Fourier Analysis of Functional Time Series with Applications to DNA Dynamics

THIS IS A TEMPORARY TITLE PAGE
It will be replaced for the final print by a version
provided by the service academique.

Thèse n. 6320 (2014)
présenté le 16 Juillet 2014
à la Faculté des Sciences de Base
Chaire de Statistique Mathématique
programme doctoral en Mathématiques
École Polytechnique Fédérale de Lausanne

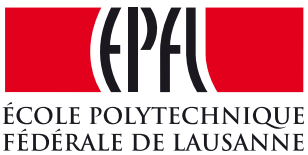
pour l'obtention du grade de Docteur ès Sciences
par

Shahin Tavakoli

acceptée sur proposition du jury:

Prof T. Mountford, président du jury
Prof V. Panaretos, directeur de thèse
Prof J. Aston, rapporteur
Prof A. C. Davison, rapporteur
Prof H.-G. Müller, rapporteur

Lausanne, EPFL, 2014



Stay close to anything that makes you glad you are alive.
— Hafez

To Baba-Shirin

Acknowledgements

I would like to thank Victor for his subtle and effective advice during these four years, sharing and teaching me his understanding and intuition in mathematics and statistics, and for his style of mentoring, from which I have learned a lot.
I would like also to thank Professors Anthony C. Davison and Stefan Morgenthaler for their kind advice and interesting discussions. I thank in advance Professor Anthony C. Davison, Professor John Aston and Professor Hans-Georg Müller for having accepted to be part of my thesis committee, and Professor Thomas Mountford for having accepted to be the president of the jury.
My thanks also go to my (current and ex-)colleagues, with whom I have had the chance to interact, both professionally and socially: Andrea, Yoav, Marie-Hélène, Mikael, Zhigang, Tung, David, Kjell, Valentina, Susan, Alix, Linda, Rafaël, Emeric, Claudio, and all the others that I may have forgotten.
I am grateful to Professors David Brillinger, Tony Cai, Soumendra Lahiri, Terry Speed, Aurore Delaigle, Rainer Dahlhaus, and Tilmann Gneiting, for giving me research or personal advice, and to the Heidelberg Laureate Forum, to which I had the chance to participate.
Last but not least, I would like to thank my family, Shirin, Marc, Farhad and Mina, for all the support they have given me, and specially thank Talissa for putting up with my strong character, and comforting me when I was too stressed.

Lausanne, August 16, 2014. Shahin Tavakoli

Preface

I have always been passionate about mathematics, and since more recently statistics. During these past four years, and specially during these last six weeks during which I wrote this thesis, I have learned that abstraction can help you a lot when confronted with a roadblock, but that it is always possible to take a further step in abstraction. In the past few weeks, I have spent a significant amount of time and effort trying to generalize some of the earlier results I had obtained. I got lost a couple of times into the spiral of abstraction, and sometimes my efforts were unfruitful. However, I was lucky enough to find answers to my questions at times. As a result, some statements are more general than versions published in our papers (Panaretos & Tavakoli 2013*a*,*b*), but might be slightly obscured by technical conditions; when this is the case, I have made an effort to add a remark or an explanation, which I hope will be helpful to the reader. A list of notation and an index have been added for helping the reading with the plethora of symbols (and different norms).

Lausanne, August 16, 2014 S.T.

Abstract

This work is about time series of functional data (*functional time series*), and consists of three main parts. In the first part (Chapter 2), we develop a doubly spectral decomposition for functional time series that generalizes the Karhunen–Loève expansion. In the second part (Chapter 3), we develop the theory of estimation for the spectral density operators, which are the main tool involved in the doubly spectral decomposition. The third part (Chapter 4) is concerned with the problem of understanding and comparing the dynamics of DNA. It proposes a methodology for comparing the dynamics of DNA minicircles that are vibrating in solution, using tools developed in this thesis.

In the first part, we develop a doubly spectral representation of a stationary functional time series that generalizes the Karhunen–Loève expansion to the functional time series setting. The representation decomposes the time series into an integral of uncorrelated frequency components (Cramér representation), each of which is in turn expanded in a Karhunen-Loève series, thus yielding a *Cramér–Karhunen–Loève* decomposition of the series. The construction is based on the spectral density operators—whose Fourier coefficients are the lag-t autocovariance operators—which characterise the second-order dynamics of the process. The spectral density operators are the functional analogues of the spectral density matrices, whose eigenvalues and eigenfunctions at different frequencies provide the building blocks of the representation. By truncating the representation at a finite level, we obtain a harmonic principal component analysis of the time series, an optimal finite dimensional reduction of the time series that captures both the temporal dynamics of the process, and the within-curve dynamics, and dominates functional PCA. The proofs rely on the construction of a stochastic integral of operator-valued functions, whose construction is similar to that of the Itō integral.

In practice, the spectral density operators are unknown. In the second part, we therefore develop the basic theory of a frequency domain framework for drawing statistical inferences on the spectral density operators of a stationary functional time series. Our main tool is the functional Discrete Fourier Transform (fDFT). We derive an asymptotic Gaussian representation of the fDFT, thus allowing the transformation of the original collection of dependent random functions into a collection of approximately independent complex-valued Gaussian random functions. Our results are then employed in order to construct estimators of the spectral density operators based on smoothed versions of the periodogram kernel, the functional generalisation of the periodogram matrix. The consistency and asymptotic law of these estimators are studied in detail. As immediate consequences, we obtain central limit theorems for the mean and the long-run covariance operator of a

stationary functional time series. Our results do not depend on structural modeling assumptions, but only functional versions of classical cumulant mixing conditions. The effect of discrete noisy observations on the consistency of the estimators is studied in a framework general enough to apply to a wide range of smoothing techniques for converting discrete noisy observations into functional data. We also perform a simulation study to assess the finite sample performance of our estimators, and give a discussion of the technical assumptions of our results, and at what cost our weak dependence assumptions could be changed or weakened, and provide examples of processes satisfying the technical assumptions of our asymptotic results.

As an application, we consider in the third part the problem of comparing the dynamics of the trajectories of two DNA minicircles that are vibrating in solution, which are obtained via Molecular Dynamics simulations. The approach we take is to view and compare the dynamics through their spectral density operators, which contain the entire second-order structure of the trajectories. As a first step, we compare the spectral density operators of the two DNA minicircles using a new test we develop, which allows us to compare the spectral density operators at a fixed frequencies. Using multiple testing procedures, we are able to localize in frequencies the differences in spectral density operators of the two DNA minicircles, while controlling a type-I error, and conduct numerical simulations to assess the performance of our method. We further investigate the differences between the two minicircles by comparing their spectral density operators within frequencies. This allows us to localize their differences both in frequencies and on the minicircles, while controlling the averaged false discovery rate over the selected frequencies. Our methodology is general enough to be applied to the comparison of the dynamics of any pair of stationary functional time series.

Keywords: functional data, Cramér representation, Karhunen–Loève expansion, discrete Fourier transform, periodogram, spectrum, spectral density, spectral density operator, DNA minicircle, mixing, weak dependence, cumulants, multiple testing, FDR.

Résumé

Ce travail porte sur des séries temporelles de données fonctionnelles (*séries temporelle fonctionnelle*), et est consitué de trois parties principales. Dans la première partie (Chapitre 2), nous développons une double décomposition spectrale pour les séries temporelles fonctionnelle qui généralise l'expansion Karhunen–Loève. Dans la deuxième partie (Chapitre 3), nous développons la théorie de l'estimation des opérateurs de densité spectrale, qui sont les outils principaux impliqués dans la double décomposition spectrale. La troisième partie (chapitre 4) s'adresse a la compréhension et a la comparaison de la dynamique de l'ADN. Nous proposons une méthode pour comparer la dynamique de minicercles d'ADN qui vibrent dans un fluide, en utilisant les outils développés dans cette thèse.

Dans la première partie, nous développons une représentation doublement spectrale d'une série temporelle stationnaire fonctionnelle qui généralise la décomposition de Karhunen–Loève au contexte des séries temporelles fonctionnelles. La représentation consiste en une décomposition de la série temporelle en une intégrale de composantes de différentes fréquences, qui sont non-corrélées (représentation de Cramér), dont chacun est à son tour décomposée a l'aide de la décomposition de Karhunen–Loève, donnant ainsi la décomposition de *CKL*. La construction s'appuie sur les opérat de densité spectrale — dont les coefficients de Fourier sont les opérateurs d'autocovariance de la série temporelle — qui caractérise la dynamique de second ordre de la série. Ce sont les valeurs propres et les vecteurs propres des opérateurs de densité spectrale qui sont les éléments principaux de la construction de notre représentation. En tronquant la représentation à un niveau fini, on obtient une analyse harmonique en composantes principales de la série temporelle, une approximation optimale de dimension finie de la série temporelle qui reflète à la fois la dynamique temporelle du processus, ainsi que la dynamique interne de la fonction, et qui domine l'analyse en composantes principales. Les preuves reposent sur la construction d'une intégrale stochastique de fonctions a valeurs dans des espaces d'opérateurs dont la construction est similaire à celle de l'intégrale d'Itō.

En pratique, les opérateurs de densité spectrale sont inconnus. Nous développons donc dans la deuxième partie de cette thèse la théorie de base d'une analyse de séries temporelles fonctionnelles stationnaires par une approche spectrale. Notre outil principal est la transformée de Fourier discrète, (fDFT). Nous obtenons une représentation gaussienne asymptotique de la fDFT, permettant ainsi la transformation de la collection originale de fonctions aléatoires dépendantes en une collection de fonctions aléatoires complexes approximativement indépendantes et gaussiennes. Nos

résultats sont ensuite utilisés pour construire des estimateurs des opérateurs de densité spectrale basés sur des versions lissées du périodogramme, le généralisation fonctionnelle de la matrice de périodogramme. La consistance et la loi asymptotique de ces estimateurs sont étudiés en détail. Comme corollaires immédiats, nous obtenons des théorèmes centraux limites pour la moyenne et l'opérateur de covariance à long terme d'une série de temporelle fonctionnelle stationnaire. Nos résultats ne dépendent pas de la structure de la séries temporelle, mais seulement de versions fonctionnelles des conditions de sommabilité des cumulants. L'effet de observations bruitées discrets sur la consistance des estimateurs est étudié dans un cadre général assez pour s'appliquer à un large éventail de techniques de lissage pour convertir des observations bruitées discrètes en données fonctionnelles. Nous effectuons également une étude de simulation pour évaluer la performance de nos estimateurs avec un échantillon fini, et donnons une discussion sur les hypothèses techniques de nos résultats, et comment elles pourraient être modifiées ou affaiblies. Nous fournissons des exemples de processus satisfaisant les hypothèses techniques de nos résultats asymptotiques.

Comme application, nous considérons dans la troisième partie le problème de la comparaison de la dynamique des trajectoires des deux minicercles d'ADN qui vibrent dans un liquide, et qui sont obtenus par simulations de dynamique moléculaire. Notre approche consiste a comparer la dynamique de ces minicercles d'ADN a travers leur spectre, qui encode toute la structure de second order de leur trajectoire. Dans un premier temps, nous comparons les spectres des deux minicercles d'ADN l'aide d'un nouveau test nous avons développé, qui nous permet de comparer les opérateurs de densité spectrale à des fréquences fixes. En utilisant des procédures de tests multiples, nous sommes en mesure de localiser dans les fréquences les différences spectres des deux minicercles d'ADN, tout en contrôlant une erreur de type I, et procédons à des simulations numériques pour évaluer la performance de notre méthode. Nous étudions de plus les différences entre les deux minicercles en comparant leurs opérateurs de densité spectrale a l'intérieur des fréquences. Cela nous permet de localiser leurs différences à la fois dans les fréquences et sur les minicercles, tout en contrôlant le taux moyen de fausses découvertes sur les fréquences sélectionnées. Notre méthodologie est assez générale pour être appliqué à la comparaison de la dynamique de n'importe quelle paire de séries temporelles fonctionnelle stationnaire.

Mots clefs : données fonctionnelles, représentation de Cramér, expansion de Karhunen–Loève, transformée de Fourier discrete, periodogramme, spectre, opérateur de densité spectrale, minicercles d'ADN, coefficients de mixage, dépendance faible, cumulants, tests multiplies, FDR.

Contents

List of Figures

List of Tables

List of Notation

$\delta_{n,m}$ the Kronecker delta, page 9

$\mathscr{S}_2(H)$ Space of Hilbert–Schmidt operators on H, page 220

$|||\cdot|||_\infty$ The operator norm, page 213

$|||\cdot|||_2$ The Hilbert–Schmidt norm, page 220

$|||\cdot|||_1$ the trace (or nuclear) norm, page 219

$\mathrm{Tr}(\cdot)$ The trace functional, page 220

$H^{\otimes k}$ The k-fold tensor product of H, page 224

càdlàg Continue à droite, avec limite à gauche, page 35

C(l,k) Weak-dependence condition, page 71

BH Benjamini-Hochberg, page 171

bp(s) base-pair(s), page 156

CLT Central Limit Theorem, page 15

DFT Discrete Fourier Transform, page 73

FAR Functional Autoregressive Model, page 15

fDFT functional Discrete Fourier Transform, page 73

FDP false discovery proportion, page 180

FLM Functional Linear Model, page 14

FLP Functional Linear Process, page 15

fPCA functional Principal Component Analysis, page 9

FTS Functional Time Series, page 14

FWER Family-wise error rate, page 170

IMSE Integrated Mean Square Error, page 90

KL Karhunen–Loève , page 23

MD Molecular Dynamics, page 154

PCA Principal Component Analysis, page 9

Introduction and Overview

David Brillinger's aphorism "*you want to be thinking of data as anything that can be mathematically expressed*" is nowadays extremely pertinent (Brillinger 2012). Partly due to technological advances over the last decades, an increasing number of modern datasets can be thought of as being sampled recordings of complex mathematical structures, such as smooth curves or surfaces. Such data are called *functional data*, and examples of datasets include growth curves, temperature curves, electricity consumption curves, gait cycle data, density functions, speech recordings, brain images, or DNA minicircles vibrating in solution (see Chapter 4 of this thesis, Ramsay & Silverman (2005), Ferraty (2011)).

From a statistical point of view, many of these datasets can be modeled as independent and identically distributed (i.i.d.) realizations of an underlying random object, such as a random curve. However, many other types of data have an intrinsic dependency structure that distinguishes them from the i.i.d. setting, and prohibits the use of "i.i.d. technology" without further justification. A particular type of dependency is given by functional time series, in which it is assumed that the collected curves $x_1, \dots, x_T$ correspond to a realization of a sequence of random curves $X_1, \dots, X_T$, where X_t represents the state of some random curve X at time t. Intuitively speaking, the dependency structure is given by the "memory" that the process $(X_t)_{t\in\mathbb{Z}}$ keeps of its preceding states, and its influence on the current configuration of the process.

The object of this thesis is the study of time series of functional data, or *functional time series*, through a frequency approach. This involves the study of the series $(X_t)_{t\in\mathbb{Z}}$ not in the (natural) time domain indexed by the time $t \in \mathbb{Z}$, but in a transformed space indexed by frequencies $\omega \in [-\pi, \pi]$. The tools used for this approach come from Fourier analysis (Fourier 1822, Edwards 1967), which explains the first part of the title of this thesis.

The advantages of taking a frequency domain approach to study functional time series are manifold (and will be described below), but perhaps the most appealing result is that the *functional discrete Fourier transform* transforms (bijectively) a dependent set of random curves $(X_t)_{t=1}^T$ into a set of complex valued random curves that are asymptotically *independent* and *Gaussian*—in other words it transforms *dependent* data into asymptotically *independent Gaussian* data!

Here is a short description of the chapters of the thesis, followed by a detailed overview. The thesis start with a compact overview of functional data analysis (Chapter 1), followed by the three chapters that are the main contributions of this thesis: Chapter 2, where we develop a doubly spectral decomposition for functional time series that generalizes the Karhunen–Loève expansion. The theory of estimation for the main tool involved this doubly spectral decomposition (the *spectral density*

operators), will be developed in Chapter 3. In Chapter 4, we study the problem of understanding the mechanics of DNA, and propose a methodology for comparing the dynamics of DNA minicircles that are vibrating in solution, using tools developed in Chapters 2 and 3. Chapters A, B and C are technical appendices presenting some of the background theory used in the thesis.

Detailed Overview of the Thesis

Chapter 1 of the thesis gives an overview of functional data analysis (FDA), and presents its major tool, the Karhunen–Loève expansion (Propositions 1.2.2 and 1.2.3). Then the basic theory for inference of the first and second-order structure of functional data in the i.i.d. case is presented in Section 1.3, where the strong law of large numbers and the central limit theorem for random elements of Hilbert spaces are presented (Theorem 1.3.1), as well as asymptotic results concerning the eigenstructure of the empirical covariance operator (Section 1.3.2). We then turn to the functional time series case (Section 1.4), first by presenting the case of functional autoregressive models and functional linear process models, under which asymptotic results for the sample mean and sample autocovariance operators are given in Theorem 1.4.2. We then move on beyond the linearity assumption and consider more general functional time series structure whose weak dependence is quantified by α-mixing or L^p-m-approximability (Section 1.4.2). We conclude the chapter by noting that the Karhunen–Loève expansion, though being a sensible and optimal tool in the i.i.d. setup, is *not* the best approach in presence of dependencies (Section 1.5).

Chapter 2 is devoted to the development of a decomposition for functional time series (FTS) that generalizes the properties that the Karhunen–Loève has in the i.i.d. setup. This is done by first developing a *functional Cramér representation* (Theorem 2.4.3), which tells us that each second-order functional time series can be approximately decomposed into a sum of uncorrelated random functions that are vibrating at distinct frequencies. Further to this first *spectral* decomposition, we can perform a second decomposition by expanding each of these uncorrelated random functions using its Karhunen–Loève expansion. As a result of this procedure, we obtain a doubly spectral decomposition of the FTS, a *Cramér–Karhunen–Loève* decomposition (Theorem 2.8.6 and Remark 2.8.7), where the first layer of decomposition is given by the functional Cramér representation, and the second layer of decomposition is given by the Karhunen–Loève expansion. Furthermore, we show that the truncation of the Cramér–Karhunen–Loève leads to a finite dimensional approximation of the FTS that *dominates* functional PCA (Theorem 2.8.2 and Remark 2.8.5). The theoretical construction of such a decomposition is closely related to linear filterings of FTSs, and by duality to a stochastic integral of operator-valued functions (presented in Section 2.5). Some technical measure theoretic considerations are presented in Section 2.7.

The basic object used in the Cramér–Karhunen–Loève are the spectral density operators (and their eigenstructure), whose Fourier coefficients are given by the autocovariance operators of the FTS. The spectral density operators are therefore usually unknown, and must be estimated from data in applications. Chapter 3 is devoted to development of the theory of estimation for the spectral density operators. The central tool for estimating the spectral density operators is the functional discrete Fourier transform (fDFT; Section 3.3). It transforms a dataset of real curves

$(X_t)_{t=1}^T$ of the time domain into a dataset of complex curves $\left(\widetilde{X}_{\omega_j}^{(T)}\right)_{j=1}^T$ in the frequency domain that are asymptotically independent and Gaussian, under appropriate cumulant mixing assumptions (Theorem 3.3.4), with the spectral density operators as asymptotic covariances. This motivates then taking the fDFT's empirical covariance, the *periodogram operators*, as estimators for the spectral density operators (Section 3.4). It turns out that the periodogram operators are asymptotically unbiased (Proposition 3.4.4), but are not consistent due to their non-vanishing variance (Proposition 3.5.4). However, by smoothing the periodogram locally (on a small window of frequencies), it is possible to construct estimators with reduced variance, the *sample spectral density operators* (Section 3.5). In particular, we can show that the sample spectral density operators are consistent estimators of the spectral density operators (Theorems 3.6.1 and 3.6.2), and that they are asymptotically Gaussian (Theorems 3.6.5 and 3.6.7), under appropriate weak dependence assumptions. As immediate consequences, we obtain central limit theorems for the mean and the long-run covariance operator of a stationary functional time series (Corollaries 3.3.6 and 3.6.8). Furthermore, we show that the eigenstructure of the sample spectral density operators consistently estimates the eigenstructure of the spectral density operators (Proposition 3.7.2), and is also asymptotically Gaussian (Theorem 3.7.3).

In practice, functional data are often discretely observed, and contaminated with noise. We therefore study in Section 3.8 sufficient conditions under which an estimator of the spectral density operators based on discrete noisy observations of the curves (X_t) will still be consistent (Theorem 3.8.3). Our result is general enough to apply to a wide range of smoothing techniques for converting discrete noisy observations into functional data (Remark 3.8.4). We perform a simulation study in Section 3.9 to assess the finite sample performance of our estimators, and give a discussion of the technical assumptions of our results, and at what cost our weak dependence assumptions could be changed or weakened. In particular, we show (Remark 3.10.3) that any functional linear process with summable coefficients (in an appropriate sense) and whose innovations have finite moments will satisfy the technical assumptions of our asymptotic results.

In Chapter 4, we study the problem of understanding the mechanics of DNA, and how the mechanical properties that are described at a fine level scale, from a statistical point of view. More specifically, we are interested in understanding and comparing the dynamics of two closed strands of DNA (*DNA minicircles*) that are vibrating in solution. To address this problem, we compare the dynamics of the DNA minicircles through the lens of the spectral density operators. The first method we propose is to compare the dynamics of two DNA minicircles by comparing their spectral density operators at the level of frequencies (Section 4.3), and localizing at which frequencies the differences occur. This is done by first testing equality of the to spectral density operators marginally at each frequency (Theorem 4.3.1), and then performing multiplicity corrections (Section 4.3.3). We conduct numerical simulations to assess the performance of our method (Section 4.3.5). The second method we propose is to first localize frequencies at which the two spectral density operators are different, and then localize *on the DNA minicircles* where the differences occur, through a multiple testing approach (Section 4.4).

On the technical side, the mathematical foundations of FDA have their roots in functional analysis and Bochner integrals.

The deterministic technical challenges with FDA are that the random elements one has to work with take values in linear spaces with *infinite* dimension, such as Banach spaces or Hilbert spaces. Contrary to finite dimensional vector spaces, not all norms are equivalent in infinite dimensions. For instance, the *trace norm* of an operator on a Hilbert space—the norm related to the total variance of random elements—is much stronger and more difficult to work with than the operator norm, which corresponds roughly to the largest singular value of an operator. In between the two is the Hilbert-Schmidt norm, which gives a Hilbert space structure to a subspace of the space of bounded operators. Often, proofs will involve a fair "mix" of these different types of norms, which might make them difficult to follow. A list of notation has therefore been added to the thesis (page xvii) to facilitate reading, while the technical background is briefly recalled in Section 1.1, and more thoroughly in Chapter A.

The concept of random variable in a Banach space is formalized by the Bochner integral, which is in some sense a generalization of the Lebesgue integral (Lebesgue 1904, Tao 2011) to Banach space valued functions. Fortunately, many of the properties of the Lebesgue integral extend to the Bochner integral; we give in Section 1.1 some of its main properties, and a more detailed exposition in Chapter B.

Further to the Bochner integral, we will be concerned with convergence in distribution for random elements in Banach spaces; Chapter C reminds some basic facts and gives a useful result for establishing tightness for the particular case of random elements in Hilbert spaces.

CHAPTER 1

Functional Data Analysis

Functional data analysis (FDA) is the field of statistics that treats complex data structures —such as smooth curves or surfaces—that depart from the conventional univariate and multivariate structure. A key feature that distinguishes functional data from multivariate or high-dimensional data is their dimension. Functional Data is assumed to belong to a (often linear) infinite dimensional space, whereas multivariate or high-dimensional data is constrained to belong in a finite dimensional space, whose dimension p is either fixed, or allowed to grow along with the sample size in asymptotic settings. Another feature that distinguishes functional data is that they are assumed to have some inherent smoothness, in contrast with multivariate data, for which smoothness is not meaningful. FDA is now a well established field. Its roots can be traced at least back to Grenander (1981), and the current literature on the subject is very rich. The main references on the subject are (Ramsay & Silverman 2005, Ramsay et al. 2009, Ferraty 2011, Horváth & Kokoszka 2012, Ferraty & Vieu 2006, Ferraty & Romain 2011).

We introduce the notation and remind basic concepts of functional analysis that shall be used throughout the chapter in Section 1.1. We then present (Section 1.2) basic aspects of FDA, and review in Section 1.3 some of the theory for inference of functional data in the i.i.d. case. We then turn to the setting of functional time series (Section 1.4), which is the main concern of this thesis. We conclude this chapter with the main motivation of this thesis, which is the need for a generalization of the Karhunen–Loève expansion to the non-i.i.d. setting (Section 1.5.

1.1 Some Background Theory

Typically, but not invariably, the random object of interest is $X : \Omega \to B$, where $(\Omega, \mathcal{O}, \mathbb{P})$ is a complete probability space, and B is a separable Banach space, or a separable Hilbert space, whose norm is denoted by $\|\cdot\|$. We usually omit the space of events Ω, and say $X \in B$ is a random element. The rigorous definition of random element in a Banach space can be done through the Bochner integral (see Chapter B).

Since the random element takes values in Banach or Hilbert spaces, we need to introduce some of their related concepts. A continuous mapping $T : B_1 \to B_2$ between two Banach spaces is called a bounded operator if it is linear and continuous. Continuity is equivalent to $|||T|||_\infty = \sup_{x \neq 0} \|Tx\|/\|x\| < \infty$. The set of bounded operators $T : B_1 \to B_2$ is denoted by $\mathcal{S}_\infty(B_1, B_2)$, and we also use the abbreviation $\mathcal{S}_\infty(B) = \mathcal{S}_\infty(B, B)$. $\mathcal{S}_\infty(B_1, B_2)$ is in fact a linear space, and is complete under the *operator norm* $|||\cdot|||_\infty$. It is therefore also a Banach space. For $T, S \in \mathcal{S}_\infty(B)$, we can define their composition $TS(x) = T(Sx), x \in B$, which is also a bounded operator whose norm satisfies $|||TS|||_\infty \leq |||T|||_\infty |||S|||_\infty$. We now turn to operators on separable Hilbert spaces. Since these are Banach spaces with additional structure (given by their inner-product $\langle \cdot, \cdot \rangle : H \times H \to \mathbb{C}$), the definition of bounded operator extends to Hilbert spaces. A bounded operator $T \in \mathcal{S}_\infty(H_1, H_2)$ between two Hilbert spaces is called *compact* if it can be written as $T = \sum_{n=1}^\infty \lambda_n(T) \varphi_n \otimes_2 \psi_n$, where $(\lambda_n(T))_{n \geq 1}$ is a sequence of decreasing positive numbers tending to zero (the *singular values* of T), the sequences $(\psi_n)_{n \geq 1}$, respectively $(\varphi_n)_{n \geq 1}$ are orthonormal sequences in H_1, respectively H_2, and $\varphi \otimes_2 \psi$ is a linear operator defined by $\varphi \otimes_2 \psi(f) = \langle f, \psi \rangle \varphi$ for $f, \psi \in H_1$, $\varphi \in H_2$. Note that compactness is a restrictive property: for instance, the identity operator on H is not compact. The class of compact operators is denoted by $\mathcal{S}_c(H_1, H_2)$. For any compact operator $T \in \mathcal{S}_c(H_1, H_2)$, we can define its *Schatten p-norm* by $|||T|||_p = \left(\sum_{n \geq 1} \lambda_n(T)^p\right)^{1/p}$, for $p \in [1, \infty)$, and $|||T|||_\infty = \sup_n \lambda_n(T)$. They are both well defined, due to the uniqueness of the singular value decomposition. The Schatten ∞-norm is equal to the operator norm when restricted to compact operators, and hence poses no conflict of notation. The Schatten p-norm is indeed a norm (for each $p \in [1, \infty]$), and the space

$$\mathcal{S}_p(H_1, H_2) = \left\{ T \in \mathcal{S}_c(H_1, H_2) : |||T|||_p < \infty \right\}$$

is a Banach space when equipped with the norm $|||\cdot|||_p$. It is called the *Schatten p space*. The abbreviation $\mathcal{S}_p(H) = \mathcal{S}_p(H, H)$ is often used. The Schatten norms follow the inequality $|||\cdot|||_p \geq |||\cdot|||_q$ if $1 \leq p \leq q \leq \infty$, which implies the following chain of inclusions,

$$\mathcal{S}_1(H_1, H_2) \subset \mathcal{S}_p(H_1, H_2) \subset \mathcal{S}_q(H_1, H_2) \subset \mathcal{S}_c(H_1, H_2) \subset \mathcal{S}_\infty(H_1, H_2),$$

for $1 \leq p \leq q \leq \infty$. A Hölder inequality also holds for Schatten spaces, and

is very useful: if $T \in \mathscr{S}_p(H)$, $S \in \mathscr{S}_q(H)$, for $p, q \in [1,\infty]$, we have

$$|||TS|||_r \leq |||T|||_p |||S|||_q, \quad r^{-1} = p^{-1} + q^{-1}.$$

The Schatten space $\mathscr{S}_2(H_1, H_2)$ is called the space of *Hilbert–Schmidt operators,* and is in fact a Hilbert space when equipped with the inner product

$$\langle T, S \rangle_{\mathscr{S}_2} = \sum_{n \geq 1} \langle Te_n, Se_n \rangle, \quad T, S \in \mathscr{S}_2(H_1, H_2),$$

where the sum extends over any orthonormal basis $(e_n)_{n\geq 1}$ of H_1, and is independent of the choice of the basis. The corresponding norm is given by $|||T|||_2^2 = \sum_{n\geq 1} \|Te_n\|^2$. The Schatten space $\mathscr{S}_1(H)$ is also quite important, and is called the space of *nuclear operators* on H, or *trace-class operators* on H, because we can define the trace for its elements:

$$\mathrm{Tr}(T) = \sum_{n \geq 1} \langle Te_n, e_n \rangle, \quad T \in \mathscr{S}_1(H),$$

where the sum extends over any orthonormal basis $(e_n)_{n\geq 1}$ of H, and is independent of the choice of the basis. In particular, $|\mathrm{Tr}(T)| \leq |||T|||_1$, and therefore the trace is a continuous linear functional on $\mathscr{S}_1(H)$. Furthermore, if T is a positive operator, i.e. $\langle Tx, x \rangle \geq 0$ for all $x \in H$, then $\mathrm{Tr}(T) = |||T|||_1$. Notice that the Hölder inequality tells us that TS and ST are trace class if $T \in \mathscr{S}_1(H)$ and S is any bounded operator. A Hilbert space that we will often use is $H = L^2([0,1], \mathbb{R})$, or $H = L^2([0,1], \mathbb{C})$. In both cases, the inner-product is given by $\langle f, g \rangle = \int_0^1 f(\tau)\overline{g(\tau)}d\tau$, where $\overline{\alpha}$ denotes the complex conjugate of $\alpha \in \mathbb{C}$. More details about operators on Banach or Hilbert spaces are given in Chapter A.

Let us now come back to random elements. Let $X \in B$ be a random element, where B is a separable Banach space. If $\mathbb{E}\|X\| < \infty$, the expectation of X exists, and is defined as the *unique* element $\mathbb{E}X \in B$ satisfying $\phi(\mathbb{E}X) = \mathbb{E}\phi(X)$ for all linear and continuous functionals on B. It satisfies the *contraction property* $\|\mathbb{E}X\| \leq \mathbb{E}\|X\|$, and commutes with any bounded operator $T \in \mathscr{S}_\infty(B, B_1)$, i.e. $T\,\mathbb{E}X = \mathbb{E}[TX]$. A random element $X \in H$, where H is a separable Hilbert space, has a mean $\mu = \mathbb{E}X$ if $\mathbb{E}\|X\| < \infty$, and a *covariance operator* $\mathscr{R} = \mathbb{E}\left[(X-\mu)\otimes_2(X-\mu)\right]$ if $\mathbb{E}\|X\|^2 < \infty$. The covariance operator is a trace-class operator, and since the trace is a continuous and linear, we have

$$\mathrm{Tr}(\mathscr{R}) = \mathbb{E}\left[\mathrm{Tr}\left((X-\mu)\otimes_2(X-\mu)\right)\right] = \mathbb{E}\left\|X-\mu\right\|^2,$$

which illustrates the fundamental relation between the total variance and the nuclear norm. If $Y \in H$ is another random element with $\mathbb{E}\|Y\|^2 < \infty$, the cross-covariance operator of X and Y is defined by

$$\mathscr{R}_{X,Y} = \mathbb{E}[(X - \mathbb{E}X)\otimes_2(Y - \mathbb{E}Y)].$$

It is a nuclear operator, and therefore

$$\mathrm{Tr}\left(\mathscr{R}_{X,Y}\right) = \mathbb{E}\left[\mathrm{Tr}\left((X-\mathbb{E}X)\otimes_2(Y-\mathbb{E}Y)\right)\right] = \mathbb{E}\left\langle (X-\mathbb{E}X),(Y-\mathbb{E}Y)\right\rangle.$$

More details and results concerning random elements in Banach or Hilbert spaces are given in Chapter C.

1.2 Basic Aspects of Functional Data Analysis

The typical setting of functional data analysis (FDA) is concerned with inference on the law of a random function $X \in L^2([0,1],\mathbb{R})$, with $\mathbb{E}\|X\|^2 < \infty$, based on a sample $X_1,\ldots,X_n \overset{\text{iid}}{\sim} X$. The random function X can be thought of as a collection of random variables $\{X(\tau) : \tau \in [0,1]\}$. A crucial assumption concerning this collection of random variables is that it is assumed to be smooth with respect to the parameter τ. Smoothness is meant in a broad sense here, and can range from assuming continuity of the sample paths of X—which is sometimes modeled by assuming that X is a random element of the Banach space $C([0,1],\mathbb{R})$—or by assuming that the sample paths are k times differentiable—by assuming X is a random element of the Sobolev space $\mathscr{W}^{k,2}([0,1],\mathbb{R})$. Most often, the chosen space is a Hilbert space (an assumption that simplifies considerably the derivation of asymptotic results), and we shall therefore assume in this Chapter that the random function X takes values in $L^2([0,1],\mathbb{R})$. We mention however that work has also been done on departing from the linear space assumption (e.g. (Chen & Müller 2012)), and is a growing area.
The main objects describing the random function X are the mean function and the covariance surface:

Definition 1.2.1.
Assuming $\mathbb{E}\|X\|^2 < \infty$*, the mean function* $\mu \in L^2([0,1],\mathbb{R})$ *is defined by*

$$\mu(\tau) = \mathbb{E}[X(\tau)], \quad \tau \in [0,1],$$

and the covariance surface*, or* covariance kernel $r \in L^2\left([0,1]^2,\mathbb{R}\right)$ *is defined by*

$$r(\tau,\sigma) = \mathbb{E}\left[\left(X(\tau)-\mu(\tau)\right)\left(X(\sigma)-\mu(\sigma)\right)\right], \quad \tau,\sigma \in [0,1].$$

Interpretation is helped by analogy to the multivariate case: The mean function is the functional analogue of the mean vector, and the covariance surface in the analogue of the covariance matrix. The covariance surface induces by right integration an linear operator $\mathscr{R}$ on $L^2([0,1],\mathbb{R})$, whose eigenstructure plays a central role in FDA:

$$\mathscr{R}h(\tau) = \int_0^1 r(\tau,\sigma)h(\sigma)d\sigma, \quad h \in L^2([0,1],\mathbb{R}); \tau \in [0,1]. \tag{1.2.1}$$

The covariance operator is well defined if $\mathbb{E}\|X\|^2 < \infty$, and is a non-negative self-adjoint and trace-class operator on $L^2([0,1],\mathbb{R})$, with singular

value decomposition

$$\mathscr{R} = \sum_{n=1}^{\infty} \lambda_n \varphi_n \otimes_2 \varphi_n, \tag{1.2.2}$$

see equation (A.2.7) on page 219 for the definition of $\otimes_2$

(see e.g. Bosq (2000)), where λ_n denotes the n-th largest eigenvalue (i.e. $\lambda_1 \geq \lambda_2 \geq \ldots 0$), φ_n denotes the corresponding eigenfunction, and the convergence of the series holds with respect to the nuclear norm $|||\cdot|||_1$. The eigenfunctions are of course not identifiable, but each *eigenprojection* $\varphi_n \otimes_2 \varphi_n$ is identifiable, provided λ_n is an eigenvalue of multiplicity one. Dimension reduction techniques play a important role in multivariate analysis, but it is no exaggeration to say that their role in FDA is central, because the random elements dealt with are intrinsically infinite dimensional. The functional counterpart of principal component analysis (PCA), *functional PCA* (fPCA), is given by the celebrated Karhunen–Loève expansion (see e.g. Karhunen (1947), Lévy (1948), Ash & Gardner (1975), Grenander (1981)):

Proposition 1.2.2 (Karhunen–Loève Expansion, L^2 version)**.**
Let $X \in L^2([0,1],\mathbb{R})$ be a random function with $\mathbb{E}\|X\|^2 < \infty$, with covariance operator given by (1.2.2)*. The random function X admits the decomposition*

$$X = \mu + \sum_{n=1}^{\infty} \xi_n \varphi_n, \tag{1.2.3}$$

where φ_n is defined in (1.2.2)*, $\xi_n = \langle \varphi_n, X - \mu \rangle$, $\mathbb{E}\xi_n = 0$ and $\mathbb{E}[\xi_n \xi_m] = \lambda_n \delta_{n,m}$, with $\delta_{n,m} = 1$ if $n = m$, and zero otherwise. The convergence of the series holds in mean square, in $L^2([0,1],\mathbb{R})$:*

$$\mathbb{E}\left\| X - \mu - \sum_{n=1}^{K} \xi_n \varphi_n \right\|^2 = \sum_{n>K} \lambda_n \overset{K\to\infty}{\longrightarrow} 0.$$

The Karhunen–Loève expansion yields a separation of the random function X into a sum of random variables (the ξ_ns, also known as the *scores*) times orthogonal deterministic functions (φ_n). Of course, any orthogonal basis of $L^2([0,1],\mathbb{R})$ would yield such a decomposition, but (1.2.3) has the additional property that the random variables (ξ_n) are uncorrelated (even independent in the Gaussian case), and that truncation of the series (1.2.3) at a finite level K yields the best K-dimensional linear approximation of X (also known as the *best basis property*). More precisely, this means that the solution to

$$\arg\min_{P \in \mathscr{P}_K} \mathbb{E}\left\| X - \mu - P(X - \mu) \right\|^2, \tag{1.2.4}$$

where $\mathscr{P}_K$ is the space of all orthogonal projections on $L^2([0,1],\mathbb{R})$ with rank at most K, is given by $P = \sum_{n=1}^{K} \varphi_n \otimes_2 \varphi_n$, the projection onto the subspace of the first K eigenfunctions of $\mathscr{R}$, or in other words, $P(X - \mu) = \sum_{n=1}^{K} \xi_n \varphi_n$. Assuming continuity of the covariance kernel of X yields an even stronger result:

Proposition 1.2.3 (e.g. Ash & Gardner (1975), Grenander (1981))**.**
Under the same setting as Proposition 1.2.2, with the additional condition that $(\tau,\sigma) \mapsto r(\tau,\sigma)$ *is continuous—or equivalently that* $\tau \mapsto X(\tau)$ *is continuous in mean square—we have:*

(i) Mercer's Lemma:

the eigenfunctions φ_n *are continuous, and the convergence of the series expansion*

$$r(\tau,\sigma) = \sum_{n=1}^{\infty} \lambda_n \varphi_n(\tau)\varphi_n(\sigma) \tag{1.2.5}$$

holds uniformly and absolutely on $[0,1]^2$.

(ii) Karhunen–Loève expansion (strong form):

the convergence of (1.2.3) *holds uniformly in mean square:*

$$\sup_{\tau\in[0,1]} \mathbb{E}\left(X(\tau) - \sum_{n=1}^{K} \xi_n \varphi_n(\tau)\right)^2 \overset{K\to\infty}{\longrightarrow} 0.$$

The Karhunen–Loève expansion is also useful for computing linear and bilinear functionals of random elements. For instance, if X, Y are random elements of $L^2([0,1],\mathbb{R})$ with Karhunen–Loève expansions $X = \sum_{n\geq 1} \xi_n \varphi_n$ and $Y = \sum_{n\geq 1} \zeta_n \psi_n$, and $f : L^2([0,1],\mathbb{R}) \to B$ is a linear and continuous mapping into a Banach space, then

$$f(X) = \sum_{n\geq 1} \xi_n f(\varphi_n),$$

where the right-hand side converges in mean. Furthermore, if

$$g : L^2([0,1],\mathbb{R}) \times L^2([0,1],\mathbb{R}) \to B$$

is a bounded multilinear mapping, then

$$g(X,Y) = \sum_{n,m\geq 1} \xi_n \zeta_m f(\varphi_n, \psi_m),$$

where the right-hand side converges in mean. In particular, the Karhunen–Loève expansion can be used to compute the covariances of two random elements (see Section 3.7.6 for a concrete example).
To use of the Karhunen–Loève expansion in practice, one needs to estimate the eigenfunctions $\varphi_1,\ldots,\varphi_K$ from a sample $X_1,\ldots,X_n \overset{\text{iid}}{\sim} X$. However, the curves $X_1,\ldots,X_n$ are usually not entirely observed, but observed on a discrete grid, possibly with some noise. We therefore describe briefly the preprocessing steps usually taken to transform the discrete noisy observations into a functional sample $X_1,\ldots,X_n$.

1.2.1 Preprocessing Steps

1.2.1.1 Projection to Functional Data

The recorded data for the curve X_i—assuming pointwise evaluation of the curve makes sense—consist usually of pairs $\{(\tau_{ij}, Y_{ij}) : j = 1, \ldots, N_i\}$, following the sampling model

$$Y_{ij} = X_i(\tau_{ij}) + \varepsilon_{ij}, \quad i = 1, \ldots, n; j = 1, \ldots, N_i, \tag{1.2.6}$$

where $\tau_{ij} \in [0,1]$ is the location at which the j-th observation of curve i occurs, and N_i is the number of observations for curve i (both can be assumed to be either deterministic or random). The variables ε_{ij} denote the noise in the observation, and are usually assumed to be independent. Both $X_i(\tau_{ij})$ and ε_{ij} are not observed.

The first step in FDA is the transformation of such data into functional data; this is usually done by some kind of smoothing, either based on least square estimation of the functions X_i through a finite basis expansion, or by localized procedures, such as kernel smoothing or localized basis and polynomial expansions (Wand & Jones 1995, Fan & Gijbels 1996, Efromovich 1999, Ramsay & Silverman 2002, 2005). Basis expansion is arguably the most used procedure, and the commonly used bases are the Fourier basis, the B-spline basis, and the wavelet basis, to cite only the most popular (Sy et al. 1997, Yao & Lee 2006, Morris & Carroll 2006, Pigoli & Sangalli 2012). The number of basis functions K is often chosen sufficiently high, to allow the expansion to capture local features of the curves X_i, and an additional penalty for the roughness of the fitted curve is often added to the least squared penalty, typically through by penalizing the norm of a differential operator applied to X_i (e.g. (Ramsay & Silverman 2005, Sangalli et al. 2009)).

1.2.1.2 The Problem of Registration

The second step in FDA is the registration of the functional data. The basic idea is that the observations $X_1, \ldots, X_n$ do not correspond to a random sample from X, but to a random sample from X^*, where $X^*(\tau) = X(\gamma(\tau))$, and γ is a random function taking values in the (non-linear) space of increasing bijective functions $[0,1] \mapsto [0,1]$. Denoting the observed sample $X_1^*, \ldots, X_n^*$, where $X_i^* = X_i \circ \gamma$, and $\circ$ denotes the composition of functions, the goal of registration is to separate the functions $X_1, \ldots, X_n$ and $\gamma_1, \ldots, \gamma_n$, to allow further inference about X, and possibly about γ (see e.g. Liu & Müller (2004), Srivastava et al. (2011)). Of course, the choice of the space in which γ takes values in intimately related to the qualitative interests in the variability of X, but is usually also constrained to avoid identifiability issues.

1.3 Inference for Functional Data: the i.i.d. Setup

We now give some of the main results concerning inference for functional data in the i.i.d. setup.

1.3.1 Estimation of the Mean Function and Covariance Surface

We now assume that we have observed a sample of curves $X_1, \ldots, X_n \overset{\text{iid}}{\sim} X$, and wish to estimate the mean function and the covariance surface (or operator) of X. The estimators for the mean function, respectively the covariance operator, are similar to those of the multivariate case, and are given by the *sample mean*

$$\overline{X} = n^{-1} \sum_{j=1}^{n} X_j,$$

and the *sample covariance operator*

$$\hat{\mathscr{R}} = \frac{1}{n} \sum_{j=1}^{n} \left(X_j - \overline{X} \right) \otimes_2 \left(X_j - \overline{X} \right),$$

respectively. Results analogous to the multivariate setup hold for asymptotics of the sample mean:

Theorem 1.3.1 (e.g. Bosq (2000))**.**
Let $X_1, \ldots, X_n \overset{\text{iid}}{\sim} X$, where X is a random function in $L^2([0,1], \mathbb{R})$ (or in any separable Hilbert space).

(i) Strong Law of Large Numbers *If $\mathbb{E}\|X\| < \infty$, then*

$$\overline{X} \xrightarrow{a.s.} \mu, \quad n \to \infty,$$

with respect to $\|\cdot\|$.

(ii) Central Limit Theorem *If $\mathbb{E}\|X\|^2 < \infty$, then*

$$n^{1/2} \left(\overline{X} - \mu \right) \xrightarrow{d} Z, \quad n \to \infty,$$

where Z is a Gaussian random element of $L^2([0,1], \mathbb{R})$ with mean zero and covariance operator $\mathscr{R}$.

This is a strong result, since it is valid for X a random variable of any separable Hilbert space. In particular, taking H to be the space of Hilbert–Schmidt operators $\mathscr{S}_2(L^2([0,1], \mathbb{R}))$, the previous theorem implies that the strong law of large numbers holds for $\hat{\mathscr{R}}$ with respect to the Hilbert–Schmidt norm (in fact also with respect to the nuclear norm by the Strong Law of Large Numbers on Banach spaces) provided $\mathbb{E}\|X\|^2 < \infty$. Furthermore, provided $\mathbb{E}\|X\|^4 < \infty$, $n^{1/2}(\hat{\mathscr{R}} - \mathscr{R})$ is asymptotically Gaussian, with

limiting covariance operator given by

$$\mathbb{E}\left[(X\otimes_2 X-\mathscr{R})\bigotimes_2(X\otimes_2 X-\mathscr{R})\right]=\sum_{i,j,k,l\geq 1}\mathbb{E}\left[\xi_i\xi_j\xi_k\xi_l\right](\varphi_i\otimes_2\varphi_j)\bigotimes_2(\varphi_k\otimes_2\varphi_l)$$
$$-\sum_{i,j\geq 1}\lambda_i\lambda_j(\varphi_i\otimes_2\varphi_i)\bigotimes_2(\varphi_j\otimes_2\varphi_j),$$

where ξ_n and φ_n are given by the Karhunen–Loève expansion of X (see Proposition 1.2.2). The last expression may be further simplified if X is Gaussian (see Dauxois et al. (1982) for details).

1.3.2 Estimation of the Eigenstructure of the Covariance Operator

As a by-product, the asymptotic normality of the sample eigenvalues and eigenprojections follows from results in perturbation theory, both with the same $n^{1/2}$ rate (see Dauxois et al. (1982), Mas & Menneteau (2003), Hall & Hosseini-Nasab (2006)), which can be improved with assumptions on higher moments, using Bernstein's exponential inequality (Mas & Ruymgaart 2014, to appear). However, estimation of the eigenfunctions is more difficult than the estimation of the eigenvalues; the eigenvalue spacings $\{\lambda_{n-1}-\lambda_n,\lambda_n-\lambda_{n+1}\}$ have a first-order effect on the estimation of φ_n, but only a second-order effect on the estimation of λ_n (Hall & Hosseini-Nasab 2006).

We also mention that work has been done on combining the projection of the discrete data and the estimation steps. For instance, Cai & Yuan (2011) derive minimax rates for the estimation of the mean function under both sparse and dense sampling. Yao, Müller & Wang (2005) investigated the estimation of the eigenvalues when the observations for each curves are sparse. Benko et al. (2009) study the estimation of the eigenvalues and eigenfunctions in the dense sampling setup.

There has been a considerable amount of work in extending models for multivariate data into the functional framework. The major idea for doing so is to project the functional data onto the linear subspace spanned by the first K empirical principal components, thus obtaining a K-dimensional approximation of the data, and then applying the ideas of multivariate analysis. When doing so, the appropriate choice of K of crucial importance, and is also a difficult problem. Roughly speaking, a K too small leads to a crude approximation of the functional data, and a too large value of K yields bad statistical properties due to ill-posedness issues.

1.3.3 Inference for Functional Data: Departure from the i.i.d. Assumption

Modeling functional data that are not i.i.d. can be done by adding a covariate. In the linear case, this can be done by the *functional linear model*

(FLM)

$$Y = \mathscr{A} X + \varepsilon$$

where Y / X are either functional/functional, multivariate/functional or functional/multivariate, X & ε are random and independent from each other, and $\mathscr{A}$ is a Hilbert–Schmidt operator. The functional linear model has been widely studied (He et al. 2000, Cai & Hall 2006, Cardot, Ferraty & Sarda 2003, Cardot et al. 1999, Cardot, Ferraty, Mas & Sarda 2003, Hall & Horowitz 2007, Hilgert et al. 2013, Crambes & Mas 2013, Cai & Yuan 2012). Various extensions to the FLM have also been proposed. For instance, James et al. (2009) present an "interpretable" extension for the case where Y is univariate and X is functional, by using a Lasso-type penalization (see Tibshirani (1996, 2011)). Aston, Chiou & Evans (2010), Hadjipantelis, Aston & Evans (2012) study the linear mixed model extension. Müller & Stadtmüller (2005) introduce the generalized functional linear models. He, Müller, Wang & Yang (2010) study the connection between the FLM and Functional canonical analysis.
Non-parametric modelling has also been extended to the functional setting. See Ferraty & Vieu (2006) and references therein for more details.

1.4 Functional Time Series

Though the i.i.d. setup is appropriate in many problems, many other problems have a natural dependency structure, that can be modeled for instance through a time series structure. Let us give an example. Let $Y_{t,h}$ denote the temperature recorded at Pully on day t and at time $h \in [0,24]$. If we define $X_t = (Y_{t,h})_{h\in[0,24]}$, then X_t is a curve describing the temperature in Pully during day t, and the series $(X_t)_{t=1,2,\dots}$ is not i.i.d., but can be modelled as a time series of functional data, or *functional time series*.

Pully is a municipality in Switzerland on the shores of Lac Léman, nearby Lausanne

In this section, we review the basic theory of functional time series, and motivate the subject of this thesis.
A *functional time series* (FTS) is a sequence $(X_t : t \in \mathbb{Z})$ where each X_t is a random element in $L^2([0,1],\mathbb{R})$. The study of functional time series is mostly done—explicitly or implicitly—under the assumption of second-order stationarity, or even strict stationarity:

Definition 1.4.1.
A functional time series $(X_t : t \in \mathbb{Z})$ *is said to be* second-order stationary *if* $\mathbb{E}\, X_s$ *and* $\mathbb{E}[X_{t+s} \otimes_2 X_s]$ *are independent of* $s \in \mathbb{Z}$*, for all* $t \in \mathbb{Z}$*, and* strictly stationary *if for any* $k = 1,2,\dots$*, and any* $t_1,\dots,t_k \in \mathbb{Z}$*, the joint law of* $\{X_{t_1+s},\dots,X_{t_k+s}\}$ *is independent of* $s \in \mathbb{Z}$.

Second-order stationarity says that the first and second-order moments of the FTS are invariant by time-shifts. Strict stationarity is stronger (and implies second-order stationarity) because it imposes that all the moments of the FTS are invariant by time-shifts (and even more if the distribution is not determined by the moments). Second-order stationarity implies

that the second-order structure of the FTS is encoded by the *lag-t autocovariance operators*

$$\mathscr{R}_t = \mathbb{E}\left[\left(X_{t+s}-\mu\right)\left(X_s-\mu\right)\right], \quad t \in \mathbb{Z}, \tag{1.4.1}$$

which are independent of s , and where $\mu = \mathbb{E}\, X_s$ is the mean function (both being independent of s due to the stationarity assumption).

1.4.1 Linear models for Functional Time Series

Perhaps the easiest and most tractable model for functional time series is the functional autoregressive model of order 1, FAR(1) (see e.g. Bosq (2000), Mas (2007)),

$$X_{t+1} - \mu = \mathscr{A}(X_t - \mu) + \varepsilon_t, \quad t \in \mathbb{Z}, \tag{1.4.2}$$

where $\mathscr{A}$ is a bounded operator on $L^2([0,1],\mathbb{R})$, $\mu \in L^2([0,1],\mathbb{R})$, and $(\varepsilon_t : t \in \mathbb{Z})$ is a mean zero and uncorrelated ($\mathbb{E}[\varepsilon_t \otimes_2 \varepsilon_{t'}] = \delta_{t,t'}$) FTS in $L^2([0,1],\mathbb{R})$. When $\left|\left|\left|\mathscr{A}^j\right|\right|\right|_\infty < 1$ for some positive integer j, then (1.4.2) has a unique stationary solution given by

$$X_t = \mu + \sum_{j=0}^{\infty} \mathscr{A}^j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \tag{1.4.3}$$

see Bosq (2000, Theorem 3.1). This casts the FAR(1) model into the more general framework of functional linear processes (FLP)

$$X_t = \mu + \sum_{j=-\infty}^{\infty} \mathscr{A}_j \varepsilon_{t-j}, \quad t \in \mathbb{Z}, \tag{1.4.4}$$

where $\left(\mathscr{A}_j : j \in \mathbb{Z}\right)$ is a sequence of bounded operators on $L^2([0,1],\mathbb{R})$. Based on a stretch of data $X_1,\ldots,X_T$, the mean function μ is estimated by the sample mean $\overline{X}_T$, and the lag-t autocovariance operator $\mathscr{R}_t$ is estimated by the *sample lag-t autocovariance operators*

$$\hat{\mathscr{R}}_t = \sum_{j=1}^{T-t} X_{j+t} \otimes_2 X_j, \quad t = 1,\ldots,T-1.$$

Under summability assumptions on the norms of the operators $\mathscr{A}_j$, there is a central limit theorem (CLT) for the mean and autocovariances of FLPs:

Theorem 1.4.2 (Merlevède et al. (1997), Mas (2002), Mas & Pumo (2011), Bosq (2000))**.**

Let X_t be a FLP satisfying $\sum_{j\in\mathbb{Z}} \left|\left|\left|\mathscr{A}_j\right|\right|\right|_\infty < \infty$, and with i.i.d. noise sequence $(\varepsilon_t : t \in \mathbb{Z})$. Let $C = \mathbb{E}[\varepsilon_0 \otimes_2 \varepsilon_0]$. Then

1. *If* $\mathbb{E}\|\varepsilon_0\|^2 < \infty$,

$$\sqrt{T}\left(\overline{X}_T - \mu\right) \xrightarrow{d} N, \quad T \to \infty \tag{1.4.5}$$

where N *is a Gaussian random element with mean zero, and covariance operator* $ACA^\dagger$, *where* $A = \sum_{j\in\mathbb{Z}} \mathscr{A}_j$.

2. *If* $\mathbb{E}\|\varepsilon_0\|^4 < \infty$, *then for any fixed positive integer* h,

$$\sqrt{T}\left(\hat{\mathscr{R}}_0 - \mathscr{R}_0, \hat{\mathscr{R}}_1 - \mathscr{R}_1, \ldots, \hat{\mathscr{R}}_h - \mathscr{R}_h\right) \xrightarrow{d} G, \quad T \to \infty,$$

where $G = (G_0, \ldots, G_h)$ *is a Gaussian random element in*

$$\left(\mathscr{S}_2(L^2([0,1],\mathbb{R}))\right)^{h+1}$$

with mean zero, and blockwise covariance structure given by

$\mathscr{S}_2(H)$ denotes the space of Hilbert–Schmidt Operators on H, see Section A.2.2.2 on page 220

$$\mathbb{E}\left[G_p \otimes_2 G_q\right] = \sum_{j\in\mathbb{Z}} \left(\mathscr{R}_{(p-q)+j} \widetilde{\bigotimes}_2 \mathscr{R}_j^\dagger + \mathscr{R}_{q+j} \widetilde{\bigotimes}_2 \mathscr{R}_{j-p}^\dagger\right) + B_q(\Lambda - \Phi) B_p, \tag{1.4.6}$$

where B_p, Λ *and* Φ *are operators on* $\mathscr{S}_2(L^2([0,1],\mathbb{R}))$, *defined by*

$$B_p = \sum_{j\in\mathbb{Z}} \mathscr{A}_{j+p} \widetilde{\bigotimes}_2 \mathscr{A}_j, \quad p = 0, \ldots, h,$$

$$\Lambda = \mathbb{E}\left[(\varepsilon_0 \otimes_2 \varepsilon_0) \bigotimes_2 (\varepsilon_0 \otimes_2 \varepsilon_0)\right],$$

and

$$\Phi(T) = C(T + T^\dagger)C + \langle T, C\rangle_{\mathscr{S}_2} C, \quad T \in \mathscr{S}_2(L^2([0,1],\mathbb{R})).$$

A few remarks:

Remark 1.4.3.

1. *For the univariate linear process* $y_t = \sum_{j\in\mathbb{Z}} a_j \epsilon_{t-j}$, *where* $(\epsilon_j : j \in \mathbb{Z})$ *is i.i.d. with mean zero and finite variance, and* $(a_j : j \in \mathbb{Z})$ *is a sequence of real numbers, the condition* $\sum_{j\in\mathbb{Z}} |a_j|^2 < \infty$ *ensures that* $\sqrt{T}\sum_{t=1}^T y_t / T$ *is asymptotically Gaussian. However, in the functional case, although the condition*

$$\sum_{j\in\mathbb{Z}} |||\mathscr{A}_j|||_\infty^2 < \infty, \tag{1.4.7}$$

ensures that the series (1.4.4) *is convergent in mean square and almost surely (see Merlevède (1996) and references therein),* (1.4.7) *is*

not a sufficient condition for the CLT (1.4.5) *to hold, and the condition* $\sum_{j\in\mathbb{Z}} \left|\!\left|\!\left|\mathscr{A}_j\right|\!\right|\!\right|_\infty < \infty$ *seems to be necessary (see Merlevède et al. (1997, Theorem 3)).*

2. *Notice that if* $\mathscr{A}_0 = \mathrm{Id}$*, the identity operator, and* $\mathscr{A}_j = 0$ *for all* $j \neq 0$*, then* $\hat{\mathscr{R}}_0, \ldots, \hat{\mathscr{R}}_h$ *are asymptotically independent. In particular, this allows developing portmanteau-type tests for checking if a functional time series is actually uncorrelated (e.g. Horváth & Kokoszka (2012), Horváth, Hušková & Rice (2013)).*

3. *As in the i.i.d. setup, as byproduct of the CLT* (1.4.6)*, the eigenvalues and eigenprojections of the sample lag-h autocovariance operators (h fixed) are asymptotically Gaussian, with* $\sqrt{T}$ *convergence rate (Mas & Menneteau 2003).*

The FLP model has been thoroughly studied (e.g. Bosq (2000), Mas (2002), Bosq & Blanke (2007), Mas & Pumo (2011)), and used for forecasting purposes (Bosq 2000, Besse et al. 2000, Antoniadis & Sapatinas 2003, Antoniadis et al. 2009), and also for change-point detection (Horváth et al. 2010). Various extensions have also been proposed. For instance, Damon & Guillas (2005) include exogenous variables for improving the prediction accuracy, and Cugliari (2013) proposes an autoregressive model where the operator $\mathscr{A}$ in (1.4.2) is allowed to depend on an exogenous vector time series.

1.4.2 Beyond Linearity Assumptions for Functional Time Series

Though linear models for FTS play an important role for modelling dependent functional data, the linearity assumption might not be appropriate in some situations (such as for modelling the errors in fMRI time series; see Aston & Kirch (2012*b*)). In order to make inferences on a stationary FTS $(X_t : t \in \mathbb{Z})$ without any linear structural assumption, one needs to assume some decay of the dependency between $(X_t : t < s_1)$ and $(X_t : t > s_2)$, as $s_2 - s_1 \to \infty$. We will call any such conditions *weak dependence* conditions. There are several ways of rigorously imposing weak dependence. Perhaps the most famous one, stemming from univariate time series analysis, is α-mixing.

Definition 1.4.4 (Rosenblatt (1985), Doukhan (1994))**.**
Let $\{X_t : t \in \mathbb{Z}\}$ *be a stationary functional time series. Let*

$$\alpha(n) = \sup_{A\in\mathscr{F}_0^-, B\in\mathscr{F}_n^+} |\mathbb{P}(A\cap B) - \mathbb{P}(A)\mathbb{P}(B)| \tag{1.4.8}$$

where $\mathscr{F}_k^- = \sigma(\ldots, X_{k-1}, X_k)$*, respectively* $\mathscr{F}_k^+ = \sigma(X_k, X_{k+1}, \ldots)$*, is the* σ*-algebra generated by all the FTS up to time k, respectively from time k onwards.* $\{X_t\}$ *is called* α-mixing*, or* strong mixing*, if*

$$\alpha(n) \to 0, \quad n \to \infty.$$

If $\alpha(n) = O(r_n)$ *with* $r_n \to 0$, *then* $\{X_t\}$ *is said to be* α-mixing with rate r_n.

Antoniadis et al. (2006) study the prediction of FTS under α-mixing conditions, and propose a nonparametric resampling method for constructing pointwise prediction intervals. Aston & Kirch (2012*b*,*a*) study the problem of change-point detection under α-mixing conditions, and apply it to the study of fMRI data.
In general, α-mixing yields very sharp results (Bradley 2007*a*,*b*,*c*), but it not easily verifiable in practice. Moreover, α-mixing fails in simple examples of linear processes, such as for the $AR(1)$ process

$$Y_{t+1} = \frac{1}{2} Y_t + \varepsilon_t,$$

where ε_t are Bernoulli innovations (Andrews 1984). Another way of imposing the decay of dependence is through L^p-m-approximability, a notion introduced by Hörmann & Kokoszka (2010):

Definition 1.4.5.
A FTS $\{X_t : t \in \mathbb{Z}\}$ *with* $\mathbb{E}\|X_t\|^p < \infty$ *is called* L^p-m-approximable *if each* X_t *admits the representation*

$$X_t = f(\varepsilon_t, \varepsilon_{t-1}, \ldots), \tag{1.4.9}$$

where the $(\varepsilon_t : t \in \mathbb{Z}) \overset{\text{iid}}{\sim} \varepsilon$ *are elements taking values in a measurable space* S,

$$f : S^\infty \to L^2([0,1], \mathbb{R})$$

is a measurable function, and

$$\sum_{t=1}^{\infty} \left(\mathbb{E}\|X_0 - X_0^{(t)}\|^p \right)^{1/p} < \infty, \tag{1.4.10}$$

where $X_0^{(t)} = f(\varepsilon_0, \varepsilon_{-1}, \ldots, \varepsilon_{-(t-1)}, \varepsilon'_{-t}, \varepsilon'_{-t-1}, \ldots)$, *and* $(\varepsilon'_t : t \le 0)$ *are i.i.d. copies of* ε, *and independent of* $(\varepsilon_t : t \in \mathbb{Z})$.

The intuition behind L^p-m-approximability is that the dependence of X_t (with respect to the stochastic L^p-norm) on the i.i.d. sequence $(\varepsilon_s : s \le u)$ should be decreasing fast enough as $t - u \to \infty$, such that (1.4.10) holds. Though L^p-m-approximability is not directly comparable with α-mixing, it seems to be a weak dependence concept that is easier to verify in practice; Hörmann & Kokoszka (2010) give conditions under which functional linear processes, the product model, or the functional bilinear process are L^p-m-approximable. Concerning inference, Horváth, Kokoszka & Reeder (2013) show the asymptotic normality of the sample mean under L^2-m-approximability, derive a consistent estimator of the long-run autocovariance operator $\sum_{t\in\mathbb{Z}} \mathscr{R}_t$ under slightly stronger assumptions, and apply these results to the problem of comparing the mean function of two FTS. Hörmann & Kokoszka (2010) study the effect of dependence

on fPCA. They show that the lag-0 autocovariance operator of a L^4-m-approximable FTS can be consistently estimated, and as a by-product they show consistency of the sample eigenvalues and eigenprojections. Kokoszka & Reimherr (2013) extend this result by showing that the L^4-m-approximability actually implies the asymptotic normality of these estimators. L^p-m-approximability has also been used for change-point analysis of FTS (Hörmann & Kokoszka 2010, Zhang et al. 2011, Aue, Hörmann, Horváth & Hušková 2014, to appear, Aston & Kirch 2012*b*,*a*), and prediction (Aue, Norinho & Hörmann 2014, to appear). Kokoszka (2012) gives an review of the literature on dependent functional data.

1.5 Is the Karhunen–Loève Expansion Sensible for Functional Time Series?

Most of the methodology for tackling functional time series relies on truncations of the Karhunen–Loève expansion

$$X_t = \sum_{n\geq 1} \sqrt{\lambda_n}\xi_n^t \varphi_n,$$

where $\mathscr{R}_0 = \sum_{n\geq 1} \lambda_n \varphi_n \otimes_2 \varphi_n$ is the eigen-decomposition of the lag-0 autocovariance operator of X_t, and $\xi_n^t = \langle X_t, \varphi_n \rangle / \sqrt{\lambda_n}$. Though this approach is sensible in the i.i.d. setting, we argue that it is not the best approach in the presence of dependence, since it is only based on the lag-0 autocovariance operator, and does not take into account any of the lag-t autocovariance operators of X_t (for $t \neq 0$). In particular, we will show that there exists a natural extension of the Karhunen–Loève expansion that dominates the Karhunen–Loève expansion when truncated at the same level (see Theorem 2.8.2 and Remark 2.8.7).

1.5.1 Objective of the Thesis

The first objective of this thesis is to develop the natural extension of the Karhunen–Loève expansion to the functional time series setup. This will be done by a frequency domain approach, by combining a *functional Cramér representation* for FTS with a Karhunen–Loève expansion (Chapter 2). The second objective (Chapter 3) is to develop the theory for estimation of the main objects involved in our frequency domain approach—the *spectra*, which play a role analogous to the covariance operator in the i.i.d. setup. The third objective (Chapter 4) of the thesis is study the dynamics of DNA strands, and propose a methodology for comparing their dynamics by comparing their spectra, using some of the theory developed in Chapters 2 and 3. An appendix containing technical and background results used in the thesis is given at the end of the manuscript.

CHAPTER 2

Doubly Spectral Decompositions of Functional Time Series

The purpose of this chapter is to develop doubly spectral decompositions for functional time series, that would generalize the properties of the Karhunen–Loève expansion to the functional time series setup. Although an earlier version of results presented in this chapter has been published (see Panaretos & Tavakoli 2013*a*), several results presented in this chapter have weaker assumptions, as discussed in Section 2.9.

2.1 Introduction

Although the Karhunen–Loève expansion plays a central role in functional data analysis, by providing a canonical decomposition of the random objects of interests as a series of orthogonal functions with random and uncorrelated amplitudes, whose truncation enjoys optimality properties, its use in the context of dependent functional data, such as functional time series (FTS), is not sensible. Indeed, focusing on the FTS case, we note that the Karhunen–Loève expansion is only based on the lag-0 autocovariance operator of the FTS, and does not take into account the lag-t autocovariance operators. This means that it only takes into account the marginal covariation structure of the series (contained in $\mathscr{R}_0$), and does not take into account any covariation across different time indices t—contained in $\mathscr{R}_t$, $t \neq 0$—which encode the dynamical properties of the

series. Furthermore, though the scores of the Karhunen–Loève expansion are uncorrelated within each time point t, they are not uncorrelated across time. The purpose of this chapter is to develop a decomposition for functional time series that takes into account all the autocovariance operators of the series, decomposes the series into a sum of components that are uncorrelated across all time lags, and which enjoys optimality properties when truncated at a finite level.

We begin this chapter with a heuristic overview of the results, which motivates the forthcoming developments, and gives some intuition, without going into the technicalities (Section 2.2). We then introduce in Section 2.3 the main objects upon which this chapter is based, the spectral density operators. In Section 2.4, we give a functional Cramér representation, which generalizes the Cramér representation to functional time series, and develop the theory for Cramér representation of linear filterings of an FTS in Section 2.5. The proofs of these two sections are given in Section 2.6, and is followed by a technical section on some measurability issues (Section 2.7). The culminating point of this chapter is Section 2.8, where the properties of the doubly spectral decomposition for functional time series (uncorrelatedness of the scores, optimality) are given. We also describe in Section 2.8.2 how an FTS can be represented as a vector time series with components that are uncorrelated across all time lags. A brief outlook of the results presented in this chapter, as well as some potential extensions, is given in Section 2.9.

2.2 Heuristics

The autocovariance operators $\{\mathscr{R}_t\}_{t\in\mathbb{Z}}$ of a second-order stationary functional time series X_t (taking values in a real part of the complexified Hilbert space H) encode the complete second-order structure of the time series $\{X_t\}_{t\in\mathbb{Z}}$, assumed to have mean zero. Corresponding to this sequence of operators, there may exist a collection of *operators* $\{\mathscr{F}_\omega\}_{\omega\in[-\pi,\pi]}$, called the *weak spectral density operators*, which satisfies

See Section A.2.3 on page 221 for the definition of complexified Hilbert space

$$\mathscr{R}_t = \int_{-\pi}^{\pi} \mathscr{F}_\omega e^{\mathbf{i}t\omega} d\omega, \quad \forall\, t \in \mathbb{Z}. \tag{2.2.1}$$

Provided the $\{\mathscr{R}_t\}_{t\in\mathbb{Z}}$ are summable in an appropriate sense, the weak spectral density operators exist and are defined as the discrete-time Fourier transform of the autocovariance operators,

The autocovariance operators $\mathscr{R}_t$ are defined on page 15

$$\mathscr{F}_\omega = \frac{1}{2\pi} \sum_{t\in\mathbb{Z}} e^{-\mathbf{i}\omega t} \mathscr{R}_t, \quad \omega \in [-\pi,\pi]. \tag{2.2.2}$$

(a rigorous definition—and sufficient conditions for its validity—will be given in Proposition 2.3.5). Now, assume that we can approximate the

integral in (2.2.1) by a Riemann sum, to get

$$\mathscr{R}_s = \int_{-\pi}^{\pi} e^{\mathbf{i}s\omega} \mathscr{F}_\omega d\omega \approx \sum_{j=1}^{J} \mathscr{F}_{\omega_j} e^{\mathbf{i}s\omega_j} (\omega_{j+1} - \omega_j),$$

where $-\pi = \omega_1 < \cdots < \omega_{J+1} = \pi$ is a partition. Then, we are naturally tempted to conjecture that X_t ought to be decomposable into a sum of distinct and uncorrelated frequency components,

$$X_t \approx \sum_{j=1}^{J} e^{\mathbf{i}\omega_j t} X_t(\omega_j), \tag{2.2.3}$$

where each $X_t(\omega_j)$ would be a mean-zero functional time series taking values in $L^2([0,1],\mathbb{C})$ with covariance operator close to $\mathscr{F}_{\omega_j}(\omega_{j+1} - \omega_j)$, since, in this case, X_t would indeed have covariance

$$\mathscr{R}_t = \sum_{j=1}^{J} \mathscr{F}_{\omega_j} e^{\mathbf{i}t\omega_j} (\omega_{j+1} - \omega_j).$$

We pursue such a decomposition in Section 2.4, where we formalize it as the functional Cramér representation (Theorem 2.4.3),

$$X_t = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} dZ_\omega, \quad \text{a.s.},$$

for a functional orthogonal increment process Z (independent of t), thus extending the classical Cramér representation of multivariate stationary processes (e.g. Brillinger (2001)).

