Eigenvalues of Large Random Matrices with Dependent Entries and Strong Solutions of SDEs

OLIVER PFAFFEL

Dissertation



Technische Universität München Fakultät für Mathematik

Technische Universität München Lehrstuhl für Mathematische Statistik

Eigenvalues of Large Random Matrices with Dependent Entries and Strong Solutions of SDEs

Oliver Pfaffel

Vollständiger Abdruck der von der Fakultät für Mathematik der Technischen Universität München zur Erlangung des akademischen Grades eines

Doktors der Naturwissenschaften (Dr. rer. nat.)

genehmigten Dissertation.

Vorsitzender:		UnivProf. Dr. Folkmar Bornemann
Prüfer der Dissertation:	1.	UnivProf. Dr. Robert Stelzer Universität Ulm
	2.	Prof. Richard A. Davis, Ph.D. Columbia University, New York, USA
	3.	UnivProf. Dr. Claudia Klüppelberg

Die Dissertation wurde am 05.07.2012 bei der Technischen Universität München eingereicht und durch die Fakultät für Mathematik am 26.10.2012 angenommen.

Zusammenfassung

Die vorliegende Arbeit ist in drei Abschnitte gegliedert. Die beiden ersten Abschnitte befassen sich mit den Eigenwerten der empirischen Kovarianzmatrix eines multivariaten linearen Prozesses, basierend auf der Annahme dass sowohl die Anzahl der Beobachtungen als auch die Dimension des Prozesses gegen unendlich konvergieren. Der letzten Abschnitt beschäftigt sich mit der Modellierung stochastischer Kovarianzmatrizen in stetiger Zeit.

Zu Beginn des ersten Abschnitts der Arbeit wird angenommen dass die Komponenten des stationären linearen Prozesses durch unabhängige Kopien eines univariaten linearen Prozesses gegeben sind. Die Randverteilung des univariaten Prozesses ist regulär variierend mit Index $\alpha \in (0, 4)$. Es wird gezeigt dass der Punktprozess der Eigenwerte der empirischen Kovarianzmatrix in Verteilung gegen einen Poisson-Punktprozess konvergiert, dessen Intensitätsmaß von α und der Summe der quadrierten Koeffizienten des zugrunde liegenden linearen Prozesses abhängt. Dies impliziert sofort die gemeinsame Konvergenz der *k*-größten Eigenwerte von XX^{T} . Das erwähnte Resultat wird im Folgenden für Modelle mit stochastischen Koeffizienten verallgemeinert. Falls die Komponenten des linearen Prozesses abhängig sind, lassen sich obere und untere Schranken für die Spektralnorm von XX^{T} angeben.

Der zweite Abschnitt befasst sich mit der Spektralverteilung von XX^{T} in dem Fall dass der lineare Prozess unabhängige Komponenten und eine Randverteilung mit endlichen vierten Momenten besitzt. In diesem Fall konvergiert die Spektralverteilung der empirischen Kovarianzmatrix gegen ein deterministisches Wahrscheinlichkeitsmaß, welches von der Spektraldichte des linearen Prozesses abhängt. Das deterministische Maß lässt sich eindeutig durch eine implizite Gleichung für seine Stieltjes-Transformierte charakterisieren. Das gleiche Resultat erhält man falls die Matrix X zeilenweise durch einen einzigen univariaten linearen Prozess konstruiert wird.

Im letzten Abschnitt wird die Existenz und Eindeutigkeit von globalen starken Lösungen für eine Klasse von Stochastischen Differentialgleichungen auf den symmetrisch positiv definiten Matrizen gezeigt. Diese Klasse enthält Wishart Prozesse und allgemeinere affine Sprungprozesse deren Diffusionskoeffizient durch den *a*-ten Matrixexponenten des Prozesses gegeben ist, wobei $0.5 \le a < 1$.

Abstract

This thesis consists of three major topics. The first two parts study the eigenvalues of a large sample covariance matrix XX^{T} of a multivariate linear process. Corresponding to a statistical setting where the sample size and the dimension of the process are large, we give asymptotic results in case where the number of rows and the number of columns of *X* both tend to infinity. The third part concerns the modeling of stochastic covariance matrices in time.

In the first part we begin with the assumption that the components of the stationary linear process are independent copies of some univariate linear process with a marginal distribution that is regularly varying with tail index $\alpha \in (0, 4)$. It is shown that the point process of the eigenvalues of the sample covariance matrix converges in distribution to a Poisson point process with intensity measure depending on α and the sum of the squared coefficients of the underlying linear process. This implies the joint convergence of the *k*-largest eigenvalues. The result will further be generalized to random coefficient models. In the case where the components of the linear process are dependent, we give an asymptotic upper and lower bound for the spectral norm of XX^{T} .

The second part studies the limiting spectral distribution of XX^{T} in case where the linear process has independent components with marginal distributions which have finite fourth moments. If the ratio of the number of rows and columns of *X* tends to a positive finite constant *y*, then the spectral distribution of the sample covariance matrix converges to a non-random distribution which only depends on *y* and the spectral density of the underlying linear process. The limiting distribution is uniquely determined by an implicit equation for its Stieltjes transform. Moreover, it is shown that the same result applies if the linear process is replaced by a univariate linear process which fills the matrix *X* row-wise.

In the final part of this thesis, we show the existence of unique global strong solutions of a class of stochastic differential equations on the cone of symmetric positive definite matrices. Our result includes Wishart processes and more general affine jump diffusion processes where the diffusion coefficient is given by the *a*-th positive semidefinite power of the process itself with $0.5 \le a < 1$.

Acknowledgements

First of all, I want to thank my supervisors Robert Stelzer and Richard Davis for their continuing support during my PhD studies. Roberts guidance from my undergraduate through my PhD studies is greatly appreciated. He gave me a lot of advice and constant encouragement. Richard is a terrific mentor who knows to combine work and joy in a successful way. He gave me very helpful advice with regard to academia and life in general.

I also want to thank Claudia Klüppelberg for being in my examination committee and creating a good working environment at TUM. My position at her department gave me the possibility to attend various conferences around the world and to meet many interesting people along the way.

I appreciate the financial support of the Technische Universität München - Institute for Advanced Study, funded by the German Excellence Initiative, and the International Graduate School of Science and Engineering. Their accountants did a tremendous job at handling my sometimes extensive applications of refund.

Finally, I'm grateful for having a great family and awesome friends. Without them the last couple of years would have been very poor.

Contents

1.	Introduction					
	1.1.	Backgi	round and motivation	1		
	1.2.	Outline	e of the thesis	3		
I.	Ex	treme	Singular Values of Heavy-tailed Random Matrices	7		
2.	Limit Theory for the largest eigenvalues of sample covariance matrices					
	2.1.	Introdu	iction	9		
	2.2. Main results on heavy-tailed random matrices with dependent entries .					
		2.2.1.	A first result on the largest eigenvalue	12		
		2.2.2.	Examples and discussion	15		
		2.2.3.	Extension to random coefficient models	16		
	2.3.	Proofs	and auxiliary results	17		
		2.3.1.	A large deviation result and its consequences	17		
		2.3.2.	Convergence in Operator Norm	21		
		2.3.3.	Extremes of diag (XX^T)	25		
		2.3.4.	Proof of Theorem 2.1	29		
		2.3.5.	Proof of Theorem 2.2	31		
3.	Obs	ervatio	ns with finite variance but infinite fourth moment	35		
	ction and main results	35				
	3.2.	Proof		38		
		3.2.1.	Approximation of S by its diagonal \ldots \ldots \ldots \ldots \ldots \ldots	38		
		3.2.2.	Point process convergence and the proof of Theorem 3.1	42		

		3.2.3. Proof of Theorem 3.2	47				
4.	4. Random matrices with strongly dependent rows and columns						
	4.1.	Introduction and main results	49				
	4.2.	Dependence of successive rows	52				
	4.3.	Proof of the theorem	55				
II.	Sp	ectral Distribution of Light-tailed Random Matrices	61				
5.	Sam	ple covariance matrices of linear processes	63				
	5.1.	Introduction and main result	63				
	5.2.	Proofs	67				
	5.3.	Illustrative examples	75				
		5.3.1. Autoregressive moving average processes	75				
		5.3.2. Fractionally integrated ARMA processes	78				
6.	A ne	ew random matrix model with dependent rows and columns	79				
	6.1.	Introduction	79				
	6.2.	A new random matrix model	82				
	6.3.	Proof of Theorem 6.1	83				
	6.4.	Sketch of an alternative proof of Theorem 6.1	94				
III.	Sti	rong Solutions of Stochastic Differential Equations	97				
7.	On s	strong solutions for positive definite jump diffusions	99				
	7.1.	Introduction	99				
	7.2.	Notation and general set-up	101				
	7.3.	Statement of the main results	102				
		7.3.1. Wishart diffusions with jumps	102				
		7.3.2. The general SDE and existence result	104				
		7.3.3. Positive definite extensions of generalised Cox-Ingersoll-Ross processes					
		and GARCH diffusions	105				
	7.4.	Proofs	107				
		7.4.1. McKean's argument	108				
		7.4.2. Proof of Theorem 7.1	109				
	7.5.	Conclusion	113				
Bil	bliog	raphy	115				

CHAPTER 1

Introduction

1.1. Background and motivation

A *random matrix ensemble* is a sequence of matrices with increasing dimensions and randomly distributed entries. *Random Matrix Theory* (RMT) studies the asymptotic spectrum, e.g., limiting eigenvalues and eigenvectors, of random matrix ensembles.

The original motivation to study random matrices came from applications in physics, cf. the introduction of [81] and the references therein. In *nuclear physics* one is interested in the properties of the energy levels of heavy nuclei. The energy levels of a such a system are described by the eigenvalues of some Hermitian operator in an infinite dimensional Hilbert space. In many cases this operator is not known, and even if it was, it would be too difficult to determine its eigenvalues and -vectors. Thus Eugene Wigner [112] proposed to approximate this operator by a large $n \times n$ Hermitian random matrix H which would, in the large n limit, describe the energy levels of some general heavy nucleus.

Another motivation to study large random matrices comes from *mathematical statistics*. Often one tries to reduce the dimensionality of a data set while preserving as much of the variation in the data as possible. To this end, *Principal Component Analysis* (PCA) [71] corresponds to a linear transformation of the data to a new set of variables, the *principal components*, which are ordered such that the first few retain most of the variation. Therefore one obtains a lower dimensional representation of the data by retaining only the first few principal components. If the data is observed from a multivariate time series and collected in some rectangular $p \times n$ matrix X, then the empirical variances of the first k-principal components are exactly the klargest eigenvalues of the $p \times p$ sample covariance matrix $\frac{1}{n}XX^{T}$. In this context, p refers to

1. Introduction

the dimension of the observed time series and *n* to the size of the sample. In modern statistical applications one often has access to a vast amount of high-dimensional data, thus *p* and *n* may be quite large. A reasonable asymptotic framework accounts for this by assuming that both *p* and *n* tend to infinity. This motivates to study matrices of the form XX^{T} in the context of Random Matrix Theory.

Of course, Hermitian and sample covariance matrices are not the only examples of matrices studied in RMT. Toeplitz, Circulant and Hankel matrices [27] have been studied as well. Furthermore, there exist results for the complex eigenvalues of non-symetric rectangular random matrices, cf. [104] and the references therein. However, we will not further discuss matrices of this kind.

While we have explained how applications in physics naturally lead to the study of Hermitian random matrices, and applications in statistics to the study of sample covariance matrices, the joint distribution of the matrix entries is another feature which differentiates random matrix ensembles. It is frequently assumed that the matrix entries are independent (subject to symmetry conditions) and identically distributed in order to guarantee the mathematical tractability of the model. As an example, let A be a real symmetric matrix such that all entries on and above the diagonal are independent and identically distributed (iid). A matrix of this type is called Wigner matrix in honor of the contributions of Eugene Wigner to RMT. The joint density of the entries of A is just the product density. Furthermore, if the matrix A is orthogonally *invariant*, i.e., A and OAO^{T} have the same distribution for any orthogonal matrix O, then it is possible to compute the joint density of the eigenvalues of A, see, for instance, [2]. This is then a starting point to analyze the limiting behaviour of the eigenvalues of A. An important example of an orthogonally invariant ensemble is the *Gaussian Orthogonal Ensemble* (GOE) which consists of symmetric random matrices where the entries on and above the diagonal are iid Gaussian distributed. Many of the results shown for the GOE could later be generalized to hold for Wigner matrices with more general non-Gaussian distributions satisfying some moment conditions, see e.g. [106] and [48].

Sample covariance matrices of the form XX^{T} , where X is some rectangular random matrix, can be treated similarly as GOE matrices if they are orthogonally invariant. This is again the case when X has iid Gaussian entries. The matrix XX^{T} is then said to have a Wishart distribution. If, for example, the columns of X are given by the observations of some *multivariate linear process*, i.e., a weighted sum (or series) of iid random vectors, then XX^{T} is not orthogonally invariant in general. Therefore it might be hard or even impossible to analytically obtain the joint density of the eigenvalues of a sample covariance matrix. Thus other techniques have to be employed. Many results in this thesis give results for this kind of matrices. We consider matrices with entries that have a light-tailed marginal distribution, i.e., a finite fourth moment, as well as matrices with a heavy-tailed marginal distribution which has an infinite fourth mo-

ment.

When the random matrix ensemble is specified by the symmetry class of the matrices and the joint distribution of its entries, one is able to study the eigenvalues of this ensemble. Clearly, when the dimension of a matrix goes to infinity then so does the number of its eigenvalues, so the vector of all eigenvalues is not an element of a fixed dimensional space and can therefore not be studied conveniently. If *A* is symmetric, then it has real eigenvalues, so one possibility is to investigate its *spectral distribution* F^A . For any subset *B* of the real line, $F^A(B)$ counts the fraction of eigenvalues of *A* which are located in the set *B*. Thus F^A is a random measure. Typically, F^A converges, suitably normalized, to a non-random limiting measure that is universal to a large class of random matrix ensembles [9]. For example, if *A* is a Wigner matrix with a marginal distribution which has a finite variance, then F^A converges to the semi-circle distribution. Sample covariance matrices of the form XX^T , where *X* has iid entries with finite variance, converge to the Marčenko-Pastur-law, see Chapter 2 for details.

Another possibility is to investigate some characteristic eigenvalue of *A*. For example, if *A* is from the GOE, then the normalized largest eigenvalue of *A* converges to the Tracy-Widom-distribution [107]. The same result holds true for the largest eigenvalue of a sample covariance matrix, cf. the introduction of Chapter 2. In this thesis we study the global behaviour of the eigenvalues of light-tailed sample covariance matrices via their spectral distribution as well as the extreme eigenvalues in case the matrix entries are heavy-tailed. Further possible topics as e.g. the eigenvectors of a large random matrix are not a focus of this thesis, for more information see [105] and the references therein.

The upcoming section gives an outline of this thesis. For a more specific overview on the past and current literature on Random Matrix Theory we refer to the individual introductions of Chapters 2 to 6. Furthermore, a comprehensive introduction to Random Matrix Theory can be found in the textbooks [2], [9], [81]; for a brief overview we recommend the survey articles [45], [70] and [108]. An introduction to the modelling of covariance matrices in continuous time is given in Chapter 7.

1.2. Outline of the thesis

Part I contains various results for the extreme eigenvalues of a sample covariance matrix of dependent observations with heavy-tails.

In Chapter 2 we study the joint limiting distribution of the *k* largest eigenvalues of a $p \times p$ sample covariance matrix XX^T based on a large $p \times n$ observation matrix *X*. The rows of *X* are given by independent copies of a linear process, $X_{it} = \sum_j c_j Z_{i,t-j}$, with regularly varying noise (Z_{it}) with tail index $\alpha \in (0, 2)$. It is shown, as $n \to \infty$ and $p \to \infty$ at a suitable rate, that

the point process based on the eigenvalues of XX^{T} converges in distribution to a Poisson point process with intensity measure depending on α and $\sum c_j^2$. This result is extended to random coefficient models where the coefficients of the linear processes (X_{it}) are given by $c_j(\theta_i)$, for some ergodic sequence (θ_i) , and thus vary in each row of X. As a by-product, we obtain a proof of the corresponding result for matrices with iid entries in cases where p/n goes to zero or infinity.

Subsequently, in Chapter 3, the results from Chapter 2 are extended to the case where $\alpha \in [2, 4)$. Apart from the special case $\alpha = 2$, this implies that our observations have a finite variance but an infinite fourth moment. That is, for example, often the case for financial data. Since PCA requires the existence of second moments, this assumption is also more consistent with the goal to derive a theoretical framework for analyzing high-dimensional data via PCA. In this setting, the entries on the diagonal of XX^{T} have a finite mean that has to be subtracted in order to obtain a non-trivial distributional limiting result. Hence we study the eigenvalues of the matrix $XX^{T} - n \sum_{j} c_{j}^{2} EZ_{11}^{2} I_{p}$, where I_{p} denotes the identity matrix.

In Chapter 4 we study the random matrix model $\hat{X} = (\hat{X}_{it})$ with

$$\hat{X}_{it} = \sum_{j,k=-\infty}^{\infty} c_j \theta_k Z_{i-k,t-j},$$
(1.2.1)

where (Z_{it}) is regularly varying with tail index $\alpha \in (0, 4)$. In contrast to the models from Chapter 2 and 3, the matrix \hat{X} has dependent rows and columns. The motivation is that the linear processes used in stochastic modelling typically do not have independent components. If all θ_k in (1.2.1) are zero except θ_0 being equal to one, then \hat{X} has independent rows and this model reduces to the one studied in Chapter 2 and 3. We give an asymptotic lower and upper bound for the largest eigenvalue (i.e., the spectral norm) of $\hat{X}\hat{X}^{\mathsf{T}}$ in the case $\alpha \in (0, 2)$, and for the largest eigenvalue of $\hat{X}\hat{X}^{\mathsf{T}} - nEZ_{11}^2 \sum_j c_j^2 HH^{\mathsf{T}}$ when $\alpha \in [2, 4)$, where $H = (H_{ij}) \in \mathbb{R}^{p \times 3p}$ is given by $H_{ij} = \theta_{p-(j-i)} \mathbf{1}_{\{0 \le j-i \le 2p\}}$.

As for Part II, we shift our attention from heavy to light-tailed matrices with entries which have finite fourth moments, and study the spectral distribution of a sample covariance matrix.

In Chapter 5 we derive the distribution of the eigenvalues of a sample covariance matrix when the data is modelled as a linear process $(X_{i,t})_{t=1,...,n} = (\sum_{j=0}^{\infty} c_j Z_{i,t-j})_{t=1,...,n}$, where $\{Z_{i,t}\}$ are assumed to be independent random variables with finite fourth moments satisfying a Lindeberg-type condition. If the sample size *n* and the number of variables $p = p_n$ both converge to infinity such that $y = \lim_{n\to\infty} n/p_n > 0$, then the empirical spectral distribution of $p^{-1}\mathbf{X}\mathbf{X}^T$ converges to a non-random distribution which only depends on *y* and the spectral density of $(X_{1,t})_{t\in\mathbb{Z}}$. In particular, our results apply to (fractionally integrated) ARMA processes, which we illustrate by some examples.

In Chapter 6 we introduce a random matrix model where the entries are dependent across both rows and columns. More precisely, we investigate $p \times n$ matrices of the form $\tilde{X} =$ $(X_{(i-1)n+t})_{it}$ derived from a single linear process $X_t = \sum_j c_j Z_{t-j}$, where the $\{Z_t\}$ are independent random variables with bounded fourth moments. We show that, when both p and n tend to infinity such that the ratio p/n converges to a finite positive limit, the empirical spectral distribution of $p^{-1}\tilde{X}\tilde{X}^{\mathsf{T}}$ converges almost surely to the same deterministic measure which has been derived in Chapter 5. Thus the matrix $p^{-1}\tilde{X}\tilde{X}^{\mathsf{T}}$ can be used as an approximation to the sample covariance matrix of a high-dimensional process whose components are independent copies of X_t , when only a single component is observed.

As mentioned before, if the matrix X from Part II has iid standard Gaussian entries, then XX^{T} is called Wishart matrix. It is easy to see that, for fixed dimension p and sample size $n \to \infty, \frac{1}{n}XX^{\mathsf{T}}$ converges to the identity matrix. Thus, a Wishart matrix is a consistent estimator of the covariance matrix of iid univariate Gaussian observations for a fixed point in time. In many applications, e.g. derivative pricing and hedging in finance, one needs a stochastic model for the covariance matrix in continuous time. To this end, one replaces the entries of X by independent standard Brownian motions $B_{ij,t}$. Then $X_t X_t^{\mathsf{T}}$ with $X_t = (B_{ij,t})$ is called Wishart process and is an example of a matrix variate process in the cone of symmetric positive definite matrices. In Part III we consider a large class of fixed dimensional covariance matrix models in continuous time which are defined as solutions to a stochastic differential equation (SDE) on the cone of symmetric positive definite matrices. More precisely, we show the existence of unique global strong solutions of this class of SDEs. Our result includes affine jump diffusion processes and therefore considerably extends the known statements concerning Wishart processes, which have been extensively employed in financial mathematics. Moreover, we consider stochastic differential equations where the diffusion coefficient is given by the *a*-th positive semidefinite power of the process itself with 0.5 < a < 1 and obtain existence conditions for them. In the case of a diffusion coefficient which is linear in the process we likewise get a positive definite analogue of the univariate GARCH diffusions.

Part I.

Extreme Singular Values of Heavy-tailed Random Matrices

CHAPTER 2

Limit Theory for the largest eigenvalues of sample covariance matrices with heavy-tails¹

2.1. Introduction

Recently there has been increasing interest in studying *large dimensional data sets* that arise in finance, wireless communications, genetics and other fields. Patterns in these data can often be summarized by the *sample covariance matrix*, as done in multivariate regression and dimension reduction via factor analysis. Therefore, our objective is to study the asymptotic behavior of the eigenvalues $\lambda_{(1)} \ge ... \ge \lambda_{(p)}$ of a $p \times p$ sample covariance matrix XX^T , where the *data matrix* X is obtained from *n* observations of a high-dimensional stochastic process with values in \mathbb{R}^p . Classical results in this direction often assume that the entries of X are independent and identically distributed (iid) or satisfy high moment conditions. Our goal is to weaken the moment conditions by allowing for heavy-tails, and the assumption of independent entries by allowing for dependence within the rows and columns. Potential applications arise in portfolio management in finance, where observations typically have heavy-tails and dependence.

Assuming that the data comes from a multivariate normal distribution, one is able to compute the joint distribution of the eigenvalues $(\lambda_{(1)}, \ldots, \lambda_{(p)})$, see [68]. Under the additional premise that the dimension p is fixed while the sample size n goes to infinity, Anderson [4] obtains a central limit like theorem for the largest eigenvalue. Clearly, it is not possible to derive the joint distribution in a general setting where the distribution of X is not invariant with respect to orthogonal transformations. Furthermore, since in modern applications with

¹This chapter is based on Richard A. Davis, Oliver Pfaffel and Robert Stelzer: Limit Theory for the largest eigenvalues of sample covariance matrices with heavy-tails. Preprint.

large dimensional data sets, p might be of similar or even larger order than n, it might be more suitable to assume that both p and n go to infinity, so Anderson's result may not be a good approximation in this setting. For example, considering a financial index like the S&P 500, the number of stocks is p = 500, whereas, if daily returns of the past 5 years are given, n is only around 1300. In genetic studies, the number of investigated genes p might easily exceed the number of participating individuals n by several orders of magnitude. In this *large n*, *large* p framework results differ dramatically from the corresponding *fixed p*, *large n* results - with major consequences for the statistical analysis of large data sets [69].

Spectral properties of large dimensional random matrices is one of many topics that has become known under the banner *Random Matrix Theory (RMT)*. The original motivation for RMT comes from mathematical physics [44], [112], where large random matrices serve as a finite-dimensional approximation of infinite-dimensional operators. Its importance for statistics comes from the fact that RMT may be used to correct traditional tests or estimators which fail in the 'large *n*, large *p*' setting. For example, Bai et al. [8] gives corrections on some likelihood ratio tests that fail even for moderate *p* (around 20), and El Karoui [46] consistently estimates the spectrum of a large dimensional covariance matrix using RMT. Thus statistical considerations will be our motivation for a random matrix model with heavy-tailed and dependent entries.

Before describing our results, we will give a brief overview of some of the key results from RMT for real-valued sample covariance matrices XX^{T} . A more detailed account on RMT can be found, for instance, in the textbooks [2], [9], or [81]. Here X is a real $p \times n$ random matrix, and p and n go to infinity simultaneously. Let us first assume that the entries of X are iid with variance 1. Results on the *global behavior of the eigenvalues* of XX^{T} mostly concern the *spectral distribution*, that is the random probability measure of its eigenvalues $p^{-1} \sum_{i=1}^{p} \epsilon_{n^{-1}\lambda_{(i)}}$, where ϵ denotes the Dirac measure. The spectral distribution converges, as $n, p \to \infty$ with $p/n \to \gamma \in (0, 1]$, to a deterministic measure with density function

$$\frac{1}{2\pi x\gamma}\sqrt{(x_+-x)(x-x_-)}\mathbf{1}_{(x_-,x_+)}(x), \quad x_{\pm} \coloneqq (1 \pm \sqrt{\gamma})^2,$$

where **1** denotes the indicator function. This is the so called *Marčenko–Pastur law* [77], [110]. One obtains a different result if XX^{T} is perturbed via an affine transformation [77], [86]. Based on these results, [89] treats the case where the rows of *X* are given by independent copies of a linear process. Apart from a few special cases, the limiting spectral distribution is not known in a closed form if the entries of *X* are not independent.

Although the eigenvalues of XX^{T} offer various interesting local properties to be studied, we will only focus on the joint asymptotic behavior of the *k* largest eigenvalues $(\lambda_{(1)}, \ldots, \lambda_{(k)})$, $k \in \mathbb{N}$. This is motivated from a statistical point of view since the variances of the first *k* principal components are given by the *k* largest eigenvalues of the covariance matrix. Geman [51] shows, assuming that the entries of *X* are iid and have finite fourth moments, that $n^{-1}\lambda_{(1)}$

converges to $x_+ = (1 + \sqrt{\gamma})^2$ almost surely if $p/n \to \gamma \in (0, \infty)$. Moreover, if the entries of *X* are iid standard Gaussian, Johnstone [69] shows that

$$\frac{\sqrt{n} + \sqrt{p}}{\sqrt[3]{\frac{1}{\sqrt{n}} + \frac{1}{\sqrt{p}}}} \left(\frac{\lambda_{(1)}}{\left(\sqrt{n} + \sqrt{p}\right)^2} - 1 \right) \stackrel{D}{\longrightarrow} \xi,$$

where ξ follows the *Tracy–Widom distribution* with $\beta = 1$. Soshnikov [99] extends this to more general symmetric non-Gaussian distributions if the matrix X is nearly square, and obtains a similar result for the joint convergence of the k largest eigenvalues. The Tracy–Widom distribution first appeared as the limit of the largest eigenvalue of a Gaussian Wigner matrix [107]. Péché [87] shows that the assumption of Gaussianity in Johnstone's result can be replaced by the assumption that the entries of X have a symmetric distribution with sub-Gaussian tails, and she allows for γ being zero or infinity.

There exist results on extreme eigenvalues of XX^{T} which include dependence within the rows or columns of X, but most of them are only valid if X has complex-valued entries such that its real as well as its complex part have a non-zero variance. A notable exception, where the real-valued case is considered, is [19]. They assume that the rows of X are normally distributed with a covariance matrix which has exactly one eigenvalue not equal to one.

In contrast to the light tailed case described above, there exist only a handful of articles dealing with sample covariance matrices XX^{T} obtained from heavy-tailed observations. All these results only apply to matrices X with iid entries. Belinschi et al. [14] compute the limiting spectral distribution of sample covariance matrices based on observations with infinite variance. Regarding the *k*-largest eigenvalues, Soshnikov [100] gives the weak limit in case the underlying distribution of the matrix entries is Cauchy. Biroli et al. [18] argued, using heuristic arguments and numerical simulations, that Soshnikov's result extends to general distributions with regularly varying tails with index $0 < \alpha < 4$. A mathematically rigorous proof of this claim followed by Auffinger et al. [7].

We extend the previous results for $0 < \alpha < 2$ by allowing for dependent entries. More specifically, the rows of X are given by independent copies of some linear process. Their respective coefficients can either all be equal (Section 2.2.1) or, more generally, conditionally on a latent process, vary in each row (Section 2.2.3). In the latter case the rows of X are not necessarily independent. The limiting Poisson process of the eigenvalues of XX^T depends on the tail index α as well as the coefficients of the observed linear processes. As a by-product, we obtain an independent proof of Soshnikov's result for iid entries which also holds in cases where $\gamma \in \{0, \infty\}$.

The chapter is organized as follows. The main results will be presented in Section 2 while the proofs will be given in Section 3. Results from the theory of point processes and regular variation are required through most of this chapter. A detailed account on both topics can be found in a number of texts. We mainly adopt the setting, including notation and terminology, of Resnick [96].

2.2. Main results on heavy-tailed random matrices with dependent entries

2.2.1. A first result on the largest eigenvalue

Let $(Z_{it})_{i,t}$ be an array of iid random variables with marginal distribution that is regularly varying with tail index $\alpha > 0$ and *normalizing sequence* a_n , i.e.,

$$\lim_{n \to \infty} nP(|Z_{it}| > a_n x) = x^{-\alpha}, \quad \text{for each } x > 0.$$
(2.2.1)

Equivalently, this means that $(|Z_{it}|)$ is in the maximum domain of attraction of a Fréchet distribution with parameter $\alpha > 0$. The sequence a_n is then necessarily characterized by

$$a_n = n^{1/\alpha} L(n), \qquad (2.2.2)$$

for some slowly varying function $L : \mathbb{R}_+ \to \mathbb{R}_+$, i.e., a function with the property that, for each x > 0, $\lim_{t\to\infty} L(tx)/L(t) = 1$. In certain cases we also assume that Z_{11} satisfies the *tail balancing condition*, i.e., the existence of the limits

$$\lim_{x \to \infty} \frac{P(Z_{11} > x)}{P(|Z_{11}| > x)} = q \quad \text{and} \quad \lim_{x \to \infty} \frac{P(Z_{11} \le -x)}{P(|Z_{11}| > x)} = 1 - q \tag{2.2.3}$$

for some $0 \le q \le 1$. For each $p, n \in \mathbb{N}$, let $X = (X_{it})$ be the $p \times n$ data matrix, where, for each i,

$$X_{it} = \sum_{j=-\infty}^{\infty} c_j Z_{i,t-j}$$
(2.2.4)

is a stationary linear times series. To guarantee that the series in (2.2.4) converges almost surely, we assume that

$$\sum_{j=-\infty}^{\infty} |c_j|^{\delta} < \infty \quad \text{for some } \delta < \min\{\alpha, 1\}.$$
(2.2.5)

Thus in our model the rows of X are given by iid copies of a linear process. We denote by $\lambda_1, \ldots, \lambda_p$ the unordered, and by $\lambda_{(1)} \ge \ldots \ge \lambda_{(p)}$ the ordered eigenvalues of the $p \times p$ sample covariance matrix XX^{T} . They are studied via the induced point process

$$N_n \coloneqq \sum_{i=1}^p \epsilon_{a_{np}^n \lambda_i}.$$
(2.2.6)

We will always assume that $p = p_n$ is an integer-valued sequence in *n* that goes to infinity as $n \to \infty$ in order to obtain results in the 'large *n*, large *p*' setting. In the following we suppress the dependence of *p* on *n* so as to simplify the notation wherever this does not cause any ambiguity. In [7, 100] the iid case is considered, i.e., $X_{it} = Z_{it}$, assuming that the condition (2.2.1) holds for $0 < \alpha < 4$. They show that, if $p, n \to \infty$ with

$$\lim_{n \to \infty} \frac{p_n}{n} = \gamma \in (0, \infty), \tag{2.2.7}$$

 N_n converges in distribution to a Poisson process N with intensity measure $\hat{v}((x, \infty]) = x^{-\alpha/2}$. Our next theorem extends this result by considering the case where X has dependent entries. More precisely, the rows of X are given by independent copies of a linear process. It will turn out that the intensity measure of the limiting Poisson process depends on the sum of the squared coefficients of the underlying linear process.

Theorem 2.1. (i) Define the matrix $X = (X_{it})$ as in equations (2.2.1), (2.2.4) and (2.2.5) with $\alpha \in (0, 2)$. Suppose $p_n, n \to \infty$ such that

$$\limsup_{n \to \infty} \frac{p_n}{n^\beta} < \infty \tag{2.2.8}$$

for some $\beta > 0$ satisfying

- a) $\beta < \infty$ if $\alpha \in (0, 1]$, and
- b) $\beta < \max\left\{\frac{2-\alpha}{\alpha-1}, \frac{1}{2}\right\}$ if $\alpha \in (1, 2)$.

Further assume, in case $\alpha \in (5/3, 2)$, that Z_{11} has mean zero and satisfies the tail balancing condition (2.2.3). Then the point process N_n of the eigenvalues of $a_{np}^{-2}XX^{\mathsf{T}}$ converges in distribution to a Poisson point process N with intensity measure v which is given by

$$\nu((x,\infty]) = x^{-\alpha/2} \left| \sum_{j=-\infty}^{\infty} c_j^2 \right|^{\alpha/2}, \quad x > 0.$$

(ii) Assume that $X_{it} = Z_{it}$ and equation (2.2.1) is satisfied with $\alpha \in (0, 2)$. Further, let either

- a) $p_n = n^{\kappa} l(n)$ for some $\kappa \in [0, \infty)$, where l is a slowly varying function which converges to infinity if $\kappa = 0$, and is bounded away from zero if $\kappa = 1$, or
- b) $p_n \sim C \exp(cn^{\kappa})$ for some $\kappa, c, C > 0$.

Then N_n converges in distribution to a Poisson point process with intensity measure given by $\hat{v}((x, \infty]) = x^{-\alpha/2}$.

Theorem 2.1 (i) weakens the assumption of independent entries made so far in the literature on heavy-tailed random matrices at the expense of assumption (2.2.8), which is more restrictive

than the usual assumption (2.2.7) if $\alpha \in [1.5, 2)$. However, if $\alpha \in (0, 1.5)$, our assumption (2.2.8) is more general than (2.2.7). This is important for statistical applications, because p and n are usually fixed and there is no functional relationship between the two of them. If we restrict ourselves to the iid case, then Theorem 2.1 (ii) shows that the point process convergence result also holds in many cases where the limit γ from condition (2.2.7) is zero or infinity, for example, by assuming that p is regularly varying in n.

It is well known [96] that a Poisson process has an explicit representation as a transformation of a homogeneous Poisson process. In our case, the limiting Poisson process N with intensity measure ν from Theorem 2.1 can be written as

$$N \stackrel{D}{=} \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} \sum_{j=-\infty}^{\infty} c_j^2}, \tag{2.2.9}$$

where $\Gamma_i = \sum_{k=1}^{i} E_k$ is the successive sum of iid exponential random variables E_k with mean one. The points of N are labeled in decreasing order so that, by the continuous mapping theorem, we can easily deduce the weak limit of the k largest eigenvalues of XX^{T} .

Corollary 2.2.1. Under the assumptions of Theorem 2.1 we have, for each fixed integer $k \ge 1$, that

$$a_{np}^{-2}\left(\lambda_{(1)},\ldots,\lambda_{(k)}\right) \xrightarrow[n\to\infty]{D} \left(\Gamma_1^{-2/\alpha},\ldots,\Gamma_k^{-2/\alpha}\right) \left(\sum_{j=-\infty}^{\infty}c_j^2\right).$$

In particular, for each x > 0,

$$P\left(\frac{\lambda_{(k)}}{a_{np}^2} \leqslant x\right) \xrightarrow[n \to \infty]{} P(N(x, \infty) \leqslant k-1) = e^{-x^{-\alpha/2}} \sum_{m=0}^{k-1} \frac{x^{-m\alpha/2}}{m!} \left(\sum_j c_j^2\right)^{m\alpha/2}.$$

Equivalently, in terms of the singular values $s_{(1)} = \sqrt{\lambda_{(1)}} \ge ... \ge s_{(p)} = \sqrt{\lambda_{(p)}}$ of the matrix *X*, we obtain, for any fixed positive integer *k*, that

$$a_{np}^{-1}\left(s_{(1)},\ldots,s_{(k)}\right) \xrightarrow[n \to \infty]{D} \left(\Gamma_{1}^{-1/\alpha},\ldots,\Gamma_{k}^{-1/\alpha}\right) \sqrt{\sum_{j=-\infty}^{\infty} c_{j}^{2}}.$$

In a nutshell, the results in this section give the asymptotic behavior of the *k* largest eigenvalues of a sample covariance matrix XX^{T} when the rows of *X* are given by iid copies of some linear process with infinite variance. Our results will be generalized further in Section 2.2.3, where, conditionally on a latent process, the rows of *X* will be independent but not identically distributed.

2.2.2. Examples and discussion

Theorem 2.1 holds for any linear process which has regularly varying noise with infinite variance as long as condition (2.2.5) is satisfied. Since the coefficients of a causal ARMA process decay exponentially, (2.2.5) is trivially satisfied in this case. As two simple examples, consider an MA(1) process $X_{it} = Z_{it} + \theta Z_{i,t-1}$, which satisfies $\sum_j c_j^2 = 1 + \theta^2$; and a causal AR(1) process $X_{it} - \phi X_{i,t-1} = Z_{it}$, $|\phi| < 1$, where $\sum_j c_j^2 = (1 - \phi^2)^{-1}$. Yet another example of a linear process fitting in our framework is a fractionally integrated ARMA(p, d, q) processes with d < 0 and regularly varying noise with index $\alpha \in [1, 2)$, see, e.g., [24] for further details. In this case $|c_j| \leq C_j t^{d-1}$ is summable and therefore condition (2.2.5) is satisfied for $\alpha \ge 1$.

Regarding the normalization in (2.2.6), the sequence a_n is chosen such that the individual entries of the matrix $Z := (Z_{it})_{i,t}$ satisfy (2.2.1). Replacing the iid sequence in the rows of Z with a linear process to obtain the matrix X changes the tail behavior of its entries. Indeed, the result stated in Davis and Resnick [39, eq. (2.7)] shows, under the assumption (2.2.3), that

$$nP\left(\left|\sum_{j}c_{j}Z_{1,t-j}\right| > a_{np}^{2}x\right) \xrightarrow[n \to \infty]{} x^{-\alpha} \sum_{j} |c_{j}|^{\alpha}$$

In view of (2.2.1) this suggests the normalization $\tilde{X}_{it} = (\sum_j |c_j|^{\alpha})^{-1/\alpha} X_{it}$. Denote by $\tilde{\lambda}_1, \ldots, \tilde{\lambda}_p$ the eigenvalues of $\tilde{X}\tilde{X}^{\mathsf{T}}$, where $\tilde{X} = (\tilde{X}_{it})_{i,t}$. Since this is just a multiplication by a constant, we immediately obtain, by Theorem 2.1, that

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\tilde{\lambda}_{i}} \xrightarrow{D}_{n \to \infty} \tilde{N},$$

where \tilde{N} is a Poisson process with intensity measure $\tilde{\nu}$ given by

$$\tilde{\nu}((x,\infty]) = x^{-\alpha/2} \frac{\left|\sum_{j} c_{j}^{2}\right|^{\alpha/2}}{\sum_{j} |c_{j}|^{\alpha}}.$$
(2.2.10)

Thus $\left|\sum_{j} c_{j}^{2}\right|^{\alpha/2} (\sum_{j} |c_{j}|^{\alpha})^{-1}$ quantifies the effect of the dependence on the point process of the eigenvalues when the tail behavior of each marginal X_{it} is equivalent to the iid case.

Assume for a moment that the dimension p is fixed for any n. Then it follows easily from [39, Theorem 4.1] and arguments of this chapter that $a_n^{-2}\lambda_{(1)} \rightarrow \sum_j c_j^2 \max_{1 \le i \le p} S_i$ in distribution as $n \to \infty$, where (S_i) are independent positive stable with index $\alpha/2$, $0 < \alpha < 2$. If p is large, one would intuitively expect that $\max_{1 \le i \le p} S_i \approx p^{2/\alpha} \Gamma_1^{-2/\alpha}$, where Γ_1 is exponentially distributed with mean 1. Corollary 2.2.1 not only makes this intuition precise but also gives the correct normalization a_{np}^{-2} . The distribution of the maximum of p independent stables is not known analytically, hence 'large n, large p' in fact gives a simpler solution than the traditional 'fixed p, large n' setting.

2.2.3. Extension to random coefficient models

So far we have assumed that our observed process has independent components, each of which are modelled by the same linear process. From now on we will allow for a different set of coefficients in each row. To this end, let $(\theta_i)_{i \in \mathbb{N}}$ be a sequence of random variables *independent* of (Z_{it}) with values in some space Θ . Assume that there is a family of measurable functions $(c_i : \Theta \to \mathbb{R})_{i \in \mathbb{N}}$ such that

$$\sup_{\theta \in \Theta} |c_j(\theta)| \leq \widetilde{c}_j, \quad \text{for some deterministic } \widetilde{c}_j \text{ satisfying condition (2.2.5).}$$
(2.2.11)

Our observed processes have the form

$$X_{it} = \sum_{j=-\infty}^{\infty} c_j(\theta_i) Z_{i,t-j}$$
(2.2.12)

where (Z_{it}) is given as in (2.2.1) with $\alpha \in (0, 2)$. Thus, conditionally on the latent process (θ_i) , the rows of X are independent linear processes with different coefficients. Unconditionally, the rows of X are dependent if the sequence (θ_i) is dependent. Theorem 2.2 below covers three classes among which (θ_i) may be chosen: stationary ergodic; stationary but not necessarily ergodic; and ergodic in the Markov sense but not necessarily stationary. In the following we say that a sequence of point processes \mathcal{M}_n converges, conditionally on a sigma-algebra \mathcal{H} , in distribution to a point process \mathcal{M} , if the conditional Laplace functionals converge almost surely, i.e., if there exists a measurable set B with measure one such that for all $\omega \in B$ and all nonnegative continuous functions f with compact support,

$$E\left(e^{-\mathcal{M}_n(f)}\big|\mathcal{H}\right)(\omega) \xrightarrow[n \to \infty]{} E\left(e^{-\mathcal{M}(f)}\big|\mathcal{H}\right)(\omega) \quad \text{as } n \to \infty.$$
(2.2.13)

Theorem 2.2. Define $X = (X_{it})$ with X_{it} as given in (2.2.12). Suppose that (2.2.11) is satisfied, and $p, n \to \infty$ such that (2.2.8) holds under the same conditions as in Theorem 2.1 (i). Further assume, in case $\alpha \in (5/3, 2)$, that Z_{11} has mean zero and satisfies the tail balancing condition (2.2.3).

(i) If (θ_i) is a stationary ergodic sequence, then, both conditionally on (θ_i) as well as unconditionally, we have that

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\lambda_{i}} \xrightarrow[n \to \infty]{} \sum_{i=1}^{\infty} \epsilon_{\Gamma_{i}^{-2/\alpha}} \|_{\Sigma_{j}} c_{j}^{2}(\theta_{1})\|_{\frac{\alpha}{2}}, \qquad (2.2.14)$$

with the constant $\left\|\sum_{j} c_{j}^{2}(\theta_{1})\right\|_{\frac{\alpha}{2}} = \left(E\left|\sum_{j} c_{j}^{2}(\theta_{1})\right|^{\alpha/2}\right)^{2/\alpha}$, and (Γ_{i}) as in (2.2.9).

(ii) If (θ_i) is stationary but not necessarily ergodic, then we have, conditionally on (θ_i) , that

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\lambda_{i}} \xrightarrow[n \to \infty]{} \sum_{i=1}^{\infty} \epsilon_{\Gamma_{i}^{-2/\alpha}Y^{2/\alpha}},$$

with $Y = E(|\sum_j c_j^2(\theta_1)|^{\alpha/2}|\mathcal{G})$, where \mathcal{G} is the invariant σ -field generated by (θ_i) . In particular, Y is independent of (Γ_i) .

(iii) Suppose (θ_i) is either an irreducible Markov chain on a countable state space Θ or a positive Harris chain in the sense of Meyn and Tweedie [83]. If (θ_i) has a stationary probability distribution π then, conditionally on (θ_i) as well as unconditionally, (2.2.14) holds with

$$\left\|\sum_{j} c_{j}^{2}(\theta_{1})\right\|_{\frac{\alpha}{2}} = \left(\int_{\Theta} \left|\sum_{j} c_{j}^{2}(\theta)\right|^{\alpha/2} \pi(d\theta)\right)^{2/\alpha}.$$

One can view the assumptions (i) and (ii) of Theorem 2.2 in a Bayesian framework in which the parameters of the observed process are drawn from an unknown prior distribution. As an example, let (θ_i) be a stationary ergodic AR(1) process $\theta_i = \phi \theta_{i-1} + \xi_i$, where $|\phi| \neq 1$ and (ξ_i) is a sequence of bounded iid random variables, and set $X_{it} = Z_{it} + \theta_i Z_{i,t-1}$. Then, by Theorem 2.2 (i), we would expect, for *n* and *p* large enough, that

$$a_{np}^{-2}\lambda_{(1)} \approx \Gamma_1^{-2/\alpha} \left(E |1+\theta_1|^{\alpha/2} \right)^{2/\alpha}.$$

Models of this kind are referred to as *random coefficient models* and often used in times series analysis, see, e.g., [74] for an overview. In the setting of Theorem 2.2 (iii) one might think of a *Hidden Markov Model* where the latent Markov process (θ_i) evolves along the rows of *X*, each state θ_i defining another univariate linear model.

2.3. Proofs and auxiliary results

The first step is to show that the matrix XX^{T} is well approximated by its diagonal, see Section 2.3.2. In the second step we then derive the extremes of the diagonal of XX^{T} in Section 2.3.3. Both steps make use of a large deviation result which is presented in the upcoming section.

