots,L$, see [37, theorem 2.8.1, page 108]. We now determine the constants $\mu,\nu_0,\nu_1,\ldots,\nu_L$. Note that the random vector $\boldsymbol{x}[n;K]$ belongs to the class $\mathcal{R}(\boldsymbol{\kappa}\otimes\mathbf{1}_{K+1})$, for all $K\geq 0$. Thus, the mean value μ is given by

$$\mu=\boldsymbol{\theta}^T\,\mathrm{E}\left\{\boldsymbol{x}[n;L]\otimes\boldsymbol{x}[n;L]\right\}=\boldsymbol{\theta}^T\boldsymbol{i}_{P(L+1)}.$$

To compute ν_l we start by noticing the identities $\boldsymbol{x}[n]=\boldsymbol{S}_{P,L,l}\,\boldsymbol{x}[n;L+l]$ and $\boldsymbol{x}[n-l]=\boldsymbol{T}_{P,l,L}\,\boldsymbol{x}[n;L+l]$ (the matrices $\boldsymbol{S}_{P,L,l}$ and $\boldsymbol{T}_{P,l,L}$ were defined in page 87). It follows that

$$\begin{aligned}\mathrm{E}\{\lambda[n]\lambda[n-l]\} &= \boldsymbol{\theta}^T\,\mathrm{E}\left\{\left(\boldsymbol{x}[n]\otimes\boldsymbol{x}[n]\right)\left(\boldsymbol{x}[n-l]\otimes\boldsymbol{x}[n-l]\right)^T\right\}\boldsymbol{\theta}\\ &= \boldsymbol{\theta}^T\,\boldsymbol{S}^{[2]}_{P,L,l}\,\mathrm{corr}\left\{\boldsymbol{x}[n;L+l]\otimes\boldsymbol{x}[n;L+l]\right\}\,{\boldsymbol{T}^{[2]}_{P,l,L}}^T\boldsymbol{\theta}\\ &= \boldsymbol{\theta}^T\,\boldsymbol{S}^{[2]}_{P,L,l}\,\boldsymbol{C}(\boldsymbol{\kappa}\otimes\mathbf{1}_{L+l+1})\,{\boldsymbol{T}^{[2]}_{P,l,L}}^T\boldsymbol{\theta}\\ &= \boldsymbol{\theta}^T\,\boldsymbol{\Sigma}_l\,\boldsymbol{\theta},\end{aligned}$$

and $\nu_l=\mathrm{E}\{\lambda[n]\lambda[n-l]\}-\mu^2=\boldsymbol{\theta}^T\left(\boldsymbol{\Sigma}_l-\boldsymbol{\mu}_{P(L+1)}\boldsymbol{\mu}^T_{P(L+1)}\right)\boldsymbol{\theta}$. Thus,

$$\nu^2=\nu_0+2\sum_{l=1}^{L}\nu_l=\boldsymbol{\theta}^T\,\boldsymbol{\Sigma}\,\boldsymbol{\theta},$$

where we defined $\boldsymbol{\Sigma}=\boldsymbol{\Sigma}_0+2\sum_{l=1}^{L}\boldsymbol{\Sigma}_l-(2L+1)\boldsymbol{\mu}_{P(L+1)}\boldsymbol{\mu}^T_{P(L+1)}$. This proves that

$$\boldsymbol{\theta}^T\left\{\sqrt{N}\left(\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}[n;L]\otimes\boldsymbol{x}[n;L]-\boldsymbol{\mu}_{P(L+1)}\right)\right\}\xrightarrow{d}\boldsymbol{\theta}^T\boldsymbol{z}$$

where $\boldsymbol{z}\overset{d}{=}\mathcal{N}\left(\mathbf{0},\boldsymbol{\Sigma}\right)$. Because $\boldsymbol{\theta}$ was chosen arbitrarily, the Cramér-Wold device permits to conclude that

$$\sqrt{N}\left(\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}[n;L]\otimes\boldsymbol{x}[n;L]-\boldsymbol{\mu}_{P(L+1)}\right)\xrightarrow{d}\mathcal{N}\left(\mathbf{0},\boldsymbol{\Sigma}\right).$$

Now, since $\boldsymbol{x}[n;\boldsymbol{l}]=\boldsymbol{D}(\boldsymbol{l})\boldsymbol{x}[n;L]$, we have

$$\boldsymbol{r}^N=\boldsymbol{D}(\boldsymbol{l})^{[2]}\left(\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{x}[n;L]\otimes\boldsymbol{x}[n;L]\right).$$

Consequently,

$$\sqrt{N}\,(\,\boldsymbol{r}^N-\underbrace{\boldsymbol{D}(\boldsymbol{l})^{[2]}\boldsymbol{\mu}_{P(L+1)}}_{\boldsymbol{\mu}(\boldsymbol{\kappa};\boldsymbol{l})}\,)\xrightarrow{d}\mathcal{N}\,(\,\mathbf{0},\underbrace{\boldsymbol{D}(\boldsymbol{l})^{[2]}\boldsymbol{\Sigma}{\boldsymbol{D}(\boldsymbol{l})^{[2]}}^T}_{\boldsymbol{\Sigma}(\boldsymbol{\kappa};\boldsymbol{l})}\,),$$

as claimed □

B.9 Proof of Lemma 3.9

For a complex vector $\boldsymbol{z} = \boldsymbol{x} + i\boldsymbol{y}$, where $\boldsymbol{x}, \boldsymbol{y} \in \mathbb{R}^m$, we have the identities $\boldsymbol{x} = \boldsymbol{E}_R[m]\imath(\boldsymbol{z})$ and $\boldsymbol{y} = \boldsymbol{E}_I[m]\imath(\boldsymbol{z})$. Thus,

$$\begin{aligned} \operatorname{Re} \boldsymbol{z}\boldsymbol{z}^H &= \boldsymbol{E}_R[m]\imath(\boldsymbol{z})\imath(\boldsymbol{z})^T\boldsymbol{E}_R[m]^T + \boldsymbol{E}_I[m]\imath(\boldsymbol{z})\imath(\boldsymbol{z})^T\boldsymbol{E}_I[m]^T \\ \operatorname{Im} \boldsymbol{z}\boldsymbol{z}^H &= \boldsymbol{E}_I[m]\imath(\boldsymbol{z})\imath(\boldsymbol{z})^T\boldsymbol{E}_R[m]^T - \boldsymbol{E}_R[m]\imath(\boldsymbol{z})\imath(\boldsymbol{z})^T\boldsymbol{E}_I[m]^T. \end{aligned}$$

This implies

$$\imath\left(\boldsymbol{z}\boldsymbol{z}^H\right) = \underbrace{\left[\begin{array}{c} \boldsymbol{E}_R[m] \otimes \boldsymbol{E}_R[m] + \boldsymbol{E}_I[m] \otimes \boldsymbol{E}_I[m] \\ \boldsymbol{E}_R[m] \otimes \boldsymbol{E}_I[m] - \boldsymbol{E}_I[m] \otimes \boldsymbol{E}_R[m] \end{array}\right]}_{\boldsymbol{E}[m]} \operatorname{vec}\left(\imath(\boldsymbol{z})\imath(\boldsymbol{z})^T\right).$$