The Cramér representation provides a spectral decomposition with respect to frequency. Nevertheless, we may pursue a second "layer" of spectral decomposition in terms of dimension. Going back to the heuristic form (2.2.3), we notice that, for each $j = 1,\ldots,J$, $X_t(\omega_j)$ is a random element of $L^2([0,1],\mathbb{C})$. We may thus represent it through its Karhunen–Loève (KL) expansion, leading to the heuristic representation

$$X_t \approx \sum_{j=1}^{J} e^{\mathbf{i}\omega_j t} \sum_{i=1}^{\infty} \xi_{i,j} \varphi_{i,j}(\tau), \tag{2.2.4}$$

with $\{\varphi_{i,j}\}_{i\geq 1}$ being the eigenfunctions of the covariance operator of $X_t(\omega_j)$ and $\{\xi_{i,j}\}_{i\geq 1}$ the corresponding Fourier coefficients. Truncating the second series at some $K < \infty$ will yield a decomposition into distinct frequency elements that are uncorrelated, and finite dimensional,

$$X_t \approx \sum_{j=1}^{J} e^{\mathbf{i}\omega_j t} \sum_{i=1}^{K} \xi_{i,j} \varphi_{i,j}(\tau). \tag{2.2.5}$$

The finite dimensional subspace in which each frequency component

takes its values need not be the same for distinct j's, even though each of them is of dimension K. In fact, it will turn out that this truncated representation only possesses K degrees of freedom. One would then hope that this reduced version of $X_t(\omega_j)$ would retain the property of being the optimal (in the L^2 sense) K-dimensional reduction of the process X_t. Non-rigorous versions of the decomposition (2.2.4), and its truncated version (2.2.5), are formally carried out in Section 2.4. Specifically, we derive the *Cramér–Karhunen–Loève* decomposition

$$X_t = \sum_{n=1}^{\infty} \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \left(\varphi_n^{\omega} \otimes_2 \varphi_n^{\omega}\right) dZ_{\omega} = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \sum_{n=1}^{\infty} \langle \varphi_n^{\omega}, dZ_{\omega} \rangle \varphi_n^{\omega}, \qquad (2.2.6)$$

where the first equality is rigorous—and is a natural extension of the Karhunen–Loève expansion to the FTS setup—and the last equality is understood formally (Remark 2.4.5, Theorem 2.8.6, and Remark 2.8.7). This is a Cramér representation with respect to frequency, but also a Karhunen–Loève expansion in terms of dimension, since it can be seen that $\{\varphi_n^{\omega}\}_{n\geq 1}$ is the basis of eigenfunctions of $\mathscr{F}_{\omega}$ (the covariance operator of dZ_{ω}). Furthermore, by considering the bounded operator-valued function

$$\sum_{n=1}^{K} \varphi_n^{\omega}(\tau) \otimes_2 \varphi_n^{\omega}(\sigma)$$

as a function over $[-\pi, \pi]$, and defining the notion of its stochastic integral (Section 2.5), we show that the truncated representation

$$X_t^* = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \left(\sum_{n=1}^{K} \varphi_n^{\omega} \otimes_2 \varphi_n^{\omega} \right) dZ_{\omega} \qquad (2.2.7)$$

is well defined, possesses K degrees of freedom, and converges to X_t in mean square as $K \to \infty$ (Section 2.8). More importantly, by considering the process X_t^* for different values of K, we obtain a *harmonic principal component analysis* of X_t. That is, we prove (Theorem 2.8.2 and Remark 2.8.5) that, among all linear reductions X_t to a process W_t of only K degrees of freedom, we have

$$\mathbb{E}\|X_t - X_t^*\|^2 \leq \mathbb{E}\|X_t - W_t\|^2.$$

Section 2.8 explains how the process $\{X_t^*\}$ can be constructed explicitly, when the spectral density estimator $\mathscr{F}_{\omega}$ is known, and how it can be represented as a stationary vector valued process *with uncorrelated coordinates* in $\mathbb{R}^K$ (see Remark 2.8.11 and Proposition 2.8.12).

Parallel to the rank K reduction, one may want to have a better finite dimensional approximation of $X_t(\omega_j)$ for some j's, and a cruder one for other j's, depending on how much each ω_j contributes to the power of the signal and/or the effective dimension of each $X_t(\omega_j)$. This can be done by letting the dimension K vary with j, leading to the heuristic

approximation

$$X_t \approx \sum_{j=1}^{J} e^{\mathbf{i}\omega_j t} X_t^{K_j}(\omega_j) = \sum_{j=1}^{J} e^{\mathbf{i}\omega_j t} \sum_{i=1}^{K_j} \xi_{i,j}\varphi_{i,j}, \tag{2.2.8}$$

where $X_t^{K_j}(\omega_j)$ is K_j-dimensional. It will turn out that such a representation is also rigorously valid (Theorem 2.8.2), and of the form

$$X_t^{**} = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \left(\sum_{n=1}^{K(\omega)} \varphi_n^{\omega} \otimes_2 \varphi_n^{\omega} \right) dZ_{\omega}$$

provided that the function $K : [-\pi, \pi] \to \{0, 1, \ldots\}$ yielding the desired finite rank for each frequency component is measurable. In fact, it will be shown, that among all linear transformations of the process $\{X_t\}$ having finite rank $K(\omega)$ at each frequency component, this is the optimal one, in the L^2 sense (Theorem 2.8.2).

2.3 The weak spectral density operators and the spectral density operators

Since we want to take a frequency domain approach to study the second-order structure of functional time series, encoded by the autocovariance operators $(\mathscr{R}_t)_{t\in\mathbb{Z}}$, we need to define their Fourier transforms, which we will call the *spectral density operators*. Similarly to classical Fourier analysis, we can define the spectral density operators either *implicitly*, by assuming the existence of a collection of objects (referred to as the *weak spectral density operators*) whose Fourier coefficients are the autocovariance operators (in an sense that will be made precise Definition 2.3.1 below), or *explicitly*, by defining the spectral density operators as a Fourier series, with the autocovariance operators as its coefficients, under summability conditions of the autocovariance operators (see Proposition 2.3.5).

Definition 2.3.1 (Weak spectral density operators)**.**
Let X_t be a second order stationary FTS in the real part of a complexified separable Hilbert space H, with mean zero, and $\mathbb{E}\|X_0\|^2 < \infty$. We denote by

$$\mathscr{R}_t = \mathbb{E}[X_t \otimes_2 X_0], \quad t \in \mathbb{Z}, \tag{2.3.1}$$

the lag-t autocovariance operator of X_t.
If there exists a function $\mathscr{F}_\cdot \in L^1([-\pi,\pi], \mathscr{S}_1(H))$ such that

$$\mathscr{R}_t = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \mathscr{F}_{\omega} d\omega, \quad t \in \mathbb{Z}, \tag{2.3.2}$$

then $\mathscr{F}_\cdot$ is called the weak spectral density operators *of X_t.*

$\mathscr{S}_1(H)$ denotes the space of trace-class operators on H, see Section A.2.2 on page 218

Remark 2.3.2.

1. *Notice that any functions f, g satisfying* (2.3.2) *will be equal almost everywhere, by Proposition B.0.16, and therefore the weak spectral density operators are well defined as an element of $L^1([-\pi,\pi], \mathscr{S}_1(H))$, but cannot be evaluated at any fixed $\omega \in [-\pi,\pi]$.*

2. *Since*

$$\int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \mathscr{F}_\omega^\dagger d\omega = \left(\int_{-\pi}^{\pi} e^{-\mathbf{i}\omega t} \mathscr{F}_\omega d\omega \right)^\dagger = \mathscr{R}_{-t}^\dagger = \mathscr{R}_t,$$

the weak spectral density operators are almost everywhere self-adjoint.

If there exists a function $\mathscr{F}_\cdot \in L^1([-\pi,\pi], \mathscr{S}_1(H))$ such that $\omega \mapsto \mathscr{F}_\omega$ is continuous with respect to the operator norm $|||\cdot|||_\infty$ and $\mathscr{F}_\omega$ satisfies (2.3.2), then this function is called the *spectral density operators* of X_t. $\mathscr{F}_\omega$ will be called the *spectral density operator at* ω. The difference between the weak spectral density operators and the spectral density operators is that latter can be evaluated at any $\omega \in [-\pi,\pi]$, whereas the weak spectral density operators are defined only in a weak sense, since they belong to an L^1 space.
We now define some conditions under which the spectral density operators of an FTS exist, and is given by the Fourier series of its autocovariance operators.

Condition 2.3.3.
H is a complexified separable Hilbert space, and X_t is a second-order stationary FTS in the real part of H, with mean zero, $\mathbb{E}\|X_0\|^2 < \infty$, and satisfies

$$\sum_{t\in\mathbb{Z}} |||\mathscr{R}_t|||_\infty < \infty$$

Condition 2.3.4.

$$\sum_{t\in\mathbb{Z}} |\mathrm{Tr}(\mathscr{R}_t)| < \infty.$$

Proposition 2.3.5.
Assume Condition 2.3.3 holds. Then, the spectral density operators of X_t are given by

$$\mathscr{F}_\omega = \frac{1}{2\pi} \sum_{t\in\mathbb{Z}} e^{-\mathbf{i}\omega t} \mathscr{R}_t, \quad \omega \in [-\pi,\pi], \tag{2.3.3}$$

where the convergence holds in $|||\cdot|||_\infty$, uniformly in ω. $\mathscr{F}_\omega$ is well defined, continuous in ω (with respect to $|||\cdot|||_\infty$), non-negative and compact *for all $\omega \in [-\pi,\pi]$. It satisfies the inversion formula*

$$\mathscr{R}_t = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \mathscr{F}_\omega d\omega, \quad t \in \mathbb{Z}. \tag{2.3.4}$$

Furthermore, if Condition 2.3.4 also holds, then

1. *the nuclear norm of all the spectral density operators is uniformly bounded:*

$$\sup_{\omega\in[-\pi,\pi]} |||\mathscr{F}_\omega|||_1 \leq (2\pi)^{-1} \sum_{s\in\mathbb{Z}} |\mathrm{Tr}\,(\mathscr{R}_s)| < \infty.$$

2. *The nuclear norm of all the lag-t autocovariance operators are uniformly bounded:*

$$\sup_{t\in\mathbb{Z}} |||\mathscr{R}_t|||_1 \leq \sum_{s\in\mathbb{Z}} |\mathrm{Tr}\,(\mathscr{R}_s)| < \infty.$$

3. *$\omega \mapsto \mathscr{F}_\omega$ is $|||\cdot|||_1$-measurable.*

4. *For almost every $\omega \in [-\pi,\pi]$,*

$$\mathrm{Tr}\,(\mathscr{F}_\omega) = \sum_{t\in\mathbb{Z}} e^{-\mathbf{i}\omega t}\mathrm{Tr}\,(\mathscr{R}_t).$$

In other words, $\omega \mapsto \mathrm{Tr}\,(\mathscr{F}_\omega)$ is equal to a continuous function almost everywhere.

Proof. Assume Condition 2.3.3 holds. Let us write

$$A^{(T)}(\omega) = (2\pi)^{-1} \sum_{t<|T|} e^{-\mathbf{i}\omega t}\mathscr{R}_t.$$

$A^{(T)}(\omega)$ is uniformly continuous in ω, and by the triangle inequality, it is a Cauchy sequence in $\mathscr{S}_\infty(H)$, and converges to $\mathscr{F}_\omega$ uniformly in ω. Since $\mathscr{F}_\cdot$ is the uniform limit of uniformly continuous functions, it is also uniformly continuous (with respect to $|||\cdot|||_\infty$). Now let

$$p_\omega^{(T)} = T^{-1} \sum_{s,t=0}^{T-1} e^{-\mathbf{i}\omega(t-s)} X_t \otimes_2 X_s = T^{-1} \left(\sum_{t=0}^{T-1} e^{-\mathbf{i}\omega t} X_t\right) \otimes_2 \left(\sum_{s=0}^{T-1} e^{-\mathbf{i}\omega s} X_s\right).$$

We see that $\mathbb{E}\, p_\omega^{(T)}$ is a positive symmetric operator, and using the second-order stationarity, we get

$$\mathbb{E}\, p_\omega^{(T)} = T^{-1}\left(A^{(0)}(\omega) + A^{(1)}(\omega) + \cdots + A^{(T-1)}(\omega)\right);$$

see Lemma 3.4.2. In particular $|||\,\mathbb{E}\, p_\omega^{(T)}|||_1 < \infty$, $\mathbb{E}\, p_\omega^{(T)}$ is compact, and

$$\lim_{T\to\infty} \mathbb{E}\, p_\omega^{(T)} \to \mathscr{F}_\omega,$$

in $|||\cdot|||_\infty$, since it is a Cesaro-sum of a convergent sequence. Therefore, the spectral density operators are compact operators (since compact operators form a closed subspace of the space of bounded operators).

Furthermore, for any $\varphi \in H$,

$$\langle \mathscr{F}_\omega \varphi, \varphi \rangle = \lim_{T\to\infty} \langle (\mathbb{E}\, p_\omega^{(T)})\varphi, \varphi \rangle \geq 0,$$

which shows that the spectral density operators are all non-negative.
Let us now turn to the inversion formula. First, notice that $\int_{-\pi}^{\pi} \mathscr{F}_\omega e^{\mathbf{i}\omega s} d\omega$ is well defined, since $|||\mathscr{F}_\omega|||_\infty \leq \sum_{t\in\mathbb{Z}} |||\mathscr{R}_t|||_\infty < \infty$. Let $\phi : \mathscr{S}_\infty(H) \to \mathbb{C}$ be a continuous linear functional (an element of the dual of $\mathscr{S}_\infty(H)$). We have

$$\begin{aligned}\phi\left(\int_{-\pi}^{\pi} \mathscr{F}_\omega e^{\mathbf{i}\omega s} d\omega\right) &= \int_{-\pi}^{\pi} \phi(\mathscr{F}_\omega) e^{\mathbf{i}\omega s} d\omega \\ &= \int_{-\pi}^{\pi} \lim_{T\to\infty} \phi\big(A^{(T)}(\omega)\big) e^{\mathbf{i}\omega s} d\omega.\end{aligned}$$

Since $\left|\phi\left(A^{(T)}(\omega)\right)\right| \leq \left|\left|\left|\phi\right|\right|\right|_\infty \sum_{t\in\mathbb{Z}} |||\mathscr{R}_t|||_\infty < \infty$, the dominated convergence theorem yields

$$\begin{aligned}\phi\left(\int_{-\pi}^{\pi} \mathscr{F}_\omega e^{\mathbf{i}\omega s} d\omega\right) &= (2\pi)^{-1} \lim_{T\to\infty} \sum_{t=-T}^{T} \phi(\mathscr{R}_t) \int_{-\pi}^{\pi} e^{\mathbf{i}\omega(s-t)} d\omega \\ &= \phi(\mathscr{R}_s).\end{aligned}$$

Therefore, since the previous equality holds for all ϕ in the dual of $\mathscr{S}_\infty(H)$, we get

$$\mathscr{R}_s = \int_{-\pi}^{\pi} \mathscr{F}_\omega e^{\mathbf{i}\omega s} d\omega, \quad s \in \mathbb{Z}.$$

Now assume that Condition 2.3.4 also holds. Let $(e_n)_{n\geq 1}$ be an orthonormal basis of H. Since the spectral density operators are non-negative, the continuity of the scalar product and Fatou's lemma yield

$$\begin{aligned}|||\mathscr{F}_\omega|||_1 = \sum_{n\geq 1} \langle \mathscr{F}_\omega e_n, e_n \rangle &= \sum_{n\geq 1} \lim_{T\to\infty} \langle (\mathbb{E}\, p_\omega^{(T)}) e_n, e_n \rangle \\ &\leq \liminf_{T\to\infty} \sum_{n\geq 1} \langle (\mathbb{E}\, p_\omega^{(T)}) e_n, e_n \rangle \\ &= \liminf_{T\to\infty} \mathrm{Tr}\left(\mathbb{E}\, p_\omega^{(T)}\right) \\ &= \liminf_{T\to\infty} \mathbb{E}\,\mathrm{Tr}\left(p_\omega^{(T)}\right) \\ &= \liminf_{T\to\infty} (2\pi T)^{-1} \sum_{s,t=0}^{T-1} e^{-\mathbf{i}\omega(t-s)}\, \mathbb{E}\,\langle X_t, X_s \rangle \\ &= \liminf_{T\to\infty} (2\pi)^{-1} \sum_{|t|<T} \left(1 - \frac{|t|}{T}\right) e^{\mathbf{i}\omega t} \mathrm{Tr}\,(\mathscr{R}_t).\end{aligned}$$

Since $\sum_{t\in\mathbb{Z}} |\mathrm{Tr}\,(\mathscr{R}_t)| < \infty$, the dominated convergence theorem yields

$$|||\mathscr{F}_\omega|||_1 \leq (2\pi)^{-1} \sum_{t\in\mathbb{Z}} e^{\mathbf{i}\omega t} \mathrm{Tr}\,(\mathscr{R}_t) \leq (2\pi)^{-1} \sum_{t\in\mathbb{Z}} |\mathrm{Tr}\,(\mathscr{R}_t)| < \infty.$$

Taking the nuclear norm of this equation yields

$$|||\mathscr{R}_t|||_1 \le \int_{-\pi}^{\pi} |||\mathscr{F}_\omega|||_1 d\omega \le \sum_{s\in\mathbb{Z}} |\mathrm{Tr}(\mathscr{R}_s)| < \infty.$$

Notice that the right-hand side is independent of t. To show that $\omega \mapsto \mathscr{F}_\omega$ is $|||\cdot|||_1$-measurable, we use Lemma B.0.8: since $\mathscr{S}_1(H)$ is separable, and its topological dual consists of all the functionals $\phi_T : \mathscr{S}_1(H) \to \mathbb{C}$, $T \in \mathscr{S}_\infty(H)$, where

$$\phi_T(A) = \mathrm{Tr}(TA), \quad A \in \mathscr{S}_1(H),$$

we only need to show that for all $T \in \mathscr{S}_\infty(H)$, the complex-valued function $\omega \mapsto \phi_T(\mathscr{F}_\omega) \in \mathbb{C}$ is measurable. Since

$$\phi_T(\mathscr{F}_\omega) = \sum_{n\ge 1} \langle T\mathscr{F}_\omega e_n, e_n\rangle,$$

where the series converges everywhere, and each function $\omega \mapsto \langle T\mathscr{F}_\omega e_n, e_n\rangle$ is continuous and positive, $\omega \mapsto \phi_T(\mathscr{F}_\omega)$ is measurable.
For the final statement, we use Proposition B.0.16: since

$$\mathrm{Tr}(\mathscr{R}_s) = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega s}\mathrm{Tr}(\mathscr{F}_\omega)d\omega = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega s}G(\omega)d\omega, \quad s \in \mathbb{Z},$$

where $G(\omega) = (2\pi)^{-1}\sum_{t\in\mathbb{Z}} e^{-\mathbf{i}\omega t}\mathrm{Tr}(\mathscr{R}_t)$, we get that $\mathrm{Tr}(\mathscr{F}_\omega) = G(\omega)$ for almost every $\omega \in [-\pi,\pi]$. Since G is an absolutely and uniformly convergent series of continuous functions on a compact set, G is uniformly continuous. This completes the proof. □

2.4 Functional Cramér representation

Now that we have defined the spectral density operators, we construct a Cramér representation for functional time series, i.e., a *functional Cramér representation*. Further to being a milestone in the development of the doubly spectral decomposition that shall be given in Section 2.8, the functional Cramér representation is important on its own because it tells us that any second-order stationary functional time series admitting weak spectral density operators can be decomposed into a superposition of uncorrelated processes fluctuating at distinct frequencies.
The following condition, defined for $p \ge 1$, will be typically assumed for the results of this section.

Condition 2.4.1 (p)**.** *$(X_t)_{t\in\mathbb{Z}}$ is a second-order stationary time series in the real part of a complexified separable Hilbert space H (i.e., an FTS) with mean zero, $\mathbb{E}\|X_0\|^2 < \infty$, and X_t admits weak spectral density operators*

$\mathscr{F}_\cdot \in L^p([-\pi,\pi],\mathscr{S}_1(H))$, *i.e.*

$$\mathbb{E}[X_t \otimes_2 X_0] = \mathscr{R}_t = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t}\mathscr{F}_\omega d\omega, \quad \forall t \in \mathbb{Z}. \tag{2.4.1}$$

Remark 2.4.2.

1. *Notice that Conditions 2.3.3 and 2.3.4 imply Condition 2.4.1*(∞)*, and that if* $1 \le p < q$*, then*

$$\textit{Conditions 2.4.1}(q) \implies \textit{Conditions 2.4.1}(p).$$

2. *Notice that conditions 2.4.1*(p) *with* $p \in [1,\infty)$ *do not imply that the trace norm of the weak spectral density operators are bounded.*

3. *If* X_t *satisfies Conditions 2.4.1*(p)*, with* $p \ge 1$*, then by Hölder's inequality, the nuclear norm of the lag-*t *autocovariance operators is uniformly bounded:*

$$\sup_{t\in\mathbb{Z}} |||\mathscr{R}_t|||_1 \le \int_{-\pi}^{\pi} |||\mathscr{F}_\omega|||_1 d\omega < \infty.$$

We shall now give the functional version of the Cramér representation. For this purpose, let us denote by $\mathbb{H}$ the space of all random elements of H with finite second moment, i.e.

$$\mathbb{H} = L^2(\Omega, H, \mathbb{P}) = \left\{Y \text{ random element of } H : \mathbb{E}\,\|Y\|^2 < \infty\right\}.$$

Let us define $\langle Y, Z\rangle_{\mathbb{H}} = \mathbb{E}\langle Y, Z\rangle$ for $Y, Z \in \mathbb{H}$, and $\|Y\|_{\mathbb{H}} = \sqrt{\langle Y, Y\rangle_{\mathbb{H}}}$ the corresponding norm. This defines actually a scalar product on $\mathbb{H}$ (after identification of random elements that are almost surely equal), and $(\mathbb{H}, \langle\cdot,\cdot\rangle_{\mathbb{H}})$ is in fact a Hilbert space (in particular it is complete).

Theorem 2.4.3 (Functional Cramér Representation)**.**
Assume Conditions 2.4.1(p) *hold for some* $p \in (1,\infty]$*. Then* X_t *admits the representation*

$$X_t = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} dZ_\omega, \quad a.s., \tag{2.4.2}$$

where for fixed ω*,* Z_ω *is a random element of* H*, defined by*

$$Z_\omega = \lim_{T\to\infty} \sum_{|t|<T} \left(1 + \frac{|t|}{T}\right) g_\omega(t) X_{-t}, \quad in\ \mathbb{H}, \tag{2.4.3}$$

where

$$g_\omega(t) = (2\pi)^{-1} \int_{-\pi}^{\omega} e^{-\mathbf{i}\alpha t} d\alpha.$$

The random process $[-\pi,\pi] \ni \omega \mapsto Z_\omega \in \mathbb{H}$ *satisfies*

1. $Z_{-\pi} = 0$

2. $Z_\pi = X_0$
3. *For every* $\alpha, \beta \in [-\pi, \pi]$,

$$Z_\alpha - Z_\beta = \overline{Z_{-\beta} - Z_{-\alpha}} \tag{2.4.4}$$

and has a covariance structure given by

$$\mathbb{E}\left[Z_\alpha \otimes_2 Z_\beta\right] = \int_{-\pi}^{\min(\alpha,\beta)} \mathscr{F}_\omega d\omega. \tag{2.4.5}$$

In particular, Z_ω *has orthogonal increments:*

$$\mathbb{E}\left\langle Z_{\omega_1} - Z_{\omega_2}, Z_{\omega_3} - Z_{\omega_4}\right\rangle = 0, \quad \text{if } \omega_1 > \omega_2 \geq \omega_3 > \omega_4. \tag{2.4.6}$$

The representation (2.4.2) *is called the* Cramér representation *of* X_t.

Remark 2.4.4. *If Condition 2.4.1*(∞) *holds, then* (2.4.3) *simplifies to*

$$Z_\omega = (2\pi)^{-1} \lim_{T\to\infty} \sum_{|t|<T} X_{-t} \int_{-\pi}^{\omega} e^{-\mathbf{i}\alpha t} d\alpha, \quad \text{in } \mathbb{H}. \tag{2.4.7}$$

Furthermore, the stochastic integral (2.4.2) *can be understood as a Riemann-Stieltjes limit, in the sense that*

$$\mathbb{E}\left\| X_t - \sum_{j=1}^{J} e^{\mathbf{i}\omega_j t}(Z_{\omega_{j+1}} - Z_{\omega_j}) \right\|^2 \to 0, \quad \text{as } J \to \infty, \tag{2.4.8}$$

where $-\pi = \omega_1 < \cdots < \omega_{J+1} = \pi$ *and* $\max_{j=1,\ldots,J} |\omega_{j+1} - \omega_j| \to 0$ *as* $J \to \infty$.

This last remark formalizes the idea of decomposing X_t into distinct frequencies: setting $X_t(\omega_j) = e^{\mathbf{i}\omega_j t}(Z_{\omega_{j+1}} - Z_{\omega_j})$, we have

$$X_t \approx \sum_{j=1}^{J} X_t(\omega_j),$$

where the approximation error can be made as small as one wishes, with respect to the $\|\cdot\|_{\mathbb{H}}$ norm.

Remark 2.4.5 (Towards the Cramér–Karhunen–Loève Decomposition)**.** *Let us denote by*

$$\mathscr{F}_\omega = \sum_{n=1}^{\infty} \mu_n(\omega) \varphi_n^\omega \otimes_2 \varphi_n^\omega$$

the singular value decomposition of the spectral density operators, for almost every $\omega \in [-\pi, \pi]$. *If the spectral density operator* $\mathscr{F}_\omega$ *is strictly positive-definite almost everywhere, we may abuse notation and write*

$$X_t = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \left(\sum_{n=1}^{\infty} \varphi_n^\omega \otimes_2 \varphi_n^\omega \right) dZ_\omega, \tag{2.4.9}$$

where $\{Z_\omega\}$ is as in Theorem 2.4.3. Provided we can exchange the sum and the integral, we have

$$X_t = \sum_{n=1}^{\infty} \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} (\varphi_n^\omega \otimes_2 \varphi_n^\omega) d\omega. \tag{2.4.10}$$

Although the representation in Remark 2.4.5 is a mere reformulation of Theorem 2.4.3, and is not rigorous, it sets the scene for the question of the nature of the approximation of $\{X_t\}$ that might arise if we where able to truncate the identity operator $\sum_{n=1}^{\infty} \varphi_n^\omega \otimes_2 \varphi_n^\omega$ to have finite rank

$$X_t^* := \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \left(\sum_{n=1}^{K} \varphi_n^\omega \otimes_2 \varphi_n^\omega \right) dZ_\omega, \tag{2.4.11}$$

i.e. to consider the limiting behaviour of $\mathbb{E}\|X_t - X_t^*\|^2$ as $K \to \infty$ (See Theorem 2.8.6). It is such truncations (and their approximation error) that are at the essence of representations of the Karhunen–Loève type. We develop in Section 2.5 the formalism to make sense of an integral of the form (2.4.11) (it is not a priori clear that it is well-defined, since now the operator in the integrand depends on ω), and prove in Section 2.8 that it provides an optimal rank K approximation of the original process, yielding a *harmonic principal component analysis* of the process X_t (in fact, we will not require that $\mathscr{F}_\omega$ be strictly positive).

Remark 2.4.6.
Note that the action of the operator $\sum_{n=1}^{\infty} \varphi_n^\omega \otimes_2 \varphi_n^\omega$ on an element $g \in L^2([0,1],\mathbb{C})$ is described by

$$\left[\sum_{n=1}^{\infty} \varphi_n^\omega \otimes_2 \varphi_n^\omega \right] g = \sum_{n=1}^{\infty} \langle g, \varphi_n^\omega \rangle \varphi_n^\omega.$$

Therefore, we may formally interpret the Cramér–Karhunen–Loève representation as

$$X_t = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \sum_{n=1}^{\infty} \langle dZ_\omega, \varphi_n^\omega \rangle \varphi_n^\omega,$$

a form which emphasizes the doubly spectral decomposition of $\{X_t\}$ as discussed in Section 2.2.

2.5 Linear Filtering, Stochastic Integrals of Operator Valued Functions

Notice that if $a \in \mathscr{S}_\infty(H)$, we can show that $aX_t = \int_{-\pi}^{\pi} a e^{\mathbf{i}\omega t} dZ_\omega$, using (2.4.8). If we define a new FTS $(Y_t)_{t\in\mathbb{Z}}$ by linear filtering of X_t, i.e.

$$Y_t = \sum_{s\in\mathbb{Z}} a_s X_{t-s}, \quad t \in \mathbb{Z},$$

it would be useful to have a Cramér representation for Y_t. Formally, we would like to do

$$\begin{aligned}\sum_s a_s X_{t-s} &= \sum_s a_s \int_{-\pi}^{\pi} e^{\mathbf{i}\omega(t-s)} dZ_\omega \\ &= \int_{-\pi}^{\pi} \left(\sum_s a_s e^{-\mathbf{i}\omega s}\right) e^{\mathbf{i}\omega t} dZ_\omega.\end{aligned}$$

This, together with Remark 2.4.5, motivates giving a meaning to stochastic integrals of the form

$$\int_{-\pi}^{\pi} A(\omega) dZ_\omega, \tag{2.5.1}$$

where $A : [-\pi,\pi] \to \mathscr{S}_\infty(H)$. In fact, if $\mathscr{F}_\cdot \in L^p([-\pi,\pi], \mathscr{S}_1(H))$, we will be able to give a meaning to this stochastic integral for all $A \in \mathfrak{H}$, where $\mathfrak{H}$ is the completion of $L^{2q}([-\pi,\pi], \mathscr{S}_\infty(H))$ under the norm $\|\cdot\|_{\mathfrak{H}} = \sqrt{\langle \cdot, \cdot \rangle_{\mathfrak{H}}}$,

$$\langle A, B \rangle_{\mathfrak{H}} = \int_{-\pi}^{\pi} \mathrm{Tr}\Big(A(\omega)\mathscr{F}_\omega B^\dagger(\omega)\Big) d\omega, \quad \forall\, A, B \in \mathfrak{H}.$$

and $q \geq 1$ such that $p^{-1} + q^{-1} = 1$. We note in particular that by Hölder's inequality,

$$L^{2q}([-\pi,\pi], \mathscr{S}_\infty(H)) \subset \mathfrak{H}.$$

For a sequence $(A_n)_{n\geq 1} \subset \mathfrak{H}$, and $A \in \mathfrak{H}$, we will say that

$$\lim_{n\to\infty} A_n = A, \quad \text{in } \mathfrak{H},$$

if $\lim_{n\to\infty} \|A_n - A\|_{\mathfrak{H}} = 0$.

More details about the construction of the space $\mathfrak{H}$ can be found in Section 2.6. The definition of the stochastic integral (2.5.1) is made precise by the following Theorem.

Theorem 2.5.1. *For any $A \in \mathfrak{H}$, there exists a (not necessarily unique) triangular array $(a_{T,t})_{|t|<T} \subset \mathscr{S}_\infty(H)$, $T = 1, 2, \ldots$ such that*

$$\lim_{T\to\infty} \left\| A - \sum_{|t|<T} a_{T,t} \mathfrak{e}_{-t} \right\|_{\mathfrak{H}} = 0, \tag{2.5.2}$$

where $\mathfrak{e}_s : \mathbb{R} \to \mathbb{C}$ is defined by $\mathfrak{e}_s(\omega) = \exp(\mathbf{i}s\omega)$. For any such triangular array, we define $\int_{-\pi}^{\pi} A(\omega) dZ_\omega \in \mathbb{H}$ to be the unique element satisfying

$$\mathbb{E}\left\| \int_{-\pi}^{\pi} A(\omega) dZ_\omega - \sum_{|t|<T} a_{T,t} X_{-t} \right\|^2 \to 0, \quad T \to \infty. \tag{2.5.3}$$

In particular, the stochastic integral $\int_{-\pi}^{\pi} A(\omega) dZ_\omega$ does not depend on the choice of array that satisfies (2.5.2).

Furthermore, we have

$$\overline{\int_{-\pi}^{\pi} A(\omega)\,dZ_\omega} = \int_{-\pi}^{\pi} \overline{A(-\omega)}\,dZ_\omega, \quad a.s., \tag{2.5.4}$$

$\mathbb{E}\int_{-\pi}^{\pi} A(\omega)\,dZ_\omega = 0$*, and the covariance structure of the stochastic integral is given by*

$$\mathbb{E}\left[\int_{-\pi}^{\pi} A(\omega)\,dZ_\omega \otimes_2 \int_{-\pi}^{\pi} B(\omega)\,dZ_\omega\right] = \int_{-\pi}^{\pi} A(\omega)\mathscr{F}_\omega B^\dagger(\omega)\,d\omega, \quad A,B \in \mathfrak{H}. \tag{2.5.5}$$

Remark 2.5.2.
Notice that equation (2.5.5) *implies that*

$$\mathbb{E}\left\langle \int_{-\pi}^{\pi} A(\omega)\,dZ_\omega, \int_{-\pi}^{\pi} B(\omega)\,dZ_\omega \right\rangle = \int_{-\pi}^{\pi} \mathrm{Tr}\Big(A(\omega)\mathscr{F}_\omega B^\dagger(\omega)\Big)\,d\omega, \quad \forall\, A,B \in \mathfrak{H}, \tag{2.5.6}$$

or put more simply, $\left\langle \int_{-\pi}^{\pi} A(\omega)\,dZ_\omega, \int_{-\pi}^{\pi} B(\omega)\,dZ_\omega \right\rangle_{\mathbb{H}} = \langle A,B\rangle_{\mathfrak{H}}$*. This identity is closely related to the proof of the Theorem, which is based on the construction of a unitary transformation between the spaces* $\mathbb{H}$ *and* $\mathfrak{H}$*. Therefore, equation* (2.5.6) *will be called the* isometry property*. The unitary transformations between* $\mathbb{H}$ *and* $\mathfrak{H}$ *are denoted by*

$$\mathbb{H} \underset{\mathscr{I}}{\overset{\mathscr{T}}{\rightleftarrows}} \mathfrak{H}\ ,$$

$\mathscr{I}(A) = \int_{-\pi}^{\pi} A(\omega)\,dZ_\omega$*,* $\mathscr{T}\left(\int_{-\pi}^{\pi} A(\omega)\,dZ_\omega\right) = A$*,* $\mathscr{T}\circ\mathscr{I} = \mathrm{Id}_{\mathfrak{H}}$*, and* $\mathscr{I}\circ\mathscr{T} = \mathrm{Id}_{\mathbb{H}}$*. Details of the construction of these mappings can be found in Section 2.6.*

Understanding Theorem 2.5.1 is not straightforward, due to its generality. Indeed, it is possible to choose a triangular array $(a_{T,t})_{|t|<T} \subset \mathscr{S}_\infty(H)$ such that (2.5.2) holds, but with $\left|\!\left|\!\left| a_{T,0} \right|\!\right|\!\right|_\infty \to \infty$, and with the limit function $A \in \mathfrak{H}$ not belonging to $L^{2q}([-\pi,\pi],\mathscr{S}_\infty(H))$. However, the restriction of the stochastic integral to $L^{2q}([-\pi,\pi],\mathscr{S}_\infty(H))$ can be viewed as a Cesaro-sum of a truncated Fourier series:

Theorem 2.5.3. *Let* X_t *satisfy Conditions 2.4.1(p) for some* $p \in (1,\infty]$*, and let* $q \in [1,\infty)$ *satisfy* $p^{-1}+q^{-1}=1$ *(*$q=1$ *if* $p=\infty$*). Then for any* $A \in L^{2q}([-\pi,\pi],\mathscr{S}_\infty(H))$*, we define the stochastic integral* $\int_{-\pi}^{\pi} A(\omega)\,dZ_\omega$ *to be the unique (a.s.) random element of* $\mathbb{H}$ *satisfying*

$$\int_{-\pi}^{\pi} A(\omega)\,dZ_\omega = \lim_{T\to\infty} \sum_{|t|<T} \left(1 - \frac{|t|}{T}\right) a_t X_{-t}, \quad \text{in } \mathbb{H}. \tag{2.5.7}$$

where

$$a_t = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} A(\omega)\,d\omega \in \mathscr{S}_\infty(H), \quad t \in \mathbb{Z}. \tag{2.5.8}$$

If $\sum_{t\in\mathbb{Z}} |||a_t|||_\infty < \infty$, *then*

$$\int_{-\pi}^{\pi} A(\omega)\, dZ_\omega = \lim_{T\to\infty} \sum_{|t|<T} a_t X_{-t}, \quad \text{in } \mathbb{H} \tag{2.5.9}$$

Remark 2.5.4.
If Condition 2.4.1(∞) *holds, and if* $\omega \mapsto A(\omega)$ *is càdlàg with respect to* $|||\cdot|||_\infty$, *with a finite number of jumps, then the stochastic integral* (2.5.1) *can be understood as a Riemann-Stieltjes limit, in the sense that for every* $\varepsilon > 0$, *there exists a partition* $P = \{\omega_1, \dots, \omega_k\}$ *of* $[-\pi, \pi]$ *such that for any finer partition* $P' = \{\omega'_1, \dots, \omega'_K\}$ *of* $[-\pi, \pi]$ *(i.e.* $P \subset P'$*), and every choice of points* $\lambda_i \in [\omega'_i, \omega'_{i+1}]$, *we have*

$$\mathbb{E}\left\| \int_{-\pi}^{\pi} A(\omega)\, dZ_\omega - \sum_{j=1}^{J} A(\lambda_j)(Z_{\omega'_{j+1}} - Z_{\omega'_j}) \right\|^2 < \varepsilon^2. \tag{2.5.10}$$

We now have the tools for stating a result about linear filters of a second-order stationary FTS.

Theorem 2.5.5. *Let* X_t *satisfy Conditions 2.4.1(p) for some* $p \in (1, \infty]$, *with Cramér representation*

$$X_t = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t}\, dZ^X_\omega, \tag{2.5.11}$$

and spectral density operators $\mathscr{F}^X_\omega$.
Let $(a_{T,t})_{|t|<T} \subset \mathscr{S}_\infty(H), T = 1, 2, \dots$ *be a triangular array such that*

$$A(\omega) = \lim_{T\to\infty} \sum_{|s|<T} e^{-\mathbf{i}\omega s} a_{T,s}, \quad \text{in } \mathfrak{H}, \tag{2.5.12}$$

for some $A \in \mathfrak{H}$ *(see Remark 2.5.6 below).*
Then, $Y_t = \lim_{T\to\infty} \sum_{|s|<T} a_{T,s} X_{t-s}$ *converges in* $\mathbb{H}$, *is second-order stationary with mean zero, and*

1. Y_t *admits the representation*

$$Y_t = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} A(\omega)\, dZ^X_\omega,$$

where $A(\omega) = \lim_{T\to\infty} \sum_{|s|<T} e^{-\mathbf{i}\omega s} a_{T,s}$ *in* $\mathfrak{H}$.

2. *The weak spectral density operators of* Y_t *are given by*

$$\mathscr{F}^Y_\omega = A(\omega)\mathscr{F}^X_\omega A^\dagger(\omega), \quad \omega \in [-\pi, \pi], \tag{2.5.13}$$

$\mathscr{F}^Y_\cdot \in L^1([-\pi, \pi], \mathscr{S}_1(H))$, *and the inversion formula holds for* Y_t,

$$\mathscr{R}^Y_t = \mathbb{E}[Y_t \otimes_2 Y_0] = \int_{-\pi}^{\pi} \mathscr{F}^Y_\omega e^{\mathbf{i}\omega t}\, d\omega, \quad \forall\, t \in \mathbb{Z}.$$

3. *If for some $q \geq p/(p-1)$, or $q = \infty$,*

$$\lim_{T\to\infty} \sum_{|t|<T} e^{-\mathbf{i}\omega n} a_{T,t} = A(\omega), \quad \text{in } L^{2q}([-\pi,\pi], \mathscr{S}_\infty(H)), \tag{2.5.14}$$

then $\mathscr{F}^Y_\omega \in L^r([-\pi,\pi], \mathscr{S}_1(H))$, where $r^{-1} = p^{-1} + q^{-1}$.

Remark 2.5.6.

1. *A condition equivalent to* (2.5.12) *is that, for every $\varepsilon > 0$, there exists an $N > 0$ such that for all $T' \geq T > N$,*

$$\sum_{|s|,|u|<T'} \mathrm{Tr}\left((a_{T',s} - a_{T,s})\mathscr{R}^X_{u-s}(a_{T',u} - a_{T,u})^\dagger\right) < \varepsilon, \tag{2.5.15}$$

where we let $a_{T,s} = 0$ if $s \geq T$. The sum on the left-hand side is guaranteed to be real and non-negative.

2. *If the triangular array is actually just a sequence $(a_t)_{t\in\mathbb{Z}}$, then condition* (2.5.15) *simplifies to*

$$\sum_{T\leq|s|,|u|<T'} \mathrm{Tr}\left(a_s \mathscr{R}^X_{u-s} a_u^\dagger\right) < \varepsilon. \tag{2.5.16}$$

A sufficient condition for this to hold is

$$\sum_{s\in\mathbb{Z}} |||a_s|||_\infty < \infty, \tag{2.5.17}$$

which implies that (2.5.14) *holds for $q = \infty$.*

3. *If X_t is m-correlated, i.e. $\mathscr{R}^X_t = 0$ for $|t| > m$, then a sufficient condition for* (2.5.16) *is*

$$\sum_{s\in\mathbb{Z}} |||a_s|||^2_\infty < \infty. \tag{2.5.18}$$

4. *If Conditions 2.3.3 and 2.3.4 hold, and the triangular array is just a sequence $(a_t)_{t\in\mathbb{Z}}$ satisfying* (2.5.17)*, then $\sum_{t\in\mathbb{Z}} \left|\left|\left|\mathscr{R}^Y_t\right|\right|\right|_\infty < \infty$.*

This result tells us that any linear filtering of the X_t with a filter satisfying (2.5.14) with $q > p/(p-1)$ will fulfill Conditions 2.4.1(r) with $r > 1$, and admits therefore its own Cramér representation

$$Y_t = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t}\, dZ^Y_\omega.$$

We can therefore say formally that

$$dZ^Y_\omega = A(\omega)\, dZ^X_\omega.$$

2.6 Proofs of Sections 2.4 and 2.5

In this section, we shall prove the results of Sections 2.4 and 2.5. We assume throughout that X_t satisfies Conditions 2.4.1(p), for some $p \in (1,\infty]$, and that q is conjugate to p, i.e. $p^{-1}+q^{-1}=1$. Notice that $q \in [1,\infty)$. The letter c shall denote the constant

$$c = \left(\int_{-\pi}^{\pi} |||\mathscr{F}_\omega|||_1^p \, d\omega \right)^{1/p} < \infty.$$

2.6.1 Definition of the Spaces Involved

We define the following subspace of $\mathbb{H}$,

$$\mathbb{M}_0 = \left\{ \sum_{l=-n}^{n} a_l X_{-l} : a_l \in \mathscr{S}_\infty(H) \text{ for all } |l| \le n; n = 1,2,\dots \right\}, \tag{2.6.1}$$

which consists of all the finite linear filterings of the time series X_t with bounded operators. Let $\mathbb{M} \subset \mathbb{H}$ be the completion of $\mathbb{M}_0$ under the norm $\|\cdot\|_{\mathbb{H}}$, and $\mathfrak{L}$ be the Banach space $L^{2q}([-\pi,\pi], \mathscr{S}_\infty(H))$, with norm

$$\|A\|_{\mathfrak{L}} = \left(\int_{-\pi}^{\pi} |||A(\omega)|||_\infty^{2q} \, d\omega \right)^{1/2q}, \quad A \in \mathfrak{L}. \tag{2.6.2}$$

We also define the mapping $\mathfrak{C}_0 : \mathfrak{L} \times \mathfrak{L} \to L^1([-\pi,\pi], \mathscr{S}_1(H))$ by

$$\mathfrak{C}_0(A,B)(\omega) = A(\omega)\mathscr{F}_\omega B^\dagger(\omega), \quad A,B \in \mathfrak{L}; \quad \omega \in [-\pi,\pi]. \tag{2.6.3}$$

The following Lemma gives some properties of $\mathfrak{C}_0$.

Lemma 2.6.1.
The mapping $\mathfrak{C}_0$ defined by (2.6.3) *is well defined. It satisfies, for all $A_1, A_2, B \in \mathfrak{L}$, $a \in \mathscr{S}_\infty(H)$, the following properties:*

1. $\mathfrak{C}_0(A_1 + aA_2, B) = \mathfrak{C}_0(A_1,B) + a\mathfrak{C}_0(A_2,B)$,
2. $\mathfrak{C}_0(A_1,B)^\dagger = \mathfrak{C}_0(B,A_1)$,
3. $\mathrm{Tr}\,(\mathfrak{C}_0(A_1,A_1)) \in L^1([-\pi,\pi],\mathbb{R})$, *and is almost everywhere non-negative.*

Proof. The fact that $\mathfrak{C}_0(A,B) \in L^1([-\pi,\pi], \mathscr{S}_1(H))$ if $A,B \in \mathfrak{L}$ follows from Proposition B.0.14, and since the other statements are shown easily, their proof is omitted. □

This Lemma essentially implies that $\mathfrak{C}_0(A,B)$, behaves like a cross-covariance operator of $A,B \in \mathfrak{L}$: for each $\omega \in [-\pi,\pi]$, $\mathfrak{C}_0(A,B)(\omega)$ is like a cross-covariance operator of $A(\omega)$ and $B(\omega)$. By taking the integral of the trace

of $\mathfrak{C}_0$, we can define a semi-scalar product on $\mathfrak{L}$:

$$\langle A, B\rangle_{\mathfrak{H}} := \int_{-\pi}^{\pi} \mathrm{Tr}\left(\mathfrak{C}_0(A,B)\right)(\omega)\,d\omega = \int_{-\pi}^{\pi} \mathrm{Tr}\Big(A(\omega)\mathscr{F}_\omega B^\dagger(\omega)\Big)d\omega, \quad A, B \in \mathfrak{L}. \tag{2.6.4}$$

We denote its corresponding norm by $\|A\|_{\mathfrak{H}}^2 = \langle A, A\rangle_{\mathfrak{H}}$.

Lemma 2.6.2.
The norm $\|\cdot\|_{\mathfrak{L}}$ is stronger that $\|\cdot\|_{\mathfrak{H}}$, i.e. any sequence $(A_n)_{n\geq 1} \subset \mathfrak{L}$ converging in $\|\cdot\|_{\mathfrak{L}}$ converges also in $\|\cdot\|_{\mathfrak{H}}$; but the converse does not always hold.

Proof. The first statement follows from Hölder's inequality:

$$\begin{aligned}
\|A\|_{\mathfrak{H}}^2 &\leq \int_{-\pi}^{\pi} |||A(\omega)|||_\infty^2 |||\mathscr{F}_\omega|||_1\, d\omega \\
&\leq \left(\int_{-\pi}^{\pi} |||A(\omega)|||_\infty^{2q} d\omega\right)^{1/q} \left(\int_{-\pi}^{\pi} |||\mathscr{F}_\omega|||_1^p d\omega\right)^{1/p} \\
&\leq c\|A\|_{\mathfrak{L}}^2.
\end{aligned}$$

The second statement follows from the fact that $\mathscr{F}_\omega$ is trace-class. □

Definition 2.6.3. *We denote by $\mathfrak{H}$ the completion of $\mathfrak{L}$ with respect to $\|\cdot\|_{\mathfrak{H}}$— constructed by taking the equivalence classes of Cauchy sequences whose distance converge to zero.*

Lemma 2.6.4. *$(\mathfrak{H}, \langle\cdot,\cdot\rangle_{\mathfrak{H}})$ is a Hilbert space, $\mathfrak{L} \subset \mathfrak{H}$, and any dense subset of $(\mathfrak{L}, \|\cdot\|_{\mathfrak{L}})$ is also dense in $(\mathfrak{H}, \|\cdot\|_{\mathfrak{H}})$.*

Proof. The fact that $\mathfrak{H}$ is a Hilbert space and $\mathfrak{L} \subset \mathfrak{H}$ follows from the construction of $\mathfrak{H}$ and Lemma 2.6.2. let $\{B_i : i \in I\} \subset \mathfrak{L}$ be a dense subset (with respect to $\|\cdot\|_{\mathfrak{L}}$). If $A \in \mathfrak{H}$, then for any $\varepsilon > 0$, there is an $A_\varepsilon \in \mathfrak{L}$ such that $\|A - A_\varepsilon\|_{\mathfrak{H}} \leq \varepsilon/2$. Since $A_\varepsilon \in \mathfrak{L}$, there is a $j \in I$ such that $\|A_\varepsilon - B_j\|_{\mathfrak{L}} \leq \varepsilon/(2\sqrt{c})$. The triangle inequality now yields

$$\|A - B_j\|_{\mathfrak{H}} \leq \|A - A_\varepsilon\|_{\mathfrak{H}} + c^{1/2}\|A_\varepsilon - B_j\|_{\mathfrak{H}} \leq \varepsilon.$$

□

Notice that $\int \mathrm{Tr}(\mathfrak{C}_0(A,B)(\omega))d\omega \leq \|A\|_{\mathfrak{H}}\|B\|_{\mathfrak{H}}$ by the Cauchy-Schwarz inequality. In fact, we also have that the mapping $\mathfrak{C}_0$ is continuous with respect to the norm $\|\cdot\|_{\mathfrak{H}}$:

Proposition 2.6.5.
For all $A, B \in \mathfrak{L}$,

$$\int_{-\pi}^{\pi} |||\mathfrak{C}_0(A,B)(\omega)|||_1\, d\omega \leq \|A\|_{\mathfrak{H}}\|B\|_{\mathfrak{H}} \tag{2.6.5}$$

Proof. Using Hölder's inequality, we get

$$\begin{aligned}
\int_{-\pi}^{\pi} |||\mathfrak{C}_0(A,B)(\omega)|||_1 d\omega &\le \int_{-\pi}^{\pi} \left|\left|\left| A(\omega)\mathscr{F}_\omega B^\dagger(\omega)\right|\right|\right|_1 d\omega \\
&= \int_{-\pi}^{\pi} \left|\left|\left| A(\omega)\mathscr{F}_\omega^{1/2}(B(\omega)\mathscr{F}_\omega^{1/2})^\dagger\right|\right|\right|_1 d\omega \\
&\le \int_{-\pi}^{\pi} |||A(\omega)\mathscr{F}_\omega^{1/2}|||_2 \left|\left|\left|(B(\omega)\mathscr{F}_\omega^{1/2})^\dagger\right|\right|\right|_2 d\omega \\
&\le \left(\int_{-\pi}^{\pi} |||A(\omega)\mathscr{F}_\omega^{1/2}|||_2^2 d\omega \int_{-\pi}^{\pi} |||B(\omega)\mathscr{F}_\omega^{1/2}|||_2^2 d\omega\right)^{1/2} \\
&= \left(\int_{-\pi}^{\pi} \left|\left|\left| A(\omega)\mathscr{F}_\omega A^\dagger(\omega)\right|\right|\right|_1 d\omega \int_{-\pi}^{\pi} \left|\left|\left| B(\omega)\mathscr{F}_\omega B^\dagger(\omega)\right|\right|\right|_1 d\omega\right)^{1/2} \\
&= \|A\|_{\mathfrak{H}}\|B\|_{\mathfrak{H}}.
\end{aligned}$$

□

We can now extend the domain of the mapping $\mathfrak{C}_0$ to a mapping $\mathfrak{C}:\mathfrak{H}\times\mathfrak{H}\to L^1([-\pi,\pi],\mathscr{S}_1(H))$ by continuity: if $A_n\to A\in\mathfrak{H}$ and $B_n\to B\in\mathfrak{H}$, where $(A_n),(B_n)\subset\mathfrak{L}$, we define

$$\mathfrak{C}(A,B)=\lim_{n\to\infty}\mathfrak{C}_0(A_n,B_n),\quad \text{in } L^1([-\pi,\pi],\mathscr{S}_1(H)). \tag{2.6.6}$$

Proposition 2.6.6.
The mapping $\mathfrak{C}:\mathfrak{H}\times\mathfrak{H}\to L^1([-\pi,\pi],\mathscr{S}_1(H))$ is well defined. Furthermore, it is continuous, i.e. if $(A_n),(B_n)\subset\mathfrak{H}$ such that $\|A_n-A\|_{\mathfrak{H}}\to 0$ and $\|B_n-B\|_{\mathfrak{H}}\to 0$ as $n\to\infty$, for some $A,B\in\mathfrak{H}$, then

$$|||\mathfrak{C}(A_n,B_n)-\mathfrak{C}(A,B)|||_1\to 0,\quad n\to\infty,$$

Furthermore, the properties of $\mathfrak{C}_0$ extend to $\mathfrak{C}$: for all $A_1,A_2,B\in\mathfrak{H}$, $a\in\mathscr{S}_\infty(H)$,

1. $\mathfrak{C}(A_1+aA_2,B)=\mathfrak{C}(A_1,B)+a\mathfrak{C}(A_2,B)$,
2. $\mathfrak{C}(A_1,B)^\dagger=\mathfrak{C}(B,A_1)$,
3. $\mathfrak{C}(A_1,A_1)$ *is a non-negative operator.*

Proof. The extension is well defined by the linearity and continuity $\mathfrak{C}_0$. Indeed,

$$\begin{aligned}
|||\mathfrak{C}_0(A_m,B_m)-\mathfrak{C}_0(A_n,B_n)|||_1 &= |||\mathfrak{C}_0(A_m-A_n,B_m)+\mathfrak{C}_0(A_n,B_m-B_n)|||_1 \\
&\le |||\mathfrak{C}_0(A_m-A_n,B_m)|||_1+|||\mathfrak{C}_0(A_n,B_m-B_n)|||_1.
\end{aligned}$$

Therefore, by Proposition 2.6.5,

$$\int_{-\pi}^{\pi}|||\mathfrak{C}_0(A_m,B_m)(\omega)-\mathfrak{C}_0(A_n,B_n)(\omega)|||_1 d\omega\le\|A_m-A_n\|_{\mathfrak{H}}\|B_m\|_{\mathfrak{H}}+\|B_m-B_n\|_{\mathfrak{H}}\|A_n\|_{\mathfrak{H}}.$$

Since $(A_n), (B_n)$ are converging in $\mathfrak{H}$, they are also Cauchy sequences, which implies that $(\mathfrak{C}_0(A_n, B_n))_{n\geq 1}$ is also a Cauchy sequence in $L^1([-\pi,\pi], \mathscr{S}_1(H))$. The uniqueness of the limit, and the continuity of $\mathfrak{C}_0$ follow from the same kind of argument. The proof of the other properties follow directly from Lemma 2.6.1 and the definition of $\mathfrak{C}$. □

2.6.2 *Isometry Between* $\mathbb{M}$ *and* $\mathfrak{H}$

Recall that $\mathfrak{e}_n \in \mathfrak{L}$ is the function defined by $\mathfrak{e}_n(\omega) = e^{\mathbf{i}\omega n}$, for $n \in \mathbb{Z}$. We now define a mapping $\mathscr{T} : \mathbb{M}_0 \to \mathfrak{H}$ by linear extension of the mappings $X_n \mapsto \mathfrak{e}_n$, or explicitly

$$\mathscr{T}\left(\sum_{|n|<N} a_n X_{-n}\right) = \sum_{|n|<N} a_n \mathfrak{e}_{-n}, \tag{2.6.7}$$

where $(a_n\mathfrak{e}_{-n})(\omega) = \mathfrak{e}_{-n}(\omega)a_n$. The properties of the mapping $\mathscr{T}$ are given in the following proposition.

Proposition 2.6.7.
The mapping $\mathscr{T} : \mathbb{M}_0 \to \mathfrak{H}$ *be defined by* (2.6.7) *has the following properties:*

1. $\mathscr{T}$ *is* $\mathscr{S}_\infty(H)$*-linear:*

$$\mathscr{T}(aY_1 + Y_2) = a\mathscr{T}(Y_1) + \mathscr{T}(Y_2), \quad Y_1, Y_2 \in \mathbb{M}_0, a \in \mathscr{S}_\infty(H). \tag{2.6.8}$$

2. *The mapping* $\mathscr{T}$ *preserves the second-order structure, i.e.*

$$\mathbb{E}[Y_1 \otimes_2 Y_2] = \int_{-\pi}^{\pi} \mathfrak{C}(\mathscr{T}(Y_1), \mathscr{T}(Y_2))(\omega)\, d\omega, \quad \forall\, Y_1, Y_2 \in \mathbb{M}_0. \tag{2.6.9}$$

In particular, the mapping $\mathscr{T} : \mathbb{M}_0 \to \mathbb{H}$ *is an isomorphism:*

$$\langle \mathscr{T}(Y_1), \mathscr{T}(Y_1)\rangle_{\mathfrak{H}} = \langle Y_1, Y_2\rangle_{\mathbb{H}}, \quad \forall\, Y_1, Y_2 \in \mathbb{M}_0. \tag{2.6.10}$$

Proof. From its construction, $\mathscr{T}$ is $\mathscr{S}_\infty(H)$-linear. Now take $Y_1, Y_2 \in \mathbb{M}_0$. Without loss of generality,

$$Y_j = \sum_{|n|<N} a_{j,n} X_{-n}, \quad j = 1, 2.$$

Using the fact that bounded operators commute with the expectation and with Bochner integrals, and using the inversion formula,

$$\begin{aligned}
\mathbb{E}[Y_1 \otimes_2 Y_2] &= \mathbb{E}\left[\left(\sum_{|n|<N} a_{1,n} X_{-n}\right) \otimes_2 \left(\sum_{|m|<N} a_{1,m} X_{-m}\right)\right] \\
&= \sum_{|n|,|m|<N} a_n \mathscr{R}_{m-n} b_m^\dagger \\
&= \sum_{|n|,|m|<N} a_n \left[\int_{-\pi}^{\pi} \mathscr{F}_\omega e^{\mathbf{i}\omega(m-n)}\, d\omega\right] b_m^\dagger
\end{aligned}$$

$$= \int_{-\pi}^{\pi} \left(\sum_{|n|<N} a_{1,n} e^{-\mathbf{i}\omega n} \right) \mathscr{F}_\omega \left(\sum_{|m|<N} a_{2,m} e^{-\mathbf{i}\omega m} \right)^\dagger d\omega$$
$$= \int_{-\pi}^{\pi} \mathfrak{C}(\mathscr{T}(Y_1), \mathscr{T}(Y_2))(\omega) d\omega.$$

Taking the trace yields $\langle Y_1, Y_2 \rangle_{\mathbb{H}} = \langle \mathscr{T}(Y_1), \mathscr{T}(Y_2) \rangle_{\mathfrak{H}}$ □

Since the mapping $\mathscr{T}$ is an isometry between $\mathbb{M}_0$ and $\mathfrak{H}$, it can be extended by continuity to a mapping $\mathscr{T} : \mathbb{M} \to \mathfrak{H}$. More precisely, if $Y \in \mathbb{M}$, and $(Y_n) \subset \mathbb{M}_0$ is a sequence converging to Y, we define

$$\mathscr{T}(Y) = \lim_{n\to\infty} Y_n, \quad \text{in } \mathfrak{H}. \tag{2.6.11}$$

Proposition 2.6.8.

The mapping $\mathscr{T} : \mathbb{M} \to \mathfrak{H}$ defined by (2.6.11) *is well-defined, linear, and surjective. Therefore it is an isometric isomorphism between $\mathbb{M}$ and $\mathfrak{H}$, has the following properties*

$$\mathscr{T}(aY_1 + Y_2) = a\mathscr{T}(Y_1) + \mathscr{T}(Y_2), \qquad Y_1, Y_2 \in \mathbb{M}, a \in \mathscr{S}_\infty(H), \tag{2.6.12}$$

$$\mathbb{E}[Y_1 \otimes_2 Y_2] = \int_{-\pi}^{\pi} \mathfrak{C}(\mathscr{T}(Y_1), \mathscr{T}(Y_2))(\omega) d\omega, \qquad Y_1, Y_2 \in \mathbb{M}. \tag{2.6.13}$$

$$\langle Y_1, Y_2 \rangle_{\mathbb{H}} = \langle \mathscr{T}(Y_1), \mathscr{T}(Y_1) \rangle_{\mathfrak{H}}, \qquad Y_1, Y_2 \in \mathbb{M}. \tag{2.6.14}$$

Moreover, $\mathscr{T}$ admits an inverse $\mathscr{T}^{-1} : \mathfrak{H} \to \mathbb{M}$ that is also a isometry.

Proof. Let us verify that $\mathscr{T}$ is well defined. Let $(Y_{1,n}), (Y_{2,n}) \subset \mathbb{M}$ be two sequences converging to $Y \in \mathbb{M}$, and define $y_j = \lim_{n\to\infty} \mathscr{T}(Y_{j,n})$, $j = 1, 2$. Then $y_1 = y_2$ by the isometry property of $\mathscr{T}$:

$$\|y_1 - y_2\|_{\mathfrak{H}} = \lim_n \|\mathscr{T}(Y_{1,n}) - \mathscr{T}(Y_{2,n})\|_{\mathfrak{H}} = \lim_n \|Y_{1,n} - Y_{2,n}\|_{\mathbb{H}} = 0.$$

The proof of the linearity and of property (2.6.12) is shown directly by taking convergent sequences in $\mathbb{M}_0$, and is omitted.

Since the subspace $\mathscr{T}(\mathbb{M}_0)$ is dense in $\mathfrak{L}$ (by Lemma B.0.13, since $2q \neq \infty$), Lemma 2.6.4 implies that it is also dense in $\mathfrak{H}$. Therefore $\mathscr{T}(\mathbb{M}) = \mathfrak{H}$ and $\mathscr{T}$ is an isometric isomorphism, and admits an inverse $\mathscr{T}^{-1} : \mathfrak{H} \to \mathbb{M}$ that is also linear and isometric. Furthermore, Lemma C.1.2 and Proposition 2.6.6 imply properties (2.6.13) and (2.6.14). □

2.6.3 The Process Z_ω and the Stochastic Integral

We now define $Z_\omega = \mathscr{T}^{-1}(\mathbf{1}_{[-\pi,\omega)} I)$, for all $\omega \in (-\pi, \pi]$, and $Z_{-\pi} = 0$. Notice that $Z_\pi = X_0$, and that by the isometry property (2.6.13),

$$\mathbb{E}\left[(aZ_\alpha) \otimes_2 (bZ_\beta)\right] = \int_{-\pi}^{\min(\alpha,\beta)} a\mathscr{F}_\omega b^\dagger d\omega. \tag{2.6.15}$$

In particular, setting $a = b = I$ and taking the trace of this expression, we see that

$$\left\langle Z_\alpha, Z_\beta \right\rangle_{\mathbb{H}} = \int_{-\pi}^{\min(\alpha,\beta)} \mathrm{Tr}\,(\mathscr{F}_\omega) d\omega,$$

i.e. $\omega \mapsto Z_\omega$ is an orthogonal increment process. We now define the integral with respect to this orthogonal process. Let $D \subset \mathfrak{H}$ be the subspace of càdlàg step functions, i.e. functions of the form

$$A = \sum_{n=1}^{N} a_n \mathbf{1}_{[\omega_n, \omega_{n+1})}, \tag{2.6.16}$$

where $-\pi = \omega_1 < \omega_2 < \cdots < \omega_{N+1} = \pi$ and $a_n \in \mathscr{S}_\infty(H)$ for $n = 1, \ldots, N$, and define the mapping $\mathscr{I} : D \to \mathbb{M}$ by

$$\mathscr{I}\left(\sum_{n=1}^{N} a_n \mathbf{1}_{[\omega_n,\omega_{n+1})}\right) = \sum_{n=1}^{N} a_n (Z_{\omega_{n+1}} - Z_{\omega_n}). \tag{2.6.17}$$

Notice that $\mathscr{I}$ is a linear mapping, and that

$$\mathscr{T}(\mathscr{I}(a\mathbf{1}_{[\alpha,\beta)})) = \mathscr{T}(a(Z_\beta - Z_\alpha)) = a\mathscr{T}(Z_\beta - Z_\alpha) = a\mathbf{1}_{[\alpha,\beta)}, \quad a \in \mathscr{S}_\infty(H). \tag{2.6.18}$$

Therefore $\mathscr{I} = \mathscr{T}^{-1}$ on D, and is an isometry on D. Since D is dense in $\mathfrak{H}$ (by Lemma B.0.13 and Lemma 2.6.4), we can extend $\mathscr{I}$ to $\mathfrak{H}$ by continuity. We denote the extension by $\mathscr{I} : \mathfrak{H} \to \mathbb{H}$, and in fact $\mathscr{I} = \mathscr{T}^{-1}$. This gives a meaning to the stochastic integral

$$\int_{-\pi}^{\pi} A(\omega) dZ_\omega = \mathscr{I}(A), \tag{2.6.19}$$

for all $A \in \mathfrak{H}$.

2.6.4 Proof of the Stated Results

Proof of Theorem 2.4.3 on page 30. We have

$$X_t = \mathscr{I}(\mathscr{T}(X_t)) = \mathscr{I}(\mathfrak{e}_t) = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} dZ_\omega, \tag{2.6.20}$$

which proves (2.4.2). The proof of (2.4.3) is a consequence of Theorem 2.5.3, which is proved below, independently from Theorem (2.4.3). To

prove (2.4.4), recall that

$$g_\omega(n) = (2\pi)^{-1}\int_{-\pi}^{\omega} e^{\mathbf{i}\alpha n}d\alpha = (2\pi)^{-1}\int_{-\pi}^{\pi} \mathbf{1}_{[-\pi,\omega)}e^{\mathbf{i}\alpha n}d\alpha,$$

and by (2.4.3),

$$Z_\beta - Z_\alpha = \lim_{T\to\infty}\sum_{|t|<T}(1-|t|/T)\left(g_\beta(n)-g_\alpha(n)\right)X_{-n}.$$

Now

$$g_\beta(n)-g_\alpha(n) = \int_\alpha^\beta e^{\mathbf{i}\omega n}d\omega = \overline{\int_{-\beta}^{-\alpha} e^{\mathbf{i}\lambda n}d\lambda} = \overline{g_{-\alpha}(n)-g_{-\beta}(n)},$$

where the change of variable $\lambda = -\omega$ was used. Therefore, since X_t is a real-valued series,

$$\begin{aligned} Z_\beta - Z_\alpha &= \lim_{T\to\infty}\sum_{|t|<T}(1-|t|/T)\,\overline{\left(g_\beta(n)-g_\alpha(n)\right)}X_{-n} \\ &= \overline{\lim_{T\to\infty}\sum_{|t|<T}(1-|t|/T)\left(g_\beta(n)-g_\alpha(n)\right)X_{-n}} \\ &= \overline{Z_{-\alpha}-Z_{-\beta}}. \end{aligned}$$

Equation (2.4.5) follows directly from the definition of Z_ω and the isometry property (2.6.13), and (2.4.6) follows from (2.4.5). □

Proof of Remark 2.4.4. We only prove (2.4.7), because (2.4.8) is proved is a similar fashion to Remark 2.5.4, which is proved independently of this result. Recall that

$$g_\omega(n) = (2\pi)^{-1}\int_{-\pi}^{\omega} e^{\mathbf{i}\alpha n}d\alpha = (2\pi)^{-1}\int_{-\pi}^{\pi} \mathbf{1}_{[-\pi,\omega)}e^{\mathbf{i}\alpha n}d\alpha,$$

and let us introduce some notation: we define

$$Z_{\omega,N} = \sum_{|n|<N} g_\omega(n)X_{-n},$$

and

$$f_{\omega,N} = \left(\mathbf{1}_{[-\pi,\omega)} - \sum_{|n|<N} g_\omega(n)\mathfrak{e}_{-n}\right),$$

where $\mathfrak{e}_n(\alpha) = e^{\mathbf{i}\alpha n}$. We get, using the isometry property (2.5.6), and Lemma 2.6.2,

$$\begin{aligned} \left\|Z_\omega - Z_{\omega,N}\right\|_{\mathbb{H}}^2 &= \left\|\mathscr{T}(Z_\omega)-\mathscr{T}(Z_{\omega,N})\right\|_{\mathfrak{H}}^2 \\ &= \left\|f_{\omega,N}I\right\|_{\mathfrak{H}}^2 \\ &\le c\left\|f_{\omega,N}I\right\|_{\mathfrak{L}}^2 \end{aligned}$$

$$= c\int_{-\pi}^{\pi} \left|\left| f_{\omega,N}(\alpha) I \right|\right|_{\infty}^2 d\alpha$$
$$= c\int_{-\pi}^{\pi} \left| f_{\omega,N}(\alpha) \right|^2 d\alpha \to 0, \quad N \to \infty,$$

where the convergence to zero comes from the fact that $f_{\omega,N}$ is the difference between $\mathbf{1}_{[-\pi,\omega)}$ and its truncated Fourier series. □

Proof of Theorem 2.5.1. Since the trigonometric polynomials

$$\sum_{|n|<N} a_n \mathfrak{e}_{-n}, \quad a_n \in \mathscr{S}_\infty(H), N \in \mathbb{N}, \tag{2.6.21}$$

are dense in $\mathfrak{L}$, and therefore also in $\mathfrak{H}$ (see Lemmas B.0.13 and 2.6.4), any $A \in \mathfrak{H}$ can be written as a limit $A = \lim_{N\to\infty} A_N$, where A_N can be taken to be (without loss of generality) to be of the form given in (2.6.21). This proves (2.5.2). Now (2.5.3), as well as the uniqueness of the limit, follow directly from the isometry property, and (2.5.5) corresponds to (2.6.13), which has already been proved.