2.3.1. A large deviation result and its consequences

The next theorem gives the joint large deviations of the sum and the maximum of iid nonnegative random variables with infinite variance. **Proposition 2.3.1.** Let $(x_n)_{n \in \mathbb{N}}$ and $(y_n)_{n \in \mathbb{N}}$ be sequences of nonnegative numbers with $x_n \to \infty$ such that $x_n/y_n \to \gamma \in (0, \infty]$. Suppose $(Y_t)_{t \in \mathbb{N}}$ is an iid sequence of nonnegative random variables with tail index $\alpha \in (0, 2)$ and normalizing sequence b_n . If $1 \leq \alpha < 2$, we assume that $b_n x_n/n^{1+\delta} \to \infty$ for some $\delta > 0$. Then

$$\lim_{n \to \infty} \frac{P\left(\sum_{t=1}^{n} Y_t > b_n x_n, \max_{1 \le t \le n} Y_t > b_n y_n\right)}{n P(Y_1 > b_n \max\{x_n, y_n\})} = 1.$$
 (2.3.1)

Proof. Let us first assume that $0 < \alpha < 1$. Using standard arguments from the theory of regularly varying functions, see e.g. [96], it can be easily seen that for any positive sequence $z_n \rightarrow \infty$ we have

$$\lim_{n \to \infty} \frac{P(\max_{1 \le t \le n} Y_t > b_n z_n)}{nP(Y_1 > b_n z_n)} = 1.$$
(2.3.2)

Obviously the limit in (2.3.1) is greater or equal than one because $\sum_{t=1}^{n} Y_t \ge \max_{1 \le t \le n} Y_t$. Thus it is only left to prove that it is also smaller. Denote by $Y_{(1)} \ge \ldots \ge Y_{(n)}$ the order statistics of Y_1, \ldots, Y_n . By decomposing $\sum_t Y_t$ into the sum of $\max_t Y_t$ and lower order terms we see that, for any $\theta \in (0, 1)$,

$$\frac{P(\sum_{t=1}^{n} Y_t > b_n x_n, \max_{1 \le t \le n} Y_t > b_n y_n)}{nP(Y_1 > b_n \max\{x_n, y_n\})} \le \frac{P(\max_{1 \le t \le n} Y_t > b_n \max\{\theta x_n, y_n\})}{nP(Y_1 > b_n \max\{x_n, y_n\})} + \frac{P(\sum_{t=2}^{n} Y_{(t)} > b_n x_n(1-\theta))}{nP(Y_1 > b_n \max\{x_n, y_n\})}.$$

By an application of [96, Proposition 0.8 (iii)] one can show similarly as in the proof of (2.3.2) that

$$\lim_{\theta \to 1} \lim_{n \to \infty} \frac{P(\max_{1 \le t \le n} Y_t > b_n \max\{\theta x_n, y_n\})}{nP(Y_1 > b_n \max\{x_n, y_n\})} = 1.$$

Hence, it is only left to show that the second summand vanishes as $n \to \infty$. To this end we partition the underlying probability space into $\{Y_{(2)} \le \epsilon b_n x_n\} \cup \{Y_{(2)} > \epsilon b_n x_n\}, \epsilon > 0$, to obtain

$$\frac{P\left(\sum_{t=2}^{n} Y_{(t)} > b_{n} x_{n}(1-\theta)\right)}{nP(Y_{1} > b_{n} \max\{x_{n}, y_{n}\})} \leqslant \frac{P\left(\sum_{t=2}^{n} Y_{(t)} \mathbf{1}_{\{Y_{(2)} \leqslant \epsilon b_{n} x_{n}\}} > b_{n} x_{n}(1-\theta)\right)}{nP(Y_{1} > b_{n} \max\{x_{n}, y_{n}\})} + \frac{P\left(Y_{(2)} > \epsilon b_{n} x_{n}\right)}{nP(Y_{1} > b_{n} \max\{x_{n}, y_{n}\})} = \Sigma_{1} + \Sigma_{2}$$

Denote by $M_n = \max_{1 \le t \le n} Y_t$ and $z_n = \max\{x_n, y_n\}$. Then easy combinatorics and (2.3.2)

yield

$$\begin{split} \Sigma_2 &= \frac{1 - P\left(Y_{(2)} \leqslant \epsilon b_n x_n\right)}{n P(Y_1 > b_n z_n)} \\ &= \frac{1 - P\left(M_n \leqslant \epsilon b_n x_n\right)}{n P(Y_1 > b_n z_n)} - \frac{n P\left(M_{n-1} \leqslant \epsilon b_n x_n\right) P(Y_1 > \epsilon b_n x_n)}{n P(Y_1 > b_n z_n)} \\ &= \frac{P\left(M_n > \epsilon b_n x_n\right)}{n P(Y_1 > \epsilon b_n x_n)} \frac{P(Y_1 > \epsilon b_n x_n)}{P(Y_1 > b_n z_n)} - \frac{P(Y_1 > \epsilon b_n x_n)}{P(Y_1 > b_n z_n)} P\left(M_{n-1} \leqslant \epsilon b_n x_n\right) \\ &\sim \frac{P(Y_1 > \epsilon b_n x_n)}{P(Y_1 > b_n z_n)} \left(1 - P\left(M_{n-1} \leqslant \epsilon b_n x_n\right)\right) \xrightarrow[n \to \infty]{} 0. \end{split}$$

The convergence to zero follows from $P(M_{n-1} \le \epsilon b_n x_n) \to 1$ and, by [96, Proposition 0.8 (iii)], $\frac{P(Y_1 > \epsilon b_n x_n)}{P(Y_1 > b_n z_n)} \to \epsilon^{-\alpha} \max\{1, \gamma^{-\alpha}\}$. Thus it is only left to show that Σ_1 goes to zero. By Markov's inequality and Karamata's Theorem [96, Theorem 0.6] we have that

$$\Sigma_{1} \leq \frac{P\left(\sum_{t=1}^{n} Y_{t} \mathbf{1}_{\{Y_{t} \leq \epsilon b_{n} x_{n}\}} > b_{n} x_{n} (1-\theta)\right)}{nP(Y_{1} > b_{n} \max\{x_{n}, y_{n}\})}$$

$$\leq \frac{1}{b_{n} x_{n} (1-\theta)} \frac{E(Y_{1} \mathbf{1}_{\{Y_{1} \leq \epsilon b_{n} x_{n}\}})}{P(Y_{1} > b_{n} z_{n})}$$

$$\sim \frac{1}{(1-\theta)} \frac{\alpha}{1-\alpha} \frac{\epsilon P(Y_{1} > \epsilon b_{n} x_{n})}{P(Y_{1} > b_{n} z_{n})} \xrightarrow[n \to \infty]{} \frac{1}{(1-\theta)} \frac{\alpha}{1-\alpha} \epsilon^{1-\alpha} \max\{1, \gamma^{-\alpha}\}.$$

which converges to zero as ϵ goes to zero, since $\alpha < 1$. Thus for $0 < \alpha < 1$ the proof is complete. If $1 \leq \alpha < 2$, only Σ_1 has to be treated differently. The truncated mean $\mu_n = E(Y_1 \mathbf{1}_{\{Y_1 \leq \epsilon b_n x_n\}})$ either converges to a constant or is a slowly varying function. In either case, we have that $b_n x_n / (n\mu_n) = b_n x_n n^{-1-\delta} n^{\delta} / \mu_n \to \infty$ by assumption. Thus, a mean-correction argument and Karamata's Theorem imply

$$\begin{split} \limsup_{n \to \infty} \Sigma_1 &\leq \limsup_{n \to \infty} \frac{P\left(\sum_{t=1}^n Y_t \mathbf{1}_{\{Y_t \leq \epsilon b_n x_n\}} - n\mu_n > b_n x_n (1-\theta) - n\mu_n\right)}{nP(Y_1 > b_n z_n)} \\ &\leq \frac{1}{(1-\theta)^2} \limsup_{n \to \infty} \frac{1}{b_n^2 x_n^2} \frac{\operatorname{Var}(Y_1 \mathbf{1}_{\{Y_1 \leq \epsilon b_n x_n\}})}{P(Y_1 > b_n z_n)} \\ &\leq \frac{1}{(1-\theta)^2} \frac{\alpha/2}{1-\alpha/2} \epsilon^{2-\alpha} \max\{1, \gamma^{-\alpha}\} \underset{\epsilon \to 0}{\longrightarrow} 0, \end{split}$$

since $\alpha < 2$. This completes the proof.

We finish this section with a few consequences of Proposition 2.3.1. Note that (2.2.1) implies

$$pnP(Z_{11}^2 > a_{np}^2 x) \xrightarrow[n \to \infty]{} x^{-\alpha/2} \quad \text{for each } x > 0.$$
(2.3.3)

2. Limit Theory for the largest eigenvalues of sample covariance matrices

Choosing $Y_t = Z_{1t}^2$, $b_n = a_n^2$, $x_n = xa_{np}^2/a_n^2$ and $y_n = ya_{np}^2/a_n^2$, we have from Proposition 2.3.1 and (2.3.3), for $\alpha \in (0, 2)$, that

$$pP\left(\sum_{t=1}^{n} Z_{1t}^{2} > a_{np}^{2} x, \max_{1 \le t \le n} Z_{1t}^{2} > a_{np}^{2} y\right) \xrightarrow[n \to \infty]{} \max\{x, y\}^{-\alpha/2} \quad \text{for each } x, y > 0.$$

Therefore, by [96, Proposition 3.21], we obtain the point process convergence

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}(\sum_{t=1}^{n} Z_{it}^{2}, \max_{1 \leq t \leq n} Z_{it}^{2})} \xrightarrow{D}_{n \to \infty} \sum_{i=1}^{\infty} \epsilon_{\Gamma_{i}^{-2/\alpha}(1,1)}, \qquad (2.3.4)$$

with (Γ_i) as in (2.2.9). For another application of Proposition 2.3.1, set $Y_t = |Z_{1t}|$, $b_n = a_n$, $x_n = xa_{np}/a_n$ and $y_n = ya_{np}/a_n$. Under the additional assumption

$$\liminf_{n\to\infty}\frac{p}{n}\in(0,\infty]$$

we have $b_n x_n / n^{1+\gamma} \to \infty$ for some $\gamma < (2 - \alpha) / \alpha$, thus, for $\alpha \in (0, 2)$,

$$pP\left(\sum_{t=1}^{n} |Z_{1t}| > a_{np}x, \max_{1 \le t \le n} |Z_{1t}| > a_{np}y\right) \xrightarrow[n \to \infty]{} \max\{x, y\}^{-\alpha} \quad \text{for each } x, y > 0.$$

Therefore we obtain as before

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-1}(\sum_{t=1}^{n} |Z_{it}|, \max_{1 \leq t \leq n} |Z_{it}|)} \xrightarrow[n \to \infty]{D} \sum_{i=1}^{\infty} \epsilon_{\Gamma_{i}^{-1/\alpha}(1,1)}.$$
(2.3.5)

The result of the following proposition is also a consequence of Proposition 2.3.1.

Proposition 2.3.2. Let (Z_{it}) be as in (2.2.1) with $0 < \alpha < 2$. Suppose that (2.2.8) is satisfied for some $0 < \beta < \infty$. Then

$$a_{np}^{-2} \max_{1 \leq i < j \leq p} \sum_{t=1}^{n} |Z_{it}Z_{jt}| \xrightarrow{P}_{n \to \infty} 0.$$

Proof. By [47], the iid random variables $Y_t = |Z_{1t}Z_{2t}|$ are regularly varying with tail index α with some normalizing sequence b_n . Thus, there exists a slowly varying L_1 such that $P(Y_1 > x) = x^{-\alpha}L_1(x)$. Using (2.2.2) this implies

$$p^2 n P(Y_1 > a_{np}^2 \epsilon) = n^{-1} \epsilon^{-\alpha} L(np)^{-2\alpha} L_1\left((np)^{2/\alpha} L(np)^2 \epsilon\right)$$

By Potter's bound, see, e.g., [96, Proposition 0.8 (ii)], for any slowly varying function \tilde{L} and any $\delta > 0$ there exist $c_1, c_2 > 0$ such that $c_1 n^{-\delta} < \tilde{L}(n) < c_2 n^{\delta}$ for *n* large enough. An application of this bound together with assumption (2.2.8) shows that

$$p^2 n P(Y_1 > a_{np}^2 \epsilon) \xrightarrow[n \to \infty]{} 0.$$
 (2.3.6)

Hence, using Proposition 2.3.1 with $x_n = a_{np}^2/b_n\epsilon$ and $y_n = 0$ yields

$$P\left(\max_{1\leqslant i < j\leqslant p} \sum_{t=1}^{n} |Z_{it}Z_{jt}| > a_{np}^{2}\epsilon\right) \leqslant p^{2} P\left(\sum_{t=1}^{n} Y_{t} > a_{np}^{2}\epsilon\right) \underset{n \to \infty}{\longrightarrow} 0,$$

since $b_n x_n / n^{1+\gamma} = a_{np}^2 / n^{1+\gamma} \to \infty$ for $\alpha < 2$ and some $\gamma < (2-\alpha) / \alpha$.

2.3.2. Convergence in Operator Norm

Denote by $D = \text{diag}(XX^{\mathsf{T}})$ the diagonal of the matrix XX^{T} , i.e., $D_{ii} = (XX^{\mathsf{T}})_{ii}$ and $D_{ij} = 0$ for $i \neq j$. In this section we show that $a_{np}^{-2}XX^{\mathsf{T}}$ converges in probability to $a_{np}^{-2}D$ in operator norm. This implies that the off-diagonal elements of $a_{np}^{-2}XX^{\mathsf{T}}$ do not contribute to the limiting eigenvalue spectrum. Recall that, for a real $p \times n$ matrix A, the operator 2-norm $||A||_2$ is the square root of the largest eigenvalue of AA^{T} , and the infinity-norm is given by $||A||_{\infty} = \max_{1 \leq i \leq p} \sum_{t=1}^{n} |A_{it}|$. The following result holds under a much more general setting than assumed in Theorem 2.1 by allowing for an arbitrary dependence structure within the rows of X.

Proposition 2.3.3. Let $X = (X_{it})_{i,t}$ be a $p \times n$ random matrix whose entries are identically distributed with tail index $\alpha \in (0, 2)$ and normalizing sequence (a_n) . Assume that the rows of X are independent. Suppose that (2.2.8) holds for some $\beta > 0$. If $1 < \alpha < 2$, assume additionally that $\beta < \frac{2-\alpha}{\alpha-1}$. Then we have

$$a_{np}^{-2} \left\| XX^{\mathsf{T}} - D \right\|_{2} \xrightarrow[n \to \infty]{P} 0.$$
(2.3.7)

Proof. Since $\|XX^{\mathsf{T}} - D\|_2 \leq \|XX^{\mathsf{T}} - D\|_{\infty}$, it is enough to show that for every $\epsilon \in (0, 1)$,

$$P\left(\max_{\substack{i=1,\dots,p\\j\neq i}}\sum_{\substack{j=1\\j\neq i}}^{p}\left|\sum_{t=1}^{n}X_{it}X_{jt}\right| > a_{np}^{2}\epsilon\right) \leqslant pP\left(\sum_{j=2}^{p}\sum_{t=1}^{n}|X_{1t}X_{jt}| > a_{np}^{2}\epsilon\right) \underset{n\to\infty}{\longrightarrow} 0.$$

By partitioning the underlying probability space into $\{\max_{j,t} |X_{1t}X_{jt}| \le a_{np}^2\}$ and its complement, we obtain that

$$pP\left(\sum_{j=2}^{p}\sum_{t=1}^{n}|X_{1t}X_{jt}| > a_{np}^{2}\epsilon\right) \leq pP\left(\sum_{j=2}^{p}\sum_{t=1}^{n}|X_{1t}X_{jt}|\mathbf{1}_{\{|X_{1t}X_{jt}| \leq a_{np}^{2}\}} > a_{np}^{2}\epsilon\right) + pP\left(\max_{2 \leq j \leq p}\max_{1 \leq t \leq n}|X_{1t}X_{jt}| > a_{np}^{2}\epsilon\right) = \mathbf{I} + \mathbf{II}.$$

The same argument used for (2.3.6) shows that $II \leq p^2 n P(|X_{11}X_{21}| > a_{np}^2) \xrightarrow[n \to \infty]{} 0$ by independence of the rows of X. To deal with term I we first assume that $\alpha > 1$ and choose some $\gamma \in (\alpha, 2)$. Hölder's inequality shows that

$$\left(\sum_{j=2}^{p}\sum_{t=1}^{n}|X_{1t}X_{jt}|\right)^{\gamma} \leq \left(\sum_{j=2}^{p}\sum_{t=1}^{n}|X_{1t}X_{jt}|^{\gamma}\right)(np)^{\gamma-1},$$

and therefore

$$\mathbf{I} \leqslant pP\left(\sum_{j=2}^{p} \sum_{t=1}^{n} |X_{1t}X_{jt}|^{\gamma} \mathbf{1}_{\{|X_{1t}X_{jt}| \leqslant a_{np}^{2}\}} > \frac{a_{np}^{2\gamma}}{(np)^{\gamma-1}}\epsilon\right).$$

Note that $|X_{1t}X_{jt}|^{\gamma}$ has regularly varying tails with index $\alpha/\gamma < 1$. Hence we can apply Markov's Inequality and Karamata's Theorem to infer that

$$\mathbf{I} \leq c_1 \frac{p^2 n(np)^{\gamma - 1}}{a_{np}^{2\gamma}} E\left(|X_{11}X_{21}|^{\gamma} \mathbf{1}_{\{|X_{11}X_{21}| \leq a_{np}^2\}}\right) \sim c_2 p^2 n(np)^{\gamma - 1} P\left(|X_{11}X_{21}| > a_{np}^2\right).$$
(2.3.8)

Therefore, the proof of Proposition 2.3.2 shows that the term in (2.3.8) goes to zero if $(np)^{\gamma-1}/n$ does. In view of assumption (2.2.8) this is true for $\beta < (2 - \gamma)/(\gamma - 1)$. Since we can choose γ arbitrary close to α it suffices that $\beta < (2 - \alpha)/(\alpha - 1)$. If $\alpha < 1$ we do not need Hölder's inequality since the above argument can be applied with $\gamma = 1$, thus it suffices that (2.2.8) holds for some $\beta < \infty$. For the remaining case $\alpha = 1$, observe that, for any given $\beta < \infty$, we choose γ arbitrarily close to 1 so that $(np)^{\gamma-1}/n \to 0$.

The above result can be improved for $5/3 < \alpha < 2$ if we assume the rows of X to be realizations of a linear process.

Proposition 2.3.4. The assumptions of Theorem 2.1 (i) imply (2.3.7).

Proof. By Proposition 2.3.3 it suffices to show that, for $\alpha \in (5/3, 2)$, the assumption

$$\lim_{n \to \infty} \frac{p}{\sqrt{n}} = 0 \tag{2.3.9}$$

implies convergence in operator norm in the sense of (2.3.7). In this proof, *c* denotes a positive constant that may vary from expression to expression. Define

$$Z_{it}^{L} = Z_{it} \mathbf{1}_{\{|Z_{it}| \le a_{np}\}}, \quad X_{it}^{L} = \sum_{k} c_{k} Z_{i,t-k}^{L},$$
$$Z_{it}^{U} = Z_{it} \mathbf{1}_{\{|Z_{it}| > a_{np}\}}, \quad X_{it}^{U} = \sum_{k} c_{k} Z_{i,t-k}^{U}.$$
Using $||XX^{\mathsf{T}} - D||_2 \leq ||XX^{\mathsf{T}} - D||_{\infty}$ as before we have

$$\begin{split} P\left(\left\|XX^{\mathsf{T}} - D\right\|_{2} > a_{np}^{2}\epsilon\right) &\leq pP\left(\sum_{j=2}^{p} \left|\sum_{t=1}^{n} X_{1t}X_{jt}\right| > a_{np}^{2}\epsilon\right) \\ &\leq pP\left(\sum_{j=2}^{p} \left|\sum_{t=1}^{n} X_{1t}^{L}X_{jt}^{L}\right| > \frac{a_{np}^{2}}{4}\epsilon\right) + pP\left(\sum_{j=2}^{p} \left|\sum_{t=1}^{n} X_{1t}^{L}X_{jt}^{U}\right| > \frac{a_{np}^{2}}{4}\epsilon\right) \\ &+ pP\left(\sum_{j=2}^{p} \left|\sum_{t=1}^{n} X_{1t}^{U}X_{jt}^{L}\right| > \frac{a_{np}^{2}}{4}\epsilon\right) + pP\left(\sum_{j=2}^{p} \left|\sum_{t=1}^{n} X_{1t}^{U}X_{jt}^{U}\right| > \frac{a_{np}^{2}}{4}\epsilon\right) \\ &= \mathbf{I} + \mathbf{II} + \mathbf{III} + \mathbf{IV}. \end{split}$$

We will show that each of theses terms converges to zero. To this end, note that $E|Z_{11}^L|$ converges to a constant, and, by Karamata's Theorem,

$$E|Z_{11}^U| \sim ca_{np}P(|Z_{11}| > a_{np}) \sim ca_{np}(np)^{-1}, \quad n \to \infty.$$

Therefore, by Markov's inequality, we have

$$II \leq \frac{4p}{a_{np}^2 \epsilon} \sum_{j=2}^p \sum_{t=1}^n \sum_{k,l} |c_k c_l| \, E|Z_{1,t-k}^L| \, E|Z_{j,t-l}^U| \sim c \left(\sum_k |c_k|\right)^2 \frac{p^2 n}{a_{np}^2} a_{np} (np)^{-1} = c \frac{p}{a_{np}},$$

and, by (2.2.2), we obtain that this is equal to $c L(np)^{-1}p^{1-1/\alpha}n^{-1/\alpha} \to 0$ as $n \to \infty$. By symmetry, III can be handled the same way. It is easy to see that term IV is of even lower order, namely

$$c \frac{p^2 n}{a_{np}^2} (a_{np} (np)^{-1})^2 = cn^{-1} \to 0.$$

Thus it is only left to show that I converges to zero. To this end, we use Karamata's Theorem to obtain

$$E\left[(Z_{11}^L)^2\right] = E\left[Z_{11}^2 \mathbf{1}_{\{|Z_{11}| \le a_{np}\}}\right] \sim ca_{np}^2 P(|Z_{11}| > a_{np}) \sim ca_{np}^2 (np)^{-1}.$$

Since Z_{11} satisfies the tail balancing condition (2.2.3), and $EZ_{11} = 0$, we can apply Karamata's Theorem to the positive and the negative tail of Z_{11}^L , thus, for $q \notin \{0, \frac{1}{2}, 1\}$,

$$\begin{split} \xi_n &\coloneqq E[Z_{11}^L] = E[Z_{11} \mathbf{1}_{\{|Z_{11}| \leqslant a_{np}\}}] = -E[Z_{11} \mathbf{1}_{\{|Z_{11}| > a_{np}\}}] \\ &= -E[Z_{11} \mathbf{1}_{\{Z_{11} > a_{np}\}}] + E[-Z_{11} \mathbf{1}_{\{-Z_{11} > a_{np}\}}] \\ &\sim -q \frac{\alpha}{\alpha - 1} a_{np} P(|Z_{11}| > a_{np}) + (1 - q) \frac{\alpha}{\alpha - 1} a_{np} P(|Z_{11}| > a_{np}) \\ &\sim (1 - 2q) \frac{\alpha}{\alpha - 1} a_{np} (np)^{-1}. \end{split}$$

2. Limit Theory for the largest eigenvalues of sample covariance matrices

Clearly, for any $0 \le q \le 1$, one therefore has

$$\frac{np\xi_n}{a_{np}} \to (1-2q)\frac{\alpha}{\alpha-1}.$$

As a consequence we obtain for $\mu_n = E(X_{11}^L X_{21}^L) = (EX_{11}^L)^2 = (\sum_k c_k)^2 \xi_n^2$ that

$$\frac{\mu_n pn}{a_{np}^2} = (np)^{-1} \left(\frac{np\xi_n}{a_{np}}\right)^2 \left(\sum_k c_k\right)^2 \to 0.$$

Therefore we obtain for summand I that

$$\mathbf{I} = pP\left(\sum_{j=2}^{p} \left|\sum_{t=1}^{n} X_{1t}^{L} X_{jt}^{L}\right| > \frac{a_{np}^{2}}{4}\epsilon\right) \leq p^{2}P\left(\left|\sum_{t=1}^{n} X_{1t}^{L} X_{2t}^{L}\right| > \frac{a_{np}^{2}}{4p}\epsilon\right)$$
$$\sim p^{2}P\left(\left|\sum_{t=1}^{n} X_{1t}^{L} X_{2t}^{L} - n\mu_{n}\right| > \frac{a_{np}^{2}}{4p}\epsilon\right)$$

Since we correct by the mean, Markov's inequality yields

$$p^{2}P\left(\left|\sum_{t=1}^{n} X_{1t}^{L} X_{2t}^{L} - n\mu_{n}\right| > \frac{a_{np}^{2}}{4p}\epsilon\right)$$

$$\leq \frac{16p^{4}}{a_{np}^{4}\epsilon^{2}} \operatorname{Var}\left(\sum_{t=1}^{n} X_{1t}^{L} X_{2t}^{L}\right)$$

$$= \frac{16p^{4}}{a_{np}^{4}\epsilon^{2}} \sum_{t,t'=1}^{n} \sum_{k,k',l,l'} c_{k}c_{k'}c_{l}c_{l'} \operatorname{Cov}\left(Z_{1,t-k}^{L} Z_{2,t-l}^{L}, Z_{1,t'-k'}^{L} Z_{2,t'-l'}^{L}\right).$$
(2.3.10)
$$(2.3.10)$$

Due to the independence of the Z's, the covariance in the last expression is non-zero iff t - k = t' - k' or t - l = t' - l'. This gives us three distinct cases we deal with separately. First, assume that both t - k = t' - k' and t - l = t' - l'. Then the covariance in (2.3.11) is equal to $Var(Z_{11}^L Z_{2.1}^L)$ and so bounded by

$$E[(Z_{11}^L)^2(Z_{21}^L)^2] = (E(Z_{11}^L)^2)^2 \sim (ca_{np}^2(np)^{-1})^2 \sim ca_{np}^4(np)^{-2}.$$

Second, let t - k = t' - k' but $t - l \neq t' - l'$. Then the covariance becomes

$$\begin{aligned} \operatorname{Cov}(Z_{1,t-k}^{L}Z_{2,t-l}^{L}, Z_{1,t'-k'}^{L}Z_{2,t'-l'}^{L}) &= E((Z_{1,t-k}^{L})^{2}Z_{2,t-l}^{L}Z_{2,t'-l'}^{L}) - \xi_{n}^{4} \\ &= E((Z_{1,t-k}^{L})^{2})\xi_{n}^{2} - \xi_{n}^{4} \\ &\sim ca_{np}^{2}(np)^{-1}(\pm c\,a_{np}(np)^{-1})^{2} - (\pm c\,a_{np}(np)^{-1})^{4} \\ &\sim ca_{np}^{4}(np)^{-3} - ca_{np}^{4}(np)^{-4} \sim ca_{np}^{4}(np)^{-3}, \end{aligned}$$

which is of lower order than in the case considered before. By symmetry, the third case, where t - l = t' - l' but $t - k \neq t' - k'$, can be dealt with in exactly the same way. In all cases t' can be assumed to be fixed, thus we can bound (2.3.11) by

$$c\frac{p^{4}}{a_{np}^{4}}\left(\sum_{k}|c_{k}|\right)^{4}\sum_{t=1}^{n}a_{np}^{4}(np)^{-2}=c\frac{p^{2}}{n}\to 0, \quad n\to\infty.$$

2.3.3. Extremes of $diag(XX^T)$

In this section we analyze the extremes of the diagonal entries of XX^T , which are partial sums of squares of linear processes. To this end, we start with an auxiliary result.

Proposition 2.3.5. Let (Z_t) be an iid sequence such that $nP(|Z_1| > a_n x) \rightarrow x^{-\alpha}$ with $\alpha \in (0, 2)$. For any sequence (c_j) satisfying (2.2.5) we have, if p and n go to infinity, that

$$pP\left(\sum_{t=1}^{n}\sum_{j=-\infty}^{\infty}c_j^2 Z_{t-j}^2 > a_{np}^2 x\right) \to \left(\sum_{j=-\infty}^{\infty}c_j^2\right)^{\frac{\mu}{2}} x^{-\alpha/2}$$

Proof. Fix some x > 0. Observe that Proposition 2.3.1 and (2.3.3) imply for $n \to \infty$ that $pP(\sum_{t=1}^{n} Z_t^2 > a_{np}^2 x) \to x^{-\alpha/2}$. We begin by showing the claim for a linear process of finite order. For any $\eta > 0$ we have

$$P\left(\left|\sum_{j=-m}^{m} c_{j}^{2} \sum_{t=1}^{n} Z_{t}^{2} - \sum_{t=1}^{n} \sum_{j=-m}^{m} c_{j}^{2} Z_{t-j}^{2}\right| > a_{np}^{2} \eta\right) \leqslant P\left(\sum_{j=-m}^{m} c_{j}^{2} \sum_{t=1-j}^{j} Z_{t}^{2} > a_{np}^{2} \eta\right) \underset{n \to \infty}{\longrightarrow} 0.$$

Consequently,

$$\lim_{n \to \infty} pP\left(\sum_{t=1}^{n} \sum_{j=-m}^{m} c_j^2 Z_{t-j}^2 > a_{np}^2 x\right) = x^{-\alpha/2} \left(\sum_{j=-m}^{m} c_j^2\right)^{\frac{\alpha}{2}}.$$
 (2.3.12)

This and the positivity of the summands implies

$$\liminf_{n \to \infty} pP\left(\sum_{t=1}^{n} \sum_{j=-\infty}^{\infty} c_j^2 Z_{t-j}^2 > a_{np}^2 x\right) \ge x^{-\alpha/2} \left(\sum_{j=-\infty}^{\infty} c_j^2\right)^{\frac{\alpha}{2}}.$$
 (2.3.13)

Thus it is only left to show that the limsup is bounded by the right hand side of (2.3.13). Using Markov's inequality yields

$$pP\left(\sum_{t=1}^{n}\sum_{j=-\infty}^{\infty}c_{j}^{2}Z_{t-j}^{2} > a_{np}^{2}x\right) \leq \sum_{j=-\infty}^{\infty}pnP\left(c_{j}^{2}Z_{1}^{2} > a_{np}^{2}x\right) + \sum_{j=-\infty}^{\infty}c_{j}^{2}\frac{pn}{a_{np}^{2}x}E\left(Z_{1}^{2}\mathbf{1}_{\{c_{j}^{2}Z_{1}^{2} \leq a_{np}^{2}x\}}\right).$$

Since $E\left(Z_1^2 \mathbf{1}_{\{Z_1^2 \le \cdot\}}\right)$ is a regularly varying function with index $\alpha/2 - 1$ we obtain, by Potter's bound, Karamata's Theorem and (2.2.5), that, for some constant $C_1 > 0$,

$$c_{j}^{2} \frac{pn}{a_{np}^{2} x} E\left(Z_{1}^{2} \mathbf{1}_{\{c_{j}^{2} Z_{1}^{2} \leqslant a_{np}^{2} x\}}\right) = \frac{c_{j}^{2}}{x} \frac{E\left(Z_{1}^{2} \mathbf{1}_{\{c_{j}^{2} Z_{1}^{2} \leqslant a_{np}^{2} x\}}\right)}{E\left(Z_{1}^{2} \mathbf{1}_{\{Z_{1}^{2} \leqslant a_{np}^{2} x\}}\right)} \frac{pn}{a_{np}^{2}} E\left(Z_{1}^{2} \mathbf{1}_{\{Z_{1}^{2} \leqslant a_{np}^{2} x\}}\right)$$
$$\leq C_{1} \frac{c_{j}^{2}}{x} \left(c_{j}^{-2}\right)^{1-\alpha/2+(\alpha/2-\delta/2)} x^{1-\alpha/2} = C_{1} x^{-\alpha/2} |c_{j}|^{\delta}.$$

Likewise, $pnP(a_{np}^{-2}Z_1^2 > \cdot)$ is a regularly varying function with index $\alpha/2$, thus we obtain, by the same arguments as before, that

$$pnP\left(c_j^2 Z_1^2 > a_{np}^2 x\right) \leq C_2 x^{-\alpha/2} |c_j|^{\delta}.$$

With $C = C_1 + C_2$ this therefore implies

$$\limsup_{n \to \infty} pP\left(\sum_{t=1}^{n} \sum_{j=-\infty}^{\infty} c_j^2 Z_{t-j}^2 > a_{np}^2 x\right) \le C \sum_{j=-\infty}^{\infty} |c_j|^{\delta} x^{-\alpha/2}.$$
(2.3.14)

Hence, by (2.3.12) and (2.3.14), we finally have, for some $\epsilon \in (0, 1)$, that

$$\limsup_{n \to \infty} pP\left(\sum_{t=1}^{n} \sum_{j=-\infty}^{\infty} c_j^2 Z_{t-j}^2 > a_{np}^2 x\right) \leq \limsup_{n \to \infty} pP\left(\sum_{t=1}^{n} \sum_{j=-m}^{m} c_j^2 Z_{t-j}^2 > (1-2\epsilon) a_{np}^2 x\right) + \limsup_{n \to \infty} pP\left(\sum_{t=1}^{n} \sum_{j=-\infty}^{\infty} c_j^2 Z_{t-j}^2 > \epsilon a_{np}^2 x\right) + \limsup_{n \to \infty} pP\left(\sum_{t=1}^{n} \sum_{j=-\infty}^{-m-1} c_j^2 Z_{t-j}^2 > \epsilon a_{np}^2 x\right) \\ \leq x^{-\alpha/2} \left((1-2\epsilon)^{-\alpha/2} \left(\sum_{j=-m}^{m} c_j^2\right)^{\frac{\alpha}{2}} + C\epsilon^{-\alpha/2} \sum_{j=m+1}^{\infty} |c_j|^{\delta} + C\epsilon^{-\alpha/2} \sum_{j=-\infty}^{-m-1} |c_j|^{\delta}\right). \quad (2.3.15)$$

Assumption (2.2.5) shows that the last two terms in (2.3.15) vanish for $m \to \infty$. Letting $\epsilon \to 0$ thereafter completes the proof.

By virtue of the previous proposition we obtain the point process convergence of the diagonal elements of the sample covariance matrix XX^{T} . This immediately characterizes their extremal behavior. Note that this result holds without any restriction on β even if $\alpha \ge 1$.

Proposition 2.3.6. *Let* $X = (X_{it})$ *be as in equations* (2.2.1), (2.2.4) *and* (2.2.5) *with* $\alpha \in (0, 2)$, *and suppose that* (2.2.8) *holds for some* $\beta > 0$. *Then*

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2} \sum_{t=1}^{n} X_{it}^{2}} \xrightarrow{D}_{n \to \infty} N = \sum_{i=1}^{\infty} \epsilon_{\Gamma_{i}^{-2/\alpha} \sum_{j=-\infty}^{\infty} c_{j}^{2}}, \qquad (2.3.16)$$

with (Γ_i) as in (2.2.9).

Proof. For notational simplicity we assume without loss of generality that

$$X_{it} = \sum_{j=0}^{\infty} c_j Z_{i,t-j}.$$

The extension to the non-causal case is obvious. We first prove the claim for finite linear processes $X_{it,m} = \sum_{i=0}^{m} c_j Z_{i,t-j}$. From Proposition 2.3.5 we already have that

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2} \sum_{t=1}^{n} \sum_{j=0}^{m} c_j^2 Z_{i,t-j}^2} \xrightarrow{D}_{n \to \infty} \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} \sum_{j=0}^{m} c_j^2}.$$
(2.3.17)

Thus it is only left to show that all terms involving cross products are negligible. By [72, Theorem 4.2] it suffices to show, for any $\eta > 0$, that

$$\lim_{n \to \infty} P\left(\sum_{i=1}^{p} \left| f\left(a_{np}^{-2} \sum_{t=1}^{n} X_{it,m}^{2}\right) - f\left(a_{np}^{-2} \sum_{t=1}^{n} \sum_{j=0}^{m} c_{j}^{2} Z_{i,t-j}^{2}\right) \right| > \eta \right) = 0$$
(2.3.18)

for any continuous function $f : \mathbb{R}_+ \to \mathbb{R}_+$ with compact support supp $(f) \subset [c, \infty]$ and c > 0. Choose some $0 < \gamma < c$ and let $K = [c - \gamma, \infty]$. On the set

$$A_{n}^{\gamma} = \left\{ \max_{1 \le i \le p} \left| \sum_{t=1}^{n} X_{it,m}^{2} - \sum_{t=1}^{n} \sum_{j=0}^{m} c_{j}^{2} Z_{i,t-j}^{2} \right| \le a_{np}^{2} \gamma \right\}$$

the following is true: if $a_{np}^{-2} \sum_{t=1}^{n} \sum_{j=0}^{m} c_j^2 Z_{i,t-j}^2 \notin K$, then the absolute difference in (2.3.18) is zero, else it is bounded by the modulus of continuity $\omega(\gamma) = \sup\{|f(x) - f(y)| : |x - y| \le \gamma\}$. Hence, the probability in (2.3.18) is bounded by

$$P\left(\omega(\gamma)\sum_{i=1}^{p}\epsilon_{a_{np}^{-2}\sum_{t=1}^{n}\sum_{j=0}^{m}c_{j}^{2}Z_{i,t-j}^{2}}(K)>\eta\right)+P\left(\left(A_{n}^{\gamma}\right)^{c}\right).$$

By (2.3.17), the first summand converges to

$$P\bigg(\omega(\gamma)\sum_{i=1}^{\infty}\epsilon_{\sum_{j=0}^{m}c_{j}^{2}\Gamma_{i}^{-2/\alpha}}(K)>\eta\bigg).$$

Since $\sum_{i=1}^{\infty} \epsilon_{\sum_{j=0}^{m} c_{j}^{2} \Gamma_{i}^{-2/\alpha}}(K) < \infty$ and $\omega(\gamma) \to 0$ as $\gamma \to 0$, this probability approaches zero as γ tends to zero. To show that

$$P((A_n^{\gamma})^c) \le P\left(2\sum_{j=0}^{m-1}\sum_{k=j+1}^m |c_j c_k| \max_{i=1:p} \sum_{t=1}^n |Z_{i,t-j} Z_{i,t-k}| > a_{np}^2 \gamma\right) \underset{n \to \infty}{\longrightarrow} 0$$
(2.3.19)

2. Limit Theory for the largest eigenvalues of sample covariance matrices

we use the following observation for fixed $j \in \{0, ..., m-1\}$ and $k \in \{j+1, ..., m\}$: the product $Z_{i,t-j}Z_{i,t-k}$ has, because of independence, tail index α , and $Z_{i,t-j}Z_{i,t-k}$ and $Z_{i,s-j}Z_{i,s-k}$ are independent if and only if $|s-t| \neq k-j$. Thus, we partition the natural numbers \mathbb{N} into k-j+1 pairwise disjoint sets $s + (k-j+1)\mathbb{N}_0$, $s \in \{0, ..., k-j\}$. Then we have, by Proposition 2.3.2 and the independence of the summands, that

$$a_{np}^{-2} \max_{1 \leq i \leq p} \sum_{t \in s+(k-j+1)\mathbb{N}_0} |Z_{i,t-j}Z_{i,t-k}| \xrightarrow{P}_{n \to \infty} 0,$$

for each $s \in \{0, ..., k - j\}$. Since j, k only vary over finite sets this implies (2.3.19). Therefore we have shown (2.3.16) for a finite order moving average $X_{it,m}$.

Now we let m go to infinity. Clearly, we have that

$$\sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} \sum_{j=0}^m c_j^2} \xrightarrow{D}_{m \to \infty} \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} \sum_{j=0}^\infty c_j^2}.$$
(2.3.20)

Thus, by [17, Theorem 3.2], it is only left to show that

$$\lim_{m \to \infty} \limsup_{n \to \infty} P\left(\sum_{i=1}^{p} \left| f\left(a_{np}^{-2} \sum_{t=1}^{n} X_{it}^{2}\right) - f\left(a_{np}^{-2} \sum_{t=1}^{n} X_{it,m}^{2}\right) \right| > \eta\right) = 0.$$
(2.3.21)

By repeating the previous arguments, it suffices to show

$$\limsup_{n \to \infty} P\left(a_{np}^{-2} \max_{1 \le i \le p} \sum_{t=1}^{n} |X_{it}^2 - X_{it,m}^2| > \gamma\right) \le \limsup_{n \to \infty} pP\left(a_{np}^{-2} \sum_{t=1}^{n} |X_{1t}^2 - X_{1t,m}^2| > \gamma\right) \to 0,$$

as $m \to \infty$. Clearly, we have that

$$X_{1t}^{2} - X_{1t,m}^{2} = \sum_{j=m+1}^{\infty} c_{j}^{2} Z_{1,t-j}^{2} + 2 \sum_{j=m+1}^{\infty} \sum_{k=0}^{m} c_{j} c_{k} Z_{1,t-j} Z_{1,t-k} + \sum_{\substack{j=m+1\\k\neq j}}^{\infty} \sum_{k=m+1}^{\infty} c_{j} c_{k} Z_{1,t-j} Z_{1,t-k}.$$
(2.3.22)

For the first summand on the right hand side of equation (2.3.22) we have, by Proposition 2.3.5, that

$$pP\left(\sum_{t=1}^{n}\sum_{j=m+1}^{\infty}c_{j}^{2}Z_{1,t-j}^{2} > \eta a_{np}^{2}\right) \xrightarrow[n \to \infty]{} \left(\sum_{j=m+1}^{\infty}c_{j}^{2}\right)^{\alpha/2} \eta^{-\alpha/2} \xrightarrow[m \to \infty]{} 0.$$

Using Proposition 2.3.5 and the elementary inequality $2|ab| \le a^2 + b^2$, we obtain for the second term in equation (2.3.22) that

$$pP\left(2\sum_{t=1}^{n}\sum_{j=m+1}^{\infty}\sum_{k=0}^{m}|c_{j}c_{k}Z_{1,t-j}Z_{1,t-k}| > \eta a_{np}^{2}\right) \leq pP\left(\sum_{t=1}^{n}\sum_{j=m+1}^{\infty}\sum_{k=0}^{m}|c_{j}c_{k}|Z_{1,t-j}^{2} > \frac{\eta}{2}a_{np}^{2}\right) + pP\left(\sum_{t=1}^{n}\sum_{j=m+1}^{\infty}\sum_{k=0}^{m}|c_{j}c_{k}|Z_{1,t-k}^{2} > \frac{\eta}{2}a_{np}^{2}\right) \sim 2\frac{\eta}{4}^{-\alpha/2}\left(\sum_{k=0}^{m}|c_{k}|\right)^{\alpha/2}\left(\sum_{j=m+1}^{\infty}|c_{j}|\right)^{\alpha/2},$$

and since $\sum_{j=0}^{\infty} |c_j| < \infty$, this term converges to zero as $m \to \infty$. The third term in equation (2.3.22) can be handled similarly. Thus the proof is complete.

2.3.4. Proof of Theorem 2.1

In this section we use the foregoing results to complete the proof of Theorem 2.1.

Proof of Theorem 2.1. (i) Denote by $S_k = (XX^T)_{kk} = \sum_{t=1}^n X_{kt}^2$ the diagonal entries of XX^T and by $S_{(1)} \ge \ldots \ge S_{(p)}$ their order statistics. Then Weyl's Inequality, cf. [15, Corollary III.2.6], and Proposition 2.3.4 imply that

$$a_{np}^{-2} \max_{1 \le k \le p} |\lambda_{(k)} - S_{(k)}| \le a_{np}^{-2} \left\| X X^{\mathsf{T}} - D \right\|_{2} \xrightarrow{P}_{n \to \infty} 0,$$
(2.3.23)

where $D = \text{diag}(XX^{\mathsf{T}})$. From Proposition 2.3.6 we have

$$\widehat{N}_n = \sum_{i=1}^p \epsilon_{a_{np}^{-2} \mathcal{S}_{(i)}} \xrightarrow[n \to \infty]{D} N.$$
(2.3.24)

Thus, by [72, Theorem 4.2], it suffices to show that

$$P(|\widehat{N}_n(f) - N_n(f)| > \eta) \le P\left(\sum_{i=1}^p \left| f\left(\frac{S_{(i)}}{a_{np}^2}\right) - f\left(\frac{\lambda_{(i)}}{a_{np}^2}\right) \right| > \eta\right) \underset{n \to \infty}{\longrightarrow} 0$$

for a nonnegative continuous function f with compact support supp $(f) \in [c, \infty]$, for some c > 0. Since $N((c/2, \infty]) < \infty$ almost surely, we can choose some $i \in \mathbb{N}$ large enough such that the probability $P(N((c/2, \infty]) \ge i) < \delta/2$. By (2.3.24), it follows that $P(a_{np}^{-2}S_{(i)} > c/2) = P(\widehat{N}_n((c/2, \infty]) \ge i) \rightarrow P(N((c/2, \infty]) \ge i)$ and thus, for n large enough, $P(a_{np}^{-2}S_{(i)} > c/2) < c/2$ and $a_{np}^{-2}\lambda_{(i)} < c$ imply that both $f(a_{np}^{-2}M_{(k)}) = 0$ and $f(a_{np}^{-2}\lambda_{(k)}) = 0$ for all $k \ge i$, we obtain

$$\begin{split} &P\left(\sum_{j=1}^{p}\left|f\left(\frac{S_{(j)}}{a_{np}^{2}}\right) - f\left(\frac{\lambda_{(j)}}{a_{np}^{2}}\right)\right| > \eta\right) \leqslant P\left(\sum_{j=1}^{p}\left|f\left(\frac{S_{(j)}}{a_{np}^{2}}\right) - f\left(\frac{\lambda_{(j)}}{a_{np}^{2}}\right)\right| > \eta, a_{np}^{-2}S_{(i)} < \frac{c}{2}\right) \\ &+ P\left(\sum_{j=1}^{p}\left|f\left(\frac{S_{(j)}}{a_{np}^{2}}\right) - f\left(\frac{\lambda_{(j)}}{a_{np}^{2}}\right)\right| > \eta, a_{np}^{-2}S_{(i)} < \frac{c}{2}, a_{np}^{-2}\lambda_{(i)} > c\right) \\ &+ P\left(\sum_{j=1}^{p}\left|f\left(\frac{S_{(j)}}{a_{np}^{2}}\right) - f\left(\frac{\lambda_{(j)}}{a_{np}^{2}}\right)\right| > \eta, a_{np}^{-2}S_{(i)} < \frac{c}{2}, a_{np}^{-2}\lambda_{(i)} < c\right) \\ \leqslant 3\delta + P\left(\sum_{j=1}^{i-1}\left|f\left(\frac{S_{(j)}}{a_{np}^{2}}\right) - f\left(\frac{\lambda_{(j)}}{a_{np}^{2}}\right)\right| > \eta\right), \end{split}$$

which becomes arbitrarily small due to equation (2.3.23) and the fact that f is uniformly continuous.