Now,

$$\begin{aligned} \imath\left(\boldsymbol{R}^N\right) &= \frac{1}{N}\sum_{n=1}^{N} \imath\left(\boldsymbol{x}[n;\boldsymbol{l}]\boldsymbol{x}[n;\boldsymbol{l}]^H\right) \\ &= \frac{1}{N}\sum_{n=1}^{N} \boldsymbol{E}[L] \operatorname{vec}\left(\imath\left(\boldsymbol{x}[n;\boldsymbol{l}]\right)\imath\left(\boldsymbol{x}[n;\boldsymbol{l}]\right)^T\right) \\ &= \boldsymbol{E}[L] \operatorname{vec}\left(\frac{1}{N}\sum_{n=1}^{N} \imath\left(\boldsymbol{x}[n;\boldsymbol{l}]\right)\imath\left(\boldsymbol{x}[n;\boldsymbol{l}]\right)^T\right) \\ &= \boldsymbol{E}[L] \operatorname{vec}\Bigg(\underbrace{\frac{1}{N}\sum_{n=1}^{N} \boldsymbol{y}[n;\boldsymbol{l}^{(2)}]\boldsymbol{y}[n;\boldsymbol{l}^{(2)}]^T}_{\boldsymbol{R}_{\boldsymbol{y}}^N}\Bigg), \end{aligned} \tag{B.35}$$

where we defined $\boldsymbol{y}[n] = \imath\left(\boldsymbol{x}[n]\right)$. Since $\boldsymbol{y}[n]$ belongs to the class $\mathcal{R}_{\mathbb{Z}}\left(\boldsymbol{\kappa}^{(2)}\right)$, lemma 3.8 asserts that

$$\operatorname{vec}\left(\boldsymbol{R}_{\boldsymbol{y}}^N\right) \sim \sqrt{N} - \mathcal{AN}\left(\operatorname{vec}\left(\boldsymbol{I}_L\right), \boldsymbol{R}\left(\boldsymbol{\kappa}^{(2)};\boldsymbol{l}^{(2)}\right)\right).$$

Thus, from (B.35), we have

$$\imath\left(\boldsymbol{R}^N\right) \sim \sqrt{N} - \mathcal{AN}\Bigg(\underbrace{\boldsymbol{E}[L]\operatorname{vec}\left(\boldsymbol{I}_L\right)}_{\imath\left(2\boldsymbol{I}_L\right)}, \underbrace{\boldsymbol{E}[L]\boldsymbol{R}\left(\boldsymbol{\kappa}^{(2)};\boldsymbol{l}^{(2)}\right)\boldsymbol{E}[L]^T}_{\boldsymbol{C}(\boldsymbol{\kappa};\boldsymbol{l})}\Bigg). \tag{B.36}$$

By definition, equation (B.36) means that $\boldsymbol{R}^N \sim \sqrt{N} - \mathcal{AN}\left(2\boldsymbol{I}_L, \boldsymbol{C}(\boldsymbol{\kappa};\boldsymbol{l})\right)$ □

B.10 Proof of Lemma 3.10

Let $E_{1p}, \ldots, E_{np}$ denote an orthonormal basis for H_pM and let $\omega_i = E_{ip}^\flat$. Define $F_{i\varrho(p)} = \varrho_*(E_{ip})$. Note that $F_{1\varrho(p)}, \ldots, F_{n\varrho(p)}$ denotes an orthonormal basis for $T_{\varrho(p)}N$, because ϱ is a Riemannian submersion. Define $\sigma_{i\varrho(p)} = F_{i\varrho(p)}^\flat$. Thus, $\sigma_{1\varrho(p)}, \ldots, \sigma_{n\varrho(p)}$ denotes an orthonormal basis for $T^*_{\varrho(p)}N$. Thus, using (3.28), we have

$$\operatorname{tr} \Upsilon = \sum_{i=1}^{n} \Upsilon\left(\sigma_{i\varrho(p)}, \sigma_{i\varrho(p)}\right).$$

But, it is easily seen that $\varrho^* \sigma_{i\varrho(p)} = \omega_{ip}$. Thus,

$$\begin{aligned} \operatorname{tr} \Upsilon &= \sum_{i=1}^{m} \Upsilon \left(\sigma_{i\varrho(p)}, \sigma_{i\varrho(p)} \right) \\ &= \sum_{i=1}^{m} \left(\varrho_* \Sigma \right) \left(\sigma_{i\varrho(p)}, \sigma_{i\varrho(p)} \right) \\ &= \sum_{i=1}^{m} \Sigma \left(\varrho^* \sigma_{i\varrho(p)}, \varrho^* \sigma_{i\varrho(p)} \right) \\ &= \sum_{i=1}^{m} \Sigma \left(\omega_{ip}, \omega_{ip} \right) \square \end{aligned}$$

Appendix C

Derivative of ψ_4, ψ_3, ψ_2 and ψ_1

C.1 Derivative of ψ_4

In this section, we calculate the derivative of ψ_4 at an arbitrary point $(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_P)$ of its domain $\mathcal{U}_4$, denoted

$$D\psi_4\left(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_P\right). \tag{C.1}$$

Given our definitions for derivatives of complex mappings in page 60, and the theory of differentials in [39], it is easily seen that (C.1) is the unique matrix satisfying

$$\imath\left(d\psi_4\right) = D\psi_4\left(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_P\right)\, \imath\left(d\boldsymbol{Z}_0, d\boldsymbol{Z}_1, \ldots, d\boldsymbol{Z}_P\right),$$

or, equivalently,

$$\imath\left(\mathrm{vec}(d\psi_4)\right) = D\psi_4\left(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_P\right) \begin{bmatrix} \imath\left(\mathrm{vec}(d\boldsymbol{Z}_0)\right) \\ \imath\left(\mathrm{vec}(d\boldsymbol{Z}_1)\right) \\ \vdots \\ \imath\left(\mathrm{vec}(d\boldsymbol{Z}_P)\right) \end{bmatrix},$$

where the symbol d stands for the differential. From (3.71), we have

$$d\psi_4 = d\boldsymbol{Z}_0\, [\, \boldsymbol{Z}_1 \cdots \boldsymbol{Z}_P\,]\, \boldsymbol{R}_{\boldsymbol{s}}[0; \boldsymbol{d}_0]^{-1/2} + \boldsymbol{Z}_0\, [\, d\boldsymbol{Z}_1 \cdots d\boldsymbol{Z}_P\,]\, \boldsymbol{R}_{\boldsymbol{s}}[0; \boldsymbol{d}_0]^{-1/2}. \tag{C.2}$$

Vectorizing both sides of (C.2) yields

$$\mathrm{vec}\left(d\psi_4\right) = \underbrace{\begin{bmatrix} \boldsymbol{R}_{\boldsymbol{s}}[0; \boldsymbol{d}_0]^{-1/2}\, [\, \boldsymbol{Z}_1 \cdots \boldsymbol{Z}_P\,]^T \otimes \boldsymbol{I}_Q & \boldsymbol{R}_{\boldsymbol{s}}[0; \boldsymbol{d}_0]^{-1/2} \otimes \boldsymbol{Z}_0 \end{bmatrix}}_{\boldsymbol{\mathcal{M}}\left(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_P\right)} \begin{bmatrix} \mathrm{vec}\left(d\boldsymbol{Z}_0\right) \\ \mathrm{vec}\left(d\boldsymbol{Z}_1\right) \\ \vdots \\ \mathrm{vec}\left(d\boldsymbol{Z}_P\right) \end{bmatrix}. \tag{C.3}$$