We now turn to the proof of (2.5.4). Let $Y_1 = \int_{-\pi}^{\pi} A(\omega) dZ_\omega$, and $Y_2 = \int_{-\pi}^{\pi} \overline{A(-\omega)} dZ_\omega$. We will show that $\left\| \overline{Y_1} - Y_2 \right\|_{\mathbb{H}} = 0$. Fix $\varepsilon > 0$. Using (2.5.2), we take a sequence $(a_t) \subset \mathscr{S}_\infty(H)$ such that

$$\left\| Y_1 - Y' \right\|_{\mathbb{H}} < \varepsilon/2,$$

where $Y' = \sum_{|t|<T} a_t X_{-t}$. By the triangle inequality,

$$\left\| \overline{Y_1} - Y_2 \right\|_{\mathbb{H}} \le \left\| \overline{Y_1} - \overline{Y'} \right\|_{\mathbb{H}} + \left\| \overline{Y'} - Y_2 \right\|_{\mathbb{H}}.$$

We know that the first term is bounded by $\varepsilon/2$, so let us turn to the second term. By the isometry property,

$$\left\| \overline{Y'} - Y_2 \right\|_{\mathbb{H}} = \left\| \mathscr{T}\left(\overline{Y'}\right) - \mathscr{T}(Y_2) \right\|_{\mathfrak{H}}.$$

Now notice that

$$\mathscr{T}\left(\overline{Y'}\right)(\omega) = \sum_{|t|<T} \overline{a_t} \mathfrak{e}_{-t}(\omega) = \overline{\sum_{|t|<T} a_t \mathfrak{e}_{-t}(-\omega)} = \overline{\mathscr{T}(Y')(-\omega)}.$$

Thus (2.5.4) is proved provided we show that $\left\| A^{\mathfrak{c}} \right\|_{\mathfrak{H}} = \|A\|_{\mathfrak{H}}$ for all $A \in \mathfrak{H}$, where $A^{\mathfrak{c}}(\omega) = \overline{A(-\omega)}$. We have

$$\begin{aligned}
\left\| A^{\mathfrak{c}} \right\|_{\mathfrak{H}}^2 &= \int_{-\pi}^{\pi} \mathrm{Tr}\left(A^{\mathfrak{c}}(\omega) \mathscr{F}_\omega (A^{\mathfrak{c}}(\omega))^\dagger \right) d\omega \\
&= \int_{-\pi}^{\pi} \mathrm{Tr}\left(\overline{A(-\omega)} \mathscr{F}_\omega \overline{A(-\omega)}^\dagger \right) d\omega \\
&= \int_{-\pi}^{\pi} \mathrm{Tr}\left(\overline{A(-\omega) \mathscr{F}_{-\omega} A(-\omega)^\dagger} \right) d\omega \qquad \text{(by Proposition 2.8.8)}
\end{aligned}$$

$$
\begin{aligned}
&= \int_{-\pi}^{\pi} \operatorname{Tr}\Big(\overline{A(\omega)\mathscr{F}_\omega A(\omega)^\dagger}\Big) d\omega && \text{(by a change of variables)} \\
&= \int_{-\pi}^{\pi} \operatorname{Tr}\Big(A(\omega)\mathscr{F}_\omega A(\omega)^\dagger\Big) d\omega && \text{(by direct calculation)} \\
&= \|A\|_{\mathfrak{H}}^2.
\end{aligned}
$$

This completes the proof. □

Proof of Theorem 2.5.3. The isometry property and Lemma 2.6.2 yield

$$
\begin{aligned}
\mathbb{E}\left\| \int_{-\pi}^{\pi} A(\omega)\, dZ_\omega - \sum_{|t|<T} \left(1 - \frac{|t|}{T}\right) a_t X_{-t} \right\|^2 &= \left\| A - \sum_{|t|<T} \left(1 - \frac{|t|}{T}\right) a_t \mathfrak{e}_{-t} \right\|_{\mathfrak{H}}^2 \\
&\le c \left\| A - \sum_{|t|<T} \left(1 - \frac{|t|}{T}\right) a_t \mathfrak{e}_{-t} \right\|_{\mathfrak{L}}^2
\end{aligned}
$$

Noticing that $\sum_{|t|<T} (1 - |t|/T)\, a_t \mathfrak{e}_{-t} = K_T * A$, where K_T is the Fejér kernel (B.0.14), and recalling that $\mathfrak{L} = L^{2q}([-\pi,\pi], \mathscr{S}_\infty(H))$, where $2q \in [1,\infty)$, Proposition B.0.16 yields (2.5.7). To show (2.5.9), we need to show that

$$
\lim_{T\to\infty} \left\| A_T - A'_T \right\|_{\mathbb{H}} = 0,
$$

where $A_T = \sum_{|t|<T} \left(1 - \frac{|t|}{T}\right) a_t X_{-t}$ and $A'_T = \sum_{|t|<T} a_t X_{-t}$, and then the result will follow from (2.5.7) and the triangle inequality. Direct calculations yield

$$
\begin{aligned}
\mathbb{E}\left\| A_T - A'_T \right\|^2 &= \sum_{|s|,|t|<T} \frac{|s||t|}{T^2} \operatorname{Tr}\Big(a_t \mathscr{R}_{s-t} a_s^\dagger\Big) \\
&\le \sup_u |||\mathscr{R}_u|||_1 \sum_{|s|,|t|<T} \frac{|s||t|}{T^2} |||a_t|||_\infty |||a_s|||_\infty \\
&= \sup_u |||\mathscr{R}_u|||_1 \left(\sum_{|t|<T} \frac{|t|}{T} |||a_t|||_\infty \right)^2.
\end{aligned}
$$

Since the nuclear norm of the autocovariance operators are uniformly bounded (Remark 2.4.2), and $\sum_{t\in\mathbb{Z}} |||a_t|||_\infty < \infty$ by assumption, the dominated convergence theorem tells us that the right-hand side converges to zero as $T \to \infty$. □

Proof of Remark 2.5.4. Let $A : \omega \to \mathscr{S}_\infty(H)$ be càdlàg with a finite number of jumps, say at $\Xi = \{\xi_1, \dots, \xi_m\}$. This implies that A is uniformly continuous within each interval not containing a jump, or possibly having a ξ_j at one of its extremities. More precisely, for all $\varepsilon > 0$, there is a $\delta > 0$ such that

$$
0 < \omega - \omega' < \delta \quad \& \quad (\omega', \omega) \cap \Xi = \emptyset \qquad \Longrightarrow \qquad \Big|\Big| A(\omega') - A(\omega) \Big|\Big|_\infty < \varepsilon. \tag{2.6.22}
$$

Now for a $\varepsilon > 0$ fixed, we choose $\delta > 0$ such that (2.6.22) holds for

$$(\varepsilon')^2 = \varepsilon^2 / \left[2\pi \sup_{\omega\in[-\pi,\pi]} |||\mathscr{F}_\omega|||_1 \right],$$

(recall that this supremum is finite since Condition 2.4.1(∞) holds). We choose $P = \{\omega_1, \ldots, \omega_K\}$ to be any partition finer than Ξ, satisfying

$$\max_{i=1,\ldots,K-1} \omega_{i+1} - \omega_i < \delta.$$

Now for any partition $P' = \left\{\omega'_1, \ldots, \omega'_{J+1}\right\}$ finer than P, and any $\lambda_j \in [\omega'_j, \omega'_{j+1}]$, if we let $S_{P'} = \sum_{j=1}^J A(\lambda_j)\mathbf{1}_{[\omega'_j,\omega'_{j+1})}$, we will have

$$\begin{aligned}
&\left\| \int_{-\pi}^{\pi} A(\omega)\, dZ_\omega - \sum_{j=1}^{J} A(\lambda_j)(Z_{\omega_{j+1}} - Z_{\omega_j}) \right\|_{\mathbb{H}}^2 \\
&= \|\mathscr{I}(A) - \mathscr{I}(S_{P'})\|_{\mathbb{H}} \\
&= \|A - S_{P'}\|_{\mathfrak{H}} \\
&= \int_{-\pi}^{\pi} \left| \left| \left| (A(\omega) - S_{P'}(\omega))\mathscr{F}_\omega (A(\omega) - S_{P'}(\omega))^\dagger \right| \right| \right|_1 d\omega \\
&\leq c \int_{-\pi}^{\pi} |||A(\omega) - S_{P'}(\omega)|||_\infty^2 d\omega
\end{aligned}$$

where $c = \sup_{\omega\in[-\pi,\pi]} |||\mathscr{F}_\omega|||_1$,

$$= c \sum_{j=1}^{J} \int_{\omega'_j}^{\omega'_{j+1}} \left| \left| \left| A(\omega) - A(\lambda_j) \right| \right| \right|_\infty^2 d\omega$$

using (2.6.22) and the fact that $P' \supset P \supset \Xi$,

$$\begin{aligned}
&\leq c \sum_{j=1}^{J} (\varepsilon')^2 (\omega'_{j+1} - \omega'_j) \\
&= 2\pi c \cdot (\varepsilon')^2 \\
&= \varepsilon^2,
\end{aligned}$$

which completes the proof. □

Proof of Theorem 2.5.5. Let us define $A_T = \sum_{|s|<T} \mathfrak{e}_{-s} a_{T,s} \in \mathfrak{H}$, and notice that by assumption,

$$\|A - A_T\|_{\mathfrak{H}} \to 0, \quad T \to \infty,$$

for some $A \in \mathfrak{H}$. In particular, this implies that $(A_T)_{T\geq 1}$ is a Cauchy se-

quence in $\mathfrak{H}$. Let $Y_{T,t} = \sum_{|s|<T} a_{T,s} X_{t-s}$, and notice that

$$\mathscr{T}(Y_{T,t}) = \sum_{|s|<T} a_{T,s} \mathfrak{e}_{t-s} = \mathfrak{e}_t A_T, \quad t \in \mathbb{Z}.$$

By the isometry property,

$$\begin{aligned}
\left\| Y_{T',t} - Y_{T,t} \right\|_{\mathbb{H}}^2 &= \left\| \mathscr{T}(Y_{T',t}) - \mathscr{T}(Y_{T,t}) \right\|_{\mathfrak{H}}^2 \\
&= \| \mathfrak{e}_t (A_{T'} - A_T) \|_{\mathfrak{H}}^2 \\
&= \int_{-\pi}^{\pi} \mathrm{Tr}\Big([e^{\mathbf{i}\omega t}(A_{T'}(\omega) - A_T(\omega))] \mathscr{F}_\omega [e^{\mathbf{i}\omega t}(A_{T'}(\omega) - A_T(\omega))]^\dagger \Big) d\omega \\
&= \int_{-\pi}^{\pi} \mathrm{Tr}\Big((A_{T'}(\omega) - A_T(\omega)) \mathscr{F}_\omega (A_{T'}(\omega) - A_T(\omega))^\dagger \Big) d\omega \\
&= \| A_{T'} - A_T \|_{\mathfrak{H}}^2.
\end{aligned}$$

Therefore $(Y_{T,t})_{T \geq 1}$ is a Cauchy sequence in $\mathbb{M}_0$, and converges to some $Y_t \in \mathbb{M}$. Furthermore,

$$\begin{aligned}
Y_t = \mathscr{I}(\mathscr{T}(Y_t)) &= \mathscr{I}(\mathscr{T}(\lim_T Y_{T,t}) \\
&= \mathscr{I}(\lim_T \mathscr{T}(Y_{T,t})) \\
&= \mathscr{I}(\lim_T \mathfrak{e}_t A_T) \\
&= \mathscr{I}(\mathfrak{e}_t A) \\
&= \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} A(\omega) \, dZ_\omega,
\end{aligned}$$

for all $t \in \mathbb{Z}$. Hence by Theorem 2.5.1, Y_t has mean zero,

$$\begin{aligned}
\mathbb{E}[Y_t \otimes_2 Y_s] &= \mathbb{E}\left[\int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} A(\omega) d\omega \otimes_2 \int_{-\pi}^{\pi} e^{\mathbf{i}\omega s} A(\omega) d\omega \right] \\
&= \int_{-\pi}^{\pi} e^{\mathbf{i}\omega(t-s)} A(\omega) \mathscr{F}_\omega A^\dagger(\omega) d\omega,
\end{aligned}$$

hence Y_t is second-order stationary, and its autocovariance operators satisfy

$$\mathscr{R}_t^Y = \mathbb{E}[Y_t \otimes_2 Y_0] = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} A(\omega) \mathscr{F}_\omega A^\dagger(\omega) d\omega, \quad t \in \mathbb{Z}. \tag{2.6.23}$$

Therefore the weak spectral density operators of Y_t are given by

$$\mathscr{F}_\omega^Y = A(\omega) \mathscr{F}_\omega^X A(\omega),$$

and $\mathscr{F}_\cdot^Y = \mathfrak{C}(A, A) \in L^1([-\pi, \pi], \mathscr{S}_1(H))$ by Proposition 2.6.6.

Now suppose that $A_T \to A$ in $L^{2q}([-\pi, \pi], \mathscr{S}_\infty(H))$ for some $q \geq p/(p-1)$. Recall that $\mathscr{F}_\cdot \in L^p([-\pi, \pi], \mathscr{S}_1(H))$, and let $r = (p^{-1} + q^{-1})^{-1}$. Notice that $r \geq 1$, and therefore, by Hölder's inequality (similarly to the proof of

Lemma 2.6.2),

$$\begin{aligned}\int_{-\pi}^{\pi} \left|\!\left|\!\left| \mathscr{F}_\omega^Y \right|\!\right|\!\right|_1^r d\omega &\leq \int_{-\pi}^{\pi} |||A(\omega)|||_\infty^{2r} |||\mathscr{F}_\omega|||_1^r d\omega \\ &\leq \left(\int_{-\pi}^{\pi} |||A(\omega)|||_\infty^{2q} d\omega\right)^{r/q} \left(\int_{-\pi}^{\pi} |||\mathscr{F}_\omega|||_1^p d\omega\right)^{r/p} \\ &< \infty,\end{aligned}$$

hence $\mathscr{F}_\omega^Y \in L^r([-\pi,\pi], \mathscr{S}_1(H))$. □

Proof of Remark 2.5.6. 1. Letting $A_T = \sum_{|s|<T} \mathfrak{e}_{-s} a_{T,s}$, condition (2.5.12) is equivalent to $(A_T)_{T\geq 1}$ being a Cauchy sequence in $\mathfrak{H}$, which is equivalent to $(Y_{T,0})_{T\geq 1}$ being a Cauchy sequence in $\mathbb{H}$, where $Y_{T,0} = \sum_{|s|<T} a_{T,s} X_{-s}$. Letting $T' > T$, we see that

$$\begin{aligned}\left\| Y_{T',0} - Y_{T,0} \right\|_{\mathbb{H}}^2 &= \mathrm{Tr}\left(\left(\sum_{|s|<T'} (a_{T',s} - a_{T,s}) X_{-s}\right) \otimes_2 \left(\sum_{|u|<T'} (a_{T',u} - a_{T,u}) X_{-u}\right)\right) \\ &= \sum_{|s|,|u|<T'} \mathrm{Tr}\left((a_{T',s} - a_{T,s}) \mathscr{R}_{u-s}^X (a_{T',u} - a_{T,u})^\dagger\right).\end{aligned}$$

Therefore (2.5.15) is real non-negative, and is equivalent to $(A_T)_{T\geq 1}$ being a Cauchy sequence.

2. (2.5.16) is obvious since $a_{T',s} - a_{T,s}$ is equal to zero if $|s| < T$ and equal to a_t if $T \leq |s| < T'$. Furthermore, since

$$\left|\mathrm{Tr}\left(a_s \mathscr{R}_{u-s} a_u^\dagger\right)\right| < |||a_s|||_\infty |||a_u|||_\infty \left|\!\left|\!\left| \mathscr{R}_{u-s}^X \right|\!\right|\!\right|_1,$$

by Hölder's inequality, and $\left|\!\left|\!\left| \mathscr{R}_t^X \right|\!\right|\!\right|_1 < M < \infty$ for all $t \in \mathbb{Z}$, (2.5.16) is implied by

$$\sum_{T\leq|s|<T'} |||a_s|||_\infty < \varepsilon,$$

which is equivalent to (2.5.17).

3. By Hölder's inequality, and the fact that $\mathscr{R}_t^X = 0$ if $|t| > m$,

$$\begin{aligned}\sum_{T\leq|s|,|u|<T'} \left|\mathrm{Tr}\left(a_s \mathscr{R}_{u-s} a_u^\dagger\right)\right| &\leq \sum_{|s|\geq T, u\in\mathbb{Z}} |||a_s|||_\infty |||a_u|||_\infty \left|\!\left|\!\left| \mathscr{R}_{u-s}^X \right|\!\right|\!\right|_1 \\ &= \sum_{|s|\geq T; |u-s|\leq m} |||a_s|||_\infty |||a_u|||_\infty \left|\!\left|\!\left| \mathscr{R}_{u-s}^X \right|\!\right|\!\right|_1\end{aligned}$$

the change of variables $u - s = k$ yields

$$\leq M \sum_{|k|\leq m} \sum_{|s|\geq T} |||a_s|||_\infty |||a_{s+k}|||_\infty$$

and the Cauchy-Schwarz inequality yields

$$\leq M \sum_{|k|\leq m} \sqrt{\sum_{|s|>T} |||a_s|||_\infty^2} \sqrt{\sum_{|s|>T} |||a_{s+k}|||_\infty^2}.$$

If (2.5.18) holds, taking T large enough in the last expression implies that (2.5.16) holds. □

4. By Proposition 3.12.11, $\mathscr{R}_t^Y = \sum_{s,l\in\mathbb{Z}} A_s \mathscr{R}_{t-s+l} A_l^\dagger$, and thus

$$\sum_{t\in\mathbb{Z}} \left|\left|\left|\mathscr{R}_t^Y\right|\right|\right|_\infty \leq \sum_s |||A_s|||_\infty \sum_l |||A_l|||_\infty \left(\sum_t \left|\left|\left|\mathscr{R}_{t-s+l}^X\right|\right|\right|_\infty\right) < \infty.$$

2.7 A Short Note About Measurability

The doubly spectral decomposition we will present in Section 2.8 will be based on integrals of the form $\int_{-\pi}^{\pi} \left(\varphi_j^\omega \otimes_2 \varphi_j^\omega\right) d\omega$, where φ_j^ω is the j-th eigenfunction of the spectral density operator $\mathscr{F}_\omega$. In this (technical) section, we investigate conditions under which such integrals are well defined.

Let $\mathscr{F}_\cdot$ be the weak spectral density operators of an FTS X_t. This implies that there is a set $E \subset [-\pi,\pi]$ of measure 2π such that

$$|||\mathscr{F}_\omega|||_1 < \infty, \quad \omega \in E.$$

For each $\omega \in E$, we can write the singular value decomposition of $\mathscr{F}_\omega$:

$$\mathscr{F}_\omega = \sum_{i=1}^{\infty} \mu_i(\omega) \varphi_i^\omega \otimes_2 \varphi_i^\omega$$

For any fixed $\omega \in E$, $\{\mu_i(\omega)\}_{i\geq 1}$ is a non-increasing positive sequences tending to zero. We denote by $\{\lambda_i(\omega)\}_{i\geq 1}$ the decreasing sequence of distinct elements of $\{\mu_i(\omega)\}_{i\geq 1}$, define the set

$$I_k(\omega) = \{i \geq 1 : \mu_i(\omega) = \lambda_k(\omega)\},$$

and we denote its cardinality by $m_k(\omega) = |I_k(\omega)|$. We will also write

$$I(\omega) = \left\{i \geq 1 : \mu_i(\omega) > 0\right\} = \bigcup_{k\geq 1 \& \lambda_k(\omega)>0} I_k(\omega) \tag{2.7.1}$$

for the set of indices of the repeated non-zero eigenvalues of $\mathscr{F}_\omega$, and

$$J(\omega) = \left\{j \geq 1 : \lambda_j(\omega) > 0\right\}, \tag{2.7.2}$$

the set of indices of the non-repeated non-zero eigenvalues of $\mathscr{F}_\omega$. Notice that

$$I(\omega) = J(\omega) = \{1,2,3,\ldots\}$$

unless $\mathscr{F}_\omega$ is of finite rank, in which case

$$\{\lambda_i(\omega)\}_{i\geq 1} = \{\lambda_1(\omega), \ldots, \lambda_N(\omega)\},$$

where $N-1 = \text{rank}(\mathscr{F}_\omega)$ and $\lambda_N(\omega) = 0$. We can now define, for $k \in J(\omega)$,

$$\Pi_k(\omega) = \sum_{i \in I_k(\omega)} \varphi_i^\omega \otimes_2 \varphi_i^\omega,$$

which is the projection onto the k^{th} eigenspace of $\mathscr{F}_\omega$, also called the *k-th eigenprojector* of $\mathscr{F}_\omega$. This way, we can rewrite the eigen-decomposition of the spectral density operators as

$$\mathscr{F}_\omega = \sum_{j \in J(\omega)} \lambda_j(\omega)\Pi_j(\omega).$$

This notation is useful since it exhibits the identifiable parts of the singular value decomposition of $\mathscr{F}_\omega$. However, though the functions $\omega \mapsto \lambda_j(\omega)$ and $\omega \mapsto \Pi_j(\omega)$ are defined almost everywhere, their measurability is not necessarily guaranteed. For instance, consider the case where $\text{rank}(\mathscr{F}_\omega) = 2$ for $\omega \in [0, \pi/2]$ and $\text{rank}(\mathscr{F}_\omega) = 1$ if $\omega \in (\pi/2, \pi]$, and all the non-zero eigenvalues of the spectral density operators are distinct. Then the eigenprojector $\Pi_2(\omega)$ is well defined and trace-class (and in fact continuous) on $[0, \pi/2]$, but it is not trace-class on $(\pi/2, \pi]$. In order to avoid such complications, we define the set

$$J = \bigcap_{\omega \in [-\pi,\pi]} J(\omega) = \left\{ j \geq 1 : \lambda_j(\omega) > 0 \text{ for all } \omega \in [-\pi, \pi] \right\},$$

which corresponds to the set of indices of the non-repeated eigenvalues that never vanish. The following Theorem ensures the measurability of the eigenstructure of the spectral density operators under (weak) summability conditions.

Theorem 2.7.1. *If $\omega \mapsto \mathscr{F}_\omega$ is continuous on $[-\pi, \pi]$, with respect to $|||\cdot|||_\infty$, and $\mathscr{F}_\omega$ is compact for all $\omega \in [-\pi, \pi]$, then*

1. *For each $i \geq 1$, the functions*

$$\mu_i : [-\pi, \pi] \longrightarrow \mathbb{R}$$

are uniformly continuous, and therefore measurable.

2. *For all $j \in J$, the following functions are measurable:*

$$m_j : [-\pi, \pi] \longrightarrow \mathbb{R}$$
$$\lambda_j : [-\pi, \pi] \longrightarrow \mathbb{R}$$
$$\sum_{i \in I_j(\cdot)} \mu_i(\cdot) : [-\pi, \pi] \longrightarrow \mathbb{R}$$

3. *If $E \subset [-\pi,\pi]$ is an open subset and for some $k \in \cap_{\omega\in E} J(\omega)$, the function*

$$E \ni \omega \mapsto \sum_{j=1}^{k} m_j(\omega) \quad \text{is constant,} \tag{2.7.3}$$

then

$$\sum_{j=1}^{k} \Pi_j(\cdot) : E \longrightarrow \mathscr{S}_\infty(H)$$
$$\omega \longmapsto \sum_{j=1}^{k} \Pi_j(\omega)$$

is continuous, and therefore measurable on E.

Remark 2.7.2.

1. *Sufficient conditions for the assumptions of the Theorem is Condition 2.3.3.*

2. *Notice that the non-repeated eigenvalues $\lambda_j(\cdot)$ are always measurable, and that repeated eigenvalues $\mu_i(\cdot)$ are continuous.*

3. *Continuity of the eigenprojections a delicate matter. Indeed, each time two distinct eigenvalues $\mu_i(\cdot), \mu_{i+1}(\cdot)$ become equal, meaning that $\mu_i(\omega_0) = \mu_{i+1}(\omega_0)$ but $\mu_i(\cdot) \neq \mu_{i+1}(\cdot)$ on $E \setminus \omega_0$, for some set E containing ω_0, then there is a jump at $m_j(\cdot)$, for some $j \geq 1$, and therefore*

$$\omega \mapsto \sum_{i\in I_j(\omega)} \varphi_i^\omega \otimes_2 \varphi_i^\omega$$

is not continuous at ω_0. However, if the set $I_j(\omega)$ is constant on some open interval $E \subset [-\pi,\pi]$, then

$$\omega \mapsto \sum_{i\in I_j(\omega)} \varphi_i^\omega \otimes_2 \varphi_i^\omega$$

is continuous on E.

4. *Condition 2.7.3 could be replaced by*

$$\cup_{l=1}^{s} I_{j_l}(\cdot) \text{ is constant on the open set } E,$$

for some $j_1, \ldots, j_s \in \cap_{\omega\in E} J(\omega)$, which would give continuity of

$$\omega \mapsto \sum_{l=1}^{s} \Pi_{j_l}(\omega).$$

5. *If all non-zero eigenvalues of $\mathscr{F}_\omega$ are distinct, then for all $j \in J$, the eigenprojections $\Pi_j(\cdot)$ are measurable.*

6. *I conjecture that Condition 2.3.3 is sufficient for the mappings*

$$\omega \mapsto \sum_{i \in I_j(\omega)} \varphi_i^\omega \otimes_2 \varphi_i^\omega$$

to be measurable.

Proof of Theorem 2.7.1. We divide the proof into small steps:

(i). The functions $\mu_i(\omega)$ are continuous since

$$|\mu(\omega+\delta) - \mu(\omega)| \leq |||\mathscr{F}_{\omega+\delta} - \mathscr{F}_\omega|||_\infty \to 0, \quad |\delta| \to 0,$$

by Lemma A.2.4. Uniform continuity follows since $[-\pi, \pi]$ is compact.

(ii). Let us prove that the functions $\omega \mapsto m_k(\omega)$ are measurable, for $k \in J$. We proceed by induction. If $1 \in J$, the function $m_1(\omega)$ is measurable since

$$\begin{aligned} m_1(\omega) &= \left|\left\{i \geq 1 : \mu_i(\omega) = \mu_1(\omega)\right\}\right| \\ &= \sum_{i \geq 1} \mathbf{1}_{\{(\mu_i - \mu_1)^{-1}(0)\}}(\omega), \end{aligned}$$

and the latter sum is finite. Now if $m_1, \ldots, m_k$ are measurable, and $k+1 \in J$, we define the function $s_k(\omega) = 1 + \sum_{j=1}^k m_k(\omega)$. By induction, it is measurable, and we have

$$\begin{aligned} m_{k+1}(\omega) &= \left|\left\{i \geq 1 : \mu_i(\omega) = \mu_{s_k(\omega)}(\omega)\right\}\right| \\ &= \sum_{i \geq 1} \mathbf{1}_{\{g_i^{-1}(0)\}}(\omega), \end{aligned}$$

where $g_i(\omega) = \mu_i(\omega) - \mu_{s_k(\omega)}(\omega)$. Notice that the last sum is finite for each ω because $\mu_{s_k(\omega)}(\omega) > 0$. Since

$$g_i(\omega) = \mu_i(\omega) - \sum_{l \geq 1} \mu_l(\omega) \mathbf{1}_{\{s_k^{-1}(l)\}}(\omega),$$

each g_i is measurable, m_{k+1} is measurable. This completes the induction.

(iii). For each $k \in J$, the function $\lambda_k(\cdot)$ is measurable since

$$\lambda_k(\omega) = \mu_{s_k(\omega)}(\omega) = \sum_{l \geq 1} \mu_l(\omega) \mathbf{1}_{\{s_k^{-1}(l)\}}(\omega).$$

(iv). The function $\omega \mapsto \sum_{i \in I_k(\omega)} \mu_i(\omega)$ is measurable since

$$\sum_{i \in I_k(\omega)} \mu_i(\omega) = \sum_{i \geq 1} \mathbf{1}_{\{i \geq s_k(\omega)\}} \mathbf{1}_{\{i < s_k(\omega) + m_k(\omega)\}} \mu_i(\omega).$$

(v). If $\omega \mapsto \sum_{j=1}^{k} m_j(\omega)$ is constant, say equal to $k' \geq 1$ on the open set $E \subset [-\pi,\pi]$, then for all $\omega \in E$ and $(\omega_n)_{n\geq 1} \subset E$ converging to ω, we have

$$
\begin{aligned}
&\left|\!\left|\!\left| \sum_{j=1}^{k} \Pi_j(\omega_n) - \sum_{j=1}^{k} \Pi_j(\omega) \right|\!\right|\!\right|_\infty \\
&= \left|\!\left|\!\left| \sum_{j=1}^{k} \sum_{i\in I_j(\omega_n)} \varphi_i^{\omega_n} \otimes_2 \varphi_i^{\omega_n} - \sum_{j=1}^{k} \sum_{i\in I_j(\omega)} \varphi_i^{\omega} \otimes_2 \varphi_i^{\omega} \right|\!\right|\!\right|_\infty \\
&= \left|\!\left|\!\left| \sum_{i=1}^{\sum_{j=1}^{k} m_j(\omega_n)} \varphi_i^{\omega_n} \otimes_2 \varphi_i^{\omega_n} - \sum_{i=1}^{\sum_{j=1}^{k} m_j(\omega)} \varphi_i^{\omega} \otimes_2 \varphi_i^{\omega} \right|\!\right|\!\right|_\infty \\
&= \left|\!\left|\!\left| \sum_{i=1}^{k'} \left[\varphi_i^{\omega_n} \otimes_2 \varphi_i^{\omega_n} - \varphi_i^{\omega} \otimes_2 \varphi_i^{\omega} \right] \right|\!\right|\!\right|_\infty \\
&= \left|\!\left|\!\left| \sum_{j=1}^{k} \sum_{i\in I_j(\omega)} \left[\varphi_i^{\omega_n} \otimes_2 \varphi_i^{\omega_n} - \varphi_i^{\omega} \otimes_2 \varphi_i^{\omega} \right] \right|\!\right|\!\right|_\infty \\
&\leq \sum_{j=1}^{k} \left|\!\left|\!\left| \sum_{i\in I_j(\omega)} \left[\varphi_i^{\omega_n} \otimes_2 \varphi_i^{\omega_n} - \varphi_i^{\omega} \otimes_2 \varphi_i^{\omega} \right] \right|\!\right|\!\right|_\infty \\
&= \sum_{j=1}^{k} \left|\!\left|\!\left| \sum_{i\in I_j(\omega)} \varphi_i^{\omega_n} \otimes_2 \varphi_i^{\omega_n} - \Pi_j(\omega) \right|\!\right|\!\right|_\infty \\
&\to 0,
\end{aligned}
$$

by Theorem A.2.5. Since the limit is zero for arbitrary sequences $\omega_n \to \omega$, the continuity is shown. □

The previous result motivates introducing the following condition:

Condition 2.7.3. *$(X_t)_{t\in\mathbb{Z}}$ is a second-order stationary FTS with mean zero, $\mathbb{E}\|X_0\|^2 < \infty$, admitting a weak spectral density operators $\omega \mapsto \mathscr{F}_\omega$ that is $|||\cdot|||_\infty$-continuous. Furthermore, we assume that all the non-zero eigenvalues of $\mathscr{F}_\omega$ are all distinct, for each $\omega \in [-\pi,\pi]$.*

Notice that Condition 2.7.3 is actually stronger than Condition 2.4.1(∞) or Condition 2.3.3. Under Condition 2.7.3, the eigenprojectors $\omega \mapsto \Pi_j(\omega)$ are measurable (in fact continuous) if $\mu_j(\omega) > 0$ for all $\omega \in [-\pi,\pi]$.

2.8 Doubly Spectral Decomposition, Optimal Finite Dimensional Reduction

2.8.1 Harmonic Principal Component Analysis

It follows from the discussion in the previous section that the stochastic integral (2.4.11) defined by a truncation of the Cramér–Karhunen–Loève representation is well-defined. The purpose of this section is to prove that

the truncation at a level K of the Cramér–Karhunen–Loève representation provides in fact the best linear approximation of X with K degrees of freedom, in a sense that generalizes functional PCA.

Given the stationary functional time series $\{X_t\}_{t\in\mathbb{Z}}$ in the real part of H, and a sequence $\{a_s\}_{s\in\mathbb{Z}}$ of bounded operators on H, we can construct a new functional time series

$$Y_t = \sum_{s\in\mathbb{Z}} a_{t-s} X_s,$$

where Y_t is a random element of $L^2([0,1],\mathbb{R})$, which is said to be obtained by *(linear) filtering* of X_t, or refered to as filtered versions of X_t. Notice that the rank K approximation of X_t based on the PCA of $\mathscr{R}_0$ can also be expressed as a filtered version of X_t, by choosing $a_s = 0$ for all $s \neq 0$.

We already know from Theorem 2.5.5 that if Condition 2.7.3 holds and $\sum_s |||a_s|||_\infty < \infty$, then

$$Y_t = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} A(\omega) dZ_\omega^X, \quad A(\omega) = \sum_{s\in\mathbb{Z}} e^{-\mathbf{i}\omega s} a_s, \tag{2.8.1}$$

and

$$\mathscr{F}_\omega^Y = A(\omega)\mathscr{F}_\omega^X A^\dagger(\omega), \tag{2.8.2}$$

where we have denoted by $\mathscr{F}_\omega^X$, respectively Z^X, the spectral density operator at ω, respectively the orthogonal increment process, associated with X. This implies in particular that $\mathscr{F}_\omega^Y$ is continuous with respect to $|||\cdot|||_\infty$.

Consider now the problem of reducing the functional time series X_t to a finite dimensional vector series (say of dimensions q), by filtering X_t:

$$Y_t = \sum_s a_s X_{t-s} \in \mathbb{C}^q, \quad a_s \in \mathscr{S}_\infty(H,\mathbb{C}^q), \tag{2.8.3}$$

where $\mathscr{S}_\infty(H,\mathbb{C}^q)$ denotes the space of bounded operators from H to $\mathbb{C}^q$. Though the series Y_t is no longer interpretable in a functional sense, it may be filtered anew to yield a functional process

$$X_t^* = \sum_s b_s Y_{t-s}, \quad b_s \in \mathscr{S}_\infty(\mathbb{C}^q, H), \tag{2.8.4}$$

which is interpretable in a functional sense, and is in fact a rank q approximation of X_t. Fixing q orthonormal vectors $f_1,\dots,f_q \in H$, we can identify either $\mathbb{R}^q$ or $\mathbb{C}^q$ with the linear span (with real scalars, respectively with complex scalars) of $f_1,\dots,f_q$. This allows us to recast (2.8.3) and (2.8.4) into the framework developed in Section 2.5, and yields the following Lemma:

Lemma 2.8.1.

Assume Condition 2.7.3 holds. Let $\{a_s\}_{s\in\mathbb{Z}} \subset \mathscr{S}_\infty(H,\mathbb{C}^q)$ *and* $\{b_s\}_{s\in\mathbb{Z}} \subset$

$\mathscr{S}_\infty(\mathbb{C}^q, H)$ *such that*

$$\sum_{s\in\mathbb{Z}} \left(|||a_s|||_\infty + |||b_s|||_\infty \right) < \infty.$$

Then, the Cramér representation of X_t^**, defined in* (2.8.4)*, is*

$$X_t^* = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} B(\omega) A(\omega)\, dZ_\omega^X, \quad \text{in } \mathbb{H},$$

where $A(\omega) = \sum_{s\in\mathbb{Z}} e^{-\mathbf{i}\omega s} a_s$ *and* $B(\omega) = \sum_{s\in\mathbb{Z}} e^{-\mathbf{i}\omega s} b_s$*. Furthermore,*

$$\mathscr{F}_\omega^{X^*} = B(\omega) A(\omega) \mathscr{F}_\omega^\dagger A^\dagger(\omega) B^\dagger(\omega). \tag{2.8.5}$$

Proof. The proof follows easily from the identification of $\mathbb{C}^q$ as a closed linear subspace of H, and is omitted. □

Therefore, the Cramér representation of X_t^* is given by

$$X_t^* = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} C(\omega)\, dZ_\omega^X,$$

where $C(\omega) = B(\omega)A(\omega)$, and is hence of rank at most q. We are now in a position to show that the truncated Cramér–Karhunen–Loève expansion (2.4.11) provides a *Harmonic Principal Component Analysis* of X_t: under the mean square error approximation criterion

$$\mathbb{E} \left\| X_t - X_t^* \right\|^2,$$

which is independent of t by stationarity, the optimal choice of $\widetilde{C}_\omega$ is given by $\sum_{n=1}^{q(\omega)} \varphi_n^\omega \otimes_2 \varphi_n^\omega$, where we recall that $(\varphi_n^\omega)_{n\geq 1}$ are the eigenfunctions of $\mathscr{F}_\omega$.

Theorem 2.8.2 (Harmonic Principal Component Analysis)**.** *Assume Conditions 2.7.3 hold. Let*

$$X_t = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t}\, dZ_\omega$$

be a stationary time series in the real part of H*, and let* $X_t^* = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} C(\omega)\, dZ_\omega$*, with* $C \in \mathfrak{H}$*. Let* $q : [-\pi, \pi] \to \mathbb{N}$ *be a measurable function such that*

$$\mu_{q(\omega)}(\omega) > 0, \quad \omega \in [-\pi, \pi]. \tag{2.8.6}$$

Then, the solution to

$$\begin{aligned} &\min_{C\in\mathfrak{H}} \mathbb{E} \left\| X_t - X_t^* \right\|^2 \\ \text{subject to} \quad &\operatorname{rank}(C(\omega)) \leq q(\omega), \end{aligned}$$

is given by

$$C(\omega) = \sum_{j=1}^{q(\omega)} \varphi_j^\omega \otimes_2 \varphi_j^\omega,$$

where $\mathscr{F}_\omega = \sum_{j=1}^{\infty} \mu_j(\omega) \varphi_j^\omega \otimes_2 \varphi_j^\omega$ *is the spectral decomposition of* $\mathscr{F}_\omega$*. The approximation error is given by*

$$\mathbb{E} \left\| X_t - X_t^* \right\|^2 = \int_{-\pi}^{\pi} \left\{ \sum_{j > q(\omega)} \mu_j(\omega) \right\} d\omega.$$

Proof. The proof is an adaptation of Brillinger (2001, Theorem 9.3.1) to our case. Since

$$X_t - X_t^* = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} (I - C(\omega)) \, dZ_\omega,$$

Theorem 2.5.1 yields

$$\begin{aligned}
\mathbb{E} \left\| X_t - X_t^* \right\|^2 &= \int_{-\pi}^{\pi} \mathrm{Tr}\Big([I - C(\omega)] \mathscr{F}_\omega [I - C(\omega)]^\dagger \Big) d\omega \\
&= \int_{-\pi}^{\pi} \mathrm{Tr}\Big((I - C(\omega)) \mathscr{F}_\omega^{1/2} \left[(I - C(\omega)) \mathscr{F}_\omega^{1/2} \right]^\dagger \Big) d\omega \\
&= \int_{-\pi}^{\pi} \left| \left| \left| (I - C(\omega)) \mathscr{F}_\omega^{1/2} \right| \right|_2^2 d\omega, \qquad (2.8.7)
\end{aligned}$$

where $\mathscr{F}_\omega^{1/2} = \sum_{j=1}^{\infty} \sqrt{\mu_j(\omega)} \varphi_j^\omega \otimes_2 \varphi_j^\omega$. The term (2.8.7) is minimized by minimizing

$$\left| \left| \left| (I - C(\omega)) \mathscr{F}_\omega^{1/2} \right| \right|_2$$

for each ω. This is achieved, under our constraints, by

$$C(\omega) = \sum_{j=1}^{q(\omega)} \varphi_j^\omega \otimes_2 \varphi_j^\omega.$$

Notice that $C(\omega)$ is a $|||\cdot|||_\infty$-measurable function, since

$$\begin{aligned}
C(\omega) &= \sum_{j=1}^{q(\omega)} \varphi_j^\omega \otimes_2 \varphi_j^\omega \\
&= \sum_{i \geq 1} \mathbf{1}_{\{i \leq q(\omega)\}} \varphi_i^\omega \otimes_2 \varphi_i^\omega
\end{aligned}$$

and the latter sum is finite since $q(\omega) \in \mathbb{N}$. For the error term, notice that

$$\begin{aligned}
(I - C(\omega)) \mathscr{F}_\omega^{1/2} &= \left(\sum_{j \geq 1} - \sum_{j \leq q(\omega)} \right) \sqrt{\mu_j(\omega)} \varphi_j^\omega \otimes_2 \varphi_j^\omega \\
&= \sum_{j > q(\omega)} \sqrt{\mu_j(\omega)} \varphi_j^\omega \otimes_2 \varphi_j^\omega
\end{aligned}$$

and hence

$$\mathbb{E}\left\|X_t - X_t^*\right\|^2 = \int_{-\pi}^{\pi} \left(\sum_{j>q(\omega)} \mu_j(\omega) \right) d\omega.$$

The function in parentheses is measurable since it is an infinite sum of measurable functions, which is absolutely convergent. Indeed,

$$\int_{-\pi}^{\pi} \sum_{j>q(\omega)} |\mu_j(\omega)| d\omega \leq \int_{-\pi}^{\pi} |||\mathscr{F}_\omega|||_1 d\omega < \infty.$$

□

Remark 2.8.3.

1. *Condition 2.8.6 is here to ensure that the eigenprojections $\varphi_i(\omega) \otimes_2 \varphi_i(\omega)$ are identifiable for $i \leq q(\omega)$.*

2. *Contrary to the classical finite-dimensional results (e.g. Brillinger (2001)), we do not restrict $q(\omega)$ to be constant over ω. In this sense, when restricted to finite-dimensional Hilbert spaces, our results are more general than analogous results for vector-valued time series.*

Remark 2.8.4 (Representation as vector time series)**.**
Restricting $q(\omega)$ to be a constant functions—say $q(\omega) = q \in \mathbb{N}$ for all ω—yields a rank q version X_t^* of X_t. This can be represented in a 1-1 fashion by the filtered vector series $Y_t \in \mathbb{R}^q$ of (2.8.3)*, whose important characteristic is the lack of correlation between its coordinates, and also across time—just as one expects with the scores obtained in a traditional principal component analysis. The Y_t can therefore serve as the harmonic principal component scores. (see Section 2.8.2, in particular Remark 2.8.11 and Proposition 2.8.12)*

Remark 2.8.5 (Harmonic PCA dominates fPCA)**.**
Since the q-dimensional linear approximation of X_t given by fPCA can be written as PX_t, where $P \in \mathscr{S}_\infty(H)$ is an orthogonal projection operator of rank q, Theorem 2.8.2 tells us that

$$\mathbb{E}\left\|X_t - X_t^*\right\|^2 \leq \mathbb{E}\|X_t - PX_t\|^2. \tag{2.8.8}$$

In other words, the harmonic PCA dominates fPCA. Let us give another derivation of (2.8.8)*. Since $\mathbb{E}\|X_t\|^2 < \infty$, we just need to show that $\mathbb{E}\|PX_t\|^2 \leq \mathbb{E}\left\|X_t^*\right\|^2$. Let $\mathscr{R}_0 = \sum_{n\geq 1} \gamma_n \psi_n \otimes_2 \psi_n$ be the eigen-decomposition of the lag-0 autocovariance operator. We have that $P = \sum_{n=1}^{q} \psi_n \otimes_2 \psi_n$, and therefore,*

if $\mu_q(\omega) > 0$ for all $\omega \in [-\pi, \pi]$, we have

$$\begin{aligned}
\mathbb{E}\|PX_t\|^2 &= \mathbb{E}\,\mathrm{Tr}\,(PX_t \otimes_2 PX_t) \\
&= \mathrm{Tr}\,(P\mathscr{R}_0 P) \\
&= \mathrm{Tr}\left(\sum_{n=1}^{q} \gamma_n \psi_n \otimes_2 \psi_n\right) \\
&= \sum_{n=1}^{q} \left\langle \mathscr{R}_0 \psi_n, \psi_n \right\rangle \\
&= \sum_{n=1}^{q} \left\langle \left(\int_{-\pi}^{\pi} \mathscr{F}_\omega d\omega\right) \psi_n, \psi_n \right\rangle \quad \text{(using the inversion formula)} \\
&= \int_{-\pi}^{\pi} \left(\sum_{n=1}^{q} \left\langle \mathscr{F}_\omega \psi_n, \psi_n \right\rangle\right) d\omega \\
&\leq \int_{-\pi}^{\pi} \left(\sum_{n=1}^{q} \mu_n(\omega)\right) d\omega \qquad \text{(by Theorem A.2.3)} \\
&= \mathbb{E}\left\|X_t^*\right\|^2, \qquad \text{(by Theorem 2.5.1)}
\end{aligned}$$

and the claim is proven.

Let us now discuss the case where $\mathbb{E}\|PX_t\|^2 = \mathbb{E}\left\|X_t^*\right\|$. From the above calculations, this happens for $q \geq 1$ fixed if, and only if, for almost every $\omega \in [-\pi, \pi]$, we have the equality

$$\sum_{n=1}^{q} \psi_n \otimes_2 \psi_n = \sum_{n=1}^{q} \varphi_n^\omega \otimes_2 \varphi_n^\omega. \tag{2.8.9}$$

(recall that we are assuming that Condition 2.7.3 holds). Equation (2.8.9) *means that the eigenspace spanned by the first q eigenfunctions of the spectral density operators are constant in ω. Now if we assume that all the spectral density operators $\mathscr{F}_\omega$ are strictly positive definite, and that for each $\omega \in [-\pi, \pi]$, all their eigenvalues are distinct, then $\mathbb{E}\|PX_t\|^2 = \mathbb{E}\left\|X_t^*\right\|$ holds for all $q \geq 1$ if and only if*

$$\varphi_n^\omega \otimes_2 \varphi_n^\omega = \psi_n \otimes_2 \psi_n, \quad n \geq 1, \, \omega\text{-a.e.}$$

This implies that the eigen-decomposition of the spectral density operator at ω is given by

$$\mathscr{F}_\omega = \sum_{n \geq 1} \mu_n(\omega) \psi_n \otimes_2 \psi_n, \quad \omega\text{-a.e.}$$

and therefore, using the inversion formula, we get, for all $t \in \mathbb{Z}$,

$$\begin{aligned}
\mathscr{R}_t &= \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \mathscr{F}_\omega d\omega \\
&= \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \left(\sum_{n \geq 1} \mu_n(\omega) \psi_n \otimes_2 \psi_n\right) d\omega
\end{aligned}$$

$$= \sum_{n\geq 1} \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \mu_n(\omega) d\omega \cdot \left(\psi_n \otimes_2 \psi_n\right)$$

(by the dominated convergence Theorem)

$$= \sum_{n\geq 1} 2 \int_{0}^{\pi} \cos(\omega t) \mu_n(\omega) d\omega \cdot \left(\psi_n \otimes_2 \psi_n\right),$$

thus

$$\mathscr{R}_t = \sum_{n\geq 1} \gamma_n^t \psi_n \otimes_2 \psi_n, \quad t \in \mathbb{Z}, \tag{2.8.10}$$

where $\gamma_n^t = 2\int_0^\pi \cos(\omega t)\mu_n(\omega)d\omega$. *In particular,* (2.8.10) *implies that an eigenfunction of* $\mathscr{R}_0$ *will also be an eigenfunction of all lag-*t *autocovariance operators, and that the scores* $\xi_n^t = \langle X_t, \psi_n \rangle$ *of the Karhunen–Loève expansion of* X_t,

$$X_t = \sum_{n\geq 1} \xi_n^t \psi_n,$$

are uncorrelated in space, $\mathbb{E}\left[\xi_n^t \xi_m^t\right] = 0$ *for* $n \neq m$, *and also across time,*

$$\mathbb{E}\left[\xi_n^t \xi_m^s\right] = \left\langle \mathscr{R}_{t-s}\psi_m, \psi_n \right\rangle = 0, \quad t \neq s; n, m \geq 1.$$

Essentially, this means that $X_t \equiv \left(\xi_t^1, \xi_t^2, \ldots\right)^\mathsf{T}$, *where each* $(\xi_t^n)_{t\in\mathbb{Z}}$ *is a univariate time series that is uncorrelated with every* $(\xi_t^m)_{t\in\mathbb{Z}}$, *for* $m \neq n$. *Equation* (2.8.10) *also implies that*

$$\mathscr{R}_{-t} = \mathscr{R}_t^\dagger = \mathscr{R}_t, \quad \forall\, t \in \mathbb{Z},$$

which can be described as second-order time reversibility *of* X_t.
Therefore, harmonic PCA strictly dominates fPCA unless the FTS X_t *is second-order time reversible, or if all its autocovariance operators do not share the same eigenfunctions.*

We now make precise the way in which a Cramér–Karhunen–Loève representation of the form (2.4.9) of (2.4.10) holds.

Theorem 2.8.6 (Cramér–Karhunen–Loève decomposition)**.** *Under the conditions of Theorem 2.8.2, we have:*

$$\mathbb{E}\left\| X_t - \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \left(\sum_{n=1}^{q} \varphi_n^\omega \otimes_2 \varphi_n^\omega \right) dZ_\omega \right\|^2 = \int_{-\pi}^{\pi} \left\{ \sum_{n>q} \mu_n(\omega) \right\} d\omega. \tag{2.8.11}$$

Furthermore, if $\mathscr{F}_\omega$ *is strictly positive-definite for all* $\omega \in [-\pi, \pi]$, *we have that*

$$X_t = \sum_{n\geq 1} \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} (\varphi_n^\omega \otimes_2 \varphi_n^\omega) dZ_\omega \quad \text{in } \mathbb{H}, \tag{2.8.12}$$

where each of the summands

$$\int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} (\varphi_n^\omega \otimes_2 \varphi_n^\omega) dZ_\omega, \quad n \geq 1, t \in \mathbb{Z},$$

is almost surely in the real part of H, and these are orthogonal in space (n), across all time lags, i.e.,

$$\mathbb{E}\left[\int_{-\pi}^{\pi} e^{\mathbf{i}\omega t}(\varphi_n^\omega \otimes_2 \varphi_n^\omega)\, dZ_\omega \otimes_2 \int_{-\pi}^{\pi} e^{\mathbf{i}\omega s}(\varphi_n^\omega \otimes_2 \varphi_n^\omega)\, dZ_\omega\right] = \delta_{n,m} \cdot \int_{-\pi}^{\pi} e^{\mathbf{i}\omega(t-s)} \mu_n(\omega) \varphi_n^\omega \otimes_2 \varphi_n^\omega d\omega.$$

Remark 2.8.7 (The Cramér–Karhunen–Loève is a natural extension of the Karhunen–Loève expansion to FTS)**.**
This Theorem shows that (2.8.12) *decomposes the FTS X_t into a sum of components that are not only uncorrelated across n, within each timepoint t—a property already present in the Karhunen–Loève expansion—but also across time lags. In this sense, the Cramér–Karhunen–Loève decomposition is the natural extension of the Karhunen–Loève expansion for functional time series.*

Proof of Theorem 2.8.6. The first statement is a corollary of Theorem 2.8.2. Let $\Phi_N(\omega) = \sum_{n=1}^N \varphi_n^\omega \otimes_2 \varphi_n^\omega$ for any $N \geq 1$, including for $N = \infty$. Notice that

$$\|\Phi_\infty - \Phi_N\|_{\mathcal{H}}^2 = \int_{-\pi}^{\pi} \mathrm{Tr}\left(\sum_{n>N} \mu_n(\omega) \varphi_n^\omega \otimes_2 \varphi_n^\omega\right) d\omega \to 0, \quad N \to \infty, \quad (2.8.13)$$

by the dominated convergence Theorem. Therefore, by the isometry property (2.5.6),

$$\begin{aligned}
X_t &= \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \Phi_\infty(\omega)\, dZ_\omega \\
&= \mathscr{I}\left(\lim_{N\to\infty} \mathfrak{e}_t \Phi_N\right) \\
&= \lim_{N\to\infty} \mathscr{I}(\mathfrak{e}_t \Phi_N) \\
&= \lim_{N\to\infty} \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t}\left(\sum_{n=1}^N \varphi_n^\omega \otimes_2 \varphi_n^\omega\right) dZ_\omega \\
&= \lim_{N\to\infty} \sum_{n=1}^N \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t}\left(\varphi_n^\omega \otimes_2 \varphi_n^\omega\right) dZ_\omega \\
&= \sum_{n=1}^{\infty} \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t}\left(\varphi_n^\omega \otimes_2 \varphi_n^\omega\right) dZ_\omega.
\end{aligned}$$

Let us now show that $Y = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t}\left(\varphi_j^\omega \otimes_2 \varphi_j^\omega\right) dZ_\omega$ is almost surely in the real part of H, or equivalently, that $\overline{Y} = Y$ almost surely. Using Theorem 2.5.1, we get

$$\begin{aligned}
\overline{Y} &= \int_{-\pi}^{\pi} \overline{e^{\mathbf{i}(-\omega)t}\left(\varphi_j^{-\omega} \otimes_2 \varphi_j^{-\omega}\right)} dZ_\omega \quad \text{a.s.} \\
&= \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t}\left(\overline{\varphi_j^{-\omega} \otimes_2 \varphi_j^{-\omega}}\right) dZ_\omega
\end{aligned}$$

$$
\begin{aligned}
&= \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \left(\varphi_j^{\omega} \otimes_2 \varphi_j^{\omega}\right) dZ_{\omega} && \text{(by Proposition 2.8.8)} \\
&= Y.
\end{aligned}
$$

Turning to the last statement of the Theorem, Proposition 2.5.1 yields

$$
\begin{aligned}
&\mathbb{E}\left[\int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \left(\varphi_n^{\omega} \otimes_2 \varphi_n^{\omega}\right) dZ_{\omega} \otimes_2 \int_{-\pi}^{\pi} e^{\mathbf{i}\omega s} \left(\varphi_m^{\omega} \otimes_2 \varphi_m^{\omega}\right) dZ_{\omega}\right] \\
&= \int_{-\pi}^{\pi} e^{\mathbf{i}\omega(t-s)} \left(\varphi_n^{\omega} \otimes_2 \varphi_n^{\omega}\right) \mathscr{F}_{\omega} \left(\varphi_m^{\omega} \otimes_2 \varphi_m^{\omega}\right)^{\dagger} d\omega \\
&= \delta_{n,m} \int_{-\pi}^{\pi} e^{\mathbf{i}\omega(t-s)} \mu_n(\omega) \left(\varphi_n^{\omega} \otimes_2 \varphi_n^{\omega}\right) d\omega.
\end{aligned}
$$

This finishes the proof. □

2.8.2 Representation as a Vector Time Series

In this section, we will provide the expressions of the filters A, B involved in (2.8.3) and (2.8.4). We first give some additional properties of the weak spectral density operators.

Proposition 2.8.8.
Assume $\mathscr{F}_{\cdot} \in L^1([-\pi,\pi], \mathscr{S}_1(H))$ are the weak spectral density operators of the functional time series (X_t), taking values in the real part of H. Then

$$
\overline{\mathscr{F}_{\omega}} = \mathscr{F}_{-\omega}, \quad \text{for almost every } \omega \in [-\pi,\pi] .
$$

Furthermore, $\mu_j \geq 0$ is an eigenvalue of $\mathscr{F}_{\omega}$ if and only if it is an eigenvalue of $\mathscr{F}_{-\omega}$, i.e.,

$$
\mu_j \Pi_j(\omega) = \mathscr{F}_{\omega} \Pi_j(\omega) \quad \text{and} \quad \mu_j \Pi_j(-\omega) = \mathscr{F}_{-\omega} \Pi_j(-\omega),
$$

where $\Pi_j(\pm\omega)$ are the corresponding eigenprojections corresponding to the eigenvalue μ_j, and these satisfy

$$
\overline{\Pi_j(\omega)} = \Pi_j(-\omega).
$$

Proof. Assume without loss of generality that $\mathscr{F}_{\omega}$ is defined for all $\omega \in [-\pi,\pi]$. For any $t \in \mathbb{Z}$, we have

$$
\begin{aligned}
\int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \overline{\mathscr{F}_{\omega}} d\omega &= \overline{\int_{-\pi}^{\pi} e^{\mathbf{i}\omega(-t)} \mathscr{F}_{\omega} d\omega} \\
&= \overline{\mathscr{R}_{-t}} \\
&= \mathscr{R}_{-t} && \text{(since } X_t \text{ is real)} \\
&= \int_{-\pi}^{\pi} e^{\mathbf{i}(-\omega)t} \mathscr{F}_{\omega} d\omega \\
&= \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \mathscr{F}_{-\omega} d\omega,
\end{aligned}
$$

where we did the change of variables $\alpha = -\omega$ for the last equality. Proposition B.0.16 therefore yields the first statement of the proof. For the other statements, letting $\mathscr{F}_\omega = \sum_{j\geq 1} \mu_j(\omega)\Pi_j(\omega)$ be the eigendecomposition of the spectral density operator at ω, we have

$$\mathscr{F}_{-\omega} = \overline{\mathscr{F}_\omega} = \sum_{j\geq 1} \mu_j(\omega)\overline{\Pi_j(\omega)}.$$

Since $\overline{\Pi_i(\omega)}\,\overline{\Pi_j(\omega)} = \overline{\Pi_i(\omega)\Pi_j(\omega)} = \delta_{i,j}$, and the last statements follow from the uniqueness of the eigendecompositions of a compact operator. □

We now turn to the representation of X_t^* as a vector time series. We will do this assuming that Conditions 2.7.3 hold, and for a truncation function $q : [-\pi,\pi] \to \mathbb{N}$ that is càdlàg with a finite number of jumps, i.e. $q(\omega) = \sum_{l=1}^{L} \mathbf{1}_{[\omega_l,\omega_{l+1})} q_l$, where

$$-\pi = \omega_1 < \cdots < \omega_{L+1} = \pi,$$

and the q_ls are non-negative integers satisfying $q(\omega) \in I(\omega)$ for all $\omega \in [-\pi,\pi]$. Note that allowing $q_l = 0$ for some l corresponds to filtering out the frequencies in the range $[\omega_l,\omega_{l+1})$. First we need to find, for each $\omega \in [-\pi,\pi]$, operators $A(\omega) \in \mathscr{S}_\infty(H,\mathbb{C}^q)$ and $B(\omega) \in \mathscr{S}_\infty(\mathbb{C}^q,H)$ such that

$$B(\omega)A(\omega) = \sum_{j=1}^{q(\omega)} \varphi_j^\omega \otimes_2 \varphi_j^\omega,$$

where $q = \max_\omega q(\omega)$, Letting $v_1,\ldots,v_q$ be the canonical orthonormal basis of $\mathbb{C}^q$ (or any orthonormal basis), we see that choosing

$$A(\omega) = \begin{pmatrix} \varphi_1^\omega \\ \varphi_2^\omega \\ \vdots \\ \varphi_{q(\omega)}^\omega \\ 0 \\ \vdots \\ 0 \end{pmatrix} = \sum_{j=1}^{q(\omega)} v_j \otimes_2 \varphi_j^\omega,$$

and $B(\omega) = A(\omega)^\dagger$ will work. The next step is to take the Fourier transform of $A(\cdot)$ and $B(\cdot)$ to define filters $(a_s)_{s\in\mathbb{Z}}$ and $(b_s)_{s\in\mathbb{Z}}$ that define the vector series (Y_t), and the reconstructed series (X_t^*). In order to do so, we will assume that $\omega \mapsto \varphi_j^\omega$ is measurable, as a $[-\pi,\pi] \mapsto H$ mapping, for all $j = 1,\ldots,q$. The need for such an assumption is explained in the following remark.

Remark 2.8.9.
Assuming measurability of the mapping $\omega \mapsto \varphi_j^\omega$ is needed, even under Condition 2.7.3, because the eigenfunctions $(\varphi_j^\omega)_{j\geq 1}$ are not identifiable. Indeed, if φ is an eigenfunction of $\mathscr{F}_\omega$, $\alpha\varphi$ will also be an eigenfunction with the same eigenvalue, for any $\alpha \in \mathbb{C}$ of modulus one. Therefore the function $\omega \mapsto \varphi_j^\omega$ could be modified to $\omega \mapsto \alpha(\omega)\varphi_j^\omega$ for any (possibly non-measurable) function $\omega \mapsto \alpha(\omega) \in \mathbb{C}$ satisfying $|\alpha(\omega)| = 1$ for all $\omega \in [-\pi, \pi]$.

We now turn to the filters (a_s). These need to satisfy $A(\omega) = \sum_{s\in\mathbb{Z}} e^{-\mathbf{i}\omega s} a_s$. Applying Proposition B.0.17, we can take

$$a_s = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{\mathbf{i}\omega s} A(\omega)\, d\omega.$$

More precisely, we have for all $s \in \mathbb{Z}$,

$$\begin{aligned}
a_s &= (2\pi)^{-1} \int_{-\pi}^{\pi} e^{\mathbf{i}\omega s} A(\omega)\, d\omega \\
&= (2\pi)^{-1} \int_{-\pi}^{\pi} e^{\mathbf{i}\omega s} \left[\sum_{j=1}^{q(\omega)} v_j \otimes_2 \varphi_j^\omega \right] d\omega \\
&= (2\pi)^{-1} \sum_{l=1}^{L} \int_{\omega_l}^{\omega_{l+1}} e^{\mathbf{i}\omega s} \left[\sum_{j=1}^{q_l} v_j \otimes_2 \varphi_j^\omega \right] d\omega \\
&= (2\pi)^{-1} \sum_{l=1}^{L} \sum_{j=1}^{q_l} v_j \otimes_2 \left[\int_{\omega_l}^{\omega_{l+1}} e^{-\mathbf{i}\omega s} \varphi_j^\omega\, d\omega \right].
\end{aligned}$$

The filters (b_s) are then given by the relation $b_s = a_{-s}^\dagger$. Explicitly,

$$b_s = (2\pi)^{-1} \sum_{l=1}^{L} \sum_{j=1}^{q_l} \left[\int_{\omega_l}^{\omega_{l+1}} e^{\mathbf{i}\omega s} \varphi_j^\omega\, d\omega \right] \otimes_2 v_j.$$

Let us now give some remarks about the filtered series Y_t:

Remark 2.8.10 (Conditions for Y_t to be real valued)**.**
The series Y_t, constructed by filtering X_t with the filter (a_s) is not necessarily real valued. Sufficient conditions for it to be real valued is that $a_s X_{t-s}$ takes values in the real part of H, or equivalently that $a_s = \overline{a_s}$. Since

$$a_s = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{\mathbf{i}\omega s} A(\omega)\, d\omega$$

and

$$\overline{a_s} = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{-\mathbf{i}\omega s} \overline{A(\omega)}\, d\omega = (2\pi)^{-1} \int_{-\pi}^{\pi} e^{\mathbf{i}\omega s} \overline{A(-\omega)}\, d\omega,$$

we would need $\overline{A(\omega)} = A(-\omega)$ for almost every ω, or equivalently that

(i) $q(\omega) = q(-\omega)$ almost everywhere, and

(ii) $\overline{\varphi_j^\omega} = \varphi_j^{-\omega}$, for each $j = 1, 2, \ldots, q(\omega)$.

Condition (i) is easy to satisfy. Let us discuss Condition (ii). Since the eigenspaces of the spectral density operators corresponding to non-zero eigenvalues are assumed to be of multiplicity one, and $\overline{\Pi_j(\omega)} = \Pi_j(-\omega)$ *(see Proposition 2.8.8), we know that there exists a complex number* $\alpha(\omega)$ *of modulus one such that* $\overline{\varphi_j^\omega} = \alpha(\omega)\cdot\varphi_j^{-\omega}$. *It is possible to choose the* φ_j^ω*s such that* $\alpha(\omega) = 1$, *yielding therefore a* real valued *vector time series* Y_t. *Indeed, just construct* $\omega \mapsto \varphi_j^\omega$ *for* $\omega \in [0,\pi]$, *and then extend it to* $\omega \in [-\pi, 0)$ *using the relation* $\varphi_j^{-\omega} = \overline{\varphi_j^\omega}$.

Remark 2.8.11.
In the particular case where $q(\omega) = q$ *is constant, we have*

$$a_s = (2\pi)^{-1}\sum_{j=1}^{q} v_j \otimes_2 \left[\int_{-\pi}^{\pi} e^{-\mathbf{i}\omega s}\varphi_j^\omega d\omega\right] = (2\pi)^{-1}\int_{-\pi}^{\pi} e^{-\mathbf{i}\omega s}\begin{pmatrix}\varphi_1^\omega\\ \vdots\\ \varphi_q^\omega\end{pmatrix} d\omega$$

and

$$b_s = (2\pi)^{-1}\sum_{j=1}^{q}\left[\int_{-\pi}^{\pi} e^{\mathbf{i}\omega s}\varphi_j^\omega d\omega\right]\otimes_2 v_j = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega s}\left(\varphi_1^\omega,\ldots,\varphi_q^\omega\right) d\omega.$$

In this case, the series Y_t *is given by*

$$Y_t = \sum_{j=1}^{q} v_j \otimes_2 \left[\sum_{s\in\mathbb{Z}}\int_{-\pi}^{\pi} e^{\mathbf{i}\omega s}\left\langle X_{t-s}, \varphi_j^\omega\right\rangle d\omega\right] = \sum_{s\in\mathbb{Z}}\int_{-\pi}^{\pi} e^{\mathbf{i}\omega s}\begin{pmatrix}\left\langle X_{t-s},\varphi_1^\omega\right\rangle\\ \vdots\\ \left\langle X_{t-s},\varphi_q^\omega\right\rangle\end{pmatrix} d\omega. \tag{2.8.14}$$

In particular, we only need condition (ii) *to hold for* Y_t *to be real-valued.*

Proposition 2.8.12.
Assume Conditions 2.7.3 holds for X_t, *and let* Y_t *be defined by* (2.8.14). *Then, the spectral density of* Y_t *is given by*

$$\mathscr{F}_\omega^Y = \begin{pmatrix}\mu_1(\omega) & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & \mu_q(\omega)\end{pmatrix} = \operatorname{diag}\left(\mu_1(\omega),\ldots,\mu_q(\omega)\right),$$

where $\mu_j(\omega)$ *is the* j*-th eigenvalue of* $\mathscr{F}_\omega^X$. *Furthermore, if we write* $Y_t = \left(Y_{t,1},\ldots,Y_{t,q}\right)$, *we have*

$$\operatorname{cov}\left(Y_{t,i}, Y_{0,j}\right) = \delta_{i,j}\cdot\int_{-\pi}^{\pi}\exp\left[\mathbf{i}\omega t\right]\mu_i(\omega)d\omega, \quad \forall i,j = 1,\ldots,q; \quad t\in\mathbb{Z}.$$

In other words, all the distinct coordinates of Y_t *are uncorrelated across all time lags*

Proof. Let a_s be defined as in Remark 2.8.11. We can assume without

loss of generality, that $(v_j)_{j=1}^q \subset H$, and therefore that $a_s \in \mathscr{S}_2(H)$, and $A \in L^2([-\pi,\pi], \mathscr{S}_2(H))$. By Proposition B.0.17, we know that

$$\sum_{|n|<N} a_s \mathfrak{e}_{t-s} \to \mathfrak{e}_t A \quad \text{as } N \to \infty, \tag{2.8.15}$$

in $L^2([-\pi,\pi], \mathscr{S}_2(H))$, where $A(\omega) = \sum_{j=1}^q v_j \otimes_2 \varphi_j^\omega$. Notice that the functions $\sum_{|n|<N} a_s \mathfrak{e}_{t-s}$ belong to the linear space

$$V = \left\{ \sum_{j=1}^q v_j \otimes_2 g_j : g_j \in L^2([-\pi,\pi], H) \right\}.$$

Furthermore, for any $g \in V$, we have

$$\begin{aligned} \|g\|_{\mathfrak{H}}^2 &= \int_{-\pi}^{\pi} \left|\!\left|\!\left| g(\omega) \mathscr{F}_\omega g^\dagger(\omega) \right|\!\right|\!\right|_1 d\omega \\ &\le \int_{-\pi}^{\pi} \left|\!\left|\!\left| g(\omega) \right|\!\right|\!\right|_2^2 |||\mathscr{F}_\omega|||_1 d\omega \\ &\le c \int_{-\pi}^{\pi} \left|\!\left|\!\left| g(\omega) \right|\!\right|\!\right|_2^2 d\omega, \end{aligned}$$

where $c = \sup_\omega |||\mathscr{F}_\omega|||_1 < \infty$. Therefore (2.8.15) also holds in $\mathfrak{H}$, and Theorem 2.5.5 can be applied. We get

$$\begin{aligned} \mathscr{F}_\omega^Y &= A(\omega) \mathscr{F}_\omega^X A^\dagger(\omega) \\ &= \left(\sum_{j=1}^q v_j \otimes_2 \varphi_j^\omega \right) \left(\sum_{i \ge 1} \mu_i(\omega) \varphi_i^\omega \otimes_2 \varphi_i^\omega \right) \left(\sum_{j=1}^q \varphi_j^\omega \otimes_2 v_j \right) \\ &= \sum_{j=1}^q \mu_j(\omega) \left(v_j \otimes_2 v_j \right). \end{aligned}$$

Using the inversion formula, we get

$$\operatorname{cov}\left(Y_{t,i}, Y_{0,j}\right) = \left\langle \mathscr{R}_t^Y v_j, v_i \right\rangle = \delta_{i,j} \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} \mu_i(\omega) d\omega,$$

which finishes the proof. □

2.9 Outlook

As far as I know, the present Chapter is the first attempt to generalize the Karhunen–Loève expansion to the setting of functional time series (apart from the article (Panaretos & Tavakoli 2013*a*), see discussion below). The Cramér–Karhunen–Loève decomposition has several potential applications. Since an FTS X_t can be approximately represented by a vector time series Y_t whose coordinates are uncorrelated across all time lags (see Section 2.8.2), the Cramér–Karhunen–Loève decomposition can

be used for modeling purposes. It would be preferable to current methodology (based on truncation of the Karhunen–Loève expansion;see, e.g., Aue, Norinho & Hörmann (2014, to appear)) because truncation of the Cramér–Karhunen–Loève decomposition explains more of the series total variance, and the models under consideration would have fewer parameters (because of the uncorrelatedness of the coordinates of Y_t across all time lags). The Cramér–Karhunen–Loève decomposition could also be used to simulate an FTS with specific properties, or for comparing the second-order dynamics of functional time series (see Chapter 4).
Extensions of the results of this chapter to general second-order stationary functional time series, not necessarily admitting a spectral density, would be of interest. For univariate time series, this is done through Herglotz's Theorem (which is a particular case of Bochner's theorem, see Rudin (1991)), which states that for any second-order stationary time series, there exists a spectral measure F on $[-\pi,\pi]$ such that $\operatorname{cov}(X_t, X_0) = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t} dF(\omega)$. The extension to the multivariate (finite dimensional) setting causes no problem, because the results can be derived for each coordinate and then put together, without any convergence problem since the number of coordinates is finite. The functional case is however more complicated, because of the infinite number of coordinates. However, it seems that Milnos' Theorem (Kuo 1996) could allow the extension of our results to general second-order stationary functional time series.

Differences with Panaretos & Tavakoli (2013a)

Although an earlier version of results presented in this chapter has been published (Panaretos & Tavakoli 2013*a*), several results presented in this chapter have weaker assumptions than those of Panaretos & Tavakoli (2013*a*). Indeed, Panaretos & Tavakoli (2013*a*) give a functional Cramér representation for FTS in $L^2([0,1],\mathbb{R})$, under summability conditions on the *nuclear* norm of the autocovariance operators, continuity of the autocovariance kernels, and their summability in the supremum norm. The functional Cramér representation given in Section 2.4 only assumes the existence of weak spectral density operators, which always exists under weaker summability conditions (see Proposition 2.3.5). Furthermore, we derive in Section 2.5 the Cramér representation of linear filterings of a functional time series, under summability assumptions of the *operator norm* of the filtering operators, whereas Panaretos & Tavakoli (2013*a*) derive the same result under summability of the Hilbert–Schmidt norm of the filters, a much more stringent assumption.

CHAPTER 3

Inference for the Spectral Density Operators

By combining a functional Cramér representation with the Karhunen–Loève expansion, we developed in Chapter 2 a doubly spectral decomposition (the Cramér–Karhunen–Loève decomposition) for second-order stationary functional time series. This decomposition generalizes the Karhunen–Loève expansion, and provides a way of approximating the functional series by a vector series which dominates traditional fPCA. The main objects involved in this construction were the spectral density operators—whose Fourier coefficients are given by the autocovariance operators of the functional time series—and their eigenstructure. In practice, the spectral density operators and their eigenstructure must be estimated from data. The purpose of this chapter is to address this problem. We begin by introducing some notation (Section 3.1), and defining the main objects that will be used throughout this chapter (Section 3.2). We then present, in Section 3.3, the functional Discrete Fourier Transform, and study its asymptotic properties. This will play a major role in the estimation of the spectral density operators. We will see in particular that its asymptotic covariance is given by the spectral density operators, a property that motivates the study of their empirical covariance, the *periodogram operators* (Section 3.4). We shall see that these are asymptotically unbiased—but inconsistent—estimators of the spectral density operators. Nevertheless, by smoothing the periodogram (Section 3.5), we can circumvent this problem. We will show in Section 3.6 that the resulting estimators, the *sample spectral density operators*, are consistent and asymptotically Gaussian estimators of the spectral density operators, un-

der suitable weak dependence assumptions. We then show, in Section 3.7, consistency and asymptotic normality for estimators of the eigenstructure of the spectral density operators. Since Functional Data are usually recorded from discrete and noisy observations, we provide in Section 3.8 some conditions under which estimators constructed under noisy and discrete observations are still consistent. We present in Section 3.9 some numerical simulations to assess the finite sample performance of our estimators, and give in Section 3.10 a discussion of the technical conditions involved in our asymptotic results, and how they might be weakened. A brief outlook (Section 3.11) concludes this chapter. Some technical results used in this chapter are gathered in Section 3.12.
Although an earlier version of this chapter has been published (Panaretos & Tavakoli 2013*a*,*b*), some results present in this chapter are new or stronger than those already published (e.g. Theorem 3.6.1, Proposition 3.10.1 and Theorem 3.8.3).