(ii) By assumption $X = (Z_{it})$. First we consider the case (a) and assume that $\kappa \ge 1$. We will show that, for any fixed positive integer *k*,

$$\frac{\lambda_{(k)}}{S_{(k)}} \xrightarrow[n \to \infty]{P} 1.$$
(2.3.25)

Equations (2.3.4) and (2.3.25) then imply

$$\left|\frac{S_{(k)}}{a_{np}^2} - \frac{\lambda_{(k)}}{a_{np}^2}\right| = \left|1 - \frac{\lambda_{(k)}}{S_{(k)}}\right| \frac{S_{(k)}}{a_{np}^2} \xrightarrow{P}_{n \to \infty} 0,$$

and hence $N_n \to N$ as in the proof of Theorem 2.1 (i). Define $M_i = \max_{1 \le t \le n} X_{it}^2$ and denote by $M_{(1)} \ge \ldots \ge M_{(p)}$ the order statistics of M_1, \ldots, M_p . Observe that the continuous mapping theorem applied to (2.3.4) and (2.3.5) yields, for any fixed k,

$$\frac{S_{(k)}}{M_{(k)}} \xrightarrow{P} 1, \text{ and } \frac{\|X\|_{\infty}^2}{M_{(1)}} \xrightarrow{P} 1,$$

because $\kappa \ge 1$. Now we start showing (2.3.25) by induction. For k = 1 we have, on the one hand, that

$$\frac{\lambda_{(1)}}{S_{(1)}} = \frac{\left\| X X^{\mathsf{T}} \right\|_{2}}{S_{(1)}} \leq \frac{\| X \|_{2}^{2}}{S_{(1)}} \leq \frac{\| X \|_{\infty}^{2}}{S_{(1)}} = \frac{\| X \|_{\infty}^{2}}{M_{(1)}} \frac{M_{(1)}}{S_{(1)}} \xrightarrow{P}{n \to \infty} 1.$$

Let us denote by e_1, \ldots, e_p the standard Euclidean orthonormal basis in \mathbb{R}^p and by i_1 the (random) index that satisfies $S_{i_1} = S_{(1)}$. Then we have, on the other hand, by the Minimax Principle [15, Corollary III.1.2], that

$$\frac{\lambda_{(1)}}{S_{(1)}} = \frac{\max_{v \in \mathbb{R}^p} \left\langle v, XX^{\mathsf{T}}v \right\rangle}{S_{(1)}} \ge \frac{\left\langle e_{i_1}, XX^{\mathsf{T}}e_{i_1} \right\rangle}{S_{(1)}} = \frac{S_{i_1}}{S_{(1)}} = 1$$

This shows (2.3.25) for k = 1. To keep the notation simple, we describe the induction step only for k = 2. The arguments for the general case are exactly the same. Denote by i_2 the random index such that $S_{i_2} = S_{(2)}$. Let $X^{(2)}$ be the $(p-1) \times n$ matrix which is obtained from removing row i_1 from X_n and denote by $\varrho_{(1)}$ the largest eigenvalue of $X^{(2)}(X^{(2)})^{\mathsf{T}}$. Since we have already shown the claim for the largest eigenvalue, it follows that $\varrho_{(1)}/S_{(2)} \to 1$ in probability. By the Cauchy Interlacing Theorem [15, Corollary III.1.5] this implies $\lambda_{(2)}/S_{(2)} \leq \varrho_{(1)}/S_{(2)} \to 1$. Another application of the Minimax Principle yields

$$\begin{split} \lambda_{(2)} &= \max_{\substack{\mathcal{M} \subset \mathbb{R}^{p} \\ \dim(\mathcal{M}) = 2}} \min_{\substack{v \in \mathcal{M} \\ \|v\| = 1}} v^{\mathsf{T}} X X^{\mathsf{T}} v \ge \min_{\substack{v \in \operatorname{span}\{e_{i_{1}}, e_{i_{2}}\} \\ \|v\| = 1}} v^{\mathsf{T}} X X^{\mathsf{T}} v \\ &= \min_{\mu_{1}, \mu_{2} \in \mathbb{R}} (\mu_{1}^{2} + \mu_{2}^{2})^{-1} \left(\mu_{1}^{2} S_{(1)} + \mu_{2}^{2} S_{(2)} + 2\mu_{1} \mu_{2} (X X^{\mathsf{T}})_{i_{1} i_{2}} \right). \end{split}$$

Since, by Proposition 2.3.2 and equation (2.3.4),

$$\frac{\left|\frac{2\mu_{1}\mu_{2}}{\mu_{1}^{2}+\mu_{2}^{2}}(XX^{\mathsf{T}})_{i_{1}i_{2}}\right|}{S_{(2)}} \leqslant \frac{a_{np}^{-2}\max_{1\leqslant i< j\leqslant p}\sum_{t=1}^{n}|Z_{it}Z_{jt}|}{a_{np}^{-2}S_{(2)}} \xrightarrow{P}_{n\to\infty} 0.$$

uniformly in $\mu_1, \mu_2 \in \mathbb{R}$, an application of the the continuous mapping theorem finally yields that $\lambda_{(2)}/S_{(2)} \ge 1 + o_P(1)$, where $o_P(1) \to 0$ in probability as $n \to \infty$. Thus the proof for $\kappa \ge 1$ is complete. Now let $\kappa \in (0, 1)$. Since $X^T X$ and XX^T have the same non-trivial eigenvalues, we consider the transpose X^T of X. This inverts the roles of p and n. Therefore, using Potter's bounds and $1/\kappa > 1$, the result follows from the same arguments as before. Note that we are in a special case of Theorem 2.1 (i) if $\kappa = 0$. In case (b) we have that $n \sim (1/c \log(p/C))^{1/\kappa}$ is a slowly varying function in p, thus an application of Theorem 2.1 (ii) (a) to X^T gives the result.

2.3.5. Proof of Theorem 2.2

As we shall see, the proof of Theorem 2.2 will more or less follow the same lines of argument as given for Theorem 2.1. We focus on the setting of Theorem 2.2 (i) here and mention (ii) and (iii) later. The next result is a generalization of Proposition 2.3.6 allowing for random coefficients.

Proposition 2.3.7. Define $X = (X_{it})$ with X_{it} satisfying (2.2.11) and (2.2.12). Suppose (2.2.8) holds for some $\beta > 0$. If (θ_i) is a stationary ergodic sequence, then, conditionally on (θ_i) as well as unconditionally, we have

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2} \sum_{t=1}^{n} X_{it}^{2}} \xrightarrow{D}_{n \to \infty} \sum_{i=1}^{\infty} \epsilon_{\Gamma_{i}^{-2/\alpha} \left(E \left| \sum_{j=-\infty}^{\infty} c_{j}^{2}(\theta_{1}) \right|^{\alpha/2} \right)^{2/\alpha}}$$
(2.3.26)

with (Γ_i) as in (2.2.9).

Proof. We first prove that, conditionally on (θ_i) ,

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2} \sum_{t=1}^{n} \sum_{j} c_{j}^{2}(\theta_{i}) Z_{i,t-j}^{2}} \xrightarrow[n \to \infty]{} \sum_{i=1}^{\infty} \sum_{r=1}^{\infty} \epsilon_{\Gamma_{i}^{-2/\alpha} \left(E \left| \sum_{j} c_{j}^{2}(\theta_{1}) \right|^{\alpha/2} \right)^{2/\alpha}}$$
(2.3.27)

by showing a.s. convergence of the Laplace functionals. By arguments from the proof of [96, Proposition 3.17] it suffices to show (2.2.13) only for a countable subset of the space of all nonnegative continuous functions with compact support. Thus we fix one nonnegative continuous function f with compact support supp $(f) \subset [c, \infty], c > 0$. Conditionally on the

2. Limit Theory for the largest eigenvalues of sample covariance matrices

process (θ_m) , the points of the point process are independent, and thus

$$E\left(e^{-\sum_{i=1}^{p} f(a_{np}^{-2} \sum_{i=1}^{n} \sum_{j} c_{j}^{2}(\theta_{i}) Z_{i,t-j}^{2})} \middle| (\theta_{m})\right)$$
(2.3.28)
$$= \prod_{i=1}^{p} \left(1 - \frac{1}{p} \int (1 - e^{-f(x)}) p P\left(a_{np}^{-2} \sum_{t=1}^{n} \sum_{j} c_{j}^{2}(\theta_{i}) Z_{1,t-j}^{2} \in dx \middle| \theta_{i}\right)\right)$$
$$= \prod_{i=1}^{p} \left(1 - \frac{1}{p} B_{i,p}\right),$$
(2.3.29)

where $B_{i,p} = \int (1 - e^{-f(x)}) p P(a_{np}^{-2} \sum_{t=1}^{n} \sum_{j} c_{j}^{2}(\theta_{i}) Z_{1,t-j}^{2} \in dx | \theta_{i})$. First assume

$$\frac{1}{p} \sum_{i=1}^{p} B_{i,p} \xrightarrow[n \to \infty]{a.s.} B \coloneqq \int (1 - e^{-f(x)}) \nu(dx)$$
(2.3.30)

with ν given by $\nu((x, \infty]) \coloneqq x^{-\alpha/2} E \left| \sum_j c_j^2(\theta_1) \right|^{\alpha/2}$, and

$$\frac{1}{p^2} \sum_{i=1}^{p} B_{i,p}^2 \xrightarrow[n \to \infty]{a.s.} 0.$$
(2.3.31)

Both claims will be justified later. By assumption (2.2.11), we have, using Proposition 2.3.5, almost surely

$$B_{i,p} \leq pP\left(a_{np}^{-2}\sum_{t=1}^{n}\sum_{j}\tilde{c}_{j}^{2}Z_{1,t-j}^{2} > c\right) \underset{n \to \infty}{\longrightarrow} c^{-\alpha/2} \left|\sum_{j}\tilde{c}_{j}^{2}\right|^{\alpha/2}$$

and hence there exists a C > 0 such that $B_{i,p} \leq C$ for all $i, p \in \mathbb{N}$ a.s. The elementary inequality $e^{\frac{-x}{1-x}} \leq 1-x \leq e^{-x} \forall x \in [0,1]$, equivalently $e^{\frac{-x^2}{1-x}} \leq (1-x)e^x \leq 1 \forall x \in [0,1]$, implies together with (2.3.31), for some $c_1 > 0$, that

$$1 \ge \prod_{i=1}^{p} \left(1 - \frac{B_{i,p}}{p}\right) e^{\frac{B_{i,p}}{p}} \ge \prod_{i=1}^{p} e^{-\frac{B_{i,p}^{2}}{p^{2} - pB_{i,p}}} \ge \prod_{i=1}^{p} e^{-\frac{B_{i,p}^{2}}{p^{2} - pC}} \ge e^{\frac{-c_{1}}{p^{2}} \sum_{i=1}^{p} B_{i,p}^{2}} \xrightarrow[n \to \infty]{a.s.}$$

As a consequence we have that the product in (2.3.29) is asymptotically equivalent to

$$\prod_{i=1}^{p} e^{-\frac{1}{p}B_{i,p}} = e^{-\frac{1}{p}\sum_{i=1}^{p}B_{i,p}} \xrightarrow[n \to \infty]{a.s.} e^{-B} = e^{-\int (1-e^{-f(x)})\nu(dx)},$$

where the convergence follows from (2.3.30). This implies the almost sure convergence of the conditional Laplace functionals, therefore (2.3.27) holds conditionally on (θ_i) . Using (2.2.11) one shows similarly as in the proof of Proposition 2.3.6, conditionally on (θ_i) , that (2.3.27) implies (2.3.26). Taking the expectation yields that (2.3.26) also holds unconditionally.

Proof of (2.3.30) *and* (2.3.31). As a function in x, $pP(\sum_{t=1}^{n} Z_{1t}^2 > a_{np}^2 x)$ is decreasing and converges pointwise to the continuous function $x^{-\alpha/2}$ as $n \to \infty$. Therefore this convergence is uniform on compact intervals of the form $[x_0, \infty]$ with $x_0 > 0$. Now fix x > 0 and let $d_i = \sum_j c_j^2(\theta_i)$. Since $d_i \le d = \sum_j \tilde{c}_j^2 < \infty$ for all $i \in \mathbb{N}$, $\frac{x}{d_i} \ge \frac{x}{d} > 0$ is bounded from below, and thus

$$\sup_{i \in \mathbb{N}} \left| pP\left(\sum_{t=1}^{n} Z_{1t}^2 > a_{np}^2 \frac{x}{d_i} \middle| d_i \right) - x^{-\alpha/2} d_i^{\alpha/2} \right| \xrightarrow[n \to \infty]{a.s.} 0.$$
(2.3.32)

Since (d_i) is an instantaneous function of the ergodic sequence (θ_i) , it is also ergodic and thus

$$\frac{1}{p} \sum_{i=1}^{p} d_i^{\alpha/2} \xrightarrow[n \to \infty]{a.s.} E|d_1|^{\alpha/2}.$$
(2.3.33)

As a consequence of (2.3.32) and (2.3.33) we obtain

$$\left|\frac{1}{p}\sum_{i=1}^{p} pP\left(\sum_{t=1}^{n} Z_{1t}^{2} > a_{np}^{2} \frac{x}{d_{i}} \middle| d_{i}\right) - x^{-\alpha/2} E|d_{1}|^{\alpha/2} \left| \frac{a.s.}{n \to \infty} 0.\right|^{\alpha/2} \right| d_{1} d_{1}$$

Then it is straightforward to show, as in the proof of Proposition 2.3.5, using (2.2.11), that

$$\frac{1}{p} \sum_{i=1}^{p} pP\left(\sum_{t=1}^{n} \sum_{j} c_{j}(\theta_{i}) Z_{1,t-j}^{2} > a_{np}^{2} x \middle| \theta_{i} \right) \xrightarrow[n \to \infty]{a.s.} x^{-\alpha/2} E|d_{1}|^{\alpha/2}$$

The vague convergence of above sequence of measures implies $p^{-1} \sum_{i=1}^{p} B_{i,p} \to B$ almost surely. In exactly the same way one can show that $p^{-1} \sum_{i=1}^{p} B_{i,p}^2$ converges, thus

$$p^{-2} \sum_{i=1}^{p} B_{i,p}^2 \to 0$$
 a.s.,

which establishes (2.3.30) and (2.3.31) as claimed.

Proof of Theorem 2.2. Proof of (i). If we condition on (θ_i) , the proofs of Propositions 2.3.3 and 2.3.4 easily carry over to this more general setting when we make use of assumption (2.2.11). Taking the expectation then yields convergence in operator norm unconditionally. A combination of this together with Proposition 2.3.7 completes the proof.

Proof of (ii). Note that (2.3.33) is the only step in the proof of Proposition 2.3.7 where we use the ergodicity of the sequence (θ_i) . But also if (θ_i) is just stationary, the ergodic theorem implies that the average in (2.3.33) converges to the random variable $Y = E(|d_1|^{\alpha/2}|\mathcal{G})$, where \mathcal{G} is the invariant σ -field generated by (θ_i) . By construction, Y depends on α and $c_j(\cdot)$, but it is independent of (Γ_i) , since (θ_i) is independent of (Z_{it}) .

2. Limit Theory for the largest eigenvalues of sample covariance matrices

Proof of (iii). In this setting (θ_i) is a Markov chain which may not be stationary. But since we derive all results in the proof of Theorem 2.2 (i) conditionally on (θ_i) and then take the expectation, stationarity is in fact not needed. The theory on Markov chains, see [83], in particular their Theorem 17.1.7 for Markov chains on uncountable state spaces, shows that (2.3.33) holds if the expectation is taken with respect to the stationary distribution of the Markov chain.

CHAPTER 3

Limit theory for sample covariance matrices of observations with finite variance but infinite fourth moment

3.1. Introduction and main results

In the statistical analysis of high-dimensional data one often tries to reduce its dimensionality while preserving as much of the variation in the data as possible. One important example of such an approach is the Principal Component Analysis (PCA). PCA makes a linear transformation of the data to a new set of variables, the principal components, which are ordered such that the first few retain most of the variation. Therefore one obtains a lower dimensional representation of the data by retaining only the first few principal components.

The variances of the first k principal components are given by the k-largest eigenvalues of the covariance matrix. Let us collect the samples of our multivariate data in a $p \times n$ matrix X, where we refer to p as the dimension of the data and to n as the sample size. In practice, the true underlying covariance matrix is not available, thus one usually replaces it with the sample covariance matrix $\frac{1}{n}XX^{T}$. To account for large high-dimensional data sets, we study the k-largest eigenvalues of the sample covariance matrix when both the dimension of the data as well as the sample size go to infinity. In Chapter 2 this has been done for observations which are regularly varying with index $\alpha \in (0, 2)$, i.e., observations with infinite variance. However, in many applications where data typically exhibits heavy-tails, like finance, the assumption of an infinite variance might be too strong. Therefore we focus in this article on the case where the observations have finite variances but infinite fourth moments. This assumption is also consistent with the motivation to derive a theoretical framework for the use of PCA for high-dimensional data.

3. Observations with finite variance but infinite fourth moment

For more details on PCA we refer the reader to one of the many textbooks available on this topic, see [5] or [71], for example. The field of research that investigates the spectral properties of large dimensional random matrices has become known as Random Matrix Theory (RMT). There exist several survey articles which stress the close relationship between Random Matrix Theory and multivariate statistics, including PCA, see e.g. [45] and [70]. Some authors have already employed tools from Random Matrix Theory to correct traditional tests or estimators which fail when the dimension of the data cannot be assumed to be negligible compared to the sample size. For example, Bai et al. [8] gives corrections on some likelihood ratio tests that even fail even for moderate dimension (around 20), and El Karoui [46] consistently estimates the spectrum of a large dimensional covariance matrix using Random Matrix Theory.

In the following we assume that *X* is a $p \times n$ matrix with entries

$$X_{it} = \sum_{j=-\infty}^{\infty} c_j Z_{i,t-j}, \quad m \in \mathbb{N}$$
(3.1.1)

where the sequence (c_j) is absolutely summable, $\sum_{j=-\infty}^{\infty} |c_j| < \infty$, and $(Z_{it})_{i,t}$ is an array of iid mean zero random variables with marginal distribution that is regularly varying with tail index $\alpha \in [2, 4)$ and *normalizing sequence* a_n , i.e.,

$$EZ_{11} = 0$$
 and $\lim_{n \to \infty} nP(|Z_{it}| > a_n x) = x^{-\alpha}$, for each $x > 0$. (3.1.2)

In other words, for each $i \in \mathbb{N}$, $(X_{it})_t$ is a infinite order moving average process driven by some regularly varying noise with finite variance but infinite fourth moment. Note that it follows from classical extreme value theory that the sequence a_n is necessarily characterized by

$$a_n = n^{1/\alpha} L(n), \tag{3.1.3}$$

for some slowly varying function $L : \mathbb{R}_+ \to \mathbb{R}_+$, i.e., a function with the property that, for each x > 0, $\lim_{t\to\infty} L(tx)/L(t) = 1$. Moreover we assume that Z_{11} satisfies the *tail balancing condition* given by

$$\lim_{x \to \infty} \frac{P(Z_{11} > x)}{P(|Z_{11}| > x)} = q = 1 - \lim_{x \to \infty} \frac{P(Z_{11} \le -x)}{P(|Z_{11}| > x)}$$
(3.1.4)

for some $0 \le q \le 1$.

Definition 3.1.1. The (normalized) sample covariance matrix of the sample X is defined as the $p \times p$ matrix $S = a_{np}^{-2} (XX^{\mathsf{T}} - n\mu_X I_p)$, where $\mu_X = EZ_{11}^2 \sum_j c_j^2$ and I_p is the identity. We denote by $\lambda_1, \ldots, \lambda_p$ the unordered, and by $\lambda_{(1)} \ge \ldots \ge \lambda_{(p)}$ the ordered eigenvalues of S.

In the following we are going to show that XX^{T} is dominated by its diagonal entries. If $\alpha > 2$, the diagonal entries have a finite mean $n\mu_X = nEZ_{11}^2 \sum_j c_j^2$, which has to be subtracted in order to obtain a non-trivial limiting result.

Remark 3.1.1. If $\alpha = 2$ it is possible that $EZ_{11}^2 = \infty$. In this case we replace μ_X in the above definition by the sequence of truncated means $\mu_X^n = \sum_j c_j^2 E(Z_{11}^2 \mathbf{1}_{\{Z_{11}^2 \leq a_{np}^2\}})$.

We will always assume that $p = p_n$ is an integer-valued sequence in *n* that goes to infinity as $n \to \infty$ in order to obtain results for high-dimensional data. In the following we suppress the dependence of *p* on *n* so as to simplify the notation wherever this does not cause any ambiguity. The following theorem is a generalization of the result of [7] to non-independent entries, except that [7] assume that p/n goes to some positive finite constant, while we assume that *p* is bounded by some small power of *n*.

Theorem 3.1. Define the matrix $X = (X_{it})$ as in equations (3.1.1), (3.1.2) and (3.1.4) with $\alpha \in [2, 4)$. Suppose $n, p \to \infty$ such that

$$\limsup_{n \to \infty} \frac{p_n}{n^\beta} < \infty \tag{3.1.5}$$

for some $\beta > 0$ satisfying

- (i) $\beta < \max\left\{\frac{1}{3}, \frac{4-\alpha}{4(\alpha-1)}\right\}$ if $2 \le \alpha < 3$, or
- (ii) $\beta < \frac{4-\alpha}{3\alpha-4}$ if $3 \le \alpha < 4$.

Then the point process $N_n := \sum_{i=1}^p \epsilon_{\lambda_i}$ of the eigenvalues of *S* converges in distribution to a Poisson point process *N* with intensity measure *v* which is given by

$$v((x,\infty]) = EN(x,\infty] = x^{-\alpha/2} \left| \sum_{j} c_{j}^{2} \right|^{\alpha/2}, \quad x > 0.$$

In particular, the theorem shows that the *k* largest eigenvalues $\lambda_{(1)} \ge ... \ge \lambda_{(k)}$ of *S*, i.e., the variances of the *k*-largest principal components, converge jointly to a random vector with a distribution that only depends on *k*, the tail index α and the coefficients (c_j) . Let (E_j) be an iid sequence of exponentially distributed random variables with mean one, i.e., $P(E_j > x) = e^{-x}$ for $x \ge 0$, and denote by $\Gamma_i = E_1 + ... + E_i$ their successive sum. Then we have that

$$(\lambda_{(1)},\ldots,\lambda_{(k)}) \xrightarrow[n\to\infty]{D} (\Gamma_1^{-2/\alpha},\ldots,\Gamma_k^{-2/\alpha}) \left(\sum_j c_j^2\right).$$

The proof of the theorem splits up into two parts: the convergence of S to its diagonal D_S , and the point process convergence of the entries of D_S . Note that some of the techniques and methods employed in the proof of Theorem 3.1 have also been used in Chapter 2.

Theorem 3.1 can be extended to a random matrix model where, conditionally on a latent process, the processes in the rows have varying coefficients. To this end, let $(\theta_i)_{i \in \mathbb{N}}$ be a

stationary ergodic process that is independent of (Z_{it}) . Further assume that there is a family of measurable functions $(c_j : \Theta \to \mathbb{R})_{j \in \mathbb{N}}$ such that

$$\sup_{\theta \in \Theta} |c_j(\theta)| \leq \widetilde{c}_j, \quad \text{for some absolutely summable } (\widetilde{c}_j). \tag{3.1.6}$$

The matrix $X = (X_{it})$ is then defined by

$$X_{it} = \sum_{j=-\infty}^{\infty} c_j(\theta_i) Z_{i,t-j}$$
(3.1.7)

where (Z_{it}) is given as in (3.1.2) with $\alpha \in [2, 4)$.

We say that a sequence of point processes \mathcal{M}_n converges, conditionally on a sigma-algebra \mathcal{H} , in distribution to a point process \mathcal{M} , if the conditional Laplace functionals converge almost surely, i.e., if there exists a measurable set B with measure one such that for all $\omega \in B$ and all nonnegative continuous functions f with compact support,

$$E\left(e^{-\mathcal{M}_n(f)}|\mathcal{H}\right)(\omega) \to E\left(e^{-\mathcal{M}(f)}|\mathcal{H}\right)(\omega) \text{ as } n \to \infty.$$
 (3.1.8)

Theorem 3.2. Define $X = (X_{it})$ with X_{it} as given in (3.1.7) and suppose that (3.1.6) is satisfied. Under the conditions of Theorem 3.1 we have that the point process $\sum_{i=1}^{p} \epsilon_{\lambda_i}$ of the eigenvalues of S converges in distribution to a Poisson point process with intensity measure given by

$$\nu((x,\infty]) = x^{-\alpha/2} E \left| \sum_{j=-\infty}^{\infty} c_j^2(\theta_1) \right|^{\alpha/2}, \quad x > 0.$$

The convergence also holds true if we condition on (θ_i) *.*

Remark 3.1.2. Suppose that (θ_i) is an irreducible Markov chain on a countable state space or a positive Harris process in the sense of [83]. If (θ_i) possesses a stationary distribution π , then above theorem also holds true for (θ_i) when the expectation is taken with respect to π .

One can view Theorem 3.2 in a Bayesian framework in which the parameters of the observed process are drawn from an unknown prior distribution. Models of this kind are referred to as random coefficient models, see, e.g., [74] for an overview. In the case when (θ_i) is a Markov chain these models are called Hidden Markov Models.

3.2. Proof

3.2.1. Approximation of *S* by its diagonal

Proposition 3.2.1. Let D_S be the diagonal of S, i.e., $(D_S)_{ii} = S_{ii}$ and $(D_S)_{ij} = 0$ for $i \neq j$. Under the conditions of Theorem 3.1 we have that $||S - D_S||_2 \rightarrow 0$ in probability. Proof. First we define the truncated process

$$X_{it}^{L} = \sum_{k} c_k Z_{i,t-k}^{L}, \quad Z_{it}^{L} = Z_{it} \mathbf{1}_{\{|Z_{it}| \leq a_{np}\}}.$$

This proof is parallel to the proof of Proposition 2.3.4 up to the point where it is left to show that

$$p^2 P\left(\left|\sum_{t=1}^n X_{1t}^L X_{2t}^L - n\mu_n\right| > \frac{a_{np}^2}{4p}\epsilon\right) \to 0,$$
(3.2.1)

with

$$\mu_n = (EX_{11}^L)^2 = \left(\sum_k c_k\right)^2 (EZ_{11}^L)^2 = O\left(\frac{a_{np}^2}{(np)^2}\right).$$

By Markov's inequality we have that

$$p^{2}P\left(\left|\sum_{t=1}^{n} X_{1t}^{L} X_{2t}^{L} - n\mu_{n}\right| > \frac{a_{np}^{2}}{4p}\epsilon\right)$$

$$\leq \frac{16p^{4}}{a_{np}^{4}\epsilon^{2}} \sum_{t,t'=1}^{n} \sum_{k,k',l,t'} c_{k}c_{k'}c_{l}c_{l'}\operatorname{Cov}\left(Z_{1,t-k}^{L} Z_{2,t-l}^{L}, Z_{1,t'-k'}^{L} Z_{2,t'-l'}^{L}\right).$$
(3.2.2)

The above covariance is only non-zero if t - k = t' - k' or t - l = t' - l'. In this case it converges to a constant if $\alpha > 2$. If $\alpha = 2$ with $EZ_{11}^2 = \infty$, then it is a slowly varying function. In any case (3.2.2) is of order

$$O\left(\frac{p^4}{a_{np}^4}n\,s(np)\right) \leq O\left(n^{\beta(4-4/\alpha)}n^{1-4/\alpha}s(np)\right) \to 0,$$

since $\beta < \frac{4-\alpha}{4(\alpha-1)}$, where $s(\cdot)$ is some slowly varying function. Let us now assume that $\alpha \in (2,3)$. By Markov's inequality applied to (3.2.1) we have

$$p^{2}P\left(\left|\sum_{t=1}^{n} X_{1t}^{L} X_{2t}^{L} - n\mu_{n}\right| > \frac{a_{np}^{2}}{4p}\epsilon\right)$$

$$\leq \frac{64}{\epsilon^{3}} \frac{p^{5}}{a_{np}^{6}} \sum_{t_{1},t_{2},t_{3}=1}^{n} E\left(\prod_{i=1}^{3} \left(X_{1,t_{i}}^{L} X_{2,t_{i}}^{L} - \mu_{n}\right)\right)$$

$$= \frac{64}{\epsilon^{3}} \frac{p^{5}}{a_{np}^{6}} \sum_{t_{1},t_{2},t_{3}=1}^{n} \sum_{k_{1},k_{2},k_{3}} \sum_{l_{1},l_{2},l_{3}} \prod_{j=1}^{3} (c_{k_{j}}c_{l_{j}}) E\left(\prod_{i=1}^{3} \left(Z_{1,t_{i}-k_{i}}^{L} Z_{2,t_{i}-l_{i}}^{L} - \xi_{n}^{2}\right)\right), \quad (3.2.3)$$

where

$$\xi_n^2 = \frac{\mu_n}{\left(\sum_k c_k\right)^2} = (EZ_{11}^L)^2 = O\left(\frac{a_{np}^2}{(np)^2}\right).$$
(3.2.4)

To determine the order of the expectation in (3.2.3) we have to distinguish various cases. In the following we say that two index pairs (a, b) and (c, d) overlap if a = c or b = d. If there exists a j = 1, 2, 3 such that the index pair $(t_j - k_j, t_j - l_j)$ does not overlap with both the other two, then, due to independence, we are able to factor out the corresponding term and obtain

$$E\left(\prod_{i=1}^{3} \left(Z_{1,t_i-k_i}^{L} Z_{2,t_i-l_i}^{L} - \xi_n^2\right)\right) = E\left(\prod_{i\neq j} \left(Z_{1,t_i-k_i}^{L} Z_{2,t_i-l_i}^{L} - \xi_n^2\right)\right) E\left(Z_{1,t_j-k_j}^{L} Z_{2,t_j-l_j}^{L} - \xi_n^2\right) = 0,$$

since $\xi_n^2 = (EZ_{11}^L)^2 = E\left(Z_{1,t_j-k_j}^L Z_{2,t_j-l_j}^L\right)$. Thus, in any non-trivial case, each index pair does overlap with (at least) one of the other two. Therefore we have at least two equalities of the form $t_i - k_i = t_{(i+1) \mod 3} - k_{(i+1) \mod 3}$ or $t_i - l_i = t_{(i+1) \mod 3} - l_{(i+1) \mod 3}$ for i = 1, 2, 3. Hence t_2 and t_3 are immediately determined by some linear combination of $t = t_1$ and the $k'_i s$ or $l'_i s$. Therefore the triple sum $\sum_{t_1, t_2, t_3 = 1}^n$ is, if we only count terms where the covariance is non-zero, in fact a simple sum $\sum_{t_{i=1}}^n$ and so only has a contribution of order n. Now we have to determine the order of the products $E\left(\prod_{i=1}^3 Z_{1,t_i-k_i}^L Z_{2,t_i-l_i}^L\right)$. If we only have a single power then, by (3.2.4), this gives us

$$E\left(Z_{1,t_i-k_i}^L Z_{2,t_i-l_i}^L\right) = \xi_n^2 = o(1)$$

Since $\alpha > 2$, powers of order two converge to a constant,

$$E\left(\left(Z_{1,t_i-k_i}^L Z_{2,t_i-l_i}^L\right)^2\right) \to \operatorname{Var}(Z_{11})^2.$$

An application of Karamata's theorem yields that

$$E\left(\left(Z_{1,t_i-k_i}^L Z_{2,t_i-l_i}^L\right)^3\right) \sim a_{np}^6 (np)^{-2}.$$

Using the above facts, it is easy to see that

$$E\left(\prod_{i=1}^{3} Z_{1,t_i-k_i}^L Z_{2,t_i-l_i}^L\right) = O\left(a_{np}^6 (np)^{-2}\right).$$
(3.2.5)

Thus we have, using (3.2.4) and (3.2.5), for the expectation in (3.2.3) that

$$\begin{split} E\bigg(\prod_{i=1}^{3} \left(Z_{1,t_{i}-k_{i}}^{L} Z_{2,t_{i}-l_{i}}^{L} - \xi_{n}^{2}\right)\bigg) &= \sum_{k=0}^{3} (-1)^{k} \sum_{J \subseteq \{1,2,3\}, |J|=k} E\bigg(\prod_{i \in \{1,2,3\}\setminus J} Z_{1,t_{i}-k_{i}}^{L} Z_{2,t_{i}-l_{i}}^{L}\bigg) \xi_{n}^{2|J|} \\ &= O\bigg(a_{np}^{6} (np)^{-2} - \frac{a_{np}^{2}}{(np)^{2}} - \frac{a_{np}^{6}}{(np)^{6}}\bigg) \\ &= O\bigg(a_{np}^{6} (np)^{-2}\bigg). \end{split}$$

The last calculation shows that the expectation in (3.2.3) is equal to $E\left(\prod_{i=1}^{3} Z_{1,t_i-k_i}^L Z_{2,t_i-l_i}^L\right)$ plus lower order terms, and that the leading term is of order $a_{np}^6(np)^{-2}$. With this observation we can finally conclude for (3.2.3) that

$$\frac{64}{\epsilon^3} \frac{p^5}{a_{np}^6} \sum_{t_1, t_2, t_3=1}^n \sum_{k_1, k_2, k_3} \sum_{l_1, l_2, l_3} \prod_{j=1}^3 (c_{k_j} c_{l_j}) E\left(\prod_{i=1}^3 \left(Z_{1, t_i-k_i}^L Z_{2, t_i-l_i}^L - \xi_n^2\right)\right)$$
$$= O\left(\frac{64}{\epsilon^3} \frac{p^5}{a_{np}^6} n a_{np}^6 (np)^{-2}\right) = \frac{64}{\epsilon^3} O\left(\frac{p^3}{n}\right) \to 0,$$

which goes to zero by assumption. This completes the proof for $\alpha \in [2,3)$. The method to deal with $\alpha \in [3,4)$ is similar and thus only described briefly. We use Markov's inequality with power four to obtain that the term in (3.2.1) is bounded by

$$\frac{256}{\epsilon^4} \frac{p^6}{a_{np}^8} \sum_{t_1, t_2, t_3, t_4=1}^n \sum_{k_1, k_2, k_3, k_4} \sum_{l_1, l_2, l_3, l_4} \prod_{j=1}^4 (c_{k_j} c_{l_j}) E\left(\prod_{i=1}^4 \left(Z_{1, t_i - k_i}^L Z_{2, t_i - l_i}^L - \xi_n^2\right)\right).$$
(3.2.6)

Observe that the expectation in (3.2.6) is only non-zero if either

- (i) all index pairs $\{(t_i k_i, t_i l_i)\}_{i=1,2,3,4}$ overlap, or
- (ii) there exist exactly two sets of overlapping index pairs, such that no index pair from one set overlaps with an index pair from the other set. We call these two sets disjoint.

Case (i) is similar to the previous case, so that one can see that

$$E\left(\prod_{i=1}^{4} \left(Z_{1,t_i-k_i}^L Z_{2,t_i-l_i}^L - \xi_n^2 \right) \right) = O\left((E((Z_{11}^L)^4))^2 \right) = O\left(\frac{a_{np}^8}{(np)^2} \right),$$

and that the contribution of $\sum_{t_1,t_2,t_3,t_4=1}^{n}$ is of order *n*. Therefore, in this case, the term in (3.2.6) is of the order

$$\frac{256}{\epsilon^4} \frac{p^6}{a_{np}^8} O\left(n \frac{a_{np}^8}{(np)^2}\right) = \frac{256}{\epsilon^4} O\left(\frac{p^4}{n}\right) \to 0.$$

Thus, we only have to determine the contribution in case (ii). Since the two sets of overlapping index pairs are disjoint, we obtain that

$$E\left(\prod_{i=1}^{4} \left(Z_{1,t_i-k_i}^L Z_{2,t_i-l_i}^L - \xi_n^2 \right) \right) = E\left(\left(Z_{11}^L Z_{21}^L - \xi_n^2 \right)^2 \right)^2.$$

Since $\alpha > 2$ this converges to a constant. In contrast to case (i), the contribution of $\sum_{t_1,t_2,t_3,t_4=1}^{n}$ is of order n^2 . This is due to the fact that the two sets of overlapping index pairs are disjoint,

hence only two out of the four indices t_1, \ldots, t_4 are given by linear combinations of the other two and the k's and l's. Therefore (3.2.6) is of the order

$$\frac{256}{\epsilon^4} \frac{p^6}{a_{np}^8} O\left(n^2\right) \to 0$$

The convergence to zero is justified by

$$\frac{p^{6}}{a_{np}^{8}}n^{2} = n^{2-8/\alpha}p^{6-8/\alpha}L(np)^{-8} \le O\left(n^{2-8/\alpha+\beta(6-8/\alpha)}L(n^{\beta+1})^{-8}\right) \to 0,$$

since $\beta < \frac{4-\alpha}{3\alpha-4}$. This completes the proof of Proposition 3.2.1.

3.2.2. Point process convergence and the proof of Theorem 3.1

For any iid sequence (Z_t) with tail index $2 < \alpha < 4$ we have that

$$pP\left(\sum_{t=1}^{n} Z_{t}^{2} - n\mu_{Z} > a_{np}^{2} x\right) \to x^{-\alpha/2}$$
(3.2.7)

where $\mu_Z = EZ_1^2$. Indeed, [62], and in greater generality also [37], show that, for any x > 0,

$$\frac{P\left(\sum_{t=1}^{n} Z_{t}^{2} - n\mu_{Z} > a_{np}^{2} x\right)}{nP\left(Z_{1}^{2} - \mu_{Z} > a_{np}^{2} x\right)} \to 1.$$
(3.2.8)

With $P(Z_1^2 - \mu_Z > a_{np}^2 x) \sim P(Z_1^2 > a_{np}^2 x) \sim p^{-1} x^{-\alpha/2}$, the result follows. Note that (3.2.8) also holds for $\alpha = 2$ if $EZ_{11}^2 < \infty$. In case $EZ_{11}^2 = \infty$ (which can only happen if $\alpha = 2$), one has to replace μ_Z by the sequence of truncated means $\mu_z^n = E(Z_{11}^2 \mathbf{1}_{\{Z_{11}^2 \le a_{np}^2\}})$. For notational simplicity, we exclude infinite variance case in the following. It is treated analogously to the finite variance case, except that everywhere μ_Z has to be replaced by μ_Z^n , μ_X by $\mu_X^n = \sum_k c_k^2 \mu_Z^n$, and finally $\mu_{X,m}$ by $\mu_{X,m}^n = \sum_{|k| \le m} c_k^2 \mu_Z^n$.

First we show the Point process convergence of the entries of D_S in the case where the observations are *m*-dependent.

Lemma 3.2.1 (Finite moving-average). Assume that there exists an $m \in \mathbb{N}$ such that $c_j = 0$ if |j| > m. Then we have, for $2 < \alpha < 4$ and p, n going to infinity, that

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\left(\sum_{t=1}^{n} X_{it}^{2} - n\mu_{X}\right)} \to \sum_{i=1}^{\infty} \epsilon_{\Gamma_{i}^{-2/\alpha} \sum_{j=-m}^{m} c_{j}^{2}},$$
(3.2.9)

where $\mu_X = EX_{11}^2 = \sum_{j=-m}^m c_j^2 \mu_Z = \sum_{j=-m}^m c_j^2 EZ_{11}^2$.

Proof. By the stationary of the Z's we have that

$$P\left(\left|\sum_{j=-m}^{m} c_{j}^{2} \sum_{t=1}^{n} (Z_{1,t}^{2} - \mu_{Z}) - \sum_{t=1}^{n} \sum_{j=-m}^{m} c_{j}^{2} (Z_{1,t-j}^{2} - \mu_{Z})\right| > a_{np}^{2} \eta\right)$$

$$\leq P\left(\sum_{j=-m}^{m} c_{j}^{2} \sum_{t=1-j}^{j} Z_{1,t}^{2} > a_{np}^{2} \eta\right) \to 0.$$

Hence, using (3.2.7), this yields

$$pP\left(\left|\sum_{t=1}^{n}\sum_{j=-m}^{m}c_{j}^{2}(Z_{1,t-j}^{2}-\mu_{Z})\right| > a_{np}^{2}x\right) \to x^{-\alpha/2}\left|\sum_{j=-m}^{m}c_{j}^{2}\right|^{\alpha/2}.$$
(3.2.10)

This immediately implies that

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2} \sum_{t=1}^{n} \sum_{j=-m}^{m} c_{j}^{2}(Z_{i,t-j}^{2} - \mu_{Z})} \rightarrow \sum_{i=1}^{\infty} \epsilon_{\Gamma_{i}^{-2/\alpha} \left| \sum_{j=-m}^{m} c_{j}^{2} \right|^{\alpha/2}}$$

Thus it is only left to show that, for any continuous $f : \mathbb{R}_+ \to \mathbb{R}_+$ with compact support,

$$\lim_{n \to \infty} P\left(\sum_{i=1}^{p} \left| f\left(a_{np}^{-2} \left(\sum_{t=1}^{n} X_{it}^{2} - n\mu_{X}\right)\right) - f\left(a_{np}^{-2} \sum_{t=1}^{n} \sum_{j=-m}^{m} c_{j}^{2} (Z_{i,t-j}^{2} - \mu_{Z})\right) \right| > \eta \right) = 0.$$

Clearly, we have that

$$\left|\sum_{t=1}^{n} X_{it}^{2} - n\mu_{X} - \sum_{t=1}^{n} \sum_{j=-m}^{m} c_{j}^{2} (Z_{i,t-j}^{2} - \mu_{Z})\right| \leq 2 \sum_{j=-m}^{m-1} \sum_{k=j+1}^{m} |c_{j}c_{k}| \left|\sum_{t=1}^{n} Z_{i,t-j}Z_{i,t-k}\right|.$$

Hence, by the arguments of the proof of Proposition 2.3.6, it suffices to show that

$$a_{np}^{-2} \max_{1 \le i \le p} \left| \sum_{t \in J_s} Z_{i,t-j} Z_{i,t-k} \right| \to 0$$

for each fixed $j \in \{-m, ..., m-1\}$, $k \in \{j+1, ..., m\}$ and $s \in \{0, ..., k-j\}$, where $J_s := s + (k - j + 1)\mathbb{N}_0$. Note that $(Z_{i,t-j}Z_{i,t-k})_{t \in J_s}$ is a sequence of iid random variables with mean zero. Therefore we have, by Markov's inequality,

$$\begin{split} P\left(\max_{1\leqslant i\leqslant p}\left|\sum_{t\in J_s}Z_{i,t-j}Z_{i,t-k}\right| > a_{np}^2\eta\right) \leqslant pP\left(\left|\sum_{t\in J_s}Z_{1,t-j}Z_{1,t-k}\right| > a_{np}^2\eta\right) \\ &\leqslant \frac{p}{\eta^2 a_{np}^4}\sum_{t\in J_s}\operatorname{Var}(Z_{1,t-j}Z_{1,t-k}) \\ &\leqslant \frac{pn}{\eta^2 a_{np}^4}(EZ_{11}^2)^2 \\ &= O\left(\frac{pn}{a_{np}^4}\right) = O\left((pn)^{1-4/\alpha}L(pn)^{-4}\right) \to 0 \end{split}$$

since $\alpha < 4$.

3. Observations with finite variance but infinite fourth moment

The previous Lemma can be extended to the general case where $(X_{it})_t$ is an infinite order moving-average.

Proposition 3.2.2 (General case). For $2 < \alpha < 4$ and p, n going to infinity we have that

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\left(\sum_{t=1}^{n} X_{it}^{2} - n\mu_{X}\right)} \rightarrow \sum_{i=1}^{\infty} \epsilon_{\Gamma_{i}^{-2/\alpha} \sum_{j=-\infty}^{\infty} c_{j}^{2}}$$
(3.2.11)

where $\mu_X = EX_{11}^2 = \sum_{j=-\infty}^{\infty} c_j^2 \mu_Z$.