Embedding both sides of (C.3), and using properties (3.21) and (3.23), gives

$$\imath\left(d\psi_4\right) = \jmath\left(\boldsymbol{\mathcal{M}}\left(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_P\right)\right) \boldsymbol{\Pi}_{QD, D(D_0+1), \ldots, D(D_0+1)} \begin{bmatrix} \imath\left(\mathrm{vec}(d\boldsymbol{Z}_0)\right) \\ \imath\left(\mathrm{vec}(d\boldsymbol{Z}_1)\right) \\ \vdots \\ \imath\left(\mathrm{vec}(d\boldsymbol{Z}_P)\right) \end{bmatrix}. \tag{C.4}$$

We recall that the permutation matrix $\boldsymbol{\Pi}_{m_1, \ldots, m_k}$ was defined in page 60. Thus, by inspection of (C.4), the derivative of ψ_4 at $(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_P)$ is

$$D\psi_4\left(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_P\right) = \jmath\left(\boldsymbol{\mathcal{M}}\left(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_P\right)\right) \boldsymbol{\Pi}_{QD, D(D_0+1), \ldots, D(D_0+1)}.$$

C.2 Derivative of ψ_3

In this section, we compute the derivative of ψ_3 at an arbitrary point $(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M)$ of its domain $\mathcal{U}_3$, denoted $D\psi_3(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M)$. Express ψ_3 in component mappings as $\psi_3 = (\xi, \nu_1, \ldots, \nu_P)$, where

$$\xi(\boldsymbol{Z}_0) = \boldsymbol{Z}_0 \tag{C.5}$$

and

$$\nu_p(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M) = \vartheta_p \circ \eta_p(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M), \tag{C.6}$$

for $p = 1, \ldots, P$. It is clear that the derivative of ψ_3 at the point $(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M)$ has the block structure

$$D\psi_3(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M) = \begin{bmatrix} D\xi(\boldsymbol{Z}_0) & \boldsymbol{0} \\ \boldsymbol{0} & D\nu_1(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M) \\ \vdots & \vdots \\ \boldsymbol{0} & D\nu_P(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M) \end{bmatrix}.$$

From (C.5), we have trivially $D\xi(\boldsymbol{Z}_0) = \boldsymbol{I}_{2QD}$. From (C.6), it follows that

$$D\nu_p(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M) = D\vartheta_p(\eta_p(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M))\, D\eta_p(\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M). \tag{C.7}$$

To compute $D\vartheta_p(\boldsymbol{Z})$ we further write $\vartheta_p = \varrho_p \circ \zeta$, where

$$\zeta(\boldsymbol{Z}) = \boldsymbol{Z}^H \boldsymbol{Z} \tag{C.8}$$

and

$$\varrho_p(\boldsymbol{Z}) = \sqrt{D_0 + 1}\, \mathrm{vec}^{-1}(\boldsymbol{q}(\boldsymbol{Z}; \lambda_{\min}(\boldsymbol{Z}); \boldsymbol{s}_p)). \tag{C.9}$$

Thus, $D\vartheta_p(\boldsymbol{Z}) = D\varrho_p(\zeta(\boldsymbol{Z}))\, D\zeta(\boldsymbol{Z})$. From (C.8),

$$d\zeta = (d\boldsymbol{Z})^H \boldsymbol{Z} + \boldsymbol{Z}^H d\boldsymbol{Z}. \tag{C.10}$$

Embedding both sides of (C.10), and using properties (3.20) and (3.21), yields

$$\imath(d\zeta) = \underbrace{\jmath\left(\boldsymbol{Z}^T \otimes \boldsymbol{I}_{D(D_0+1)}\right) \boldsymbol{C}_{2MD(D_0+1),D(D_0+1)} + \jmath\left(\boldsymbol{I}_{D(D_0+1)} \otimes \boldsymbol{Z}^H\right)}_{D\zeta(\boldsymbol{Z})} \imath(d\boldsymbol{Z}),$$

which directly exposes $D\zeta(\boldsymbol{Z})$. From (C.9), we have

$$\mathrm{vec}(\varrho_p)(\boldsymbol{Z}) = \sqrt{D_0 + 1}\, \boldsymbol{q}(\boldsymbol{Z}; \lambda_{\min}(\boldsymbol{Z}); \boldsymbol{s}_p).$$

Thus,

$$D\varrho_p(\boldsymbol{Z}) = D\mathrm{vec}(\varrho_p)(\boldsymbol{Z}) = \sqrt{D_0 + 1}\, D\boldsymbol{q}(\boldsymbol{Z}; \lambda_{\min}(\boldsymbol{Z}); \boldsymbol{s}_p).$$

We now turn to the derivative of η_p at the point $(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M)$, which is needed in (C.7). Given (3.77), we have

$$\mathrm{vec}\left((d\eta_p)^H\right) = \begin{bmatrix} \mathrm{vec}\left(\boldsymbol{I}_{D_0+1} \otimes (d\boldsymbol{Z}_1)^H\right) \\ \mathrm{vec}\left(\boldsymbol{I}_{D_0+1} \otimes d\boldsymbol{Z}_1\right) \\ \vdots \\ \mathrm{vec}\left(\boldsymbol{I}_{D_0+1} \otimes (d\boldsymbol{Z}_M)^H\right) \\ \mathrm{vec}\left(\boldsymbol{I}_{D_0+1} \otimes d\boldsymbol{Z}_M\right) \end{bmatrix}. \tag{C.11}$$

We recall that for $\boldsymbol{A} : m \times n$ and $\boldsymbol{B} : p \times q$ we have $\operatorname{vec}(\boldsymbol{A} \otimes \boldsymbol{B}) = (\boldsymbol{G} \otimes \boldsymbol{I}_p) \operatorname{vec}(\boldsymbol{B})$, where $\boldsymbol{G} = (\boldsymbol{I}_n \otimes \boldsymbol{K}_{q,m})(\operatorname{vec}(\boldsymbol{A}) \otimes \boldsymbol{I}_q)$, see [39, page 48]. Using this property in (C.11) yields

$$\operatorname{vec}\left((d\eta_p)^H\right) = (\boldsymbol{I}_{2M} \otimes \boldsymbol{\mathcal{N}}) \begin{bmatrix} \operatorname{vec}\left((d\boldsymbol{Z}_1)^H\right) \\ \operatorname{vec}(d\boldsymbol{Z}_1) \\ \vdots \\ \operatorname{vec}\left((d\boldsymbol{Z}_M)^H\right) \\ \operatorname{vec}(d\boldsymbol{Z}_M) \end{bmatrix}, \tag{C.12}$$

where $\boldsymbol{\mathcal{N}} = \boldsymbol{G} \otimes \boldsymbol{I}_D$ and $\boldsymbol{G} = (\boldsymbol{I}_{D_0+1} \otimes \boldsymbol{K}_{D,D_0+1})(\operatorname{vec}(\boldsymbol{I}_{D_0+1}) \otimes \boldsymbol{I}_D)$. Using properties (3.20), (3.21) and (3.23) in (C.12) gives

$$\imath\left(\operatorname{vec}\left((d\eta_p)^H\right)\right) = \jmath(\boldsymbol{I}_{2M} \otimes \boldsymbol{\mathcal{N}})\, \boldsymbol{\Pi}_{D^2,\ldots,D^2} \left(\boldsymbol{I}_M \otimes \begin{bmatrix} \boldsymbol{C}_{D,D} \\ \boldsymbol{I}_{2D^2} \end{bmatrix}\right) \begin{bmatrix} \imath(d\boldsymbol{Z}_1) \\ \imath(d\boldsymbol{Z}_2) \\ \vdots \\ \imath(d\boldsymbol{Z}_M) \end{bmatrix}.$$