3.1 Notation

In this chapter, we will study the problem of estimation of the spectral density operators and their eigenstructure, for a functional time series $(X_t)_{t\in\mathbb{Z}}$ taking values in $L^2([0,1],\mathbb{R})$. Although we focus here on the case $H = L^2([0,1],\mathbb{R})$, the proofs can be straightforwardly extended to more general complexified separable Hilbert spaces. Since

$$L^2([0,1],\mathbb{R}) \subset L^2([0,1],\mathbb{C}),$$

we can view our functional time series $X_t \in L^2([0,1],\mathbb{R})$ as an FTS in $L^2([0,1],\mathbb{C})$ that takes only real values.
We shall denote the inner product of $L^2([0,1],\mathbb{C})$ by $\langle\cdot,\cdot\rangle$, and denote its corresponding norm by $\|\cdot\|$. We will denote the imaginary number by $\mathbf{i} \in \mathbb{C}$, i.e. $\mathbf{i}^2 = -1$, and denote the complex conjugate of $\alpha \in \mathbb{C}$ by $\overline{\alpha}$. We also define $\Delta^{(T)}(\omega) = \sum_{t=0}^{T-1} \exp(-\mathbf{i}\omega t)$.
We will use the letters τ, σ to denote the parametrization index of the curves X_t, i.e.

$$\tau \mapsto X_t(\tau), \quad \tau \in [0,1].$$

Of course, since $X_t \in L^2([0,1],\mathbb{C})$, the pointwise evaluation $X_t(\tau)$ does not make sense, unless the curves are smooth, but writing the results in such a way makes it easier to understand the statements , by allowing analogies with the multivariate case. Therefore, statements such as

$$\mathbb{E}\, X_t(\tau) = \mu(\tau), \quad \tau \in [0,1]$$

should be understood as

$$\left\| \mathbb{E}\, X_t - \mu \right\| = 0.$$

In order to emphasize this, we will often write

$$\mathbb{E}\, X_t(\tau) = \mu(\tau), \quad \text{in } L^2.$$

At times, we will need to express equalities between elements of the Hilbert space $L^2\left([0,1]^2, \mathbb{R}\right)$, or more generally $L^2([0,1]^p, \mathbb{C})$. We will abuse notation and write

$$\langle f, g \rangle_2 = \int_{[0,1]^p} f(y)\overline{g(y)}dy, \quad f, g \in L^2([0,1]^p, \mathbb{C}), \tag{3.1.1}$$

and $\|\cdot\|_2 = \langle \cdot, \cdot \rangle_2$, without changing the notation for distinct dimensions p. When stating an asymptotic result, we shall use the small-o, big-O notation: for a sequence $(a_n) \subset H$, where $(H, \|\cdot\|)$ is a Hilbert space, $a_n = o(1)$ means $a_n \to 0$, $a_n = O(1)$ means that there exists some $M > 0$ such that

$$\|a_n\| < M, \quad n \text{ large enough},$$

and $a_n = o(r_n)$, respectively $a_n = O(r_n)$, means $a_n r_n^{-1} = o(1)$, respectively $a_n r_n^{-1} = O(1)$. For instance,

$$a^{(T)}(\tau, \sigma) = b(\tau, \sigma) + o(T^{-1}), \quad \text{as } T \to \infty, \text{in } L^2,$$

will mean

$$T\left\| a^{(T)} - b \right\|_2 \to 0, \quad T \to \infty.$$

For a function $g : D \subset \mathbb{R}^n \to \mathbb{C}$, we denote $\left\| g \right\|_\infty = \sup_{\mathbf{x} \in D} |g(\mathbf{x})|$.
A function $a \in L^2\left([0,1]^2, \mathbb{C}\right)$ induces an operator A on $L^2([0,1], \mathbb{C})$, through *right-integration*,

$$Af(\tau) = \int_0^1 a(\tau, \sigma)d\sigma, \quad f \in L^2([0,1], \mathbb{C}).$$

The function a is called the *kernel* of A, and A is called an *integral operator*. We have $\|a\|_2 = |||A|||_2$, and there is in fact a bijective correspondence between such kernels and Hilbert–Schmidt operators.

recall that $|||\cdot|||_2$ denotes the Hilbert–Schmidt norm, see Section A.2.2.2 on page 220

3.2 Basic Definitions and Main Assumptions

We will assume in this chapter that X_t is a second-order FTS in $L^2([0,1], \mathbb{R})$, with $\mathbb{E}\|X_0\|^2 < \infty$. The mean function of X_t is defined by

$$\mu(\tau) = \mathbb{E}\, X_t(\tau), \quad \text{in } L^2,$$

and the lag-t autocovariance kernel of X_t by

$$r_t(\tau, \sigma) = \mathbb{E}\left[\left(X_t(\tau) - \mu(\tau)\right)\left(X_t(\sigma) - \mu(\sigma)\right)\right], \quad t \in \mathbb{Z}, \quad \text{in } L^2. \tag{3.2.1}$$

If continuity in mean square of X_t is assumed, then the autocovariance kernels are also well defined pointwise. Each kernel r_t induces a corresponding operator $\mathscr{R}_t : L^2([0,1],\mathbb{C}) \to L^2([0,1],\mathbb{C})$ by right integration, the *autocovariance operator at lag t*,

$$\mathscr{R}_t h(\tau) = \int_0^1 r_t(\tau,\sigma)h(\sigma)\mathrm{d}\sigma = \mathrm{cov}\left(\langle X_0, h\rangle, X_t(\tau)\right), \qquad h \in L^2([0,1],\mathbb{C}).$$

One of the notions we will employ to quantify the weak dependence among the observations $\{X_t\}$ is that of a *cumulant kernel* of the series.

Definition 3.2.1.
The k-th order cumulant kernel *of* $X_{t_1},\ldots,X_{t_k}$ *is the unique element* $A \in L^2\left([0,1]^k,\mathbb{C}\right)$ *that satisfies*

$$\left\langle A, \varphi_1 \otimes \varphi_2 \otimes \cdots \otimes \varphi_k \right\rangle = \mathrm{cum}\left(\left\langle X_{t_1}, \varphi_1\right\rangle, \ldots, \left\langle X_{t_k}, \varphi_k\right\rangle\right), \tag{3.2.2}$$

for all $\varphi_1,\ldots,\varphi_n \in L^2([0,1],\mathbb{C})$. *We write*

$$A = \mathrm{cum}\left(X_{t_1},\ldots,X_{t_k}\right).$$

We recall that for random variables $Y_1,\ldots,Y_k \in \mathbb{C}$, *the cumulant is defined by*

$$\mathrm{cum}\left(Y_1,\ldots,Y_k\right) = \sum_{\nu=(\nu_1,\ldots,\nu_p)} (-1)^{p-1}(p-1)! \prod_{l=1}^{p} \mathbb{E}\left[\prod_{j\in\nu_l} Y_j\right],$$

where the sum extends over all unordered partitions of $\{1,\ldots,k\}$.

An intuitive way of understanding the previous definition is that

$$\begin{aligned}\mathrm{cum}\left(X_{t_1},\ldots,X_{t_k}\right)(\tau_1,\ldots,\tau_k) &= \mathrm{cum}\left(X_{t_1}(\tau_1),\ldots,X_{t_k}(\tau_k)\right)\\ &= \sum_{\nu=(\nu_1,\ldots,\nu_p)} (-1)^{p-1}(p-1)! \prod_{l=1}^{p} \mathbb{E}\left[\prod_{j\in\nu_l} X_{t_j}(\tau_j)\right],\end{aligned}$$

in L^2, where the sum extends over all unordered partitions of $\{1,\ldots,k\}$. Assuming $\mathbb{E}\|X_0\|^k < \infty$, $k \geq 1$, guarantees that the cumulant kernels of order k are well defined (see Proposition 3.12.6).
A cumulant kernel of order $2k$ gives rise to a corresponding *$2k$-th order cumulant operator* $\mathscr{R}_{t_1,\ldots,t_{2k}} : L^2([0,1]^k,\mathbb{C}) \to L^2([0,1]^k,\mathbb{C})$, also denoted $\mathrm{cumop}\left(X_{t_1},\ldots,X_{t_{2k}}\right)$, and defined by

$$\begin{aligned}\mathscr{R}_{t_1,\ldots,t_{2k}}(\varphi_{k+1}\otimes\ldots\otimes\varphi_{2k}) &= \mathrm{cum}\left(X_{t_1},\ldots,X_{t_k},\left\langle\varphi_{k+1},\overline{X_{t_{k+1}}}\right\rangle,\ldots,\left\langle\varphi_{2k},\overline{X_{t_{2k}}}\right\rangle\right)\\ &= \mathrm{cum}\left(X_{t_1},\ldots,X_{t_k},\left\langle\varphi_{k+1},X_{t_{k+1}}\right\rangle,\ldots,\left\langle\varphi_{2k},X_{t_{2k}}\right\rangle\right)\end{aligned} \tag{3.2.3}$$

for all $\varphi_{k+1},\ldots,\varphi_{2k} \in L^2([0,1],\mathbb{C})$. The equivalent pointwise definition of

this operator is given by

$$\mathscr{R}_{t_1,\ldots,t_{2k}} h(\tau_1,\ldots,\tau_k) = \int_{[0,1]^k} \operatorname{cum}\left(X_{t_1}(\tau_1),\ldots,X_{t_{2k-1}}(\tau_{2k-1}),X_{t_{2k}}(\tau_{2k})\right) \times h(\tau_{k+1},\ldots,\tau_{2k})\mathrm{d}\tau_{k+1}\ldots\mathrm{d}\tau_{2k}.$$

The following weak dependence conditions will be used in this chapter:

Condition C(l,k): X_t is k-th order stationary, and for each $j = 1,\ldots,k-1$,

$$\sum_{t_1,\ldots,t_{k-1}=-\infty}^{\infty} (1+|t_j|^l)\left\|\operatorname{cum}\left(X_{t_1},\ldots,X_{t_{k-1}},X_0\right)\right\|_2 < \infty.$$

For completeness, we recall here the definition of k-th order stationarity:

Definition 3.2.2.
A sequence of random elements $(\varepsilon_t)_{t\in\mathbb{Z}}$ in H with $\mathbb{E}\|\varepsilon_t\|^k < \infty$ for all $t \in \mathbb{Z}$ is called k-th order stationary *(k a positive integer) if for all $t_1,\ldots,t_l \in \mathbb{Z}$, and $l = 1,\ldots,k$,*

$$\mathbb{E}\left[\varepsilon_{t_1+s}\otimes\varepsilon_{t_2+s}\otimes\cdots\otimes\varepsilon_{t_l+s}\right]$$

is independent of $s \in \mathbb{Z}$.

Notice that condition **C(0,2)** is equivalent to

$$X_t \text{ is second-order stationary} \quad \textit{and} \quad \sum_{t\in\mathbb{Z}} \|r_t\|_2 = \sum_{t\in\mathbb{Z}} |||\mathscr{R}_t|||_2 < \infty.$$

Recall that if X_t is a second-order stationary FTS with $\sum_{t\in\mathbb{Z}} |||\mathscr{R}_t|||_\infty$ and $\sum_{t\in\mathbb{Z}} |\operatorname{Tr}(\mathscr{R}_t)| < \infty$, Proposition 2.3.5 tells us that the spectral density operators of X_t are given by

$$\mathscr{F}_\omega = (2\pi)^{-1}\sum_{t\in\mathbb{Z}} e^{-\mathbf{i}\omega t}\mathscr{R}_t, \quad \omega\in[-\pi,\pi],$$

that they are continuous in ω (with respect to $|||\cdot|||_\infty$), and that they are uniformly bounded: $|||\mathscr{F}_\omega|||_1 \le M < \infty$ for all $\omega \in [-\pi,\pi]$. Under the additional condition that $\omega \mapsto \operatorname{Tr}(\mathscr{F}_\omega)$ is continuous, we have an even stronger result:

Proposition 3.2.3.
Suppose Conditions 2.3.3 and 2.3.4 hold, and that

$$\omega \mapsto \operatorname{Tr}(\mathscr{F}_\omega), \quad \omega\in[-\pi,\pi],$$

is continuous. Then, for any orthonormal basis $(e_n)_{n\ge1}$ of $L^2([0,1],\mathbb{C})$, the convergence of the following series is uniform in $\omega \in [-\pi,\pi]$:

$$\operatorname{Tr}(\mathscr{F}_\omega) = \sum_{n=1}^{\infty} \langle \mathscr{F}_\omega e_n, e_n\rangle.$$

Proof. Let $g_n(\omega) = \langle \mathscr{F}_\omega e_n, e_n \rangle$. Notice that g_n is continuous on $[-\pi,\pi]$, and $g_n(\omega) \geq 0$ for all $\omega \in [-\pi,\pi]$. Defining

$$G_N(\omega) = \sum_{n=1}^{N} g_n(\omega),$$

we have that $G_N(\omega) \leq G_{N+1}(\omega)$ for all $N \geq 1$, and all $\omega \in [-\pi,\pi]$. Furthermore, every G_N is continuous on $[-\pi,\pi]$, and

$$\mathrm{Tr}(\mathscr{F}_\omega) = \lim_{N\to\infty} G_N(\omega), \quad \omega \in [-\pi,\pi].$$

Since $\omega \mapsto \mathrm{Tr}(\mathscr{F}_\omega)$ is continuous on the compact set $[-\pi,\pi]$, Dini's Theorem (Rudin 1976, Theorem 7.13) tells us that the convergence is uniform on $[-\pi,\pi]$. □

The following proposition gives some properties of the spectral density operators, under some summability conditions that are stronger than those seen in Chapter 2.

Proposition 3.2.4.
Suppose $p = 2$ or $p = \infty$, and consider the following conditions:

I(p). *the autocovariance kernels satisfy $\sum_{t\in\mathbb{Z}} \|r_t\|_p < \infty$,*

II. *the autocovariance operators satisfy $\sum_{t\in\mathbb{Z}} |||\mathscr{R}_t|||_1 < \infty$,*

*where $|||\mathscr{R}_t|||_1$ is the nuclear norm. Then, under **I(p)**, for any $\omega \in \mathbb{R}$, the following series converges in $\|\cdot\|_p$:*

$$f_\omega(\cdot,\cdot) = \frac{1}{2\pi} \sum_{t\in\mathbb{Z}} \exp(-\mathbf{i}\omega t) r_t(\cdot,\cdot). \tag{3.2.4}$$

We call the limiting kernel f_ω the spectral density kernel *at frequency ω. It is uniformly bounded and also uniformly continuous in ω with respect to $\|\cdot\|_p$, i.e. given any $\varepsilon > 0$, there exists a $\delta > 0$ such that*

$$\forall \omega_1, \omega_2, \quad |\omega_1 - \omega_2| < \delta \implies \left\| f_{\omega_1} - f_{\omega_2} \right\|_p < \varepsilon.$$

The spectral density operator $\mathscr{F}_\omega$, the operator induced by the spectral density kernel through right-integration, is self-adjoint and non-negative definite for all $\omega \in \mathbb{R}$. Furthermore, $\omega \mapsto f_\omega$ is 2π-periodic, and the following inversion formula holds in $\|\cdot\|_p$:

$$\int_0^{2\pi} f_\alpha(\tau,\sigma) e^{\mathbf{i}t\alpha} \mathrm{d}\alpha = r_t(\tau,\sigma), \quad t \in \mathbb{Z}; \tau,\sigma \in [0,1]. \tag{3.2.5}$$

*Under only **II**, we have*

$$\mathscr{F}_\omega = \sum_{t\in\mathbb{Z}} e^{-\mathbf{i}\omega t} \mathscr{R}_t, \tag{3.2.6}$$

where the convergence is uniform in nuclear norm. In particular, the spectral density operators are nuclear, $\omega \mapsto \mathscr{F}_\omega$ is uniformly continuous (with respect to $|||\cdot|||_1$), and $|||\mathscr{F}_\omega|||_1 \leq \sum_t |||\mathscr{R}_t|||_1 < \infty$.

Proof. The convergence of (3.2.6) in $\|\cdot\|_p$ and the uniform boundedness of the spectral density kernel follows from the triangle inequality. Using the property $r_t(\tau,\sigma) = [r_{-t}(\sigma,\tau)]^\mathsf{T}$, we obtain $f_\omega(\tau,\sigma) = \left[f_\omega(\sigma,\tau)\right]^\dagger$, so that the spectral density operator $\mathscr{F}_\omega$ is self-adjoint. For uniform continuity, notice that

$$\begin{aligned}\left\|f_{\omega_1} - f_{\omega_2}\right\|_p &\leq \sum_{t\in\mathbb{Z}} \left|e^{-\mathbf{i}t\omega_1} - e^{-\mathbf{i}t\omega_2}\right| \|r_t\|_p \\ &\leq C \sum_{|t|\leq N} \left|e^{-\mathbf{i}t\omega_1} - e^{-\mathbf{i}t\omega_2}\right| + 2\sum_{|t|>N} \|r_t\|_p, \end{aligned} \tag{3.2.7}$$

where $C = \max_{t\in\mathbb{Z}} \|r_t\|_p$. Fixing $\varepsilon > 0$, since I(p) holds, we can choose $N = N(\varepsilon) > 0$ such that the right hand summand of (3.2.7) is smaller than $\varepsilon/2$. Now since for each t, the function $\omega \mapsto e^{-\mathbf{i}t\omega}$ is uniformly continuous, we can choose a $\delta = \delta(N,\varepsilon) > 0$ such that the left hand summand of (3.2.7) is smaller than $\varepsilon/2$. Since $\delta = \delta(N(\varepsilon),\varepsilon) = \delta(\varepsilon)$, uniform continuity follows. The non-negativity and the inversion formula follow from Proposition 2.3.5.
If **II** holds, then $I(2)$ holds and the spectral density kernel is defined in an L^2 sense. Equation (3.2.6) then follows from triangle inequality. The uniform convergence and uniform continuity follow from the triangle inequality and an argument similar to (3.2.7). □

3.3 The Functional Discrete Fourier Transform

Recall the Cramér representation, which tells us that $X_t = \int_{-\pi}^{\pi} e^{\mathbf{i}\omega t}\, dZ_\omega$. We wish to estimate $\mathscr{F}_\omega$, which is the covariance operator of the infinitesimal increment dZ_ω. This motivates the following definition.

Definition 3.3.1.
Based on a finite stretch $X_0,\ldots,X_{T-1}$ of the FTS X_t, we define the functional Discrete Fourier Transform (fDFT) *of $\{X_t\}_{t=0}^{T-1}$,*

$$\widetilde{X}_\omega^{(T)}(\tau) = (2\pi T)^{-1/2} \sum_{t=0}^{T-1} X_t(\tau)\exp(-\mathbf{i}\omega t).$$

which is the functional version of the discrete Fourier transform (DFT).

Intuitively, the fDFT estimates the increment dZ, and we might hope that its covariance will be close to the spectral density operators. It turns out that this is true asymptotically, under some weak dependence conditions, as we shall see later on.
Notice that the construction of the fDFT does not require the representation of the data in a particular basis. The fDFT transforms the T functional

observations to a mapping from $\mathbb{R}$ into $L^2([0,1],\mathbb{C})$. It straightforwardly inherits some basic analytical properties that its finite dimensional counterpart satisfies; for example, it is 2π-periodic and Hermitian with respect to ω, and linear with respect to the series $\{X_t\}$.
Since the fDFT will be the building block of our estimation procedure, we will give a few results concerning its various moments. For this purpose, we need to introduce the notion of a higher order spectral density, the *cumulant spectral density of order k*, defined in L^2:

$$f_{\omega_1,\ldots,\omega_{k-1}}(\tau_1,\ldots,\tau_k)=\frac{1}{(2\pi)^{k-1}}\sum_{t_1,\ldots,t_{k-1}=-\infty}^{\infty}\exp\Big(-\mathbf{i}\sum_{j=1}^{k-1}\omega_j t_j\Big)\operatorname{cum}\big(X_{t_1}(\tau_1),\ldots,X_{t_{k-1}}(\tau_{k-1}),X_0(\tau_k)\big).$$

In shorthand, we will write

$$f_{\omega_1,\ldots,\omega_{k-1}}=\frac{1}{(2\pi)^{k-1}}\sum_{t_1,\ldots,t_{k-1}=-\infty}^{\infty}\exp\Big(-\mathbf{i}\sum_{j=1}^{k-1}\omega_j t_j\Big)\operatorname{cum}\big(X_{t_1},\ldots,X_{t_{k-1}},X_0\big).$$

This density is well defined and is bounded under **C(0, k)**. In fact, the convergence of the series defining the higher order density can be described explicitly as follows:

Lemma 3.3.2.
We have

$$f_{\omega_1,\ldots,\omega_{k-1}}=\frac{1}{(2\pi)^{k-1}}\sum_{t_1,\ldots,t_{k-1}=-(T-1)}^{T-1}\exp\Big(-\mathbf{i}\sum_{j=1}^{k-1}\omega_j t_j\Big)\operatorname{cum}\big(X_{t_1},\ldots,X_{t_{k-1}},X_0\big)+\varepsilon_T,$$

where the equality is in L^2. The error term is uniform in ω, and satisfies $\varepsilon_T\sim o(1)$ as $T\to\infty$ under **C(0,k)**. *Under the stronger condition* **C(1,k)** *we have $\varepsilon_T\sim o(T^{-1})$, as $T\to\infty$.*

Proof. Direct consideration of the expression for $f_{\omega_1,\ldots,\omega_{k-1}}$ yields that

$$|\varepsilon_T|\le\frac{1}{(2\pi)^{k-1}}\sum_{\nu=1}^{k-1}\left[\sum_{|t_\nu|\ge T}\sum_{\substack{u\neq\nu\\ t_u\in\mathbb{Z}}}\big\|\operatorname{cum}\big(X_{t_1},\ldots,X_{t_{k-1}},X_0\big)\big\|_2\right],$$

and the estimates of the error follow directly. In particular, the error is independent of the ω's. □

With regards to higher order moments, we may establish an asymptotic representation of the cumulant kernel of the functional discrete Fourier transform in terms of the cumulant spectral density of the same order:

Theorem 3.3.3.

Let $\Delta^{(T)}(\omega) = \sum_{t=0}^{T-1} \exp(-\mathbf{i}\omega t)$. *We have*

$$T^{k/2}\operatorname{cum}\left(\widetilde{X}_{\omega_1}^{(T)}(\tau_1),\ldots,\widetilde{X}_{\omega_k}^{(T)}(\tau_k)\right) = (2\pi)^{k/2-1}\Delta^{(T)}(\omega_1+\cdots+\omega_k) f_{\omega_1,\ldots,\omega_{k-1}}(\tau_1,\ldots,\tau_k) + \varepsilon_T, \quad \text{in } L^2,$$

where the error term is uniform in ω. *In particular,* $\varepsilon_T \sim o(T)$ *under* **C(0,k)** *and* $\varepsilon_T \sim O(1)$ *under* **C(1,k)**. *Notice that in the case* $k = 1$, $f(\tau) = \mu(\tau)$.

Proof. We have

$$\operatorname{cum}\left(\widetilde{X}_{\omega_1}^{(T)}(\tau_1),\ldots,\widetilde{X}_{\omega_k}^{(T)}(\tau_k)\right) = (2\pi T)^{-k/2}\sum_{t_1,\ldots,t_k=0}^{T-1} e^{-\mathbf{i}\sum_{j=1}^{k-1}(t_j-t_k)\omega_j} e^{-\mathbf{i}t_k(\omega_1+\cdots+\omega_k)} \times \operatorname{cum}\left(X_{t_1-t_k}(\tau_1),\ldots,X_{t_{k-1}-t_k},X_0(\tau_k)\right)$$

With the change of variables $t = t_k, u_j = t_j - t_k$ for $j = 1,\ldots,k-1$, and defining $h^{(T)}(t) = 1$ if $0 \le t \le T$, and 0 otherwise, we can re-express the last expression as

$$\begin{aligned}
&(2\pi T)^{k/2}\operatorname{cum}\left(\widetilde{X}_{\omega_1}^{(T)}(\tau_1),\ldots,\widetilde{X}_{\omega_k}^{(T)}(\tau_k)\right)\\
&= \sum_{u_1,\ldots,u_{k-1}=-(T-1)}^{T-1} e^{-\mathbf{i}\sum_{j=1}^{k-1}u_j\omega_j}\operatorname{cum}\left(X_{u_1}(\tau_1),\ldots,X_{u_{k-1}},X_0(\tau_k)\right)\\
&\qquad\times\sum_{t\in\mathbb{Z}} h^{(T)}(u_1+t)\cdots h^{(T)}(u_{k-1}+t)h^{(T)}(t)e^{-\mathbf{i}t(\omega_1+\cdots+\omega_k)}\\
&= \sum_{u_1,\ldots,u_{k-1}=-(T-1)}^{T-1} e^{-\mathbf{i}\sum_{j=1}^{k-1}u_j\omega_j}\Delta^{(T)}\left(\sum_{j=1}^{k}\omega_j\right)\operatorname{cum}\left(X_{u_1}(\tau_1),\ldots,X_{u_{k-1}},X_0(\tau_k)\right)+\varepsilon_{1,T},
\end{aligned}$$

where $\Delta^{(T)}(\omega) = \sum_{t=0}^{T-1} e^{-\mathbf{i}\omega t}$. Now, $\varepsilon_{1,T}$ is an error term that we can bound using Lemma 3.12.16:

$$\left\|\varepsilon_{1,T}\right\| \le 2\sum_{u_1,\ldots,u_{k-1}=-(T-1)}^{T-1}(|u_1|+\cdots+|u_{k-1}|)\left\|\operatorname{cum}\left(X_{u_1},\ldots,X_{u_{k-1}},X_0\right)\right\|.$$

Using the dominated convergence theorem, we find that $\varepsilon_{1,T} \sim o(T)$ under the first mixing condition, and $\varepsilon_{1,T} \sim O(1)$ under the second one, in both cases independently of ω. For the rest of the proof, we shall omit the τ_j's. Using Lemma 3.3.2, we have

$$\begin{aligned}
T^{k/2}\operatorname{cum}\left(\widetilde{X}_{\omega_1}^{(T)},\ldots,\widetilde{X}_{\omega_k}^{(T)}\right) &= (2\pi)^{k/2-1}\Delta^{(T)}\left(\sum_{j=1}^{k}\omega_j\right) f_{\omega_1,\ldots,\omega_{k-1}}\\
&\quad+(2\pi)^{k/2-1}\Delta^{(T)}\left(\sum_{j=1}^{k}\omega_j\right)\varepsilon_{2,T}+(2\pi)^{-k/2}\varepsilon_{1,T},
\end{aligned}$$

where $\varepsilon_{2,T}$ is the error term of Lemma 3.3.2. Since $\Delta^{(T)}$ is $O(T)$, we obtain that the global error term is $o(T)$ (resp. $O(1)$) under the first (resp. second) cumulant mixing condition. □

By assuming summability of higher-order moments, we can show that the fDFT is asymptotically Gaussian, with independent coordinates.

Theorem 3.3.4 (Asymptotic Distribution of the fDFT)**.**
Let $\{X_t\}_{t=0}^{T-1}$ be a strictly stationary sequence of random elements of $L^2([0,1],\mathbb{R})$, of length T. Assume the following conditions hold:

(i) $\mathbb{E}\|X_0\|_2^k < \infty$,

(ii) **C(0,k)** *holds for all $k \geq 2$,*

(iii) $\sum_{t\in\mathbb{Z}} |\mathrm{Tr}(\mathscr{R}_t)| < \infty$,

(iv) $\omega \mapsto \mathrm{Tr}(\mathscr{F}_\omega)$ *is continuous.*

Then, for $\omega_{1,T} := \omega_1 = 0$, $\omega_{2,T} := \omega_2 = \pi$, and distinct *integers*

$$s_{3,T}, \ldots, s_{J,T} \in \{1, \ldots, \lfloor (T-1)/2 \rfloor\}$$

such that

$$\omega_{j,T} := \frac{2\pi s_{j,T}}{T} \overset{T\to\infty}{\longrightarrow} \omega_j, \quad j = 3, \ldots J,$$

we have

$$\widetilde{X}_{\omega_1}^{(T)} - \sqrt{\frac{T}{2\pi}}\mu \xrightarrow{d} \widetilde{X}_{\omega_1}, \quad \text{as } T \to \infty, \tag{3.3.1}$$

and $\widetilde{X}_{\omega_{j,T}}^{(T)} \xrightarrow{d} \widetilde{X}_{\omega_j}$, as $T \to \infty$, $j = 2, \ldots, J$ where $\{\widetilde{X}_{\omega_j}\}$ are independent mean zero Gaussian elements of $L^2([0,1],\mathbb{R})$ for $j = 1, 2$, and of $L^2([0,1],\mathbb{C})$ for $j = 3, \ldots, J$. Their covariance operators are given by

$$\mathbb{E}\left[\left(\widetilde{X}_{\omega_j} - \mu \mathbf{1}_{\{j=1\}}\right) \otimes_2 \overline{\left(\widetilde{X}_{\omega_j} - \mu \mathbf{1}_{\{j=1\}}\right)}\right] = \mathscr{F}_{\omega_j},$$

and for $j = 3, \ldots, J$, their relation operator vanishes:

$$\mathbb{E}\left[\widetilde{X}_{\omega_j} \otimes_2 \overline{\widetilde{X}_{\omega_j}}\right] = 0, \quad j = 3, \ldots, J.$$

Remark 3.3.5. *Though the $\{\omega_{j,T}\}_{j=3}^{J}$ are distinct for every T, the limiting frequencies $\{\omega_j : j = 3, ..., J\}$ need not be distinct.*

Condition (ii) for $k \geq 3$ is the generalisation of the standard multivariate cumulant condition to the functional case (Brillinger 2001, Condition 2.6.1), and reduces to that condition if the data are finite-dimensional. Conditions (ii) for $k = 2$ and (iii) imply that the spectral density operator is a nuclear operator at each ω, see Proposition 2.3.5. With Condition (iv), these will imply the tightness of $\widetilde{X}_\omega^{(T)}$. A sufficient condition for (iii) and (iv) is $\sum_{t\in\mathbb{Z}} |||\mathscr{R}_t|||_1 < \infty$.

Proof of Theorem 3.3.4. Define $p_\omega^{(T)}(\tau,\sigma) = \widetilde{X}_\omega^{(T)}(\tau)\widetilde{X}_{-\omega}^{(T)}(\sigma)$, and assume initially that $\mu = 0$. We will treat the case $\mu \neq 0$ at the end of the proof. First we show that for any sequence ω_T, the sequence of random elements $\widetilde{X}_{\omega_T}^{(T)}, T = 1,2,\ldots$ is tight. To do this, we shall use Lemma C.2.3. Fix an orthonormal basis $\{\varphi_n\}_{n\geq 1}$ of $L^2([0,1],\mathbb{R})$. We notice that $p_{\omega_T}^{(T)}$ is a random element of the Hilbert space $L^2\left([0,1]^2,\mathbb{C}\right)$, with $\mathbb{E}\left\|p_{\omega_T}^{(T)}\right\|_2 < \infty$, and that the projection $\mathscr{P}_n : L^2\left([0,1]^2,\mathbb{R}\right) \to \mathbb{C}$ defined by $\mathscr{P}_n(h) = \left\langle h\varphi_n, \varphi_n\right\rangle$ is continuous and linear. Therefore,

$$\begin{aligned}
\mathbb{E}\sum_{n\geq N}\left|\left\langle \widetilde{X}_{\omega_T}^{(T)}, \varphi_n\right\rangle\right|^2 &= \sum_{n\geq N}\mathbb{E}\mathscr{P}_n p_{\omega_T}^{(T)} \\
&= \sum_{n\geq N}\mathscr{P}_n\,\mathbb{E}\,p_{\omega_T}^{(T)} \\
&= \sum_{n\geq N}\mathscr{P}_n[(F_T * \mathscr{F}_\cdot)(\omega_T)] \\
&= \sum_{n\geq N}(F_T * [\mathscr{P}_n\mathscr{F}_\cdot](\omega_T)) \\
&= (F_T * [\sum_{n\geq N}\mathscr{P}_n\mathscr{F}_\cdot](\omega_T)) \\
&\leq \sup_{\alpha\in[-\pi,\pi]}\sum_{n\geq N}\left\langle \mathscr{F}_\alpha\varphi_n,\varphi_n\right\rangle
\end{aligned}$$

where F_T is the Fejér kernel,

$$F_T(\omega) = \frac{1}{2\pi T}\left(\frac{\sin(T\omega/2)}{\sin(\omega/2)}\right)^2.$$

The third equality comes from Proposition 3.4.3 (which is independent of previous results), the fourth equality follows from the continuity of the convolution (see Section B.0.4.3), the fifth equality comes from the monotone convergence Theorem, and the last inequality follows from Young's inequality (Hunter & Nachtergaele 2001, Theorem 12.58). Therefore,

$$\begin{aligned}
\sup_{T\geq 1}\mathbb{E}\left\|\widetilde{X}_\omega^{(T)}\right\|^2 &\leq \sup_{\alpha\in[-\pi,\pi]}\sum_{n\geq 1}\left\langle \mathscr{F}_\alpha\varphi_n,\varphi_n\right\rangle \\
&= \sup_{\alpha\in[-\pi,\pi]}\operatorname{Tr}(\mathscr{F}_\alpha) \\
&\leq \sum_{t\in\mathbb{Z}}|\operatorname{Tr}(\mathscr{R}_t)| \\
&< \infty.
\end{aligned}$$

Furthermore, since the convergence of $\sum_{n\geq 1}\left\langle \mathscr{F}_\alpha\varphi_n,\varphi_n\right\rangle$ is uniform in α, by Proposition 3.2.3, for all $\varepsilon > 0$, there is an $N' > 0$ such that for all $N > N'$,

$$\sum_{n\geq N}\left\langle \mathscr{F}_\alpha\varphi_n,\varphi_n\right\rangle < \varepsilon, \quad \alpha\in[-\pi,\pi].$$

Therefore,

$$\lim_{N\to\infty}\sup_{T\geq 1}\mathbb{E}\sum_{n\geq N}\left|\left\langle\widetilde{X}_{\omega}^{(T)},\varphi_n\right\rangle\right|^2\leq\lim_{N\to\infty}\sup_{\alpha\in[-\pi,\pi]}\sum_{n\geq N}\left\langle\mathscr{F}_{\alpha}\varphi_n,\varphi_n\right\rangle=0,$$

and $\widetilde{X}_{\omega}^{T}$ is tight by Lemma C.2.3. Consequently, the random element

$$\left(\widetilde{X}_{\omega_{1,T}}^{(T)},\ldots,\widetilde{X}_{\omega_{J,T}}^{(T)}\right)\in\left(L^2([0,1],\mathbb{C})\right)^J$$

is also tight. Its asymptotic distribution is therefore determined by the convergence of its finite dimensional distributions (see e.g. Ledoux & Talagrand (2011, Par. 2.1)). Thus, to complete the proof, it suffices to show that for any $\psi_1,\ldots,\psi_J\in L^2([0,1],\mathbb{C})$,

$$\left(\left\langle\widetilde{X}_{\omega_{1,T}}^{(T)},\psi_1\right\rangle,\ldots,\left\langle\widetilde{X}_{\omega_{J,T}}^{(T)},\psi_J\right\rangle\right)\xrightarrow{d}\left(\left\langle\widetilde{X}_{\omega_1},\psi_1\right\rangle,\ldots,\left\langle\widetilde{X}_{\omega_J},\psi_J\right\rangle\right),\quad(3.3.2)$$

where $\widetilde{X}_{\omega_j}\sim\mathcal{N}_H\left(0,\mathscr{F}_{\omega_j}\right)$ are independent random elements of H, where $H=L^2([0,1],\mathbb{R})$ if $j=1,2$ and $H=L^2([0,1],\mathbb{C})$ if $j=3,\ldots,J$. . This is a consequence of the following claim, which is justified by Brillinger (2001, Theorem 4.4.1):

$\mathcal{N}_H(\cdot)$ is defined in Section 3.12.1 on page 134

(I) For $j=1,\ldots,J$, let $\psi_j=\varphi_{2j-1}+\mathbf{i}\varphi_{2j}$, where $\varphi_1,\ldots,\varphi_{2J}\in L^2([0,1],\mathbb{R})$, and let

$$\mathbf{Y}_t=(Y_t(1),\ldots,Y_t(2J))\in\mathbb{R}^{2J}$$

be the vector time series with coordinates $Y_t(l)=\left\langle X_t,\varphi_l\right\rangle$. Then $\widetilde{\mathbf{Y}}_{\omega_{j,T}}^{(T)}\xrightarrow{d}\widetilde{\mathbf{Y}}_{\omega_j}$, where $\left\{\widetilde{\mathbf{Y}}_{\omega_j}\right\}$ are independent mean zero complex Gaussian random vectors with covariance matrix $\mathbf{F}_{\omega_j}$, $(\mathbf{F}_{\omega_j})_{sl}:=F_{\omega_j}(s,l)=\left\langle\mathscr{F}_{\omega_j}\varphi_l,\varphi_s\right\rangle$.

For the case $\mu\neq 0$, we only need to consider $j=1,2$, since $\widetilde{(X-\mu)}_{\omega_{j,T}}^{(T)}=\widetilde{X}_{\omega_{j,T}}^{(T)}$ for $j=3,\ldots,J$. We need to show that

$$\widetilde{X}_{\omega_1}^{(T)}-\sqrt{\frac{T}{2\pi}}\mu=(2\pi T)^{-1/2}\sum_{t=0}^{T-1}(X_t-\mu)\xrightarrow{d}\widetilde{X}_0,\quad(3.3.3)$$

and also that

$$\widetilde{X}_{\omega_2}^{(T)}=(2\pi T)^{-1/2}\sum_{t=0}^{T-1}(-1)^t X_t\xrightarrow{d}\widetilde{X}_{\pi}.\quad(3.3.4)$$

The weak convergence in (3.3.3) follows immediately from the case $\mu=0$. For (3.3.4), notice that

$$\widetilde{X}_{\omega_2}^{(T)}=(2\pi T)^{-1/2}\sum_{t=0}^{T-1}(-1)^t(X_t-\mu)+\mu(2\pi T)^{-1/2}\sum_{t=0}^{T-1}(-1)^t.$$

The first summand is the discrete Fourier transform of a zero mean random process, and converges to $\widetilde{X}_{\omega_2}$. The second summand is determin-

istic and bounded by $\|\mu\|(2\pi T)^{-1/2}$, which tends to zero. Finally, the continuous mapping Theorem for metric spaces (Pollard 1984) yields (3.3.4). □

We also remark that the weak convergence relation in equation (3.3.1) can be re-expressed to trivially yield the corollary:

Corollary 3.3.6 (Central Limit Theorem for Cumulant Mixing Functional Series)**.** *Let $\{X_t\}_{t=0}^{T}$ be a strictly stationary sequence of random elements of $L^2([0,1],\mathbb{R})$ of length T satisfying conditions (i) and (ii) of Theorem 3.3.4. Then,*

$$\sqrt{T}\left(\frac{1}{T}\sum_{t=0}^{T-1} X_t(\tau) - \mu(\tau)\right) \xrightarrow{d} \mathcal{N}_{L^2([0,1],\mathbb{R})}\left(0, \sum_{t\in\mathbb{Z}} \mathscr{R}_t\right)$$

This provides one of the first instances of central limit theorems for functional series under no structural modeling assumptions beyond weak dependence. To our knowledge, the only other similar result is given in recent work by Horváth, Kokoszka & Reeder (2013), who obtain the same limit under different mixing conditions, namely L^p-m-approximability.

3.4 The Periodogram Kernel and its Properties

Since the asymptotic covariance operator of the fDFT are equal to the spectral density operators, it is natural to study properties of the empirical covariance of the fDFT, which we call the *periodogram*:

Definition 3.4.1.
The periodogram kernel *at ω is the random element of $L^2([0,1]^2,\mathbb{C})$ defined by*

$$p_\omega^{(T)}(\tau,\sigma) = \left[\widetilde{X}_\omega^{(T)}(\tau)\right]\left[\widetilde{X}_\omega^{(T)}(\sigma)\right]^\dagger = \widetilde{X}_\omega^{(T)}(\tau)\widetilde{X}_{-\omega}^{(T)}(\sigma), \quad \text{in } L^2.$$

The operator on $L^2([0,1],\mathbb{C})$ induced by this kernel through right integration will be called the periodogram operator*, and denoted $P_\omega^{(T)}$.*

We have $\left\|p_\omega^{(T)}\right\|_2 = \left\|\widetilde{X}_\omega^{(T)}\right\|_2^2$, and hence $\mathbb{E}\left\|p_\omega^{(T)}\right\|_2^l < \infty, \quad l \geq 1$. The expectation of the periodogram kernel is thus well defined, and we have

Lemma 3.4.2.
$\mathbb{E}\, p_\omega^{(T)} = T^{-1}(a_0 + \cdots + a_{T-1})$, *where* $a_T(\tau,\sigma) = \sum_{t=-T}^{T} e^{-\mathbf{i}\omega t} r_t(\tau,\sigma)$.

Proof. Make the change of variables $u = t - s$, and observe that, by stationarity of X_t,

$$\mathbb{E}\, p_\omega^{(T)}(\tau,\sigma) = T^{-1}\sum_{u=-(T-1)}^{T-1}(T-|u|)e^{-\mathbf{i}\omega u} r_u(\tau,\sigma),$$

which then yields the result. □

That is, the expectation of the periodogram kernel is a Cesàro-sum of the partial sums of the series defining the spectral density kernel. Therefore, in order to probe the properties of the periodogram kernel, we can make use of the Fejér kernel

$$F_T(\omega) = \frac{1}{2\pi T}\left(\frac{\sin(T\omega/2)}{\sin(\omega/2)}\right)^2 = \frac{1}{2\pi T}|\Delta^{(T)}(\omega)|^2.$$

It will thus be useful to recall some properties of F_T: $\int_{-\pi}^{\pi} F_T = 2\pi, F_T(0) = T, F_T(\omega) \sim O(T)$ uniformly in ω, and $F_T(2\pi s/T) = 0$ for s an integer with $s \not\equiv 0 \mod T$. This last property will be used often.
The next proposition gives the mean of the periodogram kernel:

Proposition 3.4.3.
*Assuming that **C(0,2)** holds true, we have, for each $\omega \in \mathbb{R}$,*

$$\mathbb{E}\left[p_\omega^{(T)}(\tau,\sigma)\right] = \frac{1}{2\pi}\int_{-\pi}^{\pi} F_T(\omega-\alpha) f_\alpha(\tau,\sigma)\,d\alpha + \mu(\tau)\mu(\sigma)F_T(\omega), \quad \text{in } L^2.$$

In particular, if $\omega = 2\pi s/T$, with s an integer such that $s \not\equiv 0 \mod T$,

$$\mathbb{E}\left[p_\omega^{(T)}(\tau,\sigma)\right] = \frac{1}{2\pi}\int_{-\pi}^{\pi} F_T(\omega-\alpha) f_\alpha(\tau,\sigma)\,d\alpha, \quad \text{in } L^2.$$

Proof. Using the definition of the periodogram operator,

$$\begin{aligned}
\mathbb{E}\, p_\omega^{(T)}(\tau,\sigma) &= \frac{1}{2\pi}\mathrm{cov}\left(\widetilde{X}_\omega^{(T)}, \widetilde{X}_\omega^{(T)}\right) + \frac{1}{2\pi}\,\mathbb{E}\left[\widetilde{X}_\omega^{(T)}(\tau)\right]\mathbb{E}\left[\widetilde{X}_{-\omega}^{(T)}(\sigma)\right] \\
&= \frac{1}{2\pi T}\sum_{s,t=0}^{T-1} e^{-\mathbf{i}\omega(t-s)}\mathrm{cov}\left(X_t(\tau), X_s(\sigma)\right) + \frac{1}{2\pi T}\mu(\tau)\mu(\sigma)|\Delta^{(T)}(\omega)|^2 \\
&= \frac{1}{2\pi T}\sum_{s,t=0}^{T-1} e^{-\mathbf{i}\omega(t-s)} r_{t-s}(\tau,\sigma) + \mu(\tau)\mu(\sigma)F_T(\omega).
\end{aligned}$$

Using the inversion formula of Proposition 3.2.4, we obtain

$$\begin{aligned}
\mathbb{E}\, p_\omega^{(T)}(\tau,\sigma) &= \frac{1}{2\pi T}\sum_{s,t=0}^{T-1} e^{-\mathbf{i}\omega(t-s)}\int_{-\pi}^{\pi} e^{\mathbf{i}\alpha(t-s)} f_\alpha(\tau,\sigma)\,d\alpha + \mu(\tau)\mu(\sigma)F_T(\omega) \\
&= \frac{1}{2\pi T}\int_{-\pi}^{\pi} \underbrace{\left[\sum_{s,t=0}^{T-1} e^{-\mathbf{i}(t-s)(\omega-\alpha)}\right]}_{=|\Delta^{(T)}(\omega-\alpha)|^2} f_\alpha(\tau,\sigma)\,d\alpha + \mu(\tau)\mu(\sigma)F_T(\omega) \\
&= \frac{1}{2\pi}\int_{-\pi}^{\pi} F_T(\omega-\alpha) f_\alpha(\tau,\sigma)\,d\alpha + \mu(\tau)\mu(\sigma)F_T(\omega).
\end{aligned}$$

For the case $\omega = 2\pi s/T$, with s an integer with $s \not\equiv 0 \mod T$, the result follows from the fact that $F_T(\omega) = 0$. □

In particular, the periodogram kernel is asymptotically unbiased:

Proposition 3.4.4.
Let s be an integer with $s \not\equiv 0 \mod T$. Then, we have

$$\mathbb{E}\left[p^{(T)}_{2\pi s/T}(\tau,\sigma)\right] = f_{2\pi s/T}(\tau,\sigma) + \varepsilon_T, \quad \text{in } L^2.$$

*The error term ε_T is $o(1)$ under **C(0,2)** and $O(T^{-1})$ under **C(1,2)**. In either case, the error term is uniform in integers $s \not\equiv 0 \mod T$.*

Proof. Since $s \not\equiv 0 \mod T$,

$$\mathbb{E}\left[p^{(T)}_{2\pi s/T}(\tau,\sigma)\right] = \operatorname{cum}\left(\widetilde{X}_{2\pi s/T}(\tau), \widetilde{X}_{-2\pi s/T}(\sigma)\right) = f_{2\pi s/T}(\tau,\sigma) + \varepsilon_t,$$

and the result follows from Theorem 3.3.3. □

Notice that this result can also be proven using the fact that the Fejér kernel is an approximate identity, see Proposition B.0.15. Having established the mean structure of the periodogram, we turn to the determination of its covariance structure.

Theorem 3.4.5.
Assume ω_1 and ω_2 are of the form $2\pi s(T)/T$, where $s(T)$ is an integer, $s(T) \not\equiv 0 \mod T$. We have

$$\begin{aligned}\operatorname{cov}\left(p^{(T)}_{\omega_1}(\tau_1,\sigma_1), p^{(T)}_{\omega_2}(\tau_2,\sigma_2)\right) = \eta(\omega_1-\omega_2) f_{\omega_1}(\tau_1,\tau_2) f_{-\omega_1}(\sigma_1,\sigma_2) + \\ + \eta(\omega_1+\omega_2) f_{\omega_1}(\tau_1,\sigma_2) f_{-\omega_1}(\sigma_1,\tau_2) + \varepsilon_T, \quad \text{in } L^2,\end{aligned}$$

*where the function $\eta(x)$ equals one if $x \in 2\pi\mathbb{Z}$, and zero otherwise. The error term ε_T is $o(1)$ under **C(0,2)** and **C(0,4)**; $\varepsilon_T \sim O(T^{-1})$ under **C(1,2)** and **C(1,4)**. In each case, the error term is uniform in ω_1, ω_2 of the form $2\pi s(T)/T$ with $s(T) \not\equiv 0 \mod T$.*

Proof. In this proof, in order to compactify notation, we will say that the error term is $\Big|_{O(1)}^{o(T)}$ under assumption **A/B**, meaning that the error term is of order $o(T)$ under condition **A**, and of order $O(1)$ under condition **B**.

Since

$$\begin{aligned}\operatorname{cov}\big(p^{(T)}_{\omega_1}(\tau_1,\sigma_1),p^{(T)}_{\omega_2}(\tau_2,\sigma_2)\big) &= \operatorname{cov}\big(\widetilde{X}^{(T)}_{\omega_1}(\tau_1)\widetilde{X}^{(T)}_{-\omega_1}(\sigma_1),\widetilde{X}^{(T)}_{\omega_2}(\tau_2)\widetilde{X}^{(T)}_{-\omega_2}(\sigma_2)\big)\\ &= \mathbb{E}\Big[\underbrace{\widetilde{X}^{(T)}_{\omega_1}(\tau_1)}_{A}\underbrace{\widetilde{X}^{(T)}_{-\omega_1}(\sigma_1)}_{B}\underbrace{\widetilde{X}^{(T)}_{-\omega_2}(\tau_2)}_{C}\underbrace{\widetilde{X}^{(T)}_{\omega_2}(\sigma_2)}_{D}\Big]\\ &\quad-\mathbb{E}\big[\widetilde{X}^{(T)}_{\omega_1}(\tau_1)\widetilde{X}^{(T)}_{-\omega_1}(\sigma_1)\big]\,\mathbb{E}\big[\widetilde{X}^{(T)}_{-\omega_2}(\tau_2)\widetilde{X}^{(T)}_{\omega_2}(\sigma_2)\big]\\ &= \mathbb{E}[ABCD]-\mathbb{E}[AB]\,\mathbb{E}[CD],\end{aligned}$$

using the notation $(A)=\operatorname{cum}(A)=\mathbb{E}A$, $(A,B)=\operatorname{cum}(A,B)$, $(A,B,C)=\operatorname{cum}(A,B,C)$, and so on, and invoking Lemma 3.12.5, we deduce

$$\begin{aligned}&\operatorname{cov}\big(p^{(T)}_{\omega_1}(\tau_1,\sigma_1),p^{(T)}_{\omega_2}(\tau_2,\sigma_2)\big)=(A,B,C,D)+\\ &\quad+(A)(B,C,D)+(B)(A,C,D)+(C)(A,B,D)+(D)(A,B,C)\\ &\quad+(A,C)(B,D)+(A,D)(B,C)\\ &\quad+(A)(C)(B,D)+(A)(D)(B,C)+(B)(C)(A,D)+(B)(D)(A,C)\\ &=(A,B,C,D)+(A,C)(B,D)+(A,D)(B,C).\end{aligned}\tag{3.4.1}$$

The last equality comes from the fact that $(A)=(B)=(C)=(D)=0$, given that ω_1,ω_2 are of the form $2\pi s(T)/T$, with $s(T)\not\equiv 0 \mod T$. We now approximate each term of (3.4.1) using Theorem 3.3.3:

$$(A,B,C,D)=\frac{2\pi}{T}f_{\omega_1,-\omega_1,-\omega_2}(\tau_1,\sigma_1,\tau_2,\sigma_2)+T^{-2}\varepsilon_T,$$

where ε_T is the error term of Theorem 3.3.3. Thus $(A,B,C,D)=O(T^{-1})$, uniformly in ω under either **C(0,4)** or **C(1,4)**.

For the next term, we have under the assumption **C(1,2)/C(0,2)**:

$$\begin{aligned}(A,C)(B,D)&=T^{-2}\left\{\Delta^{(T)}(\omega_1-\omega_2)f_{\omega_1}(\tau_1,\tau_2)+\Big|^{o(T)}_{O(1)}\right\}\left\{\Delta^{(T)}(\omega_2-\omega_1)f_{-\omega_1}(\sigma_1,\sigma_2)+\Big|^{o(T)}_{O(1)}\right\}\\ &=T^{-2}|\Delta^{(T)}(\omega_1-\omega_2)|^2f_{\omega_1}(\tau_1,\tau_2)f_{-\omega_1}(\sigma_1,\sigma_2)\\ &\quad+T^{-2}\Delta^{(T)}(\omega_1-\omega_2)f_{\omega_1}(\tau_1,\tau_2)\Big|^{o(T)}_{O(1)}\\ &\quad+T^{-2}\Delta^{(T)}(\omega_2-\omega_1)f_{-\omega_1}(\sigma_1,\sigma_2)\Big|^{o(T)}_{O(1)}+\Big|^{o(1)}_{O(T^{-2})}.\end{aligned}$$

Using the fact that $\Delta^{(T)}=O(T), f_\omega=O(1)$ uniformly in ω, we obtain

$$(A,C)(B,D)=\eta(\omega_1-\omega_2)f_{\omega_1}(\tau_1,\tau_2)f_{-\omega_1}(\sigma_1,\sigma_2)+\Big|^{o(1)}_{O(T^{-1})},$$

where the function $\eta(x)$ equals one if $x \in 2\pi\mathbb{Z}$, and zero otherwise. A similar calculation yields

$$(A,D)(B,C) = \eta(\omega_1+\omega_2) f_{\omega_1}(\tau_1,\sigma_2) f_{-\omega_1}(\sigma_1,\tau_2) + \begin{vmatrix} o(1) \\ O(T^{-1}) \end{vmatrix}$$

and in each of these cases, the error term is uniform in ω_1, ω_2 of the form $2\pi s(T)/T$, with $s(T) \not\equiv 0 \mod T$.
Piecing the results back together, we conclude that

$$\begin{aligned} \operatorname{cov}\left(p^{(T)}_{\omega_1}(\tau_1,\sigma_1), p^{(T)}_{\omega_2}(\tau_2,\sigma_2)\right) &= \eta(\omega_1-\omega_2) f_{\omega_1}(\tau_1,\tau_2) f_{-\omega_1}(\sigma_1,\sigma_2) \\ &\quad + \eta(\omega_1+\omega_2) f_{\omega_1}(\tau_1,\sigma_2) f_{-\omega_1}(\sigma_1,\tau_2) + \begin{vmatrix} o(1) \\ O(T^{-1}) \end{vmatrix}, \end{aligned}$$

with the error term being uniform in ω_1, ω_2 of the form $2\pi s(T)/T$, with $s(T) \not\equiv 0 \mod T$. □

3.5 *Smoothing the Periodogram: The Sample Spectral Density Operators*

The results in the previous section show that the asymptotic covariance of the periodogram is not zero, and hence, as for vector time series, the periodogram kernel itself is not a consistent estimator of the spectral density. In this section, we will see that we can nevertheless construct a consistent class of estimators by smoothing the periodogram at neighboring frequencies. The key idea is that although the periodogram is inconsistent (because its asymptotic variance does not vanish), the asymptotic variance of the periodogram operator is continuous in ω (under weak dependence assumptions). Therefore, by taking local averages of the periodogram, we can reduce the variance of the estimator while controlling the increased bias, so that the resulting estimators of the spectral density operators are consistent. To give some intuition, let us explain a bit in more detail this idea for estimating the spectral density operators in the context of *univariate* time series (Einstein 1914, Blackman & Tukey 1959, Brillinger 2001, Bloomfield 2000, Priestley 2001, Percival & Walden 1993, Fan & Yao 2003). The following paragraph is inspired by Fan & Yao (2003). For univariate time series, the periodogram is given by $p_\omega = (2\pi T)^{-1} \left|\sum_{t=0}^{T-1} e^{-\mathbf{i}\omega t} x_t\right|^2$, and is typically computed at the *Fourier frequencies* $\omega_j = 2\pi j/T$, where $j = 1, \ldots, T-1$, both for computational reasons (the Fast-Fourier Transform can be used to compute them with $O(T\log(T))$ complexity ; see, e.g., Cooley & Tukey (1965), Duhamel & Vetterli (1990)), and for statistical reasons (the variance is reduced at Fourier frequencies; see the proof of Theorem 3.4.5). Therefore, estimation of the spectral density operators can be viewed as a non-parametric functional regression problem (Ferraty & Vieu 2006), where one wishes to estimate

the spectral density operators using the scatter-plot

$$\left(\omega_j, p_{\omega_j}\right), \quad j = 1, \ldots, T-1.$$

Several approaches are possible. One can either smooth the periodogram, or the log-periodogram, and for each of these approaches, a smoothing technique, such as kernel/linear/polynomial smoothing, can be chosen. Each of these techniques is based on a *bandwidth* parameter (or a parameter playing an equivalent role), which can be chosen globally (constant in ω) or locally. A global bandwidth is easy to implement, but is inefficient if the spectral density operators are very variable. On the other hand, a local bandwidth will adapt to varying spectral density operators, but the choice of the bandwidth parameter and implementation will be more complicated.
In this section, we choose to construct estimators of the spectral by smoothing the periodogram operators using a kernel with a fixed bandwidth.
Let $W(x)$ be a real function defined on $\mathbb{R}$ such that

1. W is positive, even, and bounded in variation,
2. $W(x) = 0$ if $|x| \geq 1$,
3. $\int_{-\infty}^{\infty} W(x)\mathrm{d}x = 1$,
4. $\int_{-\infty}^{\infty} W(x)^2\mathrm{d}x < \infty$.

The assumption of compact support is not necessary, but will simplify proofs. For a bandwidth $B_T > 0$, write

$$W^{(T)}(x) = \sum_{j\in\mathbb{Z}} \frac{1}{B_T} W\left(\frac{x + 2\pi j}{B_T}\right). \tag{3.5.1}$$

Some properties of W can be found in Lemma 3.12.18. We define the *spectral density estimator* $f_\omega^{(T)}$ of f_ω at frequency ω as the weighted average of the periodogram evaluated at frequencies of the form $\{2\pi s/T\}_{s=1}^{T-1}$, with weight function $W^{(T)}$:

$$f_\omega^{(T)}(\tau,\sigma) = \frac{2\pi}{T}\sum_{s=1}^{T-1} W^{(T)}\left(\omega - \frac{2\pi s}{T}\right) p_{\frac{2\pi s}{T}}^{(T)}(\tau,\sigma).$$

We also denote the operator induced by the kernel $f_\omega^{(T)}$ by $\mathscr{F}_\omega^{(T)}$. For convenience, we might also refer to these objects as the *sample spectral density kernels* and *sample spectral density operators*, respectively.
A consequence of the assumption of compact support worth mentioning is that, in fact, at most $O(TB_T)$ summands of this expression are non-zero. We will show in this section that, under appropriate conditions on the asymptotic behavior of B_T, this estimator retains the property of asymptotic unbiasedness that the periodogram enjoys. We will determine

the behaviour of its asymptotic covariance structure and establish strong consistency in mean square (with respect to the Hilbert–Schmidt norm). Finally, we will determine the asymptotic law of the estimator.
Concerning the mean of the spectral density estimator, we have:

Proposition 3.5.1.
Under ***C(1,2)****, if* $B_T \to 0$ *and* $B_T T \to \infty$ *as* $T \to \infty$*, then*

$$\mathbb{E} f_\omega^{(T)}(\tau,\sigma) = \int_{\mathbb{R}} W(x) f_{\omega - xB_T}(\tau,\sigma)\mathrm{d}x + O(B_T^{-1}T^{-1}),$$

where the equality holds in L^2*, and the error term is uniform in* ω*.*

Proof. We use Proposition 3.4.4 to write

$$\begin{aligned}\mathbb{E} f_\omega^{(T)}(\tau,\sigma) &= \frac{2\pi}{T}\sum_{s=1}^{T-1} W^{(T)}\left(\omega - \frac{2\pi s}{T}\right)\left\{f_{2\pi s}(\tau,\sigma) + O(T^{-1})\right\} \\ &= \frac{2\pi}{T}\sum_{s=1}^{T-1} W^{(T)}\left(\omega - \frac{2\pi s}{T}\right) f_{2\pi s}(\tau,\sigma) + O(T^{-1})\left\{\frac{2\pi}{T}\sum_{s=1}^{T-1} W^{(T)}\left(\omega - \frac{2\pi s}{T}\right)\right\},\end{aligned}$$

where the error term is uniform in s. Using Lemmas 3.12.15, 3.12.17 and 3.12.19, we obtain

$$\frac{2\pi}{T}\sum_{s=1}^{T-1} W^{(T)}\left(\omega - \frac{2\pi s}{T}\right) f_{2\pi s}(\tau,\sigma) = \int_0^{2\pi} W^{(T)}(\omega - \alpha) f_\alpha(\tau,\sigma)\, d\alpha + \varepsilon_T,$$

where $\varepsilon_T \sim O(B_T^{-1}T^{-1})$, uniformly in ω. Using Lemma 3.12.19 again,

$$\frac{2\pi}{T}\sum_{s=1}^{T-1} W^{(T)}\left(\omega - \frac{2\pi s}{T}\right) = O(1)$$

if $B_T T \to \infty$. Combining these facts, we may write

$$\mathbb{E} f_\omega^{(T)}(\tau,\sigma) = \int_0^{2\pi} W^{(T)}(\omega - \alpha) f_\alpha(\tau,\sigma)\, d\alpha + O(B_T^{-1}T^{-1}) + O(T^{-1}).$$

With a change of variables, we obtain

$$\int_0^{2\pi} W^{(T)}(\omega - \alpha) f_\alpha\, d\alpha = \int_{\mathbb{R}} W(x) f_{\omega - xB_T}\, dx.$$

This completes the proof. □

Concerning the covariance of the spectral density estimator, we have:

Theorem 3.5.2. *Under **C(1,2)** and **C(1,4)**,*

$$\begin{aligned}\operatorname{cov}\big(f_{\omega_1}^{(T)}(\tau_1,\sigma_1),f_{\omega_2}^{(T)}(\tau_2,\sigma_2)\big) = \frac{2\pi}{T}\int_{-\pi}^{\pi}\Big\{&W^{(T)}(\omega_1-\alpha)W^{(T)}(\omega_2-\alpha)f_{\alpha}(\tau_1,\tau_2)f_{-\alpha}(\sigma_1,\sigma_2)\\ &+W^{(T)}(\omega_1-\alpha)W^{(T)}(\omega_2+\alpha)f_{\alpha}(\tau_1,\sigma_2)f_{-\alpha}(\sigma_1,\tau_2)\Big\}\mathrm{d}\alpha\\ &+O(B_T^{-2}T^{-2})+O(T^{-1}),\end{aligned}$$

where the equality holds in L^2, and the error terms are uniform in ω.

Proof. Using Theorem 3.4.5, conditions **C(1,2)** and **C(1,4)**, yield

$$\begin{aligned}&\operatorname{cov}\big(f_{\omega_1}^{(T)}(\tau_1,\sigma_1),f_{\omega_2}^{(T)}(\tau_2,\sigma_2)\big)\\ &=(2\pi/T)^2\sum_{s,l=1}^{T-1}W^{(T)}(\omega_1-2\pi s/T)W^{(T)}(\omega_2-2\pi l/T)\times\\ &\quad\times\Big\{\eta(2\pi(s-l)/T)f_{\frac{2\pi s}{T}}(\tau_1,\tau_2)f_{\frac{-2\pi s}{T}}(\sigma_1,\sigma_2)+\\ &\qquad+\eta(2\pi(s+l)/T)f_{\frac{2\pi s}{T}}(\tau_1,\sigma_2)f_{\frac{-2\pi s}{T}}(\sigma_1,\tau_2)+O(T^{-1})\Big\}\\ &=\left[\frac{2\pi}{T}\sum_{s=1}^{T-1}W^{(T)}(\omega_1-2\pi s/T)\right]^2O(T^{-1})\\ &\quad+\left(\frac{2\pi}{T}\right)^2\sum_{s=1}^{T-1}W^{(T)}(\omega_1-2\pi s/T)W^{(T)}(\omega_1-2\pi s/T)f_{\frac{2\pi s}{T}}(\tau_1,\tau_2)f_{\frac{-2\pi s}{T}}(\sigma_1,\sigma_2)\\ &\quad+\left(\frac{2\pi}{T}\right)^2\sum_{s=1}^{T-1}W^{(T)}(\omega_1-2\pi s/T)W^{(T)}(\omega_1+2\pi s/T)f_{\frac{2\pi s}{T}}(\tau_1,\sigma_2)f_{\frac{-2\pi s}{T}}(\sigma_1,\tau_2),\end{aligned}$$

where the error term is uniform in s,l. An application of Lemmas 3.12.15, 3.12.17, 3.12.18 and 3.12.19 now completes the proof. □

Noting that $\|W^{(T)}\|_\infty = O(B_T^{-1})$ and $\|f_\cdot\|_\infty = O(1)$, a direct consequence of the last result is the following approximation of the asymptotic covariance of the spectral density estimator:

Corollary 3.5.3. *Under **C(1,2)** and **C(1,4)**,*

$$\operatorname{cov}\big(f_{\omega_1}^{(T)}(\tau_1,\sigma_1),f_{\omega_2}^{(T)}(\tau_2,\sigma_2)\big) = O(B_T^{-2}T^{-1}),$$

where the equality holds in L^2, uniformly in the ω's.

This bound is not sharp. A better bound is given in the next statement, which, however, is not uniform in ω.

Proposition 3.5.4.

Assume conditions ***C(1,2)***, ***C(1,4)*** *hold, and that* $B_T \to 0$ *and* $B_T T \to \infty$ *as* $T \to \infty$. *Then,*

$$\lim_{T\to\infty} B_T T \operatorname{cov}\left(f_{\omega_1}^{(T)}(\tau_1,\sigma_1), f_{\omega_2}^{(T)}(\tau_2,\sigma_2)\right) = 2\pi \int_{\mathbb{R}} W(\alpha)^2 \mathrm{d}\alpha \times$$
$$\times \left\{\eta(\omega_1-\omega_2) f_{\omega_1}(\tau_1,\tau_2) f_{-\omega_1}(\sigma_1,\sigma_2) + \eta(\omega_1+\omega_2) f_{\omega_1}(\tau_1,\sigma_2) f_{-\omega_1}(\sigma_1,\tau_2)\right\},$$

where the function $\eta(x)$ *equals one if* $x \in 2\pi\mathbb{Z}$, *and zero otherwise. The convergence is in* L^2 *for any fixed* ω_1, ω_2. *If* ω_1, ω_2 *depend on* T, *then the convergence is in* L^2, *provided* $(\omega_1 \pm \omega_2)$ *are at a distance of at least* $2B_T$ *from any multiples of* 2π, *if not exactly equal to a multiple of* 2π.

Proof. Let $d(x,y)$ denote the distance in $\mathbb{R}/2\pi\mathbb{Z}$. We shall abuse notation and omit the (τ,σ)'s, for the sake of clarity. Theorem 3.5.2 yields

$$B_T T \operatorname{cov}\left(f_{\omega_1}^{(T)}, f_{\omega_2}^{(T)}\right) = 2\pi B_T \int_{-\pi}^{\pi} W^{(T)}(\omega_1-\omega_2-\alpha) W^{(T)}(\alpha) f_{\omega_2+\alpha} f_{-(\omega_2+\alpha)} d\alpha \tag{3.5.2}$$

$$+ 2\pi B_T \int_{-\pi}^{\pi} W^{(T)}(\omega_1+\omega_2-\alpha) W^{(T)}(\alpha) f_{-(\omega_2-\alpha)} f_{\omega_2-\alpha} d\alpha \tag{3.5.3}$$

$$+ O(B_T^{-1} T^{-1}) + O(B_T).$$

We have employed a change of variables, the fact that $W^{(T)}$ is even, and the fact that both $W^{(T)}$ and $f.$ are 2π-periodic. Notice that the error terms tend to zero as $B_T \to 0$, $TB_T \to \infty$.
First we show that (3.5.2) tends to

$$\eta(\omega_1-\omega_2) f_{\omega_1}(\tau_1,\tau_2) f_{-\omega_1}(\sigma_1,\sigma_2) 2\pi \int_{\mathbb{R}} W(\alpha)^2 \mathrm{d}\alpha, \tag{3.5.4}$$

in L^2, uniformly in all $\omega_1 = \omega_{1,T}, \omega_2 = \omega_{2,T}$ such that $\omega_{1,T} \equiv \omega_{2,T}$ or $d(\omega_{1,T} - \omega_{2,T}, 0) \geq 2B_T$ for large T. If $d(\omega_1-\omega_2,0) \geq 2B_T$, (3.5.2) is exactly equal to zero. If $\omega_1 \equiv \omega_2$, we claim that (3.5.2) tends to

$$f_{\omega}(\tau_1,\tau_2) f_{-\omega}(\sigma_1,\sigma_2) 2\pi \int_{\mathbb{R}} W(\alpha)^2 d\alpha. \tag{3.5.5}$$

Notice that in this case, (3.5.2) can be written as

$$\int_{-\pi}^{\pi} K_T(\alpha) f_{\omega+\alpha} f_{-(\omega+\alpha)} d\alpha \times \left\{\int_{\mathbb{R}} W(\alpha)^2 d\alpha\right\},$$

where $K_T(\alpha) = \frac{2\pi}{B_T} [W(\alpha/B_T)]^2 \left\{\int_{\mathbb{R}} W(\alpha)^2 d\alpha\right\}^{-1}$ is an approximate identity on $[-\pi,\pi]$ (see Edwards (1967, §3.2)). Since the spectral density kernel is uniformly continuous with respect to $\|\cdot\|_2$ (see Proposition 3.2.4), Lemma 3.12.21 implies that (3.5.2) tends indeed to (3.5.5) uniformly in ω with respect to $\|\cdot\|_2$. Hence (3.5.2) tends to (3.5.4) in $\|\cdot\|_2$, uniformly in ω's

satisfying

$$\omega_{1,T} \equiv \omega_{2,T} \text{ or } d(\omega_{1,T} - \omega_{2,T}, 0) \geq 2B_T \text{ for large } T.$$

Similarly, we may show that (3.5.3) tends to

$$\eta(\omega_1 + \omega_2) f_{\omega_1}(\tau_1, \sigma_2) f_{-\omega_1}(\sigma_1, \tau_2) 2\pi \int_{\mathbb{R}} W(\alpha)^2 \mathrm{d}\alpha,$$

uniformly in ω's if $\omega_{1,T} \equiv -\omega_{2,T}$ or $d(\omega_{1,T} + \omega_{2,T}) \geq 2B_T$ for large T. Piecing these results together, we obtain the desired convergence, provided for each T large enough, either $\omega_{1,T} - \omega_{2,T} \equiv 0$, $\omega_{1,T} + \omega_{2,T} \equiv 0$, or

$$d(\omega_{1,T} - \omega_{2,T}, 0) \geq 2B_T \quad \text{and} \quad d(\omega_{1,T} + \omega_{2,T}, 0) \geq 2B_T.$$

□

Remark 3.5.5. *In practice, functional data are assumed to be smooth in addition to square-integrable. In such cases, one may hope to obtain stronger results, for example with respect to uniform rather than L^2 norms. Indeed, if the conditions* **C(l,k)** *are replaced by the stronger conditions*

$$\textbf{C}'\textbf{(l,k):} \quad \sum_{t_1,\ldots,t_{k-1} \in \mathbb{Z}} (1+|t_j|^l) \left\| \mathrm{cum}\left(X_{t_1}, \ldots, X_{t_{k-1}}, X_0\right) \right\|_\infty < \infty, \quad j = 1, \ldots, k-1,$$

then the results of Lemma 3.3.2, Theorem 3.3.3, Propositions 3.4.3, 3.4.4, Theorem 3.4.5, Proposition 3.5.1, Theorem 3.5.2, Corollary 3.5.3, and Proposition 3.5.4 hold in the supremum norm with respect to τ, σ.

3.6 Consistency and Asymptotic Normality of the Sample Spectral Density Operators

Combining the results on the asymptotic bias and variance of the spectral density operator, we may now derive the consistency in the supremum norm of the Hilbert–Schmidt distance for the sample spectral density operators. Recall that $\mathscr{F}_\omega$ is the integral operator with kernel f_ω, and, similarly let $\mathscr{F}_\omega^{(T)}$ be the operator with kernel $f_\omega^{(T)}$. We have:

Theorem 3.6.1.
Provided assumptions ***C(1,2)*** *and* ***C(1,4)*** *hold, $B_T \to 0$, $B_T^2 T \to \infty$, the spectral density operator estimator $\mathscr{F}_\omega^{(T)}$ is consistent in supremum norm, with respect to the expected squared Hilbert–Schmidt norm, that is,*

$$\sup_{\omega \in [-\pi,\pi]} \mathbb{E} \left| \left| \left| \mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega \right| \right| \right|_2^2 = O(B_T^2) + O(B_T^{-2} T^{-1}),$$

where $|||\cdot|||_2$ is the Hilbert–Schmidt norm.