Proof. For notational convenience we assume without loss of generality that $c_j = 0$ for j < 0. To build upon the result of Lemma 3.2.1 we introduce the truncated process

$$X_{it,m} = \sum_{k=0}^{m} c_k Z_{i,t-k}, \quad \mu_{X,m} = E X_{11,m}^2 = \sum_{j=0}^{m} c_j^2 \mu_Z.$$

Due to the fact that

$$\sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} \sum_{j=-m}^m c_j^2} \xrightarrow{D}_{m \to \infty} \sum_{i=1}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} \sum_{j=-\infty}^\infty c_j^2},$$

we only have to show that

$$\lim_{m \to \infty} \limsup_{n \to \infty} P\left(\sum_{i=1}^{p} \left| f\left(a_{np}^{-2} \sum_{t=1}^{n} (X_{it}^{2} - \mu_{X})\right) - f\left(a_{np}^{-2} \sum_{t=1}^{n} (X_{it,m}^{2} - \mu_{X,m})\right) \right| > \gamma\right) = 0$$

for any continuous f with compact support and $\gamma > 0$. By the arguments given in Proposition 2.3.6 it suffices to show that

$$\lim_{m \to \infty} \limsup_{n \to \infty} pP\left(\left| \sum_{t=1}^{n} (X_{1t}^2 - X_{1t,m}^2 - (\mu_X - \mu_{X,m})) \right| > a_{np}^2 \gamma \right) = 0.$$

Clearly, we have that

$$\begin{split} pP\left(\left|\sum_{t=1}^{n} (X_{1t}^{2} - X_{1t,m}^{2} - (\mu_{X} - \mu_{X,m}))\right| > a_{np}^{2}\gamma\right) \\ \leqslant pP\left(\left|\sum_{k=m+1}^{\infty} c_{k}^{2} \sum_{t=1}^{n} (Z_{1,t-k}^{2} - \mu_{Z})\right| > a_{np}^{2} \frac{\gamma}{3}\right) \\ &+ pP\left(2\left|\sum_{k=m+1}^{\infty} \sum_{l=0}^{m} c_{k}c_{l} \sum_{t=1}^{n} Z_{1,t-k}Z_{1,t-l}\right| > a_{np}^{2} \frac{\gamma}{3}\right) \\ &+ pP\left(2\left|\sum_{k=m+1}^{\infty} \sum_{l=k+1}^{\infty} c_{k}c_{l} \sum_{t=1}^{n} Z_{1,t-k}Z_{1,t-l}\right| > a_{np}^{2} \frac{\gamma}{3}\right) \\ &= I + II + III \end{split}$$

We will show in turn that I, II, III $\rightarrow 0$. We begin with I. Clearly, there either exist a *t* and a *k* such that $|c_k Z_{1,t-k} > a_{np}|$, or $|c_k Z_{1,t-k} \leq a_{np}|$ for all *t*, *k*. This simple fact and Chebyshev's inequality yield

$$\begin{split} \mathbf{I} &= pP\left(\left|\sum_{k=m+1}^{\infty} c_k^2 \sum_{t=1}^n (Z_{1,t-k}^2 - \mu_Z)\right| > a_{np}^2 \frac{\gamma}{3}\right) \\ &\leqslant \sum_{k=m+1}^{\infty} pnP(|c_k Z_{1,1-k} > a_{np}|) + \frac{3}{\gamma} \frac{p}{a_{np}^4} \operatorname{Var}\left(\sum_{k=m+1}^{\infty} c_k^2 \sum_{t=1}^n Z_{1,t-k}^2 \mathbf{1}_{\{|c_k Z_{1,t-k}| \leqslant a_{np}\}}\right) \\ &+ p \mathbf{1}_{\{\sum_{k=m+1}^{\infty} c_k^2 n E\left(Z_{11}^2 \mathbf{1}_{\{|c_k Z_{1,t-k}| > a_{np}\}}\right) > a_{np}^2 \frac{\gamma}{3}\}} = \mathbf{I}_1 + \mathbf{I}_2 + \mathbf{I}_3 \end{split}$$

For the first term we have by Karamata's theorem that

$$\lim_{m\to\infty}\limsup_{n\to\infty}\mathrm{I}_1=\lim_{m\to\infty}\sum_{k=m+1}^\infty c_k^\alpha=0.$$

Another application of Karamata's theorem shows that

$$E\left(Z_{11}^2 \mathbf{1}_{\{|c_k Z_{1,t-k}| > a_{np}\}}\right) \sim |c_k|^{\alpha/2-1} \frac{a_{np}^2}{np},$$

therefore

$$\lim_{k \to \infty} \frac{p \mathbf{1}_{\{\sum_{k=m+1}^{\infty} c_k^{\alpha/2+1} > p_{\overline{3}}^{\gamma}\}}}{\mathbf{I}_3} = 1.$$

However, $p\mathbf{1}_{\{\sum_{k=m+1}^{\infty}c_k^{\alpha/2+1}>p_3^{\frac{\gamma}{3}}\}}=0$ for *n* sufficiently large, since $p=p_n\to\infty$ and

$$\sum_{k=m+1}^{\infty} c_k^{\alpha/2+1} < \infty.$$

As a consequence, $I_3 \rightarrow 0$. Regarding I_2 , observe that the covariance in

$$\mathbf{I}_{2} = \frac{3}{\gamma} \frac{p}{a_{np}^{4}} \sum_{k=m+1}^{\infty} \sum_{k'=m+1}^{\infty} c_{k}^{2} c_{k'}^{2} \sum_{t=1}^{n} \sum_{t'=1}^{n} \operatorname{Cov}\left(Z_{1,t-k}^{2} \mathbf{1}_{\{|c_{k}Z_{1,t-k}| \leq a_{np}\}}, Z_{1,t'-k'}^{2} \mathbf{1}_{\{|c_{k}'Z_{1,t'-k'}| \leq a_{np}\}}\right)$$

is zero if $t - k \neq t' - k'$. In the case of equality, t - k = t' - k', we have that

$$\sum_{t=1}^{n} \sum_{t'=1}^{n} \operatorname{Cov} \left(Z_{1,t-k}^{2} \mathbf{1}_{\{|c_{k}Z_{1,t-k}| \leq a_{np}\}}, Z_{1,t'-k'}^{2} \mathbf{1}_{\{|c_{k}'Z_{1,t'-k'}| \leq a_{np}\}} \right)$$
$$= \sum_{t=1}^{n} \operatorname{Var} \left(Z_{1,t-k}^{2} \mathbf{1}_{\{|c_{k}Z_{1,t-k}| \leq a_{np}\}} \mathbf{1}_{\{|c_{k}'Z_{1,t'-k'}| \leq a_{np}\}} \right)$$
$$\leq n E \left(Z_{1,1-k}^{4} \mathbf{1}_{\{|\min\{c_{k},c_{k'}\}Z_{1,1-k}| \leq a_{np}\}} \right)$$

Using Karamata's theorem and Potter's bound we obtain that there exists a C > 0 and an $\epsilon > 0$ such that

$$\frac{pn}{a_{np}^4} E\left(Z_{1,1-k}^4 \mathbf{1}_{\{|\min\{c_k,c_{k'}\}Z_{1,1-k}|\leqslant a_{np}\}}\right) \leqslant C\min\{c_k,c_{k'}\}^{\alpha/4-\epsilon-1}.$$

For m sufficiently large the coefficients become smaller than one, thus

$$\min\{c_k, c_{k'}\}^{\alpha/4-\epsilon-1} \leq c_k^{\alpha/4-\epsilon-1} c_{k'}^{\alpha/4-\epsilon-1}.$$

All in all we obtain

$$\lim_{m\to\infty}\limsup_{n\to\infty}\operatorname{I}_2\leqslant\frac{3C}{\gamma}\lim_{m\to\infty}\left(\sum_{k=m+1}^{\infty}c_k^{1+\alpha/4-\epsilon}\right)^2=0,$$

since $\sum_{k=0}^{\infty} c_k < \infty$. For the second term observe that it follows, using Chebyshev's inequality, $EZ_{11} = 0$ and the independence of the Z's, that

$$\begin{split} \Pi &= pP\left(2\left|\sum_{k=m+1}^{\infty}\sum_{l=0}^{m}c_{k}c_{l}\sum_{t=1}^{n}Z_{1,t-k}Z_{1,t-l}\right| > a_{np}^{2}\frac{\gamma}{3}\right) \\ &\leq \frac{6}{\gamma}\frac{p}{a_{np}^{4}}\operatorname{Var}\left(\sum_{k=m+1}^{\infty}\sum_{l=0}^{m}c_{k}c_{l}\sum_{t=1}^{n}Z_{1,t-k}Z_{1,t-l}\right) \\ &= \frac{6}{\gamma}\frac{p}{a_{np}^{4}}\sum_{k,k'=m+1}^{\infty}\sum_{l,l'=0}^{m}c_{k}c_{k'}c_{l}c_{l'}\sum_{t,t'=1}^{n}E\left(Z_{1,t-k}Z_{1,t-l}Z_{1,t'-k'}Z_{1,t'-l'}\right) \\ &\leq \frac{6}{\gamma}\frac{p}{a_{np}^{4}}\sum_{k,k'=m+1}^{\infty}\sum_{l,l'=0}^{m}c_{k}c_{k'}c_{l}c_{l'}nE\left(Z_{11}^{2}\right)^{2} \\ &\leq O\left(\left(\sum_{k=m+1}^{\infty}c_{k}\right)^{2}\left(\sum_{l=0}^{m}c_{l}\right)^{2}\frac{pn}{a_{np}^{4}}\right) \xrightarrow{n\to\infty} 0, \end{split}$$

since $2 < \alpha < 4$. The remaining term III can be dealt with similarly to the previous term II. Hence the proof is complete.

Proof of Theorem 3.1. The proof is essentially the proof of Theorem 2.1 (i). In short, Proposition 3.2.1 and Weyl's inequality ([15, Theorem III.2.6]) imply that the distance of the eigenvalues of S and D_S converges to zero uniformly in probability. Of course, the eigenvalues of the diagonal matrix D_S are just its entries. By Proposition 3.2.2, the point process of the entries of D_S converges to N. Therefore N_n also converges to N.

3.2.3. Proof of Theorem 3.2

Proposition 3.2.3. Under the conditions of Theorem 3.2 we have, conditionally on (θ_i) as well as unconditionally, that

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\left(\sum_{t=1}^{n} X_{it}^{2} - n\mu_{X}\right)} \to \sum_{i=1}^{\infty} \epsilon_{\Gamma_{i}^{-2/\alpha}\left(E\left|\sum_{j=-\infty}^{\infty} c_{j}^{2}\left(\theta_{1}\right)\right|^{\alpha/2}\right)^{2/\alpha}}$$
(3.2.12)

Proof. As in Proposition 2.3.7 one can show, for any $m < \infty$, that

$$\frac{1}{p}\sum_{i=1}^{p}pP\left(\left|\sum_{t=1}^{n}\sum_{j=-m}^{m}c_{j}(\theta_{i})(Z_{i,t-j}^{2}-\mu_{Z})\right| > a_{np}^{2}x\left|\theta_{i}\right|\right) \xrightarrow{a.s.}{\xrightarrow{\alpha/2}} E|d_{1}^{m}|^{\alpha/2},$$

where $d_1^m = \sum_{j=-m}^m c_j^2(\theta_1)$. Hence, an adaptation of the proof of Lemma 3.2.1 yields, for the truncated process

$$X_{it,m} = \sum_{k=-m}^{m} c_k Z_{i,t-k}, \quad \mu_{X,m} = E X_{11,m}^2 = \sum_{j=-m}^{m} c_j^2 \mu_Z,$$

that, conditionally on the sequence (θ_i) ,

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}\left(\sum_{t=1}^{n} X_{it,m}^{2} - n\mu_{X,m}\right)} \to \sum_{i=1}^{\infty} \epsilon_{\Gamma_{i}^{-2/\alpha}\left(E\left|\sum_{j=-m}^{m} c_{j}^{2}\left(\theta_{1}\right)\right|^{\alpha/2}\right)^{2/\alpha}}.$$
(3.2.13)

It is only left to show that this result extends to the more general setting where $m = \infty$. By Proposition 3.2.2 it suffices to show that

$$\lim_{m\to\infty}\limsup_{n\to\infty}\sum_{i=1}^p P\left(\left|\sum_{t=1}^n (X_{it}^2 - X_{it,m}^2 - (\mu_X - \mu_{X,m}))\right| > a_{np}^2 \gamma |(\theta_r)\right) = 0.$$

To proof this claim, follow the string of arguments of Proposition 3.2.2 and make use of the fact that

$$\left|\sum_{i=1}^p c_j(\theta_i)\right| \leq p\tilde{c}_j.$$

Taking expectations shows that (3.2.12) also holds unconditionally.

Proof of Theorem 3.2. By conditioning on (θ_i) , the proofs of Proposition 3.2.1 carry over to this more general setting. The result then follows immediately from the previous proposition.

CHAPTER 4

On the spectral norm of heavy-tailed random matrices with strongly dependent rows and columns

4.1. Introduction and main results

In Chapter 2 we studied the asymptotic properties of the extreme singular values of a heavytailed random matrix X the rows of which are given by independent copies of some linear process. This was motivated by the statistical analysis of observations of a high-dimensional linear process with independent components. Typically, the linear processes used in multivariate stochastic modeling have the more general form

$$\mathbf{X}_t = \sum_j A^{(j)} \mathbf{Z}_{t-j}, \quad t = 1, \dots, n,$$

where $A^{(j)}$ is a sequence of deterministic $p \times p$ matrices and \mathbf{Z}_t is a noise vector containing p independent and identically distributed (iid) random variables Z_{1t}, \ldots, Z_{pt} . Of course, the process \mathbf{X} does not have independent components except when $A^{(j)}$ is a multiple of the identity matrix. Let us denote by \tilde{X} the matrix with columns $\mathbf{X}_1, \ldots, \mathbf{X}_n$. Then the *it*-th entry of \tilde{X} is given by

$$\tilde{X}_{it} = \sum_{j} \sum_{k=1}^{p} A_{ik}^{(j)} Z_{k,t-j}.$$

This motivates to study the general random matrix model

$$ilde{X}_{it} = \sum_{j} \sum_{k} d(i, j, k) Z_{i-k,t-j}$$

4. Random matrices with strongly dependent rows and columns

with some iid array (Z_{it}) and some function $d : \mathbb{N} \times \mathbb{Z}^2 \to \mathbb{R}$, $(i, j, k) \mapsto d(i, j, k)$ such that the above double sum converges. This can be seen as a two dimensional filter applied to some noise matrix and has been studied for $d(i, j, k) = \tilde{d}(j, k)$ and Gaussian matrices by [60], and for more general distributions by [3] under the assumption that $\tilde{d}(j, k) = 0$ if j or k is larger than some fixed constant.

It will be our objective to investigate the random matrix model $\hat{X} = (\hat{X}_{it}) \in \mathbb{R}^{p \times n}$ with

$$\hat{X}_{it} = \sum_{j} \sum_{k} c_j \theta_k Z_{i-k,t-j}, \qquad (4.1.1)$$

for two real sequences (c_j) and (θ_k) , i.e., the case where the function *d* can be factorized in the form $d(i, j, k) = c_j \theta_k$. In contrast to the model $X = (X_{it})$ considered in Chapter 2, with

$$X_{it} = \sum_{j} c_j Z_{i,t-j},$$

the matrix \hat{X} has not only dependent columns but also dependent rows. Indeed, writing the model (4.1.1) in the form

$$\hat{X}_{it} = \sum_{j} c_{j} \xi_{i,t-j},$$
(4.1.2)

$$\xi_{it} = \sum_{k} \theta_k Z_{i-k,t}, \qquad (4.1.3)$$

one can see that, by going from X to \hat{X} , the noise sequence Z in the processes along the rows is replaced by a linear process ξ along the columns.

Since we want to investigate heavy-tailed random matrix models we assume that $(Z_{it})_{i,t}$ is an array of iid random variables with tail index $\alpha \in (0, 4)$ satisfying

$$nP(|Z_{11}| > a_n x) \to x^{-\alpha}.$$
 (4.1.4)

Furthermore, let (c_i) and (θ_k) be sequences of real numbers such that

$$\sum_{j} |c_{j}|^{\delta} < \infty, \text{ and}$$

$$(4.1.5)$$

$$\sum_{k} |\theta_k|^{\delta} < \infty \quad \text{for some } \delta < \min\{\alpha, 1\}.$$
(4.1.6)

For some values of α we also require that Z_{11} satisfies the tail balancing condition, i.e., the existence of the limits

$$\lim_{x \to \infty} \frac{P(Z_{11} > x)}{P(|Z_{11}| > x)} = q \quad \text{and} \quad \lim_{x \to \infty} \frac{P(Z_{11} \le -x)}{P(|Z_{11}| > x)} = 1 - q \tag{4.1.7}$$

for some $0 \le q \le 1$. By the above definitions, \hat{X} is a $p \times n$ random matrix with heavy-tailed entries. Based on our statistical motivation we will refer to p as the dimension and to n as the sample size of our observations. In order to get results which are useful for the analysis of high-dimensional data we assume that both $p = p_n$ and n go to infinity such that

$$\limsup_{n \to \infty} \frac{p_n}{n^\beta} < \infty \tag{4.1.8}$$

for some $\beta > 0$ satisfying

$$\beta < \infty \quad \text{if} \quad \alpha \in (0, 1],$$

$$\beta < \max\left\{\frac{2-\alpha}{\alpha-1}, \frac{1}{2}\right\} \quad \text{if} \quad \alpha \in (1, 2),$$

$$\beta < \max\left\{\frac{1}{3}, \frac{4-\alpha}{4(\alpha-1)}\right\} \quad \text{if} \quad 2 \le \alpha < 3, \quad \text{or}$$

$$\beta < \frac{4-\alpha}{3\alpha-4} \quad \text{if} \quad 3 \le \alpha < 4.$$

For \hat{X} given by (4.1.1), our main theorem investigates the asymptotic behaviour of the largest eigenvalue (spectral norm) $\lambda_{\text{max}} = \|\hat{X}\hat{X}^{\mathsf{T}}\|_2$ of $\hat{X}\hat{X}^{\mathsf{T}}$ when *p* and *n* jointly go to infinity.

Theorem 4.1. Consider the random matrix model given by equations (4.1.2) to (4.1.6) with $\alpha \in (0, 4)$. If $\alpha \in (5/3, 4)$ then assume that Z_{11} has zero mean and satisfies the tail balancing condition (4.1.7).

For $\alpha < 2$ let λ_{\max} be the largest eigenvalue of $\hat{X}\hat{X}^{\mathsf{T}}$. In case $\alpha > 2$, denote by λ_{\max} the largest eigenvalue of $(\hat{X}\hat{X}^{\mathsf{T}} - n\mu_{\hat{X}}HH^{\mathsf{T}})$, where $\mu_{\hat{X}} = EZ_{11}^2 \sum_j c_j^2$ and $H = (H_{ij}) \in \mathbb{R}^{p \times 3p}$ is given by $H_{ij} = \theta_{p-(j-i)} \mathbf{1}_{\{0 \le j-i \le 2p\}}$. If $\alpha = 2$ and $EZ_{11}^2 = \infty$, then replace $\mu_{\hat{X}}$ by $\mu_{\hat{X}}^n = \sum_j c_j^2 E(Z_{11}^2 \mathbf{1}_{\{Z_{11}^2 \le a_{np}^2\}})$. Let Γ_1 be an exponentially distributed random variable with mean one and x > 0. If p and n go to infinity such that condition (4.1.8) is satisfied then we have for the largest eigenvalue λ_{\max} that

$$P\left(\Gamma_{1}^{-2/\alpha}\max_{k}\theta_{k}^{2}\sum_{j}c_{j}^{2}>x\right) \leq \liminf_{n\to\infty}P\left(\lambda_{\max}>a_{np}^{2}x\right)$$
$$\leq \limsup_{n\to\infty}P\left(\lambda_{\max}>a_{np}^{2}x\right)$$
$$\leq P\left(\Gamma_{1}^{-2/\alpha}\max_{l}|\theta_{l}|\sum_{k}|\theta_{k}|\sum_{j}c_{j}^{2}>x\right)$$
(4.1.9)

Remark 4.1.1. (i) If all θ_k 's except one are zero, one has equality and therefore recovers the result from Theorem 2.1 (i). If two or more θ_k are non-zero, then

$$P\left(\Gamma_1^{-2/\alpha}\max_k\theta_k^2\sum_j c_j^2 > x\right) < P\left(\Gamma_1^{-2/\alpha}\max_l|\theta_l|\sum_k|\theta_k|\sum_j c_j^2 > x\right).$$

4. Random matrices with strongly dependent rows and columns

Whether the lim inf and lim sup are equal in this case and attain one of its boundaries remain open problems.

(ii) Since $P(\Gamma_1^{-2/\alpha} \le x) = e^{-x^{-\alpha/2}}$, inequality (4.1.9) can equivalently be written as

$$\exp\left(-x^{-\alpha/2}\max_{l}|\theta_{l}|^{\alpha/2}\left(\sum_{k}|\theta_{k}|\sum_{j}c_{j}^{2}\right)^{\alpha/2}\right) \leq \liminf_{n\to\infty}P\left(\lambda_{\max}\leq a_{np}^{2}x\right)$$
$$\leq \limsup_{n\to\infty}P\left(\lambda_{\max}\leq a_{np}^{2}x\right)$$
$$\leq \exp\left(-x^{-\alpha/2}\max_{k}|\theta_{k}|^{\alpha}\left(\sum_{j}c_{j}^{2}\right)^{\alpha/2}\right).$$

(iii) Clearly, for $0 < \alpha < 2$, the theorem can easily be rephrased for the largest singular value of *X*, that is, the spectral norm of *X*. In this case one has that

$$\begin{split} P \Biggl(\Gamma_1^{-1/\alpha} \max_k |\theta_k| \sqrt{\sum_j c_j^2} > x \Biggr) &\leq \liminf_{n \to \infty} P\left(||X||_2 > a_{np} x \right) \\ &\leq \limsup_{n \to \infty} P\left(||X||_2 > a_{np} x \right) \\ &\leq P \Biggl(\Gamma_1^{-1/\alpha} \max_l \sqrt{|\theta_l|} \sqrt{\sum_k |\theta_k|} \sqrt{\sum_j c_j^2} > x \Biggr). \end{split}$$

4.2. Dependence of successive rows

To understand the basic principle of our method it is beneficial to first investigate the case where only successive rows of \hat{X} are dependent and where $\alpha \in (0, 2)$. Thus we start with the model

$$\hat{X}_{it} = \sum_{j} c_{j} \xi_{i,t-j}, \tag{4.2.1}$$

$$\xi_{it} = Z_{it} + \theta Z_{i-1,t}. \tag{4.2.2}$$

It is easy to see that $\hat{X}_{it} = X_{it} + \theta X_{i-1,t}$, where $X_{it} = \sum_j c_j Z_{i,t-j}$ for i = 0, 1, ..., p, and t = 1, ..., n. To proceed further we define the matrices $\hat{X} = (\hat{X}_{it}) \in \mathbb{R}^{p \times n}$, $X = (X_{(i-1),t}) \in \mathbb{R}^{(p+1) \times n}$ and $H = (H_{ij}) \in \mathbb{R}^{p \times (p+1)}$, where all entries of H are zero except $H_{ii} = \theta$ and $H_{i,i+1} = 1$. Then we clearly have the matrix equality

$$\hat{X} = HX. \tag{4.2.3}$$

Moreover, we denote by $D = (D_i) = \text{diag}(XX^{\mathsf{T}}) \in \mathbb{R}^{(p+1)\times(p+1)}$ the diagonal of XX^{T} , that is the diagonal matrix which consists of the diagonal entries of XX^{T} . For the convenience of the reader, we restate the result from Proposition 2.3.4.

Proposition 4.2.1. Under the conditions of Theorem 4.1 we have that

$$a_{np}^{-2} \left\| XX^{\mathsf{T}} - D \right\|_2 \xrightarrow[n \to \infty]{P} 0.$$

Thus, since $||H||_2 \leq ||H||_{\infty} \leq 1 + |\theta|$, we immediately conclude, by (4.2.3), that

$$a_{np}^{-2} \left\| \hat{X} \hat{X}^{\mathsf{T}} - H D H^{\mathsf{T}} \right\|_{2} \le \|H\|_{2}^{2} a_{np}^{-2} \left\| X X^{\mathsf{T}} - D \right\|_{2} \to 0.$$
(4.2.4)

Hence, by Weyl's inequality ([15, Theorem III.2.6]), the largest eigenvalue λ_{max} of the sample covariance matrix $\hat{X}\hat{X}^{\mathsf{T}}$ based on the observations \hat{X} is asymptotically equal to the largest eigenvalue of the tridiagonal matrix

$$HDH^{\mathsf{T}} = \begin{pmatrix} D_1 + \theta^2 D_2 & \theta D_2 & 0 \\ \theta D_2 & D_2 + \theta^2 D_3 & \theta D_3 \\ 0 & \ddots & \ddots & 0 \\ & & D_{p-1} + \theta^2 D_p & \theta D_p \\ 0 & & \theta D_p & D_p + \theta^2 D_{p+1} \end{pmatrix} \in \mathbb{R}^{p \times p}. \quad (4.2.5)$$

It is our goal to find an asymptotic upper and lower bound for λ_{max} . First we prove a lower bound. Clearly, λ_{max} is asymptotically larger or equal than the largest diagonal entry of HDH^{T} , i.e.,

$$\lambda_{\max} \ge \max_{1 \le i \le p} \left(D_i + \theta^2 D_{i+1} \right) + o_P(1), \tag{4.2.6}$$

where $o_P(1)$ denotes some generic random variable that converges to zero in probability as *n* goes to infinity. Since $D_{i+1} = \sum_{t=1}^{n} X_{it}^2$, we have to find the maximum of an MA(1) process of partial sums of linear processes. By Proposition 2.3.6 we already know that

$$\sum_{i=0}^{p} \epsilon_{a_{np}^{-2}D_{i+1}} = \sum_{i=0}^{p} \epsilon_{a_{np}^{-2}\sum_{t=1}^{n} X_{it}^{2}} \xrightarrow{D}_{n \to \infty} \sum_{i=1}^{\infty} \epsilon_{\Gamma_{i}^{-2/\alpha} \sum_{j} c_{j}^{2}}.$$
(4.2.7)

Since (D_i) is an iid sequence, this result can be generalized as follows.

Lemma 4.2.1. Under the conditions of Theorem 4.1 we have that

$$I_p = \sum_{i=1}^p \epsilon_{a_{np}^{-2}(D_{i+1},D_i)} \xrightarrow[n \to \infty]{D} I = \sum_{i=1}^\infty \Big(\epsilon_{\Gamma_i^{-2/\alpha} \sum_j c_j^2(1,0)} + \epsilon_{\Gamma_i^{-2/\alpha} \sum_j c_j^2(0,1)} \Big).$$

4. Random matrices with strongly dependent rows and columns

Proof. By the continuous mapping theorem applied to (4.2.7), we immediately conclude that

$$I_p^* = \sum_{i=1}^p \left(\epsilon_{a_{np}^{-2}(D_{i+1},0)} + \epsilon_{a_{np}^{-2}(0,D_i)} \right) \xrightarrow[n \to \infty]{D} I.$$

Thus, we only have to show that $|I_p(f) - I_p^*(f)| \to 0$ in probability for any continuous function with $\operatorname{supp}(f) \subset \{x = (x_1, x_2) \in \mathbb{R}^2 : \max\{|x_1|, |x_2|\} \ge \delta\}$. To this end, let $L = \{x : \min\{|x_1|, |x_2|\} < \delta\}$ and observe that, by independence of (D_i) ,

$$EI_p(L^c) \leq pP(|D_{i+1}| \geq a_{np}^2\delta, |D_i| \geq a_{np}^2\delta) = O(\delta^{-\alpha}p^{-1}) \to 0.$$

Thus $I_p(f) = \int_L f dI_p + o_P(1)$ and, by definition of I_p^* , $I_p^*(f) = \int_L f dI_p^*$. Since f(z) = 0 if $\max\{|x_1|, |x_2|\} < \delta$, it suffices to show that

$$A + B = \sum_{i=1}^{p} \left| f(a_{np}^{-2}(D_{i+1}, D_i)) \mathbf{1}_{\{a_{np}^{-2}|D_{i+1}| \ge \delta\} \cap \{a_{np}^{-2}|D_i| < \delta\}} - f(a_{np}^{-2}(D_{i+1}, 0)) \mathbf{1}_{\{a_{np}^{-2}|D_{i+1}| \ge \delta\}} \right|$$

+
$$\sum_{i=1}^{p} \left| f(a_{np}^{-2}(D_{i+1}, D_i)) \mathbf{1}_{\{a_{np}^{-2}|D_{i+1}| < \delta\} \cap \{a_{np}^{-2}|D_i| \ge \delta\}} - f(a_{np}^{-2}(0, D_i)) \mathbf{1}_{\{a_{np}^{-2}|D_i| \ge \delta\}} \right| \xrightarrow{P}_{n \to \infty} 0.$$

We only treat term A, as B can be handled essentially the same way. To this end, observe that

$$A \leq \sum_{i=1}^{p} \left| f(a_{np}^{-2}(D_{i+1}, D_{i})) - f(a_{np}^{-2}(D_{i+1}, 0)) \right| \mathbf{1}_{\{a_{np}^{-2}|D_{i+1}| \ge \delta\} \cap \{a_{np}^{-2}|D_{i}| < \delta\}} + \sum_{i=1}^{p} \left| f(a_{np}^{-2}(D_{i+1}, 0)) \right| \mathbf{1}_{\{a_{np}^{-2}|D_{i+1}| \ge \delta\} \cap \{a_{np}^{-2}|D_{i}| \ge \delta\}} = I + II.$$

Clearly, by independence,

$$E(II) \leq \sup f(x)pP(a_{np}^{-2}|D_{i+1}| \geq \delta)P(a_{np}^{-2}|D_i| \geq \delta) = O(p^{-1}) \to 0.$$

Furthermore, we have, for any $0 < \eta < \delta$, that

$$\mathbf{1}_{\{a_{np}^{-2}|D_{i+1}| \ge \delta\} \cap \{a_{np}^{-2}|D_{i}| < \delta\}} \leq \mathbf{1}_{\{a_{np}^{-2}|D_{i+1}| \ge \delta\} \cap \{a_{np}^{-2}|D_{i}| < \eta\}} + \mathbf{1}_{\{a_{np}^{-2}|D_{i+1}| \ge \eta\} \cap \{a_{np}^{-2}|D_{i}| \ge \eta\}}.$$

Thus, for some c > 0,

$$E(I) \leq \sup\{|f(x_1, x_2) - f(x_1, 0)| : |x_1| > \delta, |x_2| < \eta\} p P(|D_{i+1}| \ge a_{np}^2 \eta) + c p P(|D_{i+1}| \ge a_{np}^2 \eta) P(|D_i| \ge a_{np}^2 \eta).$$

Obviously, the second summand converges, for fixed $\eta > 0$, to zero as $n \to \infty$. The first summand can be made arbitrarily small by choosing η small enough, since f is uniformly continuous.

The continuous mapping theorem applied to Lemma 4.2.1 gives

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2}(\theta^{2}D_{(i+1)}+D_{i})} \xrightarrow[n \to \infty]{} \sum_{i=1}^{\infty} \left(\epsilon_{\Gamma_{i}^{-2/\alpha} \sum_{j} c_{j}^{2} \theta^{2}} + \epsilon_{\Gamma_{i}^{-2/\alpha} \sum_{j} c_{j}^{2}} \right).$$

Therefore, by (4.2.6), the asymptotic lower bound of λ_{max} is given by

$$a_{np}^{-2} \max_{1 \le i \le p} \left(D_i + \theta^2 D_{i+1} \right) \xrightarrow[n \to \infty]{D} \max\{1, \theta^2\} \Gamma_1^{-2/\alpha} \sum_j c_j^2.$$

$$(4.2.8)$$

Regarding the upper bound, we make use of the fact that $\|HDH^{\mathsf{T}}\|_{2} \leq \|HDH^{\mathsf{T}}\|_{\infty}$. Observe that

$$\begin{split} \left\| HDH^{\mathsf{T}} \right\|_{\infty} &= \max_{1 \leqslant i \leqslant p} \left(\mathbf{1}_{\{i \neq 1\}} |\theta| D_i + D_i + \theta^2 D_{i+1} + |\theta| D_{i+1} \mathbf{1}_{\{i \neq p\}} \right) \\ &= \max_{1 \leqslant i \leqslant p} \left((1 + |\theta| \mathbf{1}_{\{i \neq 1\}}) D_i + (|\theta| \mathbf{1}_{\{i \neq p\}} + \theta^2) D_{i+1} \right). \end{split}$$

So once again we have to determine the maximum of an MA(1) of partial sums of linear processes. An application of Lemma 4.2.1 yields that

$$a_{np}^{-2} \left\| HDH^{\mathsf{T}} \right\|_{\infty} \xrightarrow[n \to \infty]{} \max\{1 + |\theta|, |\theta| + \theta^2\} \Gamma_1^{-2/\alpha} \sum_j c_j^2.$$

$$(4.2.9)$$

The lower and upper bound (4.2.8) and (4.2.9) together with equation (4.2.4) finally yield that

$$P\left(\max\{1,\theta^2\}\Gamma_1^{-2/\alpha}\sum_j c_j^2 > x\right) \leq \liminf_{n \to \infty} P\left(\lambda_{\max} > a_{np}^2 x\right)$$
$$\leq \limsup_{n \to \infty} P\left(\lambda_{\max} > a_{np}^2 x\right)$$
$$\leq P\left(\left(|\theta| + \max\{1,\theta^2\}\right)\Gamma_1^{-2/\alpha}\sum_j c_j^2 > x\right).$$

Clearly, this result is a special case of Theorem 4.1 when the process ξ_{it} is a moving average process of order one.

4.3. Proof of the theorem

In this section we will proof Theorem 4.1 in its full generality. We start with the case where $\alpha < 2$. To this end we define an approximation $\hat{X}^{(p)}$ of X and so that

(i)
$$a_{np}^{-2} \| \hat{X}^{(p)} (\hat{X}^{(p)})^{\mathsf{T}} - HDH^{\mathsf{T}} \|_{2} \xrightarrow{P}{n \to \infty} 0,$$
 (4.3.1)

(ii)
$$a_{np}^{-2} \left\| \hat{X} \hat{X}^{\mathsf{T}} - \hat{X}^{(p)} (\hat{X}^{(p)})^{\mathsf{T}} \right\|_{2} \xrightarrow{P}_{n \to \infty} 0,$$
 (4.3.2)

(iii) and finally we derive upper and lower bounds for $||HDH^{\mathsf{T}}||_2$.

Note that, for notational convenience, we will assume that $\theta_k = 0$ for k < 0, since the extension of the proof to the case where the dependence in (4.1.3) is two-sided is analogous.

(i). First we define the approximation $\hat{X}^{(p)} = (\hat{X}_{it}^{(p)}) \in \mathbb{R}^{p \times n}$ by $\hat{X}_{it}^{(p)} = \sum_{k=0}^{p} \theta_k X_{i-k,t}$, where $X_{it} = \sum_j c_j Z_{i,t-j}$. Furthermore we define $X = (X_{i-p,t}) \in \mathbb{R}^{2p \times n}$, and $H = (H_{ij}) \in \mathbb{R}^{p \times 2p}$ by

$$H_{ij} = \begin{cases} \theta_{p-(j-i)} & \text{if } 0 \le j-i \le p, \\ 0 & \text{else.} \end{cases}$$
(4.3.3)

Then we have that $HX = \hat{X}^{(p)}$. Indeed,

$$(HX)_{it} = \sum_{l=0}^{2p} H_{il} X_{l-p,t} = \sum_{l=i}^{i+p} H_{il} X_{l-p,t} = \sum_{l=0}^{p} H_{i,i+l} X_{i+l-p,t} = \sum_{l=0}^{p} \theta_{p-l} X_{i-(p-l),t}$$
$$= \sum_{k=0}^{p} \theta_k X_{i-k,t} = \hat{X}_{it}^{(p)}.$$

Thus, if we let $D = (D_i) = \text{diag}(XX^{\mathsf{T}}) \in \mathbb{R}^{2p \times 2p}$, then we obtain (4.3.1) by virtue of Proposition 4.2.1 and $||H||_2 \leq ||H||_{\infty} \leq \sum_{k=0}^{\infty} |\theta_k| < \infty$.

(ii). In order to proceed we will require the following lemma.

Lemma 4.3.1. Under the conditions of Theorem 4.1 we have that

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2} \sum_{k=0}^{\infty} \theta_k \sum_{t=1}^{n} X_{i-k,t}^2} \xrightarrow{D} \sum_{i=1}^{\infty} \sum_{k=0}^{\infty} \epsilon_{\Gamma_i^{-2/\alpha} \theta_k \sum_j c_j^2}.$$

Proof. A straight-forward generalization of Lemma 4.2.1 yields, for any $m < \infty$, that

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2} \sum_{t=1}^{n} (X_{it}^{2}, X_{i-1,t}^{2}, \dots, X_{i-m,t}^{2})} \xrightarrow{D}_{n \to \infty} \sum_{k=0}^{m} \sum_{i=1}^{\infty} \epsilon_{\Gamma_{i}^{-2/\alpha} \sum_{j} c_{j}^{2} e_{k+1}},$$
(4.3.4)

where e_k denotes the k-th unit vector of \mathbb{R}^{∞} , i.e, the k-th component of e_k is one and all others are zero. By an application of the continuous mapping theorem we obtain the claim for a finite order moving average of the partial sums $(\sum_{t=1}^{n} X_{it}^2)_i$, i.e.,

$$\sum_{i=1}^{p} \epsilon_{a_{np}^{-2} \sum_{k=0}^{m} \theta_k \sum_{t=1}^{n} X_{i-k,t}^2} \xrightarrow{D}_{n \to \infty} \sum_{i=1}^{\infty} \sum_{k=0}^{m} \epsilon_{\Gamma_i^{-2/\alpha} \theta_k \sum_j c_j^2},$$

On the other hand we have, for $m \to \infty$, that

$$\sum_{i=1}^{\infty}\sum_{k=0}^{m}\epsilon_{\Gamma_{i}^{-2/\alpha}\theta_{k}\sum_{j}c_{j}^{2}} \xrightarrow{D}_{m\to\infty}\sum_{i=1}^{\infty}\sum_{k=0}^{\infty}\epsilon_{\Gamma_{i}^{-2/\alpha}\theta_{k}\sum_{j}c_{j}^{2}}.$$

To finish the proof of the lemma it is therefore only left so show that

$$\lim_{m\to\infty}\limsup_{n\to\infty}\rho\left(\sum_{i=1}^{p}\epsilon_{a_{np}^{-2}\sum_{k=0}^{m}\theta_{k}\sum_{t=1}^{n}X_{i-k,t}^{2}},\sum_{i=1}^{p}\epsilon_{a_{np}^{-2}\sum_{k=0}^{\infty}\theta_{k}\sum_{t=1}^{n}X_{i-k,t}^{2}}\right)=0,$$

where ρ denotes a metric of the vague topology on the space of point processes. To this end, observe that

$$\left|\sum_{k=0}^{m} \theta_{k} \sum_{t=1}^{n} X_{i-k,t}^{2} - \sum_{k=0}^{\infty} \theta_{k} \sum_{t=1}^{n} X_{i-k,t}^{2}\right| \leq \sum_{k>m} |\theta_{k}| \sum_{t=1}^{n} X_{i-k,t}^{2}.$$

Therefore, by the arguments of the proof of Proposition 2.3.6, we only have to show, for any $\gamma > 0$, that

$$\lim_{m\to\infty}\limsup_{n\to\infty}P((A_n^{\gamma})^c)=0,$$

where

$$A_n^{\gamma} = \left\{ \max_{1 \le i \le p} \sum_{l > m} |\theta_l| \sum_{t=1}^n X_{i-l,t}^2 \le a_{np}^2 \gamma \right\}.$$

Observe that

$$P((A_{n}^{\gamma})^{c}) \leq pP\left(\sum_{l>m} |\theta_{l}| \sum_{t=1}^{n} X_{lt}^{2} > a_{np}^{2}\gamma\right) \leq pP\left(\sum_{l>m} |\theta_{l}| \sum_{j} c_{j}^{2} \sum_{t=1}^{n} Z_{l,t-j}^{2} > a_{np}^{2} \frac{\gamma}{2}\right) + pP\left(\sum_{l>m} |\theta_{l}| \sum_{j} \sum_{k>j} |c_{j}c_{k}| \sum_{t=1}^{n} |Z_{l,t-j}Z_{l,t-k}| > a_{np}^{2}\gamma\right) = I + II.$$
(4.3.5)

We have

$$\lim_{m \to \infty} \limsup_{n \to \infty} I = \lim_{m \to \infty} \left(\sum_{l > m} |\theta_l| \right)^{\alpha/2} \left(2 \sum_j c_j^2 \right)^{\alpha/2} \gamma^{-\alpha/2} = 0$$

by a slight modification of the proof of Proposition 2.3.5. In fact, one can also map the array (Z_{it}) to a sequence and then apply Proposition 2.3.5 directly. Regarding the second term, note that

$$\begin{split} II \leqslant pP \left(\sum_{l>m} |\theta_l| \sum_j \sum_{k>j} |c_j c_k| \sum_{t=1}^n Z_{l,t-j}^2 > a_{np}^2 \gamma \right) \\ + pP \left(\sum_{l>m} |\theta_l| \sum_j \sum_{k>j} |c_j c_k| \sum_{t=1}^n Z_{l,t-k}^2 > a_{np}^2 \gamma \right) = \mathrm{II}_1 + \mathrm{II}_2. \end{split}$$

4. Random matrices with strongly dependent rows and columns

As before we conclude that

$$\lim_{m \to \infty} \limsup_{n \to \infty} \operatorname{II}_1 = \lim_{m \to \infty} \left(\sum_{l > m} |\theta_l| \right)^{\alpha/2} \left(\sum_j \sum_{k > j} |c_j c_k| \right)^{\alpha/2} \gamma^{-\alpha/2} = 0,$$

and clearly term II₂ can be handled similarly.

We will now prove equation (4.3.2). By definition of the matrices \hat{X} and $\hat{X}^{(p)}$ we have that

$$(\hat{X}\hat{X}^{\mathsf{T}} - \hat{X}^{(p)}(\hat{X}^{(p)})^{\mathsf{T}})_{ij} = \sum_{l,l'k,k' \in \mathbb{Z}^2 \times (\mathbb{N}_0 \setminus \{0,1,\dots,p\})^2} c_l c_{l'} \theta_k \theta_{k'} \sum_{t=1}^n Z_{i-k,t-l} Z_{j-k',t-l'} + C_{i-k,t-l} Z_{j-k',t-l'}$$

Therefore we have the bound

$$\begin{split} \left\| \hat{X} \hat{X}^{\mathsf{T}} - \hat{X}^{(p)} (\hat{X}^{(p)})^{\mathsf{T}} \right\|_{2} &\leq \left\| \hat{X} \hat{X}^{\mathsf{T}} - \hat{X}^{(p)} (\hat{X}^{(p)})^{\mathsf{T}} \right\|_{\infty} \\ &= \max_{1 \leq i \leq p} \sum_{j=1}^{p} \sum_{l,l',k,k' \in \mathbb{Z}^{2} \times (\mathbb{N}_{0} \setminus \{0,1,\dots,p\})^{2}} |c_{l} c_{l'} \theta_{k} \theta_{k'}| \sum_{t=1}^{n} |Z_{i-k,t-l} Z_{j-k',t-l'}|. \end{split}$$

Observe that the product $|Z_{i-k,t-l}Z_{j-k',t-l'}|$ has tail index $\alpha/2$ if and only if j - k' = i - k and l = l'. In this case we can treat this term like the first term in I in (4.3.5) and obtain

$$a_{np}^{-2} \max_{1 \leq i \leq p} \sum_{l,k,k' \in \mathbb{Z} \times \{p+1,p+2,\ldots\}^2} |c_l^2 \theta_k \theta_{k'}| \sum_{t=1}^n |Z_{i-k,t-l}^2| \xrightarrow{P}_{n \to \infty} 0,$$

since $\sum_{k>p} |\theta_k| \to 0$. If the product $|Z_{i-k,t-l}Z_{j-k',t-l'}|$ does not have tail index $\alpha/2$, i.e., $j-k' \neq i-k'$ or $l \neq l'$, the product has only tail index α and can then be treated similarly as the second term II in (4.3.5).

(iii). By a combination of (i) and (ii) we have that

$$a_{np}^{-2} \left\| \hat{X} \hat{X}^{\mathsf{T}} - H D H^{\mathsf{T}} \right\|_{2} \xrightarrow[n \to \infty]{P} 0.$$

Thus, by Weyl's inequality, the difference of the largest eigenvalue of $\hat{X}\hat{X}^{\mathsf{T}}$ and HDH^{T} converges to zero. As in the previous section, the final step is to find lower and upper bounds on $||HDH^{\mathsf{T}}||_2$. By definition of *H*, we have

$$(HDH^{\mathsf{T}})_{ij} = \sum_{l=\max\{i,j\}}^{\min\{i,j\}+p} \theta_{p-(l-i)} \theta_{p-(l-j)} D_l.$$

58
Hence HDH^{T} is no longer a tridiagonal matrix. Recall that the entries of the diagonal matrix D are given by $D_i = \sum_{t=1}^n X_{i-p,t}^2$. By virtue of Lemma 4.3.1 an asymptotic lower bound is given by

$$\begin{aligned} a_{np}^{-2} \left\| HDH^{\mathsf{T}} \right\|_{2} \geqslant a_{np}^{-2} \max_{1 \le i \le p} (HDH^{\mathsf{T}})_{ii} \\ = a_{np}^{-2} \max_{1 \le i \le p} (\theta_{p}^{2}D_{i} + \ldots + \theta_{0}^{2}D_{i+p}^{2}) \xrightarrow{D}_{n \to \infty} \Gamma_{1}^{-2/\alpha} \max_{k} \theta_{k}^{2} \sum c_{j}^{2}. \end{aligned}$$

Regarding the upper bound, observe that

$$\begin{split} \left\| HDH^{\mathsf{T}} \right\|_{2} &\leq \left\| HDH^{\mathsf{T}} \right\|_{\infty} = \max_{1 \leq i \leq p} \sum_{j=1}^{p} |(HDH^{\mathsf{T}})_{ij}| \\ &\leq \max_{1 \leq i \leq p} \sum_{j=1}^{p} \sum_{l=1}^{l=\min\{i,j\}+p} |\theta_{p-(l-i)}\theta_{p-(l-j)}| D_{l} \\ &= \max_{1 \leq i \leq p} \sum_{l=1}^{2p} D_{l} \sum_{j=1}^{p} \mathbf{1}_{\{l-p \leq j \leq l, i \leq l \leq i+p\}} |\theta_{p-(l-i)}\theta_{p-(l-j)}| \\ &= \max_{1 \leq i \leq p} \sum_{l=i}^{i+p} D_{l} |\theta_{p-(l-i)}| \sum_{j=l-p}^{l} |\theta_{p-(l-j)}| \\ &= \max_{1 \leq i \leq p} \sum_{l=0}^{p} D_{i+l} |\theta_{p-l}| \sum_{k=0}^{p} |\theta_{k}|, \end{split}$$

so we have to determine the maximum of a moving average of order p of (D_i) , with coefficients $|\theta_{p-l}| \sum_{k=0}^{p} |\theta_k|$. By Lemma 4.3.1,

$$a_{np}^{-2} \max_{1 \le i \le p} \sum_{l=0}^{p} D_{i+l} |\theta_{p-l}| \sum_{k=0}^{p} |\theta_k| \xrightarrow{D}_{n \to \infty} \Gamma_1^{-2/\alpha} \max_{0 \le l \le \infty} |\theta_l| \sum_{k=0}^{\infty} |\theta_k| \sum_j c_j^2.$$
(4.3.6)

This completes the proof of Theorem 4.1.