Finally, using the fact that

$$\imath\left(\operatorname{vec}\left((d\eta_p)^H\right)\right) = \imath\left((d\eta_p)^H\right) = \boldsymbol{C}_{2MD(D_0+1),D(D_0+1)}\, \imath(d\eta_p),$$

we have

$$\imath(d\eta_p) = \underbrace{\boldsymbol{C}^T_{2MD(D_0+1),D(D_0+1)} \jmath(\boldsymbol{I}_{2M} \otimes \boldsymbol{\mathcal{N}})\, \boldsymbol{\Pi}_{D^2,\ldots,D^2} \left(\boldsymbol{I}_M \otimes \begin{bmatrix} \boldsymbol{C}_{D,D} \\ \boldsymbol{I}_{2D^2} \end{bmatrix}\right)}_{D\eta_p(\boldsymbol{Z}_1,\ldots,\boldsymbol{Z}_M)} \begin{bmatrix} \imath(d\boldsymbol{Z}_1) \\ \imath(d\boldsymbol{Z}_2) \\ \vdots \\ \imath(d\boldsymbol{Z}_M) \end{bmatrix}.$$

C.3 Derivative of ψ_2

In this section, we compute the derivative of ψ_2 at an arbitrary point $(\boldsymbol{z}, \boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M)$ of its domain $\mathcal{U}_2$, denoted $D\psi_2(\boldsymbol{z}, \boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M)$. Due to (3.80), the derivative of ψ_2 has the block structure

$$D\psi_2(\boldsymbol{z}, \boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M) = \begin{bmatrix} D\tau(\boldsymbol{z}, \boldsymbol{Z}_0) & \boldsymbol{0} & \cdots & \boldsymbol{0} \\ D\upsilon_1(\boldsymbol{z}, \boldsymbol{Z}_0) & D\varsigma(\boldsymbol{Z}_1) & \cdots & \boldsymbol{0} \\ \vdots & \vdots & \ddots & \vdots \\ D\upsilon_M(\boldsymbol{z}, \boldsymbol{Z}_0) & \boldsymbol{0} & \cdots & D\varsigma(\boldsymbol{Z}_M) \end{bmatrix},$$

where

$$\upsilon_m(\boldsymbol{z}, \boldsymbol{Z}) = \upsilon(\boldsymbol{z}, \boldsymbol{Z})\, \boldsymbol{Z}_m \upsilon(\boldsymbol{z}, \boldsymbol{Z})^H \tag{C.13}$$

and

$$\varsigma(\boldsymbol{Z}) = \upsilon(\boldsymbol{z}, \boldsymbol{Z}_0)\, \boldsymbol{Z} \upsilon(\boldsymbol{z}, \boldsymbol{Z}_0)^H. \tag{C.14}$$

We start with $D\tau(\boldsymbol{z}, \boldsymbol{Z})$. Let $\boldsymbol{Z}$ be fixed and compute the differential $d\tau$ with respect to $\boldsymbol{z}$. Given (3.78), it can be shown that

$$d\tau = \frac{1}{2} \boldsymbol{Z} \operatorname{Diag}\left(\operatorname{Re}\boldsymbol{z} - \sigma^2 \boldsymbol{1}_D\right)^{-1/2} \operatorname{Diag}(\operatorname{Re} d\boldsymbol{z}). \tag{C.15}$$

Vectorizing both sides of (C.15) and simplifying leads to

$$\operatorname{vec}(d\tau) = \underbrace{\frac{1}{2}\left[\begin{array}{cc} \boldsymbol{I}_D \boxtimes \boldsymbol{Z}\operatorname{Diag}\left(\operatorname{Re}\boldsymbol{z} - \sigma^2\mathbf{1}_D\right)^{-1/2} & \mathbf{0}\end{array}\right]}_{\boldsymbol{\mathcal{O}}(\boldsymbol{z},\boldsymbol{Z})} \imath(d\boldsymbol{z}). \tag{C.16}$$

Here, for matrices $\boldsymbol{A} = [\,\boldsymbol{a}_1\,\boldsymbol{a}_2\,\cdots\,\boldsymbol{a}_n\,]$ and $\boldsymbol{B} = [\,\boldsymbol{b}_1\,\boldsymbol{b}_2\,\cdots\,\boldsymbol{b}_n\,]$ with the same number of columns, the symbol $\boldsymbol{A}\boxtimes\boldsymbol{B}$ denotes their Khatri-Rao product [47],

$$\boldsymbol{A}\boxtimes\boldsymbol{B} = \left[\begin{array}{cccc}\boldsymbol{a}_1\otimes\boldsymbol{b}_1 & \boldsymbol{a}_2\otimes\boldsymbol{b}_2 & \cdots & \boldsymbol{a}_n\otimes\boldsymbol{b}_n\end{array}\right].$$

From (C.16) we conclude that

$$\imath(d\tau) = \underbrace{\left[\begin{array}{c}\operatorname{Re}\boldsymbol{\mathcal{O}}(\boldsymbol{z},\boldsymbol{Z})\\ \operatorname{Im}\boldsymbol{\mathcal{O}}(\boldsymbol{z},\boldsymbol{Z})\end{array}\right]}_{\boldsymbol{\mathcal{P}}(\boldsymbol{z},\boldsymbol{Z})}\imath(d\boldsymbol{z}). \tag{C.17}$$

Holding now $\boldsymbol{z}$ fixed and calculating the differential $d\tau$ with respect to $\boldsymbol{Z}$ in (3.78) yields, after some computations,

$$\imath(d\tau) = \jmath\left(\operatorname{Diag}\left(\operatorname{Re}\boldsymbol{z} - \sigma^2\mathbf{1}_D\right)^{1/2}\otimes\boldsymbol{I}_Q\right)\imath(d\boldsymbol{Z}). \tag{C.18}$$

Thus, using (C.17) and (C.18), we have

$$D\tau(\boldsymbol{z},\boldsymbol{Z}_0) = \left[\begin{array}{cc}\boldsymbol{\mathcal{P}}(\boldsymbol{z},\boldsymbol{Z}_0) & \jmath\left(\operatorname{Diag}\left(\operatorname{Re}\boldsymbol{z} - \sigma^2\mathbf{1}_D\right)^{1/2}\otimes\boldsymbol{I}_Q\right)\end{array}\right]. \tag{C.19}$$

We now compute $D\upsilon_m(\boldsymbol{z},\boldsymbol{Z}_0)$. Given (C.13), we have

$$d\upsilon_m = d\upsilon\boldsymbol{Z}_m\upsilon^H + \upsilon\boldsymbol{Z}_m(d\upsilon)^H.$$

As a consequence,

$$\begin{aligned}\imath(d\upsilon_m) &= \jmath\left(\overline{\upsilon(\boldsymbol{z},\boldsymbol{Z})}\boldsymbol{Z}_m^T\otimes\boldsymbol{I}_D\right)\imath(d\upsilon) + \jmath\left(\boldsymbol{I}_D\otimes\upsilon(\boldsymbol{z},\boldsymbol{Z})\,\boldsymbol{Z}_m\right)\imath\left((d\upsilon)^H\right)\\ &= \underbrace{\left[\jmath\left(\overline{\upsilon(\boldsymbol{z},\boldsymbol{Z})}\boldsymbol{Z}_m^T\otimes\boldsymbol{I}_D\right) + \jmath\left(\boldsymbol{I}_D\otimes\upsilon(\boldsymbol{z},\boldsymbol{Z})\,\boldsymbol{Z}_m\right)\boldsymbol{C}_{D,Q}\right]}_{\boldsymbol{\mathcal{Q}}(\boldsymbol{z},\boldsymbol{Z},\boldsymbol{Z}_m)}\imath(d\upsilon).\end{aligned} \tag{C.20}$$