Proof. Notice that

$$\mathbb{E}\left|\left|\left|\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\right|\right|\right|_2^2 = \mathbb{E}\left|\left|\left|\mathscr{F}_\omega^{(T)} - \mathbb{E}\mathscr{F}_\omega^{(T)}\right|\right|\right|_2^2 + \left|\left|\left|\mathscr{F}_\omega - \mathbb{E}\mathscr{F}_\omega^{(T)}\right|\right|\right|_2^2,$$

which is essentially the usual bias/variance decomposition of the mean square error. Initially, we focus on the variance term. Lemma C.1.3 yields

$$\mathbb{E}\left|\left|\left|\mathscr{F}_\omega^{(T)} - \mathbb{E}\mathscr{F}_\omega^{(T)}\right|\right|\right|_2^2 \mathrm{d}\omega = \iint_{[0,1]^2} \mathrm{var}\left(f_\omega^{(T)}(\tau,\sigma)\right) d\tau\, d\sigma.$$

Corollary 3.5.3 yields

$$\mathbb{E}\left|\left|\left|\mathscr{F}_\omega^{(T)} - \mathbb{E}\mathscr{F}_\omega^{(T)}\right|\right|\right|_2^2 = O(B_T^{-2}T^{-1}),$$

where the error term is uniform in $\omega \in [-\pi,\pi]$. Therefore,

$$\sup_\omega \mathbb{E}\left|\left|\left|\mathscr{F}_\omega^{(T)} - \mathbb{E}\mathscr{F}_\omega^{(T)}\right|\right|\right|_2^2 = O(B_T^{-2}T^{-1}).$$

Turning to the squared bias, Proposition 3.5.1 yields

$$\left|\left|\left|\mathscr{F}_\omega - \mathbb{E}\mathscr{F}_\omega^{(T)}\right|\right|\right|_2^2 \leq 3\left|\left|\left|\left\{\int_{\mathbb{R}} W(x) f_{\omega - xB_T}\mathrm{d}x - f_\omega\right\}\right|\right|\right|_2^2 + O(T^{-2}) + O(B_T^{-2}T^{-2}),$$

where the error terms are uniform in $\omega \in [-\pi,\pi]$. We have used Jensen's inequality here, and $\{\int_{\mathbb{R}} W(x) f_{\omega - xB_T}\mathrm{d}x - f_\omega\}$ denotes the operator with kernel

$$\int_{\mathbb{R}} W(x) f_{\omega - xB_T}(\tau,\sigma)\mathrm{d}x - f_\omega(\tau,\sigma).$$

Lemma 3.12.13 implies that this difference is of order $O(B_T)$, uniformly in ω. Hence,

$$\sup_\omega \left|\left|\left|\mathscr{F}_\omega - \mathbb{E}\mathscr{F}_\omega^{(T)}\right|\right|\right|_2^2 = O(B_T^2) + O(B_T^{-2}T^{-2}).$$

In summary, we have

$$\sup_{\omega\in[-\pi,\pi]} \mathbb{E}\left\|\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\right\|^2 = O(B_T^2) + O(B_T^{-2}T^{-1}).$$

□

This implies of course that $\int_{-\pi}^{\pi} \left|\left|\left|\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\right|\right|\right|_2^2 d\omega \to 0$ at the rate $O(B_T^2) + O(B_T^{-2}T^{-1})$. In fact, the rate under this integrated error criterion can be slightly improved:

Theorem 3.6.2.
Provided assumptions ***C(1,2)*** *and* ***C(1,4)*** *hold, $B_T \to 0$, $B_T T \to \infty$, the spectral density operator estimator $\mathscr{F}_\omega^{(T)}$ is consistent in integrated mean square, that is*

$$IMSE(\mathscr{F}^{(T)}) = \int_{-\pi}^{\pi} \mathbb{E}\left|\left|\left|\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\right|\right|\right|_2^2 \mathrm{d}\omega = O(B_T^2) + O(B_T^{-1}T^{-1}),$$

where $|||\cdot|||_2$ is the Hilbert–Schmidt norm.
We also have pointwise mean square convergence for every fixed $\omega \in [-\pi, \pi]$:

$$\mathbb{E}\left|\left|\left|\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\right|\right|\right|_2^2 = O(B_T^2) + O(B_T^{-1}T^{-1}),$$

as $T \to \infty$.

Proof. For a kernel operator K with a complex-valued kernel $k(\tau,\sigma)$, we will denote by $\overline{K}$ the operator with kernel $\overline{k(\tau,\sigma)}$. Let $|||\cdot|||_2$ be the Hilbert–Schmidt norm. Proposition A.2.11 yields $|||K|||_2 = \left|\left|\left|\overline{K}\right|\right|\right|_2$. Further, notice that $f_{-\omega}(\tau,\sigma) = \overline{f_\omega(\tau,\sigma)}$, hence $\mathscr{F}_{-\omega} = \overline{\mathscr{F}_\omega}$. Similarly, $\mathscr{F}_{-\omega}^{(T)} = \overline{\mathscr{F}_\omega^{(T)}}$. Thus, via a change of variables, the IMSE of the spectral density estimator can be written as

$$\begin{aligned}\int_{-\pi}^{\pi} \mathbb{E}\left|\left|\left|\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\right|\right|\right|_2^2 \mathrm{d}\omega &= 2\int_0^{\pi} \mathbb{E}\left|\left|\left|\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\right|\right|\right|_2^2 \mathrm{d}\omega \\ &= 2\int_0^{\pi} \mathbb{E}\left|\left|\left|\mathscr{F}_\omega^{(T)} - \mathbb{E}\mathscr{F}_\omega^{(T)}\right|\right|\right|_2^2 \mathrm{d}\omega + 2\int_0^{\pi} \left|\left|\left|\mathscr{F}_\omega - \mathbb{E}\mathscr{F}_\omega^{(T)}\right|\right|\right|_2^2 \mathrm{d}\omega,\end{aligned}$$

which is essentially the usual bias/variance decomposition of the mean square error. Initially, we focus on the variance term. Lemma C.1.3 yields

$$\int_0^{\pi} \mathbb{E}\left|\left|\left|\mathscr{F}_\omega^{(T)} - \mathbb{E}\mathscr{F}_\omega^{(T)}\right|\right|\right|_2^2 \mathrm{d}\omega = \int_0^{\pi} \iint_{[0,1]^2} \mathrm{var}\left(f_\omega^{(T)}(\tau,\sigma)\right) d\tau\, d\sigma\, \mathrm{d}\omega.$$

Decomposing the outer integral into three terms,

$$\int_0^{\pi} = \int_0^{B_T} + \int_{B_T}^{\pi - B_T} + \int_{\pi - B_T}^{\pi},$$

we can use Corollary 3.5.3 for the first and last summands, and Proposition 3.5.4 for the second summand to obtain

$$\int_0^{\pi} \mathbb{E}\left|\left|\left|\mathscr{F}_\omega^{(T)} - \mathbb{E}\mathscr{F}_\omega^{(T)}\right|\right|\right|_2^2 \mathrm{d}\omega = O(B_T^{-1}T^{-1}).$$

Turning to the squared bias, Proposition 3.5.1 yields

$$\begin{aligned}&\int_0^{\pi} \left|\left|\left|\mathscr{F}_\omega - \mathbb{E}\mathscr{F}_\omega^{(T)}\right|\right|\right|_2^2 \mathrm{d}\omega \\ &\leq 3\int_0^{\pi} \left|\left|\left|\left\{\int_{\mathbb{R}} W(x) f_{\omega - xB_T}\,\mathrm{d}x - f_\omega\right\}\right|\right|\right|_2^2 \mathrm{d}\omega + O(T^{-2}) + O(B_T^{-2}T^{-2}),\end{aligned}$$

where we have used Jensen's inequality and where

$$\left\{\int_{\mathbb{R}} W(x) f_{\omega - xB_T}\,\mathrm{d}x - f_\omega\right\}$$

denotes the operator with kernel $\int_{\mathbb{R}} W(x) f_{\omega - xB_T}(\tau,\sigma)\mathrm{d}x - f_\omega(\tau,\sigma)$. Lemma 3.12.13 implies that this difference is of order $O(B_T)$, uniformly in ω. Hence,

$$3\int_0^\pi \left|\!\left|\!\left| \left\{ \int_{\mathbb{R}} W(x) f_{\omega - xB_T}\mathrm{d}x - f_\omega \right\} \right|\!\right|\!\right|_2^2 \mathrm{d}\omega \leq O(B_T^2).$$

In summary, we have

$$\int_{-\pi}^{\pi} \mathbb{E}\left\| \mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega \right\|^2 \mathrm{d}\omega \leq O(B_T^2) + O(B_T^{-1}T^{-1}).$$

The spectral density estimator $\mathscr{F}_\cdot^{(T)}$ is therefore consistent in integrated mean square if $B_T \to 0$ and $B_T T \to \infty$ as $T \to \infty$.
A careful examination of the proof reveals that the pointwise statement of the Theorem follows from a directly analogous argument. Indeed, the error term for the squared bias is uniform in ω. For the variance term, Proposition 3.5.4 tells us that for any fixed $\omega \in [-\pi,\pi]$, the Hilbert–Schmidt norm of the variance term is of order $O(B_T^{-1}T^{-1})$, which completes the proof. □

Let us now study the asymptotic distribution of the sample spectral density operators. In order to do this, we first need a result that is analogue to Proposition 3.5.4, but for the projections of the sample spectral density operators. We introduce some notation that will be used hereafter. Let (φ_n) be an orthonormal basis of $L^2([0,1],\mathbb{R})$. Then $\{\varphi_n \otimes_2 \varphi_m\}_{n,m\geq 1}$ is an orthonormal basis of the complex Hilbert space $L^2([0,1]^2,\mathbb{C})$ (e.g. Kadison & Ringrose (1997)). The representation of the spectral density operator $\mathscr{F}_\omega$ in this basis will be called the *spectral density matrix* (it is an infinite matrix in fact), and denoted by Φ_ω. We denote by $\Phi_\omega(m,n)$ its $(m,n)^{th}$ coordinate, i.e.

$$\Phi_\omega(m,n) = \left\langle \mathscr{F}_\omega \varphi_n, \varphi_m \right\rangle.$$

We also define the periodogram matrix $P_\omega^{(T)}(m,n)$ and the estimator of the spectral density matrix $\Phi_\omega^{(T)}(m,n)$ as the scalar product of $p_\omega^{(T)}$, respectively $f_\omega^{(T)}$, with $\varphi_m \otimes_2 \varphi_n$ in the space $HSL^2([0,1],\mathbb{C})$.

Proposition 3.6.3.
For T large enough, we have the following bound:

$$\sum_{m,n\geq 1} TB_T \mathrm{var}\left(\Phi_\omega^{(T)}(m,n)\right) \leq 24\|W\|_\infty^2 \cdot \left[\left(\sum_{t\in\mathbb{Z}} (1+|t|) \left|\!\left|\!\left| \mathscr{R}_t \right|\!\right|\!\right|_1 \right)^2 + \right.$$
$$\left. + \sum_{t_1,t_2,t_3\in\mathbb{Z}} \left|\!\left|\!\left| \mathscr{R}_{t_1,t_2,t_3} \right|\!\right|\!\right|_1 \right], \quad (3.6.1)$$

Provided the terms on the right-hand side converge. Here, $\mathscr{R}_{t_1,t_2,t_3}$ is the operator on $L^2\left([0,1]^2,\mathbb{C}\right)$ defined in (3.2.3)

Proof. Notice that

$$\Phi_{\omega}^{(T)}(m,n) = \frac{2\pi}{T}\sum_{s=1}^{T-1} W^{(T)}(\omega - 2\pi s/T) P_{2\pi s/T}^{(T)}(m,n),$$

where $W^{(T)}$ is defined in (3.5.1). Furthermore, let $\xi_t(m) = \langle X_t, \varphi_m \rangle$, and define the k^{th}-order cumulant spectra array $\Phi_{\omega_1,\ldots,\omega_{k-1}}$ by

$$\Phi_{\omega_1,\ldots,\omega_{k-1}}(m_1,\ldots,m_k) = \int_{[0,1]^k} f_{\omega_1,\ldots,\omega_{k-1}}(\tau_1,\ldots,\tau_k)\varphi_{m_1}(\tau_1)\cdots\varphi_{m_k}(\tau_k)\,d\boldsymbol{\tau}.$$

In other words, the k^{th}-order cumulant spectra array is the scalar product in $L^2([0,1]^k,\mathbb{C})$ between $f_{\omega_1,\ldots,\omega_{k-1}}$ and $\varphi_{m_1}\otimes\cdots\otimes\varphi_{m_k}$.

The proof is divided into two parts:

(I) For T large enough, the following holds uniformly in $n, m \geq 1$:

$$TB_T \operatorname{var}\left(\Phi_{\omega}^{(T)}(m,n)\right) \leq 8\|W\|_{\infty}^2 \Big[\operatorname{sc}_0(m,n,m,n) + \operatorname{sc}_1(m,m)\operatorname{sc}_1(n,n) + \operatorname{sc}_0(m,m)\operatorname{sc}_0(n,n) + \operatorname{sc}_0(m,n)^2 \Big], \quad (3.6.2)$$

where

$$\operatorname{sc}_0(m_1,\ldots,m_k) = \sum_{t_1,\ldots,t_{k-1}\in\mathbb{Z}} \left|\operatorname{cum}\left(\xi_{t_1}(m_1),\ldots,\xi_{t_{k-1}}(m_{k-1}),\xi_0(m_k)\right)\right|,$$

$$\operatorname{sc}_1(m_1,\ldots,m_k) = \sum_{j=1}^{k-1}\sum_{t_1,\ldots,t_{k-1}\in\mathbb{Z}} |t_j| \left|\operatorname{cum}\left(\xi_{t_1}(m_1),\ldots,\xi_{t_{k-1}}(m_{k-1}),\xi_0(m_k)\right)\right|.$$

(II) We have the following bound, for T large enough:

$$\sum_{m,n\geq 1} TB_T \operatorname{var}\left(\Phi_{\omega}^{(T)}(m,n)\right) \leq 24\|W\|_{\infty}^2 \left[\left(\sum_{t\in\mathbb{Z}} (1+|t|) \left|\!\left|\!\left| \mathscr{R}_t \right|\!\right|\!\right|_1 \right)^2 + \sum_{t_1,t_2,t_3\in\mathbb{Z}} \left|\!\left|\!\left| \mathscr{R}_{t_1,t_2,t_3} \right|\!\right|\!\right|_1 \right]. \quad (3.6.3)$$

Recall that $\mathscr{R}_{t_1,t_2,t_3}$ is the operator on $L^2\left([0,1]^2,\mathbb{C}\right)$ with kernel

$$r_{t_1,t_2,t_3}((\tau_1,\tau_2),(\tau_3,\tau_4)) = \operatorname{cum}\left(X_{t_1},X_{t_2},X_{t_3},X_0\right)(\tau_1,\tau_2,\tau_3,\tau_4).$$

That is,

$$\mathscr{R}_{t_1,t_2,t_3} f(\tau_1,\tau_2) = \iint_{[0,1]^2} r_{t_1,t_2,t_3}((\tau_1,\tau_2),(\tau_3,\tau_4))\, f(\tau_3,\tau_4)\,d\tau_3 d\tau_4$$

for $f \in L^2([0,1]^2,\mathbb{C})$.

First we concentrate on establishing (3.6.2). Recall that

$$\operatorname{var}\left(\Phi_{\omega}^{(T)}(m,n)\right) = (2\pi/T)^2 \sum_{s,l=1}^{T-1} W^{(T)}(\omega - 2\pi s/T) W^{(T)}(\omega - 2\pi l/T) \times \operatorname{cov}\left(P_{\frac{2\pi s}{T}}^{(T)}(m,n), P_{\frac{2\pi l}{T}}^{(T)}(m,n)\right)$$

We need to find an explicit bound on the error terms of Lemma 3.3.2, Theorem 3.3.3, and Theorem 3.4.5. An examination of the proof of Lemma 3.3.2 yields

$$\Phi_{\omega_1,\ldots,\omega_{k-1}}(m_1,\ldots,m_k) = (2\pi)^{-(k-1)} \sum_{t_1,\ldots,t_{k-1}=-(T-1)}^{T-1} \exp\Big(-\mathbf{i}\sum_{j=1}^{k-1}\omega_j t_j\Big) \times \times \operatorname{cum}\left(\xi_{t_1}(m_1),\ldots,\xi_{t_{k-1}}(m_{k-1}),\xi_0(m_k)\right) + + \varepsilon_T^{(3.3.2)}(m_1,\ldots,m_k),$$

and $\left|\varepsilon_T^{(3.3.2)}(m_1,\ldots,m_k)\right| \le (2\pi)^{-(k-1)}(k-1)\,\mathrm{sc}_0(m_1,\ldots,m_k)$. We have used the notation $\varepsilon_T^{(3.3.2)}(m_1,\ldots,m_k)$ to denote the error term of Lemma 3.3.2, and we shall do likewise for the error term in Theorem 3.3.3:

$$T^{k/2}\operatorname{cum}\left(\widetilde{\xi}_{\omega_1}^{(T)}(m_1),\ldots,\widetilde{\xi}_{\omega_k}^{(T)}(m_k)\right) = (2\pi)^{k/2-1}\Delta^{(T)}\left(\sum_{j=1}^{k}\omega_j\right)\Phi_{\omega_1,\ldots,\omega_{k-1}}(m_1,\ldots,m_k) + \varepsilon_T^{(3.3.3)}\left(\sum_{j=1}^{k}\omega_j; m_1,\ldots,m_k\right),$$

where

$$\begin{aligned}\left|\varepsilon_T^{(3.3.3)}(\omega; m_1,\ldots,m_k)\right| &\le 2(2\pi)^{-k/2} \sum_{t_1,\ldots,t_{k-1}=-(T-1)}^{T-1} (|t_1|+\cdots+|t_{k-1}|)\left|\operatorname{cum}\left(\xi_{t_1}(m_1),\ldots,\xi_{t_{k-1}}(m_{k-1}),\xi_0(m_k)\right)\right| \\ &\quad + (2\pi)^{k/2-1}\Delta^{(T)}(\omega)\left|\varepsilon_T^{(3.3.2)}(m_1,\ldots,m_k)\right| \\ &\le 2(2\pi)^{-k/2}\,\mathrm{sc}_1(m_1,\ldots,m_k) + (2\pi)^{-k/2}(k-1)\Delta^{(T)}(\omega)\,\mathrm{sc}_0(m_1,\ldots,m_k).\end{aligned}$$

A less sharp bound (but independent of the frequency) will also be useful:

$$\left|\varepsilon_T^{(3.3.3)}(\cdot; m_1,\ldots,m_k)\right| \le 3(2\pi)^{-k/2}(k-1)T\,\mathrm{sc}_0(m_1,\ldots,m_k).$$

We will also need a bound on the spectral density matrix:

$$\left|\Phi_{\omega_1,\ldots,\omega_{k-1}}(m_1,\ldots,m_k)\right| \le (2\pi)^{-(k-1)}\,\mathrm{sc}_0(m_1,\ldots,m_k).$$

We now turn to Theorem 3.4.5: for $s, l = 1, \dots, T-1$,

$$\begin{aligned}\operatorname{cov}\left(P^{(T)}_{\frac{2\pi s}{T}}(m,n), P^{(T)}_{\frac{2\pi l}{T}}(m,n)\right) &= (2\pi/T)\Phi_{\frac{2\pi s}{T},-\frac{2\pi s}{T},\frac{2\pi l}{T}}(m,n,m,n) + T^{-2}\varepsilon^{(3.3.3)}_T(\cdot;m,n,m,n)\\ &+ \delta_{s,l}\Big[\Phi_{\frac{2\pi s}{T}}(m,m)\Phi_{-\frac{2\pi s}{T}}(n,n) + \Phi_{\frac{2\pi s}{T}}(m,m)T^{-1}\varepsilon^{(3.3.3)}_T(\cdot;n,n)\\ &\quad + \Phi_{-\frac{2\pi s}{T}}(n,n)T^{-1}\varepsilon^{(3.3.3)}_T(\cdot;m,m)\Big]\\ &+ \delta_{s+l,T}\Big[\Phi_{\frac{2\pi s}{T}}(m,n)\Phi_{-\frac{2\pi s}{T}}(n,m) + \Phi_{\frac{2\pi s}{T}}(m,n)T^{-1}\varepsilon^{(3.3.3)}_T(\cdot;n,m)\\ &\quad + \Phi_{-\frac{2\pi s}{T}}(n,m)T^{-1}\varepsilon^{(3.3.3)}_T(\cdot;m,n)\Big]\\ &+ T^{-2}\Bigg[\varepsilon^{(3.3.3)}_T\left(\frac{2\pi(s-l)}{T};m,m\right)\varepsilon^{(3.3.3)}_T\left(-\frac{2\pi(s-l)}{T};n,n\right)\\ &\quad + \varepsilon^{(3.3.3)}_T\left(\frac{2\pi(s+l)}{T};m,n\right)\varepsilon^{(3.3.3)}_T\left(-\frac{2\pi(s+l)}{T};n,m\right)\Bigg],\end{aligned}$$

where $\delta_{s,l} = 1$ if $s = l$, and zero otherwise. Using the previous bounds, and the fact that $\mathrm{sc}_0(m,n) = \mathrm{sc}_0(n,m)$, we obtain

$$\begin{aligned}\left|\operatorname{cov}\left(P^{(T)}_{\frac{2\pi s}{T}}(m,n), P^{(T)}_{\frac{2\pi l}{T}}(m,n)\right)\right| &\le \frac{1}{4\pi^2}\left[4T^{-2}\mathrm{sc}_1(m,m)\mathrm{sc}_1(n,n) + 10T^{-1}\,\mathrm{sc}_0(m,n,m,n)\right.\\ &\left. + 8\,\delta_{s,l}\mathrm{sc}_0(m,m)\mathrm{sc}_0(n,n) + 8\,\delta_{s+l,T}\mathrm{sc}_0(m,n)^2\right],\end{aligned}$$

and hence

$$TB_T\left|\operatorname{var}\left(\Phi^{(T)}_\omega(m,n)\right)\right| \tag{3.6.4}$$

$$\le B_T\left[T^{-1}\sum_{s=1}^{T-1} W^{(T)}(\omega - 2\pi s/T)\right]^2\left[4T^{-1}\mathrm{sc}_1(m,m)\mathrm{sc}_1(n,n) + 10\,\mathrm{sc}_0(m,n,m,n)\right]$$

$$+ 8\,\mathrm{sc}_0(m,m)\mathrm{sc}_0(n,n)\,B_T T^{-1}\sum_{s=1}^{T-1}\left(W^{(T)}(\omega - 2\pi s/T)\right)^2 \tag{3.6.5}$$

$$+ 8\,\mathrm{sc}_0(m,n)^2\,B_T T^{-1}\sum_{s=1}^{T-1} W^{(T)}(\omega - 2\pi s/T)W^{(T)}(\omega + 2\pi s/T). \tag{3.6.6}$$

Since at most $\frac{TB_T}{\pi} + 1$ of the summands are non-zero, and

$$\left\|W^{(T)}\right\|_\infty \le B_T^{-1}\|W\|_\infty$$

(see Lemma 3.12.18), we obtain

$$\left[T^{-1}\sum_{s=1}^{T-1} W^{(T)}(\omega - 2\pi s/T)\right]^2 \le \pi^{-2}\|W\|^2_\infty,$$

and

$$B_T T^{-1} \sum_{s=1}^{T-1} (W^{(T)}(\omega - 2\pi s/T))^2 \leq \pi^{-1} \|W\|_\infty^2,$$

for large T. Similarly

$$\left| B_T T^{-1} \sum_{s=1}^{T-1} W^{(T)}(\omega - 2\pi s/T) W^{(T)}(\omega + 2\pi s/T) \right| \leq \pi^{-1} \|W\|_\infty^2$$

for large T. Since $B_T \to 0$, for T large enough, we have the following, uniformly in $n, m \geq 1$:

$$\begin{aligned} TB_T \left| \mathrm{var}\left(\Phi_\omega^{(T)}(m,n)\right) \right| \leq \|W\|_\infty^2 \cdot \big[& \mathrm{sc}_0(m,n,m,n) + \mathrm{sc}_1(m,m)\mathrm{sc}_1(n,n) \\ & + 8\mathrm{sc}_0(m,m)\mathrm{sc}_0(n,n) + 8\mathrm{sc}_0(m,n)^2 \big], \end{aligned}$$

and (3.6.2) follows immediately.
To prove (3.6.3), notice that, for large T, inequality (3.6.2) gives us

$$\begin{aligned} \sum_{m,n\geq 1} TB_T \mathrm{var}\left(\Phi_\omega^{(T)}(m,n)\right) \leq 8\|W\|_\infty^2 \Bigg[& \sum_{m,n\geq 1} \mathrm{sc}_0(m,n,m,n) + \left(\sum_{m\geq 1} \mathrm{sc}_1(m,m) \right)^2 + \\ & + \left(\sum_{m\geq 1} \mathrm{sc}_0(m,m) \right)^2 + \sum_{m,n\geq 1} \mathrm{sc}_0(m,n)^2 \Bigg] \end{aligned}$$

Notice that

$$\mathrm{cum}\left(\xi_{t_1}(m), \xi_{t_2}(n), \xi_{t_3}(m), \xi_0(n)\right) = \left\langle \mathscr{R}_{t_1,t_2,t_3} \varphi_m \otimes_2 \varphi_n, \varphi_m \otimes_2 \varphi_n \right\rangle,$$

so

$$\sum_{m,n\geq 1} \mathrm{sc}_0(m,n,m,n) \leq \sum_{t_1,t_2,t_3 \in \mathbb{Z}} \left|\!\left|\!\left| \mathscr{R}_{t_1,t_2,t_3} \right|\!\right|\!\right|_1.$$

We also have $\mathrm{cum}\left(\xi_t(m), \xi_0(n)\right) = \left\langle \mathscr{R}_t \varphi_n, \varphi_m \right\rangle$, hence

$$\sum_{m\geq 1} \mathrm{sc}_0(m,m) \leq \sum_{t\in\mathbb{Z}} |||\mathscr{R}_t|||_1.$$

Using the Cauchy-Schwarz inequality and Parseval's identity, we also obtain

$$\sum_{m,n\geq 1} \mathrm{sc}_0(m,n)^2 \leq \left(\sum_{t\in\mathbb{Z}} |||\mathscr{R}_t|||_1 \right)^2.$$

Similarly,

$$\sum_{m,n\geq 1} \mathrm{sc}_1(m,m)\mathrm{sc}_1(n,n) \leq \left(\sum_{t\in\mathbb{Z}} |t| |||\mathscr{R}_t|||_1 \right)^2.$$

Inequality (3.6.3) follows, and the proof is finished. □

We automatically get the following corollary, that gives sufficient conditions for the convergence in distribution of the sample spectral density

operators.

Corollary 3.6.4. *If*

$$\sum_{t\in\mathbb{Z}}(1+|t|)\|\mathscr{R}_t\|_1<\infty \quad \textit{and} \quad \sum_{t_1,t_2,t_3\in\mathbb{Z}}\left\|\mathscr{R}_{t_1,t_2,t_3}\right\|_1<\infty, \tag{3.6.7}$$

the sequence of random elements

$$\left(\sqrt{TB_T}(\mathscr{F}_\omega^{(T)}-\mathbb{E}\mathscr{F}_\omega^{(T)})\right)_{T=1,2,\ldots}\subset\mathscr{S}_2(H)$$

is tight in $\mathscr{S}_2(H)$. *In particular, if under some assumptions* ***A****, and for some random element* $\mathscr{G}$ *of* $\mathscr{S}_2(H)$, *we have*

$$\left\langle\sqrt{TB_T}(\mathscr{F}_\omega^{(T)}-\mathbb{E}\mathscr{F}_\omega^{(T)}),\varphi\right\rangle_{\mathscr{S}_2}\xrightarrow{d}\langle\mathscr{G},\varphi\rangle_{\mathscr{S}_2},\quad \varphi\in\mathscr{S}_2(H),$$

then ***A*** *and* (3.6.7) *imply that*

$$\sqrt{TB_T}(\mathscr{F}_\omega^{(T)}-\mathscr{F}_\omega^{(T)})\xrightarrow{d}\mathscr{G},\quad \textit{in } \mathscr{S}_2(H).$$

Proof. Since

$$\mathbb{E}\left|\left\langle\sqrt{TB_T}\left(f_{\omega_j}^{(T)}-\mathbb{E}f_{\omega_j}^{(T)}\right),\varphi_m\otimes_2\varphi_n\right\rangle\right|^2=TB_T\operatorname{var}\left(\Phi_\omega^{(T)}(m,n)\right),$$

An application of Lemma C.2.3 and the previous Proposition yields the result. □

Finally, we may obtain the asymptotic distribution of our estimator as being Gaussian, if we include some higher-order cumulant mixing conditions.

Theorem 3.6.5.
Assume that $\mathbb{E}\|X_0\|^k<\infty$ *for all* $k\geq 2$ *and*

(i) $\sum_{t_1,\ldots,t_{k-1}=-\infty}^{\infty}\left\|\operatorname{cum}\left(X_{t_1},\ldots,X_{t_{k-1}},X_0\right)\right\|_2<\infty,\quad$ *for all* $k\geq 2$,

(i') $\sum_{t_1,\ldots,t_{k-1}=-\infty}^{\infty}(1+|t_j|)\left\|\operatorname{cum}\left(X_{t_1},\ldots,X_{t_{k-1}},X_0\right)\right\|_2<\infty$, *for* $k\in\{2,4\}$ *and* $j<k$,

(ii) $\sum_{t\in\mathbb{Z}}(1+|t|)\|\mathscr{R}_t\|_1<\infty$,

(iii) $\sum_{t_1,t_2,t_3\in\mathbb{Z}}\left\|\mathscr{R}_{t_1,t_2,t_3}\right\|_1<\infty$,

Then, for any frequencies $\omega_1,\ldots,\omega_J$, *with* $J<\infty$, *if* $B_T\to 0$ *and* $TB_T\to\infty$,

$$\sqrt{B_TT}\left(f_{\omega_j}^{(T)}-\mathbb{E}f_{\omega_j}^{(T)}\right)\xrightarrow{d}\check{f}_{\omega_j},\quad j=1,\ldots,J,$$

where $\check{f}_{\omega_j}$, $j = 1,\dots,J$ *are jointly mean zero complex Gaussian elements in* $L^2\left([0,1]^2,\mathbb{C}\right)$, *with covariance kernel*

$$\operatorname{cov}\left(\check{f}_{\omega_i}(\tau_1,\sigma_1),\check{f}_{\omega_j}(\tau_2,\sigma_2)\right) = 2\pi\int_{\mathbb{R}} W(\alpha)^2 d\alpha \times$$
$$\times\left\{\eta(\omega_i-\omega_j)f_{\omega_i}(\tau_1,\tau_2)f_{-\omega_i}(\sigma_1,\sigma_2)+\eta(\omega_i+\omega_j)f_{\omega_i}(\tau_1,\sigma_2)f_{-\omega_i}(\sigma_1,\tau_2)\right\}.$$

In particular, we see that $\check{f}_{\omega_i}$ *and* $\check{f}_{\omega_j}$ *are independent if* $\omega_i \pm \omega_j \not\equiv 0 \mod 2\pi$, *and* $\check{f}_{\omega}$ *is real Gaussian if* $\omega \equiv 0 \mod \pi$.

Notice that the limiting random elements $\check{f}_{\omega}$ are entirely determined by their covariance operators $\mathbb{E}\left[\check{\mathscr{F}}_{\omega}\otimes_2\check{\mathscr{F}}_{\omega}\right]$ and their relation operator $\mathbb{E}\left[\check{\mathscr{F}}_{\omega}\otimes_2\overline{\check{\mathscr{F}}_{\omega}}\right]$, see Section 3.12.1

Proof of Theorem 3.6.5. By Corollary 3.6.4, we know that the rescaled sample spectral density operator

$$\sqrt{TB_T}\left(f^{(T)}_{\omega_j}-\mathbb{E}f^{(T)}_{\omega_j}\right)$$

is tight. Therefore the vector

$$\sqrt{TB_T}\left(f^{(T)}_{\omega_1}-\mathbb{E}f^{(T)}_{\omega_1},\dots,f^{(T)}_{\omega_J}-\mathbb{E}f^{(T)}_{\omega_J}\right)$$

is also tight in $\left(L^2\left([0,1]^2,\mathbb{C}\right)\right)^J$. Applying Brillinger (2001, Theorem 7.4.4) to the finite dimensional distributions of this vector completes the proof. □

Condition (ii) guarantees that $\omega \mapsto \mathscr{F}_{\omega}$ is continuous with respect to the nuclear norm. If in addition we want it to be continuous in continuous in τ,σ, we need to assume the stronger conditions $\sum_{t\in\mathbb{Z}}\|r_t\|_{\infty}<\infty$ and that each r_t is continuous.

Remark 3.6.6. *It can be useful to write the second-order structure of the random element* $\check{\mathscr{F}}_{\omega}$ *using Kronecker product notation; indeed, this reveals the structure of the random element if* H *is taken to be an abstract complexified separable Hilbert space. If* $\omega\in(0,\pi)$, *then* $\check{\mathscr{F}}_{\omega}$ *is a complex Gaussian random element. Using results from Section A.3.4, we get that its covariance operator can be written*

$$\mathbb{E}\left[\check{\mathscr{F}}_{\omega}\otimes_2\check{\mathscr{F}}_{\omega}\right] = \kappa^2\cdot\mathscr{F}_{\omega}\,\widetilde{\bigotimes}_2\,\mathscr{F}_{\omega},\quad \omega\in(0,\pi), \tag{3.6.8}$$

and its relation operator can be written

$$\mathbb{E}\left[\check{\mathscr{F}}_{\omega}\otimes_2\overline{\check{\mathscr{F}}_{\omega}}\right] = \kappa^2\cdot\mathscr{F}_{\omega}\,\widetilde{\bigotimes}_{\mathsf{T}}\,\overline{\mathscr{F}_{\omega}},\quad \omega\in(0,\pi), \tag{3.6.9}$$

If $\omega\in\{0,\pi\}$, $\check{\mathscr{F}}_{\omega}$ *is a real Gaussian random element with covariance opera-*

tor

$$\mathbb{E}\left[\check{\mathscr{F}}_\omega \otimes_2 \check{\mathscr{F}}_\omega\right] = \kappa^2 \cdot [C \widetilde{\bigotimes}_2 C + C \widetilde{\bigotimes}_\tau C], \quad \omega \in \{0, \pi\} \tag{3.6.10}$$

where $\kappa^2 = 2\pi \int_{\mathbb{R}} W(x)^2 dx$.

We have seen that for the sample spectral density operators to be asymptotically Gaussian around their mean, we need to smooth at a rate $B_T \to 0$ slow enough such that $TB_T \to \infty$. If we furthermore assume that $TB_T^3 \to 0$ (which can be understood as $B_T \to 0$ fast enough), the sample spectral density operators are asymptotically Gaussian around the true spectral density operators:

Theorem 3.6.7.
Under the conditions of Theorem 3.6.5, for any frequencies $\omega_1, \ldots, \omega_J$, *with* $J < \infty$, *if* $B_T \to 0$, $TB_T \to \infty$ *and* $TB_T^3 \to 0$,

$$\sqrt{B_T T}\left(f_{\omega_j}^{(T)} - f_{\omega_j}\right) \xrightarrow{d} \check{f}_{\omega_j}, \quad j = 1, \ldots, J,$$

where $\check{f}_{\omega_j}$, $j = 1, \ldots, J$ *are jointly mean zero complex Gaussian elements in* $L^2\left([0,1]^2, \mathbb{C}\right)$, *with covariance kernel*

$$\operatorname{cov}\left(\check{f}_{\omega_i}(\tau_1, \sigma_1), \check{f}_{\omega_j}(\tau_2, \sigma_2)\right) = 2\pi \int_{\mathbb{R}} W(\alpha)^2 d\alpha \times$$
$$\times \left\{\eta(\omega_i - \omega_j) f_{\omega_i}(\tau_1, \tau_2) f_{-\omega_i}(\sigma_1, \sigma_2) + \eta(\omega_i + \omega_j) f_{\omega_i}(\tau_1, \sigma_2) f_{-\omega_i}(\sigma_1, \tau_2)\right\}.$$

In particular, we see that $\check{f}_{\omega_i}$ *and* $\check{f}_{\omega_j}$ *are independent if* $\omega_i \pm \omega_j \not\equiv 0 \mod 2\pi$, *and* $\check{f}_\omega$ *is real Gaussian if* $\omega \equiv 0 \mod \pi$.

Again, notice that the limiting random elements $\check{f}_\omega$ are entirely determined by their covariance operators $\mathbb{E}\left[\check{\mathscr{F}}_\omega \otimes_2 \check{\mathscr{F}}_\omega\right]$ and their relation operator $\mathbb{E}\left[\check{\mathscr{F}}_\omega \otimes_2 \overline{\check{\mathscr{F}}_\omega}\right]$, see Section 3.12.1.

Proof. Write

$$\sqrt{TB_T}\left(\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\right) = \underbrace{\sqrt{TB_T}\left(\mathscr{F}_\omega^{(T)} - \mathbb{E}\,\mathscr{F}_\omega^{(T)}\right)}_{(i)} + \underbrace{\sqrt{TB_T}\left(\mathbb{E}\,\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\right)}_{(ii)}.$$

The term (i) converges in distribution to $\check{\mathscr{F}}_\omega$, jointly for ωs satisfying the condition of the Theorem, and the limiting random elements are independent (see Theorem 3.6.5). For (ii), earlier results yield

$$\mathbb{E}\,\mathscr{F}_\omega = \int_{\mathbb{R}} W(x) f_{\omega - \alpha B_T} d\alpha + O([TB_T]^{-1}) = \mathscr{F}_\omega + O(B_T) + O([TB_T]^{-1}),$$

uniformly in τ, σ, and hence

$$\sqrt{TB_T}\left|\left|\left| \mathbb{E}\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega \right|\right|\right|_2 = O\left(\sqrt{TB_T^3}\right) + O([TB_T]^{-1/2}),$$

which converges to zero under the conditions of the Theorem. The proof is finished by applying Slutsky's lemma. □

When $\omega = 0$, the operator $2\pi\mathscr{F}_\omega$ reduces to the long-run covariance operator $\sum_{t\in\mathbb{Z}} \mathscr{R}_t$, the limiting covariance operator of the empirical mean. Correspondingly, $2\pi\mathscr{F}_0^{(T)}$ is an estimator of the long-run covariance operator that is consistent in mean square for the long-run covariance, under no structural modeling assumptions. A similar estimator was also considered in Horváth, Kokoszka & Reeder (2013), who derived weak consistency under L^4-m-approximability weak dependence conditions. Hörmann & Kokoszka (2010) studied this problem by projecting onto a finite-dimensional subspace. However, neither of these papers consider functional central limit theorems for the estimator of the long-run covariance operator; taking $\omega = 0$, in Theorem 3.6.5, we obtain such a result:

Corollary 3.6.8. *Under the conditions of Theorem 3.6.5, we have*

$$\sqrt{B_T T}\left(2\pi\mathscr{F}_0^{(T)} - 2\pi\mathscr{F}_0\right) \xrightarrow{d} \mathscr{N}\left(0, (2\pi)^{3/2}\|W\|_2^2\mathfrak{C}\right),$$

where $\mathfrak{C}$ *is the integral operator on* $L^2([0,1]^2, \mathbb{R})$ *with kernel*

$$c(\tau_1, \sigma_1, \tau_2, \sigma_2) = \left\{f_0(\tau_1, \tau_2)f_0(\sigma_1, \sigma_2) + f_0(\tau_1, \sigma_2)f_0(\sigma_1, \tau_2)\right\}.$$

We remark that the limiting Gaussian random operator is purely real.

3.7 Estimation of the Eigenstructure of the Spectral Density Operators

Now that we have seen that the sample spectral density operators are consistent and asymptotically Gaussian estimators of the spectral density operators, we shall study the asymptotic behavior of their eigenstructure.

3.7.1 Basic definitions

Before stating our results concerning the asymptotic distribution of the estimators of the eigenvalues/eigenfunctions, we need to introduce some necessary notation. For any $\omega \in [0, \pi]$, let

$$\mathscr{F}_\omega^{(T)} = \sum_{i=1}^{\infty} \mu_{i,T}(\omega)\varphi_{i,T}^\omega \otimes_2 \varphi_{i,T}^\omega$$

be the singular value decomposition of $\mathscr{F}_\omega^{(T)}$, and recall that

$$\mathscr{F}_\omega = \sum_{i=1}^{\infty} \mu_i(\omega) \varphi_i^\omega \otimes_2 \varphi_i^\omega$$

is the singular decomposition of $\mathscr{F}_\omega$. For any fixed ω, $\{\mu_{i,T}(\omega)\}_{i\geq 1}$ and $\{\mu_i(\omega)\}_{i\geq 1}$ are non-increasing positive sequences tending to zero. We denote by $\{\lambda_i(\omega)\}_{i\geq 1}$ the decreasing sequence of distinct elements of $\{\mu_i(\omega)\}_{i\geq 1}$, define the sets

$$I_k(\omega) = \{i \geq 1 : \mu_i(\omega) = \lambda_k(\omega)\}, \quad k \geq 1,$$

and we denote its cardinality by $m_k(\omega) = |I_k(\omega)|$. We will also write

$$I(\omega) = \big\{i \geq 1 : \mu_i(\omega) > 0\big\} = \bigcup_{k\geq 1 \& \lambda_k(\omega)>0} I_k(\omega) \tag{3.7.1}$$

which is the set of indices of the repeated non-zero eigenvalues of $\mathscr{F}_\omega$. Notice that the sets $I_j(\omega)$ satisfy

$$k < l \implies I_k(\omega) < I_l(\omega),$$

and that

$$I(\omega) = J(\omega) = \{1, 2, 3, \ldots\}$$

unless $\mathscr{F}_\omega$ is of finite rank, in which case

$$\{\lambda_i(\omega)\}_{i\geq 1} = \{\lambda_1(\omega), \ldots, \lambda_N(\omega)\},$$

where $N-1 = \operatorname{rank}(\mathscr{F}_\omega)$ and $\lambda_N(\omega) = 0$. Furthermore, $m_N(\omega) = |I_N(\omega)| = \infty$. We can now define, for $k \in J(\omega)$,

$$\Pi_k(\omega) = \sum_{i\in I_k(\omega)} \varphi_i^\omega \otimes_2 \varphi_i^\omega,$$

which is the projection onto the k^{th} eigenspace of $\mathscr{F}_\omega$. This way,

$$\mathscr{F}_\omega = \sum_{j\in J(\omega)} \lambda_j(\omega)\Pi_j(\omega).$$

The estimator of $\Pi_k(\omega)$ is defined by

$$\Pi_{k,T}(\omega) = \sum_{i\in I_k(\omega)} \varphi_{i,T}^\omega \otimes_2 \varphi_{i,T}^\omega.$$

We also define

$$S_k(\omega) = \sum_{j\in J(\omega): j\neq k} (\lambda_k(\omega) - \lambda_j(\omega))^{-1}\Pi_j(\omega), \quad k \in J(\omega). \tag{3.7.2}$$

Notice that the sum is over all $j \neq k$ such that $\lambda_j(\omega) \neq 0$. We define the operator

$$\eta_k^\omega = S_k(\omega) \widetilde{\bigotimes}_2 \Pi_k(\omega) + \Pi_k(\omega) \widetilde{\bigotimes}_2 S_k(\omega), \quad k \in J(\omega), \tag{3.7.3}$$

which belongs to $\mathscr{S}_\infty(\mathscr{S}_2(H))$, and the bounded operator $p_k^\omega : \mathscr{S}_2(H) \to \mathbb{C}$ by

$$p_k^\omega(A) = \langle A, \Pi_k(\omega) \rangle_{\mathscr{S}_2}, \quad \omega \in J(\omega). \tag{3.7.4}$$

Notice that if $A \in \mathscr{S}_2(H)$ is a self-adjoint operator, then $p_k^\omega(A) \in \mathbb{R}$. The following Lemma establishes that η_k^ω is indeed a bounded operator.

Lemma 3.7.1.
Letting $\delta_k(\omega) = \min(\lambda_k(\omega) - \lambda_{k+1}(\omega), \lambda_{k-1}(\omega) - \lambda_k(\omega))$, *we have*

$$\left|\left|\left| \eta_k^\omega \right|\right|\right|_\infty = \delta_k^{-1}(\omega) < \infty, \quad k \in J(\omega).$$

Proof. We omit writing ω to simplify the notation.
Letting $A = \sum_{j \in J(\omega); j \neq k}^\infty \alpha_{jk} \varphi_j \otimes_2 \varphi_k \in \mathscr{S}_2(L^2([0,1], \mathbb{C}))$,

$$\eta_k A = \sum_{j \in J(\omega); j \neq k} \alpha_{jk} (\lambda_k - \lambda_j)^{-1} \varphi_j \otimes_2 \varphi_k + \sum_{j \in J(\omega); j \neq k} \alpha_{kj} (\lambda_k - \lambda_j)^{-1} \varphi_k \otimes_2 \varphi_j,$$

thus

$$\left|\left|\left| \eta_k A \right|\right|\right|_2^2 = \sum_{j \in J(\omega); j \neq k} \frac{|\alpha_{kj}|^2 + |\alpha_{jk}|^2}{(\lambda_k - \lambda_j)^2} \leq |||A|||_2^2 / \delta_k^2,$$

with equality if we choose the α_{kj} and α_{jk} properly. □

3.7.2 Consistency

The following proposition establishes the consistency of the sample eigenvalues and sample eigenprojectors.

Proposition 3.7.2.
Under ***C(1,2)*** *and* ***C(1,4)****,*

$$\sup_{\omega \in [-\pi, \pi]} \mathbb{E}\left[\sup_{i \in I(\omega)} |\mu_{i,T}(\omega) - \mu_i(\omega)|^2 \right] = O(B_T^2) + O(B_T^{-2} T^{-1}),$$
$$\int_{-\pi}^{\pi} \mathbb{E}\left[\sup_{i \in I(\omega)} |\mu_{i,T}(\omega) - \mu_i(\omega)|^2 \right] d\omega = O(B_T^2) + O(B_T^{-1} T^{-1}).$$

Furthermore, for each $j \in J(\omega)$, *and each* $\omega \in [-\pi, \pi]$,

$$\left|\left|\left| \Pi_{j,T}(\omega) - \Pi_j(\omega) \right|\right|\right|_2 \xrightarrow{p} 0. \tag{3.7.5}$$

if $TB_T \to \infty$, $B_T \to 0$.

Proof. Using Proposition A.2.4, we get

$$\sup_{i\in I(\omega)} |\mu_{i,T}(\omega) - \mu_i(\omega)|^2 < \left|\left|\left| \mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega \right|\right|\right|_2^2$$

and therefore the first two statements follow from Theorems 3.6.1 and 3.6.2.
The proof of (3.7.5) is based on the "linear approximation" of the difference between the eigenprojector, found in Mas & Menneteau (2003). We fix ω and write Π_j instead of $\Pi_j(\omega)$, and similarly for η_j, λ_j, to ease the notation. Mas & Menneteau (2003) yields

$$\Pi_{j,T} - \Pi_j = \eta_j \left(\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega \right) + r_{j,T},$$

where $\eta_j = \eta_j^\omega$ has been defined in (3.7.3), and the remainder term $r_{i,T}$ satisfies

$$\left|\left|\left| r_{j,T} \mathbf{1}_{O_{j,T}} \right|\right|\right|_2 \leq \frac{8}{\delta_j^2} \left|\left|\left| \mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega \right|\right|\right|_2, \tag{3.7.6}$$

with $O_{j,T} = \left\{ \left|\left|\left| \mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega \right|\right|\right|_2 < \delta_j/4 \right\}$ and $\delta_j = \min\left(\lambda_j - \lambda_{j+1}, \lambda_{j-1} - \lambda_j\right)$. The operator η_j is continuous, with operator norm $1/\delta_i$ (See Lemma 3.7.1), hence

$$\mathbb{E} \left|\left|\left| \eta_j \left(\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega \right) \right|\right|\right|_2^2 \to 0.$$

We now turn to the remainder term:

$$r_{j,T} = r_{j,T} \mathbf{1}_{O_{j,T}} + r_{j,T} \mathbf{1}_{O_{j,T}^c}.$$

From (3.7.6), the first term converges in probability to zero. For the second term,

$$\mathbb{P}\left(\left|\left|\left| r_{j,T} \mathbf{1}_{O_{j,T}^c} \right|\right|\right|_2 > \varepsilon \right) \leq \mathbb{P}\left(O_{j,T}^c \right) \to 0, \quad \text{for any } \varepsilon > 0,$$

therefore $\Pi_{j,T} \xrightarrow{p} \Pi_j$ as $T \to \infty$. □

The results of this proposition are stronger for the eigenvalues of the sample spectral density operators than for its eigenprojectors. This is compatible with our intuition, since eigenprojections are less stable to perturbations compared to eigenvalues.

3.7.3 Asymptotic Normality

Assuming some decay of the higher order moments, we can show that the asymptotic distribution of the eigenvalues and eigenprojectors of the sample spectral density operators are Gaussian:

Theorem 3.7.3.
Let $\omega_1, \ldots, \omega_K \in [0, \pi]$ be distinct, $K < \infty$. Let $L_i \subset J(\omega_i)$, be a set of cardinality $|L_i| < \infty$, for $i = 1, \ldots, K$.

Provided the conditions of Theorem 3.6.5 hold, and $B_T \to 0$ such that $TB_T \to \infty$ and $TB_T^3 \to 0$, then

$$\sqrt{TB_T}\{\Pi_{j,T}(\omega_i) - \Pi_j(\omega_i) : j \in L_i\} \xrightarrow{d} \eta_{L_i}^{\omega_i}(\check{\mathscr{F}}_{\omega_i}), \quad i = 1, \ldots, K.$$

and

$$\sqrt{TB_T}\left\{\sum_{s \in I_j(\omega_i)} [\mu_{s,T}(\omega_i) - \lambda_j(\omega_i)] : j \in L_i\right\} \xrightarrow{d} p_{L_i}^{\omega_i}(\check{\mathscr{F}}_{\omega_i}), \quad i = 1, \ldots, K.$$

The limiting random elements $\{\eta_{L_i}^{\omega_i}(\check{\mathscr{F}}_{\omega_i})\}_{i=1,\ldots,K}$ and $\{p_{L_i}^{\omega_i}(\check{\mathscr{F}}_{\omega_i})\}_{i=1,\ldots,K}$ are all independent Gaussian random elements with mean zero. In particular, the random elements $\eta_{L_i}^{\omega_i}\left(\check{\mathscr{F}}_{\omega_i}\right)$ are complex, but the variables $p_{L_i}^{\omega_i}\left(\check{\mathscr{F}}_{\omega_i}\right)$ are real. Their covariances are given by the following formulas, in which we have written λ_k instead of $\lambda_k(\omega)$ for clarity, and similarly for Π_k, φ_k,

$$\mathbb{E}\left[\eta_k^{\omega}(\check{\mathscr{F}}_{\omega}) \bigotimes_2 \eta_l^{\omega}(\check{\mathscr{F}}_{\omega})\right] = \begin{cases} -\kappa^2 \lambda_k \lambda_l (\lambda_k - \lambda_l)^{-2} \left[\Pi_k \,\widetilde{\otimes}_2\, \Pi_l + \Pi_l \,\widetilde{\otimes}_2\, \Pi_k + A_{kl}^{\omega} + \left(A_{kl}^{\omega}\right)^{\dagger}\right], & k \neq l; \\ \kappa^2 \sum_{s \in J \setminus \{k\}} \lambda_k \lambda_s (\lambda_k - \lambda_s)^{-2} \left[\Pi_k \,\widetilde{\otimes}_2\, \Pi_s + \Pi_s \,\widetilde{\otimes}_2\, \Pi_k + A_{ks}^{\omega} + \left(A_{ks}^{\omega}\right)^{\dagger}\right], & k = l, \end{cases} \tag{3.7.7}$$

where $\kappa^2 = 2\pi \int_{\mathbb{R}} W(x)^2 dx$, $A_{ks}^{\omega} = \mathbf{1}_{\{0,\pi\}}(\omega) \cdot \Pi_k(\omega) \,\widetilde{\otimes}_{\mathsf{T}}\, \Pi_s(\omega)$. The relation operator *of $\check{\mathscr{F}}_{\omega}$, for $\omega \in (0,\pi)$, is given by*

$$\mathbb{E}\left[\eta_k^{\omega}(\check{\mathscr{F}}_{\omega}) \bigotimes_2 \overline{\eta_l^{\omega}(\check{\mathscr{F}}_{\omega})}\right] = \begin{cases} -\kappa^2 \lambda_k \lambda_l (\lambda_k - \lambda_l)^{-2} \left[\Pi_k \,\widetilde{\otimes}_{\mathsf{T}}\, \overline{\Pi_l} + \Pi_l \,\widetilde{\otimes}_{\mathsf{T}}\, \overline{\Pi_k}\right] & \text{if } k \neq l; \\ \kappa^2 \sum_{s \in J \setminus \{k\}} \lambda_k \lambda_s (\lambda_k - \lambda_s)^{-2} \left[\Pi_k \,\widetilde{\otimes}_{\mathsf{T}}\, \overline{\Pi_s} + \Pi_s \,\widetilde{\otimes}_{\mathsf{T}}\, \overline{\Pi_k}\right] & \text{if } k = l. \end{cases} \tag{3.7.8}$$

The covariance of the sample eigenvalues is given by

$$\operatorname{cov}\left(p_l^{\omega}(\check{\mathscr{F}}_{\omega}), p_k^{\omega}(\check{\mathscr{F}}_{\omega})\right) = \left(1 + \mathbf{1}_{\{0,\pi\}}(\omega)\right) \kappa^2 \delta_{lk} \lambda_l^2 m_l,$$

where $\kappa^2 = 2\pi \int_{\mathbb{R}} W^2(x) dx$.

Notice that the estimators of the eigenspaces are not asymptotically independent, which is expected since they are constrained to be mutually orthogonal.

Before turning to the proof of the Theorem, let us illustrate the above expression of the covariance by an example. Let $\omega \in (0,\pi)$, $L = \{1\}$, $m_1(\omega) = 1$, and let

$$\mathscr{C} = \mathbb{E}\left[\eta_1^{\omega}(\check{\mathscr{F}}_{\omega}) \bigotimes_2 \eta_1^{\omega}(\check{\mathscr{F}}_{\omega})\right]$$

denote the covariance operator of $\eta_1^{\omega}(\check{\mathscr{F}}_{\omega})$.

If $\psi_i, \psi_j, \psi_k, \psi_l \in L^2([0,1], \mathbb{C})$, we have

$$\langle \mathscr{C}(\psi_i \otimes_2 \psi_j), \psi_k \otimes_2 \psi_l \rangle_{\mathscr{S}_2} = \kappa^2 \sum_{s \in J(\omega)\setminus\{1\}} \frac{\lambda_1 \lambda_s}{(\lambda_s - \lambda_1)^2} \Big[\langle \varphi_1, \psi_l \rangle \langle \psi_i, \varphi_1 \rangle \langle \varphi_s, \psi_j \rangle \langle \psi_k, \varphi_s \rangle + \\ + \langle \varphi_s, \psi_l \rangle \langle \psi_j, \varphi_s \rangle \langle \varphi_1, \psi_j \rangle \langle \psi_k, \varphi_1 \rangle \Big],$$

where we wrote $\Pi_s(\omega) = \varphi_s \otimes_2 \varphi_s$, $\lambda_j = \lambda(\omega)$. In particular, the variance of $\left\langle \eta_1^{\omega}(\check{\mathscr{F}}_{\omega})\psi_i, \psi_j \right\rangle$ is given by

$$\langle \mathscr{C}(\psi_i \otimes_2 \psi_j), \psi_i \otimes_2 \psi_j \rangle_{\mathscr{S}_2} = \kappa^2 \sum_{s \in J(\omega)\setminus\{1\}} \frac{\lambda_1 \lambda_s}{(\lambda_s - \lambda_1)^2} \Big[|\langle \varphi_1, \psi_i \rangle \langle \varphi_s, \psi_j \rangle|^2 + |\langle \varphi_s, \psi_i \rangle \langle \varphi_1, \psi_j \rangle|^2 \Big]. \tag{3.7.9}$$

Proof of Theorem 3.7.3. The proof rests on the adaptation of Theorem 1.3 of Mas & Menneteau (2003) to our case (we want the convergence in distribution of the sample eigenvalues and eigenprojectors, jointly in a finite number of frequencies), and therefore we give only a sketch. For $l \geq 1$, we denote by $\mathscr{S}^l$ the l-fold product space $\mathscr{S}_2(H) \times \cdots \times \mathscr{S}_2(H)$, equipped with the norm

$$\|(A_1, \ldots, A_l)\|_{\mathscr{S}^l} = \max_{j=1,\ldots,l} \left|\!\left|\!\left| A_j \right|\!\right|\!\right|_2,$$

and equip $\mathbb{C}^l$ with the norm

$$|(\alpha_1, \ldots, \alpha_l)|_{\infty} = \max_{j=1,\ldots,l} |\alpha_j|.$$

We now endow the space $\mathscr{S}^l \times \mathbb{C}^l$ with the norm

$$\|(A_1, \ldots, A_l, \alpha_1, \ldots, \alpha_l)\|_* = \max_{j=1,\ldots,l} \{ \left|\!\left|\!\left| A_j \right|\!\right|\!\right|_2, |\alpha_j| \}.$$

To simplify the notation, we assume that $L_i = L$, $|L| = l$, for all $i = 1, \ldots, K$. Defining the bounded linear operator $\Phi : \mathscr{S}^K \to \mathscr{S}^{Kl} \times \mathbb{C}^{Kl}$ by

$$\Phi(A_1, \ldots, A_K) = \big(\eta_L(A_1), \ldots, \eta_L(A_K), p_L(A_1), \ldots, p_L(A_K)\big),$$

we show that

$$\sqrt{TB_T} \left[\left\{ \Pi_{j,T}(\omega_i) - \Pi_j(\omega_i) \right\}_{j \in L; i=1,\ldots,K}, \left\{ \sum_{s \in I_j(\omega_i)} [\mu_{s,T}(\omega_i) - \lambda_j(\omega_i)] \right\}_{j \in L; i=1,\ldots,K} \right] \\ = \Phi(\sqrt{TB_T}\{\mathscr{F}_{\omega_i}^{(T)} - \mathscr{F}_{\omega_i}\}_{i=1,\ldots,K}) + \sqrt{TB_T}\mathscr{R},$$

where $\mathscr{R} = \big(\{R_{L,T}(\omega_i)\}_{i=1,\ldots,K}, \{r_{L,T}(\omega_i)\}_{i=1,\ldots,K}\big)$, where $R_{L,T}(\omega_i), r_{L,T}(\omega_i)$ are given by Mas & Menneteau (2003, Proposition 2.3). The proof is

completed by showing that $\sqrt{TB_T}\mathscr{R} \xrightarrow{p} 0$—which is equivalent to showing that $R_{L,T}(\omega_i) \xrightarrow{p} 0$ and $r_{L,T}(\omega_i) \xrightarrow{p} 0$ for all $i = 1,\dots,K$—and applying the continuous mapping Theorem. The determination of the covariance structure of the limiting random elements is given separately in Section 3.7.6. □

3.7.4 Some Comments on the Estimation of Eigenspace Projectors

Let us give a short comment on the estimation of the projectors $\Pi_j(\omega)$. Fix j and ω, and for simplicity let Π denote $\Pi_j(\omega)$, whose associated eigenvalue λ is assumed to have multiplicity one. We thus have $\Pi = \varphi \otimes_2 \varphi$, where φ is an eigenfunction associated to the eigenvalue λ of the operator $\mathscr{F}_\omega$. Recall that for any $\alpha \in \mathbb{C}$ of modulus one, $\alpha\varphi$ is also an eigenfunction with eigenvalue λ, so the eigenfunctions are not uniquely defined. We talk about φ whereas the true object we are dealing with is a representative of the set $\{\alpha\varphi : |\alpha| = 1\}$. The eigenprojectors do not have this problem; they are uniquely defined since

$$(\alpha\varphi) \otimes_2 (\alpha\varphi) = \alpha\overline{\alpha}\varphi \otimes_2 \varphi = \varphi \otimes_2 \varphi = \Pi.$$

Let us now address a question that will occur in practice. Given two estimators $\hat{\varphi}^1$ and $\hat{\varphi}^2$ of φ, *how can we construct a good estimator of* Π*?* Two obvious approaches are possible: we could either take $\hat{\Pi} = \frac{1}{2}(\hat{\varphi}^1 \otimes_2 \hat{\varphi}^1 + \hat{\varphi}^2 \otimes_2 \hat{\varphi}^2)$, or we could first estimate φ by $\tilde{\varphi} = \frac{1}{2}(\hat{\varphi}^1 + \hat{\varphi}^2)$ and then estimate Π by $\hat{\Pi} = \frac{1}{2}(\tilde{\varphi} \otimes_2 \tilde{\varphi})$. Our point is that the first estimator is better, because it is invariant to the choice of representatives $\hat{\varphi}^1, \hat{\varphi}^2$. However, the second estimator doesn't satisfy this property. To see this, let α, β be complex numbers of modulus one. If we choose $\tilde{\varphi} = \frac{1}{2}(\alpha\hat{\varphi}^1 + \beta\hat{\varphi}^2)$, the conjugate-bilinearity of the tensor product yields

$$\frac{1}{2}(\tilde{\varphi} \otimes_2 \tilde{\varphi}) = \frac{1}{4}\left(\hat{\varphi}^1 \otimes_2 \hat{\varphi}^1 + \hat{\varphi}^2 \otimes_2 \hat{\varphi}^2 + \alpha\overline{\beta}\hat{\varphi}^1 \otimes_2 \hat{\varphi}^2 + \overline{\alpha}\beta\hat{\varphi}^2 \otimes_2 \hat{\varphi}^1\right),$$

which depends on α and β. Thus when trying to estimate Π from a sample of empirical eigenfunctions $\hat{\varphi}^i$, $i = 1,\dots,n$, we shall use

$$\hat{\Pi} = \frac{1}{n}\left(\sum_{i=1}^{n} \hat{\varphi}^i \otimes_2 \hat{\varphi}^i\right).$$

Applying this *invariance principle* to the plug-in estimator of the variance expression obtained in (3.7.9), we see that the expressions of the form $|\langle\varphi, \psi\rangle|^2$ should be estimated by

$$\frac{\left|\langle\hat{\varphi}^1, \hat{\psi}^1\rangle\right|^2 + \left|\langle\hat{\varphi}^2, \hat{\psi}^2\rangle\right|^2}{2}.$$

3.7.5 *Karhunen–Loève-type Expansions for the Asymptotic Sample Spectral Density Operators*

In this section, we derive Karhunen–Loève-type expansions for the random element $\check{\mathscr{F}}_\omega$, to which the rescaled sample spectral density operator converges in distribution. By a Karhunen–Loève-type expansion, we mean a decomposition of the random element into a sum of orthogonal deterministic functions, multiplied by uncorrelated random variables. Expansions of the sort will be useful in Section 3.7.
Recall that Remark 3.6.6 tells us that $\check{\mathscr{F}}_\omega$ is either a complex Gaussian random element, with covariance operator

$$\mathscr{G} = \kappa^2 \cdot C \widetilde{\bigotimes}_2 C, \quad \omega \in (0,\pi), \tag{3.7.10}$$

and relation operator

$$\mathscr{C} = \kappa^2 \cdot C \widetilde{\bigotimes}_{\mathsf{T}} \overline{C}, \quad \omega \in (0,\pi), \tag{3.7.11}$$

or a real Gaussian random element with covariance operator

$$\mathscr{C} = \kappa^2 \cdot [C \widetilde{\bigotimes}_2 C + C \widetilde{\bigotimes}_{\mathsf{T}} C], \quad \omega \in \{0,\pi\} \tag{3.7.12}$$

where $\kappa^2 = 2\pi \int_{\mathbb{R}} W(\omega)^2 d\omega$, C is a nuclear operator on the complexified Hilbert space H in the first case, and on the real Hilbert space $H_{\mathbb{R}}$ in the second case.
In this Section, we let $C = \sum_i \mu_i \varphi_i \otimes_2 \varphi_i$ be the singular value decomposition of C.

Lemma 3.7.4.
Let H be a complexified Hilbert space, and C be a nuclear and self-adjoint operator on H, with spectral decomposition $C = \sum_i \mu_i \varphi_{ii}$, where $\varphi_{ij} = \varphi_i \otimes_2 \varphi_j$. If Y is a complex Gaussian random element on $\mathscr{S}_2(H)$ with mean 0*, covariance operator*

$$\mathscr{G} = \mathbb{E}[Y \otimes_2 Y] = C \widetilde{\bigotimes}_2 C$$

and relation operator

$$\mathscr{C} = \mathbb{E}\left[Y \otimes_2 \overline{Y}\right] = C \widetilde{\bigotimes}_{\mathsf{T}} \overline{C},$$

Then the following expansion is a Karhunen–Loève-type expansion for Y:

$$Y = \sum_i \xi_i \varphi_{ii} + \sum_{i<j} \left[\xi_{ij} e_{ij} + \zeta_{ij} \mathbf{i} \tilde{e}_{ij}\right], \tag{3.7.13}$$

where the convergence holds in $L^2(\Omega, \mathscr{S}_2(H), \mathbb{P})$,

$$e_{ij} = 2^{-1/2}(\varphi_{ij} + \varphi_{ji}),$$

$$\tilde{e}_{ij} = 2^{-1/2}(\varphi_{ij} - \varphi_{ji}),$$

and

$$\{\xi_i\}_i \cup \{\xi_{ij}\}_{i<j} \cup \{\zeta_{ij}\}_{i<j}$$

are independent real *Gaussian random variables, defined by*

$$\xi_i = \langle Y, \varphi_{ii} \rangle_{\mathscr{S}_2}, \quad \xi_{ij} = \langle Y, e_{ij} \rangle_{\mathscr{S}_2}, \quad \zeta_{ij} = -\mathbf{i}\langle Y, \tilde{e}_{ij} \rangle_{\mathscr{S}_2}. \tag{3.7.14}$$

They have mean zero and variance

$$\operatorname{var}(\xi_i) = \mu_i^2, \qquad \operatorname{var}(\xi_{ij}) = \operatorname{var}(\zeta_{ij}) = \mu_i \mu_j, \quad i \neq j.$$

Notice that we have *not* assumed that Y takes self-adjoint values in the statement, but that the Lemma tells us that Y takes necessarily self-adjoint values:

Corollary 3.7.5. *If Y satisfies the conditions of Lemma 3.7.4, then $Y = Y^\dagger$ almost surely.*

Proof. Taking the adjoint of (3.7.13) yields the result. □

We also point out that the random element Y can also be written in the following form, which will be useful later on.