Proof of Theorem 4.1 for $\alpha \ge 2$ and $EZ_{11}^2 < \infty$. Since we now consider the largest eigenvalue λ_{\max} of $\hat{X}\hat{X}^{\mathsf{T}} - n\mu_X H H^{\mathsf{T}}$, one has to replace D by the centralized diagonal matrix $\tilde{D} = D - n\mu_X I_p$, i.e,

$$\tilde{D}_i = \sum_{t=1}^n (X_{i-p,t}^2 - \mu_X).$$

Then one has that

$$\begin{aligned} a_{np}^{-2} \left\| (\hat{X}\hat{X}^{\mathsf{T}} - n\mu_{X}HH^{\mathsf{T}}) - H\tilde{D}H^{\mathsf{T}} \right\|_{2} = & a_{np}^{-2} \left\| H(XX^{\mathsf{T}} - n\mu_{X}I_{p})H^{\mathsf{T}} - H(D - n\mu_{X}I_{p})H^{\mathsf{T}} \right\|_{2} \\ \leqslant & a_{np}^{-2} \left\| H \right\|_{2}^{2} \left\| (XX^{\mathsf{T}} - n\mu_{X}I_{p}) - (D - n\mu_{X}I_{p}) \right\|_{2} \xrightarrow{P}_{n \to \infty} 0, \end{aligned}$$

by an application of Proposition 3.2.1. The remainder of the proof is then a straightforward combination of the results from the foregoing chapter with the methods used in the proof of Theorem 4.1. The case where $\alpha = 2$ and $EZ_{11}^2 = \infty$ is treated analogously with $\mu_{\hat{X}}$ replaced by $\mu_{\hat{X}}^n$.

Part II.

Spectral Distribution of Light-tailed Random Matrices

CHAPTER 5

Eigenvalue distribution of large sample covariance matrices of linear processes²

5.1. Introduction and main result

A typical object of interest in many fields is the sample covariance matrix $(n-1)^{-1}\mathbf{X}\mathbf{X}^T$ of a data matrix $\mathbf{X} = (X_{i,t})_{it}$, i = 1, ..., p, t = 1, ..., n. The matrix \mathbf{X} can be seen as a sample of size *n* of *p*-dimensional data vectors. For fixed *p* one can show, as *n* tends to infinity, that under certain assumptions the eigenvalues of the sample covariance matrix converge to the eigenvalues of the true underlying covariance matrix [5]. However, the assumption $p \ll n$ may not be justified if one has to deal with high dimensional data sets, so that it is often more suitable to assume that the dimension *p* is of the same order as the sample size *n*, that is $p = p_n \rightarrow \infty$ such that

$$\lim_{n \to \infty} \frac{n}{p} \eqqcolon y \in (0, \infty).$$
(5.1.1)

For a symmetric matrix A with eigenvalues $\lambda_1, \ldots, \lambda_p$, we denote by

$$F^A = \frac{1}{p} \sum_{i=1}^p \delta_{\lambda_i}$$

the spectral distribution of A, where δ_x denotes the Dirac measure located at x. This means that $pF^A(B)$ is equal to the number of eigenvalues of A that lie in the set B. From now on we will

²This chapter is based on *O. Pfaffel and E. Schlemm:* Eigenvalue distribution of large sample covariance matrices of linear processes, Probab. Math. Statist., 31(2), 313–329, 2011.

5. Sample covariance matrices of linear processes

call $p^{-1}\mathbf{X}\mathbf{X}^T$ the sample covariance matrix. Due to Eq. (5.1.1), this change of normalization can be reversed by a simple transformation of the limiting spectral distribution. For notational convenience we suppress the explicit dependence of the occurring matrices on *n* and *p* where this does not cause ambiguity.

The distribution of Gaussian sample covariance matrices of fixed size was first computed in [113]. Several years later, it was Marčenko and Pastur [77] who considered the case where the random variables $\{X_{i,t}\}$ are more general i. i. d. random variables with finite second moments $\mathbb{E}X_{11}^2 = 1$, and the number *p* of variables is of the same order as the sample size *n*. They showed that the empirical spectral distribution (ESD) $F^{p^{-1}XX^T}$ of $p^{-1}XX^T$ converges, as $n \to \infty$, to a non-random distribution \hat{F} , called limiting spectral distribution (LSD), given by

$$\hat{F}(dx) = \frac{1}{2\pi x} \sqrt{(x_{+} - x)(x - x_{-})} \mathbf{1}_{\{x_{-} \le x \le x_{+}\}} dx,$$
(5.1.2)

and point mass $\hat{F}(\{0\}) = 1 - y$ if y < 1; in this formula, $x_{\pm} = (1 \pm \sqrt{y})^2$. Here and in the following, convergence of the ESD means almost sure convergence as a random element of the space of probability measures on \mathbb{R} equipped with the weak topology. In particular, the eigenvalues of the sample covariance matrix of a matrix with independent entries do not converge to the eigenvalues of the true covariance matrix, which is the identity matrix and therefore only has eigenvalue one. This leads to the failure of statistics that rely on the eigenvalues of $p^{-1}\mathbf{X}\mathbf{X}^T$ which have been derived under the assumption of fixed p, and random matrix theory is a tool to correct these statistics [8, 69]. In the case where the true covariance matrix is not the identity matrix, the LSD can in general only be given in terms of a non-linear equation for its Stieltjes transform, which is defined by

$$m_{\hat{F}}(z) = \int \frac{1}{\lambda - z} d\hat{F} \quad \forall z \in \mathbb{C}^+ \coloneqq \{z = u + iv \in \mathbb{C} : \Im_z = v > 0\}$$

Conversely, the distribution \hat{F} can be obtained from its Stieltjes transform $m_{\hat{F}}$ via the Stieltjes– Perron inversion formula ([9, Theorem B.8]), which states that

$$\hat{F}([a,b]) = \frac{1}{\pi} \lim_{\epsilon \to 0^+} \int_a^b \mathfrak{I}_{m_{\hat{F}}}(x+i\epsilon) \mathrm{d}x.$$
(5.1.3)

for all continuity points a < b of \hat{F} . For a comprehensive account of random matrix theory we refer the reader to [2], [9], [81], and the references therein.

Our aim in this chapter is to obtain a Marčenko–Pastur type result in the case where there is dependence within the rows of **X**. More precisely, for i = 1, ..., p, the *i*th row of **X** is given by a linear process of the form

$$(X_{i,t})_{t=1,\ldots,n} = \left(\sum_{j=0}^{\infty} c_j Z_{i,t-j}\right)_{t=1,\ldots,n}, \quad c_j \in \mathbb{R}.$$

Here, $(Z_{i,t})_{it}$ is an array of independent random variables that satisfies

$$\mathbb{E}Z_{i,t} = 0, \quad \mathbb{E}Z_{i,t}^2 = 1, \quad \text{and} \quad \sigma_4 \coloneqq \sup_{i,t} \mathbb{E}Z_{i,t}^4 < \infty, \tag{5.1.4}$$

as well as the Lindeberg-type condition that, for each $\epsilon > 0$,

$$\frac{1}{pn}\sum_{i=1}^{p}\sum_{j=1}^{n}\mathbb{E}\left(Z_{i,t}^{2}\mathbf{1}_{\{Z_{i,t}^{2} \ge \epsilon n\}}\right) \to 0, \quad \text{as} \quad n \to \infty.$$
(5.1.5)

Clearly, Eq. (5.1.5) is satisfied if all $\{Z_{i,t}\}$ are identically distributed.

The novelty of our result is that we allow for dependence within the rows, and that the equation for $m_{\hat{F}}$ is given in terms of the spectral density

$$f(\omega) = \sum_{h \in \mathbb{Z}} \gamma(h) e^{-ih\omega}, \quad \omega \in [0, 2\pi],$$

of the linear processes X_i only, which is the Fourier transform of the autocovariance function

$$\gamma(h) = \sum_{j=0}^{\infty} c_j c_{j+|h|}, \quad h \in \mathbb{Z}.$$

Potential applications arise whenever data is not independent in time such that the well-known Marčenko–Pastur law is not a good approximation. This includes e. g. wireless communications [109] and mathematical finance [23, 93]. Note that a similar question is also discussed in [10]. However, they have a different proof which relies on a moment condition to be verified. Furthermore, they assume that the random variables $\{Z_{i,t}\}$ are identically distributed so that the processes within the rows are independent copies of each other. More importantly, their results do not yield concrete formulas except in the AR(1) case and are therefore not directly applicable. In the context of free probability theory, the limiting spectral distribution of large sample covariance matrices of Gaussian ARMA processes is investigated in [30].

Before we present the main result of this article, we explain the notation used in this article. The symbols \mathbb{Z} , \mathbb{N} \mathbb{R} , and \mathbb{C} denote the sets of integers, natural, real, and complex numbers, respectively. For a matrix A, we write A^T for its transpose and trA for its trace. Finally, the indicator of an expression \mathcal{E} is denoted by $I_{\{\mathcal{E}\}}$ and defined to be one if \mathcal{E} is true, and zero otherwise; for a set S, we also write $I_S(x)$ instead of $I_{\{x \in S\}}$.

Theorem 5.1. For each i = 1, ..., p, let $X_{i,t} = \sum_{j=0}^{\infty} c_j Z_{i,t-j}$, $t \in \mathbb{Z}$, be a linear stochastic process with continuously differentiable spectral density f. Assume that

- *i)* the array $(Z_{i,t})_{it}$ satisfies conditions (5.1.4) and (5.1.5),
- *ii)* there exist positive constants C and δ such that $|c_j| \leq C(j+1)^{-1-\delta}$ for all $j \geq 0$,

iii) for almost all $\lambda \in \mathbb{R}$, $f(\omega) = \lambda$ for at most finitely many $\omega \in [0, 2\pi]$, and

iv) $f'(\omega) \neq 0$ for almost every ω .

Then the empirical spectral distribution $F^{p^{-1}\mathbf{X}\mathbf{X}^T}$ of $p^{-1}\mathbf{X}\mathbf{X}^T$ converges, as n tends to infinity, almost surely to a non-random probability distribution \hat{F} with bounded support. Moreover, there exist positive numbers λ_-, λ_+ such that the Stieltjes transform $z \mapsto m_{\hat{F}}(z)$ of \hat{F} is the unique mapping $\mathbb{C}^+ \to \mathbb{C}^+$ satisfying

$$\frac{1}{m_{\hat{F}}(z)} = -z + \frac{y}{2\pi} \int_{\lambda_{-}}^{\lambda_{+}} \frac{\lambda}{1 + \lambda m_{\hat{F}}(z)} \sum_{\omega \in [0, 2\pi]: f(\omega) = \lambda} \frac{1}{|f'(\omega)|} d\lambda.$$
(5.1.6)

The assumptions of the theorem are met, for instance, if $(X_{i,t})_t$ is an ARMA or fractionally integrated ARMA process; see Section 5.3 for details.

Theorem 5.1, as it stands, does not contain the classical Marčenko–Pastur law as a special case. For if the entries $X_{i,t}$ of the matrix **X** are i. i. d., the corresponding spectral density f is identically equal to the variance of $X_{1,1}$, and thus condition iv is not satisfied. We therefore also present a version of Theorem 5.1 that holds if the rows of the matrix **X** have a piecewise constant spectral density.

Theorem 5.2. For each i = 1, ..., p, let $X_{i,t} = \sum_{j=0}^{\infty} c_j Z_{i,t-j}$, $t \in \mathbb{Z}$, be a linear stochastic process with spectral density f of the form

$$f: [0, 2\pi] \to \mathbb{R}^+, \quad \omega \mapsto \sum_{j=1}^k \alpha_j \mathbf{1}_{A_j}(\omega), \quad k \in \mathbb{N},$$
 (5.1.7)

for some positive real numbers α_j and a measurable partition $A_1 \cup \cdots \cup A_k$ of the interval $[0, 2\pi]$. If conditions *i* and *ii* of Theorem 5.1 hold, then the empirical spectral distribution $F^{p^{-1}\mathbf{X}\mathbf{X}^T}$ of $p^{-1}\mathbf{X}\mathbf{X}^T$ converges, as $n \to \infty$, almost surely to a non-random probability distribution \hat{F} with bounded support. Moreover, the Stieltjes transform $z \mapsto m_{\hat{F}}(z)$ of \hat{F} is the unique mapping $\mathbb{C}^+ \to \mathbb{C}^+$ that satisfies

$$\frac{1}{m_{\hat{F}}(z)} = -z + \frac{y}{2\pi} \sum_{j=1}^{k} \frac{|A_j|\alpha_j}{1 + \alpha_j m_{\hat{F}}(z)},$$
(5.1.8)

where $|A_j|$ denotes the Lebesgue measure of the set A_j . In particular, if the entries of **X** are *i*. *i*. *d*. with unit variance, one recovers the limiting spectral distribution (5.1.2) of the Marčenko–Pastur law.

Remark 5.1.1. In applications one often considers processes of the form

$$X_{i,t} = \mu + \sum_{j=0}^{\infty} c_j Z_{i,t-j}$$

with mean $\mu \neq 0$. If we denote by $x_t \in \mathbb{R}^p$ the *t*th column of the matrix **X**, and define the empirical mean by $\overline{x} = p^{-1} \sum_{t=1}^{n} x_t$, then the sample covariance matrix is given by the expression $p^{-1} \sum_{t=1}^{n} (x_t - \overline{x})(x_t - \overline{x})^T$ instead of $p^{-1}\mathbf{X}\mathbf{X}^T$. However, by [9, Theorem A.44], the subtraction of the empirical mean does not change the LSD, and thus Theorems 5.1 and 5.2 remain valid if the underlying linear process has a non-zero mean.

Remark 5.1.2. The proof of Theorems 5.1 and 5.2 can easily be generalized to cover noncausal linear processes, which are defined as $X_{i,t} = \sum_{j=-\infty}^{\infty} c_j Z_{i,t-j}$. For this case one obtains the same result except that the autocovariance function is now given by $\sum_{j=-\infty}^{\infty} c_j c_{j+|h|}$.

Remark 5.1.3. If one considers a matrix **X** which has independent linear processes in its columns instead of its rows, one obtains the same formulas as in Theorems 5.1 and 5.2 except that y is replaced by y^{-1} . This is due to the fact that $\mathbf{X}^T \mathbf{X}$ and $\mathbf{X} \mathbf{X}^T$ have the same non-trivial eigenvalues.

In Section 5.2 we proceed with the proofs of Theorems 5.1 and 5.2. Thereafter we present some interesting examples in Section 5.3.

5.2. Proofs

In this section we present our proofs of Theorems 5.1 and 5.2. Dealing with infinite-order moving average processes directly is difficult, and we therefore first prove a variant of these theorems for the truncated processes $\widetilde{X}_{i,t} = \sum_{j=0}^{n} c_j Z_{i,t-j}$. We define the $p \times n$ matrix $\widetilde{\mathbf{X}} = (\widetilde{X}_{i,t})_{it}, i = 1, \dots, p, t = 1, \dots, n$.

Theorem 5.3. Under the assumptions of Theorem 5.1 (Theorem 5.2), the empirical spectral distribution of the sample covariance matrix of the truncated process \tilde{X} converges, as n tends to infinity, to a deterministic distribution with bounded support. Its Stieltjes transform is uniquely determined by Eq. (5.1.6) (Eq. (5.1.8)).

Proof. The proof starts from the observation that one can write $\widetilde{\mathbf{X}} = \mathbf{Z}H$, where $\mathbb{R}^{p \times 2n} \ni \mathbf{Z} = (Z_{i,t})_{it}, i = 1, \dots, p, t = 1 - n, \dots, n$, and

$$H = \begin{pmatrix} c_n & c_{n-1} & \dots & c_1 & c_0 & 0 & \dots & 0 \\ 0 & c_n & \dots & c_2 & c_1 & c_0 & & \vdots \\ \vdots & & \ddots & \vdots & \vdots & & \ddots & 0 \\ 0 & \dots & 0 & c_n & c_{n-1} & \dots & \dots & c_0 \end{pmatrix}^T \in \mathbb{R}^{2n \times n}.$$
 (5.2.1)

In particular, $\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^T = \mathbf{Z}HH^T\mathbf{Z}^T$. In order to prove convergence of the empirical spectral distribution $F^{p^{-1}\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^T}$ and to obtain a characterization of the limiting distribution, it suffices, by

5. Sample covariance matrices of linear processes

[86, Theorem 1], to prove that the spectral distribution F^{HH^T} of HH^T converges to a non-trivial limiting distribution. This will be done in Lemma 5.2.1, where the LSD of HH^T is shown to be $\hat{F}^{HH^T} = \frac{1}{2}\delta_0 + \frac{1}{2}\hat{F}^{\Gamma}$; the distribution \hat{F}^{Γ} is computed in Lemma 5.2.2 if we impose the assumptions of Theorem 5.1, respectively in Lemma 5.2.3 if we impose the assumptions of Theorem 5.2. Inserting this expression for \hat{F}^{HH^T} into equation (1.2) of [86] shows that the ESD $F^{p^{-1}\tilde{X}\tilde{X}^T}$ converges, as $n \to \infty$, almost surely to a deterministic distribution, which is determined by the requirement that its Stieltjes transform $z \mapsto m(z)$ satisfies

$$\frac{1}{m(z)} = -z + 2y \int_{\lambda_{-}}^{\lambda_{+}} \frac{\lambda}{1 + \lambda m(z)} d\hat{F}^{HH^{T}} = -z + y \int_{\lambda_{-}}^{\lambda_{+}} \frac{\lambda}{1 + \lambda m(z)} d\hat{F}^{\Gamma}.$$
 (5.2.2)

Using the explicit formulas of \hat{F}^{Γ} computed in Lemmas 5.2.2 and 5.2.3, one obtains Eqs. (5.1.6) and (5.1.8). Uniqueness of a mapping $m : \mathbb{C}^+ \to \mathbb{C}^+$ solving Eq. (5.2.2) was shown in [9, p. 88]. We complete the proof by arguing that the LSD of $p^{-1}\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^T$ has bounded support. For this it is enough, by [9, Theorem 6.3], to show that the spectral norm of HH^T is bounded in n, which is also done in Lemma 5.2.1.

Lemma 5.2.1. Let $H = (c_{n-i+j}\mathbf{1}_{\{0 \le n-i+j \le n\}})_{ij}$ be the matrix appearing in Eq. (5.2.1), and assume that there exist positive constants C, δ such that $|c_j| \le C(j+1)^{-1-\delta}$ (assumption ii of Theorem 5.1). Then the spectral norm of the matrix HH^T is bounded in n. If, moreover, the spectral distribution of the Toeplitz matrix $\Gamma = (\gamma(i-j))_{ij}$ converges weakly to some limiting distribution \hat{F}^{Γ} , then the spectral distribution F^{HH^T} converges weakly, as $n \to \infty$, to $\frac{1}{2}\delta_0 + \frac{1}{2}\hat{F}^{\Gamma}$.

Proof. We first introduce the notation $\mathcal{H} \coloneqq HH^T \in \mathbb{R}^{2n \times 2n}$ as well as the block decomposition $\mathcal{H} = \begin{bmatrix} \mathcal{H}_{11} & \mathcal{H}_{12} \\ \mathcal{H}_{12}^T & \mathcal{H}_{22} \end{bmatrix}$, $\mathcal{H}_{ij} \in \mathbb{R}^{n \times n}$. We prove the second part of the lemma first. There are several ways to show that the spectral distributions of two sequences of matrices converge to the same limit. In our case it is convenient to use [9, Corollary A.41] which states that two sequences A_n and B_n , either of whose empirical spectral distribution converges, have the same limiting spectral distribution if $n^{-1} \text{tr}(A_n - B_n)(A_n - B_n)^T$ converges to zero as n tends to infinity. We shall employ this result twice: first to show that the LSDs of $\mathcal{H} = HH^T$ and $\widetilde{\mathcal{H}} \coloneqq \text{diag}(0, \mathcal{H}_{22})$ agree, and then to prove equality of the LSDs of \mathcal{H}_{22} and Γ . Let $\Delta_{\mathcal{H}} = n^{-1} \text{tr}(\mathcal{H} - \widetilde{\mathcal{H}})(\mathcal{H} - \widetilde{\mathcal{H}})^T$; a direct calculation shows that $\Delta_{\mathcal{H}} = n^{-1} [\text{tr}\mathcal{H}_{11}\mathcal{H}_{11}^T + 2\text{tr}\mathcal{H}_{12}\mathcal{H}_{12}^T]$, and we will consider each of the two terms in turn. From the definition of H it follows that the (i, j)th entry of \mathcal{H} is given by

$$\mathcal{H}^{ij} = \sum_{k=1}^{n} c_{n-i+k} c_{n-j+k} \mathbf{1}_{\{\max(i,j)-n \leq k \leq \min(i,j)\}}$$

The trace of the square of the upper left block of $\mathcal H$ therefore satisfies

$$\operatorname{tr}\mathcal{H}_{11}\mathcal{H}_{11}^{T} = \sum_{i,j=1}^{n} \left\{ \mathcal{H}^{ij} \right\}^{2} = \sum_{i,j=1}^{n} \left[\sum_{k=1}^{\min(i,j)} c_{n-i+k} c_{n-j+k} \right]^{2}$$
$$\leq \sum_{i,j,k,l=1}^{n} |c_{i+k-1}|| c_{j+k-1} ||c_{i+l-1}|| c_{j+l-1}|$$
$$\leq C^{4} \sum_{i,j,k,l=2}^{n+1} i^{-1-\delta} j^{-1-\delta} l^{-1-\delta} k^{-1-\delta}$$
$$< \left[C\zeta(1+\delta) \right]^{4} < \infty,$$

where $\zeta(z)$ denotes the Riemann zeta function. As a consequence, the limit of n^{-1} tr $\mathcal{H}_{11}\mathcal{H}_{11}^{T}$ as *n* tends to infinity is zero. Similarly, we obtain for the trace of the square of the off-diagonal block of \mathcal{H} the bound

$$\operatorname{tr}\mathcal{H}_{12}\mathcal{H}_{12}^{T} = \sum_{i=1}^{n} \sum_{j=n+1}^{2n} \left\{ \mathcal{H}^{ij} \right\}^{2} = \sum_{i=1}^{n} \sum_{j=n+1}^{n+i} \left[\sum_{k=j-n}^{i} c_{n-i+k} c_{n-j+k} \right]^{2}$$

$$\leq \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{i=1}^{n-i+1} \sum_{l=j}^{n-i+1} c_{i+k-1} c_{k-j} c_{i+l-1} c_{l-j}$$

$$\leq \sum_{i=1}^{n} \sum_{j=1}^{n} \sum_{r=0}^{n} \sum_{s=0}^{n} |c_{i+r+j-1}|| c_{r} || c_{s+j-1} || c_{s} |$$

$$\leq C^{4} \sum_{i,j,r,s=1}^{n+1} i^{-1-\delta} r^{-1-\delta} s^{-1-\delta} j^{-1-\delta}$$

$$< \left[C\zeta (1+\delta) \right]^{4} < \infty,$$

which shows that the limit of n^{-1} tr $\mathcal{H}_{12}\mathcal{H}_{12}^{T}$ is zero. It follows, as *n* goes to infinity, that $\Delta_{\mathcal{H}}$, as defined in Lemma 5.2.1, converges to zero, and therefore that the LSDs of \mathcal{H} and $\widetilde{\mathcal{H}} = \text{diag}(0, \mathcal{H}_{22})$ coincide. The latter distribution is clearly given by $F^{\widetilde{\mathcal{H}}} = \frac{1}{2}\delta_0 + \frac{1}{2}F^{\mathcal{H}_{22}}$, and we show next that the LSD of \mathcal{H}_{22} agrees with the LSD of $\Gamma = (\gamma(i-j))_{ij}$. As before it suffices to show, by [9, Corollary A.41], that $\Delta_{\Gamma} = n^{-1}$ tr $(\mathcal{H}_{22} - \Gamma)(\mathcal{H}_{22} - \Gamma)^{T}$ converges to zero as *n* tends to infinity. It follows from the definitions of \mathcal{H} and Γ that $n\Delta_{\Gamma}$ can be estimated

as

$$\begin{split} n\Delta_{\Gamma} &= \sum_{i,j=1}^{n} \left[\sum_{k=\max(i,j)}^{n} c_{k-i}c_{k-j} - \sum_{k=1}^{\infty} c_{k-1}c_{k+|i-j|-1} \right]^{2} \\ &= \sum_{i,j=1}^{n} \left[\sum_{k=\max(i,j)}^{n} c_{k-i}c_{k-j} - \sum_{k=\max(i,j)}^{\infty} c_{k-i}c_{k-j} \right]^{2} \\ &= \sum_{i,j=1}^{n} \sum_{k,l=1}^{\infty} c_{k+i-1}c_{k+j-1}c_{l+i-1}c_{l+j-1} \\ &\leq C^{4} \sum_{i,j=2}^{n+1} \sum_{k,l=2}^{\infty} i^{-1-\delta}j^{-1-\delta}k^{-1-\delta}l^{-1-\delta} < \left[C\zeta(1+\delta) \right]^{4} < \infty. \end{split}$$

Consequently, Δ_{Γ} converges to zero as *n* goes to infinity, and it follows that $\hat{F}^{\mathcal{H}} = \frac{1}{2}\delta_0 + \frac{1}{2}\hat{F}^{\Gamma}$.

In order to show that the spectral norm of $\mathcal{H} = HH^T$ is bounded in *n*, we use Gerschgorin's circle theorem ([52, Theorem 2]), which states that every eigenvalue of \mathcal{H} lies in at least one of the balls $B(\mathcal{H}^{ii}, R_i)$ with centre \mathcal{H}^i and radius R_i , i = 1, ..., 2n, where the radii R_i are defined as $R_i = \sum_{j \neq i} |\mathcal{H}^{ij}|$. We first note that the centres \mathcal{H}^{ii} satisfy

$$\mathcal{H}^{ii} = \sum_{k=\max\{1,i-n\}}^{\min\{i,n\}} c_{n-i+k}^2 \leq \sum_{k=0}^n c_k^2 \leq [C\zeta(2+2\delta)]^2 < \infty.$$

To obtain a uniform bound for the radii R_i we first assume that i = 1, ..., n. Then

$$\begin{aligned} |R_i| &\leq \sum_{j=1}^n \sum_{k=1}^{\min\{i,j\}} |c_{n-i+k}|| c_{n-j+k}| + \sum_{j=n+1}^{2n} \sum_{k=j-n}^i |c_{n-i+k}|| c_{n-j+k}| \\ &\leq \sum_{j,k=1}^n |c_{n-i+k}|| c_{j+k-1}| + \sum_{j=n+1-i}^{2n-i} \sum_{k=0}^{n-j} |c_{k+j}|| c_k| \leq 2 \left[C\zeta(1+\delta) \right]^2 < \infty. \end{aligned}$$

Similarly we find that, for i = n + 1, ..., 2n,

$$\begin{aligned} |R_i| &\leq \sum_{j=1}^n \sum_{k=i-n}^j |c_{n-i+k}| |c_{n-j+k}| + \sum_{j=n+1}^{2n} \sum_{k=\max\{i,j\}-n}^n |c_{n-i+k}| |c_{n-j+k}| \\ &\leq \sum_{j=i-n}^{i-1} \sum_{k=0}^{n+1-j} |c_{k+j}| |c_k| + \sum_{j=n+1}^{2n} \sum_{k=0}^{n-\max\{i,j\}} |c_k| |c_{k+|j-i|}| \leq 3 \left[C\zeta(1+\delta) \right]^2 \end{aligned}$$

is bounded, which completes the proof.

In the following two lemmas, we argue that the distribution \hat{F}^{Γ} exists and we prove explicit formulas for it in the case that the assumptions of Theorem 5.1 or Theorem 5.2 are satisfied.

Lemma 5.2.2. Let $(c_j)_j$ be a sequence of real numbers, $\gamma : h \mapsto \sum_{j=0}^{\infty} c_j c_{j+|h|}$, and $f : \omega \mapsto \sum_{h \in \mathbb{Z}} \gamma(h) e^{-ih\omega}$. Under the assumptions of Theorem 5.1 it holds that the spectral distribution F^{Γ} of $\Gamma = (\gamma(i-j))_{ij}$ converges weakly, as $n \to \infty$, to an absolutely continuous distribution \hat{F}^{Γ} with bounded support and density

$$g: (\lambda_{-}, \lambda_{+}) \to \mathbb{R}^{+}, \quad \lambda \mapsto \frac{1}{2\pi} \sum_{\omega: f(\omega) = \lambda} \frac{1}{|f'(\omega)|}.$$
 (5.2.3)

Proof. We first note that under assumption if of Theorem 5.1 the autocovariance function γ is absolutely summable because

$$\sum_{h=0}^{\infty} |\gamma(h)| \leq \sum_{h=0}^{\infty} \sum_{j=0}^{\infty} |c_j| |c_{j+h}| \leq C^2 \sum_{h,j=1}^{\infty} h^{-1-\delta} j^{-1-\delta} < \left[C\zeta(1+\delta)^2 < \infty \right].$$

Szegő's first convergence theorem ([59] and [58, Corollary 4.1]) then implies that \hat{F}^{Γ} exists, and that the cumulative distribution function of the eigenvalues of the Toeplitz matrix Γ associated with the sequence $h \mapsto \gamma(h)$ is given by

$$G(\lambda) \coloneqq \frac{1}{2\pi} \int_0^{2\pi} \mathbf{1}_{\{f(\omega) \le \lambda\}} \mathrm{d}\omega = \frac{1}{2\pi} \operatorname{Leb}(\{\omega \in [0, 2\pi] : f(\omega) \le \lambda\}), \qquad (5.2.4)$$

for all λ such that the level sets { $\omega \in [0, 2\pi] : f(\omega) = \lambda$ } have Lebesgue measure zero. By assumption iii of Theorem 5.1, Eq. (5.2.4) holds for almost all λ . In order to prove that the LSD \hat{F}^{Γ} is absolutely continuous with respect to the Lebesgue measure, it suffices to prove that the cumulative distribution function *G* is differentiable almost everywhere. Clearly, for $\Delta \lambda > 0$,

$$G(\lambda + \Delta \lambda) - G(\lambda) = \frac{1}{2\pi} \operatorname{Leb}(\{\omega \in [0, 2\pi] : \lambda < f(\omega) \leq \lambda + \Delta \lambda\}).$$

Due to assumption iv of Theorem 5.1, the set of all $\lambda \in \mathbb{R}$ such that the set $\{\omega : \in [0, 2\pi] : f(\omega) = \lambda$ and $f'(\omega) = 0\}$ is non-empty is a Lebesgue null-set. Hence it is enough to consider only λ for which this set is empty. Let $f^{-1}(\lambda) = \{\omega : f(\omega) = \lambda\}$ be the pre-image of λ , which is a finite set by assumption iii. The implicit function theorem then asserts that, for every $\omega \in f^{-1}(\lambda)$, there exists an open interval I_{ω} around ω such that f restricted to I_{ω} is invertible. It is no restriction to assume that these I_{ω} are disjoint. By choosing $\Delta\lambda$ sufficiently small it can be ensured that the interval $[\lambda, \Delta\lambda]$ is contained in $\bigcap_{\omega \in f^{-1}(\lambda)} f(I_{\omega})$, and from the continuity of f it follows that outside of $\bigcup_{\omega \in f^{-1}(\lambda)} I_{\omega}$, the values of f are bounded away from λ , so that

$$\lim_{\Delta \lambda \to 0} \frac{1}{\Delta \lambda} \left[G(\lambda + \Delta \lambda) - G(\lambda) \right]$$

= $\frac{1}{2\pi} \lim_{\Delta \lambda \to 0} \frac{1}{\Delta \lambda} \operatorname{Leb} \left(\bigcup_{\omega \in f^{-1}(\lambda)} \{ \omega' \in I_{\omega} : \lambda < f(\omega') \leq \lambda + \Delta \lambda \} \right)$
= $\frac{1}{2\pi} \sum_{\omega \in f^{-1}(\lambda)} \lim_{\Delta \lambda \to 0} \frac{1}{\Delta \lambda} \operatorname{Leb} \left(\{ \omega' \in I_{\omega} : \lambda < f(\omega') \leq \lambda + \Delta \lambda \} \right).$

5. Sample covariance matrices of linear processes

In order to further simplify this expression, we denote the local inverse functions by f_{ω}^{-1} : $f(I_{\omega}) \to [0, 2\pi]$. Observing that the Lebesgue measure of an interval is given by its length, and that the derivatives of f_{ω}^{-1} are given by the inverse of the derivative of f, it follows that

$$\begin{split} \lim_{\Delta\lambda\to 0} \frac{1}{\Delta\lambda} \left[G(\lambda + \Delta\lambda) - G(\lambda) \right] &= \frac{1}{2\pi} \sum_{\omega \in f^{-1}(\lambda)} \lim_{\Delta\lambda\to 0} \frac{1}{\Delta\lambda} \left| f_{\omega}^{-1}(\lambda + \Delta\lambda) - f_{\omega}^{-1}(\lambda) \right| \\ &= \frac{1}{2\pi} \sum_{\omega \in f^{-1}(\lambda)} \left| \frac{\mathrm{d}}{\mathrm{d}\lambda} f_{\omega}^{-1}(\lambda) \right| \\ &= \frac{1}{2\pi} \sum_{\omega \in f^{-1}(\lambda)} \frac{1}{|f'(\omega)|}. \end{split}$$

This shows that G is differentiable almost everywhere with derivative

$$g: \lambda \mapsto \frac{1}{2\pi} \sum_{\omega \in f^{-1}(\lambda)} \frac{1}{|f'(\omega)|}$$

It remains to argue that the support of \hat{F}^{Γ} is bounded. The absolute summability of $\gamma(\cdot)$ implies boundedness of its Fourier transform f. The claim then follows from Eq. (5.2.4), which shows that the support of g is equal to the range of f.

Lemma 5.2.3. Let $f : \omega \mapsto \sum_{j=1}^{k} \alpha_j \mathbf{1}_{A_j}(\omega)$ be the piecewise constant spectral density of the linear process $X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}$, and denote the corresponding autocovariance function by $\gamma : h \mapsto \sum_{j=0}^{\infty} c_j c_{j+|h|}$. Under the assumptions of Theorem 5.2 it holds that the spectral distribution F^{Γ} of $\Gamma = (\gamma(i-j))_{ij}$ converges weakly, as $n \to \infty$, to the distribution $\hat{F}^{\Gamma} = (2\pi)^{-1} \sum_{i=1}^{k} |A_j| \delta_{\alpha_j}$.

Proof. Without loss of generality we may assume that $0 < \alpha_1 < ... < \alpha_k$. As in the proof of Lemma 5.2.2 one sees that \hat{F}^{Γ} exists, and that $\hat{F}^{\Gamma}(-\infty, \lambda)$ is given by

$$G(\lambda) \coloneqq \frac{1}{2\pi} \operatorname{Leb}(\{\omega \in [0, 2\pi] : f(\omega) \leq \lambda\}), \quad \forall \lambda \in [0, 2\pi] \setminus \bigcup_{j=1}^{k} \{\alpha_j\}.$$

The special structure of f thus implies that $G(\lambda) = (2\pi)^{-1} \sum_{j=1}^{k_{\lambda}} |A_j|$, where k_{λ} is the largest integer such that $\alpha_{k_{\lambda}} \leq \lambda$. Since G must be right-continuous, this formula holds for all λ in the interval $[0, 2\pi]$. It is easy to see that the function G is the cumulative distribution function of the discrete measure $(2\pi)^{-1} \sum_{j=1}^{k} |A_j| \delta_{\alpha_j}$, which completes the proof.

Proof. of Theorems 5.1 and 5.2 It is only left to show that the truncation performed in Theorem 5.3 does not alter the LSD, i. e. that the difference of $F^{p^{-1}XX^{T}}$ and $F^{p^{-1}\tilde{X}\tilde{X}^{T}}$ converges to zero almost surely. By [9, Corollary A.42], this means that we have to show that

$$\underbrace{\frac{1}{p^2} \operatorname{tr}(\mathbf{X}\mathbf{X}^T + \widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^T)}_{=\mathrm{I}} \underbrace{\frac{1}{p^2} \operatorname{tr}((\mathbf{X} - \widetilde{\mathbf{X}})(\mathbf{X} - \widetilde{\mathbf{X}})^T)}_{=\mathrm{II}}$$
(5.2.5)

converges to zero. To this end we show that I has a limit, and that II converges to zero, both almost surely. By the definition of X and \tilde{X} we have

$$II = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{t=1}^{n} \sum_{k=n+1}^{\infty} \sum_{m=n+1}^{\infty} c_k c_m Z_{i,t-k} Z_{i,t-m}.$$

We shall prove that the variances of II are summable. For this purpose we need the following two estimates which are implied by the Cauchy–Schwarz inequality, the assumption that $\sigma_4 = \sup_{i,t} \mathbb{E}Z_{i,t}^4$ is finite, and the assumed absolute summability of the coefficients $(c_j)_j$:

$$\mathbb{E}\sum_{i=1}^{p}\sum_{t=1}^{n}\sum_{k,m=1}^{\infty}|c_{k}c_{m}Z_{i,t-k}Z_{i,t-m}| \le pn\left(\sum_{k=1}^{\infty}|c_{k}|\right)^{2} < \infty,$$
(5.2.6a)

$$\mathbb{E}\sum_{i,i'=1=1}^{p}\sum_{t,t'=1}^{n}\sum_{k,k',m,m'=1}^{\infty}|c_{k}c_{m}c_{k'}c_{m'}Z_{i,t-k}Z_{i,t-m}Z_{i',t'-k'}Z_{i',t'-m'}|$$

$$\leq (np)^{2}\sigma_{4}\left(\sum_{k=1}^{\infty}|c_{k}|\right)^{4}<\infty.$$
(5.2.6b)

Therefore we can, by Fubini's theorem, interchange expectation and summation to bound the variance of II as

$$\operatorname{Var}(\mathrm{II}) \leq \frac{1}{p^4} \sum_{i,i'=1}^{p} \sum_{t,t'=1}^{n} \sum_{\substack{k,k'\\m,m'}=n+1}^{\infty} c_k c_m c_{k'} c_{m'} \mathbb{E}(Z_{i,t-k} Z_{i,t-m} Z_{i',t'-k'} Z_{i',t'-m'}).$$

Considering separately the terms where i = i' and $i \neq i'$, we can write

$$\operatorname{Var}(\mathrm{II}) \leq \frac{1}{p^{4}} \sum_{\substack{i,i'=1\\i\neq i'}}^{p} \sum_{t,t'=1}^{n} \sum_{\substack{k,k'\\m,m'}=n+1}^{\infty} c_{k} c_{m} c_{k'} c_{m'} \mathbb{E}(Z_{i,t-k} Z_{i,t-m} Z_{i',t'-k'} Z_{i',t'-m'}) + \frac{1}{p^{4}} \sum_{i=1}^{p} \sum_{t,t'=1}^{n} \sum_{\substack{k,k'\\m,m'}=n+1}^{\infty} c_{k} c_{m} c_{k'} c_{m'} \mathbb{E}(Z_{i,t-k} Z_{i,t-m} Z_{i,t'-k'} Z_{i,t'-m'}).$$

For the expectation in the first sum not to be zero, k must equal m and k' must equal m', in which case its value is unity. The expectation in the second term can always be bounded by σ_4 , so that we obtain

$$\operatorname{Var}(\mathrm{II}) \leq \frac{p^2 - p}{p^4} n^2 \left(\sum_{k=n+1}^{\infty} c_k^2 \right)^2 + \sigma_4 \frac{p n^2}{p^4} \left(\sum_{k=n+1}^{\infty} |c_k| \right)^4.$$

5. Sample covariance matrices of linear processes

Due to Eq. (5.1.1) and the assumed polynomial decay of c_k there exists a constant K such that the right hand side is bounded by $Kn^{-1-4\delta}$, which implies that

$$\sum_{n=1}^{\infty} \operatorname{Var}\left(\operatorname{II}\right) \leqslant K \sum_{n=1}^{\infty} n^{-1-4\delta} < \infty,$$

and therefore, by the first Borel–Cantelli lemma, that II converges to a constant almost surely. In order to show that this constant is zero, it suffices to shows that the expectation of II converges to zero. Since $\mathbb{E}Z_{i,t} = 0$, and the $\{Z_{i,t}\}$ are independent, one sees, using Eq. (5.2.6a) and again Fubini's theorem, that $\mathbb{E}(II) = np^{-1} \sum_{k=n+1}^{\infty} c_k^2$, which converges to zero because the $\{c_k\}$ are square-summable.

We now consider factor I of expression (5.2.5) and define $\Delta_X = \mathbf{X}\mathbf{X}^T - \widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^T$. Then

$$\mathbf{I} = \underbrace{\frac{1}{p^2} \mathrm{tr}(\Delta_X)}_{=\mathbf{I}_a} + 2 \underbrace{\frac{1}{p^2} \mathrm{tr}(\widetilde{\mathbf{X}} \widetilde{\mathbf{X}}^T)}_{=\mathbf{I}_b}.$$
(5.2.7)

Because of

$$(\mathbf{X}\mathbf{X}^{T})_{ii} = \sum_{t=1}^{n} X_{i,t}^{2} = \sum_{t=1}^{n} \sum_{k=0}^{\infty} \sum_{m=0}^{\infty} c_{k}c_{m}Z_{i,t-k}Z_{i,t-m},$$

and similarly $(\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^T)_{ii} = \sum_{t=1}^n \sum_{k=0}^n \sum_{m=0}^n c_k c_m Z_{i,t-k} Z_{i,t-m}$, we have that

$$\operatorname{tr}(\Delta_{X}) = \sum_{i=1}^{p} \left[(\mathbf{X}\mathbf{X}^{T})_{ii} - (\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{T})_{ii} \right]$$

$$= \sum_{i=1}^{p} \sum_{t=1}^{n} \sum_{k=n+1}^{\infty} \sum_{m=n+1}^{\infty} c_{k}c_{m}Z_{i,t-k}Z_{i,t-m}$$

$$= \operatorname{II} \rightarrow 0 \text{ a.s.}$$

$$+ 2 \sum_{i=1}^{p} \sum_{t=1}^{n} \sum_{k=n+1}^{\infty} \sum_{m=1}^{n} c_{k}c_{m}Z_{i,t-k}Z_{i,t-m}.$$
 (5.2.8)

Equation (5.2.6b) allows us to apply Fubini's theorem to compute the variance of the second term in the previous display as

$$\frac{4}{p^4} \sum_{i,i'=1}^{p} \sum_{t,t'=1}^{n} \sum_{k,k'=n+1}^{\infty} \sum_{m,m'=1}^{n} c_k c_m c_{k'} c_{m'} \mathbb{E}(Z_{i,t-k} Z_{i,t-m} Z_{i',t'-k'} Z_{i',t'-m'}),$$

which is, by the same reasoning as we did for II, bounded by

$$4\sigma_4 \frac{p}{p^4} n^2 \left(\sum_{k=n+1}^{\infty} |c_k| \right)^2 \left(\sum_{m=1}^n |c_m| \right)^2 \le K n^{-1-2\delta},$$



Figure 5.1.: Limiting spectral densities $\lambda \mapsto p(\lambda)$ of $p^{-1}\mathbf{X}\mathbf{X}^T$ for the MA(1) process $X_t = Z_t + \vartheta Z_{t-1}$ for different values of ϑ and y = n/p

for some positive constant *K*. Clearly, this is summable in *n*. Having, by Eq. (5.2.6a), expected value zero, the second term of Eq. (5.2.8) and, therefore, also $tr(\Delta_X)$ both converge to zero almost surely. Thus, we only have to look at the contribution of I_b in expression (5.2.7). From Theorem 5.3 we know that $F^{p^{-1}\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^T}$ converges almost surely weakly to some non-random distribution \hat{F} with bounded support. Hence, denoting by $\lambda_1, \ldots, \lambda_p$ the eigenvalues of $p^{-1}\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^T$,

$$\mathbf{I}_{\mathbf{b}} = \frac{1}{p} \operatorname{tr} \left(\frac{1}{p} \widetilde{\mathbf{X}} \widetilde{\mathbf{X}}^{T} \right) = \frac{1}{p} \sum_{i=1}^{p} \lambda_{i} = \int \lambda \mathrm{d} F^{\frac{1}{p}} \widetilde{\mathbf{X}} \widetilde{\mathbf{X}}^{T} \to \int \lambda \mathrm{d} \hat{F} < \infty,$$

almost surely. It follows that, in Eq. (5.2.5), factor I is bounded, and factor II converges to zero, and so the proof of Theorems 5.1 and 5.2 is complete. \Box

5.3. Illustrative examples

For several classes of widely employed linear processes, Theorem 5.1 can be used to obtain an explicit description of the limiting spectral distribution. In this section we consider the class of autoregressive moving average (ARMA) processes as well as fractionally integrated ARMA models. The distributions we obtain in the case of AR(1) and MA(1) processes can be interpreted as one-parameter deformations of the classical Marčenko–Pastur law.