From (C.20) we conclude that

$$D\upsilon_m(\boldsymbol{z},\boldsymbol{Z}_0) = \boldsymbol{\mathcal{Q}}(\boldsymbol{z},\boldsymbol{Z},\boldsymbol{Z}_m)\;D\upsilon(\boldsymbol{z},\boldsymbol{Z}_0).$$

Now, given (3.79), and performing computations similar to those leading to (C.19), it can be shown that

$$D\upsilon(\boldsymbol{z},\boldsymbol{Z}_0) = \left[\begin{array}{cc}\boldsymbol{\mathcal{S}}(\boldsymbol{z},\boldsymbol{Z}_0) & \jmath\left(\boldsymbol{I}_Q\otimes\operatorname{Diag}\left(\operatorname{Re}\boldsymbol{z} - \sigma^2\mathbf{1}_D\right)^{-1/2}\right)\boldsymbol{C}_{Q,D}\end{array}\right],$$

where

$$\boldsymbol{\mathcal{S}}(\boldsymbol{z},\boldsymbol{Z}) = \left[\begin{array}{c}\operatorname{Re}\boldsymbol{\mathcal{R}}(\boldsymbol{z},\boldsymbol{Z})\\ \operatorname{Im}\boldsymbol{\mathcal{R}}(\boldsymbol{z},\boldsymbol{Z})\end{array}\right]$$

and

$$\boldsymbol{\mathcal{R}}(\boldsymbol{z},\boldsymbol{Z}) = -\frac{1}{2}\left[\begin{array}{cc}\overline{\boldsymbol{Z}}\operatorname{Diag}\left(\operatorname{Re}\boldsymbol{z} - \sigma^2\mathbf{1}_D\right)^{-3/2}\boxtimes\boldsymbol{I}_D & \mathbf{0}\end{array}\right].$$

Finally, given (C.14), we have

$$D\varsigma(\boldsymbol{Z}_m) = \jmath\left(\overline{\upsilon(\boldsymbol{z},\boldsymbol{Z}_0)}\otimes\upsilon(\boldsymbol{z},\boldsymbol{Z}_0)\right).$$

C.4 Derivative of ψ_1

In this section, we compute the derivative of ψ_1 at an arbitrary point $(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M)$ of its domain $\mathcal{U}_1$, denoted $D\psi_1(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M)$. Given (3.83), the derivative of ψ_1 has the block structure

$$D\psi_1(\boldsymbol{Z}_0, \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_M) = \begin{bmatrix} D\nu(\boldsymbol{Z}_0) & \mathbf{0} & \cdots & \mathbf{0} \\ D\xi(\boldsymbol{Z}_0) & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \boldsymbol{I}_{2Q^2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \boldsymbol{I}_{2Q^2} \end{bmatrix}$$

where

$$\nu(\boldsymbol{Z}) = \boldsymbol{\lambda}(\rho(\boldsymbol{Z})) \tag{C.21}$$

and

$$\xi(\boldsymbol{Z}) = \boldsymbol{Q}(\rho(\boldsymbol{Z}); \boldsymbol{r}_1, \ldots, \boldsymbol{r}_D). \tag{C.22}$$

We start by noticing that, due to (3.82), we have

$$D\rho(\boldsymbol{Z}) = \frac{1}{2}\left(\boldsymbol{I}_{2Q^2} + \boldsymbol{C}_{Q,Q}\right).$$

Since

$$\nu(\boldsymbol{Z}) = \begin{bmatrix} \lambda_1(\rho(\boldsymbol{Z})) \\ \lambda_2(\rho(\boldsymbol{Z})) \\ \vdots \\ \lambda_D(\rho(\boldsymbol{Z})) \end{bmatrix},$$

it follows that

$$D\nu(\boldsymbol{Z}) = \begin{bmatrix} D\lambda(\rho(\boldsymbol{Z}); \lambda_1(\rho(\boldsymbol{Z})); \boldsymbol{r}_1) \\ D\lambda(\rho(\boldsymbol{Z}); \lambda_2(\rho(\boldsymbol{Z})); \boldsymbol{r}_2) \\ \vdots \\ D\lambda(\rho(\boldsymbol{Z}); \lambda_D(\rho(\boldsymbol{Z})); \boldsymbol{r}_D) \end{bmatrix} D\rho(\boldsymbol{Z}).$$

We recall that the definition of $D\lambda(\boldsymbol{Z}; \lambda; \boldsymbol{c})$ was introduced in page 82.

Similarly, due to (C.22) and (3.81), we have

$$\operatorname{vec}(\xi(\boldsymbol{Z})) = \begin{bmatrix} \boldsymbol{q}(\rho(\boldsymbol{Z}); \lambda_1(\rho(\boldsymbol{Z})); \boldsymbol{r}_1) \\ \boldsymbol{q}(\rho(\boldsymbol{Z}); \lambda_2(\rho(\boldsymbol{Z})); \boldsymbol{r}_2) \\ \vdots \\ \boldsymbol{q}(\rho(\boldsymbol{Z}); \lambda_D(\rho(\boldsymbol{Z})); \boldsymbol{r}_D) \end{bmatrix}.$$

Thus,

$$D\xi(\boldsymbol{Z}) = \boldsymbol{\Pi}_{D,\ldots,D} \begin{bmatrix} D\boldsymbol{q}(\rho(\boldsymbol{Z}); \lambda_1(\rho(\boldsymbol{Z})); \boldsymbol{r}_1) \\ D\boldsymbol{q}(\rho(\boldsymbol{Z}); \lambda_2(\rho(\boldsymbol{Z})); \boldsymbol{r}_2) \\ \vdots \\ D\boldsymbol{q}(\rho(\boldsymbol{Z}); \lambda_D(\rho(\boldsymbol{Z})); \boldsymbol{r}_D) \end{bmatrix} D\rho(\boldsymbol{Z}),$$

where the definition of $D\boldsymbol{q}(\boldsymbol{Z}; \lambda; \boldsymbol{c})$ is given in page 82.

Appendix D

Proofs for chapter 4

D.1 Proof of Theorem 4.1

Before proving the intrinsic variance lower bound (IVLB), we need a technical lemma.

Lemma D.1. *Let the sectional curvature of M be bounded above by $C \geq 0$ on the geodesic ball $B_\epsilon(m)$. That is, $K(\Pi) \leq C$, for all planes $\Pi \subset T_nM$, $n \in B_\epsilon(m)$. Suppose that $\sqrt{C}\epsilon < T \equiv \sqrt{3/2}$ holds. Then, the function $k_m : M \to \mathbb{R}$, $k_m(n) = \frac{1}{2}d(m,n)^2$, is smooth on $B_\epsilon(m)$ and we have*

$$|\mathrm{grad} k_m(n)| = d(m,n) \tag{D.1}$$

$$\mathrm{Hess}\, k_m(X_n, X_n) \geq \left(1 - \frac{2}{3}Cd(m,n)^2\right)|X_n|^2, \tag{D.2}$$

for all $X_n \in T_nM$, and $n \in B_\epsilon(m)$.