Remark 3.7.6. *Under the conditions of Lemma 3.7.4, we can write*

$$Y = \sum_{i,j=1}^{\infty} \eta_{ij} \varphi_{ij},$$

where the $(\eta_{ij})_{i,j\geq 1}$ are all random variables with mean zero. The random variables η_{ii} are real Gaussian random variables, with variance μ_i^2, and the variables $\{\eta_{ij},\ i \neq j\}$ are complex circular Gaussian random variables, i.e., $\Re(\eta_{ij})$ and $\Im(\eta_{ij})$ are independent real Gaussian random variables, with variance

Circular Gaussian random variable are defined in Definition 3.12.1 on page 134; $\Re(\cdot)$ and $\Im(\cdot)$ denote the real and imaginary parts of a complex number

$$\operatorname{var}(\Re(\eta_{ij})) = \operatorname{var}(\Im(\eta_{ij})) = \mu_i \mu_j / 2, \quad i \neq j.$$

Furthermore, they satisfy

$$\eta_{ij} = \overline{\eta_{ji}}, \quad i \neq j,$$

and the following random variables are all independent:

$$\{\eta_{ij} : 1 \leq i \leq j\}.$$

Proof of Lemma 3.7.4. We assume, without loss of generality, that $\{\varphi_j : j \geq 1\}$ is an orthonormal basis of H. We want to look for a Karhunen–Loève-type expansion for Y. The idea behind a Karhunen–Loève-type expansion is to

find a sequence of unit length elements (vectors) $(\phi_n)_{n\geq 1} \subset \mathcal{S}_2(H)$ such that the scores $(\xi_n)_{n\geq 1}, \xi_n = \langle Y, \phi_n \rangle_{\mathcal{S}_2}$, are all strongly uncorrelated, with $\mathrm{var}(\xi_n) \geq \mathrm{var}(\xi_m)$ if $n > m$. In our case, strong uncorrelatedness translates into the following conditions, for $n \neq m$:

the definition of strong uncorrelatedness is given in Definition 3.12.3 on page 135

$$\mathbb{E}\left[\langle Y, \phi_n \rangle_{\mathcal{S}_2} \overline{\langle Y, \phi_m \rangle_{\mathcal{S}_2}}\right] = \langle \mathcal{G}\phi_m, \phi_n \rangle = 0, \tag{3.7.15}$$

$$\mathbb{E}\left[\langle Y, \phi_n \rangle_{\mathcal{S}_2} \langle Y, \phi_m \rangle_{\mathcal{S}_2}\right] = \left\langle \mathcal{C}\overline{\phi_m}, \phi_n \right\rangle = 0, \tag{3.7.16}$$

for all $n \neq m \geq 1$, and the variance condition translated to

$$\mathbb{E}|\langle Y, \phi_n \rangle|^2 \geq \mathbb{E}|\langle Y, \phi_m \rangle|^2, \quad n > m, \tag{3.7.17}$$

or equivalently,

$$\langle \mathcal{G}\phi_n, \phi_n \rangle \geq \langle \mathcal{G}\phi_m, \phi_m \rangle, \quad n > m. \tag{3.7.18}$$

To simplify notation, we shall write $\Phi = \phi_n$, $\Psi = \phi_m$. Assuming without loss of generality that $(\varphi_{ij})_{i,j\geq 1}$ is an orthonormal basis of $\mathcal{S}_2(H)$, we write Φ, Ψ in this basis:

$$\Phi = \sum_{i,j\geq 1} c_{ij}\varphi_{ij},$$
$$\Psi = \sum_{i,j\geq 1} d_{kl}\varphi_{kl},$$

Here and in the following, sums will all be over $\{1, 2, \ldots\}$, e.g. $\sum_{i,j}$ will mean $\sum_{i,j\geq 1}$. The following fact, justified by Proposition A.2.10, will be useful:

$$\varphi_{ij}\varphi_{kl} = (\varphi_i \otimes_2 \varphi_j)(\varphi_k \otimes_2 \varphi_l) = \delta_{jk}\varphi_{il}.$$

Since

$$\begin{aligned}
\left\langle (C \widetilde{\bigotimes}_2 C)\Phi, \Psi \right\rangle &= \langle C\Phi C, \Psi \rangle \\
&= \left\langle \left(\sum_r \mu_j \varphi_{rr}\right)\left(\sum_{i,j} c_{ij}\varphi_{ij}\right)\left(\sum_s \mu_s \varphi_{ss}\right), \Psi \right\rangle \\
&= \left\langle \sum_{r,i,j,s} \mu_r \mu_s c_{ij} \delta_{ri}\delta_{js}\varphi_{rs}, \Psi \right\rangle \\
&= \left\langle \sum_{i,j} \mu_i \mu_j c_{ij}\varphi_{ij}, \sum_{k,l} d_{kl}\varphi_{kl} \right\rangle \\
&= \sum_{i,j,k,l} \mu_i \mu_j c_{ij} \overline{d_{kl}} \delta_{ik}\delta_{jl} \\
&= \sum_{i,j} \mu_i \mu_j c_{ij} \overline{d_{ij}},
\end{aligned}$$

equation (3.7.15) becomes

$$\sum_{i,j} \mu_i \mu_j c_{ij} \overline{d_{ij}} = 0. \tag{3.7.19}$$

Similarly, since

$$\begin{aligned}
\left\langle (C \widetilde{\bigotimes}_{\mathsf{T}} \overline{C}) \overline{\Phi}, \Psi \right\rangle &= \left\langle C \overline{\Phi}^{\mathsf{T}} \overline{C}^{\mathsf{T}}, \Psi \right\rangle \\
&= \left\langle C \Phi^{\dagger} C, \Psi \right\rangle \\
&= \left\langle \left(\sum_r \mu_j \varphi_{rr} \right) \left(\sum_{i,j} \overline{c_{ji}} \varphi_{ij} \right) \left(\sum_s \mu_s \varphi_{ss} \right), \Psi \right\rangle \\
&= \sum_{i,j} \mu_i \mu_j \overline{c_{ji}} \overline{d_{ij}},
\end{aligned}$$

using the fact that the eigenvalues μ_i are all real, equation (3.7.16) becomes

$$\sum_{i,j} \mu_i \mu_j c_{ji} d_{ij} = 0. \tag{3.7.20}$$

The variance condition (3.7.18) translates into

$$\sum_{i,j} \mu_i \mu_j |c_{ij}|^2 \geq \sum_{i,j} \mu_i \mu_j |d_{ij}|^2, \tag{3.7.21}$$

and the unit length condition is

$$\sum_{i,j} |c_{ij}|^2 = 1 = \sum_{i,j} |d_{ij}|^2. \tag{3.7.22}$$

We now have all the equations necessary to find the vectors ϕ_n. We start with $\Phi = \phi_1$: it must maximize the variance of $\langle Y, \phi_1 \rangle$, i.e., maximize the left-hand side of (3.7.21) subject to condition (3.7.22). Since the eigenvalues of C are decreasing, $\mu_1 \geq \mu_2 \geq \cdots$, the solution is $c_{ij} = \delta_i j$, or in other words $\phi_1 = \varphi_{11}$ (if $\mu_1 = \mu_2$, the solution is not unique, but the final form of the expansion will be similar). Now to $\Psi = \phi_2$: by (3.7.19), we must have $d_{11} = 0$. To maximize the variance, we then see that we should take $\delta_{ij} = 0$ unless (i, j) is either equal to $(1,2)$ or equal to $(2,1)$. We see that the equations are not enough to determine $\Psi = \phi_2$; we therefore also consider $\Phi = \phi_3$ simultaneously. Similar considerations lead to $c_{ij} = 0$ unless (i, j) is either $(1,2)$ or $(2,1)$. Writing $\alpha = c_{12}$, $\beta = c_{21}$, $\gamma = d_{12}$, $\delta = d_{21}$, equations (3.7.19), (3.7.20), and (3.7.22) yield the system of equations

$$\begin{cases}
\alpha\delta + \gamma\beta & = 0, \\
\alpha\overline{\gamma} + \beta\overline{\delta} & = 0, \\
|\alpha|^2 + |\beta|^2 & = 1, \\
|\gamma|^2 + |\delta|^2 & = 1.
\end{cases}$$

Solving this system yields $\alpha = \overline{\beta}$, $\gamma = \overline{\delta}$, $\gamma = \pm \mathbf{i}\alpha$, with $|\alpha| = 1/\sqrt{2}$. Notice that the matrices (c_{ij}) and (d_{ij}) of any of these solutions are self-adjoint. We choose without loss of generality $\alpha = \beta = 1/\sqrt{2}$, $\gamma = \overline{\delta} = \mathbf{i}/\sqrt{2}$. Using the notation of the statement of the Lemma, we have $\phi_2 = e_{12}$ and $\phi_3 = \mathbf{i}\tilde{e}_{12}$. Now we write $\xi_{12} = \langle Y, e_{12}\rangle_{\mathscr{S}_2}$ and $\zeta_{12} = \langle Y, \mathbf{i}\tilde{e}_{12}\rangle_{\mathscr{S}_2}$. We can now continue with the same arguments to get the following ϕ_n's. If $\mu_1\mu_3 \geq \mu_2\mu_2$,

$$\phi_3 = e_{13}, \quad \phi_4 = \tilde{e}_{13}, \quad \phi_5 = \varphi_{22}.$$

Otherwise,

$$\phi_3 = \varphi_{22}, \phi_4 = e_{13}, \phi_5 = \tilde{e}_{13},$$

and so on. We define $\xi_i = \langle Y, \varphi_{ii}\rangle_{\mathscr{S}_2}$ for all $i \geq 1$, and also $\xi_{ij} = \langle Y, e_{ij}\rangle_{\mathscr{S}_2}$ and $\zeta_{ij} = \langle Y, \mathbf{i}\tilde{e}_{ij}\rangle_{\mathscr{S}_2}$ for all $j > i \geq 1$. By construction, these are all strongly orthogonal, and since Y is complex Gaussian, they are also independent. Furthermore, the covariance of ξ_{ij} is

$$\mathbb{E}\left[\xi_{ij}\overline{\xi_{ij}}\right] = \langle \mathscr{G} e_{ij}, e_{ij}\rangle_{\mathscr{S}_2} = \langle Ce_{ij}C, e_{ij}\rangle_{\mathscr{S}_2} = \mu_i\mu_j$$

and the relation of ξ_{ij} is

$$\mathbb{E}\left[\xi_{ij}\xi_{ij}\right] = \langle \mathscr{C}\overline{e_{ij}}, e_{ij}\rangle_{\mathscr{S}_2} = \left\langle Ce_{ij}^{\dagger}C, e_{ij}\right\rangle_{\mathscr{S}_2} = \langle Ce_{ij}C, e_{ij}\rangle_{\mathscr{S}_2} = \mu_i\mu_j.$$

Therefore, by Lemma 3.12.2, ξ_{ij} is real Gaussian with mean zero and variance $\mu_i\mu_j$. The exact same reasoning applies to ξ_i. For ζ_{ij}, things are similar: the covariance is

$$\mathbb{E}\left[\zeta_{ij}\overline{\zeta_{ij}}\right] = \langle \mathscr{G}\tilde{e}_{ij}, \tilde{e}_{ij}\rangle_{\mathscr{S}_2} = \langle C\tilde{e}_{ij}C, \tilde{e}_{ij}\rangle_{\mathscr{S}_2} = \mu_i\mu_j$$

and the relation of ζ_{ij} is

$$\mathbb{E}\left[\zeta_{ij}\zeta_{ij}\right] = \left\langle \mathscr{C}\overline{\mathbf{i}\tilde{e}_{ij}}, \mathbf{i}\tilde{e}_{ij}\right\rangle_{\mathscr{S}_2} = -\mathbf{i}\left\langle C(\mathbf{i}\tilde{e}_{ij})^{\dagger}C, \tilde{e}_{ij}\right\rangle_{\mathscr{S}_2} = -\langle C(-\tilde{e}_{ij})C, \tilde{e}_{ij}\rangle_{\mathscr{S}_2} = \mu_i\mu_j.$$

Therefore, ζ_{ij} is also real Gaussian. To complete the proof, notice that $e_{ij} + \tilde{e}_{ij} = \sqrt{2}\varphi_{ij}$ and $e_{ij} - \tilde{e}_{ij} = \sqrt{2}\varphi_{ji}$, and therefore

$$\{\varphi_{ii} : i \geq 1\} \cup \{\{e_{ij}, \tilde{e}_{ij}\} : 1 \leq i < j\}$$

forms a complete orthonormal basis of H, and

$$Y = \sum_i \xi_i\varphi_{ii} + \sum_{i<j}\left[\xi_{ij}e_{ij} + \zeta_{ij}\mathbf{i}\tilde{e}_{ij}\right]. \qquad \square$$

We now turn to the real case.

Lemma 3.7.7.

Let $H_{\mathbb{R}}$ be a real Hilbert space, and C be a nuclear and self-adjoint operator

on $H_{\mathbb{R}}$, with spectral decomposition

$$C = \sum_i \mu_i \varphi_{ii},$$

where $\varphi_{ij} = \varphi_i \otimes_2 \varphi_j$.
If Y is a Gaussian random element on $\mathscr{S}_2(H_{\mathbb{R}})$, with mean 0 and covariance operator

$$\mathscr{C} = C \widetilde{\bigotimes}_2 C + C \widetilde{\bigotimes}_{\mathsf{T}} C, \tag{3.7.23}$$

then

$$Y = \sum_i \xi_i \varphi_{ii} + \sum_{i<j} \xi_{ij} e_{ij}, \tag{3.7.24}$$

where the convergence holds in $L^2(\Omega, \mathscr{S}_2(H_{\mathbb{R}}), \mathbb{P})$, $e_{ij} = 2^{-1/2}(\varphi_{ij} + \varphi_{ji})$, and

$$(\xi_i)_i \cup (\xi_{ij})_{i<j}$$

are independent real *Gaussian random variables, defined by*

$$\xi_i = \langle Y, \varphi_{ii} \rangle_{\mathscr{S}_2}, \quad \xi_{ij} = \langle Y, e_{ij} \rangle_{\mathscr{S}_2}. \tag{3.7.25}$$

They have mean zero and variance

$$\operatorname{var}(\xi_i) = 2\mu_i^2, \quad \operatorname{var}(\xi_{ij}) = 2\mu_i \mu_j.$$

Proof. Assume without loss of generality that $\{\varphi_j : j \geq 1\}$ is an orthonormal basis of $H_{\mathbb{R}}$. Using results from Section A.3.4, we get

$$\mathscr{C} = \sum_i 2\mu_i^2 \varphi_{ii} \bigotimes_2 \varphi_{ii} + \sum_{i<j} 2\mu_i \mu_j e_{ij} \bigotimes_2 e_{ij}.$$

Since φ_{ii} and e_{ij} are orthonormal for all $i < j$, the random elements defined in (3.7.25) have the stated properties. Furthermore, since $\tilde{e}_{ij}$ is orthogonal to φ_{ii} and e_{ij}, for all $i < j$, $\mathscr{C}\tilde{e}_{ij} = 0$. Furthermore,

$$\{\varphi_{ii} : i \geq 1\} \cup \{\{e_{ij}, \tilde{e}_{ij}\} : 1 \leq i < j\}$$

is an orthonormal basis of $H_{\mathbb{R}}$, and therefore,

$$Y = \sum_{i<j} \left[\langle Y, \varphi_{ii} \rangle \varphi_{ii} + \langle Y, e_{ij} \rangle e_{ij} + \langle Y, \tilde{e}_{ij} \rangle \tilde{e}_{ij} \right] = \sum_i \xi_i \varphi_{ii} + \sum_{i<j} \xi_{ij} \varphi_{ij},$$

and the proof is complete. □

Similarly to the complex case, although we have *not* assumed in the statement of the Lemma that Y takes self-adjoint values, it turns out that it must take self-adjoint values:

Corollary 3.7.8. *If Y satisfies the conditions of Lemma 3.7.7, then $Y = Y^\dagger$ almost surely.*

Proof. Taking the adjoint of (3.7.24) yields the result. □

We can also re-write the Karhunen–Loève expansion of Y in the following form, which will be useful later on:

Remark 3.7.9. *Under the conditions of Lemma 3.7.7, we can write*

$$Y = \sum_{i,j=1}^{\infty} \eta_{ij}\varphi_{ij},$$

where the $(\eta_{ij})_{i,j\geq 1}$ *are all real Gaussian random variables, with mean zero, and variance*

$$\operatorname{var}(\eta_{ii}) = 2\mu_i^2, \quad \operatorname{var}(\eta_{ij}) = \mu_i\mu_j,\ i \neq j,$$

and $\eta_{ij} = \eta_{ji}$ *for all* $i \neq j$*. Furthermore,*

$$\{\eta_{ij} : 1 \leq i \leq j\}$$

are all independent.

3.7.6 Computation of Asymptotic Covariances

In this Section, we determine the asymptotic covariance of estimators of the eigenprojections and eigenvalues of $\mathscr{F}_\omega$, as stipulated in Theorem 3.7.3. We will use the same notation as in Section 3.7. Simple calculations yield

$$\eta_k^\omega(\varphi_i^\omega \otimes_2 \varphi_j^\omega) = E_k^\omega(i,j)\varphi_i^\omega \otimes_2 \varphi_j^\omega,$$

where E_k^ω is defined by

$$E_k^\omega(i,j) = \mathbf{1}_{\{k\in J(\omega)\}} \cdot \sum_{s\in J(\omega)\setminus\{k\}} \left[\mathbf{1}_{\{(i\in I_k(\omega))\cap(j\in I_s(\omega))\}} + \mathbf{1}_{\{(i\in I_s(\omega))\cap(j\in I_k(\omega))\}}\right](\lambda_k(\omega) - \lambda_s(\omega))^{-1}. \quad (3.7.26)$$

From now on, we suppress dependence on the frequency ω, for tidiness. Equation 3.7.26 says that $E_k(i,j) = (\lambda_k - \lambda_s)^{-1}$ if for some $s \neq k$; $s,k \in J$, either

$$i \in I_k \ \&\ j \in I_s$$

or

$$j \in I_k \ \&\ i \in I_s,$$

and $E_k(i,j) = 0$ otherwise. Notice that $E_k(i,i) = 0$, and $E_k(i,j) = E_k(j,i)$. Thus $\eta_k(e_{ij}) = E_k(i,j)e_{ij}$ and $\eta_k(\tilde{e}_{ij}) = E_k(i,j)\tilde{e}_{ij}$. The following Lemma gives some other useful properties of $E_k(i,j)$.

Recall the definition of e_{ij} and $\tilde{e}_{ij}$ on page 106

Lemma 3.7.10.
Let

$$\{c(i,j) : i,j \in J(\omega)\} \subset H$$

be a bounded sequence of elements of a separable Hilbert space H satisfying

$$c(i,j) = c(j,i), \quad i,j \in J(\omega).$$

Then,

$$\sum_{1\le i<j<\infty} E_k(i,j)^2 c(i,j)\mu_i\mu_j = \sum_{s\in J(\omega)\setminus\{k\}} (\lambda_k-\lambda_s)^{-2}\lambda_k\lambda_s \sum_{i\in I_k}\sum_{j\in I_s} c(i,j)\lambda_i\lambda_j$$

$$\mathbf{1}_{\{k,l\in J(\omega)\}} \sum_{1\le i<j<\infty} E_k(i,j)E_l(i,j)c(i,j)\mu_i\mu_j = -(\lambda_k-\lambda_l)^{-2}\lambda_k\lambda_l \sum_{i\in I_k}\sum_{j\in I_l} c(i,j)$$

Proof. Notice that

$$\begin{aligned}
E_k(i,j)^2 &= \mathbf{1}_{\{k\in J(\omega)\}} \cdot \sum_{s\in J(\omega)\setminus\{k\}} \sum_{t\in J(\omega)\setminus\{k\}} (\lambda_k-\lambda_s)^{-1}(\lambda_k-\lambda_t)^{-1}\times \\
&\qquad \times \left[\mathbf{1}_{\{(i\in I_k)\cap(j\in I_s)\}} + \mathbf{1}_{\{(i\in I_s)\cap(j\in I_k)\}}\right]\left[\mathbf{1}_{\{(i\in I_k)\cap(j\in I_t)\}} + \mathbf{1}_{\{(i\in I_t)\cap(j\in I_k)\}}\right] \\
&= \mathbf{1}_{\{k\in J(\omega)\}} \cdot \sum_{s\in J(\omega)\setminus\{k\}} (\lambda_k-\lambda_s)^{-2}\left[\mathbf{1}_{\{(i\in I_k)\cap(j\in I_s)\}} + \mathbf{1}_{\{(i\in I_s)\cap(j\in I_k)\}}\right].
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum_{1\le i<j} E_k(i,j)^2 c(i,j)\mu_i\mu_j &= \mathbf{1}_{\{k\in J(\omega)\}} \cdot \sum_{s\in J(\omega)\setminus\{k\}} (\lambda_k-\lambda_s)^{-2} \sum_{i=1}^{\infty}\sum_{j=i+1}^{\infty} \left[\mathbf{1}_{\{(i\in I_k)\cap(j\in I_s)\}} + \mathbf{1}_{\{(i\in I_s)\cap(j\in I_k)\}}\right] c(i,j)\mu_i\mu_j \\
&= \mathbf{1}_{\{k\in J(\omega)\}} \cdot \sum_{s>k;s\in J(\omega)} (\lambda_k-\lambda_s)^{-2} \sum_{i\in I_k}\sum_{j\in I_s} c(i,j)\mu_i\mu_j \\
&\quad + \mathbf{1}_{\{k\in J(\omega)\}} \cdot \sum_{s<k;s\in J(\omega)} (\lambda_k-\lambda_s)^{-2} \sum_{i\in I_s}\sum_{j\in I_k} c(i,j)\mu_i\mu_j \\
&= \mathbf{1}_{\{k\in J(\omega)\}} \cdot \sum_{s>k;s\in J(\omega)} (\lambda_k-\lambda_s)^{-2} \sum_{i\in I_k}\sum_{j\in I_s} c(i,j)\lambda_k\lambda_s \\
&\quad + \mathbf{1}_{\{k\in J(\omega)\}} \cdot \sum_{s<k;s\in J(\omega)} (\lambda_k-\lambda_s)^{-2} \sum_{i\in I_k}\sum_{j\in I_s} c(j,i)\lambda_s\lambda_k \\
&= \mathbf{1}_{\{k\in J(\omega)\}} \cdot \sum_{s\in J(\omega)\setminus\{k\}} (\lambda_k-\lambda_s)^{-2}\lambda_k\lambda_s \sum_{i\in I_k}\sum_{j\in I_s} c(i,j).
\end{aligned}$$

This proves the first equality. For the second one, if $i<j$,

$$\begin{aligned}
E_k(i,j)E_l(i,j) &= \mathbf{1}_{\{k,l\in J(\omega)\}} \cdot \sum_{s\in J(\omega)\setminus\{k\}} \sum_{t\in J(\omega)\setminus\{l\}} (\lambda_k-\lambda_s)^{-1}(\lambda_l-\lambda_t)^{-1}\times \\
&\qquad \times \left[\mathbf{1}_{\{(i\in I_k)\cap(j\in I_s)\}} + \mathbf{1}_{\{(i\in I_s)\cap(j\in I_k)\}}\right]\left[\mathbf{1}_{\{(i\in I_l)\cap(j\in I_t)\}} + \mathbf{1}_{\{(i\in I_t)\cap(j\in I_l)\}}\right] \\
&= \mathbf{1}_{\{k,l\in J(\omega)\}} \cdot \sum_{s\in J(\omega)\setminus\{k\}} \sum_{t\in J(\omega)\setminus\{l\}} (\lambda_k-\lambda_l)^{-1}(\lambda_l-\lambda_k)^{-1}\times \\
&\qquad \times \left[\delta_{k,t}\delta_{s,l}\mathbf{1}_{\{(i\in I_k)\cap(j\in I_l)\}} + \delta_{k,t}\delta_{s,l}\mathbf{1}_{\{(i\in I_l)\cap(j\in I_k)\}}\right] \\
&= -\mathbf{1}_{\{k,l\in J(\omega)\}} \cdot (\lambda_k-\lambda_l)^{-2} \cdot \left[\mathbf{1}_{\{(i\in I_k)\cap(j\in I_l)\}} + \mathbf{1}_{\{(i\in I_l)\cap(j\in I_k)\}}\right]
\end{aligned}$$

Therefore, if $k < l$,

$$\begin{aligned}
&\sum_{i<j} E_k(i,j)E_l(i,j)c(i,j)\mu_i\mu_j \\
&= -\mathbf{1}_{\{k,l\in J(\omega)\}} \cdot (\lambda_k - \lambda_l)^{-2} \cdot \sum_{i<j} \left[\mathbf{1}_{\{(i\in I_k)\cap(j\in I_l)\}} + \mathbf{1}_{\{(i\in I_l)\cap(j\in I_k)\}} \right] c(i,j)\mu_i\mu_j \\
&= -\mathbf{1}_{\{k,l\in J(\omega)\}} \cdot (\lambda_k - \lambda_l)^{-2}\lambda_k\lambda_l \cdot \sum_{i\in I_k} \sum_{j\in I_l} c(i,j)
\end{aligned}$$

because $i \in I_l \implies j \notin I_k$, since $I_k < I_l$. Interchanging k and l yields the result same result for $k > l$.

Finally, notice that there is no problem with the convergence of the series since all their absolute sums are all bounded by

$$\lambda_k m_k \sup \left\| c(i,j) \right\| \cdot \max(\lambda_k - \lambda_{k+1}, \lambda_{k-1} - \lambda_k)^{-2} \cdot |||\mathscr{F}_\omega|||_1 < \infty. \qquad \square$$

If Y is a random element of $\mathscr{S}_2(H)$ of the form (3.7.13), we have (omitting the ωs)

$$\eta_k(Y) = \sum_{i<j} E_k(i,j) \left[\xi_{ij} e_{ij} + \mathbf{i}\zeta_{ij}\tilde{e}_{ij} \right], \tag{3.7.27}$$

and thus

$$\begin{aligned}
\eta_k(Y) \bigotimes_2 \eta_l(Y) = \sum_{i<j} \sum_{s<t} E_k(i,j)E_l(s,t) \times \\
\times \left[\xi_{ij}\xi_{st} e_{ij} \bigotimes_2 e_{st} - \mathbf{i}\xi_{ij}\zeta_{st} e_{ij} \bigotimes_2 \tilde{e}_{st} + \mathbf{i}\zeta_{ij}\xi_{st}\tilde{e}_{ij} \bigotimes_2 e_{st} + \zeta_{ij}\zeta_{st}\tilde{e}_{ij} \bigotimes_2 \tilde{e}_{st} \right].
\end{aligned}$$

The equality comes from the fact that the cross-terms cancel out, and the properties of the tensor products $.\otimes_2.$ and $.\,\widetilde{\otimes}_2\,.$, see Section A.3.4. The covariance operator between $\eta_k(Y)$ and $\eta_l(Y)$ is therefore given by

$$\begin{aligned}
\mathbb{E}\left[\eta_k(Y) \bigotimes_2 \eta_l(Y) \right] &= \sum_{i<j} E_k(i,j)E_l(i,j) \left[\operatorname{var}\left(\xi_{ij}\right) e_{ij} \bigotimes_2 e_{ij} + \operatorname{var}\left(\zeta_{ij}\right) \tilde{e}_{ij} \bigotimes_2 \tilde{e}_{ij} \right] \\
&= \sum_{i<j} E_k(i,j)E_l(i,j)\mu_i\mu_j \left[\varphi_{ii} \,\widetilde{\bigotimes}_2\, \varphi_{jj} + \varphi_{jj} \,\widetilde{\bigotimes}_2\, \varphi_{ii} \right].
\end{aligned}$$

We now distinguish the cases $k = l$ and $k \neq l$. If $k = l$, Lemma 3.7.10 yields

$$\begin{aligned}
\mathbb{E}\left[\eta_k(Y)\bigotimes_2\eta_k(Y)\right] &= \sum_{i<j}E_k(i,j)^2\mu_i\mu_j\left[\varphi_{ii}\,\widetilde{\bigotimes}_2\,\varphi_{jj}+\varphi_{jj}\,\widetilde{\bigotimes}_2\,\varphi_{ii}\right].\\
&= \sum_{s\in J\setminus\{k\}}\sum_{i\in I_k}\sum_{j\in I_s}\lambda_k\lambda_s(\lambda_k-\lambda_s)^{-2}\left[\varphi_{ii}\,\widetilde{\bigotimes}_2\,\varphi_{jj}+\varphi_{jj}\,\widetilde{\bigotimes}_2\,\varphi_{ii}\right]\\
&= \sum_{s\in J\setminus\{k\}}\lambda_k\lambda_s(\lambda_k-\lambda_s)^{-2}\left[\left(\sum_{i\in I_k}\varphi_{ii}\right)\widetilde{\bigotimes}_2\left(\sum_{j\in I_s}\varphi_{jj}\right)+\left(\sum_{j\in I_s}\varphi_{jj}\right)\widetilde{\bigotimes}_2\left(\sum_{i\in I_k}\varphi_{ii}\right)\right]\\
&= \sum_{s\in J\setminus\{k\}}\lambda_k\lambda_s(\lambda_k-\lambda_s)^{-2}\left[\Pi_k\,\widetilde{\bigotimes}_2\,\Pi_s+\Pi_s\,\widetilde{\bigotimes}_2\,\Pi_k\right]
\end{aligned}$$

If $k\neq l$, Lemma 3.7.10 yields

$$\begin{aligned}
\mathbb{E}\left[\eta_k(Y)\bigotimes_2\eta_l(Y)\right] &= -\sum_{i\in I_k}\sum_{j\in I_l}\lambda_k\lambda_l(\lambda_k-\lambda_l)^{-2}\left[\varphi_{ii}\,\widetilde{\bigotimes}_2\,\varphi_{jj}+\varphi_{jj}\,\widetilde{\bigotimes}_2\,\varphi_{ii}\right]\\
&= -\lambda_k\lambda_l(\lambda_k-\lambda_l)^{2}\left[\Pi_k\,\widetilde{\bigotimes}_2\,\Pi_l+\Pi_l\,\widetilde{\bigotimes}_2\,\Pi_k\right]
\end{aligned}$$

Therefore, for $k,l\in J(\omega)$,

$$\mathbb{E}\left[\eta_k(Y)\bigotimes_2\eta_l(Y)\right]=\begin{cases}-\lambda_k\lambda_l(\lambda_k-\lambda_l)^2\left[\Pi_k\,\widetilde{\bigotimes}_2\,\Pi_l+\Pi_l\,\widetilde{\bigotimes}_2\,\Pi_k\right] & \text{if } k\neq l\\ \sum_{s\in J(\omega)\setminus\{k\}}\lambda_k\lambda_s(\lambda_k-\lambda_s)^2\left[\Pi_k\,\widetilde{\bigotimes}_2\,\Pi_s+\Pi_s\,\widetilde{\bigotimes}_2\,\Pi_k\right] & \text{if } k=l.\end{cases}$$

We now turn to the computation of the relation operator of a random element $Y\in\mathscr{S}_2(H)$ of the form (3.7.13). We have

$$\begin{aligned}
\eta_k(Y)\bigotimes_2\overline{\eta_l(Y)} &= \sum_{i<j}\sum_{s<t}E_k(i,j)E_l(s,t)\times\\
&\quad\times\left[\xi_{ij}\xi_{st}e_{ij}\bigotimes_2\overline{e_{st}}+\mathbf{i}\xi_{ij}\zeta_{st}e_{ij}\bigotimes_2\overline{\tilde{e}_{st}}+\mathbf{i}\zeta_{ij}\xi_{st}\tilde{e}_{ij}\bigotimes_2\overline{e_{st}}-\zeta_{ij}\zeta_{st}\tilde{e}_{ij}\bigotimes_2\overline{\tilde{e}_{st}}\right].
\end{aligned}$$

The relation operator between $\eta_k(Y)$ and $\eta_l(Y)$ is therefore given by

$$\begin{aligned}
\mathbb{E}\left[\eta_k(Y)\bigotimes_2\overline{\eta_l(Y)}\right] &= \sum_{i<j}E_k(i,j)E_l(i,j)\left[\operatorname{var}\left(\xi_{ij}\right)e_{ij}\bigotimes_2\overline{e_{ij}}-\operatorname{var}\left(\zeta_{ij}\right)\tilde{e}_{ij}\bigotimes_2\overline{\tilde{e}_{ij}}\right]\\
&= \sum_{i<j}E_k(i,j)E_l(i,j)\mu_i\mu_j\left[e_{ij}\bigotimes_2\overline{e_{ij}}-\tilde{e}_{ij}\bigotimes_2\overline{\tilde{e}_{ij}}\right]\\
&= \sum_{i<j}E_k(i,j)E_l(i,j)\mu_i\mu_j\left[\varphi_{ii}\,\widetilde{\bigotimes}_{\mathsf{T}}\,\overline{\varphi_{jj}}+\varphi_{jj}\,\widetilde{\bigotimes}_{\mathsf{T}}\,\overline{\varphi_{ii}}\right],
\end{aligned}$$

where we have used Proposition A.3.5.

Using Lemma 3.7.10, we get

$$\mathbb{E}\left[\eta_k(Y)\otimes_2\overline{\eta_l(Y)}\right]=\begin{cases}-\lambda_k\lambda_l(\lambda_k-\lambda_l)^{-2}\left[\Pi_k\widetilde{\otimes}_{\mathsf{T}}\overline{\Pi_l}+\Pi_l\widetilde{\otimes}_{\mathsf{T}}\overline{\Pi_k}\right], & k\neq l,\\ \sum_{s\in J(\omega)\setminus\{k\}}\lambda_k\lambda_s(\lambda_k-\lambda_s)^{-2}\left[\Pi_k\widetilde{\otimes}_{\mathsf{T}}\overline{\Pi_s}+\Pi_s\widetilde{\otimes}_{\mathsf{T}}\overline{\Pi_k}\right], & k=l.\end{cases}$$

Notice in particular that by Proposition A.3.5,

$$\left(\Pi_k\widetilde{\bigotimes}_{\mathsf{T}}\overline{\Pi_s}\right)^{\mathsf{T}}=\left(\Pi_s\widetilde{\bigotimes}_{\mathsf{T}}\overline{\Pi_k}\right),$$

which is not surprising since

$$\left(\eta_k(Y)\otimes_2\overline{\eta_l(Y)}\right)^{\mathsf{T}}=\eta_l(Y)\otimes_2\overline{\eta_k(Y)}.$$

If Y is of the form (3.7.24) (and in particular real valued), we obtain

$$\eta_k(Y)=\sum_{i<j}E_k(i,j)\xi_{ij}e_{ij}, \tag{3.7.28}$$

and therefore

$$\begin{aligned}\mathbb{E}\left[\eta_k(Y)\otimes_2\eta_l(Y)\right]&=\sum_{i<j}E_k(i,j)E_l(i,j)\left[\operatorname{var}\left(\xi_{ij}\right)e_{ij}\otimes_2 e_{ij}\right]\\&=\begin{cases}-\lambda_k\lambda_l(\lambda_k-\lambda_l)^{-2}\left[\Pi_k\,\widetilde{\otimes}_2\,\Pi_l+\Pi_l\,\widetilde{\otimes}_2\,\Pi_k+\Pi_k\widetilde{\otimes}_{\mathsf{T}}\Pi_l+\Pi_l\widetilde{\otimes}_{\mathsf{T}}\Pi_k\right], & k\neq l\\ \sum_{s\neq k}\lambda_k\lambda_s(\lambda_k-\lambda_s)^{-2}\left[\Pi_k\,\widetilde{\otimes}_2\,\Pi_s+\Pi_s\,\widetilde{\otimes}_2\,\Pi_k+\Pi_k\widetilde{\otimes}_{\mathsf{T}}\Pi_l+\Pi_l\widetilde{\otimes}_{\mathsf{T}}\Pi_k\right], & k=l,\end{cases}\end{aligned}$$

where we have used Proposition A.3.5 and the fact the $\varphi_i=\overline{\varphi_i}$ because Y is real-valued. Notice in particular that

$$\left(\Pi_k\widetilde{\bigotimes}_{\mathsf{T}}\Pi_l\right)^{\dagger}=\Pi_l\widetilde{\bigotimes}_{\mathsf{T}}\Pi_k.$$

We now turn our attention to the joint distribution of $p_l(Y)$ and $\eta_k(Y)$. Since the Karhunen–Loève expansion of $p_l(Y)$ is given by

$$p_l(Y)=\sum_{i\in I_l}\xi_i. \tag{3.7.29}$$

in both cases (3.7.13) and (3.7.24), we see directly from (3.7.27) and (3.7.28) that $p_l(Y)$ is independent of $\eta_k(Y)$ for any $l,k\in J$.

For the covariance of the estimated eigenvalues, using the Karhunen–Loève expansion of $p_l(Y)$, we get

$$\operatorname{cov}\left(p_l(Y),p_k(Y)\right)=\delta_{lk}\lambda_l^2\operatorname{Tr}(\Pi_l),$$

if Y is of the form (3.7.13), and

$$\operatorname{cov}\big(p_l(Y), p_k(Y)\big) = \delta_{lk} 2\lambda_l^2 \operatorname{Tr}(\Pi_l),$$

if Y is of the form (3.7.24). This finishes the computation of the covariances for the asymptotic distributions of the sample eigenvalues and sample eigenprojections.

3.8 The Effect of Discrete Observation

In practice, functional data are often observed on a discrete grid, subject to measurement error, and smoothing is employed to make the transition into the realm of smooth functions. This section considers the stability of the consistency of our estimator of the spectral density operator with respect to discrete observation of the underlying stationary functional process. Since our earlier results do not a priori require any smoothness of the functional data, except perhaps smoothness that is imposed by our mixing conditions, we consider a 'minimal' scenario where the curves are only assumed to be continuous in mean square. Under this weak assumption, we formalise the asymptotic discrete observation framework via observation on an increasingly dense grid subject to noise of decreasing variance (the so called "low noise" or "decreasing noise" setup, see e.g. Hall & Vial (2006)).

Let Γ be the grid $0 = \tau_1 < \tau_2 < \cdots < \tau_M < \tau_{M+1} = 1$ on $[0,1]$, with $M = M(T)$ being a function of T such that $M(T) \to \infty$ as $T \to \infty$, and

$$|\Gamma| = \sup_{j=1,\ldots,M} \tau_{j+1} - \tau_j \to 0, \quad M \to \infty.$$

We assume we are in a fixed design setup: we observe each curve X_t on the grid Γ (except possibly at τ_{M+1}), additively corrupted by measurement error, represented by random variables $\{\varepsilon_{tj}\}$:

$$y_{tj} = X_t(\tau_j) + \varepsilon_{tj}, \quad t = 0, \ldots, T-1;\ j = 1, \ldots M.$$

We shall make the following assumptions concerning the additive noise:

Condition 3.8.1. *The noise ε_{tj} satisfies the following conditions:*

1. *Independence with respect to X: the random variables $(\varepsilon_{tj})_{t,j}$ are all independent of the X_ts.*

2. *Uniformly bounded fourth moment: $\mathbb{E}\left[\varepsilon_{tj}^4\right] < c_1$ for all t, j, for some constant $c_1 = O(1)$.*

3. *Fourth order stochastic orthogonality across time :*

$$\begin{aligned}
&\mathbb{E}[\varepsilon_t.\varepsilon_s.] = \mathbb{E}[\varepsilon_t.]\,\mathbb{E}[\varepsilon_s.], && t \neq s,\\
&\mathbb{E}[\varepsilon_t.\varepsilon_t.\varepsilon_s.] = \mathbb{E}[\varepsilon_t.\varepsilon_t.]\,\mathbb{E}[\varepsilon_s.], && t \neq s,\\
&\mathbb{E}\left[\varepsilon_{t_1}.\varepsilon_{t_2}.\varepsilon_{t_3}.\right] = \mathbb{E}\left[\varepsilon_{t_1}.\right]\mathbb{E}\left[\varepsilon_{t_2}.\right]\mathbb{E}\left[\varepsilon_{t_3}.\right], && t_1, t_2, t_3 \text{ distinct},\\
&\mathbb{E}\left[\varepsilon_{t_1}.\varepsilon_{t_2}.\varepsilon_{t_3}.\varepsilon_{t_4}.\right] = \mathbb{E}\left[\varepsilon_{t_1}.\right]\mathbb{E}\left[\varepsilon_{t_2}.\right]\mathbb{E}\left[\varepsilon_{t_3}.\right]\mathbb{E}\left[\varepsilon_{t_4}.\right], && t_1, t_2, t_3, t_4 \text{ distinct},\\
&\mathbb{E}[\varepsilon_t.\varepsilon_t.\varepsilon_s.\varepsilon_s.] = \mathbb{E}[\varepsilon_t.\varepsilon_t.]\,\mathbb{E}[\varepsilon_s.\varepsilon_s.], && t \neq s,\\
&\mathbb{E}[\varepsilon_t.\varepsilon_t.\varepsilon_t.\varepsilon_s.] = \mathbb{E}[\varepsilon_t.\varepsilon_t.\varepsilon_t.]\,\mathbb{E}[\varepsilon_s.], && t \neq s,
\end{aligned}$$

where we have simplified the notation by writing $\varepsilon_{t.}$ instead of ε_{tj}. This means that the dots in each $\varepsilon_{t.}$ can be replaced by any j, provided we do the replacement on both sides of the equations, at the corresponding locations. For instance, the first equation really means

$$\mathbb{E}\left[\varepsilon_{tj}\varepsilon_{sk}\right] = \mathbb{E}\left[\varepsilon_{tj}\right]\mathbb{E}[\varepsilon_{sk}], \quad \forall s \neq t, \forall j, k.$$

Remark 3.8.2. *The reason we make no stronger assumptions on the noise will be explained in Remark 3.8.4. Meanwhile, notice that*

1. *We do* not *make any assumption about the mean of the observation noise, nor did we assume it to be uncorrelated within a fixed time point: $(\varepsilon_{tj})_{j=1}^{m}$ may be correlated.*
2. *Assumption* 3 *of Conditions 3.8.1 could be replaced by the stronger assumptions of independence across time of the noise variables.*

Our goal is to show that our estimator of $\mathscr{F}_\omega^{(T)}$, when constructed on the basis of the y_{tj}'s, retains its strong consistency for the true spectral density operator. To construct our estimator on the basis of discrete observations, we use the following (naive) proxy of the true X_t:

$$_{\varepsilon,s}X_t(\tau) = y_{tj}, \quad \text{if } \tau_j \leq \tau < \tau_{j+1},$$

and define the step-wise version of X_t:

$$_sX_t(\tau) = X_t(\tau_j), \quad \text{if } \tau_j \leq \tau < \tau_{j+1}.$$

Just as the spectral density kernel estimator $f_\omega^{(T)}$ is a functional of the X_t's, we can define $_{\varepsilon,s}f_\omega^{(T)}$ and $_sf_\omega^{(T)}$, as the corresponding functionals of the $_{\varepsilon,s}X_t$'s, $_sX_t$, respectively. The same can also be done for f_ω, $\mathscr{F}_\omega^{(T)}$, $p_\omega^{(T)}$, $\widetilde{X}_\omega^{(T)}$. We then have the following stability result.

Theorem 3.8.3.
Assume that each r_t is continuous,

$$\sum_t \|r_t\|_\infty < \infty, \tag{3.8.1}$$

and that **C(1,2), C(1,4)** *and Conditions 3.8.1 hold. Let* $\gamma = \gamma(M,T)$ *be the upper bound on the standard deviation of the noise,*

$$\gamma = \gamma(M,T) = \sup_{j=1,\dots,M;t=0,\dots,T-1} \sqrt{\operatorname{var}\left(\varepsilon_{tj}\right)},$$

and $b = b(M,T)$ *be the upper bound on its bias,*

$$b = b(M,T) = \sup_{j=1,\dots,M;t=0,\dots,T-1} |\,\mathbb{E}\,\varepsilon_{tj}|.$$

Then, assuming $T \to \infty$, $B_T \to 0$, *we have*

$$\begin{aligned}\sup_{\omega\in[-\pi,\pi]} \mathbb{E}\left|\left|\left|{}_{\varepsilon,s}\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\right|\right|\right|_2^2 &= 8 \sup_{\omega\in[-\pi,\pi]} |||{}_s\mathscr{F}_\omega - \mathscr{F}_\omega|||_2^2 \\ &\quad + O(B_T^2) + O(T^{-2}) + O(T^{-1}B_T^{-2}) \\ &\quad + O(B_T^{-1})\left[O(T^2b^4) + O(T)O(b^2)O(\gamma^2+b^2)\right] \\ &\quad + O(B_T^{-1})\left[O(c_1 b) + O(\gamma^2)\right] \\ &\quad + O(B_T^{-1})\left[O(b^4) + O(\gamma^4)\right] \\ &\quad + O(TB_T)^{-1}O(c_1).\end{aligned}$$

Replacing the supremum norm by the mean, we get a slightly different bound (the changes are on the first two lines):

$$\begin{aligned}\int_{-\pi}^{\pi} \mathbb{E}\left|\left|\left|{}_{\varepsilon,s}\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\right|\right|\right|_2^2 d\omega &= 8\int_{-\pi}^{\pi} |||{}_s\mathscr{F}_\omega - \mathscr{F}_\omega|||_2^2 d\omega \\ &\quad + O(B_T^2) + O(T^{-2}) + O(T^{-1}B_T^{-1}) \\ &\quad + O(B_T^{-1})\left[O(T^2b^4) + O(T)O(b^2)O(\gamma^2+b^2)\right] \\ &\quad + O(B_T^{-1})\left[O(c_1 b) + O(\gamma^2)\right] \\ &\quad + O(B_T^{-1})\left[O(b^4) + O(\gamma^4)\right] \\ &\quad + O(TB_T)^{-1}O(c_1),\end{aligned}$$

Remark 3.8.4. *The reason that the statement of the Theorem is so complicated is its generality. Indeed, it relates the consistency of the estimator constructed from noisy observation to the rate at which the mean, variance, and fourth moment of the observation errors have to decrease, compared to the length* T *of the FTS, and the bandwidth parameter* B_T.

1. *Notice that the term* $\sup_{\omega\in[-\pi,\pi]} |||{}_s\mathscr{F}_\omega - \mathscr{F}_\omega|||_2^2$ *is* $o(1)$ *under the assumptions of the Theorem.*

2. *If* $\mathbb{E}\left[\varepsilon_{tj}\right] = 0$ *for all* t, j, *and* $TB_T^2 \to \infty$, $c_1 = O(1)$, *the error terms simplify to*

$$O(B_T^{-1})O(\gamma^2+\gamma^4).$$

This is essentially telling us that a small "noise variance to bandwidth"

ratio is sufficient to have consistent estimators, provided the errors have mean zero.

3. *One might want to opt for more sophisticated smoothing techniques to estimate the curves X_t from noisy observations. In this case, Theorem 3.8.3 can be applied. Indeed, if $\hat{X}_t$ is the estimated curve for X_t, we can take $y_{tj} = \hat{X}_t(\tau_j)$. Then the error terms become $\varepsilon_{tj} = \hat{X}_t(\tau_j) - X_t(\tau_j)$. The quantity b is therefore a uniform bound on the bias (conditionally on X) of the estimators $\hat{X}_t$, and γ^2 is a uniform bound on their conditional variance. Notice that the error terms for the same t will be correlated, but this is not an issue since the Theorem holds for arbitrary correlation of the errors within a time point. For a given smoothing method, the Theorem therefore tells us at what (uniform) rate the bias and the variance of the estimated curves have to decrease, relatively to B_T and T, in order to ensure the consistency of the sample spectral density operators based on noisy observations. (for instance, if $c_1 = O(1)$ but not $o(1)$, then the bias of the estimated curves should be uniformly $o(B_T)$). These conditions can then be used and give insight into how much to smooth the noisy observations; this will be investigated in future work.*

Proof of Theorem 3.8.3. By the triangle inequality,

$$\mathbb{E}\left|\left|\left|{}_{\varepsilon,s}\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega^{(T)}\right|\right|\right|_2^2 \le 2\,\mathbb{E}\left|\left|\left|{}_{\varepsilon,s}\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega^{(T)}\right|\right|\right|_2^2 + 2\,\mathbb{E}\left|\left|\left|\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega\right|\right|\right|_2^2,$$

hence, by Theorem 3.6.1 and 3.6.2, we can restrict ourselves to the first summand. Using the triangle inequality again,

$$\sup_\omega \mathbb{E}\left|\left|\left|{}_{\varepsilon,s}\mathscr{F}_\omega^{(T)} - \mathscr{F}_\omega^{(T)}\right|\right|\right|_2^2 \le \sup_\omega \iint \mathbb{E}\left|{}_{\varepsilon,s}f_\omega^{(T)} - {}_sf_\omega^{(T)}\right|^2 + \sup_\omega \iint \mathbb{E}\left|{}_sf_\omega^{(T)} - f_\omega^{(T)}\right|^2. \tag{3.8.2}$$

The integrals are on $[0,1]^2$ with respect to $d\tau d\sigma$. First, we deal with the first summand:

$$\begin{aligned}\left|{}_{\varepsilon,s}f_\omega^{(T)} - {}_sf_\omega^{(T)}\right|^2 &= 2\pi T^{-2}\left|\sum_{l=0}^{T-1} W^{(T)}(\omega - 2\pi l/T)\left({}_{\varepsilon,s}p_{2\pi l/T}^{(T)} - {}_sp_{2\pi l/T}^{(T)}\right)\right|^2 \\ &\le O(T^{-1})\sum_{l=0}^{T-1}[W^{(T)}(\omega - 2\pi l/T)]^2\left|{}_{\varepsilon,s}p_{2\pi l/T}^{(T)} - {}_sp_{2\pi l/T}^{(T)}\right|^2,\end{aligned}$$

where we have used Jensen's inequality. We claim that, if $\tau_j \le \tau < \tau_{j+1}$ and $\tau_k \le \sigma < \tau_{k+1}$,

$$\begin{aligned}\left|{}_{\varepsilon,s}p_\omega^{(T)}(\tau,\sigma) - {}_sp_\omega^{(T)}(\tau,\sigma)\right|^2 \le{}& 3|{}_s\widetilde{X}_\omega^{(T)}(\tau)|^2|\widetilde{\varepsilon}_{-\omega}^{(T)}(k)|^2 \\ &+ 3|\widetilde{\varepsilon}_\omega^{(T)}(j)\widetilde{\varepsilon}_{-\omega}^{(T)}(k)|^2 + 3|\widetilde{\varepsilon}_\omega^{(T)}(j)|^2|{}_s\widetilde{X}_{-\omega}^{(T)}(\sigma)|^2,\end{aligned}$$

where $\widetilde{\varepsilon}_{\omega}^{(T)}(j) = (2\pi T)^{-1/2}\sum_{l=0}^{T-1} e^{-\mathbf{i}\omega t}\varepsilon_{tj}$. To see this, we note that

$$\begin{aligned}
{}_{\varepsilon,s}p_{\omega}^{(T)}(\tau,\sigma) - {}_{s}p_{\omega}^{(T)}(\tau,\sigma) &= {}_{\varepsilon,s}\widetilde{X}_{\omega}^{(T)}(\tau)\cdot {}_{\varepsilon,s}\widetilde{X}_{-\omega}^{(T)}(\sigma) - {}_{s}\widetilde{X}_{\omega}^{(T)}(\tau)\cdot {}_{s}\widetilde{X}_{-\omega}^{(T)}(\sigma) \\
&= \left({}_{\varepsilon,s}\widetilde{X}_{\omega}^{(T)}(\tau) - {}_{s}\widetilde{X}_{\omega}^{(T)}(\tau)\right)\cdot {}_{\varepsilon,s}\widetilde{X}_{-\omega}^{(T)}(\sigma) + \\
&\quad + {}_{s}\widetilde{X}_{\omega}^{(T)}(\tau)\cdot\left({}_{\varepsilon,s}\widetilde{X}_{-\omega}^{(T)}(\sigma) - {}_{s}\widetilde{X}_{-\omega}^{(T)}(\sigma)\right) \\
&= {}_{s}\widetilde{X}_{\omega}^{(T)}(\tau)\widetilde{\varepsilon}_{-\omega}^{(T)}(k) + \widetilde{\varepsilon}_{\omega}^{(T)}(j)\widetilde{\varepsilon}_{-\omega}^{(T)}(k) + \\
&\quad + \widetilde{\varepsilon}_{\omega}^{(T)}(j)\,{}_{s}\widetilde{X}_{-\omega}^{(T)}(\sigma),
\end{aligned}$$

since ${}_{\varepsilon,s}\widetilde{X}_{\omega}^{(T)}(\tau) = {}_{s}\widetilde{X}_{\omega}^{(T)}(\tau) + \widetilde{\varepsilon}_{\omega}^{(T)}(j)$, and similarly if we replace σ by τ and j by k. Our claim thus follows from Jensen's inequality.

In order to bound the expectation of $\left|{}_{\varepsilon,s}p_{\omega}^{(T)}(\tau,\sigma) - {}_{s}p_{\omega}^{(T)}(\tau,\sigma)\right|^2$, we will first compute the expectation, conditional on the σ-algebra generated by the X_t's, which we will denote by $\mathbb{E}_X$, and then use the tower property. As an intermediate step, we claim that

$$\mathbb{E}_X\,|\widetilde{\varepsilon}_{\omega}^{(T)}(j)|^2 = O(Tb^2) + O(\gamma^2) =: C_1$$

and

$$\mathbb{E}_X\,|\widetilde{\varepsilon}_{\omega}^{(T)}(j)\widetilde{\varepsilon}_{-\omega}^{(T)}(k)|^2 = O(c_1T^{-1}) + O(c_1b) + O(\gamma^4) + O(b^4) + O(\gamma^2+b^2)O(Tb^2) + O(T^2b^4) =: C_2$$

uniformly in $j,k = 1,\dots,M$, and uniformly in ω (notice that all $\mathbb{E}_X$ can be replaced by $\mathbb{E}$ since the ε_{tj}'s are independent of the X_t's). For simplicity, we shall refer to these bounds by C_1 and C_2. To establish our claim, notice that

$$\mathbb{E}_X\,|\widetilde{\varepsilon}_{\omega}^{(T)}(j)|^2 = \mathrm{var}_X\left(\widetilde{\varepsilon}_{\omega}^{(T)}(j)\right) + \mathbb{E}_X^2\left(\widetilde{\varepsilon}_{\omega}^{(T)}(j)\right),$$

where the second term is $O(Tb^2)$. Since $|\widetilde{\varepsilon}_{\omega}^{(T)}(j)|^2 = \widetilde{\varepsilon}_{\omega}^{(T)}(j)\widetilde{\varepsilon}_{-\omega}^{(T)}(j)$, we get

$$\mathrm{var}_X\left(\widetilde{\varepsilon}_{\omega}^{(T)}(j)\right) = (2\pi T)^{-1}\sum_{t,s=0}^{T-1} e^{-\mathbf{i}\omega(t-s)}\mathrm{cov}_X\left(\varepsilon_{tj},\varepsilon_{sj}\right).$$

The summand is equal to γ^2 if $t = s$, and zero otherwise (by our assumptions on the ε's), hence the first statement follows directly. We can now turn to the second statement. First, notice that

$$\mathbb{E}_X\,|\widetilde{\varepsilon}_{\omega}^{(T)}(j)\widetilde{\varepsilon}_{-\omega}^{(T)}(k)|^2 \le T^{-2}\sum_{t_1,t_2,t_3,t_4=0}^{T-1}\left|\mathrm{cov}_X\left(\varepsilon_{t_1}\varepsilon_{t_2},\varepsilon_{t_3}\varepsilon_{t_4}\right)\right|,$$

where we have written ε_t instead of ε_{tj} for tidiness. We now bound the summand in each of the following cases, and count the number of such cases:

1. the summand is of the form $\mathbb{E}_X\,\varepsilon_t^4$, which is bounded by c_1. There are T such cases;

2. the summand is of the form $\mathbb{E}_X(\varepsilon_t^3\varepsilon_s) = \mathbb{E}_X(\varepsilon_t^3)\,\mathbb{E}_X\varepsilon_s$, which is bounded by $c_1 b$. There are $O(T^2)$ such cases;

3. the summand is of the form $\mathbb{E}_X\left(\varepsilon_t^2\varepsilon_s^2\right)$ with $s \neq t$. It is bounded by $O(\gamma^4) + O(b^4)$, and there are $O(T^2)$ such summands;

4. the summand is of the form $\mathbb{E}_X\left(\varepsilon_t^2\varepsilon_{s_1}\varepsilon_{s_2}\right)$, with both $s_1, s_2 \neq t$. It is bounded by $(\gamma^2 + b^2)b^2$, and there are $O(T^3)$ such summands;

5. the summands are of the form $\mathbb{E}_x\left(\varepsilon_{t_1}\varepsilon_{t_2}, \varepsilon_{t_3}\varepsilon_{t_4}\right)$, where all the t_js are distinct. It is bounded by b^4, and there are $O(T^4)$ such summands.

Piecing these bounds together, we get

$$\mathbb{E}_X|\widetilde{\varepsilon}_{\omega}^{(T)}(j)\widetilde{\varepsilon}_{-\omega}^{(T)}(k)|^2 = O(c_1 T^{-1}) + O(c_1 b) + O(\gamma^4) + O(b^4) + O(\gamma^2 + b^2)O(Tb^2) + O(T^2 b^4)$$

uniformly in ω, j, k.

We can now bound $\mathbb{E}_X\left|{}_{\varepsilon,s}p_{2\pi l/T}^{(T)} - {}_{s}p_{2\pi l/T}^{(T)}\right|^2$:

$$\begin{aligned}
\mathbb{E}_X\left|{}_{\varepsilon,s}p_{2\pi l/T}^{(T)} - {}_{s}p_{2\pi l/T}^{(T)}\right|^2 &\leq 3|{}_{s}\widetilde{X}_{2\pi l/T}^{(T)}(\tau)|^2\,\mathbb{E}_X|\widetilde{\varepsilon}_{-2\pi l/T}^{(T)}(k)|^2 \\
&\quad + 3\,\mathbb{E}_X|\widetilde{\varepsilon}_{2\pi l/T}^{(T)}(j)\widetilde{\varepsilon}_{-2\pi l/T}^{(T)}(k)|^2 \\
&\quad + 3|{}_{s}\widetilde{X}_{-2\pi l/T}^{(T)}(\sigma)|^2\,\mathbb{E}_X|\widetilde{\varepsilon}_{2\pi l/T}^{(T)}(j)|^2 \\
&\leq 3C_1\left[|{}_{s}\widetilde{X}_{2\pi l/T}^{(T)}(\tau)|^2 + |{}_{s}\widetilde{X}_{-2\pi l/T}^{(T)}(\sigma)|^2\right] + 3C_2,
\end{aligned}$$

where C_1 and C_2 are defined above. Since $|\widetilde{X}_{\omega}^{(T)}(\tau)|^2 = |p_{\omega}^{(T)}(\tau,\tau)|$, Proposition 3.4.4, Remark 3.5.5 and (3.8.1) yield that $\int \mathbb{E}|{}_{s}\widetilde{X}_{2\pi l/T}^{(T)}(\tau)|^2 d\tau = O(1)$. Using the tower property, we obtain

$$\iint \mathbb{E}\left|{}_{\varepsilon,s}p_{2\pi l/T}^{(T)} - {}_{s}p_{2\pi l/T}^{(T)}\right|^2 \leq O(C_1) + O(C_2),$$

uniformly in $l = 1, \ldots, T-1$, under the assumptions of this Theorem. Thus

$$\begin{aligned}
\iint \mathbb{E}\left|{}_{\varepsilon,s}f_{\omega}^{(T)} - {}_{s}f_{\omega}^{(T)}\right|^2 &\leq O(T^{-1})\,[O(C_1) + O(C_2)] \cdot \sum_{l=0}^{T-1}[W^{(T)}(\omega - 2\pi l/T)]^2 \\
&= O(T^{-1})\,[O(C_1) + O(C_2)]\,O(TB_T)\left\|W^{(T)}\right\|_{\infty}^2 \\
&\qquad\qquad \text{(at most } O(TB_T) \text{ non-zero summand)} \\
&= O(B_T^{-1})\,[O(C_1) + O(C_2)] \qquad \text{(by Lemma 3.12.18)} \\
&= O(B_T^{-1})\left[O(T^2 b^4) + O(c_1 b) + O(\gamma^2) + \right. \\
&\qquad \left. + O(Tb^2)O(\gamma^2 + b^2) + O(b^4) + O(\gamma^4) + O(T^{-1}c_1)\right],
\end{aligned}$$

uniformly in ω. Hence we obtain the bound on the expectation of first summand of (3.8.2).

We now turn to the second summand of (3.8.2). Using the decomposition

$$\mathbb{E}\left|{}_s f_\omega^{(T)} - f_\omega^{(T)}\right|^2 = \operatorname{cov}\left({}_s f_\omega^{(T)} - f_\omega^{(T)}, {}_s f_\omega^{(T)} - f_\omega^{(T)}\right) + \left|\,\mathbb{E}\left[{}_s f_\omega^{(T)} - f_\omega^{(T)}\right]\right|^2,$$

the covariance term can be written as sums and differences of four terms of the form

$$\operatorname{cov}\left(f_\omega^{(T)}(\sigma_1,\sigma_2), f_\omega^{(T)}(\sigma_3,\sigma_4)\right),$$

for some σ_l's. The important thing here is that each of these terms can be bounded in L^2—independently of the σ_l's—using Corollary 3.5.3 and Proposition 3.5.4:

$$\operatorname{cov}\left({}_s f_\omega^{(T)} - f_\omega^{(T)}, {}_s f_\omega^{(T)} - f_\omega^{(T)}\right) = \begin{cases} O(B_T^{-2}T^{-1}) + O(T^{-1}), & \omega \in [-\pi,\pi], \\ O(B_T^{-1}T^{-1}), & \omega \in [B_T, \pi - B_T], \end{cases}$$

in L^2, where the bound is uniform in ω in both cases. The second bound will be useful for the last statement of the Theorem. Using the decomposition $\int_0^\pi = \int_0^{B_T} + \int_{B_T}^{\pi-B_T} + \int_{\pi-B_T}^{\pi}$, we get

$$\int_0^\pi \iint \left|\operatorname{cov}\left({}_s f_\omega^{(T)} - f_\omega^{(T)}, {}_s f_\omega^{(T)} - f_\omega^{(T)}\right)\right| d\omega = O(B_T^{-1}T^{-1}),$$

and

$$\sup_\omega \iint \left|\operatorname{cov}\left({}_s f_\omega^{(T)} - f_\omega^{(T)}, {}_s f_\omega^{(T)} - f_\omega^{(T)}\right)\right| = O\left(B_T^{-2}T^{-1}\right).$$

In order to bound $\left|\,\mathbb{E}\left[{}_s f_\omega^{(T)} - f_\omega^{(T)}\right]\right|^2$, we use Proposition 3.5.1 and Lemma 3.12.13 (with $p = 1$):

$$\iint \left|\,\mathbb{E}\left[{}_s f_\omega^{(T)} - f_\omega^{(T)}\right]\right|^2 \le 4 \iint \left|{}_s f_\omega - f_\omega\right|^2 + O(B_T^2) + O(T^{-2}) + O(B_T T)^{-2},$$

uniformly in ω. Thus

$$\sup_\omega \iint \left|\,\mathbb{E}\left[{}_s f_\omega^{(T)} - f_\omega^{(T)}\right]\right|^2 \le 4 \sup_\omega |||{}_s\mathscr{F}_\omega - \mathscr{F}_\omega|||_2^2 + O(B_T^2) + O(T^{-2}) + O(B_T T)^{-2}.$$

The quantity $|||{}_s\mathscr{F}_\omega - \mathscr{F}_\omega|||_2^2 = \iint \left|{}_s f_\omega - f_\omega\right|^2$ is in fact the squared distance between ${}_s f_\omega$ and f_ω in the space $L^2([0,1]^2, \mathbb{C})$. Under (3.8.1), $f_\omega(\tau,\sigma)$ is uniformly continuous in ω, τ, σ; since ${}_s f_\omega$ is a step-wise approximation of f_ω, we obtain

$$\sup_{\omega\in[-\pi,\pi]} |||{}_s\mathscr{F}_\omega - \mathscr{F}_\omega|||_2^2 \to 0, \quad M \to \infty.$$

The proof is completed by piecing all these results together. □

Remark 3.8.5. *The use of Proposition 3.5.4 was valid in this context, but requires some attention. Indeed, it relies on Lemma 3.12.21, applied to* $g_{(\tau,\sigma)}(\alpha) = {}_s f_\alpha^{(T)}(\tau,\sigma)$. *Remark 3.12.22 tells us that the convergence of the*

convolution integral depends on the uniform continuity parameter $\delta(\varepsilon)$, which here will depend on the size of the sampling grid $M = M(T)$; in other words, $\delta(\varepsilon) = \delta(\varepsilon, M)$. But notice that since (3.8.1) *holds,*

$$\begin{aligned}\left\| {}_{s}f_{\omega_1} - {}_{s}f_{\omega_2} \right\|_2 &\leq \sup_{0\leq\tau,\sigma\leq 1} \left| {}_{s}f_{\omega_1}(\tau,\sigma) - {}_{s}f_{\omega_2}(\tau,\sigma) \right| \\ &= \sup_{\tau,\sigma=\tau_1,\ldots,\tau_M} \left| f_{\omega_1}(\tau,\sigma) - f_{\omega_2}(\tau,\sigma) \right| \\ &\leq \sup_{0\leq\tau,\sigma\leq 1} \left| f_{\omega_1}(\tau,\sigma) - f_{\omega_2}(\tau,\sigma) \right|,\end{aligned}$$

hence we can choose a $\delta(\varepsilon)$ that is independent of M, and the application of Proposition 3.5.4 is valid.

3.9 Numerical Simulations

In order to probe the finite sample performance of our estimators (in terms of IMSE), we have performed numerical simulations on stationary functional time series admitting a linear representation

$$X_t = \sum_{s=0}^{10} A_s \varepsilon_{t-s}.$$

We have taken the collection of 'innovation' functions $\{\varepsilon_t\}$ to be independent Wiener processes on $[0,1]$, which we have represented using a truncated Karhunen–Loève expansion:

$$\varepsilon_t(\tau) = \sum_{k=1}^{1000} \xi_{k,t} \sqrt{\lambda_k} e_k(\tau).$$

Here $\lambda_k = 1/[(k-1/2)^2\pi^2]$, $\xi_{k,t}$ are independent standard Gaussian random variables and $e_k(\tau) = \sqrt{2}\sin[(k-1/2)\pi\tau]$ is orthonormal system in $L^2([0,1],\mathbb{R})$ (Adler (1990)). We have constructed the operators A_s so that their image be contained within a 50-dimensional subspace of $L^2([0,1],\mathbb{R})$, spanned by an orthonormal basis $\psi_1,\ldots,\psi_{50}$. Representing ε_t in the e_k basis, and A_s in the $\psi_m \otimes_2 e_k$ basis, we obtain a matrix representation of the process X_t as $\mathbf{X}_t = \sum_{s=0}^{10} \mathbf{A}_s \boldsymbol{\varepsilon}_{t-s}$, where $\mathbf{X}_t$ is a 50×1 matrix, each $\mathbf{A}_s$ is a 50×1000 matrix, and $\boldsymbol{\varepsilon}_t$ is a 1000×1 matrix.

We simulated a stretch of X_t, $t = 0,\ldots T-1$ for $T = 2^n$, with $n = 7,8,\ldots,15$. Typical functional data sets would range between $T = 2^6$ and $T = 2^8$ data points. We constructed the matrices $\mathbf{A}_s$, as random Gaussian matrices with independent entries, such that elements in row j where $N(0, j^{-2\alpha})$ distributed. When $\alpha = 0$, the projection of each ε_t onto the subspace spanned by each ψ_m, $m = 1,\ldots,50$ has (roughly) a comparable magnitude. A positive value of α, e.g. $\alpha = 1$ means that the projection of ε_t onto the subspace spanned by ψ_j will have smaller magnitude for larger j's.

For comparison purposes, we also carried out analogous simulations, but

with $\lambda_k = 1$, that is, the variance of the innovations ε_t being equal to one in each direction $e_n, n = 1, \ldots, 1000$. In the sequel, we will refer to these as the simulations with "white noise innovations", and to the previous ones as "Wiener innovations". The white noise process is, of course, not a true white noise process, but a projection of a white noise process. However, it does represent a case of a "rough" innovation process, which we present here as an extreme scenario.

For each T, we generated 200 simulation runs which we used to compute the IMSE by approximating the integral

$$2\int_0^\pi \mathbb{E}\left|\left|\left|\mathscr{F}_\omega - \mathscr{F}_\omega^{(T)}\right|\right|\right|_2^2 d\omega$$

by a weighted sum over the finite grid $\Gamma = \{\pi j/10; \quad j = 0, \ldots, 9\}$. We chose $B_T = T^{-1/5}$ (e.g. Grenander & Rosenblatt (1957, Par. 4.7), Brillinger (2001, Par. 7.4)) and $W(x)$ to be the *Epanechnikov kernel* (e.g. Wand & Jones (1995)), $W(x) = \frac{3}{4}(1 - x^2)$ if $|x| < 1$, and zero otherwise. The results are shown on a log-log scale in Figure 3.2, for $\alpha = 2$. The slopes of the least square lines passing through the medians of the simulation results show that $IMSE(\mathscr{F}^{(T)}) \propto T^\beta$, with $\beta \approx -0.797$ for the white noise innovations, and $\beta \approx -0.796$ for the Wiener innovations. According to Theorem 3.6.2, the decay of the $IMSE(\mathscr{F}^{(T)})$ is bounded by

$$C_1 T^{-2/5} + C_2 T^{-4/5} \approx C_1 T^{-0.4} \quad \text{(if } T \text{ is large)},$$

for some constants C_1, C_2.

In order to gain a visual appreciation of the accuracy of the estimators, we construct plots to compare the true and estimated spectral density kernels in Figures 3.3 and 3.4, for the Wiener and white noise cases, respectively. For practical purposes, we set $\alpha = 2$, as for the simulation of the IMSEs. We simulated $X_t = A_0\varepsilon_t + A_1\varepsilon_{t-1}$, where $\varepsilon_t(\tau)$ lies on the subspace of $L^2([0,1], \mathbb{R})$ spanned by the basis $e_1, \ldots, e_{100}$, and the operators A_0, A_1 lie in the subspace spanned by $(\psi_m \otimes_2 e_k)_{m=1,\ldots,51;k=1,\ldots,100}$. Since the target parameter is a complex-valued function defined over a two-dimensional rectangle, some information loss must be incurred when representing it graphically. We chose to suppress the phase component of the spectral density kernel, plotting only its amplitude, $|f_\omega(\tau,\sigma)|$, for all $(\tau,\sigma) \in [0,1]^2$ and for selected frequencies ω (the spectral density kernel is seen to be smooth in ω, so this does not entail a significant loss of information). For various choices of sample size T, we have replicated the realisation of the process, and the corresponding kernel density estimator for the particular frequency. Each time, we plotted the contours in superposition, in order to be able to visually appreciate the variability in the estimators:

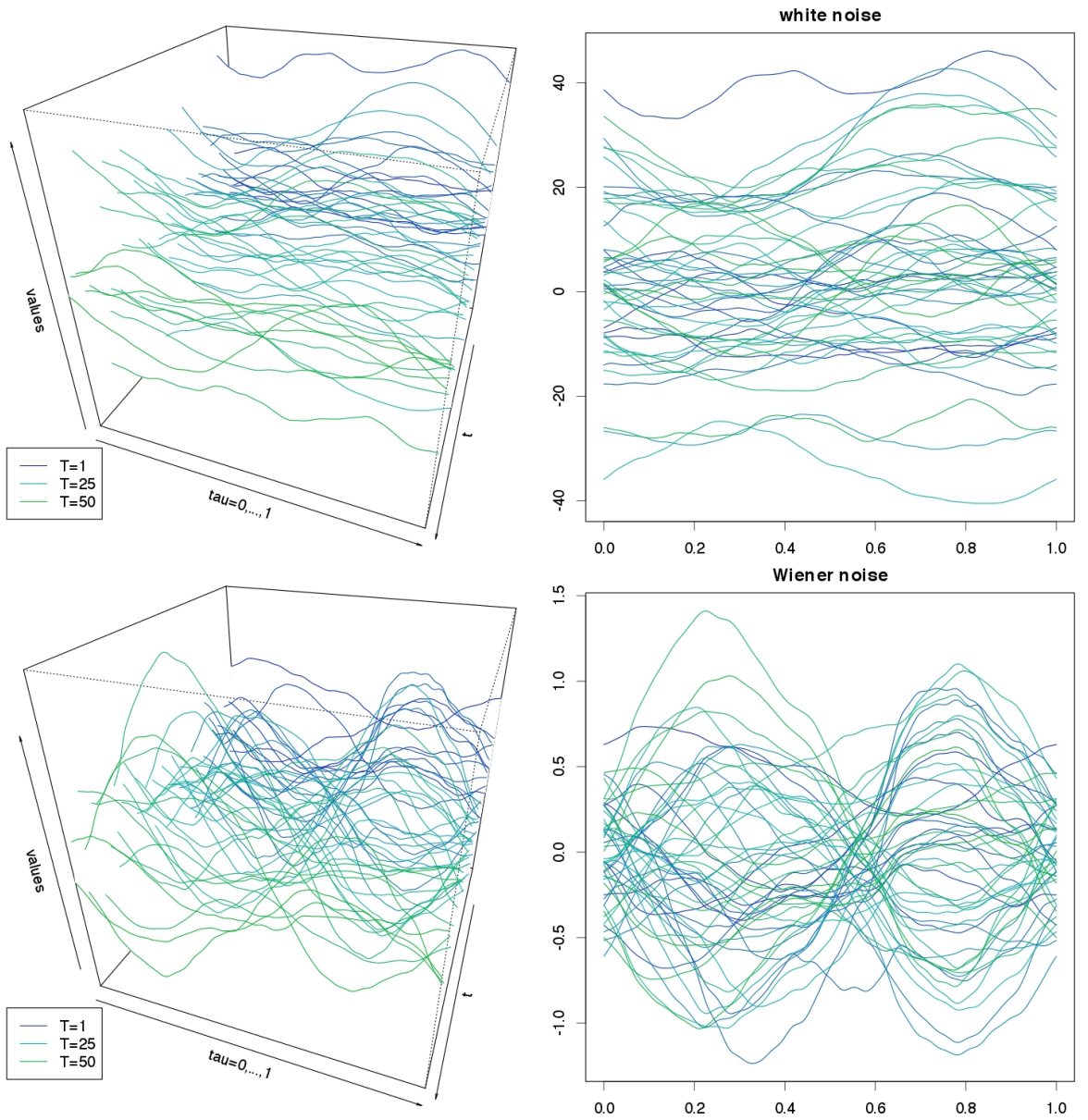


Figure 3.1 – Plots of the functional linear process used for the simulation of the IMSE. The plots in the top row are for the linear process with "Wiener noise", and the bottom ones are with "white noise". In each case we have a perspective plot, which illustrates the direction of time, and a superposition plot, where the temporal dimension is suppressed.

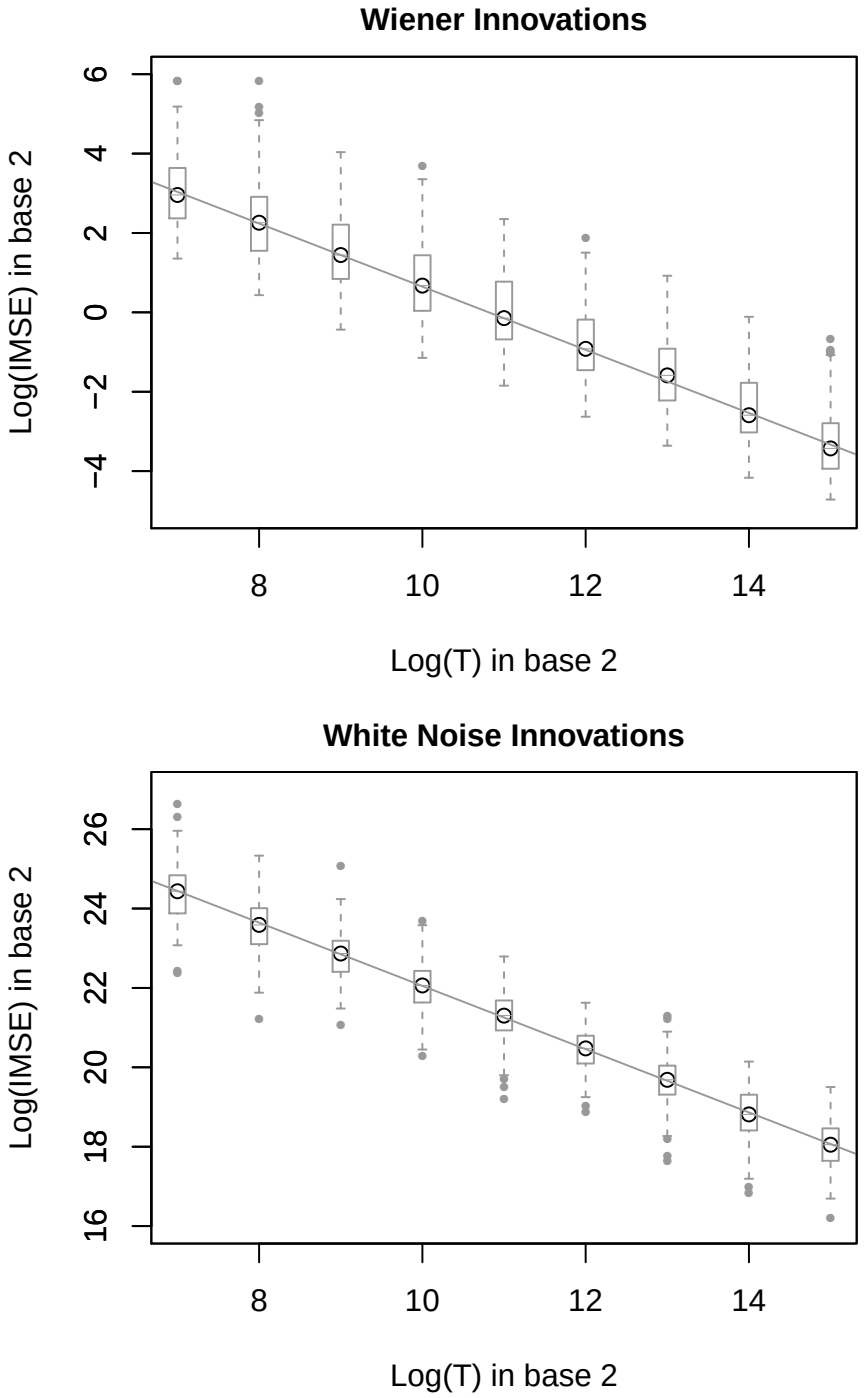


Figure 3.2 – Results of the simulated ISE on a log-log scale, with $\alpha = 2$. The upper and lower plots correspond to the Wiener Innovations and the White Noise Innovation setups, respectively. The dots correspond to the median of the results of the simulations, and the lines are the least squares lines of the medians. The boxplots summarise the distribution of the ISE for the 200 simulation runs. Though the ranges of the y-axes are different, the scales are the same, and the two least square lines are indeed almost parallel.