5.3.1. Autoregressive moving average processes

Given polynomials $a : z \mapsto 1 + a_1z + ... a_pz^p$ and $b : z \mapsto 1 + b_1z + ... + b_qz^q$, an ARMA(p,q) process X with autoregressive polynomial a and moving average polynomial b is defined as the stationary solution to the stochastic difference equation

$$X_t + a_1 X_{t-1} + \ldots + a_p X_{t-p} = Z_t + b_1 Z_{t-1} + \ldots + b_q Z_{t-q}, \quad t \in \mathbb{Z}$$



Figure 5.2.: Limiting spectral densities $\lambda \mapsto p(\lambda)$ of $p^{-1}\mathbf{X}\mathbf{X}^T$ for the AR(1) process $X_t = \varphi X_{t-1} + Z_t$ for different values of φ and y = n/p

If the zeros of *a* lie outside the closed unit disk, it is well known that *X* has an infinite-order moving average representation $X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}$, where $\{c_j\}$ are the coefficients in the power series expansion of b(z)/a(z) around zero. It is also known ([24]) that there exist positive constants $\rho < 1$ and *K* such that $|c_j| \leq K\rho^j$, so that assumption ii of Theorem 5.1 is satisfied. While the autocovariance function of a general ARMA process does not in general have a simple closed form, its Fourier transform is given by

$$f(\omega) = \left| \frac{b\left(e^{i\omega} \right)}{a\left(e^{i\omega} \right)} \right|^2, \quad \omega \in [0, 2\pi].$$
(5.3.1)

Since *f* is rational, assumptions iii and iv of Theorem 5.1 are satisfied as well. In order to compute the LSD of Γ , it is necessary, by Lemma 5.2.2, to find the roots of a trigonometric polynomial of possibly high degree, which can be done numerically.

We now consider the special case of the ARMA(1,1) process $X_t = \varphi X_{t-1} + Z_t + \vartheta Z_{t-1}$, $|\varphi| < 1$, for which one can obtain explicit results. By Eq. (5.3.1), the spectral density of X is given by

$$f(\omega) = \frac{1 + \vartheta^2 + 2\vartheta \cos \omega}{1 + \varphi^2 - 2\varphi \cos \omega}, \quad \omega \in [0, 2\pi].$$

Equation (5.2.3) implies that the LSD of the autocovariance matrix Γ has a density *g*, which is given by

$$g(\lambda) = \frac{1}{2\pi} \sum_{\omega \in [0,2\pi]: f(\omega) = \lambda} \frac{1}{|f'(\omega)|}$$
$$= \frac{1}{\pi(\vartheta + \varphi\lambda) \sqrt{[(1+\vartheta)^2 - \lambda(1-\varphi)^2] [\lambda(1+\varphi)^2 - (1-\vartheta)^2]}} \mathbf{1}_{(\lambda_-,\lambda_+)}(\lambda),$$

where

$$\lambda_{-} = \min(\lambda^{-}, \lambda^{+}), \quad \lambda_{+} = \max(\lambda^{-}, \lambda^{+}), \quad \lambda^{\pm} = \frac{(1 \pm \vartheta)^{2}}{(1 \mp \varphi)^{2}}$$



Figure 5.3.: Histograms of the eigenvalues and limiting spectral densities $\lambda \mapsto p(\lambda)$ of $p^{-1}\mathbf{X}\mathbf{X}^T$ for the ARMA(1,1) process $X_t = \frac{1}{2}X_{t-1} + Z_t + Z_{t-1}$ for different values of y = n/p, p = 1000

By Theorem 5.1, the Stieltjes transform $z \mapsto m_z$ of the limiting spectral distribution of $p^{-1}\mathbf{X}\mathbf{X}^T$ is the unique mapping $m : \mathbb{C}^+ \to \mathbb{C}^+$ that satisfies the equation

$$\frac{1}{m_z} = -z + y \int_{\lambda_-}^{\lambda_+} \frac{\lambda g(\lambda)}{1 + \lambda m_z} d\lambda$$

$$= -z + \frac{\vartheta y}{\vartheta m_z - \varphi} \qquad (5.3.2)$$

$$- \frac{(\vartheta + \varphi)(1 + \vartheta \varphi)y}{(\vartheta m_z - \varphi) \sqrt{[(1 - \varphi)^2 + m_z(1 + \vartheta)^2] [(1 + \varphi)^2 + m_z(1 - \vartheta)^2]}}.$$

This is a quartic equation in $m_z \equiv m(z)$ which can be solved explicitly. An application of the Stieltjes inversion formula (5.1.3) then yields the limiting spectral distribution of $p^{-1}\mathbf{X}\mathbf{X}^T$.

If one sets $\varphi = 0$, one obtains an MA(1) process; plots of the densities obtained in this case for different values of ϑ and y are displayed in Fig. 5.1. Similarly, the case $\vartheta = 0$ corresponds to an AR(1) process; see Fig. 5.2 for a graphical representation of the densities one obtains for different values of φ and y in this case. For the special case $\varphi = 1/2$, $\vartheta = 1$, Fig. 5.3 compares the histogram of the eigenvalues of $p^{-1}XX^T$ with the limiting spectral distribution obtained from Theorem 5.1 for different values of y.

Equation (5.3.2) for the Stieltjes transform of the limiting spectral distribution of the sample covariance matrix of an ARMA(1,1) process should be compared to [10, Eq. (2.10)], where the analogous result is obtained for an autoregressive process of order one. They use the notation $c = \lim p/n$ and consider the spectral distribution of $n^{-1}XX^{T}$ instead of $p^{-1}XX^{T}$. If one observes that this difference in the normalization amounts to a linear transformation of the corresponding Stieltjes transform, one obtains their result as a special case of Eq. (5.3.2).

5.3.2. Fractionally integrated ARMA processes

In many practical situations, data exhibit long-range dependence, which can be modelled by long-memory processes. Denote by B the backshift operator and define, for d > -1, the (fractional) difference operator by

$$abla^d = (1-\mathbf{B})^d = \sum_{j=0}^{\infty} \prod_{k=1}^j \frac{k-1-d}{k} \mathbf{B}^j, \quad \mathbf{B}^j X_t = X_{t-j}.$$

A process $(X_t)_t$ is called a fractionally integrated ARMA(p,d,q) processes with $d \in (-1/2, 1/2)$ and $p, q \in \mathbb{N}$ if $(\nabla^d X_t)_t$ is an ARMA(p,q) process. These processes have a polynomially decaying autocorrelation function and therefore exhibit long-range-dependence, cf. [24, Theorem 13.2.2] and [56, 66]. We assume that d < 0, and that the zeros of the autoregressive polynomial a of $(\nabla^d X_t)_t$ lie outside the closed unit disk. Then it follows that X has an infiniteorder moving average representation $X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}$, where the $(c_j)_j$ have, in contrast to our previous examples, not an exponential decay, but satisfy $K_1(j+1)^{d-1} \leq c_j \leq K_2(j+1)^{d-1}$, for some $K_1, K_2 > 0$. Therefore, if d < 0, one can apply Theorem 5.1 to obtain the LSD of the sample covariance matrix, using that the spectral density of $(X_t)_t$ is given by

$$f(\omega) = \left| \frac{b(\mathbf{e}^{\mathrm{i}\omega})}{a(\mathbf{e}^{\mathrm{i}\omega})} \right|^2 \left| 1 - \mathbf{e}^{-\mathrm{i}\omega} \right|^{-2d}, \quad \omega \in [0, 2\pi].$$

CHAPTER 6

Limiting spectral distribution of a new random matrix model with dependence across rows and columns³

6.1. Introduction

Random matrix theory studies the properties of large random matrices $A = (A_{i,j})_{ij} \in \mathbb{K}^{p \times n}$, for some field K. In this article, the entries A_{ij} are real random variables unless otherwise specified. Commonly, the focus is on asymptotic properties of such matrices as their dimensions tend to infinity. One particularly interesting object of study is the asymptotic distribution of their singular values. Since the squared singular values of A are the eigenvalues of AA^{T} , this is often done by investigating the eigenvalues of AA^{T} , which is called a sample covariance matrix. The spectral characteristics of a $p \times p$ matrix S are conveniently studied via its empirical spectral distribution, which is defined as $F^{S} = p^{-1} \sum_{i=1}^{p} \delta_{\lambda_i}$; here, $\{\lambda_1, \ldots, \lambda_p\}$ are the eigenvalues of S, and δ_x denotes the Dirac measure located at x. For some set $B \subset \mathbb{R}$, the figure $F^{S}(B)$ is the number of eigenvalues of S that lie in B. The measure F^{S} is considered a random element of the space of probability distributions equipped with the weak topology, and we are interested in its limit as both n and p tend to infinity such that the ratio p/n converges to a finite positive limit y.

The first result of this kind can be found in the remarkable paper of Marčenko and Pastur [77]. They showed that $F^{p^{-1}AA^{\mathsf{T}}}$ converges to a non-random limiting spectral distribution $\hat{F}^{p^{-1}AA^{\mathsf{T}}}$ if all A_{ij} are independent, identically distributed, centred random variables with finite

³This chapter is based on *O. Pfaffel and E. Schlemm:* Limiting spectral distribution of a new random matrix model with dependence across rows and columns. Linear Algebra and Its Applications 436 (2012) pp. 2960-2973.

fourth moment. Interestingly, the Lebesgue density of $\hat{F}^{p^{-1}AA^{\mathsf{T}}}$ is given by an explicit formula which only involves the ratio *y* and the common variance of A_{ij} and is therefore universal with respect to the distribution of the entries of *A*. Subsequently [110, 114], the same result was obtained under the weaker moment condition that the entries A_{ij} have finite variance. The requirement that the entries of *A* be identically distributed has later been relaxed to a Lindeberg-type condition, cf. Eq. (5.1.5). For more details and a comprehensive treatment of random matrix theory we refer the reader to the text books [2, 9, 81].

Recent research has focused on the question to what extent the assumption of independence of the entries of A can be relaxed without compromising the validity of the Marčenko–Pastur law. In [6] it was shown that for random matrices A whose rows are independent \mathbb{R}^n -valued random variables uniformly distributed on the unit ball of $l_q(\mathbb{R}^n)$, q > 1, the empirical spectral distribution $F^{p^{-1}AA^{\mathsf{T}}}$ still converges to the same law as in the i. i. d. case. The Marčenko–Pastur law is, however, not stable with respect to more substantial deviations from the independence assumptions.

A very useful tool to characterize the limiting spectral distribution in random matrix models with dependent entries is the Stieltjes transform which, for some measure μ , is defined as the map $s_{\mu}: \mathbb{C}^+ \to \mathbb{C}^+, s_{\mu}(z) = \int_{\mathbb{R}} (t-z)^{-1} \mu(\mathrm{d}t)$. A particular, very successful random matrix model exhibiting dependence within the rows was investigated already by [77] and later in greater generality by [86, 98]: they modelled dependent data as a linear transformation of independent random variables which led to the study of the eigenvalues of random matrices of the form AHA^{T} , where the entries of A are independent, and H is a positive semidefinite population covariance matrix whose spectral distribution converges to a non-random limit \hat{F}^{H} . They found that the Stieltjes transform of the limiting spectral distribution of $p^{-1}AHA^{T}$ can be characterized as the solution to an integral equation involving only \hat{F}^{H} and the ratio y = $\lim p/n$. Another approach, suggested in [10] and further pursued in Chapter 5, is to model the rows of A independently as stationary linear processes with independent innovations. This structure is interesting because the class of linear processes includes many practically relevant time series models, such as (fractionally integrated) ARMA processes, as special cases. The main result of Chapter 5 shows that for this model the limiting spectral distribution depends only on y and the second-order properties of the underlying linear process.

All results for independent rows with dependent row entries also hold with minor modifications for the case where A has independent columns with dependent column entries. This is due to the fact that the matrices AA^{T} and $A^{T}A$ have the same non-zero eigenvalues.

In contrast, there are only very few results dealing with random matrix models where the entries are dependent across both rows and columns. The case where *A* is given as the result of a two-dimensional linear filter applied to an array of independent complex Gaussian random variables is considered in [60]. They use the fact that *A* can be transformed to a random matrix with uncorrelated, non-identically distributed entries. Because of the assumption of

Gaussianity the entries are in fact independent, and so an earlier result by the same authors [61] can be used to obtain the asymptotic distribution of the eigenvalues of $p^{-1}AA^*$. In the context of operator-valued free probability theory, [95] succeeded in characterizing the limiting spectral distribution of block Wishart matrices through a quadratic matrix equation for the corresponding operator-valued Stieltjes transform.

A parallel line of research focuses on the spectral statistics of large symmetric or Hermitian square matrices with dependent entries, thus extending Wigner's [112] seminal result for the i.i.d. case. Models studied in this context include random Toeplitz, Hankel and circulant matrices [21, 27, 80, and references therein] as well as approaches allowing for a more general dependence structure [3, 63].

In Chapter 5, we considered sample covariance matrices of high-dimensional stochastic processes, the components of which are modelled by independent infinite-order moving average processes with identical second-order characteristics. In practice, it is often not possible to observe all components of such a high-dimensional process, and the sample covariance matrix can then not be computed. To solve this problem when only one component is observed, it seems reasonable to partition one long observation record of that observed component of length pn into p segments of length n, and to treat the different segments as if they were records of the unobserved components. We show that this approach is valid and leads to the correct asymptotic eigenvalue distribution of the sample covariance matrix if the components of the underlying process are modelled as independent moving averages.

We are thus led to investigate a model of random matrices **X** whose entries are dependent across both rows and columns, and which is not covered by the results mentioned above. The entries of the random matrix under consideration are defined in terms of a single linear stochastic process, see Section 6.2 for a precise definition. Without assuming Gaussianity we prove almost sure convergence of the empirical spectral distribution of $p^{-1}XX^{T}$ to a deterministic limiting measure and characterize the latter via an integral equation for its Stieltjes transform, which only depends on the asymptotic aspect ratio of the matrix and the second-order properties of the underlying linear process. Our result extends the class of random matrix models for which the limiting spectral distribution can be identified explicitly by a new, theoretically appealing model. It thus contributes to laying the ground for further research into more general random matrix models with dependent, non-identically distributed entries.

Outline In Section 6.2 we give a precise definition of the random matrix model we investigate and state the main result about its limiting spectral distribution. The proof of the main theorem as well as some auxiliary results are presented in Section 6.3. Finally, in Section 6.4, we indicate how our result could be obtained in an alternative way from a similar random matrix model with independent rows.

Notation We use \mathbb{E} and Var to denote expected value and variance. Where convenient, we also write $\mu_{1,X}$ and $\mu_{2,X}$ for the first and second moment, respectively, of a random variable X. The symbol $\mathbf{1}_m$, m a natural number, stands for the $m \times m$ identity matrix. For the trace of a matrix S we write trS. For sequences of matrices $(S_n)_n$ we will suppress the dependence on n where this does not cause ambiguity; the sequence of associated spectral distributions is denoted by F^S , and for their weak limit, provided it exists, we write \hat{F}^S . It will also be convenient to use asymptotic notation: for two sequences of real numbers $(a_n)_n$, $(b_n)_n$ we write $a_n = O(b_n)$ to indicate that there exists a constant C which is independent of n, such that $a_n \leq Cb_n$ for all n. We denote by \mathbb{Z} the set of integers and by \mathbb{N} , \mathbb{R} , and \mathbb{C} the sets of natural, real, and complex numbers, respectively. $\Im z$ stands for the imaginary part of a complex number z, and \mathbb{C}^+ is defined as $\{z \in \mathbb{C} : \Im z > 0\}$. The indicator of an expression \mathcal{E} is denoted by $I_{\{\mathcal{E}\}}$ and defined to be one if \mathcal{E} is true and zero otherwise.

6.2. A new random matrix model

For a sequence $(Z_t)_{t \in \mathbb{Z}}$ of independent real random variables and real coefficients $(c_j)_{j \in \mathbb{N} \cup \{0\}}$, the linear process $(X_t)_{t \in \mathbb{Z}}$ and the $p \times n$ matrix **X** are defined by

$$X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}$$

and

$$\mathbf{X} = (\mathbf{X}_{i,t})_{it} = (X_{(i-1)n+t})_{it} = \begin{pmatrix} X_1 & \dots & X_n \\ X_{n+1} & \dots & X_{2n} \\ \vdots & & \vdots \\ X_{(p-1)n+1} & \dots & X_{pn} \end{pmatrix} \in \mathbb{R}^{p \times n}.$$
(6.2.1)

The interesting feature about this matrix **X** is that its entries are dependent across both rows and columns. In contrast to models considered in [10, 61, 89], not all entries far away from each other are asymptotically independent, e. g., the correlation between the entries $X_{i,n}$ and $X_{i+1,1}$, i = 1, ..., p - 1, does not depend on *n*. We will investigate the asymptotic distribution of the eigenvalues of $p^{-1}XX^{T}$ as both *p* and *n* tend to infinity such that their ratio p/n converges to a finite, positive limit *y*. We assume that the sequence $(Z_t)_t$ satisfies

$$\mathbb{E}Z_t = 0, \quad \mathbb{E}Z_t^2 = 1, \quad \text{and} \quad \sigma_4 \coloneqq \sup_t \mathbb{E}Z_t^4 < \infty,$$
 (6.2.2)

and that the following Lindeberg-type condition is satisfied: for each $\epsilon > 0$,

$$\frac{1}{pn} \sum_{t=1}^{pn} \mathbb{E}\left(Z_t^2 I_{\{Z_t^2 \ge \epsilon n\}}\right) \to 0, \quad \text{as} \quad n \to \infty.$$
(6.2.3)

Condition (6.2.3) is satisfied if all $\{Z_t\}$ are identically distributed, but that is not necessary. As it turns out, the limiting spectral distribution of $p^{-1}\mathbf{X}\mathbf{X}^{\mathsf{T}}$ depends only on y and the second-order structure of the underlying linear process X_t , which we now recall: its auto-covariance function $\gamma : \mathbb{Z} \to \mathbb{R}$ is defined by $\gamma(h) = \mathbb{E}X_0 X_h = \sum_{j=0}^{\infty} c_j c_{j+|h|}$; its spectral density $f : [0, 2\pi] \to \mathbb{R}$ is the Fourier transform of γ , namely $f(\omega) = \sum_{h \in \mathbb{Z}} \gamma(h) e^{-ih\omega}$. The following is the main result of the chapter.

Theorem 6.1. Let $X_t = \sum_{j=0}^{\infty} c_j Z_{t-j}$, $t \in \mathbb{Z}$, be a linear stochastic process with continuously differentiable spectral density f, and let the matrix $\mathbf{X} \in \mathbb{R}^{p \times n}$ be given by Eq. (6.2.1). Assume that

- i) the sequence $(Z_t)_t$ satisfies conditions (6.2.2) and (6.2.3),
- *ii) there exist positive constants* C, δ *such that* $|c_j| \leq C(j+1)^{-1-\delta}$ *, for all* $j \in \mathbb{N} \cup \{0\}$ *.*

Then, as n and p tend to infinity such that the ratio p/n converges to a finite positive limit y, the empirical spectral distribution of $p^{-1}\mathbf{X}\mathbf{X}^{\mathsf{T}}$ converges almost surely to a non-random probability distribution \hat{F} with bounded support. The Stieltjes transform $z \mapsto s_{\hat{F}}(z)$ of \hat{F} is the unique mapping $\mathbb{C}^+ \to \mathbb{C}^+$ satisfying

$$\frac{1}{s_{\hat{F}}(z)} = -z + y \int_0^{2\pi} \frac{f(\omega)}{1 + f(\omega)s_{\hat{F}}(z)} d\omega.$$
(6.2.4)

Remark 6.2.1. The assumption that the coefficients $(c_j)_j$ decay at least polynomially is not very restrictive; it allows, e. g., for X_t to be an ARMA or fractionally integrated ARMA process, which exhibits long-range dependence [56, 66]. In the latter case the entries of the matrix **X** are long-range dependent as well.

Remark 6.2.2. It is possible to generalize the proof of Theorem 6.1 so that the result also holds for non-causal processes, where $X_t = \sum_{j=-\infty}^{\infty} c_j Z_{t-j}$. The required changes are merely notational, the only difference in the result is that the auto-covariance function is then given by $\sum_{j=-\infty}^{\infty} c_j c_{j+|h|}$.

The distribution \hat{F} can be obtained from $s_{\hat{F}}$ via the Perron–Frobenius inversion formula [9, Theorem B.8], which states that for all continuity point 0 < a < b of \hat{F} , it holds that $\hat{F}([a,b]) = \lim_{\epsilon \to 0^+} \int_a^b \Im s_{\hat{F}}(x + \epsilon i) dx$. In general, the analytic determination of this distribution is not feasible. It is, however, easy to check that for the special case of independent entries one recovers the classical Marčenko–Pastur law.

6.3. Proof of Theorem 6.1

The strategy in the proof of Theorem 6.1 is to show that the limiting spectral distribution of $p^{-1}\mathbf{X}\mathbf{X}^{\mathsf{T}}$ is stable under modifications of **X** which reduce the sample covariance matrix to the

6. A new random matrix model with dependent rows and columns

form $p^{-1}\mathbf{Z}H\mathbf{Z}^{\mathsf{T}}$, for a matrix \mathbf{Z} with i. i. d. entries, and some positive definite H. To this end we will repeatedly use the following lemma which presents sufficient conditions for the limiting spectral distributions of two sequences of matrices to be equal.

Lemma 6.3.1 (Trace criterion). Let $A_{1,n}$, $A_{2,n}$ be sequences of $p \times n$ matrices, where $p = p_n$ depends on *n* such that $p_n \to \infty$ as $n \to \infty$. Assume that the spectral distribution $F^{p^{-1}A_{1,n}A_{1,n}^{\mathsf{T}}}$ converges almost surely to a deterministic limit $\hat{F}^{p^{-1}A_{1,n}A_{1,n}^{\mathsf{T}}}$ as *n* tends to infinity. If there exists a positive number ϵ such that

i)
$$p^{-4}\mathbb{E}\left[\operatorname{tr}(A_{1,n}-A_{2,n})(A_{1,n}-A_{2,n})^{\mathsf{T}}\right]^{2} = O(n^{-1-\epsilon}),$$

ii)
$$p^{-2}\mathbb{E}\text{tr}A_{i,n}A_{i,n}^{\mathsf{T}} = O(1), i = 1, 2, \text{ and }$$

iii) p^{-4} Vartr $A_{i,n}A_{i,n}^{\mathsf{T}} = O(n^{-1-\epsilon}), i = 1, 2,$

then the spectral distribution of $p^{-1}A_{2,n}A_{2,n}^{\mathsf{T}}$ is convergent almost surely with the same limit $\hat{F}^{p^{-1}A_{1,n}A_{1,n}^{\mathsf{T}}}$.

Proof. The claim is a direct consequence of Chebyshev's inequality, the first Borel–Cantelli lemma, and [9, Corollary A.42]

With the constants C and δ from assumption ii of Theorem 6.1 we define

$$\overline{c}_j \coloneqq C(j+1)^{-1-\delta},$$

such that $|c_j| \leq \overline{c}_j$ for all j. Without further reference we will repeatedly use the fact that $j \mapsto \overline{c}_j$ is monotone, that $\sum_{j=1}^{\infty} \overline{c}_j^{\alpha}$ is finite for every $\alpha \geq 1$, and that $\sum_{j=n}^{\infty} \overline{c}_j^{\alpha}$ is of order $O(n^{1-\alpha(1+\delta)})$. Since it is difficult to deal with infinite-order moving averages processes directly, it is convenient to truncate the entries of the matrix **X** by defining $\widetilde{X}_t = \sum_{j=0}^n c_j Z_{t-j}$ and $\widetilde{\mathbf{X}} = (\widetilde{X}_{(i-1)n+t})_{it}$; this is different from the usual truncation of the support of the entries of a random matrix.

Proposition 6.3.1 (Truncation). *If the empirical spectral distribution of* $p^{-1}\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathsf{T}}$ *converges to a limit, then the empirical spectral distribution of* $p^{-1}\mathbf{X}\mathbf{X}^{\mathsf{T}}$ *converges to the same limit.*

Proof. The proof proceeds in two steps in which we verify conditions i to iii of Lemma 6.3.1.

Step 1 The definitions of X and \widetilde{X} imply that

$$\Delta_{\mathbf{X},\widetilde{\mathbf{X}}} \coloneqq \frac{1}{p^2} \operatorname{tr} \left(\mathbf{X} - \widetilde{\mathbf{X}} \right) \left(\mathbf{X} - \widetilde{\mathbf{X}} \right)^{\mathsf{T}} = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{t=1}^{n} \sum_{i=1}^{n} \left[\mathbf{X}_{it} - \widetilde{\mathbf{X}}_{it} \right]^2$$
$$= \frac{1}{p^2} \sum_{i,t=1}^{p,n} \sum_{k,k'=n+1}^{\infty} Z_{(i-1)n+t-k} Z_{(i-1)n+t-k'} c_k c_{k'}.$$

We shall show that the second moment of $\Delta_{\mathbf{X},\mathbf{X}}$ is of order at most $n^{-2-2\delta}$. Since

$$\sum_{\substack{k,k'\\m,m'}=n+1}^{\infty} \mathbb{E} \left| Z_{(i-1)n+t-k} Z_{(i-1)n+t-k'} Z_{(i'-1)n+t'-m} Z_{(i'-1)n+t'-m'} \right| |c_k||c_{k'}||c_m||c_{m'}|$$

$$\leq \sigma_4 \left[\sum_{k=0}^{\infty} |c_k| \right]^4 < \infty, \tag{6.3.1}$$

we can apply Fubini's theorem to interchange expectation and summation in the computation of

$$\mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}} \coloneqq \mathbb{E}\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}^{2}$$

$$= \frac{1}{p^{4}} \sum_{\substack{i,i' \\ t,t'}=1}^{p,n} \sum_{\substack{k,k' \\ m,m'}=n+1}^{\infty} \mathbb{E}\left[Z_{(i-1)n+t-k}Z_{(i-1)n+t-k'}Z_{(i'-1)n+t'-m}Z_{(i'-1)n+t'-m'}\right] c_{k}c_{k'}c_{m}c_{m'}.$$
(6.3.2)

Since the $\{Z_t\}$ are independent, the expectation in that sum is non-zero only if all four Z are the same or else one can match the indices in two pairs. In the latter case we distinguish three cases according to which factor the first Z is paired with. This leads to the additive decomposition

$$\mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}} = \mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{\Pi\Pi} + \mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{\Pi\Pi} + \mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{\Pi\Pi} + \mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{\Pi\Pi}, \tag{6.3.3}$$

where the ideograms indicate which of the four factors are equal. For the contribution from all four Z being equal it holds that k = k', m = m', and (i - 1)n + t - k = (i' - 1)n + t' - m, so that

$$\mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{\Pi\Pi} = \frac{\sigma_4}{p^4} \sum_{i,i'}^p \sum_{t,t'=1}^n \sum_{m=\max\{n+1,n+1-(i-i')n-(t-t')\}}^\infty c_{(i-i')n+(t-t')+m}^2 c_m^2$$

If we introduce the new summation variables $\delta_i := i - i'$ and $\delta_t := t - t'$, we obtain

$$\mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{\Pi\Pi} = \frac{\sigma_4}{p^4} \sum_{\delta_i=1-p}^{p-1} \underbrace{(p-|\delta_i|)}_{\leqslant p} \sum_{\delta_t=1-n}^{n-1} \underbrace{(n-|\delta_t|)}_{\leqslant n} \sum_{m=\max\{n+1,n+1-\delta_i n-\delta_t\}}^{\infty} c_{m+\delta_i n+\delta_t}^2 c_m^2.$$

If δ_i is positive, then $\delta_i n + \delta_t$ is positive as well; the fact that $|c_j|$ is bounded by \overline{c}_j and the monotonicity of $j \mapsto \overline{c}_j$ imply that $c_{m+\delta_i n+\delta_t}^2 \leq \overline{c}_{(\delta_i-1)n}\overline{c}_{\delta_t+n}$ so that the contribution from $\delta_i \geq 1$ can be estimated as

$$\mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{\Pi\Pi,+} \leq \underbrace{\frac{\sigma_4 n}{p^3}}_{=O(n^{-2})} \underbrace{\sum_{\delta_i=1}^{p-1} \overline{c}_{(\delta_i-1)n}}_{=O(n^{-1-\delta})} \underbrace{\sum_{\delta_t=1}^{2n-1} \overline{c}_{\delta_t}}_{=O(1)} \underbrace{\sum_{m=n+1}^{\infty} \overline{c}_m^2}_{=O(n^{-1-2\delta})} = O(n^{-4-3\delta}).$$

An analogous argument shows that the contribution from $\delta_i \leq -1$, denoted by $\mu_{2,\Delta_{X,\widetilde{X}}}^{\Pi\Pi,-}$, is of the same order of magnitude. The contribution to $\mu_{2,\Delta_{X,\widetilde{X}}}^{\Pi\Pi}$ from $\delta_i = 0$ is given by

$$\mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{\Pi\Pi,\varnothing} = \frac{\sigma_4 n}{p^3} \sum_{\delta_t=1-n}^{n-1} \sum_{\substack{m=\max\{n+1,n+1-\delta_t\}\\ 0 < m}}^{\infty} c_m^2 c_{m+\delta_t}^2 \\ \leqslant \underbrace{\frac{\sigma_4 n}{p^3}}_{=O(n^{-2})} \left[2 \underbrace{\sum_{\substack{\delta_t=1\\ 0 < m}}^{n-1} \overline{c}_{\delta}^2}_{=O(1)} \underbrace{\sum_{\substack{m=n+1\\ 0 < m}}^{\infty} \overline{c}_{m}^2}_{=O(n^{-1-2\delta})} + \underbrace{\sum_{\substack{m=n+1\\ 0 < m}}^{\infty} \overline{c}_{m}^4}_{=O(n^{-3-2\delta})} \right] = O(n^{-3-2\delta}).$$

By combining the last two displays, it follows that $\mu_{2,\Delta_{X,\tilde{X}}}^{\Pi\Pi}$ is of order $O(n^{-3-2\delta})$. The second term in Eq. (6.3.3) corresponds to k = k', m = m', and $(i-1)n + t - k \neq (i'-1)n + t' - m$. The restriction that not all four factors be equal is taken into account by subtracting $\mu_{2,\Delta_{X,\tilde{X}}}^{\Pi\Pi}$; consequently,

$$\mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{\Pi \Pi} = \underbrace{\frac{1}{p^4} \sum_{i,i'=1}^{p} \sum_{t,t'=1}^{n} \sum_{\substack{k,m=n+1\\ =O(1)}}^{\infty} \sum_{\substack{k,m=n+1\\ =O(n^{-2-4\delta})}}^{\infty} c_k^2 c_m^2 - \mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{\Pi \Pi} = O(n^{-2-4\delta}).$$

It remains to analyse $\mu_{2,\Delta_{\mathbf{X},\mathbf{X}}}^{[\Box]}$ which, by symmetry, is equal to $\mu_{2,\Delta_{\mathbf{X},\mathbf{X}}}^{[\Box]}$. If the first factor is paired with the third, the condition for non-vanishment becomes k = m + (i - i')n + t - t', k' = m' + (i - i')n + t - t', and $m \neq m'$. Again introducing the new summation variables $\delta_i := i - i'$ and $\delta_t := t - t'$, we obtain that

$$\mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{[\Box]} = \frac{1}{p^4} \sum_{\delta_i=1-p}^{p-1} \underbrace{(p-|\delta_i|)}_{\leqslant p} \sum_{\delta_t=1-n}^{n-1} \underbrace{(n-|\delta_t|)}_{\leqslant n} \sum_{m,m'} c_m c_{m'} c_{m+\delta_i n+\delta_t} c_{m'+\delta_i n+\delta_t} - \mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{[\Box]}$$

where the summation of $\sum_{m,m'}$ is over the set

$$\left\{ (m, m') \in \mathbb{N}^2 : \max\{n+1, n+1 - \delta_i n - \delta_t\} \le m, m' \le \infty \right\}$$

As in the analysis of $\mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{\Pi\Pi}$ we obtain the contribution from $\delta_i \neq 0$ as

$$\left| \mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{(\Box),+} \right| = \left| \mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{(\Box),-} \right| \leq \underbrace{\frac{n}{p^3}}_{=O(n^{-2})} \underbrace{\sum_{\delta_i=1}^{p-1} \overline{c}_{(\delta_i-1)n}}_{=O(n^{-1-\delta})} \underbrace{\sum_{\delta_i=1}^{2n-1} \overline{c}_{\delta_i}}_{=O(1)} \underbrace{\sum_{m,m'=n+1}^{\infty} \overline{c}_m \overline{c}_{m'}}_{=O(n^{-2\delta})} + \mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{(\Box)} = O(n^{-3-2\delta}).$$
(6.3.4)

Finally, for the contribution from $\delta_i = 0$ one finds that

$$\begin{aligned} \left| \mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{\Pi,\emptyset} \right| &\leq \frac{n}{p^3} \sum_{\delta_t=1-n}^{n-1} \sum_{m,m'=\max\{n+1,n+1-\delta_t\}}^{\infty} |c_m c_{m'} c_{m+\delta_t} c_{m'+\delta_t}| + \mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{\Pi\Pi} \\ &\leq \underbrace{\frac{n}{p^3}}_{=O(n^{-2})} \left[2 \sum_{\substack{\delta_t=1\\ =O(1)}}^{n-1} \overline{c}_{\delta_t}^2 \sum_{\substack{m,n'=n+1\\ =O(n^{-2\delta})}}^{\infty} \overline{c}_m \overline{c}_{m'} + \underbrace{\sum_{m,n'=n+1}}^{\infty} \overline{c}_m^2 \overline{c}_{m'}^2 - O(n^{-2-2\delta}) \right] + \mu_{2,\Delta}^{\Pi\Pi} = O(n^{-2-2\delta}). \quad (6.3.5)$$

The last two displays (6.3.4) and (6.3.5) imply that

$$\mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{[\Box]} = \mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{[\Box],-} + \mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{[\Box],+} + \mu_{2,\Delta_{\mathbf{X},\widetilde{\mathbf{X}}}}^{[\Box],+} = O(n^{-2-2\delta}).$$

Thus, $\mu_{2,\Delta_{\mathbf{X}\widetilde{\mathbf{X}}}}$ is of order $O(n^{-2-2\delta})$, as claimed.

Step 2 Next we verify assumptions ii and iii of Lemma 6.3.1, which means that we show that both $\Sigma_{\mathbf{X}} := p^{-2} \text{tr} \mathbf{X} \mathbf{X}^{\mathsf{T}}$ and $\Sigma_{\mathbf{X}} := p^{-2} \text{tr} \mathbf{X} \mathbf{X}^{\mathsf{T}}$ have bounded first moments and variances of order $n^{-1-\epsilon}$, for some $\epsilon > 0$; in fact, ϵ will turn out to be one. For $\Sigma_{\mathbf{X}}$ we obtain

$$\mu_{1,\Sigma_{\mathbf{X}}} \coloneqq \mathbb{E}\Sigma_{\mathbf{X}} = \frac{1}{p^2} \sum_{i=1}^{p} \sum_{t=1}^{n} \sum_{k,k'=0}^{\infty} \mathbb{E}\left[Z_{(i-1)n+t-k}Z_{(i-1)n+t-k'}\right] c_k c_{k'} = \frac{n}{p} \sum_{k=0}^{\infty} c_k^2,$$

where the change of the order of expectation and summation is valid by Fubini's theorem. Using Eq. (6.3.1) and Fubini's theorem, the second moment of $\Sigma_{\mathbf{X}}$ becomes

$$\begin{split} \mu_{2,\Sigma_{\mathbf{X}}} &\coloneqq \mathbb{E}\Sigma_{\mathbf{X}}^{2} \\ &= \frac{1}{p^{4}} \sum_{i,i'=1}^{p} \sum_{t,t'=1}^{n} \sum_{\substack{k,k'\\m,m'}=0}^{\infty} \mathbb{E}\left[Z_{(i-1)n+t-k} Z_{(i-1)n+t-k'} Z_{(i'-1)n+t'-m} Z_{(i'-1)n+t'-m'} \right] c_{k} c_{k'} c_{m} c_{m'}. \end{split}$$

This sum coincides with the expression analysed in Eq. (6.3.2), except that here the k, k', m, m' sums start at zero, and not at n + 1. A straightforward adaptation of the arguments there show that $\mu_{2,\Sigma_{\mathbf{X}}}$ equals $n^2 p^{-2} \left(\sum_{k=0}^{\infty} c_k^2 \right)^2 + O(n^{-2})$, and, consequently, that $\operatorname{Var}\Sigma_{\mathbf{X}} = \mu_{2,\Sigma_{\mathbf{X}}} - (\mu_{1,\Sigma_{\mathbf{X}}})^2 = O(n^{-2})$. Analogous computations show that $\mathbb{E}\Sigma_{\mathbf{X}}$ is bounded, and that $\operatorname{Var}\Sigma_{\mathbf{X}} = O(n^{-2})$. Thus, conditions ii and iii of Lemma 6.3.1 are verified, and the proof of the proposition is complete.

Because of Proposition 6.3.1 the problem of determining the limiting spectral distribution of the sample covariance matrix $p^{-1}\mathbf{X}\mathbf{X}^T$ has been reduced to computing the limiting spectral distribution of $p^{-1}\mathbf{\widetilde{X}}\mathbf{\widetilde{X}}^T$, where now, for fixed *n*, the matrix $\mathbf{\widetilde{X}}$ depends on only finitely many of

the noise variables Z_t . The fact that the entries of $\widetilde{\mathbf{X}}$ are finite-order moving average processes and therefore linearly dependent on the Z_t allows for $\widetilde{\mathbf{X}}$ to be written as a linear transformation of the i. i. d. matrix $\mathbf{Z} \coloneqq (Z_{(i-2)n+t})_{i=1,\dots,p+1,t=1,\dots,n}$. We emphasize that \mathbf{Z} , in contrast to \mathbf{X} and $\widetilde{\mathbf{X}}$, is a $(p+1) \times n$ matrix; this is necessary because the entries in the first row of $\widetilde{\mathbf{X}}$ depend on noise variables with negative indices, up to and including Z_{1-n} . In order to formulate the transformation that maps \mathbf{Z} to $\widetilde{\mathbf{X}}$ concisely in the next lemma, we define the matrices $K_n = \begin{pmatrix} 0 & 0 \\ \mathbf{1}_{n-1} & 0 \end{pmatrix} \in \mathbb{R}^{n \times n}$, as well as the polynomials $\chi_n(z) = c_0 + c_1 z + \ldots + c_n z^n$ and $\overline{\chi}_n(z) = z^n \chi (1/z) = c_n + c_{n-1} z + \ldots + c_0 z^n$.

Lemma 6.3.2. With $\widetilde{\mathbf{X}}$, \mathbf{Z} , K_n and χ_n , $\overline{\chi}_n$ defined as before it holds that

$$\widetilde{\mathbf{X}} = \begin{bmatrix} 0 & \mathbf{1}_p & \mathbf{1}_p & 0 \end{bmatrix} \begin{pmatrix} \mathbf{Z} & 0 \\ 0 & \mathbf{Z} \end{pmatrix} \begin{bmatrix} \chi_n \begin{pmatrix} K_n^{\mathsf{T}} \end{pmatrix} \\ \bar{\chi}_n \begin{pmatrix} K_n \end{pmatrix} \end{bmatrix}.$$
(6.3.6)

Proof. Let $s_N : \mathbb{R}^N \to \mathbb{R}^N$ be the right shift operator defined by $s_N(v_1, \ldots, v_N) = (0, v_1, \ldots, v_{N-1})$ and for positive integers r, s denote by $\operatorname{vec}_{r,s} : \mathbb{R}^{r \times s} \to \mathbb{R}^{rs}$ the bijective linear operator that transforms a matrix into a vector by horizontally concatenating its subsequent rows, starting with the first one. The operator $S_{r,s} : \mathbb{R}^{r \times s} \to \mathbb{R}^{r \times s}$ is then defined as $S_{r,s} =$ $\operatorname{vec}_{r,s}^{-1} \circ s_{rs} \circ \operatorname{vec}_{r,s}$. This operator shifts all entries of a matrix to the right except for the entries in the last column, which are shifted down and moved into the first column. For $k = 1, 2, \ldots$, the operator $S_{r,s}^k$ is defined as the k-fold composition of $S_{r,s}$. In the following, we write $S := S_{p+1,n}$. With this notation it is clear that $\widetilde{\mathbf{X}} = \begin{bmatrix} 0 & \mathbf{1}_p \\ 2n(S)\mathbf{Z} \end{bmatrix} \chi_n(S)\mathbf{Z}$. In order to obtain Eq. (6.3.6), we observe that the action of S can be written in terms of matrix multiplications as $S\mathbf{Z} = K_{p+1}\mathbf{Z}E + \mathbf{Z}K_n^T$, where the entries of the $n \times n$ matrix E are all zero except for a one in the lower left corner. Using the fact that $E(K_n^T)^m E$ is zero for every non-negative integer m it follows by induction that S^k , $k = 1, \ldots, n$, acts like

$$S^{k}\mathbf{Z} = \mathbf{Z}(K_{n}^{\mathsf{T}})^{k} + K_{p+1}\mathbf{Z}\sum_{i=1}^{k} (K_{n}^{\mathsf{T}})^{k-i} E(K_{n}^{\mathsf{T}})^{i-1} = \begin{bmatrix} \mathbf{1}_{p+1} & K_{p+1} \end{bmatrix} \begin{pmatrix} \mathbf{Z} & 0 \\ 0 & \mathbf{Z} \end{pmatrix} \begin{bmatrix} (K_{n}^{\mathsf{T}})^{k} \\ K_{n}^{n-k} \end{bmatrix}.$$

This implies that

$$\widetilde{\mathbf{X}} = \begin{bmatrix} 0 & \mathbf{1}_p \end{bmatrix} \begin{bmatrix} \mathbf{1}_{p+1} & K_{p+1} \end{bmatrix} \begin{pmatrix} \mathbf{Z} & 0 \\ 0 & \mathbf{Z} \end{pmatrix} \sum_{k=0}^n c_k \begin{bmatrix} \begin{pmatrix} K_n^\mathsf{T} \end{pmatrix}^k \\ K_n^{n-k} \end{bmatrix}$$
$$= \begin{bmatrix} 0 & \mathbf{1}_p & \mathbf{1}_p & 0 \end{bmatrix} \begin{pmatrix} \mathbf{Z} & 0 \\ 0 & \mathbf{Z} \end{pmatrix} \begin{bmatrix} \chi_n \begin{pmatrix} K_n^\mathsf{T} \end{pmatrix} \\ \bar{\chi}_n \begin{pmatrix} K_n \end{pmatrix} \end{bmatrix}$$

and completes the proof.

While the last lemma gives an explicit description of the relation between \mathbf{Z} and $\widetilde{\mathbf{X}}$, it is impractical for directly determining the limiting spectral distribution of $p^{-1}\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathsf{T}}$. The reason

is that **Z** appears twice in the central block-diagonal matrix and is moreover multiplied by some deterministic matrices from both the left and the right. The LSD of the product of three random matrices has been computed in the literature [115], but this result is not applicable in our situation due to the appearance of the random block matrix in Eq. (6.3.6). Sample covariance matrices derived from random block matrices have been considered in [95]. However, they only treat the Gaussian case and, more importantly, do not cover the case of a non-trivial population covariance matrix. We are thus not aware of any result allowing to derive the LSD of $p^{-1}\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathsf{T}}$ directly from Lemma 6.3.2.

The next proposition allows us to circumvent this problem. It is shown that, at least asymptotically and at the cost of slightly changing the size of the involved matrices, one can simplify the structure of $\widetilde{\mathbf{X}}$ so that \mathbf{Z} appears only once and is multiplied by a deterministic matrix only from the right.

Proposition 6.3.2. Let \mathbf{Z} , K_n and χ_n , $\bar{\chi}_n$ be as before and define the matrix $\widehat{\mathbf{X}} := \mathbf{Z}\Omega \in \mathbb{R}^{(p+1)\times(n+1)}$, where

$$\Omega = \begin{bmatrix} 0 & \mathbf{1}_n & \mathbf{1}_n & 0 \end{bmatrix} \begin{bmatrix} \chi_{n+1} \begin{pmatrix} K_{n+1} \\ \bar{\chi}_{n+1} & (K_{n+1}) \end{bmatrix} \in \mathbb{R}^{n \times (n+1)}.$$
(6.3.7)

If the empirical spectral distribution of $p^{-1}\widehat{\mathbf{X}}\widehat{\mathbf{X}}^{\mathsf{T}}$ converges to a limit, then the empirical spectral distribution of $p^{-1}\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathsf{T}}$ converges to the same limit.

Proof. In order to be able to compare the limiting spectral distributions of $p^{-1}\widetilde{\mathbf{X}}\widetilde{\mathbf{X}}^{\mathsf{T}}$ and $p^{-1}\widehat{\mathbf{X}}\widehat{\mathbf{X}}^{\mathsf{T}}$ in spite of their dimensions being different, we introduce the matrix

$$\overline{\mathbf{X}} = \begin{bmatrix} 0 & 0 \\ 0 & \widetilde{\mathbf{X}} \end{bmatrix} \in \mathbb{R}^{(p+1) \times (n+1)}$$

Clearly, $F^{p^{-1}\overline{\mathbf{X}\mathbf{X}}^{\mathsf{T}}} = (p+1)^{-1}\delta_0 + p(p+1)^{-1}F^{p^{-1}\overline{\mathbf{X}}\overline{\mathbf{X}}^{\mathsf{T}}}$, which implies equality of the limiting spectral distributions provided either of the two, and hence both, exists. It is therefore sufficient to show that the LSD of $p^{-1}\widehat{\mathbf{X}}\widehat{\mathbf{X}}^{\mathsf{T}}$ and $p^{-1}\overline{\mathbf{X}}\overline{\mathbf{X}}^{\mathsf{T}}$ are identical; this will be done by verifying the three conditions of Lemma 6.3.1. The remainder of the proof will be divided in two parts. In the first part we check the validity of assumption i about the difference $\widehat{\mathbf{X}} - \overline{\mathbf{X}}$, whereas in the second one we consider the terms $\operatorname{tr}\widehat{\mathbf{X}}\widehat{\mathbf{X}}^{\mathsf{T}}$ and $\operatorname{tr}\overline{\mathbf{X}}\overline{\mathbf{X}}^{\mathsf{T}}$, which appear in conditions ii and iii.