Proof. Use [32, theorem 4.6.1, page 193] together with the inequality $t\,\mathrm{ctg}(t) \geq 1 - \frac{2}{3}t^2$, for $0 \leq t \leq T$ □

Proof of Theorem 4.1 We start by restating, in our notation, the line of proof from [27]. This goes, with minor modifications, nearly up to (D.7). We then conclude by manipulating some classical inequalities and exploiting lemma D.1. Since ϑ has bias b, $b(p)$ is a global minimizer of $\rho_p : M \to \mathbb{R}$,

$$\rho_p(n) = \mathrm{E}_p\left\{k_\vartheta(n)\right\} = \int_\Omega k_{\vartheta(\omega)}(n)\, f(\omega; p)\, d\mu,$$

hence, a stationary point of ρ_p. Thus,

$$X_{b(p)}\rho_p = 0, \tag{D.3}$$

for any $X_{b(p)} \in T_{b(p)}M$. Let $X \in \mathcal{T}(M)$ and define $\phi : P \to \mathbb{R}$, $\phi(p) = X_{b(p)}\rho_p$. Thus, $\phi \equiv 0$. Note that

$$\begin{aligned}\phi(p) &= \int_\Omega X_{b(p)} k_{\vartheta(\omega)} f(\omega; p)\, d\mu \\ &= \int_\Omega \left(\psi_\omega \circ b\right)(p) f(\omega; p)\, d\mu,\end{aligned} \tag{D.4}$$

where $\psi_\omega : M \to \mathbb{R}$, $\psi_\omega(n) = dk_{\vartheta(\omega)}(X_n)$ and we recall that the symbol d denotes exterior differentiation. Let $Y_p \in T_pP$. Given $Y_p\phi = 0$ and equation (D.4), we have

$$\int_\Omega Y_p\left(\psi_\omega \circ b\right) f(\omega;p) + X_{b(p)}k_{\vartheta(\omega)}Y_p f_\omega \, d\mu = 0. \tag{D.5}$$

Note that $Y_p\left(\psi_\omega \circ b\right) = b_*\left(Y_p\right)\psi_\omega$. By the rules of covariant differentiation,

$$\begin{aligned} b_*(Y_p)\psi_\omega &= b_*(Y_p)\left(dk_{\vartheta(\omega)}(X)\right) \\ &= \left(\nabla_{b_*(Y_p)}dk_{\vartheta(\omega)}\right)\left(X_{b(p)}\right) + dk_{\vartheta(\omega)}\left(\nabla_{b_*(Y_p)}X\right) \\ &= \operatorname{Hess} k_{\vartheta(\omega)}\left(X_{b(p)}, b_*(Y_p)\right) + \nabla_{b_*(Y_p)}Xk_{\vartheta(\omega)}. \end{aligned} \tag{D.6}$$

Inserting (D.6) in (D.5) gives

$$\int_\Omega \operatorname{Hess} k_{\vartheta(\omega)}\left(X_{b(p)}, b_*(Y_p)\right) f(\omega;p) + X_{b(p)}k_{\vartheta(\omega)}Y_p l_\omega f(\omega;p)\, d\mu = 0. \tag{D.7}$$

Here, we exploited the fact that

$$\int_\Omega \nabla_{b_*(Y_p)}Xk_{\vartheta(\omega)}f(\omega;p)d\mu = \left(\nabla_{b_*(Y_p)}X\right)\rho_p = 0,$$

with the last equality following from (D.3). Moreover, in (D.7), we made use of the identity $Y_p f_\omega = Y_p l_\omega f(\omega;p)$.

Putting $X_{b(p)} = b_*(Y_p)$ in (D.7) implies

$$\int_\Omega \operatorname{Hess} k_{\vartheta(\omega)}\left(b_*(Y_p), b_*(Y_p)\right) f(\omega;p)\, d\mu = -\int_\Omega b_*(Y_p)k_{\vartheta(\omega)}Y_p l_\omega f(\omega;p)\, d\mu. \tag{D.8}$$

Hereafter, we assume $|Y_p| = 1$ in (D.8). Using inequality (D.2) on the left-hand side of (D.8) yields

$$0 \le \left(1 - \frac{2}{3}C\operatorname{var}_p(\vartheta)\right)|b_*(Y_p)|^2 \le \int_\Omega \operatorname{Hess} k_{\vartheta(\omega)}\left(b_*(Y_p), b_*(Y_p)\right) f(\omega;p)\, d\mu. \tag{D.9}$$

On the other hand, the absolute value of the right-hand side of (D.8), can be bounded as

$$\left|\int_\Omega b_*(Y_p)k_{\vartheta(\omega)}Y_p l_\omega f(\omega;p)\, d\mu\right| \le |b_*(Y_p)|\sqrt{\operatorname{var}_p(\vartheta)}\,\sqrt{\lambda_p}. \tag{D.10}$$

To establish (D.10), define

$$\alpha(\omega,p) = b_*(Y_p)k_{\vartheta(\omega)} = \left\langle\, b_*(Y_p), \operatorname{grad} k_{\vartheta(\omega)}(b(p))\,\right\rangle$$

and let $\beta(\omega,p) = Y_p l_\omega$. Then, we have

$$\left|\int_\Omega \alpha(\omega,p)\beta(\omega,p)f(\omega;p)\, d\mu\right| \le \sqrt{\int_\Omega \alpha(\omega,p)^2 f(\omega;p)\, d\mu}\,\sqrt{\int_\Omega \beta(\omega,p)^2 f(\omega;p)\, d\mu} \tag{D.11}$$

$$\le |b_*(Y_p)|\sqrt{\operatorname{var}_p(\vartheta)}\sqrt{I\left(Y_p, Y_p\right)} \tag{D.12}$$

$$\le |b_*(Y_p)|\sqrt{\operatorname{var}_p(\vartheta)}\sqrt{\lambda_p}. \tag{D.13}$$

In (D.11), we used the Cauchy-Schwarz inequality. To establish (D.12), we used the definition of the Fisher information form (4.8) on the 2nd factor. To bound the 1st factor, notice that

$$\alpha(\omega, p) \leq |b_*(Y_p)| \left|\mathrm{grad} k_{\vartheta(\omega)}(b(p))\right|,$$

by the Cauchy-Schwarz inequality. Now, use (D.1) and the definition of variance of ϑ given in (4.10). Inequality (D.13) follows from (4.9). Inequalities (D.9) and (D.10) imply

$$\left(1 - \frac{2}{3} C \mathrm{var}_p(\vartheta)\right) \sqrt{\sigma_p} \leq \sqrt{\mathrm{var}_p(\vartheta)} \sqrt{\lambda_p}, \tag{D.14}$$

by using the definition of σ_p in (4.11). If $C = 0$, then we have $\mathrm{var}_p(\vartheta) \geq 1/\eta_p$, where $\eta_p = \lambda_p/\sigma_p$. If $C > 0$, then squaring both sides of (D.14) yields $g\left(\mathrm{var}_p(\vartheta)\right) \leq 0$, where $g(t)$ denotes the quadratic polynomial

$$g(t) = \frac{4}{9} C^2 t^2 - \left(\frac{4}{3} C + \eta_p\right) t + 1.$$

Thus, $\mathrm{var}_p(\vartheta) \in [t_1, t_2]$, where $t_1 \leq t_2$ denote the two real roots of $g(t)$. In particular,

$$\mathrm{var}_p(\vartheta) \geq t_1 = \frac{4C + 3\eta_p - \sqrt{\eta_p\left(9\eta_p + 24C\right)}}{\frac{8}{3} C^2} \,\square$$

Bibliography

[1] K. Abed-Meraim, P. Loubaton, and E. Moulines, "A subspace algorithm for certain blind identification problems," *IEEE Transactions on Information Theory*, vol. 43, no. 2, pp. 499–511, March 1997.