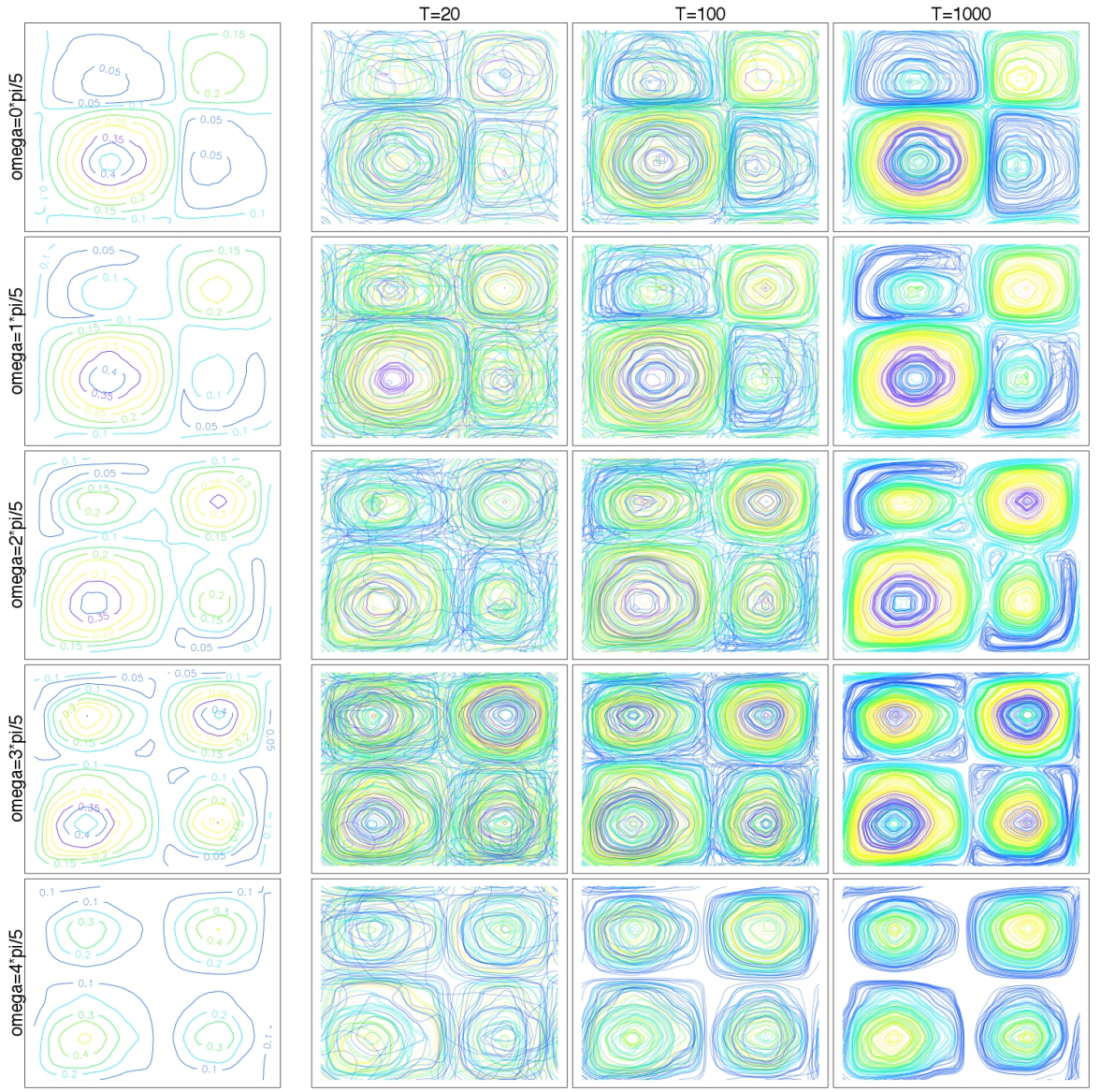


Figure 3.3 – Contour plots for the amplitude of the true and estimated spectral density kernel when the innovation process consists of Wiener processes. Each row corresponds to a different frequency ($\omega = k\pi/5$, $k = 0, 1, ..., 4$, from top to bottom). The first column contains the contour plots of the true amplitudes of the kernel at each corresponding frequency. The rest of the columns correspond to the estimated contours for different sample sizes ($T = 20, 100, 1000$ from left to right). Twenty estimates, corresponding to twenty replications of the process, have been superposed in order to provide a visual illustration of the variability. The contours plotted always correspond to the same level curves and use the same colour-coding in each row.

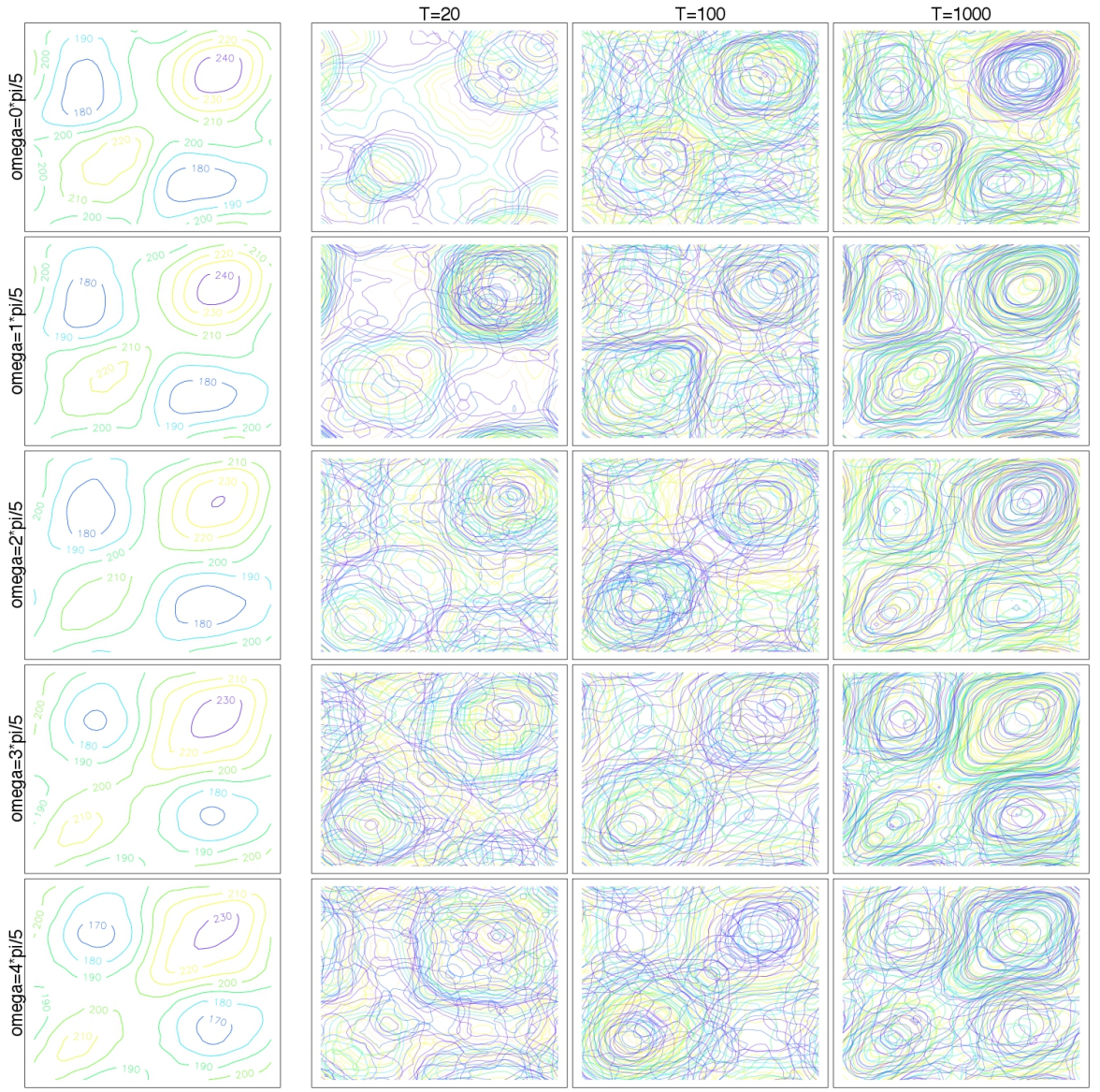


Figure 3.4 – Contour plots for the amplitude of the true and estimated spectral density kernel when the innovation process consists of white noise processes. Each row corresponds to a different frequency ($\omega = k\pi/5$, $k = 0, 1, ..., 4$, going from top to bottom). The first column contains the contour plots of the true amplitudes of the kernel at each corresponding frequency. The rest of the columns correspond to the estimated contours for different sample sizes ($T = 20, 100, 1000$ from left to right). Twenty estimates, corresponding to twenty replications of the process, have been superposed in order to provide a visual illustration of the variability. The contours plotted always correspond to the same level curves and use the same colour-coding in each row.

tangled contour lines where no clear systematic pattern emerges signify a region of high variability, whereas aligned contour lines that adhere to a recognisable shape represent regions of low variability. As is expected, the "smoother" the innovation process, the less variable the results appear to be, and the variability decreases for larger values of T.

3.10 Mixing, Tightness and Projections

Our results on the asymptotic Gaussian representations of the discrete Fourier transform and the spectral density estimator (Theorems 3.3.4 and 3.6.5) effectively rest upon two sets of mixing conditions: (1) the summability of the nuclear norms of the autocovariance operators (at various rates), and (2) the summability of the cumulant kernels of all orders (at various rates). The roles of these two sets of mixing conditions are distinct. The first is required in order to establish tightness of the sequence of discrete Fourier transforms and spectral density estimators of the underlying process. Tightness allows one to then apply the Cramér–Wold device, and to determine the asymptotic distribution by considering finite-dimensional projections (see Section C.2, Lemma C.2.3, and Corollary 3.6.4). The role of the second set of mixing conditions, then, is precisely to allow the determination of the asymptotic law of the projections, thus identifying the stipulated limiting distribution via tightness.

Therefore, in principle, one can replace the second set of mixing conditions with a set of conditions that allow for the discrete Fourier transforms and spectral density estimators of the *vector* time series of the projections to be asymptotically Gaussian, jointly in any finite number of frequencies. Our approach was to generalize the cumulant multivariate conditions of Brillinger (2001), which do not require structural assumptions further to stationarity. Alternatively, one may pursue generalizations of multivariate conditions involving α-mixing and summable cumulants of order $2, 4$, and 8 as in Hannan (1970, Chapter IV, Par. 4) and Rosenblatt (1984, 1985), though α-mixing can also be a strong condition. Adding more structure, e.g. in the context of linear processes, one can focus on extending weaker conditions requiring finite fourth moments and summable coefficients (Hannan 1970, Anderson 1994).

For the case of non-linear moving-average representations of the form

$$\xi_t = G(\varepsilon_t, \varepsilon_{t-1}, \ldots),$$

where G is a measurable function and $\{\varepsilon_j\}$ are i.i.d. random variables, several results exist; however, none of them are (yet) established for vector time series. For instance Shao & Wu (2007) show that if the second

moment of ξ_t is finite and

$$\sum_{k=0}^{\infty} \sqrt{\mathbb{E}\,|\,\mathbb{E}\,[\xi_k - \xi_{k+1}|\mathscr{F}_0]|^2} < \infty,$$

where $\mathscr{F}_0$ is the sigma-algebra generated by $\{\varepsilon_0, \varepsilon_{-1}, \ldots\}$, then the discrete Fourier transforms of ξ_t are asymptotically Gaussian, jointly for a finite number of frequencies. Furthermore, Shao & Wu (2007) establish the asymptotic normality of the spectral density estimator at distinct frequencies under the moment condition $\mathbb{E}|\xi_t|^{4+\delta} < \infty$, for some $\delta > 0$, and the following coupling condition: there exist $\alpha > 0$, $C > 0$ and $\rho \in (0,1)$ such that

$$\mathbb{E}|\xi_t - \xi_t'|^{\alpha} < C\rho^t, \quad t = 0, 1, \ldots, \tag{3.10.1}$$

where $\xi_t' = G(\varepsilon_t, \ldots, \varepsilon_1, \varepsilon_0', \varepsilon_{-1}', \ldots)$ and $(\varepsilon_k')_{k\in\mathbb{Z}}$ is an i.i.d. copy of $(\varepsilon_k)_{k\in\mathbb{Z}}$. Notice that (3.10.1) is related to (in fact stronger than) the L^p-m-approximability condition of Hörmann & Kokoszka (2010). Under the weaker conditions $\mathbb{E}|\xi_t|^4 < \infty$, and

$$\sum_{t=0}^{\infty} \left(\mathbb{E}\,|\xi_t - \check{\xi}_t|^4\right)^{1/4} < \infty,$$

where $\check{\xi}_t = G(\ldots, \varepsilon_{-1}, \varepsilon_0', \varepsilon_1, \ldots, \varepsilon_t)$ and ε_0' is an i.i.d. copy of ε_0, Liu & Wu (2010) establish that the spectral density estimator at a *fixed* frequency is asymptotically Gaussian. The idea behind these coupling conditions is to approximate the series ξ_t by m-dependent series, for which derivation of asymptotic results is easier. We also mention that, under milder conditions, Peligrad & Wu (2010) establish that for almost all $\omega \in (0, 2\pi)$, the discrete Fourier transform at ω is asymptotically normal.

The mixing conditions pursued in this chapter have the advantage of not requiring additional structure, at the price of being relatively strong if additional structure could be assumed. For example, if a process is linear, the cumulant conditions will be satisfied provided all moments exist and the coefficient operators are summable in an appropriate sense, as shown in the results below (see in particular Remark 3.10.3). As mentioned above, we conjecture that four moments and summability of the coefficients would suffice in the linear case, however a more thorough study of mixing conditions for the linear case is outside the scope of the present thesis.

Proposition 3.10.1.
Let $l \geq 1$, or $l = 0$ be a constant, and let ε_t, $t = 0, \pm 1, \ldots$ be a k-order stationary (see Definition 3.2.2) sequence of random elements of H satisfying $\mathbb{E}\,\|\varepsilon_0\|^k < \infty$, and

$$\sum_{t_1,\ldots,t_{k-1}\in\mathbb{Z}} \left(1 + |t_j|^l\right) \left\|\operatorname{cum}\left(\varepsilon_{t_1}, \ldots, \varepsilon_{t_{k-1}}, \varepsilon_0\right)\right\|_{\pi} < \infty \tag{3.10.2}$$

for all $j = 1, \ldots, k-1$.

$\|\cdot\|_{\pi}$ is defined in Section A.3.2 on page 225

If (A_s) is a sequence of bounded operators on H, satisfying

$$\sum_{s\in\mathbb{Z}}(1+|s|^l)|||A_s|||_\infty<\infty, \tag{3.10.3}$$

then the (filtered) series

$$X_t=\sum_{s\in\mathbb{Z}}A_s\varepsilon_{t-s}$$

satisfies

$$\sum_{t_1,\dots,t_{k-1}\in\mathbb{Z}}\left(1+|t_j|^l\right)\left\|\operatorname{cum}\left(X_{t_1},\dots,X_{t_{k-1}},X_0\right)\right\|_\pi<\infty. \tag{3.10.4}$$

for all $j=1,\dots,k-1$.

Proposition 3.10.2.
The statement of Proposition 3.10.1 holds when replacing all the projective norms $\|\cdot\|_\pi$ by the Hilbert tensor norm $\|\cdot\|$ in (3.10.2) *and* (3.10.4).

Remark 3.10.3.
1. *The sample spectral density operators of a filtered series $X_t=\sum_{s\in\mathbb{Z}}A_s\varepsilon_{t-s}$ satisfying the conditions of Proposition 3.10.1 for $k=2$ and $k=4$ will be tight.*

2. *If (ε_t) is an i.i.d. sequence, then* (3.10.2) *reduces to $\mathbb{E}\|\varepsilon_0\|^k<\infty$, by Proposition 3.12.8. In particular, if $\mathbb{E}\|\varepsilon_0\|^k<\infty$ for all $k\geq 1$, and* (3.10.3) *holds for $l=1$, then the linear process $X_t=\sum_{s\in\mathbb{Z}}A_s\varepsilon_{t-s}$ satisfies the conditions of Theorems 3.3.4, 3.6.5 and 3.7.3.*

Proof of Propositions 3.10.1 and 3.10.2. We will only prove Proposition 3.10.1, since the proof of Proposition 3.10.2 follows the exact same argument, with $\|\cdot\|_\pi$ replaced by $\|\cdot\|$ everywhere.
Using Propositions 3.12.10 and 3.12.11, we get

$$\begin{aligned}
&\sum_{t_1,\dots,t_{k-1}\in\mathbb{Z}}\left(1+|t_j|^l\right)\left\|\operatorname{cum}\left(X_{t_1},\dots,X_{t_{k-1}},X_0\right)\right\|_\pi\\
&=\sum_{t_1,\dots,t_{k-1}\in\mathbb{Z}}\left(1+|t_j|^l\right)\left\|\sum_{s_1,\dots,s_{k-1}\in\mathbb{Z}}\left(A_{s_1}\otimes_\pi\cdots\otimes_\pi A_{s_k}\right)\operatorname{cum}\left(\varepsilon_{t_1-s_1},\dots,\varepsilon_{t_{k-1}-s_{k-1}},\varepsilon_{-s_k}\right)\right\|_\pi\\
&\leq\sum_{t_1,\dots,t_{k-1}\in\mathbb{Z}}\left(1+|t_j|^l\right)\sum_{s_1,\dots,s_{k-1}\in\mathbb{Z}}|||A_{s_1}\otimes_\pi\cdots\otimes_\pi A_{s_k}|||_\infty\left\|\operatorname{cum}\left(\varepsilon_{t_1-s_1},\dots,\varepsilon_{t_{k-1}-s_{k-1}},\varepsilon_{-s_k}\right)\right\|_\pi\\
&\leq\sum_{t_1,\dots,t_{k-1}\in\mathbb{Z}}\left(1+|t_j|^l\right)\sum_{s_1,\dots,s_{k-1}\in\mathbb{Z}}|||A_{s_1}|||_\infty\cdots|||A_{s_k}|||_\infty\left\|\operatorname{cum}\left(\varepsilon_{t_1-s_1},\dots,\varepsilon_{t_{k-1}-s_{k-1}},\varepsilon_{-s_k}\right)\right\|_\pi\\
&=\sum_{s_1,\dots,s_{k-1}\in\mathbb{Z}}|||A_{s_1}|||_\infty\cdots|||A_{s_k}|||_\infty\sum_{t_1,\dots,t_{k-1}\in\mathbb{Z}}\left(1+|t_j|^l\right)\left\|\operatorname{cum}\left(\varepsilon_{t_1-s_1},\dots,\varepsilon_{t_{k-1}-s_{k-1}},\varepsilon_{-s_k}\right)\right\|_\pi
\end{aligned}$$

Making the changes of variables $u_l = t_l - s_l + s_k$, for $l = 1, \ldots, k-1$, we get

$$\begin{aligned}
&\sum_{t_1,\ldots,t_{k-1}\in\mathbb{Z}} \left(1+|t_j|^l\right) \left\|\operatorname{cum}\left(X_{t_1},\ldots,X_{t_{k-1}},X_0\right)\right\|_\pi \\
&= \sum_{s_1,\ldots,s_{k-1}\in\mathbb{Z}} |||A_{s_1}|||_\infty \cdots |||A_{s_k}|||_\infty \sum_{u_1,\ldots,u_{k-1}\in\mathbb{Z}} \left(1+|u_j+s_j-s_k|^l\right) \\
&\quad \times \left\|\operatorname{cum}\left(\varepsilon_{u_1-s_k},\ldots,\varepsilon_{u_{k-1}-s_k},\varepsilon_{-s_k}\right)\right\|_\pi
\end{aligned}$$

Using Jensen's inequality and the k-stationarity of ε_t, we get

$$\begin{aligned}
&\leq \sum_{s_1,\ldots,s_{k-1}\in\mathbb{Z}} |||A_{s_1}|||_\infty \cdots |||A_{s_k}|||_\infty \sum_{u_1,\ldots,u_{k-1}\in\mathbb{Z}} \left(1+2^{l-1}|u_j|^l+2^{l-1}|s_j|^l+2^{l-1}|s_k|^l\right) \\
&\quad \times \left\|\operatorname{cum}\left(\varepsilon_{u_1},\ldots,\varepsilon_{u_{k-1}},\varepsilon_0\right)\right\|_\pi \\
&< \infty,
\end{aligned}$$

since (3.10.2) and (3.10.3) hold. □

3.11 Outlook

In this chapter, we have given some theory on the estimation of the spectral density operators. The basic idea was to take the functional discrete Fourier transforms of the data, then take its empirical operator to form the periodogram operators, which are then smoothed with a kernel to form consistent estimators of the spectral density operators. We have also seen that the eigenstructure of the sample spectral density operators form consistent estimators of the eigenstructure of the spectral density operators. This in turn could be used in practice to approximate a functional series by truncation of the Cramér–Karhunen–Loève, introduced in Chapter 2. The sample spectral density operators could also be used for other purposes, such as for classification, clustering, or prediction.

Since Functional Data are usually observed on a grid, and possibly contaminated by some noise, we studied the consistency of the sample spectral density operators constructed on the basis of such observations, and gave some results relating the mean/variance of the observation noise to the stretch of the series and the bandwidth parameter. The setting we chose was general enough to be applicable to various smoothing techniques (for turning the discrete observations into functional data). Extensions to the setup of sparsely observed functional time series would be of interest.

3.12 Some Technical Results for this Chapter

This section contains some technical results that were used in this Chapter.

3.12.1 Complex Gaussian Random Elements

If X is a Gaussian random element of the real Hilbert space $H_{\mathbb{R}}$, with mean μ and covariance operator $\mathscr{C} = \mathbb{E}\left[(X-\mu)\otimes_2(X-\mu)\right]$, we will write $X \sim \mathcal{N}_{H_{\mathbb{R}}}(\mu, \mathscr{C})$. If H is a complexified Hilbert space, the covariance of a random element $X \in H$ satisfying $\mathbb{E}\|X\|^2 < \infty$ is defined by $\mathbb{E}\left[(X-\mu)\otimes_2(X-\mu)\right]$, where $\mu = \mathbb{E}X$. In the case $H = \mathbb{C}^d$, the covariance of a random element $X \in \mathbb{C}^d$ is given by $\mathbb{E}\left[(X-\mu)(X-\mu)^\dagger\right] = \mathbb{E}\left[(X-\mu)\overline{(X-\mu)}^{\mathsf{T}}\right]$. In particular, this implies that the variance of a random element $X \in \mathbb{C}$ is

See Section A.2.3 on page 221 for the definition of complexified Hilbert space

$$\operatorname{var}(X) = \mathbb{E}\left[|X - \mathbb{E}X|^2\right].$$

We now define the notion of complex Gaussian random vector, which will be used in this Chapter.

Definition 3.12.1 (e.g. Picinbono (1996), Schreier & Scharf (2010))**.**
A random vector $Z = X + \mathbf{i}Y \in \mathbb{C}^d$, where $X = \Re(Z) \in \mathbb{R}^d$ and $Y = \Im(Z) \in \mathbb{R}^d$, is said to follow the complex Gaussian distribution *if the vector $(X^{\mathsf{T}}, Y^{\mathsf{T}})^{\mathsf{T}} \in \mathbb{R}^{2d}$ follows a multivariate Gaussian distribution. The distribution of a complex Gaussian random vector $Z \in \mathbb{C}^d$ is entirely characterized by the following parameters:*

$$\begin{aligned}
\mu &= \mathbb{E}[Z], \\
\Gamma &= \mathbb{E}\left[(Z-\mu)(Z-\mu)^\dagger\right] = \mathbb{E}\left[(Z-\mu)\otimes_2(Z-\mu)\right], \\
C &= \mathbb{E}\left[(Z-\mu)(Z-\mu)^{\mathsf{T}}\right] = \mathbb{E}\left[(Z-\mu)\otimes_2\overline{(Z-\mu)}\right].
\end{aligned}$$

The parameters μ, Γ, C are called the mean*, the* covariance matrix *and the* relation matrix*, respectively, and we write $Z \sim \mathcal{N}_{\mathbb{C}^d}(\mu, \Gamma, C)$.*
If $C = 0$, then Z is said to follow a circular complex Gaussian *distribution , and we write either $Z \sim \mathcal{N}_{\mathbb{C}^d}(\mu, \Gamma, 0)$, or $Z \sim \mathcal{N}_{\mathbb{C}^d}(\mu, \Gamma)$.*

Assume that $\mu = 0$, we have

$$\Gamma = (\mathbb{E}[X\otimes_2 X] + \mathbb{E}[Y\otimes_2 Y]) + \mathbf{i}(\mathbb{E}[Y\otimes_2 X] - \mathbb{E}[X\otimes_2 Y]),$$

and

$$C = (\mathbb{E}[X\otimes_2 X] - \mathbb{E}[Y\otimes_2 Y]) + \mathbf{i}(\mathbb{E}[Y\otimes_2 X] + \mathbb{E}[X\otimes_2 Y]).$$

We therefore get the following Lemma:

Lemma 3.12.2.

Suppose $Z = X + \mathbf{i}Y \in \mathbb{C}^d$ is a random vector, with mean zero, covariance matrix Γ and relation matrix C. Then

$$\mathbb{E}[X \otimes_2 X] = \frac{1}{2}\Re(\Gamma + C), \qquad \mathbb{E}[Y \otimes_2 Y] = \frac{1}{2}\Re(\Gamma - C).$$
$$\mathbb{E}[X \otimes_2 Y] = \frac{1}{2}\Im(C - \Gamma), \qquad \mathbb{E}[Y \otimes_2 X] = \frac{1}{2}\Im(C + \Gamma).$$

We also mention some useful properties:

1. *If $\Re(\Gamma - C) = 0$, then $Y = 0$ almost surely, and Z is almost surely a real vector.*

2. *If Z is circular (i.e. $C = 0$) and $d = 1$, then $\Re(Z)$ and $\Im(Z)$ are uncorrelated, and have the same variance.*

3. *If $Z = (Z_1, Z_2, \ldots, Z_q) \in \mathbb{C}^q$ is a complex Gaussian random element, with covariance matrix $\Gamma = (\Gamma_{ij})$ and relation matrix $C = (C_{ij})$. Let $I, J \subset \{1, \ldots, q\}$ with $I \cap J = \emptyset$, define $Z_I = (Z_i)_{i \in I}$, and similarly for Z_J. Then,*

$$\left[Z_I \text{ is independent of } Z_J\right] \iff \left[\Gamma_{ij} = 0 = C_{ij}, \forall i \in I, j \in J\right].$$

The last property mentionned in the Lemma tells us that two entries Z_q and Z_r of a complex Gaussian vector $Z = (Z_1, \ldots, Z_q)$ are independent if, and only if, $\mathbb{E}\left[Z_q \overline{Z_r}\right] = 0$ and $\mathbb{E}\left[Z_q Z_r\right] = 0$, or in other words, if their covariance and relation vanishes. This motivates the following definition:

Definition 3.12.3.
Two complex random variables $Z_1, Z_2 \in \mathbb{C}$ are said to be strongly uncorrelated *if $\mathbb{E}\left[Z_1 \overline{Z_2}\right] = 0$ and $\mathbb{E}[Z_1 Z_2] = 0$. If only the first equality holds, then Z_1 and Z_2 are* uncorrelated.

We now define complex Gaussian random elements in a general complexified Hilbert space.

Definition 3.12.4.
A random element Z of a complexified Hilbert space H is said to follow a complex Gaussian distribution if $\mathbb{E}\|Z\|^2 < \infty$, and all its finite dimensional projections follow compatible complex Gaussian distribution, i.e., for all $\varphi_1, \ldots, \varphi_J \in H$, and all $J = 1, 2, \ldots$, the random vector

$$\left(\langle Z, \varphi_1 \rangle, \ldots, \langle Z, \varphi_J \rangle\right) \in \mathbb{C}^J$$

follows a Gaussian distribution $\mathcal{N}_{\mathbb{C}^J}(\mu, \Gamma, C)$ satisfying, for all $i, j \in \{1, 2, \ldots, J\}$,

the following compatibility conditions

$$\begin{aligned}\mu_i &= \langle m, \varphi_i \rangle, \\ \Gamma_{ij} &= \langle \mathscr{G}\varphi_j, \varphi_i \rangle, \\ C_{ij} &= \langle \mathscr{C}\overline{\varphi_j}, \varphi_i \rangle,\end{aligned}$$

where $\mu = (\mu_1, \ldots, \mu_J)$, $\Gamma = (\Gamma_{ij})$, $C = (C_{ij})$ *and*

$$\begin{aligned}m &= \mathbb{E}[Z], \\ \mathscr{G} &= \mathbb{E}[(Z-m)\otimes_2(Z-m)] \in \mathscr{S}_1(H), \\ \mathscr{C} &= \mathbb{E}\left[(Z-m)\otimes_2\overline{(Z-m)}\right] \in \mathscr{S}_1(H).\end{aligned}$$

We write $Z \sim \mathscr{N}_H(m, \mathscr{G}, \mathscr{C})$. *If* $\mathscr{C} = 0$, *we say that* Z *follows a* circular complex Gaussian distribution , *and we write* $Z \sim \mathscr{N}_H(m, \mathscr{G}, 0)$ *or* $Z \sim \mathscr{N}_H(m, \mathscr{G})$.

3.12.2 Cumulants

3.12.2.1 Cumulants of Random Variables

We recall the definition of a cumulant for complex valued random variables:

$$\operatorname{cum}(Y_1, \ldots, Y_k) = \sum_{\nu} (-1)^{p-1}(p-1)! \prod_{l=1}^{p} \mathbb{E}\left[\prod_{j\in\nu_l} Y_j\right],$$

where the summation extends over all unordered partitions

$$\nu = (\nu_1, \ldots, \nu_p), \quad p = 1, \ldots, k,$$

of $\{1, \ldots, k\}$. The following result is found in Rosenblatt (1985, p.34):

Lemma 3.12.5.
If $\mathbb{E}|\prod_{j\in J} Y_j| < \infty$ *for all subset of indices* $J \subset \{1, \ldots, k\}$,

$$\mathbb{E}[Y_1 \cdots Y_k] = \sum_{\nu} \prod_{l=1}^{p} \operatorname{cum}\left(Y_j; j \in \nu_l\right), \tag{3.12.1}$$

where the sum extends over all unordered partitions $\nu = (\nu_1, \ldots, \nu_p)$ *of* $\{1, \ldots, k\}$.

This result tells us that the k-th order cumulant can be thought of as a generalization of the covariance. Indeed, rewriting (3.12.1), we get

$$\operatorname{cum}(Y_1, \ldots, Y_k) = \mathbb{E}[Y_1 \cdots Y_k] - \sum_{\nu; p\neq 1} \prod_{l=1}^{p} \operatorname{cum}\left(Y_j; j \in \nu_l\right), \tag{3.12.2}$$

which gives the intuition that the cumulant of order k is equal to the k-th

moment, with all lower-order dependencies removed.

3.12.2.2 Cumulants of Random Elements of Hilbert spaces

We now define cumulants for random elements of Hilbert spaces: Let $(H, \|\cdot\|)$ be a separable Hilbert space, and for k a positive integer, let $L^k = L^k(\Omega, H, \mathbb{P})$ be the Banach space of random elements $Y \in H$ with norm

$$\|Y\|_{L^k} = (\mathbb{E}\|Y\|^k)^{1/k} < \infty.$$

See Chapter B on page 229 for the definition of the space $L^k(\Omega, H, \mathbb{P})$

Proposition 3.12.6.

For any (random) elements $Y_1, \ldots, Y_k \in L^k$, we define their k-th order cumulant $\operatorname{cum}_k(Y_1, \ldots, Y_k) \in \otimes_{j=1}^k H$ *as the unique element that satisfies*

$$\left\langle \operatorname{cum}_k(Y_1, \ldots, Y_k), \varphi_1 \otimes \cdots \otimes \varphi_k \right\rangle = \operatorname{cum}\left(\left\langle Y_1, \varphi_1 \right\rangle, \ldots, \left\langle Y_k, \varphi_k \right\rangle\right), \quad (3.12.3)$$

$\forall \varphi_1, \ldots, \varphi_k \in H$*, and where the function* $\operatorname{cum}(\cdot)$ *on the right-hand side is the usual cumulant function for complex random variables.*

Furthermore, the mapping

$$\operatorname{cum}_k(\cdot) : L^k \times \cdots \times L^k \to \bigotimes_{j=1}^k H$$

is a bounded multilinear mapping, i.e., it is linear in each coordinate, and

$$\|\operatorname{cum}_k(Y_1, \ldots, Y_k)\| \leq (k-1)! B_k \|Y_1\|_{L^k} \cdots \|Y_k\|_{L^k},$$

where B_k is the k-th Bell number, giving the number of partitions of a set of size k. We sometimes write $\operatorname{cum}(\cdot)$ *instead of* $\operatorname{cum}_k(\cdot)$.

Proof. Let H^k denote the k-fold Cartesian product of H. Fix $t_1, \ldots, t_k \in \mathbb{Z}$. We define the multilinear functional $G : H^j \to \mathbb{C}$ by

$$G(\varphi_1, \ldots, \varphi_k) = \overline{\operatorname{cum}\left(\left\langle Y_1, \varphi_1 \right\rangle, \ldots, \left\langle Y_k, \varphi_k \right\rangle\right)},$$

for all $\varphi_1, \ldots, \varphi_k \in H$. It is clear that it is a multilinear functional. Furthermore, letting $c = \sum_\nu (p-1)!$, where the sum extends over all unordered partitions $\nu = (\nu_1, \ldots, \nu_p)$ of $\{1, \ldots, k\}$, and writing $|\nu_j|$ for the size of the partition ν_j, we get (using Jensen's and Hölder's inequality)

$$
\begin{aligned}
|G(\varphi_1,\dots,\varphi_k)| &\le \sum_{\nu}(p-1)!\prod_{l=1}^{p}\mathbb{E}\left[\prod_{j\in\nu_l}|\langle Y_j,\varphi_j\rangle|\right]\\
&\le \sum_{\nu}(p-1)!\prod_{l=1}^{p}\prod_{j\in\nu_l}\mathbb{E}\left[|\langle Y_j,\varphi_j\rangle|^{|\nu_l|}\right]^{1/|\nu_l|}\\
&\le \sum_{\nu}(p-1)!\prod_{l=1}^{p}\prod_{j\in\nu_l}\|\varphi_j\|\,\mathbb{E}\left[\|Y_j\|^{|\nu_l|}\right]^{1/|\nu_l|}\\
&= \|\varphi_1\|\cdots\|\varphi_n\|\sum_{\nu}(p-1)!\prod_{l=1}^{p}\prod_{j\in\nu_l}\left(\mathbb{E}\|Y_j\|^{k}\right)^{1/k}\\
&= \|\varphi_1\|\cdots\|\varphi_n\|\sum_{\nu}(p-1)!\prod_{j=1}^{k}\left(\mathbb{E}\|Y_j\|^{k}\right)^{1/k}\\
&= c\|\varphi_1\|\cdots\|\varphi_n\|\prod_{j=1}^{k}\left(\mathbb{E}\|Y_j\|^{k}\right)^{1/k}.
\end{aligned}
$$

Furthermore, letting $I_b=(e_n)_{n\ge1}$ be an orthonormal basis of H, and using Jensen's and Hölder's inequality, as well as Parseval's identity, we get

$$
\left(\sum_{\varphi_1,\dots,\varphi_k\in I_b}|G(\varphi_1,\dots,\varphi_k)|^2\right)^{1/2}\le(k-1)!\left(\sum_{\varphi_1,\dots,\varphi_k\in I_b}\left|\sum_{\nu}\prod_{l=1}^{p}\mathbb{E}\left[\prod_{j\in\nu_l}\langle Y_j,\varphi_j\rangle\right]\right|^2\right)^{1/2}
$$

using Jensen's inequality on $\sum_\nu$, we get

$$
\begin{aligned}
\left(\sum_{\varphi_1,\dots,\varphi_k\in I_b}|G(\varphi_1,\dots,\varphi_k)|^2\right)^{1/2} &\le (k-1)!\sum_{\nu}\left(\sum_{\varphi_1,\dots,\varphi_k\in I_b}\left|\prod_{l=1}^{p}\mathbb{E}\left[\prod_{j\in\nu_l}\langle Y_j,\varphi_j\rangle\right]\right|^2\right)^{1/2}\\
&= (k-1)!\sum_{\nu}\prod_{l=1}^{p}\left(\sum_{\varphi_j\in I_b:\,j\in\nu_l}\left|\mathbb{E}\left[\prod_{j\in\nu_l}\langle Y_j,\varphi_j\rangle\right]\right|^2\right)^{1/2}
\end{aligned}
$$

again, using Jensen's inequality on the expectation yields

$$
\begin{aligned}
\left(\sum_{\varphi_1,\dots,\varphi_k\in I_b}|G(\varphi_1,\dots,\varphi_k)|^2\right)^{1/2} &\le (k-1)!\sum_{\nu}\prod_{l=1}^{p}\mathbb{E}\left[\left(\sum_{\varphi_j\in I_b:\,j\in\nu_l}\left|\prod_{j\in\nu_l}\langle Y_j,\varphi_j\rangle\right|^2\right)^{1/2}\right]\\
&= (k-1)!\sum_{\nu}\prod_{l=1}^{p}\mathbb{E}\left[\left(\prod_{j\in\nu_l}\sum_{\varphi_j\in I_b}|\langle Y_j,\varphi_j\rangle|^2\right)^{1/2}\right]\\
&= (k-1)!\sum_{\nu}\prod_{l=1}^{p}\mathbb{E}\left[\prod_{j\in\nu_l}\|Y_j\|\right]\\
&\le (k-1)!\sum_{\nu}\prod_{l=1}^{p}\prod_{j\in\nu_l}\left[\mathbb{E}\|Y_j\|^{|\nu_l|}\right]^{1/|\nu_l|}
\end{aligned}
$$

$$
\begin{aligned}
&\leq (k-1)! \sum_{\nu} \prod_{l=1}^{p} \prod_{j \in \nu_l} \left[\mathbb{E} \left\| Y_j \right\|^k \right]^{1/k} \\
&\leq (k-1)! \sum_{\nu} \prod_{j=1}^{k} \left\| Y_j \right\|_{L^k} \\
&\leq (k-1)! \prod_{j=1}^{k} \left\| Y_j \right\|_{L^k} \left(\sum_{\nu} 1 \right) \\
&\leq (k-1)! B_k \prod_{j=1}^{k} \left\| Y_j \right\|_{L^k},
\end{aligned}
$$

where B_k is the k-th Bell number, giving the number of partitions of a set of size k. Therefore G is a Hilbert–Schmidt functional, and in particular it is a weak Hilbert–Schmidt mapping, see e.g. Kadison & Ringrose (1997, Theorem 2.6.4). Therefore, by the universal property of the Hilbert tensor product, there exists a unique mapping continuous linear functional $G' : \otimes_{j=1}^{k} H \to \mathbb{C}$ such that

$$
G'(\varphi_1 \otimes \cdots \otimes \varphi_n) = G(\varphi_1, \ldots, \varphi_n), \quad \forall \varphi_1, \ldots, \varphi_n \in H.
$$

Now the Riesz representation Theorem tells us that there exists a unique $A \in \otimes_{j=1}^{k} H$ such that

$$
G'(\cdot) = \langle ., A \rangle.
$$

Therefore $\operatorname{cum}_k(Y_1, \ldots, Y_k) = A$ is well defined. The linearity of $\operatorname{cum}_k(\cdot)$ follows directly from the property (3.12.3) and the boundedness has already been shown since

$$
\left\| \operatorname{cum}_k(Y_1, \ldots, Y_k) \right\|^2 = \sum_{\varphi_1, \ldots, \varphi_k \in I_b} \left| G(\varphi_1, \ldots, \varphi_k) \right|^2,
$$

by Parseval's equality. □

Lemma 3.12.7.
Using the same notation as Proposition 3.12.6, if $A_1, \ldots, A_k$ are bounded operators on H, then

$$
\operatorname{cum}(A_1 Y_1, \ldots, A_k X_k) = \left(A_1 \widetilde{\bigotimes} \cdots \widetilde{\bigotimes} A_k \right) \operatorname{cum}(Y_1, \ldots, Y_k).
$$

Proof. The proof follows directly from (3.12.3), and is omitted. □

We can also define elements of the form $\operatorname{cum}\left(\langle \varphi, Y_1 \rangle, Y_2, \ldots, Y_k \right) \in H^{\otimes (k-1)}$, for $\varphi \in H$. This is done rigorously by defining, for any $i_1, \ldots, i_s \in \{1, \ldots, k\}$ and any $\varphi_{i_1}, \ldots, \varphi_{i_s} \in H$ the element

$$
\operatorname{cum}\left(Y_1', \ldots, Y_k' \right) \in H^{\otimes (k-s)}
$$

where $Y_{i_r}' = \langle \varphi_{i_r}, Y_{i_r} \rangle$ for $r = 1, \ldots, s$, and $Y_j' = Y_j$ if $j \notin \{i_1, \ldots, i_s\}$, to be the

unique element of $H^{\otimes(k-s)}$ satisfying

$$\left\langle \operatorname{cum}\left(Y_1', \ldots, Y_k'\right), \varphi_{j_1} \otimes \cdots \varphi_{j_{k-s}} \right\rangle = \left\langle \operatorname{cum}\left(Y_1, \ldots, Y_k\right), \varphi_1 \otimes \cdots \otimes \varphi_k \right\rangle,$$

for all $\varphi_{j_1}, \ldots, \varphi_{j_{k-s}} \in H$, where $\left\{j_1, \ldots, j_{k-s}\right\}$ is the complementary set of indices $\{i_1, \ldots, i_s\}$, i.e.

$$\left\{j_1, \ldots, j_{k-s}\right\} = \{1, \ldots, k\} \setminus \{i_1, \ldots, i_s\}.$$

We now present a way of writing cumulants using the tensor product:

Proposition 3.12.8.

Let $Y_1, \ldots, Y_k \in L^k$. The cumulant can be defined via the tensor product as

$$\operatorname{cum}(Y_1, \ldots, Y_k) = \sum_{\nu=(\nu_1,\ldots,\nu_p)} (-1)^{p-1}(p-1)! \cdot \operatorname{Perm}_{\nu^{-1}} \left(\bigotimes_{l=1}^{p} \mathbb{E}\left[\bigotimes_{j \in \nu_l} Y_j \right] \right), \tag{3.12.4}$$

and this definition is compatible with Definition 3.2.1. Furthermore, we can view $\operatorname{cum}(\cdot)$ *as a bounded linear operator*

$$\operatorname{cum}(\cdot) : \left(L^k\right)^{\otimes_\pi k} \to H^{\otimes_\pi k}. \tag{3.12.5}$$

We will first give some intuition about why the operator $\operatorname{Perm}_{\nu^{-1}}(\cdot)$ is important here. To understand why, recall that any partition $\nu = (\nu_1, \ldots, \nu_p)$ can also be viewed as a permutation $\nu : \{1, \ldots, k\} \to \{1, \ldots, k\}$, sending j to $\nu(j)$, where we write $\nu = (\nu(1), \ldots, \nu(l))$. The inverse is denoted ν^{-1}. Notice that

$$\bigotimes_{l=1}^{p} \left(\bigotimes_{j \in \nu_l} Y_j \right) = Y_{\nu(1)} \otimes Y_{\nu(2)} \otimes \cdots \otimes Y_{\nu(k)}.$$

Therefore we need to reorder the tensor product after having taken the expectations, since we want the property

$$\left\langle \operatorname{cum}(Y_1, \ldots, Y_k), \varphi_1 \otimes \varphi_2 \otimes \cdots \otimes \varphi_k \right\rangle = \operatorname{cum}\left(\left\langle Y_1, \varphi_1 \right\rangle, \ldots, \left\langle Y_k, \varphi_k \right\rangle \right), \tag{3.12.6}$$

Proof of Proposition 3.12.8. It suffices to use the properties of $\operatorname{Perm}_\nu(\cdot)$, and the characterization (3.12.3) of the cumulant to show the compatibility with Definition 3.2.1.

Let us now show that $\operatorname{cum}(Y_1, \ldots, Y_k) \in H^{\otimes_\pi k}$:

$$\|\operatorname{cum}(Y_1, \ldots, Y_k)\|_\pi \leq \sum_{\nu=(\nu_1,\ldots,\nu_p)} (p-1)! \cdot \left\| \operatorname{Perm}_{\nu^{-1}} \left(\bigotimes_{l=1}^{p} \mathbb{E}\left[\bigotimes_{j \in \nu_l} Y_j \right] \right) \right\|_\pi.$$

Using the fact that the projective tensor product is isomorphically commutative (see Section A.3.3),

$$\begin{aligned}\|\mathrm{cum}(Y_1,\dots,Y_k)\|_\pi &= \sum_{\nu=(\nu_1,\dots,\nu_p)} (p-1)!\cdot \left\| \bigotimes_{l=1}^{p} \mathbb{E}\left[\bigotimes_{j\in\nu_l} Y_j\right]\right\|_\pi \\ &= \sum_{\nu=(\nu_1,\dots,\nu_p)} (p-1)!\cdot \prod_{l=1}^{p} \left\| \mathbb{E}\left[\bigotimes_{j\in\nu_l} Y_j\right]\right\|_\pi, \quad \text{(by (A.3.5))}\end{aligned}$$

Using the contraction property of the Bochner integral,

$$\begin{aligned}\|\mathrm{cum}(Y_1,\dots,Y_k)\|_\pi &\le \sum_{\nu=(\nu_1,\dots,\nu_p)} (p-1)!\cdot \prod_{l=1}^{p} \mathbb{E}\left\| \bigotimes_{j\in\nu_l} Y_j\right\|_\pi \\ &= \sum_{\nu=(\nu_1,\dots,\nu_p)} (p-1)!\cdot \prod_{l=1}^{p} \mathbb{E}\prod_{j\in\nu_l} \|Y_j\|, \qquad \text{(by (A.3.5))}\end{aligned}$$

and using Hölder's inequality,

$$\begin{aligned}\|\mathrm{cum}(Y_1,\dots,Y_k)\|_\pi &\le \sum_{\nu=(\nu_1,\dots,\nu_p)} (p-1)!\cdot \prod_{j=1}^{k} \|Y_j\|_{L^k} \\ &< \infty.\end{aligned}$$

This shows that $\mathrm{cum}(\cdot) : \left(L^k\right)^k \to H^{\otimes_\pi k}$ is bounded. Furthermore, we already know that it is multilinear. Therefore, by the universal property of the projective tensor products (A.3.9), we can view $\mathrm{cum}(\cdot)$ as a bounded linear operator $(L^k)^{\otimes_\pi k} \to H^{\otimes_\pi k}$. □

Corollary 3.12.9. *Let $s > 1$ be an integer. If $Y_1,\dots,Y_{2s} \in L^{2s}$, then*

$$\left|\left|\left|\mathrm{cumop}(Y_1,\dots,Y_{2s})\right|\right|\right|_1 \le \|\mathrm{cum}(Y_1,\dots,Y_{2s})\|_\pi < \infty. \tag{3.12.7}$$

Proof. We denote $\Psi : H^{\otimes s} \otimes_\pi H^{\otimes s} \to \mathscr{S}_2(H^{\otimes s})$ the unitary operator $\Psi(x\otimes y) = x \otimes_2 \overline{y}$ for $x, y \in H^{\otimes s}$ (see Proposition A.3.2). Also, we denote by $\iota : H^{\otimes_\pi s} \hookrightarrow H^{\otimes s}$ the continuous inclusion defined by linear extension of $\iota(\otimes_{i=1}^s x_i) = \otimes_{i=1}^s x_i$ for all simple tensors. Since $\mathrm{cum}(Y_1,\dots,Y_{2s}) \in H^{\otimes_\pi 2s}$, and

$$H^{\otimes_\pi 2s} = \left(H^{\otimes_\pi s}\right) \otimes_\pi \left(H^{\otimes_\pi s}\right),$$

we have

$$(\iota\otimes_\pi \iota)\mathrm{cum}(Y_1,\dots,Y_{2s}) \in H^{\otimes s} \otimes_\pi H^{\otimes s},$$

and therefore $\Psi(\iota\otimes_\pi\iota)\,\mathrm{cum}(Y_1,\dots,Y_{2s})$ is well defined. Notice that

$$\mathrm{cumop}(Y_1,\dots,Y_{2s}) = \Psi(\iota\otimes_\pi\iota)\,\mathrm{cum}(Y_1,\dots,Y_{2s}),$$

since the equality holds on all simple tensors. We therefore get,

$$\left|\left|\left|\mathrm{cumop}\,(Y_1,\ldots,Y_{2s})\right|\right|\right|_1 \le |||\Psi|||_\infty |||\iota|||_\infty^2 \|\mathrm{cum}\,(Y_1,\ldots,Y_{2s})\|_\pi$$

and the proof is finished by using the isometry property of Ψ, the fact that $|||\iota|||_\infty \le 1$, and Proposition 3.12.8. □

Proposition 3.12.10.
Let $A_1,\ldots,A_k$ be bounded linear operators on H, and $Y_1,\ldots,Y_k \in L^k$. Then

$$(A_1 \otimes_\pi \cdots \otimes_\pi A_k)\,\mathrm{cum}\,(Y_1,\ldots,Y_k) = \mathrm{cum}\,(A_1Y_1,\ldots,A_kY_k)\,. \tag{3.12.8}$$

Proof.

$$\begin{aligned}
&(A_1 \otimes_\pi \cdots \otimes_\pi A_k)\,\mathrm{cum}\,(Y_1,\ldots,Y_k)\\
&= \sum_{\nu=(\nu_1,\ldots,\nu_p)} (-1)^{p-1}(p-1)!\cdot\left(\bigotimes_{i=1}^{k}{}_\pi\, A_i\right)\mathrm{Perm}_{\nu^{-1}}\left(\bigotimes_{l=1}^{p}\mathbb{E}\left[\bigotimes_{j\in\nu_l} Y_j\right]\right),\\
&= \sum_{\nu}(-1)^{p-1}(p-1)!\cdot\mathrm{Perm}_{\nu^{-1}}\left(\mathrm{Perm}_{\nu}\left(\bigotimes_{i=1}^{k}{}_\pi\, A_i\right)\left[\bigotimes_{l=1}^{p}\mathbb{E}\left[\bigotimes_{j\in\nu_l} Y_j\right]\right]\right),\\
&\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\text{(by (A.3.11))}\\
&= \sum_{\nu}(-1)^{p-1}(p-1)!\cdot\mathrm{Perm}_{\nu^{-1}}\left(\left[\bigotimes_{l=1}^{p}{}_\pi\left[\bigotimes_{j\in\nu_l}{}_\pi\, A_j\right]\right]\left[\bigotimes_{l=1}^{p}\mathbb{E}\left[\bigotimes_{j\in\nu_l} Y_j\right]\right]\right).
\end{aligned}$$

Using the definition of the tensor of bounded operators, and the commutativity of bounded operators with the expectation,

$$\begin{aligned}
&(A_1 \otimes_\pi \cdots \otimes_\pi A_k)\,\mathrm{cum}\,(Y_1,\ldots,Y_k)\\
&= \sum_{\nu}(-1)^{p-1}(p-1)!\cdot\mathrm{Perm}_{\nu^{-1}}\left(\bigotimes_{l=1}^{p}\left[\left(\bigotimes_{j\in\nu_l}{}_\pi\, A_j\right)\mathbb{E}\left[\bigotimes_{j\in\nu_l} Y_j\right]\right]\right),\\
&= \sum_{\nu}(-1)^{p-1}(p-1)!\cdot\mathrm{Perm}_{\nu^{-1}}\left(\bigotimes_{l=1}^{p}\left[\mathbb{E}\left[\left(\bigotimes_{j\in\nu_l}{}_\pi\, A_j\right)\bigotimes_{j\in\nu_l} Y_j\right]\right]\right),\\
&= \sum_{\nu}(-1)^{p-1}(p-1)!\cdot\mathrm{Perm}_{\nu^{-1}}\left(\bigotimes_{l=1}^{p}\mathbb{E}\left[\bigotimes_{j\in\nu_l} A_jY_j\right]\right),\\
&= \mathrm{cum}\,(A_1Y_1,\ldots,A_kY_k)
\end{aligned}$$

□

Proposition 3.12.11.
Let A_s be bounded operators on H, satisfying

$$\sum_{s\in\mathbb{Z}} |||A_s|||_\infty < \infty.$$

Let ε_t, $t = 0,\pm1,\ldots$ be a k-order stationary sequence of random elements of H such that $\mathbb{E}\,\|\varepsilon_0\|^k < \infty$ for some positive integer k,

Define $X_t^{(N)} = \sum_{s=-N}^{N} A_s \varepsilon_{t-s}$. *Then, the* linear process $X_t = \sum_{s\in\mathbb{Z}} A_s \varepsilon_{t-s}$ *has the following properties:*

(i) L^k *convergence:* $\lim_{N\to\infty} \left\| X_t^{(N)} - X_t \right\|_{L^k} = 0$, *and* $\mathbb{E}\|X_t\|^k < \infty$,

(ii) $\left\|\operatorname{cum}\left(\varepsilon_{t_1},\ldots,\varepsilon_{t_k}\right)\right\|_\pi < \infty$.

(iii) *We have*

$$\operatorname{cum}\left(X_{t_1},\ldots,X_{t_k}\right) = \sum_{s_1,\ldots,s_k\in\mathbb{Z}} (A_{t_1-s_1} \otimes_\pi \cdots \otimes_\pi A_{t_k-s_k}) \operatorname{cum}\left(\varepsilon_{s_1},\ldots,\varepsilon_{s_k}\right),$$

where the convergence is in $\|\cdot\|_\pi$.

Furthermore, X_t *is* k*-th order stationary.*

Remark 3.12.12. *Since the projective norm* $\|\cdot\|_\pi$ *is stronger than the Hilbert tensor norm* $\|\cdot\|$, *the statements* (ii) *and* (iii) *of the Proposition also hold for the Hilbert tensor norm.*

Proof of Proposition 3.12.11. 1. Let

$$X_t^{-(N)} = X_t - X_t^{(N)} = \sum_{|s|>N} A_s \varepsilon_{t-s},$$

for the tail of the series of X_t. Since

$$\|A_s \varepsilon_{t-s}\|_{L^k} \le |||A_s|||_\infty \|\varepsilon_{t-s}\|_{L^k},$$

we get, by stationarity of ε_s,

$$\left\| X_t^{-(N)} \right\|_{L^k} \le \sum_{|s|>N} \|A_s \varepsilon_{t-s}\|_{L^k} \le \sum_{|s|>N} |||A_s|||_\infty \|\varepsilon_{t-s}\|_{L^k} = \|\varepsilon_0\|_{L^k} \sum_{|s|>N} |||A_s|||_\infty.$$

Therefore $X_t^{(N)} \to X_t$ in L^k as $N \to \infty$.

2. By Proposition 3.12.8, and by k-stationarity of ε_t, we get that

$$\left\|\operatorname{cum}\left(\varepsilon_{t_1,\ldots,\varepsilon_{t_k}}\right)\right\|_\pi \le c\, \mathbb{E}\|\varepsilon_0\|^k < \infty,$$

where c is a constant.

3. Let us show the next statement: using Proposition 3.12.8 and Propo-

sition 3.12.10, we get

$$\begin{aligned}
\operatorname{cum}\left(X_{t_1},\dots,X_{t_k}\right) &= \operatorname{cum}\left(\lim_{N_1\to\infty} X_{t_1}^{(N_1)},\dots,\lim_{N_k\to\infty} X_{t_k}^{(N_k)}\right)\\
&= \lim_{N_1,\dots,N_k\to\infty} \operatorname{cum}\left(\sum_{|s_1|<N_1} A_{s_1}\varepsilon_{t_1-s_1},\dots,\sum_{|s_k|<N_k} A_{s_k}\varepsilon_{t_k-s_k}\right)\\
&= \lim_{N_1,\dots,N_k\to\infty} \sum_{|s_1|<N_1}\cdots\sum_{|s_k|<N_k} \operatorname{cum}\left(A_{s_1}\varepsilon_{t_1-s_1},\dots,A_{s_k}\varepsilon_{t_k-s_k}\right)\\
&= \lim_{N_1,\dots,N_k\to\infty} \sum_{|s_1|<N_1}\cdots\sum_{|s_k|<N_k} \left(A_{s_1}\otimes_\pi\cdots\otimes_\pi A_{s_k}\right)\operatorname{cum}\left(\varepsilon_{t_1-s_1},\dots,\varepsilon_{t}\right.\\
&= \sum_{s_1,\dots,s_k\in\mathbb{Z}} \left(A_{s_1}\otimes_\pi\cdots\otimes_\pi A_{s_k}\right)\operatorname{cum}\left(\varepsilon_{t_1-s_1},\dots,\varepsilon_{t_k-s_k}\right),
\end{aligned}$$

where there is no convergence problems since the last sum is bounded by

$$c\cdot\mathbb{E}\|\varepsilon_0\|^k\left(\sum_{s\in\mathbb{Z}}|||A_s|||_\infty\right)^k<\infty.$$

The k-th order stationarity follows directly from property 3. This completes the proof. □

3.12.3 Some Accessory Results

Lemma 3.12.13.
*If $\int_{-\infty}^{\infty}|x|^p W(x)\,dx<\infty$ and **C(p,2)** holds true, then*

$$\int_{\mathbb{R}} W(x) f_{\omega-xB_T}\,dx = f_\omega + \\ +\sum_{k=1}^{p-1}\frac{(-1)^k B_T^k}{k!}\frac{\partial^k f_\omega}{\partial\omega^k}\cdot\int_{\mathbb{R}} x^k W(x)\,dx + O(B_T^p),$$

in L^2, and the error term is uniform in ω. Notice that since W is even, the integral is zero if k is odd. The case $p=1$ will be useful for consistent estimation of f_ω:

$$\int W(x) f_{\omega-xB_T}\,dx = f_\omega + O(B_T),\quad \text{in } L^2,$$

the error term being uniform in ω.

Proof. In the following, all equalities are meant in the L^2 sense with respect to the variables τ,σ. Since for every $\varphi\in L^2([0,1],\mathbb{C})$, the mapping

$$\omega\mapsto\frac{\partial^k}{\partial\omega^k}\langle f_\omega,\varphi\rangle$$

is continuous, we can write the Taylor expansion of $f_{\omega - xB_T}(\tau, \sigma)$, with respect to x at $x = 0$:

$$\begin{aligned} f_{\omega - xB_T}(\tau,\sigma) = f_\omega(\tau,\sigma) &- B_T \frac{\partial f_\alpha(\tau,\sigma)}{\partial \alpha}\Big|_{\alpha=\omega} x + \cdots \\ &+ \frac{(-1)^{p-1} B_T^{p-1}}{(p-1)!} \frac{\partial^{p-1} f_\alpha(\tau,\sigma)}{\partial \alpha^{p-1}}\Big|_{\alpha=\omega} x^{p-1} \\ &+ R_p(x,\omega,\tau,\sigma), \end{aligned}$$

where

$$R_p(x,\omega,\tau,\sigma) = \frac{(-1)^p B_T^p}{p!} \frac{\partial^p f_\alpha(\tau,\sigma)}{\partial \alpha^p}\Big|_{\alpha=\omega-\theta_x B_T} x^p,$$

and $\theta_x \in [0, x]$. This expression is bounded in L^2 by

$$\left\| R_p(x,\omega,\cdot,\cdot) \right\|_2 \le \frac{B_T^p}{p!} \sup_\omega \left\| \frac{\partial^p}{\partial \omega^p} f_\omega \right\|_2 |x|^p,$$

which does not depend on ω. Hence we obtain

$$\begin{aligned} \int_{\mathbb{R}} W(x) f_{\omega - xB_T}\, dx = f_\omega &+ \sum_{k=1}^{p-1} \frac{(-1)^k B_T^k}{k!} \frac{\partial^k f_\omega}{\partial \omega^k} \cdot \int_{\mathbb{R}} x^k W(x)\, dx \\ &+ \underbrace{\frac{B_T^p}{p!} \sup_\omega \left\| \frac{\partial^p}{\partial \omega^p} f_\omega \right\|_2 \int_{\mathbb{R}} |x|^p W(x)\, dx}_{= O(B_T^p)}, \end{aligned}$$

and the error is uniform in ω. □

Lemma 3.12.14.
Let $x_1, \ldots, x_n; y_1, \ldots, y_n$ be (complex or real) numbers bounded by K. Then $\left| x_1 x_2 \cdots x_n - y_1 y_2 \cdots y_n \right| \le K^{n-1} \sum_{j=1}^n |x_j - y_j|$.

Proof. Rewriting the expression in a suitable way yields the result:

$$\begin{aligned} \left| x_1 x_2 \cdots x_n - y_1 y_2 \cdots y_n \right| &= \left| \sum_{k=1}^n \left(\prod_{j=1}^{k-1} y_j \prod_{l=k}^n x_l - \prod_{j=1}^k y_j \prod_{l=k+1}^n x_l \right) \right| \\ &\le \sum_{k=1}^n \prod_{j=1}^{k-1} \prod_{l=k+1}^n |y_j x_l| |x_k - y_k| \\ &\le K^{n-1} \sum_{k=1}^n |x_k - y_k|. \end{aligned}$$

□

Denote by $V_a^b(h)$ the total variation of a function $h : [a, b] \to \mathbb{C}$ (Wheeden & Zygmund 1977, Chapter 2.1).

Lemma 3.12.15.
Let $f, f_1, \ldots, f_n : [a,b] \to \mathbb{C}$ be bounded in variation and bounded. Let $\|f\|_\infty = \sup_{x\in[a,b]} |f(x)|$. Then,

(i) $V_a^b\left(\prod_{j=1}^n f_j\right) \le \sum_{i=1}^n V_a^b(f_i) \prod_{j\neq i} \|f_j\|_\infty$.

(ii) If $\psi : [c,d] \to [a,b]$ is a strictly increasing bijection, $V_a^b(f) = V_{\psi(c)}^{\psi(d)}(f) = V_c^d(f\circ\psi)$. If $\psi : [c,d] \to [a,b]$ is a strictly decreasing bijection, $V_a^b(f) = V_d^c(f\circ\psi)$.

(iii) For any $a < c < b$, we have $V_a^c(f) + V_c^b(f) = V_a^b(f)$, and hence for any $a \le c \le d \le b$, we have $V_c^d(f) \le V_a^b(f)$.

(iv) For any $\lambda \in \mathbb{C}$, $V_a^b(\lambda f) = |\lambda| V_a^b(f)$.

(v) Triangle inequality: $V_a^b(f_1 + f_2) \le V_a^b(f_1) + V_a^b(f_2)$.

(vi) If f is continuous on $[a,b]$, f' exists on (a,b) and is Riemann integrable on $[a,b]$, $V_a^b(f) = \int_a^b |f'(x)|dx$.

(vii) If $f : \mathbb{R} \to \mathbb{C}$ is 2π-periodic, and $g(x) = f(\omega - x)$ for some $\omega \in \mathbb{R}$, then $V_0^{2\pi}(g) = V_0^{2\pi}(f)$.

(viii) If $V_a^b(f) < \infty$, then f is bounded on $[a,b]$.

We notice that the total variation has some of the properties of a norm.

Proof. Recall the definition of total variation: let $\Gamma = \{x_0, \ldots, x_m\}$ be a partition of $[a,b]$, that is,

$$a = x_0 < x_1 < \cdots < x_m = b,$$

and define

$$S_\Gamma(f) = \sum_{i=1}^m \left|f(x_i) - f(x_{i-1})\right|.$$

The total variation of f between a and b is

$$V(f) = \sup_\Gamma S_\Gamma(f),$$

where the supremum is taken over all partitions Γ of $[a,b]$.
Hence for any such partitions, and any $f, g : [a,b] \to \mathbb{C}$,

$$\begin{aligned}
\sum_{i=1}^m \left|f(x_i)g(x_i) - f(x_{i-1})g(x_{i-1})\right| &\le \sum_{i=1}^m \left|f(x_i)\left[g(x_i) - g(x_{i-1})\right]\right| \cdot \left|g(x_{i-1})\left[f(x_i) - f(x_{i-1})\right]\right| \\
&\le \|f\|_\infty \sum_{i=1}^m \left|g(x_i) - g(x_{i-1})\right| + \|g\|_\infty \sum_{i=1}^m \left|f(x_i) - f(x_{i-1})\right| \\
&\le \|f\|_\infty V(g) + \|g\| V(f).
\end{aligned}$$

Taking the supremum over all partitions yields

$$V(f \cdot g) \leq \|f\|_\infty V(g) + \|g\|_\infty V(f).$$

We then obtain (i) by induction.

For (ii), first assume that ψ is a strictly increasing bijection. Therefore $\{x_0, \ldots, x_m\}$ is a partition of $[c,d]$ if and only if $\{\psi(x_0), \ldots, \psi(x_m)\}$ is a partition of $[a,b]$. If ψ is a strictly decreasing bijection, $\{x_0, \ldots, x_m\}$ is a partition of $[c,d]$ if and only if $\{\psi(x_m), \psi(x_{m-1}), \ldots, \psi(x_0)\}$ is a partition of $[a,b]$. Statement (ii) is a direct consequence of these one-to-one correspondences between the partitions of $[a,b]$ and $[c,d]$.

Statements (iii), (vi) and $(viii)$ are proven in Wheeden & Zygmund (1977, Chapter 2.1).

For (iv), notice that

$$\sum_{i=1}^{m} \left|(\lambda f)(x_i) - (\lambda f)(x_{i-1})\right| = |\lambda| \cdot \sum_{i=1}^{m} \left|f(x_i) - f(x_{i-1})\right|.$$

For (v), notice that

$$\left|(f_1 + f_2)(x_i) - (f_1 + f_2)(x_{i-1})\right| \leq \left|f_1(x_i) - f_1(x_{i-1})\right| + \left|f_2(x_i) - f_2(x_{i-1})\right|.$$

Taking the sum over i and the supremum on all partitions $\{x_0, \ldots, x_m\}$ of $[a,b]$ yields the (v).

For (vii), define $g(x) = f(\omega - x)$. Notice that $V_a^b(f) = V_{a+2\pi}^{b+2\pi}(f)$ because f is 2π-periodic. Hence without loss of generality, we can assume that $0 < \omega < 2\pi$. Using (iii) we obtain

$$V_0^{2\pi}(f) = V_0^{\omega}(f) + V_\omega^{2\pi}(f) = V_{2\pi}^{\omega+2\pi}(f) + V_\omega^{2\pi}(f) = V_\omega^{\omega+2\pi}(f).$$

Choosing $\psi(x) = \omega - x$, (ii) yields

$$V_\omega^{\omega+2\pi}(f) = V_{-2\pi}^{0}(g) = V_0^{2\pi}(g). \qquad \square$$

We introduce the following condition, which will be used in the following results:

T1 (Taper condition 1) Let $h(u), -\infty < u < \infty$ be a real function, which is bounded, bounded in variation and with $h(u) = 0$ for $|u| \geq 1$. We denote by $\|h\|_\infty$ its supremum, and by $V_{-1}^{1}(h)$ its total variation (between -1 and 1).

Lemma 3.12.16.
*Suppose $h_{a_1},\dots,h_{a_k}$ satisfy **T1**, set $h_{a_j}^{(T)}(t) = h_{a_j}(t/T)$ and define $H_{a_1,\dots,a_k}^{(T)}(\omega) = \sum_{t\in\mathbb{Z}}\left[\prod_{j=1}^{k} h_{a_j}^{(T)}(t)\right]\exp(-\mathbf{i}\omega t)$. We have the following inequality for all $u_1,\dots,u_{k-1}\in\mathbb{Z}$:*

$$\left|\sum_{t\in\mathbb{Z}} h_{a_1}^{(T)}(t+u_1)\cdots h_{a_{k-1}}^{(T)}(t+u_{k-1})h_{a_k}^{(T)}(t)\exp(-\mathbf{i}\omega t) - H_{a_1,\dots,a_k}^{(T)}(\omega)\right| \le K\Big(|u_1|+\cdots+|u_{k-1}|\Big),$$

where $K = \left(\max_{j=1,\dots,k}\|h_{a_j}\|_\infty\right)^{k-1}\cdot\left(\max_{j=1,\dots,k} V_{-1}^{1}(h_{a_j})\right)$ is independent of ω.

Proof. Using Lemma 3.12.14, we obtain

$$\begin{aligned}
&\left|\sum_{t\in\mathbb{Z}} h_{a_1}^{(T)}(t+u_1)\cdots h_{a_{k-1}}^{(T)}(t+u_{k-1})h_{a_k}^{(T)}(t)\exp(-\mathbf{i}\omega t) - H_{a_1,\dots,a_k}^{(T)}(\omega)\right| \\
&\le \sum_{t\in\mathbb{Z}}\left|h_{a_k}^{(T)}(t)\right|\cdot\left|h_{a_1}^{(T)}(t+u_1)\cdots h_{a_{k-1}}^{(T)}(t+u_{k-1}) - h_{a_1}^{(T)}(t)\cdots h_{a_{k-1}}^{(T)}(t)\right| \\
&\le \|h_{a_1}\|_\infty\left(\max_{j=1,\dots,k-1}\|h_{a_j}\|_\infty\right)^{k-2}\cdot\sum_{j=1}^{k-1}\sum_{t\in\mathbb{Z}}\left|h_{a_j}^{(T)}(t+u_j) - h_{a_1}^{(T)}(t)\right|.
\end{aligned}$$

Let us now bound $\sum_{t\in\mathbb{Z}}\left|h_{a_j}^{(T)}(t+u_j) - h_{a_1}^{(T)}(t)\right|$. For simplicity, we suppress the indices. If $u>0$,

$$\begin{aligned}
\sum_{t\in\mathbb{Z}}\left|h_a^{(T)}(t+u) - h_a^{(T)}(t)\right| &\le \sum_{v=0}^{u_j-1}\sum_{t\in\mathbb{Z}}\left|h_a^{(T)}(t+v+1) - h_a^{(T)}(t+v)\right| \\
&\le \sum_{v=0}^{u_j-1} V_{-T}^{T}(h_a^{(T)}) \\
&= |u_j| V_{-1}^{1}(h_a),
\end{aligned}$$

where we have used Lemma 3.12.15 for the last equality. If $u<0$, we simply replace $\sum_{v=0}^{u_j-1}$ with $\sum_{u_j+1}^{0}$, and the same bound holds. Hence

$$\left|\sum_{t\in\mathbb{Z}} h_{a_1}^{(T)}(t+u_1)\cdots h_{a_{k-1}}^{(T)}(t+u_{k-1})h_{a_k}^{(T)}(t)\exp(-\mathbf{i}\omega t) - H_{a_1,\dots,a_k}^{(T)}(\omega)\right| \le K\Big(|u_1|+\cdots+|u_{k-1}|\Big)\square$$

We shall require the following lemma to quantify the approximation error of integrals by Riemann sums.