6. A new random matrix model with dependent rows and columns

Step 1 Using the definitions of $\widehat{\mathbf{X}}$ and $\overline{\mathbf{X}}$, it follows that

$$\begin{split} \Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}} &\coloneqq \frac{1}{p^2} \operatorname{tr} \left(\widehat{\mathbf{X}} - \overline{\mathbf{X}} \right) \left(\widehat{\mathbf{X}} - \overline{\mathbf{X}} \right)^{\mathsf{T}} \\ &= \frac{1}{p^2} \sum_{i=1}^{p+1} \sum_{j=1}^{n+1} \left[\widehat{\mathbf{X}}_{ij} - \overline{\mathbf{X}}_{ij} \right]^2 \\ &\leqslant \frac{2}{p^2} \sum_{i=2}^{p+1} \sum_{j=2}^{n+1} \sum_{k,k'=j}^{n} Z_{(i-2)n+k} Z_{(i-2)n+k'} c_{j-k+n+1} c_{j-k'+n+1} \\ &+ \frac{2}{p^2} \sum_{i=2}^{p+1} \sum_{j=2}^{n+1} \sum_{k,k'=j-1}^{n} Z_{(i-3)n+k} Z_{(i-3)n+k'} c_{j-k+n-1} c_{j-k'+n-1} \\ &+ \frac{1}{p^2} \sum_{i=1}^{p+1} \sum_{k,k'=1}^{n} Z_{(i-2)n+k} Z_{(i-2)n+k'} c_{n-k+2} c_{n-k'+2} \\ &+ \frac{2}{p^2} \sum_{j=2}^{n+1} \sum_{k,k'=1}^{j-1} Z_{-n+k} Z_{-n+k'} c_{j-k-1} c_{j-k'-1} \\ &+ \frac{2}{p^2} \sum_{j=2}^{n+1} \sum_{k,k'=j}^{n} Z_{-n+k} Z_{-n+k'} c_{j-k+n+1} c_{j-k'+n+1} \\ &=: \sum_{i=1}^{5} \Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(i)}, \end{split}$$
(6.3.8)

where the elementary inequality $(a + b)^2 \le 2a^2 + 2b^2$ was used twice. In order to show that the variances of expression (6.3.8) are summable, we consider each term in turn. For the second moment of the first term of Eq. (6.3.8) we obtain

$$\begin{split} \mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(1)}} &\coloneqq \mathbb{E}\left(\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(1)}\right)^2 \\ &= \frac{4}{p^4} \sum_{i,i'=2}^{p+1} \sum_{j,j'=2}^{n+1} \sum_{k,k'=1}^{n-j+1} \sum_{m,m'=1}^{n-j'+1} \mathbb{E}\left[Z_{(i-1)n-k+1} Z_{(i-1)n-k'+1} Z_{(i'-1)n-m+1} Z_{(i'-1)n-m'+1}\right] \\ &\times c_{j+k} c_{j+k'} c_{j'+m} c_{j'+m'}. \end{split}$$

As before we consider all configurations where above expectation is not zero. The expectation equals σ_4 if i = i' and k, k', m, m' are equal, and, hence,

$$\mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(1)}}^{\Pi\Pi} \leqslant \frac{4\sigma_4}{p^4} \sum_{i=2}^{p+1} \sum_{k=1}^n \left(\sum_{j=2}^{n+1} c_{j+k}^2 \right)^2 \leqslant \frac{4\sigma_4}{p^3} \sum_{k=1}^n \overline{c}_k^2 \left(\sum_{j=2}^{n+1} \overline{c}_j \right)^2 = O(n^{-3}).$$

The expectation is one if the four Z can be collected in two non-equal pairs. The first term equals the second, and the third equals the fourth if k = k' and m = m', and thus

$$\mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(1)}}^{\Pi \Pi} = \frac{4}{p^4} \sum_{i,i'=2}^{p+1} \sum_{j,j'=2}^{n+1} \sum_{k=1}^{n-j+1} \sum_{m=1}^{n-j+1} c_{j+k}^2 c_{j'+m}^2 - \mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(1)}}^{\Pi \Pi}$$
$$= \frac{4}{p^2} \left(\sum_{j=2}^{n+1} \sum_{k=1}^{n-j+1} c_{j+k}^2 \right)^2 - \mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(1)}}^{\Pi \Pi} = O(n^{-2}).$$

Likewise, the contribution from pairing the first factor with the third, and the second with the fourth, can be estimated as

$$\begin{split} \left| \mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(1)}} \right| &\leq \frac{4}{p^4} \sum_{i'=2}^{p+1} \sum_{j,j'=2}^{n+1} \sum_{k,k'=1}^{n} |c_{j+k}c_{j+k'}c_{j'+k}c_{j'+k'}| + \mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(1)}}^{\Pi\Pi} \\ &\leq \frac{4}{p^3} \left(\sum_{j=1}^{n+1} \overline{c}_j \right)^4 + \mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(1)}}^{\Pi\Pi} = O(n^{-3}). \end{split}$$

Obviously, the configuration $\mu_{2,\Delta_{\widehat{\mathbf{X}}\overline{\mathbf{X}}}^{(1)}}$ can be handled the same way as $\mu_{2,\Delta_{\widehat{\mathbf{X}}\overline{\mathbf{X}}}^{(1)}}$ above. Thus we have shown that the second moment of $\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(1)}$, the first term in Eq. (6.3.8), is of order n^{-2} . This can be shown for the second term in Eq. (6.3.8) in the same way. We now consider the second moment of the third term in Eq. (6.3.8):

$$\mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(3)}} := \mathbb{E} \left(\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(3)} \right)^{2}$$

$$= \frac{1}{p^{4}} \sum_{i,i'=1}^{p+1} \sum_{\substack{k,k'\\m,m'}=1}^{n} \mathbb{E} \left[Z_{(i-2)n+k} Z_{(i-2)n+k'} Z_{(i'-2)n+m} Z_{(i'-2)n+m'} \right]$$

$$\times c_{n-k+2} c_{n-k'+2} c_{n-m+2} c_{n-m'+2}.$$

Distinguishing the same cases as before, we have $\mu_{2,\Delta_{\widehat{\mathbf{X}}\overline{\mathbf{X}}}}^{\Pi\Pi} = \sigma_4 \frac{p+1}{p^4} \sum_{k=1}^n c_{n-k+2}^4 = O(n^{-3})$ and, thus,

$$\mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(3)}}^{\Pi \Pi} = \frac{(p+1)^2}{p^4} \left(\sum_{k=1}^n c_{n-k+2}^2 \right)^2 - \mu_{2,\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(3)}}^{\Pi \Pi} = O(n^{-2}),$$

as well as $\mu_{2,\Delta_{\widehat{X}\overline{X}}^{(3)}}^{(7)} = \mu_{2,\Delta_{\widehat{X}\overline{X}}^{(3)}}^{(7)} = O(n^{-3})$. Thus, the second moment of the third term in Eq. (6.3.8) is of order $O(n^{-2})$; repeating the foregoing arguments, it can be seen that the

second moments of $\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(4)}$ and $\Delta_{\widehat{\mathbf{X}},\overline{\mathbf{X}}}^{(5)}$, the two last terms in Eq. (6.3.8), are of order $O(n^{-2})$ as well, so that we have shown that

$$\frac{1}{p^4} \mathbb{E}\left[\operatorname{tr}\left(\widehat{\mathbf{X}} - \overline{\mathbf{X}}\right) \left(\widehat{\mathbf{X}} - \overline{\mathbf{X}}\right)^{\mathsf{T}} \right]^2 = \mathbb{E}\left(\Delta_{\widehat{\mathbf{X}}, \overline{\mathbf{X}}} \right)^2 \leq 5 \sum_{i=1}^5 \mu_{2, \Delta_{\widehat{\mathbf{X}}, \overline{\mathbf{X}}}^{(i)}} = O(n^{-2}).$$

Step 2 In this step we shall prove that both $\Sigma_{\widehat{\mathbf{X}}} \coloneqq p^{-2} \operatorname{tr} \widehat{\mathbf{X}} \widehat{\mathbf{X}}^{\mathsf{T}}$ and $\Sigma_{\overline{\mathbf{X}}} \coloneqq p^{-2} \operatorname{tr} \overline{\mathbf{X}} \overline{\mathbf{X}}^{\mathsf{T}}$ have bounded first moments, and that their variances are summable sequences in *n*, i. e. we check conditions (ii) and (iii) of Lemma 6.3.1. Since $\operatorname{tr} \overline{\mathbf{X}} \overline{\mathbf{X}}^{\mathsf{T}}$ is equal to $\operatorname{tr} \widetilde{\mathbf{X}} \widetilde{\mathbf{X}}^{\mathsf{T}}$, the claim about $\Sigma_{\overline{\mathbf{X}}}$ has already been shown in the second step of the proof of Proposition 6.3.1. For the first term one finds, by the definition of $\widehat{\mathbf{X}}$, that

$$\begin{split} \Sigma_{\widehat{\mathbf{X}}} &= \frac{1}{p^2} \sum_{i=1}^{p+1} \sum_{j=1}^{n+1} \left(\sum_{k=1}^{j-1} Z_{(i-2)n+k} c_{j-k-1} + \sum_{k=j}^{n} Z_{(i-2)n+k} c_{j-k+n+1} \right)^2 \\ &\leqslant \frac{2}{p^2} \sum_{i=1}^{p+1} \sum_{j=1}^{n+1} \sum_{k,k'=1}^{j-1} Z_{(i-2)n+k} c_{j-k-1} Z_{(i-2)n+k'} c_{j-k'-1} \\ &+ \frac{2}{p^2} \sum_{i=1}^{p+1} \sum_{j=1}^{n+1} \sum_{k,k'=j}^{n} Z_{(i-2)n+k} c_{j-k+n+1} Z_{(i-2)n+k'} c_{j-k'+n+1} =: \Sigma_{\widehat{\mathbf{X}}}^{(1)} + \Sigma_{\widehat{\mathbf{X}}}^{(2)}. \end{split}$$

Clearly, the first two moments of $\Sigma^{(1)}_{\widehat{\mathbf{X}}}$ are given by

$$\begin{split} \mu_{1,\Sigma_{\widehat{\mathbf{X}}}^{(1)}} &\coloneqq \mathbb{E}\Sigma_{\widehat{\mathbf{X}}}^{(1)} = \frac{2}{p^2} \sum_{i=1}^{p+1} \sum_{j=1}^{n+1} \sum_{k,k'=1}^{j-1} \mathbb{E}\left[Z_{(i-2)n+k} Z_{(i-2)n+k'}\right] c_{j-k-1} c_{j-k'-1} \\ &= \frac{2(p+1)}{p^2} \sum_{j=1}^{n+1} \sum_{k=1}^{j-1} c_{k-1}^2, \end{split}$$

and

$$\begin{split} \mu_{2,\Sigma_{\widehat{\mathbf{X}}}^{(1)}} &\coloneqq \mathbb{E}\left(\Sigma_{\widehat{\mathbf{X}}}^{(1)}\right)^{2} \\ &= \frac{4}{p^{4}} \sum_{i,i'=1}^{p+1} \sum_{j,j'=1}^{n+1} \sum_{k,k'=1}^{j-1} \sum_{m,m'=1}^{j'-1} \mathbb{E}(Z_{(i-2)n+k} Z_{(i-2)n+k'} Z_{(i'-2)n+m} Z_{(i'-2)n+m'}) \\ &\times c_{j-k-1} c_{j-k'-1} c_{j'-m-1} c_{j'-m'-1}. \end{split}$$

We separately consider the case that all four factors are equal, and the three possible pairings of the four Z. If all four Z are equal, it must hold that i = i', k = k' = m = m', with

contribution

$$\mu_{2,\Sigma_{\widehat{\mathbf{X}}}^{(1)}}^{\Pi\Pi} = \frac{4\sigma_4}{p^4} \sum_{i=1}^{p+1} \sum_{j,j'=1}^{n+1} \sum_{k=1}^{\min\{j,j'\}-1} c_{j-k-1}^2 c_{j'-k-1}^2 \\ \leqslant \frac{4\sigma_4(p+1)}{p^4} \sum_{j,j'=1}^{n+1} \overline{c}_{j-\min\{j,j'\}} \overline{c}_{j'-\min\{j,j'\}} \sum_{k=1}^{\min\{j,j'\}-1} \overline{c}_{k-1}^2 \\ \leqslant \frac{4\sigma_4(p+1)}{p^4} \sum_{j,j'=1}^{n+1} \overline{c}_0 \overline{c}_{|j-j'|} \sum_{k=1}^n \overline{c}_{k-1}^2.$$

Introducing the new summation variable $\delta_j := j - j'$, one finds that

$$\mu_{2,\Sigma_{\widehat{\mathbf{X}}}^{(1)}}^{\Pi\Pi} \leqslant \frac{4\sigma_4(p+1)(n+1)}{p^4} \overline{c}_0 \left[\overline{c}_0 + 2\sum_{\delta_j=1}^n \overline{c}_{\delta_j} \right] \sum_{k=1}^n \overline{c}_{k-1}^2 = O(n^{-2}).$$
(6.3.9)

The first factor being paired with the second, and the third with the fourth, means that k = k', m = m', and $m \neq (i - i')n + k$, so that the contribution of this configuration is given by

$$\mu_{2,\Sigma_{\widehat{\mathbf{X}}}^{(1)}}^{\Pi} = \frac{4}{p^4} \sum_{i,i'=1}^{p+1} \sum_{j,j'=1}^{n+1} \sum_{k=1}^{j-1} \sum_{m=1}^{j'-1} c_{j-k-1}^2 c_{j'-m-1}^2 - \mu_{2,\Sigma_{\widehat{\mathbf{X}}}^{(1)}}^{\Pi\Pi} = \left(\mu_{1,\Sigma_{\widehat{\mathbf{X}}}^{(1)}}\right)^2 + O(n^{-2}).$$
(6.3.10)

For the \square pairing, the constraints are i = i', k = m, k' = m', $k \neq k'$, and the corresponding contribution is

$$\mu_{2,\Sigma_{\widehat{\mathbf{X}}}^{(1)}}^{[\text{Th}]} = \frac{4}{p^4} \sum_{i=1}^{p+1} \sum_{j,j'=1}^{n+1} \sum_{k,k'=1}^{\min\{j,j'\}-1} c_{j-k-1}c_{j'-k-1}c_{j'-k'-1} - \mu_{2,\Sigma_{\widehat{\mathbf{X}}}^{(1)}}^{[\text{Th}]}$$

$$\leq \frac{4(p+1)}{p^4} \sum_{j,j'=1}^{n+1} \overline{c}_{j-\min\{j,j'\}}\overline{c}_{j'-\min\{j,j'\}} \sum_{k,k'=1}^{\min\{j,j'\}-1} \overline{c}_{k-1}c_{k'-1} + O(n^{-2})$$

$$\leq \frac{4(p+1)(n+1)}{p^4} \overline{c}_0 \left[\overline{c}_0 + 2\sum_{\delta_j=1}^n \overline{c}_{\delta_j}\right] \sum_{k,k'=1}^n \overline{c}_{k-1}\overline{c}_{k'-1} + O(n^{-2}) = O(n^{-2}). \quad (6.3.11)$$

Renaming the summation indices shows that $\mu_{2,\Sigma_{\widehat{\mathbf{X}}}^{(1)}}^{[\Pi]} = \mu_{2,\Sigma_{\widehat{\mathbf{X}}}^{(1)}}^{[\Pi]}$. Combining this with the displays (6.3.9) to (6.3.11), it follows that $\operatorname{Var}\Sigma_{\widehat{\mathbf{X}}}^{(1)} = \mu_{2,\Sigma_{\widehat{\mathbf{X}}}^{(1)}} - \mu_{1,\Sigma_{\widehat{\mathbf{X}}}^{(1)}}^2 = O(n^{-2})$. Since a very similar reasoning can be applied to $\Sigma_{\widehat{\mathbf{X}}}^{(2)}$, and $p^{-4}\operatorname{Vartr}\widehat{\mathbf{X}}\widehat{\mathbf{X}}^{\mathsf{T}}$ is smaller than $2\operatorname{Var}\Sigma_{\widehat{\mathbf{X}}}^{(1)} + 2\operatorname{Var}\Sigma_{\widehat{\mathbf{X}}}^{(2)}$, we conclude that $p^{-4}\operatorname{Vartr}\widehat{\mathbf{X}}\widehat{\mathbf{X}}^{\mathsf{T}}$ is of order $O(n^{-2})$.

The intention behind Proposition 6.3.2 was to allow the application of results about the limiting spectral distribution of matrices of the form $\mathbf{Z}H\mathbf{Z}^{\mathsf{T}}$, where \mathbf{Z} is an i. i. d. matrix, and *H* is a positive semidefinite matrix. Expressions for the Stieltjes transform of the LSD of such matrices in terms of the LSD of H have been obtained by [77, 98], and, in the most general form, by [86]. The next lemma shows that in the current context the population covariance matrix H has the same LSD as the auto-covariance matrix Γ of the process X_t , which is defined in terms of the auto-covariance function $\gamma(h) = \sum_{j=0}^{\infty} c_j c_{j+|h|}$ by $\Gamma = (\gamma(i-j))_{ij}$; this correspondence is used to characterize the LSD of H by the spectral density f associated with the coefficients $(c_j)_j$.

Lemma 6.3.3. Let Ω be given by Eq. (6.3.7). The limiting spectral distribution of the matrix $\Omega\Omega^{T}$ exists and is the same as the limiting spectral distribution of the auto-covariance matrix Γ . It therefore satisfies

$$\int h(\lambda)\hat{F}^{\Omega\Omega^{\mathsf{T}}}(\mathrm{d}\lambda) = \frac{1}{2\pi} \int_{0}^{2\pi} h(f(\omega))\mathrm{d}\omega, \qquad (6.3.12)$$

for every continuous function h.

Proof. The first claim follows by standard computations from the fact that Ω is, except for one missing row, a circulant matrix with entries $\Omega_{ij} = c_{n+j-i \mod (n+1)}$, and [9, Corollaries A.41 and A.42]. The second claim is an application of Szegő's limit theorem about the LSD of Toeplitz matrices; see [103, Theorem XVIII] for the original result or, e. g., [22, Sections 5.4 and 5.5] for a modern treatment.

Proof of Theorem 6.1. According to Proposition 6.3.2, the matrix $\widehat{\mathbf{X}} \widehat{\mathbf{X}}^{\mathsf{T}}$ is of the form $\mathbf{Z} \Omega \Omega^{\mathsf{T}} \mathbf{Z}^{\mathsf{T}}$, where Ω is given by Eq. (6.3.7). Using [86, Theorem 1] and the fact that, by Lemma 6.3.3, the limiting spectral distribution of $\Omega \Omega^{\mathsf{T}}$ exists, it follows that the limiting spectral distribution $\widehat{F}^{p^{-1}\widehat{\mathbf{X}}\widehat{\mathbf{X}}^{\mathsf{T}}}$ exists. Therefore, the combination of Propositions 6.3.1 and 6.3.2 shows that the limiting spectral distribution of $p^{-1}\mathbf{X}\mathbf{X}^{\mathsf{T}}$ also exists and is the same as that of $p^{-1}\widehat{\mathbf{X}}\widehat{\mathbf{X}}^{\mathsf{T}}$. [86, equation (1.2)] thus implies that the Stieltjes transform of $\widehat{F}^{p^{-1}\mathbf{X}\mathbf{X}^{\mathsf{T}}}$ is the unique mapping $s_{\widehat{F}p^{-1}\mathbf{X}\mathbf{X}^{\mathsf{T}}: \mathbb{C}^+ \to \mathbb{C}^+$ which solves

$$\frac{1}{s_{\hat{F}^{p^{-1}\mathbf{X}\mathbf{X}^{\mathsf{T}}}(z)} = -z + y \int_{\mathbb{R}} \frac{\lambda}{1 + \lambda s_{\hat{F}^{p^{-1}\mathbf{X}\mathbf{X}^{\mathsf{T}}}(z)}} \hat{F}^{\Omega\Omega^{\mathsf{T}}}(\mathrm{d}\lambda),$$

and Eq. (6.3.12) from Lemma 6.3.3 completes the proof.

6.4. Sketch of an alternative proof of Theorem 6.1

In this section we indicate how Theorem 6.1 could be proved alternatively using the methods employed in Chapter 5. We denote by $\widetilde{\mathbf{X}}_{(\alpha)}$ the matrix which is defined as in Eq. (6.2.1) but with the linear process being truncated at $\lfloor n^{\alpha} \rfloor$ with $0 < \alpha < 1$, i. e.

$$\widetilde{\mathbf{X}}_{(\alpha)} = \left(\sum_{j=0}^{\lfloor n^{\alpha} \rfloor} c_j Z_{(i-1)n+t-j}\right)_{it}.$$
If $1 - \alpha$ is sufficiently small, then an adaptation of the proof of Proposition 6.3.1 to this setting shows that $p^{-1}\mathbf{X}\mathbf{X}^{\mathsf{T}}$ and $p^{-1}\mathbf{\widetilde{X}}_{(\alpha)}\mathbf{\widetilde{X}}_{(\alpha)}^{\mathsf{T}}$ have the same limiting spectral distribution almost surely. The next step is to partition $\mathbf{\widetilde{X}}_{(\alpha)}$ into two blocks of dimensions $p \times \lfloor n^{\alpha} \rfloor$ and $p \times (n - \lfloor n^{\alpha} \rfloor)$, respectively. If we denote these two blocks by $\mathbf{\widetilde{X}}_{(\alpha)}^{1}$ and $\mathbf{\widetilde{X}}_{(\alpha)}^{2}$, i. e. $\mathbf{\widetilde{X}}_{(\alpha)} = \left[\mathbf{\widetilde{X}}_{(\alpha)}^{1}\mathbf{\widetilde{X}}_{(\alpha)}^{2}\right]$, then clearly

$$\widetilde{\mathbf{X}}_{(\alpha)}\widetilde{\mathbf{X}}_{(\alpha)}^{\mathsf{T}} = \widetilde{\mathbf{X}}_{(\alpha)}^{1} \left(\widetilde{\mathbf{X}}_{(\alpha)}^{1} \right)^{\mathsf{T}} + \widetilde{\mathbf{X}}_{(\alpha)}^{2} \left(\widetilde{\mathbf{X}}_{(\alpha)}^{2} \right)^{\mathsf{T}},$$

and an application of [9, Theorem A.43] yields that

$$\begin{split} \sup_{\lambda \in \mathbb{R}_{\geq 0}} \left| F^{p^{-1}\mathbf{X}\mathbf{X}^{\mathsf{T}}}([0,\lambda]) - F^{p^{-1}\widetilde{\mathbf{X}}^{2}_{(\alpha)}\left(\widetilde{\mathbf{X}}^{2}_{(\alpha)}\right)^{\mathsf{T}}}([0,\lambda]) \right| &\leq \frac{1}{p} \operatorname{rank}\left(\widetilde{\mathbf{X}}^{1}_{(\alpha)}\left(\widetilde{\mathbf{X}}^{1}_{(\alpha)}\right)^{\mathsf{T}}\right) \\ &\leq \frac{1}{p} \min\left(\lfloor n^{\alpha}\rfloor, p\right) = O\left(p^{-1}n^{\alpha}\right) \to 0. \end{split}$$

It therefore suffices to derive the limiting spectral distribution of $p^{-1}\widetilde{\mathbf{X}}_{(\alpha)}^2 \left(\widetilde{\mathbf{X}}_{(\alpha)}^2\right)^{\mathsf{T}}$. Since the matrix $\widetilde{\mathbf{X}}_{(\alpha)}^2$ has independent rows, this could be done by a careful adaptation of the arguments given in Chapter 5. We chose, however, to provide a self-contained proof, which also provides intermediate results of independent interest like Proposition 6.3.2, and we therefore omit the lengthy details of this alternative proof.

Part III.

Strong Solutions of Stochastic Differential Equations

CHAPTER 7

On strong solutions for positive definite jump diffusions⁴

7.1. Introduction

A result of the general theory for affine Markov processes on the cone S_d^+ of symmetric positive semidefinite matrices developed in [32] is that for a $d \times d$ matrix-valued standard Brownian motion B, $d \times d$ matrices Q and β , a symmetric constant drift b, and a positive linear drift $\Gamma : S_d^+ \to S_d^+$, weak global solutions exist to the stochastic differential equation (SDE)

$$dX_t = \sqrt{X_t} dB_t Q + Q^{\mathsf{T}} dB_t^{\mathsf{T}} \sqrt{X_t} + (X_t \beta + \beta^{\mathsf{T}} X_t + \Gamma(X_t) + b) dt, \qquad (7.1.1)$$
$$X_0 = x \in S_d^+,$$

whenever $b - (d - 1)Q^{\top}Q \in S_d^+$. Above \sqrt{X} denotes the unique positive semidefinite square root of a matrix $X \in S_d^+$. For $\Gamma = 0$ solutions to the SDE (7.1.1) are called Wishart processes and their existence has been considered in detail in the fundamental paper by Marie-France Bru [26]. Further probabilistic investigations on properties of Wishart processes have been carried out in [41, 42, 55], for instance, and references therein.

In the present chapter, we focus on the existence of global *strong solutions* of (7.1.1) and generalisations of it including jumps and more general diffusion coefficients. Because of the non-Lipschitz diffusion at the boundary of the cone, this problem is a quite delicate one – a-priori it is only clear that a unique local solution of (7.1.1) exists until X_t hits the boundary of S_d^+ , since the SDE is locally Lipschitz in the interior of S_d^+ . Furthermore, known results for pathwise uniqueness, for instance, that of the seminal paper of Yamada and Watanabe [111,

⁴This chapter is based on *E. Mayerhofer, O. Pfaffel and R. Stelzer:* On Strong Solutions for Positive Definite Jump Diffusions, Stochastic Process. Appl., 121(9), 2072–2086, 2011.

Corollary 3], are essentially one-dimensional, and therefore do not apply. Hence, the present setting seems to be more complicated than, for instance, the canonical affine one (concerning diffusions on $\mathbb{R}^m_+ \times \mathbb{R}^n$, [50, Lemma 8.2]).

Positive semidefinite matrix valued processes are increasingly used in finance, particularly for stochastic modelling of multivariate stochastic volatility phenomena in equity and fixed income models, see [28, 29, 34, 33, 35, 36, 53, 54, 57, 91]. See also [32] and the references therein. Most papers mentioned use Bru's class of Wishart diffusions, as this results in multivariate analogues of the popular Heston stochastic volatility model and its extensions, Ornstein-Uhlenbeck type processes ([91]) giving a multivariate generalisation of the popular model of [12] or a combination of both ([73]). This motivated the research of [32] on positive semidefinite affine processes including all the aforementioned models and generalising the results of [43], which covered all of these models in the univariate setting. Appropriate multivariate models are especially important for issues like portfolio optimisation, portfolio risk management and the pricing of options depending on several underlyings, which are heavily influenced by the dependence structure.

Clearly S_d^+ -valued processes model the covariances, not the correlations, which are, however, preferable when interpreting the dependence structure. The results of the present chapter are particularly relevant, when one wants to derive correlation dynamics (see e.g., [28, 29]), because one needs to assume boundary non-attainment conditions for a rigorous derivation.

The name "Wishart process" is, unfortunately, not always used in the same way in the literature. We follow the above cited applied papers in finance and call any solution to (7.1.1) with $\Gamma = 0$ "Wishart process" whereas in most of the previous probabilistic literature "Wishart process" also means $\beta = 0$ and the "Wishart processes with drift" of [42] are not even special cases of our "Wishart processes". For $\Gamma = \beta = 0$ and $b = nQ^TQ$ with $n \in \mathbb{N}$ one may also speak of a "squared Ornstein-Uhlenbeck process". In the univariate case the name "Wishart process" is not used, instead one typically uses "Cox-Ingersoll-Ross process" in the financial and "squared Bessel process" in the probability literature.

However, in this chapter we do not limit ourselves to the analysis of (7.1.1). Instead, as a special case of a considerably more general result, we consider a similar SDE allowing for a general (not necessarily linear) drift Γ and an additional jump part of finite variation. This implies that many Lévy-driven SDEs on S_d^+ like the positive semidefinite Ornstein-Uhlenbeck (OU) type processes (see [13, 92]) or the volatility process of a multivariate COGARCH process (see [101]), where the existence of global strong solutions has previously been shown by path-wise arguments, are special cases of our setting. Thus our results allow to consider certain "jump diffusions", viz. mixtures of such jump processes and Wishart diffusions, in applications. It should be noted that [26] also contains results on strong solutions for Wishart processes (see our upcoming Proposition 7.3.1 and Remark 7.4.1), however, they are derived under strong parametric restrictions, because her method requires an application of Girsanov's theorem. The latter is based on a martingale criterion, which in the matrix valued setting seems hard to verify. Also, the general result (with a non-vanishing linear drift) only holds until the first time when two of the eigenvalues of the process collide. Our approach generalises her method of proof for the case $\beta = 0$ (vanishing linear drift) and avoids change of measure techniques.

The most general result of our chapter, Theorem 7.1, also opens the way to use positive semidefinite extensions of the univariate GARCH diffusions of [84] or of so-called generalised Cox-Ingersoll-Ross models (cf. e.g. [20, 49]), where the square root in the diffusion part of (7.1.1) is replaced by the α -th positive semidefinite power with $\alpha \in [1/2, 1]$, in applications (see Corollary 7.3.3).

The remainder of the chapter is structured as follows. In the subsequent section we summarise some notation and preliminaries. In Section 7.3 we state our main result, Theorem 7.1, and its corollaries applying to Wishart processes, matrix-variate generalised Cox-Ingersoll-Ross and GARCH diffusions. Moreover, we compare our results to the work of Bru which is recalled in Proposition 7.3.1. In the following section we gradually develop the proof of our result. Our method relies on a generalisation of the so-called *McKean's argument*, but avoids the use of Girsanov's theorem. In Section 7.4.1 we thus provide a self-contained proof of a generalisation of *McKean's argument* and then deliver the proof of Theorem 7.1 in Section 7.4.2. We conclude the chapter with some final remarks in Section 7.5.

7.2. Notation and general set-up

We assume given an appropriate filtered probability space $(\Omega, \mathcal{F}, \mathbb{P}, (\mathcal{F}_t)_{t \in \mathbb{R}_+})$ satisfying the usual hypotheses (complete and right-continuous filtration) and rich enough to support all processes occurring. For short, we sometimes write just Ω when actually referring to this filtered probability space. *B* is a $d \times d$ standard Brownian motion on Ω and $d \in \mathbb{N}$ always denotes the dimension. Furthermore, we use the following notation, definitions and setting:

- $\mathbb{R}_+ := [0, \infty), M_d$ is the set of real valued $d \times d$ matrices and I_d is the identity matrix.
- S_d ⊂ M_d is the space of symmetric matrices, and S⁺_d ⊂ S_d is the cone of symmetric positive semidefinite matrices in S_d and S⁺⁺⁺_d its interior, i.e. the positive definite matrices. The partial order on S_d induced by the cone is denoted by ≤, and x > 0, if and only if x ∈ S⁺⁺⁺_d. We endow S_d with the scalar product ⟨x, y⟩ := tr(xy), where tr(A) denotes the trace of A ∈ M_d. || · || denotes the associated norm, and d(x, ∂S⁺_d) = inf_{y∈∂S⁺_d} ||x y|| is the distance of x ∈ S⁺_d to the boundary ∂S⁺_d.
- The usual tensor (Kronecker) product of two matrices A, B is denoted by $A \otimes B$ and the vectorisation operator mapping M_d to \mathbb{R}^{d^2} by stacking the columns of a matrix A below each other is denoted by vec(A) (see [65, Chapter 4] for more details).

7. On strong solutions for positive definite jump diffusions

- A function f : S⁺⁺_d → U with U being (a subset of) a normed space is called *locally* Lipschitz if ||f(x) f(y)|| ≤ K(C)||x y|| ∀ x, y ∈ C for all compacts C ⊂ U. f is said to have *linear growth* if ||f(x)||² ≤ K(1 + ||x||²) ∀ x ∈ S⁺⁺_d.
- An S_d -valued càdlàg adapted stochastic process X is called S_d^+ -increasing, if $X_t \ge X_s$ a.s. for all $t > s \ge 0$. Such a process is necessarily of finite variation on compacts by [13, Lemma 5.21] and hence a semimartingale. We call it of *pure jump type* provided $X_t = X_0 + \sum_{0 \le s \le t} \Delta X_s$, where $\Delta X_s = X_s - X_{s-}$.

For the necessary background on stochastic analysis we refer to one of the standard references like [67, 94, 97]. Moreover, we frequently employ stochastic integrals where the integrands or integrators are matrix- or even linear-operator valued. Thus, we briefly explain how they have to be understood. Let $(A_t)_{t \in \mathbb{R}^+}$ in M_d , $(B_t)_{t \in \mathbb{R}^+}$ in M_d be càdlàg and adapted processes and $(L_t)_{t \in \mathbb{R}^+}$ in M_d be a semimartingale (i.e. each element is a semimartingale). Then we denote by $\int_0^t A_{s-d}L_sB_{s-}$ the matrix C_t in M_d which has ij-th element $C_{ij,t} = \sum_{k=1}^d \sum_{l=1}^d \int_0^t A_{ik,s-}B_{lj,s-}dL_{kl,s}$. Equivalently such an integral can be understood in the sense of [82] by identifying it with the integral $\int_0^t \mathbf{A}_{s-d}L_s$ with \mathbf{A}_t being for each fixed t the linear operator $M_d \to M_d$, $X \mapsto A_t X B_t$ and L being a semimartingale in the Hilbert space M_d . Stochastic integrals of the form $\int_0^t K(X_{s-})dJ_s$ with J being a semimartingale in M_d (coordinatewise or equivalently as in [82, Section 10] where the equivalence easily follows from [82, Section 10.9] and by noting that on a finite dimensional Hilbert space all norms are equivalent) and $K(x) : M_d \to M_d$ a linear operator for all x can be understood again as in [82]. Alternatively, one can equivalently identify M_d with \mathbb{R}^{d^2} using the vec-operator and K(x) with a matrix in M_{d^2,d^2} and then define the stochastic integral coordinatewise as above.

7.3. Statement of the main results

7.3.1. Wishart diffusions with jumps

In order to illustrate the context of our result and, because it is of most relevance in applications, we discuss first the special case of Wishart diffusions with jumps. For $Q \in M_d$, $\delta > d - 1, \beta \in M_d$ and an M_d -valued standard Brownian motion *B*, a Wishart process is the strong solution of the equation

$$dX_t = \sqrt{X_t} dB_t Q + Q^\top dB_t^\top \sqrt{X_t} + (X_t \beta + \beta^\top X_t + \delta Q^\top Q) dt, \qquad (7.3.1)$$
$$X_0 = x \in S_d^{++},$$

on the maximal stochastic interval $[0, T_x)$, where T_x is naturally defined as

$$T_x = \inf\{t > 0 : X_t \in \partial S_d^+\}.$$

That such a unique local strong solution, which does not explode before or at time T_x , exists, follows from standard SDE theory, since all the coefficients in (7.3.1) are locally Lipschitz and of linear growth on S_d^{++} . To be more precise, this follows by appropriately localising the usual results as e.g. in [94, Chapter V] or by variations of the proofs in [82, Chapter 3]. A localisation procedure adapted particularly to certain convex sets like S_d^+ is presented in detail in [102, Section 6.7].

The following is a summary of the results [26, Theorem 2, 2' and 2''] – the to the best of our knowledge only known results regarding strong existence of Wishart processes:

Proposition 7.3.1. Let $\delta \ge d + 1$.

(i) If $Q = I_d$ and $\beta = 0$, then $T_x = \infty$.

Suppose additionally that the d eigenvalues of x are distinct.

- (ii) If $Q \in S_d^{++}$, $-\beta \in S_d^+$ such that β and Q commute, then there exists a solution $(X_t)_{t \in \mathbb{R}_+}$ of (7.3.1) until the first time τ_x when two of the eigenvalues of X_t collide.
- (iii) If $\beta = \beta_0 I_d$ and $Q = \gamma I_d$, where $\beta_0, \gamma \in \mathbb{R}$, then $T_x = \infty$ for the solution of $(X_t)_{t \in \mathbb{R}_+}$ of (7.3.1).

Consequently, for the respective choice of parameters, there exist unique global strong S_d^{++} -valued solutions of the SDE (7.3.1) on $[0, \tau_x)$ resp. on all of $[0, \infty)$.

The upcoming general Theorem 7.1 implies the following result for a generalisation of the Wishart SDE allowing for additional jumps and a non-linear drift Γ .

Corollary 7.3.2. Let $b \in S_d$, $Q \in M_d$, $\beta \in M_d$, and let

- J be an S_d -valued càdlàg adapted process which is S_d^+ -increasing and of pure jump type,
- $\Gamma: S_d^{++} \to S_d^+$ be a locally Lipschitz function of linear growth and
- $K: S_d^{++} \to L(S_d^+, S_d^+)$ (the linear operators on S_d mapping S_d^+ into S_d^+) be a locally Lipschitz function of linear growth.

If $b \geq (d+1)Q^{\top}Q$, then the SDE

 $dX_{t} = \sqrt{X_{t-}} dB_{t} Q + Q^{\top} dB_{t}^{\top} \sqrt{X_{t-}} + (X_{t-}\beta + \beta^{\top} X_{t-} + \Gamma(X_{t-}) + b) dt + K(X_{t-}) dJ_{t}, \quad (7.3.2)$ $X_{0} = x \in S_{d}^{++},$

has a unique adapted càdlàg global strong solution $(X_t)_{t \in \mathbb{R}_+}$ on S_d^{++} . In particular we have $T_x := \inf\{t \ge 0 : X_{t-} \in \partial S_d^+ \text{ or } X_t \notin S_d^{++}\} = \inf\{t \ge 0 : X_{t-} \in \partial S_d^+\} = \infty$ almost surely.

Proof. For the term on the right hand side of the upcoming condition (7.3.3) we obtain

$$\operatorname{tr}(2\beta) + \operatorname{tr}(\Gamma(x)x^{-1}) + \operatorname{tr}((b - (d+1)Q^{\top}Q)x^{-1}) \ge 2\operatorname{tr}(\beta),$$

noting that x^{-1} , $\Gamma(x)$ and $b - (d+1)Q^{\top}Q$ are positive semidefinite and that S_d^+ is a selfdual cone, which implies that $\operatorname{tr}(zy) \ge 0$ for any $z, y \in S_d^+$. Setting $c(t) = 2\operatorname{tr}(\beta)$ an application of Theorem 7.1 concludes.

By choosing Γ linear and J = 0, we obtain a result for (7.1.1) which considerably generalises Proposition 7.3.1.

- **Remark 7.3.1.** (i) In the univariate case the condition $b \ge (d+1)Q^{\top}Q$ is known to be also necessary for boundary non-attainment (see [97, Chapter XI]).
 - (ii) A possible choice for J is a matrix subordinator without drift (see [11]), i.e. an S_d^+ increasing Lévy process. By choosing $\Gamma \neq 0$ in (7.3.2) appropriately our results also
 apply to SDEs involving matrix subordinators with a non-vanishing drift.
- (iii) Setting Q = 0, $\Gamma = 0$, K to the identity and b equal to the drift of the matrix subordinator, Equation (7.3.2) becomes the SDE of a positive definite OU type process, [13, 92]. Likewise, it is straightforward to see that the SDE of the volatility process Y of the multivariate COGARCH process of [101] is a special case of (7.3.2).
- (iv) An OU-type process on the positive semidefinite matrices is necessarily driven by a Lévy process of finite variation having positive semidefinite jumps only (follows by slightly adapting the arguments in the proof of [92, Theorem 4.9]). This entails that a generalisation of the above result to a more general jump behaviour requires additional technical restrictions.

7.3.2. The general SDE and existence result

The main result of this chapter is the following general theorem concerning non-attainment of the boundary of S_d^+ and the existence of a unique global strong solution for a generalisation of the SDE (7.1.1). The proof of this result is gradually developed in the next sections.

Theorem 7.1. Let

- $F, G : \mathbb{R}_+ \times S_d^{++} \to M_d$, be functions such that $G^{\top} \otimes F$ given by $G^{\top} \otimes F(t, x) = (G(t, x))^{\top} \otimes F(t, x)$ is locally Lipschitz and of linear growth,
- $H: \mathbb{R}_+ \times S_d^{++} \to S_d$ be locally Lipschitz and of linear growth,

- J be an S_d -valued càdlàg adapted process which is S_d^+ -increasing and of pure jump type,
- and $K: S_d^{++} \to L(S_d^+, S_d^+)$ (the linear operators on S_d mapping S_d^+ into S_d^+) be a locally Lipschitz function of linear growth.

Suppose that there exists a function $c : \mathbb{R}_+ \to \mathbb{R}$ which is locally integrable, i.e. $\int_0^s |c(t)| dt < \infty$ ∞ for all $s \in \mathbb{R}^+$, such that

$$c(t) \leq \operatorname{tr}(H(t,x)x^{-1}) - \operatorname{tr}(f(t,x)x^{-1})\operatorname{tr}(g(t,x)x^{-1}) - \operatorname{tr}(f(t,x)x^{-1}g(t,x)x^{-1})$$
(7.3.3)

for all $x \in S_d^{++}$ and $t \in \mathbb{R}_+$ where $f(t, x) := F(t, x)F(t, x)^\top$, $g(t, x) = G(t, x)^\top G(t, x)$. Then the SDE

$$dX_{t} = F(t, X_{t-}) dB_{t}G(t, X_{t-}) + G(t, X_{t-})^{\top} dB_{t}^{\top}F(t, X_{t-})^{\top} + H(t, X_{t-}) dt + K(X_{t-}) dJ_{t},$$

$$X_{0} = x \in S_{d}^{++},$$
(7.3.4)

has a unique adapted càdlàg global strong solution $(X_t)_{t \in \mathbb{R}_+}$ on S_d^{++} . In particular, we have $T_x := \inf\{t \ge 0 : X_{t-} \in \partial S_d^+ \text{ or } X_t \notin S_d^{++}\} = \inf\{t \ge 0 : X_{t-} \in \partial S_d^+ \text{ or } X_t \notin S_d^{++}\}$ $\partial S_d^+ = \infty$ almost surely.

7.3.3. Positive definite extensions of generalised Cox-Ingersoll-Ross processes and GARCH diffusions

In the univariate case generalised Cox-Ingersoll-Ross (GCIR) processes given by the SDE $dx_t = (b + ax_t)dt + qx_t^{\alpha}dB_t$ with $b \ge 0, q > 0, a \in \mathbb{R}$ and $\alpha \in [1/2, 1]$ are – as discussed in the introduction – of relevance in financial modelling. $\alpha = 1/2$ corresponds, of course, to the already discussed Bessel case, whereas $\alpha = 1$ gives the so-called GARCH diffusions. Given the popularity of the Wishart based models in nowadays finance, it seems natural to consider also positive semidefinite extensions of the GCIR processes. An application of our general theorem to the case where $F(X) = X^{\alpha}$, G(X) = Q with $\alpha \in [1/2, 1]$ yields:

Corollary 7.3.3. (i) Let $\alpha \in [1/2, 1]$, $b \in S_d$, $Q \in M_d$, $\beta \in M_d$, and let

- J be an S_d -valued càdlàg adapted process which is S_d^+ -increasing and of pure jump type,
- $\Gamma: S_d^{++} \to S_d^+$ be a locally Lipschitz function of linear growth and
- $K: S_d^{++} \to L(S_d^+, S_d^+)$ (the linear operators on S_d mapping S_d^+ into S_d^+) be a locally Lipschitz function of linear growth.

Suppose that for all $x \in S_d^{++}$

$$\operatorname{tr}(\Gamma(x)x^{-1} + bx^{-1}) \ge \operatorname{tr}(x^{2\alpha-1})\operatorname{tr}(Q^{\top}Qx^{-1}) + \operatorname{tr}(x^{2\alpha-2}Q^{\top}Q).$$
(7.3.5)

Then the SDE

$$dX_{t} = X_{t-}^{\alpha} B_{t} Q + Q^{\top} dB_{t}^{\top} X_{t-}^{\alpha} + (X_{t-}\beta + \beta^{\top} X_{t-} + \Gamma(X_{t-}) + b) dt + K(X_{t-}) dJ_{t}, \quad (7.3.6)$$

$$X_{0} = x \in S_{d}^{++},$$

has a unique adapted càdlàg global strong solution $(X_t)_{t \in \mathbb{R}_+}$ on S_d^{++} . In particular we have $T_x := \inf\{t \ge 0 : X_{t-} \in \partial S_d^+ \text{ or } X_t \notin S_d^{++}\} = \inf\{t \ge 0 : X_{t-} \in \partial S_d^+\} = \infty$ almost surely. (ii) Any of the following sets of conditions implies (7.3.5):

- (a) $b + \Gamma(x) \ge \operatorname{tr}(x^{2\alpha-1})Q^{\mathsf{T}}Q + x^{\alpha-1/2}Q^{\mathsf{T}}Qx^{\alpha-1/2}$ for all $x \in S_d^{++}$.
- (b) $b + \Gamma(x) \ge \operatorname{tr}(x^{2\alpha-1})Q^{\top}Q + \lambda_{Q^{\top}Q}x^{2\alpha-1}$ for all $x \in S_d^{++}$ with $\lambda_{Q^{\top}Q}$ denoting the largest eigenvalue of $Q^{\top}Q$.
- (c) $\alpha = 1$ and $b + \Gamma(x) \ge \operatorname{tr}(x)Q^{\top}Q + \lambda_{Q^{\top}Q}x$ for all $x \in S_{d}^{++}$.
- (d) $b \ge 0$ and $\Gamma(x) \ge 2\operatorname{tr}(x^{2\alpha-1})Q^{\top}Q$ for all $x \in S_d^{++}$.
- (e) $b \ge 0$ and $\Gamma(x) \ge 2\left(\operatorname{tr}(x) + d(2\alpha 1)^{2-2\alpha}\right)Q^{\top}Q$ for all $x \in S_d^{++}$ (and setting $0^0 := 1$ for $\alpha = 1/2$).