[2] K. Abed-Meraim and Y. Hua, "Blind identification of multi-input multi-output system using minimum noise subspace," *IEEE Transactions on Signal Processing*, vol.45, no. 1, January 1997.

[3] S. Andersson, M. Millnert, M. Viberg, and B. Wahlberg, "An adaptive array for mobile communications," *IEEE Transactions on Vehicular Technology*, vol. 40, pp. 230–236, January, 1991.

[4] V. Barroso, J. M. F. Moura, and J. Xavier, "Blind array channel division multiple access (AChDMA) for mobile communications," *IEEE Transactions on Signal Processing*, vol. 46, pp. 737–752, March 1998.

[5] A. Belouchrani, K. Abed-Meraim, J. F. Cardoso, and E. Moulines, "A blind source separation technique using second-order statistics," *IEEE Transactions on Signal Processing*, vol. 45, no. 2, pp. 434-444, February 1997.

[6] C. Berenstein and E. Patrick, "Exact deconvolution for multiple convolution operators: an overview plus performance characterizations for imaging sensors," *IEEE Proceedings*, vol. 78, pp. 723–734, April 1990.

[7] W. Boothby. *An Introduction to Differentiable Manifolds and Riemannian Geometry*. 2nd ed., New York: Academic Press, 1986.

[8] R. S. Bucy, J. M. F. Moura and S. Leung, "Geometrical methods for direction determination," in *Proceedings of the NATO Advanced Study Institute on Acoustic Signal Processing for Ocean Exploration*, pp. 187–191, Madeira, Portugal 1992.

[9] J. F. Cardoso and A. Souloumiac, "Blind beamforming for non-gaussian signals," *IEE Proc.-F, Radar & Signal Processing*, vol. 140, pp. 362–370, December 1993.

[10] T. Chang, "Spherical regression," *The Annals of Statistics*, vol. 14, no. 3, pp. 907–924, 1986.

[11] A. Chevreuil and P. Loubaton, "On the use of conjugate cyclostationarity: a blind second-order multi-user equalization method", in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)*, vol. 4, pp. 2441–2445, 1996.

[12] A. Chevreuil and P. Loubaton, "MIMO blind second-order equalization method and conjugate cyclostationarity," *IEEE Transactions on Signal Processing*, vol. 47, no. 2, pp. 572–578, February 1999.

[13] A. Edelman, T. Arias, and S. Smith, "The geometry of algorithms with orthogonality constraints," *SIAM Journal on Matrix Analysis and Applications*, vol. 20, no. 2, pp. 303–353, 1998.

[14] D. Forney, "Minimal bases of rational vector spaces with applications to multivariable linear systems," *SIAM Journal on Control*, vol. 13, no. 3, pp. 493–520, May 1975.

[15] G. Forney, Jr., "Maximum-likelihood sequence estimation of digital sequences in the presence of intersymbol interference," *IEEE Transactions on Information Theory*, vol. 8, no. 3, pp. 363–378, May 1972.

[16] S. Gallot, D.Hulin, and J. Lafontaine. *Riemannian Geometry*. 2nd ed., Springer-Verlag, 1990.

[17] G Giannakis, Y. Hua, P. Stoica, L. Tong, eds. *Signal Processing Advances in Wireless & Mobile Communications. Volume I: Trends in Channel Estimation and Equalization*, Prentice-Hall, 2001.

[18] G Giannakis, Y. Hua, P. Stoica, L. Tong, eds. *Signal Processing Advances in Wireless & Mobile Communications. Volume II: Trends in Single- and Multi-User Systems*, Prentice-Hall, 2001.

[19] L. Godara, ed. *Handbook of Antennas in Wireless Communications*. CRC Press, 2002.

[20] L. Godara, "Applications of antenna arrays to mobile communications. Part I: performance improvement, feasibility, and system considerations," *Proceedings IEEE*, vol. 85, no. 7, pp. 1031–1060, July 1997.

[21] L. Godara, "Applications of antenna arrays to mobile communications. Part II: beamforming direction-of-arrival considerations," *Proceedings IEEE*, vol. 85, no. 8, pp. 1195–1245, August 1997.

[22] J. Gorman and A. Hero, "Lower bounds for parametric estimation with constraints," *IEEE Transactions on Information Theory*, vol. 26. no. 6, pp. 1285–1301, November 1990.

[23] G. Golub and C. van Loan. *Matrix Computations*. 2nd ed. John Hopkins University Press, 1993.

[24] A. Gorokhov and P. Loubaton, "Subspace-based techniques for blind separation of convolutive mixtures with temporally correlated sources," *IEEE Transactions on Circuits and Systems – I: Fundamental Theory and Applications*, vol. 44, no. 9, pp. 813–820, September 1997.

[25] B. Halder, B. Ng, A. Paulraj and T. Kailath, "Unconditional naximum likelihood approach for blind estimation of digital signals", in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'96)*, vol. 2, pp. 1081–1084, 1996.

[26] S. Haykin, ed. *Unsupervised Adaptive Filtering. Volume I: Blind Source Separation.* John Wiley & Sons, 2000.

[27] H. Hicks, "A Cramér-Rao type lower bound for estimators with values in a manifold," *Journal of Multivariate Analysis*, no. 38, pp. 245–261, 1991.

[28] R. Horn and C. Johnson. *Matrix Analysis*. Cambridge University Press, 1999.

[29] Y. Hua, "Fast maximum likelihood for blind identification of multiple FIR channels," *IEEE Transactions on Signal Processing*, vol. 44, pp. 661–672, March 1996.

[30] Y. Hua, Y. Xiang, K. Abed-Meraim, "Blind identification of colored signals distorted by FIR channels," in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'00)*, vol. 5, pp. 3124–3127, 2000.

[31] Y. Hua and J. Tugnait, "Blind identifiability of FIR-MIMO systems with colored input using second-order statistics," *IEEE Signal Processing Letters*, vol. 7, no. 12, December 2000.

[32] J. Jost. *Riemannian Geometry and Geometric Analysis*. 2nd ed. Springer-Verlag, 1998.

[33] R. Kass and P. Vos. *Geometrical Foundations of Asymptotic Inference.* John Wiley & Sons, 1997.

[34] S. Kay. *Fundamentals of Statistical Signal Processing: Estimation Theory*. Prentice-Hall, 1993.

[35] H. Krim and M. Viberg, "Two decades of array signal processing research," *IEEE Signal Processing Magazine*, vol. 13, no. 4, pp. 47–94, July 1996.

[36] J. Lee. *Riemannian manifolds*. Springer-verlag, 1997.

[37] E. L. Lehmann. *Elements of Large-Sample Theory.* Springer-Verlag, 1999.

[38] C. Ma, Z. Ding and S. Yau, "A two-stage algorithm for MIMO blind deconvolution of non-stationary colored signals," *IEEE Transactions on Signal Processing*, vol. 48, no. 4, pp. 1187–1102, April 2000.

[39] J. Magnus and H. Neudecker. *Matrix Differential Calculus with Applications to Statistics and Econometrics*, revised edition. Wiley Series in Probability and Statistics, 1999.

[40] J. Marsden and M. Hoffman. *Basic Complex Analysis.* 2nd ed., W.H. Freeman and Company, 1987.

[41] T. Marzetta, "A simple derivation of the constrained multiple parameter Cramér-Rao bound," *IEEE Transactions on Signal Processing*, vol. 41, no. 6, pp. 2247–2249, June 1993.

[42] D. Morgan, J. Benesty, and M. Sondhi, "On the evaluation of estimated impulse responses," *IEEE Signal Processing Letters*, vol. 5, no. 7, pp. 174–176, July 1998.

[43] E. Moulines, P. Duhamel, J. Cardoso, and S. Mayrargue, "Subspace methods for the blind identification of multichannel FIR filters," *IEEE Transactions on Signal Processing*, vol. 43, no. 2, pp. 516–525, February 1995.

[44] J. Oller and J. Corcuera, "Intrinsic analysis of statistical estimation," *The Annals of Statistics*, vol. 23, no. 5, pp. 1562–1581, 1995.

[45] A. Paulraj and C. B. Papadias, "Array processing in mobile communications," in *Handbook of Signal Processing*, CRC Press, 1997.

[46] J. Proakis, *Digital Communications.* 3rd ed., McGraw-Hill, 1995.

[47] C. Rao and M. Rao. *Matrix Algebra and Its Applications to Statistics and Econometrics.* World Scientific Publishing, 1998.

[48] C. Rao. *Linear Statistical Inference and Its Application.* New York: Wiley, 1973.

[49] M. João Rendas and J. M. F. Moura, "Ambiguity in radar and sonar," *IEEE Transactions on Signal Processing*, vol. 46, no. 2, February 1998.

[50] R. Roy. *ESPRIT, Estimation of Signal Paramaters via Rotational Invariance Techniques.* PhD thesis, Stanford University, CA, 1987.

[51] L. Scharf. *Statistical Signal Processing: Detection, Estimation, and Time Series Analysis.* Addison-Wesley Publishing Company, 1991.

[52] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Transactions Antennas Propagation*, vol. AP-34, no. 3, pp. 276–280, March 1986.

[53] J. Shao. *Mathematical Statistics.* Springer-Verlag, 1998.

[54] S. Smith, "Intrinsic Cramér-Rao bounds and subspace estimation accuracy," *1st IEEE Workshop on Sensor Arrays and Multichannel Signal Processing*, Boston, MA, 2000.

[55] P. Stoica and B. Ng, "On the Cramér-Rao bound under parametric constraints," *IEEE Signal Processing Letters*, vol. 5, no. 7, pp. 177–179, July 1998.

[56] S. Talwar, M. Viberg, and A. Paulraj, "Blind estimation of multiple co-channel digital signals using an antenna array," *IEEE Signal Processing Letters*, vol. 1, no. 2, pp. 29–31, February 1994.

[57] S. Talwar, M. Viberg, and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array – part I: algorithms," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1184–1197, May 1996.

[58] S. Talwar and A. Paulraj, "Blind separation of synchronous co-channel digital signals using an antenna array – part II: performance analysis," *IEEE Transactions on Signal Processing*, vol. 45, no. 3, pp. 706–718, March 1997.

[59] L. Tong, G. Xu, and T. Kailath, "A new approach to blind identification and equalization of multipath channels," in *Proceedings 25th Asilomar Conference* (Pacific Grove, CA), pp. 856–860, 1991.

[60] L. Tong, G. Xu, and T. Kailath, "Fast blind equalization via antenna array," *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'93)*, vol. 3, pp. 272–275, 1993.

[61] L. Tong, G. Xu, and T. Kailath, "Blind identification and equalization based on second-order statistics: a time domain approach," *IEEE Transactions on Information Theory*, vol. 40, no. 2, pp. 340–349, March 1994.

[62] L. Tong and S. Perreau, "Multichannel blind identification: from subspace to maximum likelihood methods," *Proceedings IEEE*, vol. 86, pp. 1951–1968, October 1998.

[63] H. Van Trees. *Detection, Estimation and Modulation Theory, Part I.* New York: Wiley, 1968.

[64] J. K. Tugnait, "On linear predictors for MIMO channels and related blind identification and equalization," *IEEE Signal Processing Letters*, vol. 5, pp. 289–291, November 1998.

[65] J. K. Tugnait and B. Huang, "Blind channel estimation and equalization of multiple-input multiple-output channels," in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'99)*, vol. 5, pp. 2707–2710, 1999.

[66] J. K. Tugnait and B. Huang, "Multi-step linear predictors-based blind equalization of multiple-input multiple-output channels," in *Proceedings IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'99)*, vol. 5, pp. 2949–2952, 1999.

[67] A. van der Veen and A. Paulraj, "An analytical constant modulus algorithm," *IEEE Transactions on Signal Processing*, vol. 44, no. 5, pp. 1136–1155, May 1996.

[68] A. Van der Veen, S. Talwar, and A. Paulraj, "A subspace approach to blind space-time signal processing for wireless communication systems," *IEEE Transactions on Signal Processing*, vol. 45, no. 1, pp. 173–190, January 1997.

[69] B. Van Veen and K. Buckley, "Beamforming: a versatile approach to spatial filtering," *Signal Processing Magazine*, pp. 4–24, April 1988.

[70] A. Viterbi. *CDMA: Principles of Spread-Spectrum Communications.* Addison-Wesley, 1995.

[71] J. Xavier, V. Barroso, and J. M. F. Moura, "Closed form blind channel identification and source separation in SDMA systems through correlative coding", *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1506–1517, October 1998.

[72] J. Xavier and V. Barroso, "Correlative Coding Approach for Blind Identification of MIMO Systems," *Proceedings of the IEEE Signal Processing Workshop on Signal Processing Advances in Wireless Communications (SPAWC'99)*, pp. 263–266, Annapolis, EUA, May 1999.

[73] J. Xavier and V. Barroso, "Blind source separation, ISI cancelation and carrier phase recovery in SDMA systems for mobile communications", *Wireless Personal Communications*, special issue on *Wireless Broadband Communications*, vol. 10, pp. 36-76, Kluwer Academic Publishers, June 1999.

[74] J. Xavier and V. Barroso, "Blind Identification of MIMO Systems Based on Source Correlative Filtering", *Proceedings of the 2000 IEEE International Symposium on Information Theory (ISIT'2000)*, Sorrento, Italy, July 2000.

[75] J. Xavier, V. Barroso, and J. M. F. Moura, "Closed-form Correlative Coding (CFC_2) Blind Identification of MIMO Channels: Isometry Fitting to Second-Order Statistics," *IEEE Transactions on Signal Processing*, vol. 49, no. 5, pp. 1073 – 1086, May 2001.

[76] J. Xavier and V. Barroso, "Performance Analysis of a Second Order Statistics Based Solution for the MIMO Channel Identification Problem," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'01)*, vol. 4 , pp. 1241–1244, Salt Lake City, EUA, May 2001.

[77] J. Xavier and V. Barroso, "Second-Order Blind MIMO Channel Identification Based on Correlative Coding: Asymptotic Performance Analysis," *Proceedings of the 2001 IEEE International Symposium on Information Theory*, Washington, EUA, June 2001.

[78] J. Xavier and V. Barroso, "Intrinsic Distance Lower Bound for Unbiased Estimators on Riemannian Manifolds," *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP'02)*, Orlando, Florida, EUA, May 2002.

[79] J. Xavier and V. Barroso, "Intrinsic Variance Lower Bound for Inference Problems on Riemannian Manifolds," *Proceedings of the* 10*th Mediterranean Conference on Control and Automation (MED'2002)*, Lisbon, Portugal, July 2002.

[80] G. Xu, H. Liu, L. Tong, and T. Kailath, "A deterministic approach to blind channel identification," *IEEE Transactions on Signal Processing*, vol. 43, no. 12, pp. 2982–2993, December 1995.