(f)
$$b \ge 0$$
 and $\Gamma(x) \ge 2(\operatorname{tr}(x) + d)Q^{\top}Q$ for all $x \in S_d^{++}$.

(g)
$$\alpha > 1/2$$
, $d = 1$, $\Gamma(x) \ge 0$ for all $x \in \mathbb{R}_+$ and $b > 0$.

Proof. One easily calculates the right hand side of (7.3.3) to be equal to $tr(2\beta + \Gamma(x)x^{-1} + bx^{-1}) - tr(x^{2\alpha-1})tr(Q^{\top}Qx^{-1}) - tr(x^{2\alpha-2}Q^{\top}Q)$ and hence (i) follows from Theorem 7.1.

Turning to the proof of (ii) using the selfduality of S_d^+ as in the proof of Corollary 7.3.2 gives (a). Next we observe that $Q^\top Q \leq \lambda_{Q^\top Q} I_d$ and, hence, $x^{\alpha-1/2} Q^\top Q x^{\alpha-1/2} \leq \lambda_{Q^\top Q} x^{2\alpha-1}$. This gives (b) and (c) is simply the special case for $\alpha = 1$.

Since for $A, B \in S_d^+$ we have that $\operatorname{tr}(AB) \leq \operatorname{tr}(A)\operatorname{tr}(B)$ due to the Cauchy-Schwarz inequality and the elementary inequality $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for all $a, b \in \mathbb{R}_+$, we have that $\operatorname{tr}(x^{2\alpha-2}Q^{\top}Q) \leq \operatorname{tr}(x^{2\alpha-1})\operatorname{tr}(Q^{\top}Qx^{-1})$. Hence, (7.3.5) is implied by $\operatorname{tr}(\Gamma(x)x^{-1} + bx^{-1}) \geq 2\operatorname{tr}(x^{2\alpha-1})\operatorname{tr}(Q^{\top}Qx^{-1})$. Using once again the selfduality gives (d).

Since the trace is the sum of the eigenvalues, $\lambda \ge \lambda^{2\alpha-1}$ for all $\lambda \ge 1$ and $\alpha \in [1/2, 1]$ and $\lambda^{2\alpha-1} \le \lambda + \max_{\lambda \in [0,1]} \{\lambda^{2\alpha-1} - \lambda\}$ for all $\lambda \in [0,1)$ and $\alpha \in [1/2, 1]$, we immediately obtain (e) from (d), because $\max_{\lambda \in [0,1]} \{\lambda^{2\alpha-1} - \lambda\} = (2\alpha - 1)^{2-2\alpha}$. In turn (f) follows from (e) noting that $\max_{\lambda \in [0,1]} \{\lambda^{2\alpha-1} - \lambda\} \in [0,1]$.

Turning to (g) we have for the right hand side of (7.3.3) in the univariate case

$$\ell(x) = 2\beta + \Gamma(x)/x + b/x - 2Q^2/x^{2-2\alpha}.$$

Now one notes that the second term is non-negative and that for b > 0 the term $b/x - 2Q^2/x^{2-2\alpha}$ is bounded from below on \mathbb{R}^+ , because $\lim_{x\to 0, x>0} x^{-1}/x^{2\alpha-2} = \infty$. Hence, Theorem 7.3.2 concludes.

In the different cases of (ii) a valid choice of b and Γ is always obtained by taking them equal to the right hand side of the inequalities. It should be noted that (c) shows that in the positive semidefinite GARCH diffusion generalisation one can always take a linear drift. Likewise, (e) and (f) show that a linear drift is possible for the generalized CIR. For $\alpha = 1/2$ the case (d) is again sharp in the univariate setting, but for general dimensions it is a stronger condition than the one given in Corollary 7.3.2.

The last case (g) in particular recovers the well-known univariate result for $dx_t = (b + ax_t)dt + qx_t^{\alpha}dB_t$ with $b \ge 0, q > 0, a \in \mathbb{R}$ and $\alpha \in [1/2, 1]$. In our matrix-variate case for $\alpha > 1/2$ a result similar to the univariate one, viz. that a strictly positive constant drift is all that is needed to ensure boundary non-attainment, seems to be out of reach. When one tries to use arguments similar to (e) in general, one would need something like $\operatorname{tr}(bx^{-1}) \ge k\operatorname{tr}(x^{2\alpha-1})\operatorname{tr}(Q^{\top}Qx^{-1}) + K$ with some constants k > 0 and K to ensure (7.3.5). However, when the process comes close to the boundary of the cone, this only means that at least one eigenvalue gets close to zero. Hence, $\operatorname{tr}(bx^{-1})$ and $\operatorname{tr}(Q^{\top}Qx^{-1})$ should then go to infinity at a comparable rate. However, all the other eigenvalues of x may still be arbitrarily large and so there is no appropriate upper bound on the term $\operatorname{tr}(x^{2\alpha-1})$.

7.4. Proofs

In this section we gradually prove our main result. As a priori all processes involved are only defined up to a stopping time, we collect first some basic definitions regarding stochastic processes defined on stochastic intervals following mainly [76].

Definition 7.4.1. *Let* $A \in \mathcal{F}$ *and let* T *be a stopping time.*

- A random variable X on A is a mapping $A \to \mathbb{R}$ which is measurable with respect to the σ -algebra $A \cap \mathcal{F}$.
- A family $(X_t)_{t \in \mathbb{R}_+}$ of random variables on $\{t < T\}$ is called a stochastic process on [0, T). If X_t is $\{t < T\} \cap \mathcal{F}_t$ -measurable for all $t \in \mathbb{R}_+$, then X is said to be adapted.
- An adapted process M on [0, T) is called a continuous local martingale on the interval [0, T) if there exists an increasing sequence of stopping times $(T_n)_{n \in \mathbb{N}}$ and a sequence of

7. On strong solutions for positive definite jump diffusions

continuous martingales $(M^{(n)})_{n \in \mathbb{N}}$ (in the usual sense on $[0, \infty)$) such that $\lim_{n\to\infty} T_n = T$ a.s. and $M_t = M_t^{(n)}$ on $\{t < T_n\}$. Other local properties for adapted processes on [0, T) are defined likewise.

- A semimartingale on [0, T) is the sum of a càdlàg local martingale on [0, T) and an adapted càdlàg process of locally finite variation on [0, T).
- For a continuous local martingale on [0, T) the quadratic variation is the $\mathbb{R} \cup \{\infty\}$ -valued stochastic process [M, M] defined by

$$[M, M]_t = \sup_{n \in \mathbb{N}} [M^{(n)}, M^{(n)}]_{t \wedge T_n} \text{ for all } t \in \mathbb{R}_+.$$

7.4.1. McKean's argument

In this section we finally establish Proposition 7.4.2 which generalises an argument of [79, p. 47, Problem 7] concerning continuous local martingales on stochastic intervals used, for instance, in [25, 26, 85]. We keep the tradition of referring to it as *McKean's argument*. Since it may also be helpful in other situations, we state our result and its proof in detail.

Lemma 7.4.1. Let *M* be a continuous local martingale on a stochastic interval [0, T). Then on $\{T > 0\}$ it holds almost surely that either $\lim_{t \uparrow T} M_t$ exists in \mathbb{R} or that

$$\limsup_{t\uparrow T} M_t = -\liminf_{t\uparrow T} M_t = \infty.$$

Proof. Combine [76, Theorem 3.5] with analogous arguments to the proof of [97, Chapter V, Proposition 1.8].

Proposition 7.4.2 (McKean's Argument). Let $Z = (Z_s)_{s \in \mathbb{R}_+}$ be an adapted càdlàg $\mathbb{R}^+ \setminus \{0\}$ -valued stochastic process on a stochastic interval $[0, \tau_0)$ such that $Z_0 > 0$ a.s. and $\tau_0 = \inf\{0 < s \leq \tau_0 : Z_{s-} = 0\}$. Suppose $h : \mathbb{R}_+ \setminus \{0\} \to \mathbb{R}$ is continuous and satisfies the following:

- (i) *For all* $t \in [0, \tau_0)$ *, we have* $h(Z_t) = h(Z_0) + M_t + P_t$ *, where*
 - a) *P* is an adapted càdlàg process on $[0, \tau_0)$ such that $\inf_{t \in [0, \tau_0 \wedge T)} P_t > -\infty$ a.s. for each $T \in \mathbb{R}^+ \setminus \{0\}$,
 - b) *M* is a continuous local martingale on $[0, \tau_0)$ with $M_0 = 0$,

(ii) and $\lim_{z\downarrow 0} h(z) = -\infty$.

Then $\tau_0 = \infty a.s.$

Above, $\tau_0 = \inf\{0 < s \le \tau_0 : Z_{s-} = 0\}$ is not to be understood as the definition of τ_0 , but it means that the already defined stopping time τ_0 is also the first hitting time of Z_{s-} at zero. Since Z is only defined up to time τ_0 , one cannot take the infimum over \mathbb{R}^+ .

Proof. Since $h(Z_t)_- = h(Z_{t-}) = h(Z_0) + P_{t-} + M_{t-}$ and P_{t-} is a.s. bounded from below on compacts, we have $\tau_0 = \inf\{s > 0 : M_{s-} = -\infty\}$ and further $\tau_0 > 0$ due to the right continuity of *Z*. Assume, by contradiction, that $\tau_0 < \infty$ on a set $A \in \mathcal{F}$ with $\mathbb{P}(A) > 0$. Hence, $\lim_{t \neq \tau_0} M_t = -\infty$ on *A* and this contradicts Lemma 7.4.1.

7.4.2. Proof of Theorem 7.1

Before we provide a proof of Theorem 7.1, we recall some elementary identities from matrix calculus and provide some further technical lemmata. For a differentiable function $f: M_d \to \mathbb{R}$, we denote by ∇f the usual gradient written in coordinates as $(\frac{\partial f}{\partial x_{ij}})_{ij}$.

Lemma 7.4.2. On S_d^{++} , we have

(i)
$$\nabla \det(x) = \det(x)(x^{-1})^{\top} = \det(x)x^{-1}$$
,

(ii)
$$\frac{\partial^2}{\partial x_{ij}\partial x_{kl}} \det(x) = \det(x)[(x^{-1})_{kl}(x^{-1})_{ij} - (x^{-1})_{il}(x^{-1})_{jk}].$$

Proof. The first identity in (i) can be found in [75, Section 9.10] and the second is an immediate consequence of restricting to symmetric matrices. Now (ii) follows using $\frac{\partial}{\partial x_{kl}}x^{-1} = -x^{-1}\left(\frac{\partial}{\partial x_{kl}}x\right)x^{-1}$ and finally the symmetry:

$$\frac{\partial}{\partial x_{kl} x_{ij}} \det(x) = \frac{\partial}{\partial x_{kl}} \left(\det(x) (x^{-1})_{ji} \right) = \det(x) \left((x^{-1})_{lk} (x^{-1})_{ji} + \frac{\partial}{\partial x_{kl}} (x^{-1})_{ji} \right)$$
$$= \det(x) \left((x^{-1})_{lk} (x^{-1})_{ji} - (x^{-1})_{jk} (x^{-1})_{li} \right).$$

For a semimartingale X we denote by X^c as usual its continuous part. All semimartingales in the following will have a discontinuous part of finite variation, i.e. $\sum_{0 < s \le t} ||\Delta X_s||$ is finite for all $t \in \mathbb{R}^+$. Thus we define $X_t^c = X_t - \sum_{0 < s \le t} \Delta X_s$ and note that the quadratic variation of a semimartingale is the one of its local continuous martingale part plus the sum of its squared jumps.

The continuous quadratic variation of X solving (7.3.4) is only influenced by the Brownian terms and, hence, we have a general version of [26, Equation (2.4)] which is proved just as [1, Lemma 2]:

Lemma 7.4.3. Consider the solution X_t of (7.3.4) on $[0, T_x)$. Then

$$\frac{d[X_{ij}, X_{kl}]_t^c}{dt} = (FF^{\mathsf{T}}(t, X_{t-}))_{ik} (G^{\mathsf{T}}G(t, X_{t-}))_{jl} + (FF^{\mathsf{T}}(t, X_{t-}))_{il} (G^{\mathsf{T}}G(t, X_{t-}))_{jk} + (FF^{\mathsf{T}}(t, X_{t-}))_{jk} (G^{\mathsf{T}}G(t, X_{t-}))_{il} + (FF^{\mathsf{T}}(t, X_{t-}))_{jl} (G^{\mathsf{T}}G(t, X_{t-}))_{ik}.$$

Here $G^{\top}G(t,x) := G(t,x)^{\top}G(t,x)$ and $FF^{\top}(t,x) := F(t,x)F(t,x)^{\top}$ to ease notation.

Moreover, we shall need the following result where a Brownian motion on a stochastic interval [0, T) is defined as a continuous local martingale on [0, T) with $[\beta, \beta]_t = t$.

Lemma 7.4.4. Let X_t be a continuous S_d^+ -valued adapted càdlàg stochastic process on a stochastic interval [0, T) with T being a predictable stopping time and let $h : M_d \to M_d$. Then there exists a one-dimensional Brownian motion β^h on [0, T) such that

$$\operatorname{tr}\left(\int_{0}^{t} h(X_{u-}) dB_{u}\right) = \int_{0}^{t} \sqrt{\operatorname{tr}(h(X_{u-})^{\top} h(X_{u-}))} d\beta_{u}^{h}$$
(7.4.1)

holds on [0, T).

Proof. We define for $t \in [0, T)$,

$$\beta_t^h := \sum_{i,j=1}^d \int_0^t \frac{h(X_{u-})_{ij}}{\sqrt{\operatorname{tr}(h(X_{u-})^\top h(X_{u-}))}} dB_{u,ji},$$

and since the numerator equals zero, whenever the denominator vanishes, we use the convention $\frac{0}{0} = 1$. Clearly for each *i*, *j* and for all $u \in [0, T)$ we have

$$\left|\frac{h(X_{u-})_{,ij}}{\sqrt{\operatorname{tr}(h(X_{u-})^{\top}h(X_{u-}))}}\right| \leq 1$$

which ensures that β^h is well-defined, square-integrable and a continuous local martingale on [0, T) by stopping at a sequence of stopping times announcing T. Furthermore, by construction

$$[\beta^{h},\beta^{h}]_{t} = \sum_{i,j=1}^{d} \int_{0}^{t} \frac{h(X_{u-})_{ij}^{2}}{\operatorname{tr}(h(X_{u-})^{\top}h(X_{u-}))} du = t$$

and therefore β^h is a Brownian motion on [0, T).

Finally by the very definition of β^h , we have

$$\operatorname{tr}(h(X_{t-})dB_t) = \sum_{i,j=1}^d h(X_{t-})_{ij} dB_{t,ji} = \sqrt{\operatorname{tr}(h(X_{t-})^{\mathsf{T}}h(X_{t-}))} d\beta_t^h,$$

which proves identity (7.4.1).

Finally, we state a variant of Itô's formula which we later employ. It follows easily from the usual versions like [16, Theorem 3.9.1] by arguments similar to [76, Theorem 5.4] and [13, Proposition 3.4].

Lemma 7.4.5. Let *X* be an S_d^{++} -valued semimartingale on a stochastic interval [0, T) and $f: S_d^{++} \to \mathbb{R}$ a twice continuously differentiable function. If $X_{t-} \in S_d^{++}$ for all $t \in [0, T)$ and $\sum_{0 < s \leq t} ||\Delta X_s|| < \infty$ for $t \in [0, T)$, then f(X) is a semimartingale on [0, T) and

$$f(X_t) = f(X_0) + \operatorname{tr}\left(\int_0^t \nabla f(X_{s-})^\top dX_s^c\right) + \frac{1}{2} \sum_{i,j,k,l=1}^d \int_0^t \frac{\partial^2}{\partial x_{ij} \partial x_{kl}} f(X_{s-}) d[X_{ij}, X_{kl}]_s^c + \sum_{0 < s \le t} \left(f(X_s) - f(X_{s-})\right).$$

We are now prepared to provide a proof of Theorem 7.1. Note that to shorten our formulae we use in the following differential notation and not integral notation as above.

Proof of Theorem 7.1. Since

$$\operatorname{vec}\left(F(t, X_{t-})dB_{t}G(t, X_{t-})\right) = \left(G(t, X_{t-})^{\top} \otimes \left(F(t, X_{t-})\right)\operatorname{vec}(dB_{t})\right)$$

it is easy to see that all coefficients of (7.3.4) are locally Lipschitz and of linear growth. Hence, standard SDE theory implies again the existence of a unique càdlàg adapted non-explosive local strong solution until the first time $T_x = \inf\{t \ge 0 : X_{t-} \in \partial S_d^+ \text{ or } X_t \notin S_d^{++}\}$ when X hits the boundary or jumps out of S_d^{++} . Hence, we have to show $T_x = \infty$.

By the choice of *K* and *J*, all jumps have to be positive semidefinite and hence the solution *X* cannot jump out of S_d^{++} . This implies that $T_x = \inf\{t \ge 0 : X_{t-} \in \partial S_d^+\}$.

In the following, all statements are meant to hold on the stochastic interval $[0, T_x)$. Note that by the right continuity of X_t , a.s. $T_x > 0$. Moreover, we set $T_n = \inf\{t \in \mathbb{R}_+ : d(X_t, \partial S_d^+) \leq 1/n \text{ or } ||X_t|| \geq n\}$. Then $(T_n)_{n \in \mathbb{N}}$ is an increasing sequence of stopping times with $\lim_{n\to\infty} T_n = T_x$, hence T_x is predictable.

We define the following processes and functions according to the notation of Proposition 7.4.2:

$$Z_t := \det(X_t), \quad h(z) := \ln(z), \quad r_t := h(Z_t).$$
 (7.4.2)

Then $T_x = \inf\{t > 0 : r_{t-} = -\infty\}.$

By Lemma 7.4.2 (i) and using the abbreviation $f = FF^{\top}$, $g = G^{\top}G$, we obtain

$$tr(\nabla(\det(X_{t-}))dX_{t}^{c}) = \det(X_{t-}) \bigg[2\sqrt{tr(f(t,X_{t-})X_{t-}^{-1}g(t,X_{t-})X_{t-}^{-1})} dW_{t-} + tr(H(t,X_{t-})X_{t-}^{-1})dt \bigg],$$

with some one-dimensional Brownian motion W on $[0, T_x)$, whose existence is guaranteed by Lemma 7.4.4. Furthermore, by Lemma 7.4.2 (ii), Lemma 7.4.3 and elementary calculations we have that

$$\begin{aligned} \frac{1}{2} \sum_{i,j,k,l} \frac{\partial^2}{\partial x_{ij} \partial x_{kl}} \det(X_{t-}) d[X_{ij}, X_{kl}]_t^c \\ &= \frac{\det(X_{t-})}{2} \sum_{i,j,k,l} \left[\left((X_{t-}^{-1})_{kl} (X_{t-}^{-1})_{ij} - (X_{t-}^{-1})_{il} (X_{t-}^{-1})_{jk} \right) \left(f(t, X_{t-})_{ik} g(t, X_{t-})_{jl} \right. \\ &+ f(t, X_{t-})_{il} g(t, X_{t-})_{jk} + f(t, X_{t-})_{jk} g(t, X_{t-})_{il} + f(t, X_{t-})_{jl} g(t, X_{t-})_{ik} \right) \right] \\ &= \det(X_{t-}) \left(\operatorname{tr}(f(t, X_{t-}) X_{t-}^{-1} g(t, X_{t-}) X_{t-}^{-1}) - \operatorname{tr}(f(t, X_{t-}) X_{t-}^{-1}) \operatorname{tr}(g(t, X_{t-}) X_{t-}^{-1}) \right) dt. \end{aligned}$$

According to Itô's formula, Lemma 7.4.5, we therefore obtain by summing up the two equations,

$$dZ_{t} = 2 \det(X_{t-}) \sqrt{\operatorname{tr}(f(t, X_{t-}) X_{t-}^{-1} g(t, X_{t-}) X_{t-}^{-1}) dW_{t} + \det(X_{t}) - \det(X_{t-})} + \det(X_{t-}) \left[\operatorname{tr}(H(t, X_{t-}) X_{t-}^{-1}) + \operatorname{tr}(f(t, X_{t-}) X_{t-}^{-1} g(t, X_{t-}) X_{t-}^{-1}) - \operatorname{tr}(f(t, X_{t-}) X_{t-}^{-1}) \operatorname{tr}(g(t, X_{t-}) X_{t-}^{-1}) \right] dt.$$

Using again Itô's formula, we have

$$dr_{t} = 2\sqrt{\operatorname{tr}(f(t, X_{t-})X_{t-}^{-1}g(t, X_{t-})X_{t-}^{-1})}dW_{t} + \ln(\operatorname{det}(X_{t})) - \ln(\operatorname{det}(X_{t-})) + \left[\operatorname{tr}(H(t, X_{t-})X_{t-}^{-1}) - \operatorname{tr}(f(t, X_{t-})X_{t-}^{-1}g(t, X_{t-})X_{t-}^{-1}) - \operatorname{tr}(f(t, X_{t-})X_{t-}^{-1})\operatorname{tr}(g(t, X_{t-})X_{t-}^{-1})\right]dt.$$

Hence, we have $r_t = r_0 + M_t + P_t$, where

$$\begin{split} M_t &= 2 \int_0^t \sqrt{\operatorname{tr}(f(s, X_{s-}) X_{s-}^{-1} g(s, X_{s-}) X_{s-}^{-1})} dW_s, \\ P_t &= \int_0^t \left[\operatorname{tr}(H(s, X_{s-}) X_{s-}^{-1}) - \operatorname{tr}(f(s, X_{s-}) X_{s-}^{-1} g(s, X_{s-}) X_{s-}^{-1}) \\ &- \operatorname{tr}(f(s, X_{s-}) X_{s-}^{-1}) \operatorname{tr}(g(s, X_{s-}) X_{s-}^{-1}) \right] ds + \sum_{0 < s \leq t} \left(\ln(\det(X_s)) - \ln(\det(X_{s-})) \right). \end{split}$$

We infer that $(M_t^{(n)})_{t \ge 0}$ defined by

$$M_t^{(n)} := 2 \int_0^t \sqrt{\operatorname{tr}(f(s, X_{s-}^{T_n})(X_{s-}^{T_n})^{-1}g(s, X_{s-}^{T_n})(X_{s-}^{T_n})^{-1})} dW_s$$

is a continuous martingale. Obviously, $M_t = M_t^{(n)}$ on $\{t < T_n\}$ and thus M is a continuous local martingale on $[0, T_x)$. Furthermore, $X_s - X_{s-} \ge 0$ for all $s \in [0, T)$ and hence $det(X_s) \ge det(X_{s-})$ using [64, Corollary 4.3.3]. Therefore, we have that $P_t \ge \int_0^t c(s) ds$ on $[0, T_x)$.

Finally, by Proposition 7.4.2 we have that $T_x = \infty$ a.s. noting that *c* is assumed to be locally integrable.

Remark 7.4.1. Bru's method for proving her proposition 7.3.1 for Wishart diffusions consists of the following two steps:

- (i) First assume $\beta = 0$. By applying the original McKean's argument twice, one derives that $h(\det(X))$ is a local martingale. This is proved separately for $\delta = d + 1$ and $\delta > d + 1$ by choosing $h(z) = \ln(z)$ in the first case and $h(z) = z^{d+1-\delta}$ in the second one. Therefore, the existence of a unique global strong solution on S_d^{++} is settled.
- (ii) One may therefore suppose that X_t is an S_d^{++} -valued solution on $[0, \infty)$ of

$$dX_t = \sqrt{X_t} dB_t Q + Q^{\mathsf{T}} dB_t^{\mathsf{T}} \sqrt{X_t} + \delta Q^{\mathsf{T}} Q dt, \quad X_0 = x \in S_d^{++}.$$

where $Q \in GL(d)$ and $\delta \ge d + 1$. Now, Girsanov's Theorem is applied which allows to introduce a drift by changing to an equivalent probability measure. This step generalises a one-dimensional method by Pitman and Yor, see [26, p. 748]. The involved arguments and calculations, which are not presented in detail in [26], appear rather complicated and work seemingly only in the special case given in Proposition 7.3.1 (ii), (iii).

The technical details of [26] concerning strong solutions are explained in more detail in [88].

Our proof above circumvented the problems associated to the use of Girsanov's theorem by extending the approach outlined in (i).

7.5. Conclusion

In this chapter we have extended the previously known sufficient boundary non-attainment conditions for certain Wishart processes to more general SDEs on S_d^{++} , which include affine diffusions with state-independent jumps of finite variation. This allowed to infer the existence of strong solutions of a large class of affine matrix valued processes. Moreover, we have thus obtained strong existence results for SDEs which can be considered as positive semidefinite extensions of GARCH diffusions and generalised Cox-Ingersoll-Ross processes.

However, this results in several open questions related to our SDE (7.1.1) which will hopefully be addressed in future work. The following questions are beyond the scope of the present chapter, since they are obviously rather non-trivial and apparently need very different techniques than the ones employed here. For d > 1 and the Wishart diffusions it is not clear, whether the condition $b \ge (d+1)Q^{\top}Q$ for the drift is a necessary non-attainability condition or not. Only in the case $\beta = 0, \Gamma = 0, Q = I_d$ and $b = \delta I_d$ with $\delta \in (d-1, d+1)$ it is known from [42, Theorem 1.4] that the boundary is hit. On the other hand, one knows that in the case d = 1 pathwise uniqueness holds, hence there exists a strong solution for all $b \ge 0$ (even in the general setting of CBI processes, see [40, Theorem 5.1]). For $d \ge 2$, the situation seems in general to be rather complicated and therefore existence of global strong solutions remains an open problem when $b \not\ge (d+1)Q^{\top}Q$ (and the conditions for the existence of weak solutions of [32] are satisfied). Likewise, it is a very interesting problem in the case of the GCIR processes with $\alpha > 1/2$ whether a state dependent drift away from the boundary is really necessary and what happens if one has only a constant drift towards the interior of S_d^+ .

Finally, we remark that our method of proof could be generalised to state-spaces D other than S_d^+ , as long as the existence of an appropriate function $g : D \to \mathbb{R}_+$ is guaranteed, such that $g^{-1}(0) = \partial D$. For instance, similar (but simpler) arguments to the ones of the proof of Theorem 7.1 yield a rigorous proof of the non-attainment condition formulated in [31, Section 6] for affine jump diffusions on the canonical state space $\mathbb{R}^m_+ \times \mathbb{R}^n$. Here one takes $g(x_1, x_2, \ldots, x_m) = x_1 \cdot x_2 \cdot \cdots \cdot x_m$.

Bibliography

- A. Ahdida and A. Alfonsi. Exact and high order discretization schemes for Wishart processes and their affine extensions. Preprint, 2010. available at: http://adsabs. harvard.edu/abs/2010arXiv1006.2281A.
- [2] G. W. Anderson, A. Guionnet, and O. Zeitouni. An introduction to random matrices, volume 118 of Cambridge Studies in Advanced Mathematics. Cambridge University Press, Cambridge, 2010.
- [3] G. W. Anderson and O. Zeitouni. A law of large numbers for finite-range dependent random matrices. *Comm. Pure Appl. Math.*, 61(8):1118–1154, 2008.
- [4] T. W. Anderson. Asymptotic theory for principal component analysis. *Ann. Math. Statist.*, 34:122–148, 1963.
- [5] T. W. Anderson. *An introduction to multivariate statistical analysis*. Wiley Series in Probability and Statistics. Wiley-Interscience, Hoboken, third edition, 2003.
- [6] G. Aubrun. Random points in the unit ball of l_p^n . Positivity, 10(4):755–759, 2006.
- [7] A. Auffinger, G. Ben Arous, and S. Péché. Poisson convergence for the largest eigenvalues of heavy tailed random matrices. *Ann. Inst. Henri Poincaré Probab. Stat.*, 45(3):589–610, 2009.
- [8] Z. Bai, D. Jiang, J.-F. Yao, and S. Zheng. Corrections to LRT on large-dimensional covariance matrix by RMT. Ann. Statist., 37(6B):3822–3840, 2009.
- [9] Z. Bai and J. W. Silverstein. *Spectral analysis of large dimensional random matrices*. Springer Series in Statistics. Springer, New York, second edition, 2010.

- [10] Z. Bai and W. Zhou. Large sample covariance matrices without independence structures in columns. *Statist. Sinica*, 18(2):425–442, 2008.
- [11] O. Barndorff-Nielsen and V. Pérez-Abreu. Matrix Subordinators and related Upsilon Transformations. *Theory Probab. Appl.*, 52:1–23, 2008.
- [12] O. E. Barndorff-Nielsen and N. Shephard. Non-Gaussian Ornstein-Uhlenbeck-based models and some of their uses in Financial Economics (with discussion). J. R. Stat. Soc. B Statist. Methodol., 63:167–241, 2001.
- [13] O. E. Barndorff-Nielsen and R. Stelzer. Positive-Definite Matrix Processes of Finite Variation. *Probab. Math. Statist.*, 27:3–43, 2007.
- [14] S. Belinschi, A. Dembo, and A. Guionnet. Spectral measure of heavy tailed band and covariance random matrices. *Comm. Math. Phys.*, 289(3):1023–1055, 2009.
- [15] R. Bhatia. *Matrix analysis*, volume 169 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, 1997.
- [16] K. Bichteler. *Stochastic Integration with Jumps*, volume 89 of *Encyclopedia of Mathematics and its Applications*. Cambridge University Press, Cambridge, 2002.
- [17] P. Billingsley. Convergence of probability measures. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons Inc., New York, second edition, 1999.
- [18] G. Biroli, J.-P. Bouchaud, and M. Potters. On the top eigenvalue of heavy-tailed random matrices. *Europhys. Lett. EPL*, 78(1):Art. 10001, 5, 2007.
- [19] A. Bloemendal and B. Virág. Limits of spiked random matrices I. http://arxiv. org/abs/1011.1877.
- [20] M. Borkovec and C. Klüppelberg. Extremal behavior of diffusion models in finance. *Extremes*, 1:47–80, 1998.
- [21] A. Bose, R. S. Hazra, and K. Saha. Limiting spectral distribution of circulant type matrices with dependent inputs. *Electron. J. Probab.*, 14:2463–2491, 2009.
- [22] A. Böttcher and B. Silbermann. *Introduction to large truncated Toeplitz matrices*. Springer-Verlag, New York., 1999.
- [23] J. P. Bouchaud and M. Potters. Financial Applications of Random Matrix Theory: a short review. 2009. arXiv:0910.1205v1.

- [24] P. J. Brockwell and R. A. Davis. *Time series: theory and methods*. Springer Series in Statistics. Springer-Verlag, New York, second edition, 1991.
- [25] M.-F. Bru. Diffusions of Perturbed Principal Component Analysis. J. Multivariate Anal., 29:127–136, 1989.
- [26] M.-F. Bru. Wishart Processes. J. Theoret. Probab., 4:725–751, 1991.
- [27] W. Bryc, A. Dembo, and T. Jiang. Spectral measure of large random Hankel, Markov and Toeplitz matrices. *Ann. Probab.*, 34(1):1–38, 2006.
- [28] A. Buraschi, A. Cieslak, and F. Trojani. Correlation Risk and the Term Structure of Interest Rates. Working paper, Imperial College, UK, 2007.
- [29] A. Buraschi, P. Porchia, and F. Trojani. Correlation Risk and Optimal Portfolio Choice. J. Finance, 65 (1):393–420, 2010.
- [30] Z. Burda, A. Jarosz, M. A. Nowak, and M. Snarska. A random matrix approach to VARMA processes. *New Journal of Physics*, 12:075036, 2010.
- [31] P. Cheridito, D. Filipović, and M. Yor. Equivalent and Absolutely Continuous Measure Changes for Jump-Diffusion Processes. *Ann. Appl. Probab.*, 15:1713–1732, 2005.
- [32] C. Cuchiero, D. Filipović, E. Mayerhofer, and J. Teichmann. Affine Processes on Positive Semidefinite Matrices. Ann. Appl. Probab., 2010. to appear.
- [33] J. Da Fonseca, M. Grasseli, and F. Ielpo. Estimating the Wishart Affine Stochastic Correlation Model using the Empirical Characteristic Function. Working paper, 2008.
- [34] J. Da Fonseca, M. Grasseli, and F. Ielpo. Hedging (Co)Variance Risk with Variance Swaps. Working paper, University St.Gallen, 2008.
- [35] J. Da Fonseca, M. Grasseli, and C. Tebaldi. Option Pricing when Correlations are Stochastic: An Analytical Framework. *Review of Derivatives Research*, 10:151–180, 2007.
- [36] J. Da Fonseca, M. Grasseli, and C. Tebaldi. A Multifactor Volatility Heston model. J. *Quant. Finance*, 8:591–604, 2008.
- [37] R. Davis and T. Hsing. Point process and partial sum convergence for weakly dependent random variables with infinite variance. *The Annals of Probability*, 23(2):879–917, 1995.
- [38] R. Davis, O. Pfaffel, and R. Stelzer. Limit theory for the largest eigenvalues of sample covariance matrices with heavy-tails. 2011.

- [39] R. A. Davis and S. I. Resnick. Limit theory for moving averages of random variables with regularly varying tail probabilities. *Ann. Probab.*, 13(1):179–195, 1985.
- [40] D. A. Dawson and Z. Li. Skew Convolution Semigroups and Affine Markov processes. *Ann. Probab.*, 34:1103–1142, 2006.
- [41] C. Donati-Martin. Large Deviations for Wishart Processes. *Probab. Math. Statist.*, 28:325–343, 2008.
- [42] C. Donati-Martin, Y. Doumerc, H. Matsumoto, and M. Yor. Some Properties of the Wishart Processes and a Matrix Extension of the Hartman-Watson Laws. *Publ. Res. Inst. Math. Sci.*, 40:1385–1412, 2004.
- [43] D. Duffie, D. Filipović, and W. Schachermayer. Affine Processes and Applications in Finance. Ann. Appl. Probab., 13:984–1053, 2003.
- [44] F. J. Dyson. Statistical theory of the energy levels of complex systems. I. J. Mathematical Phys., 3:140–156, 1962.
- [45] N. El Karoui. Recent results about the largest eigenvalue of random covariance matrices and statistical application. *Acta Phys. Polon. B*, 36(9):2681–2697, 2005.
- [46] N. El Karoui. Spectrum estimation for large dimensional covariance matrices using random matrix theory. *Ann. Statist.*, 36(6):2757–2790, 2008.
- [47] P. Embrechts and C. M. Goldie. On closure and factorization properties of subexponential and related distributions. J. Austral. Math. Soc. Ser. A, 29(2):243–256, 1980.
- [48] L. Erdös. Universality of Wigner random matrices: a survey of recent results. Uspekhi Mat. Nauk, 66(3(399)):67–198, 2011.
- [49] V. Fasen, C. Klüppelberg, and A. Lindner. Extremal behavior of stochastic volatility models. In A. N. Shiryaev, M. R. Grossinho, P. E. Oliveira, and M. L. Esquível, editors, *Stochastic finance*, pages 107–155, New York, 2006. Springer.
- [50] D. Filipović and E. Mayerhofer. Affine Diffusion Processes: Theory and Applications. In Advanced Financial Modelling, volume 8 of Radon Ser. Comput. Appl. Math. Walter de Gruyter, Berlin, 2009.
- [51] S. Geman. A limit theorem for the norm of random matrices. *Ann. Probab.*, 8(2):252–261, 1980.
- [52] S. Gerschgorin. Über die Abgrenzung der Eigenwerte einer Matrix. *Izv. Akad. Nauk. USSR Otd. Fiz.-Mat. Nauk*, 7:749–754, 1931.

- [53] C. Gourieroux and R. Sufana. Wishart Quadratic Term Structure. Working paper, CREST, CEPREMAP and University of Toronto, 2007.
- [54] C. Gourieroux and R. Sufana. Derivative Pricing with Wishart Multivariate Stochastic Volatility. J. Bus. Econom. Statist., 28:438–451, 2010.
- [55] P. Graczyk and L. Vostrikova. The Moments of Wishart Processes via Itô calculus. *Theory Probab. Appl.*, 51:609–625, 2007.
- [56] C. W. J. Granger and R. Joyeux. An introduction to long-memory time series models and fractional differencing. J. Time Ser. Anal., 1(1):15–29, 1980.
- [57] M. Grasselli and C. Tebaldi. Solvable Affine Term Structure Models. *Math. Finance*, 18(1):135–153, 2008.
- [58] R. M. Gray. Toeplitz and circulant matrices: A review. Now Publishers, Boston, 2006.
- [59] U. Grenander and G. Szegő. *Toeplitz forms and their applications*. Chelsea Publishing, New York, second edition, 1984.
- [60] W. Hachem, P. Loubaton, and J. Najim. The empirical eigenvalue distribution of a Gram matrix: from independence to stationarity. *Markov Process. Related Fields*, 11(4):629– 648, 2005.
- [61] W. Hachem, P. Loubaton, and J. Najim. The empirical distribution of the eigenvalues of a Gram matrix with a given variance profile. *Ann. Inst. H. Poincaré Probab. Statist.*, 42(6):649–670, 2006.
- [62] C. C. Heyde. On large deviation probabilities in the case of attraction to a non-normal stable law. *Sankhyā Ser. A*, 30:253–258, 1968.
- [63] K. Hofmann-Credner and M. Stolz. Wigner theorems for random matrices with dependent entries: ensembles associated to symmetric spaces and sample covariance matrices. *Electron. Commun. Probab.*, 13:401–414, 2008.
- [64] R. A. Horn and C. R. Johnson. *Matrix Analysis*. Cambridge University Press, Cambridge, 1985.
- [65] R. A. Horn and C. R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, Cambridge, 1991.
- [66] J. R. M. Hosking. Fractional differencing. *Biometrika*, 68(1):165–176, 1981.

- [67] J. Jacod and A. N. Shiryaev. *Limit theorems for stochastic processes*, volume 288 of *Fundamental Principles of Mathematical Sciences*. Springer-Verlag, Berlin, second edition, 2003.
- [68] A. T. James. Distributions of matrix variates and latent roots derived from normal samples. *Ann. Math. Statist.*, 35:475–501, 1964.
- [69] I. M. Johnstone. On the distribution of the largest eigenvalue in principal components analysis. *Ann. Statist.*, 29(2):295–327, 2001.
- [70] I. M. Johnstone. High dimensional statistical inference and random matrices. In *Inter-national Congress of Mathematicians. Vol. I*, pages 307–333. Eur. Math. Soc., Zürich, 2007.
- [71] I. Jolliffe. Principal Component Analysis. Springer, 2002.
- [72] O. Kallenberg. Random measures. Akademie-Verlag, Berlin, third edition, 1983.
- [73] M. Leippold and F. Trojani. Asset Pricing with Matrix Affine Jump Diffusions. Working paper, 2010. available from: http://papers.ssrn.com/sol3/papers.cfm? abstract_id=1274482.
- [74] N. T. Longford. Random coefficient models, volume 11 of Oxford Statistical Science Series. Oxford University Press, New York, 1993.
- [75] J. R. Magnus and H. Neudecker. *Matrix differential calculus with applications in statistics and econometrics*. Wiley, Chichester, 1988.
- [76] B. Maisonneuve. Une mise au point sur les martingales locales continues définies sur un intervalle stochastique. In *Séminaire de Probabilités, XI (Univ. Strasbourg, Strasbourg, 1975/1976)*, volume 581 of *Lecture Notes in Math.*, pages 435–445, Berlin, 1977. Springer.
- [77] V. A. Marčenko and L. A. Pastur. Distribution of eigenvalues in certain sets of random matrices. *Mat. Sb.* (*N.S.*), 72 (114):507–536, 1967.
- [78] E. Mayerhofer, O. Pfaffel, and R. Stelzer. On strong solutions for positive definite jump diffusions. *Stochastic Process. Appl.*, 121(9):2072–2086, 2011.
- [79] H. P. McKean. Stochastic Integrals. Academic Press, 1969.
- [80] M. W. Meckes. On the spectral norm of a random Toeplitz matrix. *Electron. Comm. Probab.*, 12:315–325, 2007.

- [81] M. L. Mehta. *Random matrices*, volume 142 of *Pure and Applied Mathematics (Ams-terdam)*. Elsevier/Academic Press, Amsterdam, third edition, 2004.
- [82] M. Métivier and J. Pellaumail. *Stochastic Integration*. Academic Press, New York, 1980.
- [83] S. Meyn and R. L. Tweedie. *Markov chains and stochastic stability*. Cambridge University Press, Cambridge, second edition, 2009.
- [84] D. B. Nelson. ARCH models as diffusion approximations. J. Econometrics, 45:7–38, 1990.
- [85] J. R. Norris, L. C. G. Rogers, and D. Williams. Brownian Motions of Ellipsoids. *Trans. Amer. Math. Soc.*, 294:757–765, 1986.
- [86] G. Pan. Strong convergence of the empirical distribution of eigenvalues of sample covariance matrices with a perturbation matrix. *J. Multivariate Anal.*, 101(6):1330–1338, 2010.
- [87] S. Péché. Universality results for the largest eigenvalues of some sample covariance matrix ensembles. *Probab. Theory Related Fields*, 143(3-4):481–516, 2009.
- [88] O. Pfaffel. Wishart Processes. Master's thesis, Zentrum Mathematik, Technische Universität München, 2012. http://arxiv.org/abs/1201.3256.
- [89] O. Pfaffel and E. Schlemm. Eigenvalue distribution of large sample covariance matrices of linear processes. *Probab. Math. Statist.*, 31 (2):313–329, 2011.
- [90] O. Pfaffel and E. Schlemm. Limiting spectral distribution of a new random matrix model with dependence across rows and columns. *Linear Algebra and Its Applications*, 436(8):2960–2973, 2012.
- [91] C. Pigorsch and R. Stelzer. A Multivariate Ornstein-Uhlenbeck Type Stochastic Volatility Model. Submitted for publication, 2009.
- [92] C. Pigorsch and R. Stelzer. On the Definition, Stationary Distribution and Second Order Structure of Positive Semi-definite Ornstein-Uhlenbeck type Processes. *Bernoulli*, 15:754–773, 2009.
- [93] V. Plerou, P. Gopikrishnan, B. Rosenow, L. A. N. Amaral, T. Guhr, and H. E. Stanley. Random matrix approach to cross correlations in financial data. *Physical Review E*, 65(6):66126, 2002.

- [94] P. E. Protter. Stochastic Integration and Differential Equations, volume 21 of Stochastic Modelling and Applied Probability. Springer-Verlag, Berlin, 2005. Second edition. Version 2.1, Corrected third printing.
- [95] R. Rashidi Far, T. Oraby, W. Bryc, and R. Speicher. On slow-fading MIMO systems with nonseparable correlation. *IEEE Trans. Inform. Theory*, 54(2):544–553, 2008.
- [96] S. I. Resnick. *Extreme values, regular variation and point processes*. Springer Series in Operations Research and Financial Engineering. Springer, New York, 2008. Reprint of the 1987 original.
- [97] D. Revuz and M. Yor. Continuous Martingales and Brownian motion, volume 293 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]. Springer-Verlag, Berlin, 1991.
- [98] J. W. Silverstein and Z. D. Bai. On the empirical distribution of eigenvalues of a class of large-dimensional random matrices. *J. Multivariate Anal.*, 54(2):175–192, 1995.
- [99] A. Soshnikov. A note on universality of the distribution of the largest eigenvalues in certain sample covariance matrices. J. Statist. Phys., 108(5-6):1033–1056, 2002.
- [100] A. Soshnikov. Poisson statistics for the largest eigenvalues in random matrix ensembles. In *Mathematical physics of quantum mechanics*, volume 690 of *Lecture Notes in Phys.*, pages 351–364. Springer, Berlin, 2006.
- [101] R. Stelzer. Multivariate COGARCH(1,1) Processes. Bernoulli, 16:80–115, 2010.
- [102] R. J. Stelzer. Multivariate Continuous Time Stochastic Volatility Models Driven by a Lévy Process. PhD thesis, Centre for Mathematical Sciences, Munich University of Technology, Garching, Germany, 2007. available from: http://mediatum2.ub.tum.de/node?id=624065.
- [103] G. Szegő. Beiträge zur theorie der toeplitzschen formen. *Math. Z.*, 6(3-4):167–202, 1920.
- [104] T. Tao and V. Vu. Random matrices: universality of ESDs and the circular law. *Ann. Probab.*, 38(5):2023–2065, 2010.
- [105] T. Tao and V. Vu. Random matrices: Universal properties of eigenvectors. *arXiv:1103.2801v2*, 2011.
- [106] T. Tao and V. Vu. Random matrices: universality of local eigenvalue statistics. Acta Math., 206(1):127–204, 2011.

- [107] C. A. Tracy and H. Widom. Level-spacing distributions and the Airy kernel. Comm. Math. Phys., 159(1):151–174, 1994.
- [108] C. A. Tracy and H. Widom. Distribution functions for largest eigenvalues and their applications. In *Proceedings of the International Congress of Mathematicians, Vol. I* (*Beijing, 2002*), pages 587–596, Beijing, 2002. Higher Ed. Press.
- [109] A. M. Tulino and S. Verdu. *Random Matrix Theory and Wireless Communications*. Now Publishers, Boston, 2004.
- [110] K. W. Wachter. The strong limits of random matrix spectra for sample matrices of independent elements. *Ann. Probability*, 6(1):1–18, 1978.
- [111] S. Watanabe and T. Yamada. On the Uniqueness of Solutions of Stochastic Differential Equations II. *J. Math. Kyoto Univ.*, 11:553–563, 1971.
- [112] E. P. Wigner. On the distribution of the roots of certain symmetric matrices. *Ann. of Math.* (2), 67:325–327, 1958.
- [113] J. Wishart. The generalised product moment distribution in samples from a normal multivariate population. *Biometrika*, 20A:32–52, 1928.
- [114] Y. Q. Yin. Limiting spectral distribution for a class of random matrices. *J. Multivariate Anal.*, 20(1):50–68, 1986.
- [115] L. Zhang. *Spectral Analysis of large dimensional random matrices*. PhD thesis, National University of Singapore, 2006.