Lemma 3.12.17.
Let $h:[a,b]\to\mathbb{R}$ be a function of bounded variation. If $\Delta_n = \int_a^b h(t)\,dt - \frac{b-a}{n}\sum_{j=1}^{n} h(a+(b-a)j/n)$, then $|\Delta_n| \le (b-a)\frac{V_a^b(h)}{n}$. If the sum goes only from 1 to $(n-1)$, $|\Delta_n| \le (b-a)\frac{V_a^b(h)+h(b)}{n}$.

Proof. First let us prove the Lemma for $f : [0,1] \to \mathbb{R}$ of bounded variation. Since f is of bounded variation, it is bounded and has a finite number of discontinuities, and is hence Riemann integrable on $[0,1]$ (Wheeden & Zygmund 1977, see results (2.1), (2.8) and (5.54)). The following comes from Pólya & Szegő (1972, Pt.2, Chapter 1, problem 9). Set $\Delta'_n = \int_0^1 f(x)dx - \frac{1}{n}\sum_{j=1}^n f(j/n)$. Since

$$\int_0^1 f(x)dx = \sum_{j=1}^n \int_{(j-1)/n}^{j/n} f(x)dx = \sum_{j=1}^n \int_0^{1/n} f\left(\frac{j-1}{n} + x\right)dx,$$

we obtain

$$\begin{aligned}
|\Delta'_n| &= \int_0^{1/n} \sum_{j=1}^n \left| f(j/n) - f\left(\frac{j-1}{n} + x\right)\right| dx \\
&\leq \int_0^{1/n} \sum_{j=1}^n \left\{ \left| f(j/n) - f\left(\frac{j-1}{n} + x\right)\right| + \left| f\left(\frac{j-1}{n} + x\right) - f\left(\frac{j-1}{n}\right)\right| \right\} dx \\
&\leq \int_0^{1/n} V_0^1(f)dx = \frac{V_0^1(f)}{n}.
\end{aligned}$$

For $\Delta_n = \int_a^b h(t)dt - \frac{b-a}{n}\sum_{j=1}^n h(a + (b-a)j/n)$, we use the change of variables $\psi(x) = (b-a)x + a, x \in [0,1]$ and we obtain

$$\Delta_n = (b-a)\left\{ \int_0^1 (h\circ\psi)(x)dx - \frac{1}{n}\sum_{j=1}^n (h\circ\psi)(j/n)\right\}.$$

The previous results and Lemma 3.12.15 (ii) yield

$$|\Delta_n| \leq (b-a)V_0^1(h\circ\psi)/n = (b-a)V_a^b(h)/n.$$

The second statement of the Lemma follows trivially from this. □

Lemma 3.12.18.
$W^{(T)}(x)$ is 2π periodic, $\int_{-\pi}^{\pi} W^{(T)}(x)\mathrm{d}x = 1$. Furthermore, if $B_T < 1$, $\|W^{(T)}\|_\infty = \frac{1}{B_T}\|W\|_\infty$, and we have $V_{-\pi}^{\pi}(W^{(T)}) = \frac{1}{B_T}V_{-\pi}^{\pi}(W)$.

Proof. Let us first prove that $\int_{-\pi}^{\pi} W^{(T)}(x)dx = 1$. Since

$$\int_{-\pi}^{\pi} W^{(T)}(x)dx = \sum_{j\in\mathbb{Z}} \int_{-\pi}^{\pi} \frac{1}{B_T} W\left(\frac{x+2\pi j}{B_T}\right)dx,$$

the change of variables $y = (x + 2\pi j)/B_T$ yields

$$\int_{-\pi}^{\pi} W^{(T)}(x)dx = \sum_{j\in\mathbb{Z}} \int_{(2j-1)\pi/B_T}^{(2j+1)\pi/B_T} W(y)dy = \int_{\mathbb{R}} W(y)dy = 1.$$

If $B_T < 1$, then for $x \in [-\pi,\pi]$, $W^{(T)}(x) = \frac{1}{B_T}W(x/B_T)$. The third statement

follows directly because of the periodicity, and the last statement follows from

$$V_{-\pi}^{\pi}(W^{(T)}) = V_{-\pi/B_T}^{\pi/B_T}(W/B_T) = \frac{1}{B_T} V_{-\pi}^{\pi}(W).$$

We have used Lemma 3.12.15 and the fact that $W(x) = 0$ if $|x| \geq 1$. □

Using this lemma, we obtain

Lemma 3.12.19.
Provided $B_T \to 0$,

$$\frac{2\pi}{T} \sum_{s=1}^{T-1} W^{(T)}(\omega - 2\pi s/T) = 1 + O(B_T^{-1} T^{-1}),$$

and the error is uniform in ω.

Proof. Set

$$\begin{aligned}\Delta_n &= \int_{-\pi}^{\pi} W^{(T)}(\omega - \alpha) d\alpha - \frac{2\pi}{T} \sum_{s=1}^{T-1} W^{(T)}(\omega - 2\pi s/T), \\ &= \int_{-\pi}^{\pi} W'^{(T)}(\alpha) d\alpha - \frac{2\pi}{T} \sum_{s=1}^{T-1} W'^{(T)}(2\pi s/T),\end{aligned}$$

where $W'^{(T)}(\alpha) = W^{(T)}(\omega - \alpha)$. Lemmas 3.12.15, 3.12.17 and 3.12.18 yield

$$|\Delta_n| \leq \frac{2\pi}{T}(V_{-\pi}^{\pi}(W'^{(T)}) + \|W'^{(T)}\|_\infty) = \frac{2\pi}{B_T T}(V_{-\pi}^{\pi}(W) + \|W\|_\infty),$$

which does not depend on ω, and is of order $O(B_T^{-1} T^{-1})$. Since

$$\int_{-\pi}^{\pi} W^{(T)}(\omega - \alpha) d\alpha = 1,$$

we obtain

$$\frac{2\pi}{T} \sum_{s=1}^{T-1} W^{(T)}(\omega - 2\pi s/T) = 1 + O(B_T^{-1} T^{-1}),$$

and the error is uniform in ω. □

Lemma 3.12.20.
*Under condition **C(p,2)**, for each $k = 0, 1, \ldots, p$:*

$$\frac{\partial^k f_\omega}{\partial \omega^k} = \sum_{t \in \mathbb{Z}} (-\mathbf{i}t)^k e^{-\mathbf{i}\omega t} r_t,$$

and the convergence is L^2, uniformly ω. Moreover,

$$\sup_\omega \left\| \frac{\partial^k}{\partial \omega^k} f_\omega \right\|_2 < \infty \quad k = 1, 2, \ldots, p.$$

Proof. Since **C(p,2)** implies **C(k,2)** for $k = 0, 1, \ldots, p$, the result follows by an iterative application of Rudin (1976, Theorem 7.17) to the projection of the partial sums

$$\sum_{t=-N}^{N} e^{-\mathbf{i}\omega t} r_t.$$

Hence

$$\sup_{\tau,\sigma,\omega} \left\| \frac{\partial^k}{\partial \omega^k} f_\omega \right\|_2 \leq \sum_t (1 + |t|^k) \|r_t\|_2 < \infty. \qquad \square$$

The following Lemma is a straightforward extension of results on approximate identities (see Edwards 1967, §3.2) adapted to our framework:

Lemma 3.12.21 (Approximate identities)**.**
Suppose $K_T, T = 1, 2, \ldots$ is a sequence of functions defined on $[-\pi, \pi]$ satisfying, as $T \to \infty$:

(i) $\sup_T \int_{-\pi}^{\pi} |K_T(\alpha)| d\alpha < \infty$,

(ii) $\int_{-\pi}^{\pi} K_T(\alpha) d\alpha \longrightarrow 2\pi$,

(iii) for all $\delta > 0$, $\int_{\delta \leq |\alpha| \leq \pi} |K_T(\alpha)| d\alpha \to 0$.

Let $E \subset \mathbb{R}$ be an interval, let

$$g : [-\pi, \pi] \times E \to \mathbb{C}$$

be a function and, for each $e \in E$, define $g_e(\omega) = g(\omega, e)$. Let $g_\cdot(\omega)$ denote the function $e \mapsto g_e(\omega)$.
If the function $\omega \mapsto g_\cdot(\omega)$ is uniformly continuous with respect to $\|\cdot\|_p$, meaning that $\forall \varepsilon > 0, \exists \delta > 0$ such that

$$|\omega_1 - \omega_2| < \delta \implies \left\| g_\cdot(\omega_1) - g_\cdot(\omega_2) \right\|_p < \varepsilon, \tag{3.12.9}$$

and bounded with respect to $\|\cdot\|_p$,

$$\sup_\omega \left\| g_\cdot(\omega) \right\|_p < \infty,$$

then the convolution

$$K_T * g_e(\omega) = \int_{-\pi}^{\pi} K_T(\alpha) g_e(\omega - \alpha) d\alpha$$

converges in $\|\cdot\|_p$ to $g_e(\omega)$, uniformly in ω:

$$\sup_\omega \left\| K_T * g_\cdot(\omega) - g_\cdot(\omega) \right\|_p \to 0 \quad \text{as } T \to \infty.$$

Notice that if $p = \infty$, (3.12.9) *is the same as uniform equicontinuity of the family of functions $\{g_e\}_{e \in E}$.*

Proof. We follow the same strategy as in Edwards (1967, Theorem 3.2.2). We use the shorthand notation $\int = \int_{-\pi}^{\pi}$. Setting

$$a_T = \int K_T(\alpha) d\alpha,$$

we obtain via Jensen's inequality

$$\sup_{\omega} \left\| K_T * g_\cdot(\omega) - a_T g_\cdot(\omega) \right\|_p \leq \int |K_T(\alpha)| \cdot \sup_{\omega} \left\| g_\cdot(\omega - \alpha) - g_\cdot(\omega) \right\|_p d\alpha = I.$$

Fix $\varepsilon > 0$. We can choose $\delta > 0$ satisfying (3.12.9). The rest of the proof parallels that of Edwards (1967); the idea is the following: we separate the integral I into

$$\int_{0 \leq |\alpha| \leq \delta} + \int_{\delta \leq |\alpha| \leq \pi}.$$

The first integral will be small because of the continuity condition and the boundedness of the L^1-norm of K_T, and the second integral is small of the property (iii) of an approximate identity, and because the $g_\cdot$'s are bounded. The proof is then completed by noting that $a_T \to 2\pi$. □

Remark 3.12.22. *Notice that, in the previous Lemma, the rate of convergence depends only on*

1. *the properties of the approximate identity,*
2. *the bound* $\sup_\omega \left\| g_\cdot(\omega) \right\|_p$,
3. *the continuity parameter* $\delta = \delta(\varepsilon)$.

CHAPTER 4

Dynamics of DNA Minicircles

In this chapter, we study the dynamics of closed DNA strands, and propose a methodology for comparing the dynamics by comparing their spectral density operators, using some of the theory developed in Chapters 2 and 3. Since this chapter contains the applied contribution of this thesis, I have tried to write it in a self-contained manner, to facilitate independent reading.

4.1 Molecular Biophysics and Dynamics of DNA Minicircles

"What is life?" is an old and fundamental question. In trying to answer this question, Schrödinger (1944) developed some of the seminal ideas of molecular biophysics, which is the field that studies the physics of biopolymers, such as DNA strands and proteins (Noble 2010). Within this field, understanding the physics of DNA is of particular importance, because the mechanical properties of DNA are closely related to its biological functions, such as DNA packaging, replication, and transcription.
An ongoing theme of research in molecular biophysics is to understand how the mechanics of a strand of DNA—which are well described by models at the atomic level—behave at the larger scale of tens, hundreds (Walter, Gonzalez & Maddocks 2010, Gonzalez, Petkevičiūtė & Maddocks 2013) or even thousands of base-pairs (Sambriski et al. n.d.). A way of understanding the mechanics of such strands of DNA is through *cyclization* experiments. These experiments provide insight into the bending properties of DNA by estimating the *J-factor* (Amzallag et al. 2006) of a DNA strand, which is the probability that the two ends of the strand are in

a configuration where they could bind to form a loop. The resulting loop is called a *DNA minicircle* (Kahn & Crothers 1992, Shore et al. 1981, Shore & Baldwin 1983).

Recently, technological improvements have allowed static three-dimensional reconstruction of such minicircles, thus allowing the comparison of DNA minicircles with different J-factors. A specific example is that of the CAP minicircle and the TATA minicircle (see Section 4.2), whose J-factors are significantly different (Amzallag et al. 2006). Based on three-dimensional reconstructions of a sample of CAP & TATA minicircles, obtained by cryo-electron microscopy, it appeared that the differences between the two minicircles are *not* in their mean shape, but in the way they vary around their mean shape, i.e., in their covariance structure (Amzallag, Vaillant, Jacob, Unser, Bednar, Kahn, Dubochet, Stasiak & Maddocks 2006, Panaretos, Kraus & Maddocks 2010). These studies allowed however only insight into the static mechanics of DNA strands, and do *not* give information about the dynamical properties—or *dynamics*—of DNA, since they are based on static images of the DNA minicircles.

The ideal kind of data needed for understanding the dynamics of DNA would be in the form of a movie of DNA minicircles oscillating in solution. Unfortunately, empirical acquisition of such data is not yet feasible, but *in silico* surrogates can be created via Molecular Dynamics (MD) simulations (Leach 2001, Dryden et al. 2002, Gonzalez & Maddocks 2001, Lankas et al. 2009). MD simulations are used to obtain the trajectory of a DNA minicircle moving in solution, and are obtained by numerically solving a model based on the pairwise interactions between all the atoms of DNA and the water solution in which it is immersed. As such, it is not the trajectories of each individual DNA atom that are of interest, but their joint behaviour in the scaling-limit, where the mechanics of a DNA strand is similar to that of an elastic rod. It is therefore natural to adopt a functional data analysis (FDA, see Ramsay & Silverman 2005) viewpoint for the analysis of DNA minicircle trajectories.

The data we will be working with in this chapter are trajectories of CAP & TATA minicircles oscillating in solution, obtained by MD simulations. Given the scientific interest of understanding their large-scale dynamics, we will model them as *functional time series* (FTS). An FTS is a sequence $\{X_t : t = \ldots, -1, 0, 1, \ldots\}$, where t denotes the time index, and each X_t is a random function, say $X_t \in L^2\left([0,1], \mathbb{R}^3\right)$, representing for instance the shape of a minicircle at time t. The dynamics of the minicircles can be viewed through the lens of the second-order structure of the time series, which is contained in the lag-t autocovariance operators of the time series: these encode the covariation of the random function t time points apart. Understanding and comparing the dynamics of CAP and TATA minicircles can therefore be translated into the problem of estimation and inference for the second-order structure of functional time series.

Inference for functional time series is usually carried out under functional

autoregressive models, or more general linear models, see e.g. Mas (2002), Bosq (2000), Ferraty & Romain (2011); relaxing the linearity assumptions is the object of more recent work, see e.g. Hörmann & Kokoszka (2010). The problem of inference for functional time series without linear assumptions is starting to be addressed: Horváth, Kokoszka & Reeder (2013) treat the two-sample problem of testing the equality of the mean function of two functional time series. Concerning inference on the second-order structure, Horváth, Kokoszka & Reeder (2013) propose a consistent estimator for the long-run covariance operator. We proposed in Chapter 3 (see also Panaretos & Tavakoli 2013*b*) to estimate the entire second-order structure via a frequency domain approach, and showed in particular the asymptotic normality of our estimators, under cumulant mixing conditions. However, it seems that the problem of comparing the entire second-order structure of stationary functional time series has so far not been treated.

The contribution of this chapter comprises the development of a methodology for the comparison of the entire dynamics (encoded by all the lag-t autocovariance operators) of two stationary functional time series, by using a frequency domain approach. In particular, it gives a way of localizing the differences either only at the level of frequencies, or to select significant frequencies, and compare the dynamics of the curves within each frequency, while controlling the overall significance of the detected differences.

The chapter is organized as follows. In Section 4.2, we present the CAP and TATA minicircle data, the preprocessing steps, and the estimation of their dynamics through their spectral density operators. In Section 4.3, we treat the problem of detecting differences between the spectral density operators of CAP and TATA, by first comparing them at fixed frequencies—using a test for comparing their spectral density operators that we introduce in this chapter—and then adjusting for multiplicities to localize the differences in the frequencies, while controlling the overall significance of the detections. We also conduct a simulation study to assess the performance of our method for small sample sizes. The detection of the differences between the CAP and TATA is further investigated in Section 4.4, where we consider the problem of first selecting frequencies at which the spectral density operators of CAP and TATA are different, and then detecting and localizing their differences on the minicircles, within each selected frequency. We conclude this chapter by a brief outlook and some potential extensions (Section 4.5).

4.2 Description of the Data

The dataset of our case study is comprised of (time) trajectories of two DNA minicircles obtained via molecular dynamics (MD) simulations (Leach 2001, Chapter 7). Such simulations give the simulated trajectory of

a strand of DNA moving in water, and are obtained by numerical integration of a model taking into account all the pairwise interactions between atoms. The MD simulations were conducted by members of the *Laboratory for Computation and Visualization in Mathematics and Mechanics* at EPFL (http://lcvmwww.epfl.ch/), and kindly shared with us.

The two DNA minicircles are called CAP and TATA: they each consist of 158 base-pairs, and differ only at 18 base-pairs (see Table 4.1).

CAP

GATGAATTCACGGATCCGGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCCGTTTTTT
GCCCGTTTTTTCCGGATCCGTACAGGAATTCTAGACCTAGGGTGCCTAATGAGTGAGCTAACTCACA
TTAATTGCGTTGCGCCATGGAATC

TATA

GATGAATTCACGGATCCGGTTTTTTGCCCGTTTTTTGCCGTTTTTTGCCCGTTTTTTGCCGTTTTTT
GCCCGTTTTTTCCGGATCCGTACAGGAATTCTAGACCTAGGGTGCCTAATGAGTGCCCTTTTATAGC
TTAAACGCGTTGCGCCATGGAATC

Table 4.1 – The sequences of base-pairs for the CAP and TATA minicircles; the differences between the two sequences are in gray.

The MD simulation used an integration step of 2 femtoseconds ($2 \cdot 10^{-15}$ seconds) for the numerical integration algorithm, and the data were recorded every picosecond (10^{-12} seconds). Time-wise, the data consist of 50000 snapshots, where the data at each snapshot have been simplified to the three-dimensional coordinates of the 158 base-pairs centers of the minicircle.

For each minicircle, the data at hand are therefore

$$\{M_t(j) \mid t = 1, 2, \ldots, 50000;\, j = 1, \ldots, 158\} \subset \mathbb{R}^3,$$

where t denotes the time index (in picoseconds), and j is the base-pair index. In other words, $M_t(j) \in \mathbb{R}^3$ is the coordinates of base-pair j at time t. In the following, we shall view the set $\{1, \ldots, 158\}$ as the quotient group $\mathbb{Z}/158\mathbb{Z}$, so that $M_t(j+k)$ has a meaning for all $j, k \in \mathbb{Z}$.

The data are shown for various timepoints t in Figure 4.1 on the next page. Since the data can be rotated or translated without changing the information they convey, our analysis should hinge on features that are invariant to translations and rotations. We therefore choose to work with the curvature of the DNA minicircles, since curvature does not depend on the location or the orientation of the data. Further to solving the problem of data registration, the use of the curvature also reduces the dimensionality of the data, changing it from a time series of $\mathbb{R}^3$- valued curves to a time series of real-valued curves.

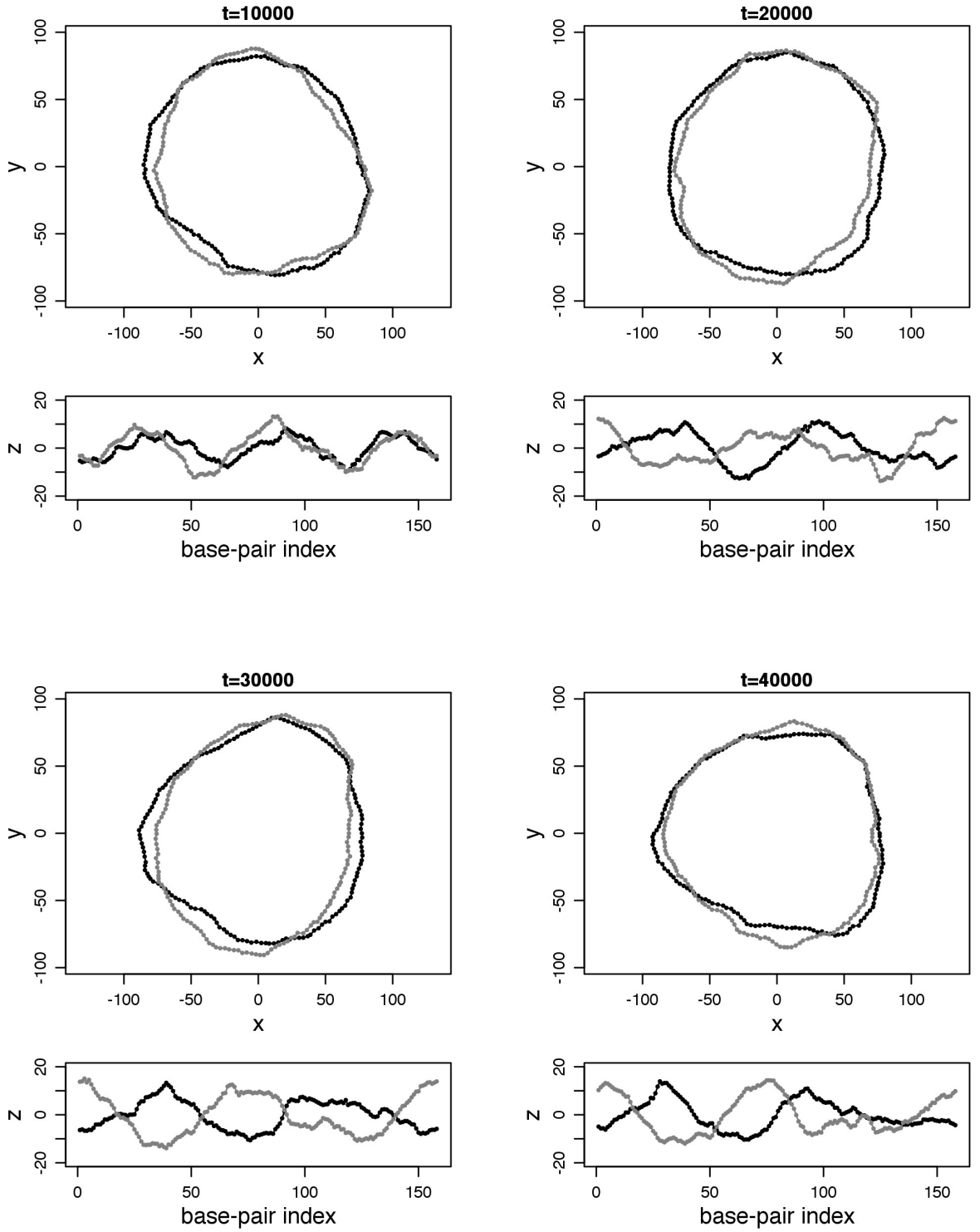


Figure 4.1 – The DNA minicircles for various timepoints t (CAP in black, TATA in gray). The top plot of the four subfigures contains the projection of the DNA minicircles onto the XY plane, and the plot below shows their projection onto the Z-axis. The units of the X, Y, Z axes are in ångström (1 ångström = 10^{-10} meters).

4.2.1 Data Preprocessing

The problem of estimating the curvature of a curve based on discrete noisy observations has already been studied (e.g. Lee et al. 1993, Sangalli et al. 2009). Sangalli et al. (2009) use free-knot regression splines to estimate the curve. Once the curve is estimated, its curvature is computed by a plug-in method, i.e., the curvature estimate is

$$c(t) = |\gamma'(t) \wedge \gamma''(t)| / |\gamma'(t)|^3, \tag{4.2.1}$$

where $\gamma(t) \in \mathbb{R}^3$ is the estimated curve, and $u \wedge v$ denotes the cross-product of two vectors $u, v \in \mathbb{R}^3$. These techniques use a smoothness parameter C that plays a crucial role in the estimation, and whose choice is rather subjective (Sangalli et al. 2009, Section 4). Furthermore, our personal experience with plug-in estimation of the curvature is that it yields estimates with too many degrees of freedom, mainly because of the renormalization factor in (4.2.1). The problem persists even with roughness penalization of the estimated curve. We therefore choose to use another method for estimating the curvature of our minicircles, that also fits better our interest of understanding the larger scale behaviour of DNA minicircles. For each minicircle trajectory, we compute the curvature trajectory

by $\mathbb{R}_+$ we mean the set of positive real numbers

$$\{c_t(j) : t = 1, 2, \ldots, 50000; j = 1, \ldots, 158\} \subset \mathbb{R}_+,$$

where $c_t(j)$ is the curvature (inverse radius) of the circle passing through the three points $M_t(j-5), M_t(j), M_t(j+5)$. We recall that the curvature of three points $p_1, p_2, p_3 \in \mathbb{R}^3$ is given by

$$\text{curvature}(p_1, p_2, p_3) = 2\big\|(p_2 - p_1) \wedge (p_3 - p_2)\big\| / \big(\big\|p_2 - p_1\big\| \cdot \big\|p_3 - p_2\big\| \cdot \big\|p_3 - p_1\big\|\big).$$

The reason why the curvature is taken between the base-pairs $\{j-5, j, j+5\}$ instead of $\{j-1, j, j+1\}$, is that since the DNA double helix performs on average a complete rotation in about 11 or 12 base-pairs; taking the curvature of the directly adjacent base-pairs would represent a very local curvature of the minicircles, whereas the curvature computed on base-pairs $\{j-5, j, j+5\}$ corresponds to a coarser version of the curvature, which fits our interest of understanding the larger scale behaviour of the minicircles. Furthermore, computing curvatures in this way yields more stable estimates (which are less sensitive to small perturbations of the base-pair center). From a statistical point of view, this procedure estimates a biased (smoothed) version of the curvature, which discards very local bends of the DNA, but keeps its larger scale bends.

Since the curvature is constrained to be positive, the curvature data $c_j(t)$ do not lie in a linear space. Nevertheless, most methodology for functional data—including ours—assumes implicitly that the data take values in a linear space (e.g. functional principal component analysis consists of finite dimensional *linear* approximations of the data). We therefore

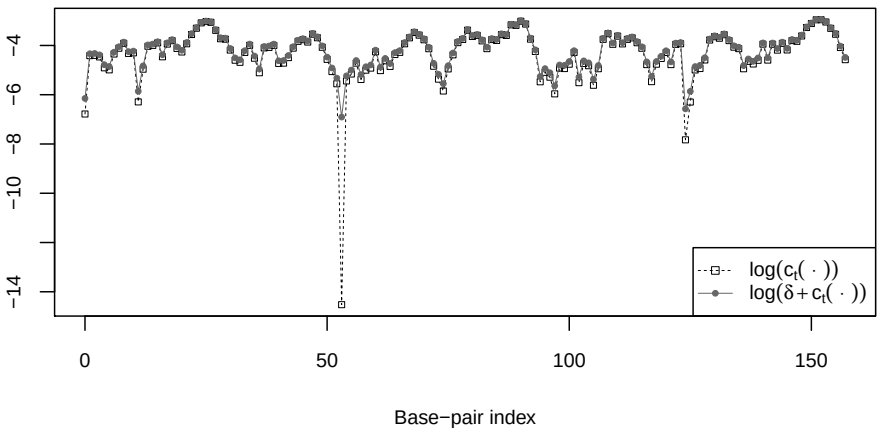


Figure 4.2 – The effect of the constant $\delta = 10^{-3}$, see (4.2.2), in the linearization of the curvature. The dashed curve is $\log(c_t(\cdot))$, and the gray solid curve is $d_t(\cdot) = \log(\delta + c_t(\cdot))$. Notice that the d_t smooths the downward peaks that are very deep, while changing the other points only a little.

convert the curvature data into elements of a linear space by using the transformation $x \mapsto \log(\delta + x)$, where $\delta > 0$ is a fixed constant (see below), and define the *linearized curvature* by

$$d_t(j) = \log(\delta + c_t(j)), \quad \text{for all } t, j. \tag{4.2.2}$$

The purpose of the constant δ is to prevent $d_j(t)$ from taking too large (negative) values if $c_j(t)$ is close to zero, which would ruin any further analysis. If δ is too small, the functions $d_j(\cdot)$ will have very large spikes, and if δ is too large, $d_j(\cdot)$ will be essentially constant. We choose $\delta = 10^{-3}$, based on exploratory analysis, which gave a good compromise between the two situations. Figure 4.2 illustrates the role of δ. Since the linearized curvatures are discretely sampled versions of smooth curves, we transform each function $j \mapsto d_t(j)$ into a smooth curve $\tau \mapsto Y_t(\tau), \tau \in [0,1]$, by smoothing the scatter plot

$$\left(\frac{j-1}{158}, d_t(j)\right), \quad j = 1, \ldots, 158, \tag{4.2.3}$$

for each fixed t, using a basis expansion (Ramsay & Silverman 2005) with 80 *periodic* cubic B-splines (King et al. 2010). We used periodic B-splines instead of the usual B-splines because we expect the smooth curve to be periodic on $[0,1]$, since it represents the curvature of a *closed* strand of DNA. The choice of the number of basis elements represents a bias-variance trade-off, and is usually either based on prior knowledge of the smoothness of the functional data, or some kind of cross-validation method. Our choice of 80 basis functions came from the combination of considerations on the postulated degrees of freedom of the curvature (which should be less than the number of base-pairs), computational considerations, and graphical goodness of fit assessment. We also conducted the analysis presented in the rest of the chapter with 40 and 60 basis functions, and the results were similar to those obtained with 80

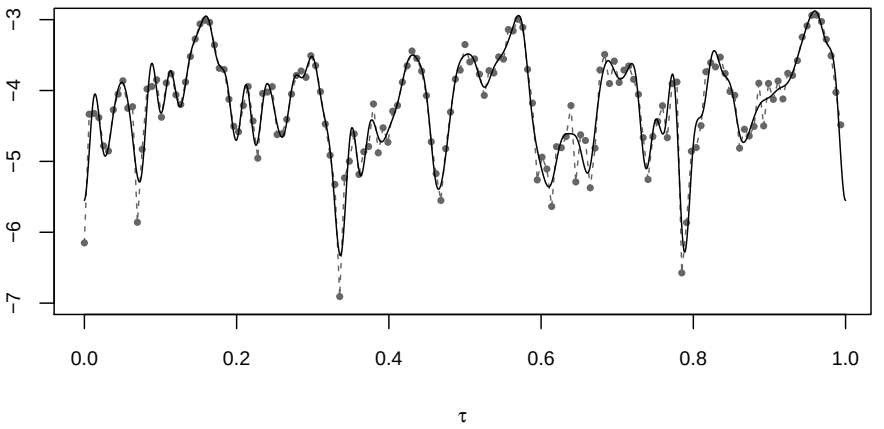


Figure 4.3 – Illustration of the smoothing process. The dashed curve with the solid grey dots represents the scatter plot (4.2.3), and the solid black curve represents its smoothed version obtained by projecting it onto a basis of 80 periodic cubic B-splines.

basis functions. We note that exploratory plots revealed no need to use penalization for smoothing the functions d_t. Figure 4.3 illustrates the smoothing process.

Exploratory analysis of the functional time series $\{Y_t : t = 1,\dots,50000\}$ revealed that the series exhibited a strong temporal dependence, together with a non-stationary behaviour. Taking the time differences $X_t = Y_{t+1} - Y_t$ circumvents this problem. We therefore chose to work with the series $\{X_t : t = 1,\dots,49999\}$ for the rest of the analysis. To put things in perspective, the model we are implicitly assuming is

$$Y_{t+1}(\tau) = Y_t(\tau) + X_t(\tau), \quad \tau \in [0,1],$$

where we recall that Y_t is the linearized curvature of the DNA minicircle. The stationary series X_t is therefore the process that governs the change in linearized curvature of the DNA minicircle. Applying all these steps to the CAP minicircles, and respectively to the TATA minicircles, we get two functional time series, X_t^1, respectively X_t^2. Figure 4.4 on the next page contains plots of X_t^a $(a = 1,2)$ for different values of t.

4.2.2 Estimation of the Dynamics

The (second-order) dynamics of a stationary FTS $\{X_t : t \in \mathbb{Z}\}$ are contained in the lag-t autocovariance operators

$$\mathscr{R}_t = \mathbb{E}\big[(X_t - \mu) \otimes_2 (X_0 - \mu)\big], \quad t \in \mathbb{Z}, \tag{4.2.4}$$

where $\mu = \mathbb{E}\, X_t$. Estimation of the dynamics could therefore be reduced to the estimation of the autocovariance operators, but we choose to take a different approach, and estimate the dynamics through a frequency domain approach (see Chapters 2 and 3, as well as Panaretos & Tavakoli (2013*b*)). Recall that in the frequency domain approach, the objects of interest are not the lag-t autocovariance operators, but their Fourier trans-

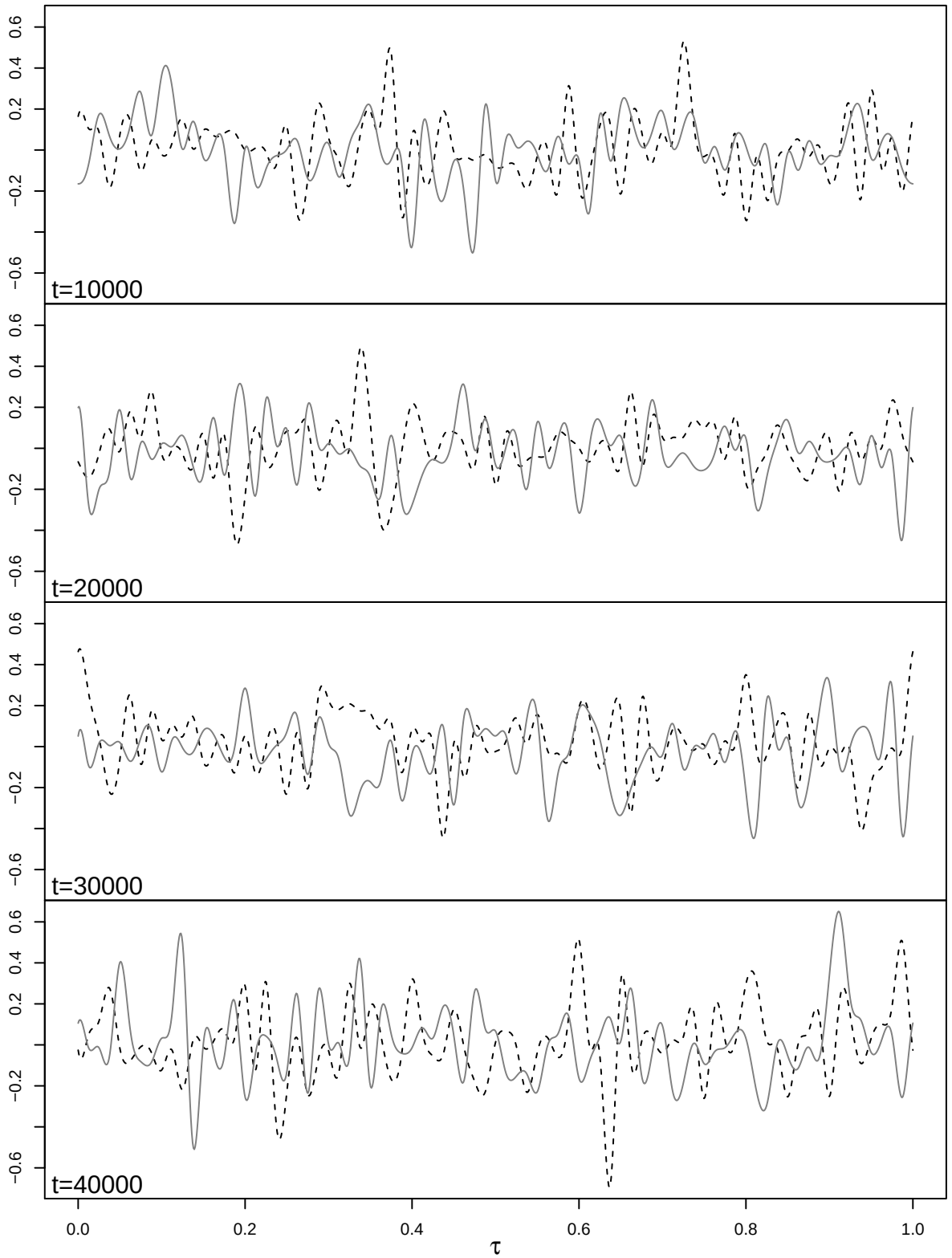


Figure 4.4 – Plot of the innovation process X_t of the linearized curvatures for CAP (dashed black curve) and TATA (solid gray curve) for various timepoints t.

forms,

$$\mathscr{F}_\omega = (2\pi)^{-1} \sum_{t\in\mathbb{Z}} \exp(-\mathbf{i}\omega t)\mathscr{R}_t, \qquad \omega \in [-\pi,\pi], \tag{4.2.5}$$

where $\mathbf{i} \in \mathbb{C}$ is the imaginary number, $\mathbf{i}^2 = -1$. Each $\mathscr{F}_\omega, \omega \in [-\pi,\pi]$, is called the spectral density operator at ω, and is well defined under a summability conditions on the autocovariance operators (see Proposition 3.2.4). At each frequency ω, the spectral density operator $\mathscr{F}_\omega$ is an operator on $L^2([0,1],\mathbb{C})$, associated with a spectral density kernel f_ω, a complex valued surface $[0,1]^2 \ni (\tau,\sigma) \mapsto f_\omega(\tau,\sigma) \in \mathbb{C}$. The spectral density operator has various properties, but we just mention that $f_{-\omega} = \overline{f_\omega}$, and we therefore restrict our interest to $\omega \in [0,\pi]$. The elements of the collection $\{\mathscr{F}_\omega : \omega \in [0,\pi]\}$ are called the *spectral density operators*. Intuitively, the spectral density operators generalize of the spectral density matrices of multivariate time series (Brillinger 2001, Priestley 2001) to the functional setting, and yield a decomposition of the variance of the FTS, given by the inversion formula

$$\mathscr{R}_t = \int_{-\pi}^{\pi} \exp(\mathbf{i}\alpha t)\mathscr{F}_\alpha d\alpha, \quad t \in \mathbb{Z}.$$

The reason we decide to estimate the dynamics of our FTS via a frequency domain approach is because the spectral density operators are closely related to optimal linear finite dimensional representations of an FTS, via harmonic principal component analysis (see Section 2.8).

Estimation of the spectral density operators is done by computing the discrete Fourier transforms of the FTS, taking their empirical covariance (called the periodogram operator), and smoothing it with a kernel of bandwidth B_T (see Panaretos & Tavakoli (2013*b*) for details). The first step can be done using the Fast Fourier Transform, whose calculation is most efficient when the length of the series is highly composite. We therefore use only the stretch of data for $t = 1,\ldots,49152$ for all further computations.

The bandwidth parameter B_T needs to satisfy the conditions $B_T \to 0$ and $TB_T \to \infty$ as $T \to \infty$, for the asymptotic results to hold. The choice of B_T governs also the bias/variance trade-off for the estimation of the spectral density operators, similarly to nonparametric regression. For inference (relying on Theorem 4.3.1 and Panaretos & Tavakoli (2013*b*, Theorem 3.7)), it is crucial in finite samples to notice that the central limit theorem effect occurs because the spectral density estimator at a given frequency is obtained by weighted averaging of $m = TB_T/2\pi$ approximately independent summands. The order of m plays therefore a role similar to the number of i.i.d. summands when applying the classical central limit theorem, and should be taken into account before making any inferential statements based on asymptotics. In our case, we use leave-one-out cross-validation of the trace of the periodogram operator as a guide to bandwidth choice (Lee 1997). This suggests taking $B_T = 0.077$

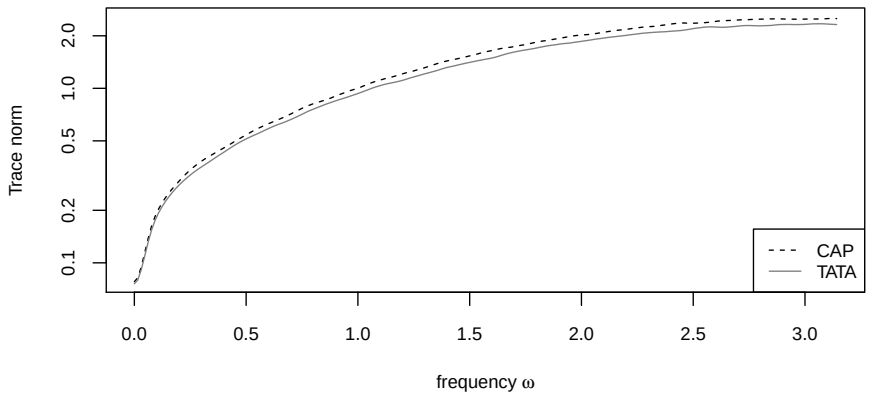


Figure 4.5 – Plot of the trace of the estimated spectral density operators. Notice that the sample spectral density operators of CAP have consistently a larger magnitude than the sample spectral density operators of TATA.

($m = 600$), a choice that is compatible with the results of our simulation studies, in which we set $B_T = T^{-1/5}$ (following the heuristic that $B_T \sim O(T^{-1/5})$ asymptotically, to minimize the mean-square error (Brillinger 2001, p.251)). We denote by $\mathscr{F}_\omega^{a,(T)}$ the estimated spectral density operator, also called the *sample spectral density operator*, of X_t^a, for $a = 1,2$.

Graphical representation of the sample spectral density operators is not straightforward: for each frequency $\omega \in [0,\pi]$, the sample spectral density operator is an operator on $L^2([0,1],\mathbb{C})$: its trace norm is shown Figure 4.5 for CAP and TATA. The modulus of the sample spectral density kernels $f_\omega^{a,(T)}$, $a = 1,2$, associated to the sample spectral density operators, are represented in Figure 4.6.

We notice that most of the trace energy is concentrated towards the higher end of the frequencies (near $\omega = \pi$), with low energy near $\omega = 0$. This means that the series X_t consists mainly of high frequency functional oscillations, and that the low frequency oscillations contained in the linearized curvature series Y_t mostly cancel out when taking its time differences to form X_t. Figure 4.6 shows that most of the energy of the spectral density operators is near the diagonal of the sample spectral density kernels, and then falls off sharply as one moves away from the diagonal. The interpretation of this is that the series X_t has strong local interactions, in the sense that $X_t(\tau)$ and $X_0(\sigma)$ are interacting strongly for $|\tau - \sigma|$ small, say $|\tau - \sigma| < \varepsilon$, and almost not interacting for $|\tau - \sigma| > \varepsilon$. This reflects the fact that the series X_t is locally smooth, but globally quite rough, as can be seen in Figure 4.4 on page 161 .

4.2.3 Comparing the Spectral Density Operators?

Even though the traces of the sample spectral density operators of CAP and TATA are different, and small differences between their sample spectral density kernels are visible, it is not a priori clear if these differences are due purely to randomness, or if the spectral density operators of the two DNA minicircles are different. The next section answers this question, by

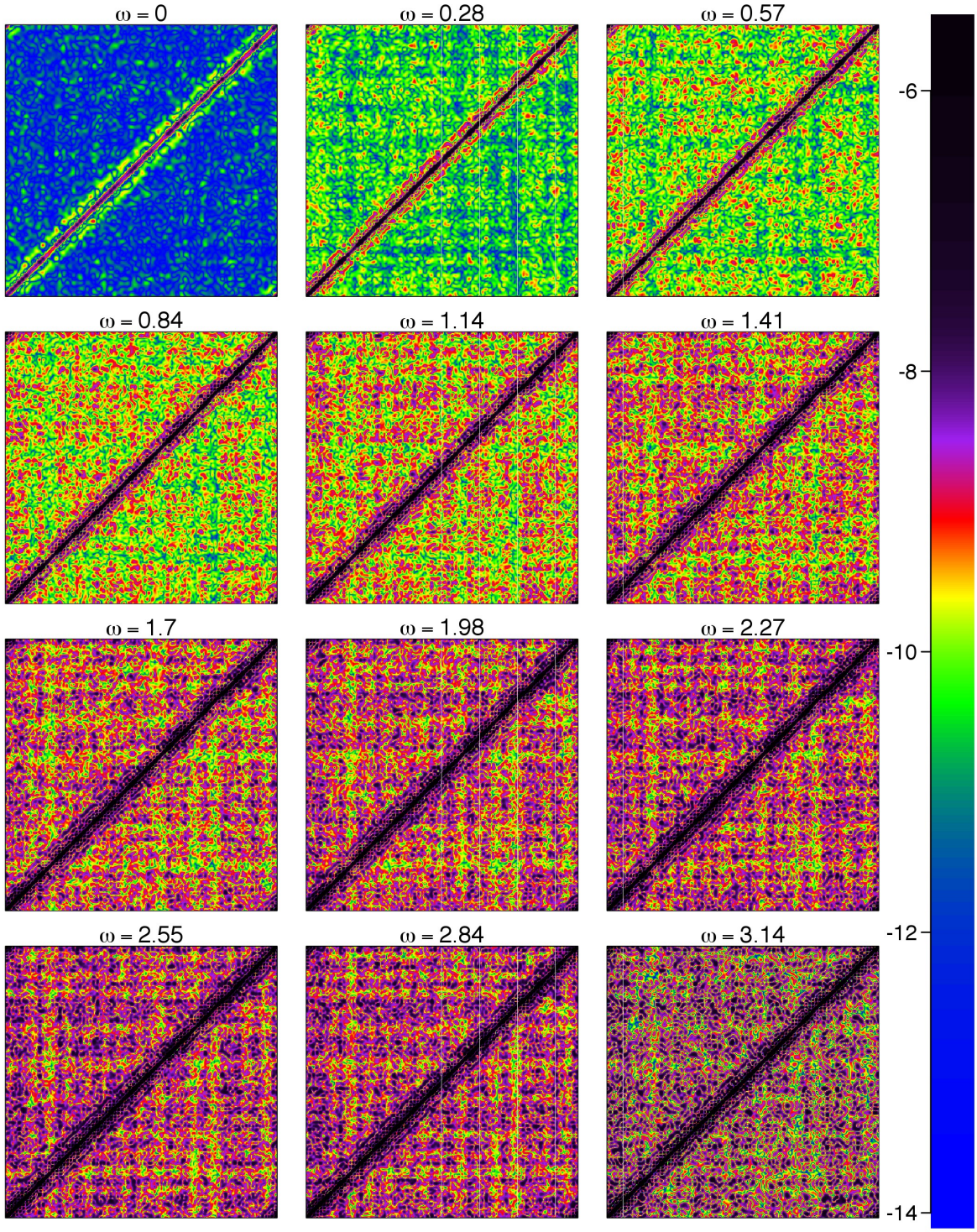


Figure 4.6 – Plot of the sample spectral density kernels for 12 equispaced frequencies on $[0, \pi]$. For each ω, the modulus of the sample spectral density kernel of the minicircles is plotted: the upper-left part of each square represents the modulus of the sample spectral density kernel of CAP, and the lower-right part represents the corresponding quantity for TATA. Lack of symmetry between the upper-left and lower-right is a sign of differences in the sample spectral density operators of CAP and TATA.

comparing the two spectral density operators on a dense grid of frequencies, and localizing frequencies at which the spectral density operator are significantly different, while adjusting for multiplicities.

4.3 Comparing the Spectral Density Operators, Localizing the Differences in Frequencies

Comparing the second order dynamics of the functional time series X_t^1 and X_t^2 can now be reduced to testing the equality of their spectral density operators. More precisely, we wish to test

$$\bigcap_{\omega\in[0,\pi]} H_\omega \quad \text{against} \quad \text{“}H_\omega \text{ is wrong for some } \omega\in[0,\pi]\text{”},$$

where H_ω :“ $\mathscr{F}_\omega^1 = \mathscr{F}_\omega^2$ ”, for $\omega \in [0,\pi]$, and detect the frequencies ω for which the null hypothesis H_ω is wrong, while controlling an overall error criterion. We will take a multiple testing approach to this problem, first by testing each H_ω marginally, and then performing multiplicity corrections.

4.3.1 Comparing the Spectral Density Operators at a Fixed Frequency

In order to test H_ω for a fixed $\omega \in [0,\pi]$, we created a test inspired by Panaretos et al. (2010). The key idea is to project the difference between the two sample spectral density operators, $\mathscr{F}_\omega^{1,(T)} - \mathscr{F}_\omega^{2,(T)}$, onto the (random) subspaces generated by the tensor products of the first K estimated eigenfunctions of $\mathscr{F}_\omega^1$, under H_ω. Let $(\tilde{\mu}_i(\omega), \tilde{\varphi}_i^\omega)_{i=1}^\infty$ be the eigenvalue and eigenfunction pairs of the pooled spectral density operators $(\mathscr{F}_\omega^{1,(T)} + \mathscr{F}_\omega^{2,(T)})/2$. We propose the following test statistic,

$$\tilde{\Delta}_K^{(T)}(\omega) = \sum_{i,j=1}^{K} \frac{\left|\left\langle D_\omega^{(T)}\tilde{\varphi}_j^\omega, \tilde{\varphi}_i^\omega \right\rangle\right|^2}{\left(1+\mathbf{1}_{\{0,\pi\}}(\omega)\right) 4\pi\kappa^2 \tilde{\mu}_i(\omega)\tilde{\mu}_j^\omega}, \tag{4.3.1}$$

where $\kappa^2 = \int_{\mathbb{R}} W(x)^2 dx$, and $D_\omega^{(T)} = \sqrt{TB_T}\left(\mathscr{F}_\omega^{1,(T)} - \mathscr{F}_\omega^{2,(T)}\right)$ is the rescaled difference between the sample spectral density operators of CAP and TATA. The indicator in the denominator is a correction term for the frequencies $\omega \in \{0,\pi\}$, at which the sample spectral density operator has an increased variance. The test measures the rescaled differences between the two sample spectral density operators at frequency ω that is contained in the space spanned by the first K sample eigenfunctions of the pooled spectral density operators. In a multivariate setting, one would want take the truncation level K equal to the dimension of the vector series; this is however not possible here, because it leads to an inverse problem since $\tilde{\mu}_i^\omega \to 0$ as $i \to \infty$. The parameter K can be therefore viewed as a regularization parameter. Its choice is discussed in Section 4.3.2.

The following Theorem gives the asymptotic distribution of our test statistic, and is the main theoretical result of this chapter.

Theorem 4.3.1.
Let $K_1, \dots, K_J$ be fixed non-negative integers, and let $\omega_1, \dots, \omega_J \in [0, \pi]$ be a fixed number of distinct frequencies. Assume that the conditions of Theorem 3.6.5 hold, and that $B_T \to 0$ & $TB_T \to \infty$ as $T \to \infty$. Furthermore, assume that for each ω_j, the first K_j eigenvalues of the spectral density operator $\mathscr{F}_{\omega_j}$ are all distinct, and strictly positive. Then, under the null hypothesis $H_0 = \cap_{j=1}^{J} H_{\omega_j}$, the test statistics $\tilde{\Delta}_{K_j}^{(T)}(\omega_j)$, $j = 1, \dots, J$, converge in distribution to independent *random variables $\Delta_{K_j}(\omega_j)$, where*

$$\Delta_{K_j}(\omega_j) \sim \begin{cases} \chi^2_{K(K+1)/2}, & \omega_j \in \{0, \pi\}, \\ \chi^2_{K^2}, & \textit{otherwise.} \end{cases} \tag{4.3.2}$$

Proof. The proof is in two parts. First we will show the result for a modified version of the test, where we assume that the eigenfunctions and eigenvalues of the spectral density operators are known. Then we will show that the sample versions of the eigenfunctions and eigenvalues are consistent, and we will conclude the proof using Slutsky's theorem for metric spaces.
We recall that under H_ω, $\mathscr{F}_\omega^1 = \mathscr{F}_\omega^2 = \mathscr{F}_\omega$, whose singular value decomposition is given by

$$\mathscr{F}_\omega = \sum_{j \geq 1} \mu_j(\omega) \varphi_j^\omega \otimes_2 \varphi_j^\omega.$$

We also denote by

$$\mathscr{F}_\omega^{(T)} = (\mathscr{F}_\omega^{1,(T)} + \mathscr{F}_\omega^{2,(T)})/2$$

the pooled sample spectral density operator, with singular value decomposition

$$\mathscr{F}_\omega^{(T)} = \sum_{j \geq 1} \tilde{\mu}_j(\omega) \tilde{\varphi}_j^\omega \otimes_2 \tilde{\varphi}_j^\omega.$$

Let us also write $\varphi_{ij}^\omega = \varphi_i^\omega \otimes_2 \varphi_j^\omega$, and similarly for $\tilde{\varphi}_{ij}^\omega$. In the following, we shall omit the superscripts "$\cdot^\omega$" and some of the "(ω)" in order to alleviate notation; this will not interfere with the validity of the proof, since ω will be fixed.
Fix $\omega \in [0, \pi]$, and let us define

$$\check{\Delta}_K^{(T)}(\omega) = \sum_{i,j=1}^{K} \frac{\left| \left\langle D_\omega^{(T)} \varphi_j, \varphi_i \right\rangle \right|^2}{\left(1 + \mathbf{1}_{\{0,\pi\}}(\omega)\right) 4\pi\kappa^2 \mu_i \mu_j}, \tag{4.3.3}$$

where $D_\omega^{(T)} = \sqrt{TB_T}\left(\mathscr{F}_\omega^{1,(T)} - \mathscr{F}_\omega^{2,(T)}\right)$. $\check{\Delta}_K^{(T)}(\omega)$ is similar to $\hat{\Delta}_K^{(T)}(\omega)$ with the eigenstructure of the pooled sample spectral density operator replaced by its true (and unknown) counterpart. By Theorem 3.6.5, we know that $D_\omega^{(T)}$ converges in distribution to a random element $\check{\mathscr{F}}_\omega$, whose Karhunen–

Loève expansion is given by $\check{\mathscr{F}}_\omega = \sum_{i,j=1}^{\infty} \eta_{ij}\varphi_{ij}$, see Remarks 3.7.6 and 3.7.9. In particular, $\{\eta_{ij} : 1 \le i \le j\}$ are independent Gaussian random variables with mean zero. Hence, by the continuous mapping theorem,

$$\check{\Delta}_K^{(T)}(\omega) \xrightarrow{d} \sum_{i,j=1}^{K} \frac{|\eta_{ij}|^2}{\left(1 + \mathbf{1}_{\{0,\pi\}}(\omega)\right) 4\pi\kappa^2 \mu_i \mu_j} =: \Delta_K(\omega) \tag{4.3.4}$$

We now need to distinguish two cases: if $\omega \in \{0,\pi\}$, then $\eta_{ji} = \eta_{ij}$ for $i < j$,

$$\left\{\eta_{ij} : j \ge i \ge 1\right\}$$

are *real* independent Gaussian random variables with mean zero,

$$\operatorname{var}\left(\eta_{ii}\right) = 8\pi\kappa^2\mu_i^2,$$

and

$$\operatorname{var}\left(\eta_{ij}\right) = 4\pi\kappa^2\mu_i\mu_j,\ i < j.$$

A direct calculation yields thus $\Delta_K(\omega) \sim \chi^2_{K(K+1)/2}$, since

$$\left(|\eta_{ij}|^2 + |\eta_{ji}|^2\right)/(8\pi\kappa^2\mu_i\mu_j) = |\eta_{ij}|^2/(4\pi\kappa^2\mu_i\mu_j) \sim \chi^2_1, \quad i < j,$$

and

$$|\eta_{ii}|^2/(8\pi\kappa^2\mu_i^2) \sim \chi^2_1.$$

If $\omega \notin \{0,\pi\}$, the random variables η_{ij} are real Gaussian variables for $i = j$, with and circular complex Gaussian for $i < j$ (see Remark 3.7.6), with $\eta_{ji} = \overline{\eta_{ij}}$ for $i < j$ and

$$\operatorname{var}\left(\eta_{ij}\right) = 4\pi\kappa^2\mu_i\mu_j, \quad \forall\, i \le j.$$

Hence, $\Delta_K(\omega) \sim \chi^2_{K^2}$, since

$$|\eta_{ii}|^2/(4\pi\kappa^2\mu_i^2) \sim \chi^2_1,$$

and

$$\left(|\eta_{ij}|^2 + |\eta_{ji}|^2\right)/(4\pi\kappa^2\mu_i\mu_j) = 2|\eta_{ij}|^2/(4\pi\kappa^2\mu_i\mu_j) \sim \chi^2_2,$$

for $i < j$.

Let us now turn to the second part of the proof. Fix $\omega \in [0,\pi]$. From Proposition 3.7.2, we know that for all $i = 1,2,\ldots,K$, $\tilde{\mu}_i$ converges in L^2 to μ_i, and $\tilde{\mu}_i$ is therefore a consistent estimator of μ_i. We now turn to the eigenfunctions φ_i. We point out that these eigenfunctions are not uniquely defined, however the eigenprojectors $\Pi_i = \varphi_i \otimes_2 \varphi_i$ are well defined. Using results from Section A.3.4, we get

$$|\langle D_\omega^{(T)}\tilde{\varphi}_j, \tilde{\varphi}_i\rangle|^2 = \langle D_\omega^{(T)} \bigotimes_2 D_\omega^{(T)}, \tilde{\Pi}_i \,\widetilde{\bigotimes}_2\, \tilde{\Pi}_j\rangle_{\mathscr{S}_2},$$

Since $\tilde{\Pi}_i$ is a consistent estimator of Π_i, $i = 1, \ldots, K$ (Proposition 3.7.2), and the Kronecker and tensor products are continuous, the continuous mapping Theorem implies that

$$\tilde{\Pi}_i \widetilde{\bigotimes}_2 \tilde{\Pi}_j \xrightarrow{p} \Pi_i \widetilde{\bigotimes}_2 \Pi_j,$$

and

$$D_\omega^{(T)} \bigotimes_2 D_\omega^{(T)} \xrightarrow{d} \check{\mathscr{F}}_\omega \bigotimes_2 \check{\mathscr{F}}_\omega.$$

Therefore, by Slutsky's Theorem,

$$\begin{aligned}
\tilde{\Delta}_K^{(T)} &\xrightarrow{d} \sum_{i,j=1}^{K} \frac{\left\langle \check{\mathscr{F}}_\omega \otimes_2 \check{\mathscr{F}}_\omega, \Pi_i \otimes_2 \Pi_j \right\rangle_{\mathscr{S}_2}}{\left(1 + \mathbf{1}_{\{0,\pi\}}(\omega)\right) 4\pi\kappa^2 \mu_i \mu_j} \\
&= \sum_{i,j=1}^{K} \frac{\left| \left\langle \check{\mathscr{F}}_\omega, \varphi_{ij} \right\rangle_{\mathscr{S}_2} \right|^2}{\left(1 + \mathbf{1}_{\{0,\pi\}}(\omega)\right) 4\pi\kappa^2 \mu_i \mu_j} \\
&= \sum_{i,j=1}^{K} \frac{\left|\eta_{ij}\right|^2}{\left(1 + \mathbf{1}_{\{0,\pi\}}(\omega)\right) 4\pi\kappa^2 \mu_i \mu_j} \\
&= \Delta_K(\omega).
\end{aligned}$$

$\langle \cdot, \cdot \rangle_{\mathscr{S}_2}$ denotes the Hilbert-Schmidt inner product, see Section A.2.2.2 on page 220

To finish the proof, notice that the independence of the $\Delta_{K_j}(\omega_j)$s follows directly from the independence of the $\check{\mathscr{F}}_{\omega_j}$s. □

The application of this test requires the choice, for each $\omega \in \Gamma$, of a parameter K, which is now discussed.

4.3.2 *Automatic Choice of the Truncation Level*

The choice of the truncation level K is a difficult problem. If it is chosen too small, then the test will respect the level, but will not be powerful. With a K too large, the test will not respect the level due to the ill-posedness of the problem.

Ideally, the choice of truncation level K should depend on the frequency ω_j, i.e., $K = K(\omega_j)$. One way of choosing $K(\omega)$ is to find the $K \geq 1$ that minimizes the following pseudo-AIC criterion

$$\mathrm{AIC}(K, \omega) = \mathrm{GOF}(K, \omega) + \mathrm{PEN}_1(K, \omega) + \mathrm{PEN}_2(K, \omega) \tag{4.3.5}$$

where $\mathrm{GOF}(K, \omega)$ is a goodness of fit criterion, and $\mathrm{PEN}_a(K, \omega)$, $a = 1, 2$, are penalizations for overfitting the spectral densities $\mathscr{F}_\omega^{a,(T)}$, $a = 1, 2$. We propose taking

$$\mathrm{GOF}(K, \omega) = \sum_{k=K+1}^{N_b} \left\langle \left(\mathscr{F}_\omega^{1,(T)} - \mathscr{F}_\omega^{2,(T)} \right) \tilde{\varphi}_k^\omega, \tilde{\varphi}_k^\omega \right\rangle \tag{4.3.6}$$

and

$$\mathrm{PEN}_a(K,\omega) = \left(\sum_{j=1}^{N_b} \tilde{\mu}_j(\omega)\right) \sum_{j=1}^{N_b} \frac{\left\langle \mathscr{F}_\omega^{a,(T)}(K)\,\hat{\varphi}_j^{a,\omega}, \hat{\varphi}_j^{a,\omega} \right\rangle}{n(\omega,m)\hat{\mu}_j^a(\omega)}, \quad a = 1,2, \quad (4.3.7)$$

where $(\hat{\mu}_j^a(\omega), \hat{\varphi}_j^{a,\omega})$ denotes the j-th eigenvalue/eigenvector pair of $\mathscr{F}_\omega^{a,(T)}$, $a = 1,2; j = 1,2,\ldots$, $(\tilde{\mu}_j(\omega), \tilde{\varphi}_j^{\omega})$ denotes the j-th eigenvalue/eigenvector pair of the pooled sample spectral density operator $\mathscr{F}_\omega^{(T)} = \left(\mathscr{F}_\omega^{(T)} + \mathscr{F}_\omega^{(T)}\right)/2$, and

$$\mathscr{F}_\omega^{a,(T)}(K) = \sum_{k_1,k_2=1}^{K} \left\langle \mathscr{F}_\omega^{a,(T)} \tilde{\varphi}_{k_1}^{\omega}, \tilde{\varphi}_{k_2}^{\omega} \right\rangle \tilde{\varphi}_{k_1}^{\omega} \otimes_2 \tilde{\varphi}_{k_2}^{\omega}, \qquad a = 1,2,$$

is the projection of the sample spectral density operator onto the first K eigenspaces of the pooled sample spectral density operator $(\mathscr{F}_\omega^{1,(T)} + \mathscr{F}_\omega^{2,(T)})/2$. The constant $n(\omega,m)$ depends only on ω and $m = TB_T/2\pi$, see (4.3.8) below. The intuition behind this criterion is that it corresponds to the AIC criterion of Panaretos et al. (2010, Section 3.3) if we had observed $n(\omega)$ i.i.d. complex curves from a random function with covariance $\mathscr{F}_\omega^a$, for $a = 1,2$. Even though these curves are not observed in our context, the choice of $n(\omega)$ should reflect the number of independent pieces of information used to construct our estimate $\mathscr{F}_\omega^{a,(T)}$. Following Brillinger (2001, p.252), we set

$$n(\omega,m) = m/\kappa^2. \quad (4.3.8)$$

We also propose a variant of this criterion, by using the following penalizations instead of (4.3.7),

$$\mathrm{PEN}_a^*(K,\omega) = \left(\sum_{j=1}^{N_b} \tilde{\mu}_j(\omega)\right) \sum_{j=1}^{N_b} \frac{\left\langle \mathscr{F}_\omega^{a,(T)}(K)\,\hat{\varphi}_j^{a,\omega}, \hat{\varphi}_j^{a,\omega} \right\rangle}{n(\omega,m)\sqrt{\hat{\mu}_j^a(\omega)\hat{\gamma}_j^a(\omega)}}, \quad a = 1,2, \quad (4.3.9)$$

where $\hat{\gamma}_1^a(\omega) = \hat{\mu}_1^a(\omega) - \hat{\mu}_2^a(\omega)$ and

$$\hat{\gamma}_l^a(\omega) = \min\left\{\hat{\mu}_{l-1}^a(\omega) - \hat{\mu}_l^a(\omega), \hat{\mu}_l^a(\omega) - \hat{\mu}_{l+1}^a(\omega)\right\}, \quad l = 2,\ldots,$$

and $a = 1,2$. The corresponding pseudo-AIC criterion is

$$\mathrm{AIC}^*(K,\omega) = \mathrm{GOF}(K,\omega) + \mathrm{PEN}_1^*(K,\omega) + \mathrm{PEN}_2^*(K,\omega). \quad (4.3.10)$$

The difference between AIC and AIC* is that the second criterion takes into account the difficulty of estimating the eigenstructure of the pooled spectral density operator, in addition to penalizing for the roughness of the pooled spectral density operator with respect to $\mathscr{F}_\omega^{1,(T)}$ and $\mathscr{F}_\omega^{a,(T)}$ (see Bosq (2000, Lemma 4.3)). We also note that both criteria are invariant to scaling of the sample spectral density operator.

Numerical simulation (see Section 4.3.5) suggest that in order to maximize

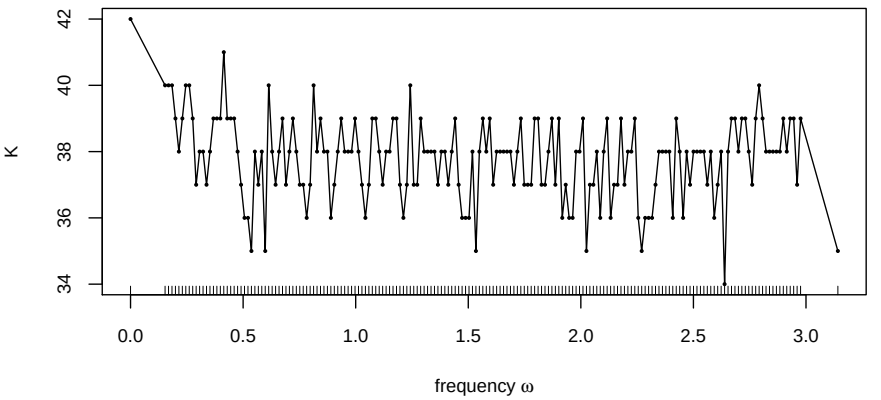


Figure 4.7 – Truncation levels $K(\omega)$ as chosen by the AIC* criterion (4.3.10). The small ticks on the horizontal axis represent the grid of frequencies Γ for which the test is computed.

power, $K(\omega)$ should be selected with AIC in settings where the eigenvalues of the spectral density operators decay quickly, and that AIC* should be used in more rough settings, where the eigenvalues of the spectral density operators decay slowly. Since our DNA minicircle data corresponds to the second case, we shall use AIC* in the rest of the chapter.

4.3.3 Localization of Differences on the Frequencies, Multiplicity Adjustments

Recall that H_ω denotes the null hypothesis $\mathscr{F}_\omega^1 = \mathscr{F}_\omega^2$. In order to test the global null hypothesis

$$H_G := \bigcap_{\omega\in[0,\pi]} H_\omega,$$

we will first obtain marginal p-values for each of the null hypotheses H_ω, $\omega \in \Gamma$, where

$$\Gamma := \{\omega_1, \ldots, \omega_J\} \subset [0,\pi]$$

is a grid of frequencies, and then adjust the p-values to account for multiplicity effects. The p-values will be based on the asymptotic distribution of the test statistic $\tilde{\Delta}_K^{(T)}(\omega)$, given by Theorem 4.3.1.
The results of applying the automatic truncation level rule (4.3.10) to our DNA minicircle dataset are shown in Figure 4.7. We notice that the selected values of $K(\omega)$ vary between 34 and 42. The corresponding (approximate) p-values are

$$p_j = \mathbb{P}\left(\chi^2_\nu(\omega_j) > \tilde{\Delta}^{(T)}_{K(\omega_j)}(\omega_j)\right), \quad j = 1, \ldots, J,$$

where $\nu(\omega_j) = K(\omega_j)[K(\omega_j)-1]/2$ if $\omega_j \in \{0,\pi\}$, and $\nu(\omega_j) = K(\omega_j)^2$ otherwise. The choice of the grid of frequencies at which the p-values are computed should be guided by a priori knowledge of the nature of the alternative hypothesis; see Section 4.3.4. In our case, we chose a grid of 187 frequencies, which is shown in Figure 4.7. Adjusting the p-values for multiplicities can be done either to control the family-wise error rate

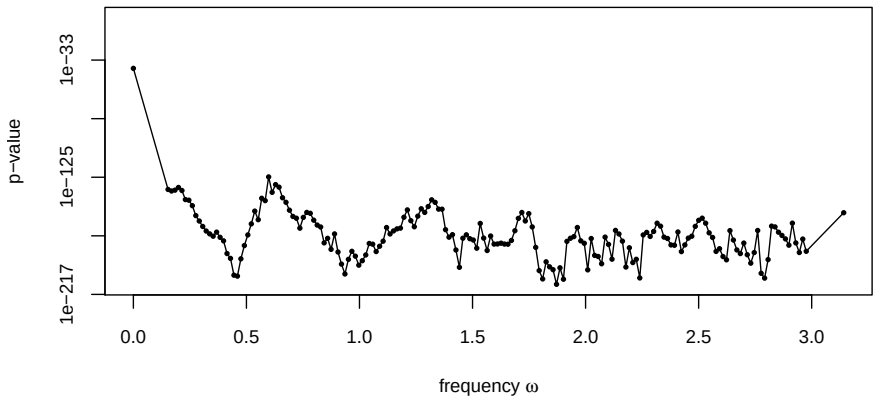


Figure 4.8 – Adjusted p-values (using Holm's procedure) for testing the equality of the spectral density operators of CAP and TATA, with the truncation level $K(\omega)$ automatically chosen at each frequency ω with the pseudo-AIC criterion (4.3.10).

(FWER; Dudoit et al. 2003), which is the probability of making at least one false rejection, or to control the less stringent false discovery rate (FDR; Benjamini & Hochberg 1995), which is the expected value of the proportion of false rejections amongst all rejections. Control of the FWER can be achieved via Holm's procedure (Dudoit et al. 2003). For controlling the FDR, since the p-values p_i and p_j are dependent for $|\omega_i - \omega_j| < 0.15$, and approximately independent for $|\omega_i - \omega_j| > 0.15$, we are in the context of dependence in finite blocks, and the original Benjamini-Hochberg (BH) algorithm for controlling the FDR is appropriate (Storey et al. 2004). In our case, since the difference between the two minicircles is quite strong, we only show the adjustment using Holm's procedure; see Figure 4.8. We notice that the two spectral density operators are very significantly different at all frequencies. More generally, our advice is to use the BH procedure for FDR adjustment, which will have more power in smaller-sample situations. We also conducted numerical simulations to assess the performance and validity of our procedure in finite sample; these suggest that both Holm's and BH procedure are valid for small sample sizes (see Section 4.3.5).

4.3.4 *Choice of the Discretization Grid* Γ

The choice of the grid Γ (and therefore J) is related to the alternative against which we wish to test the global null $H_G = \cap_{\omega\in[0,\pi]} H_\omega$:

Power for global differences between the two spectral density operators: If we believe that the true differences between the two spectral density operators are going to be on a large subinterval of $[0,\pi]$, J should be small, so that the power of the test is not lost because of multiple comparisons. If $\Gamma \subset [0,\pi]$ is chosen such that $|\omega_i - \omega_j| \geq 2B_T$, for all $\omega_i \neq \omega_j \in \Gamma$, then Hochberg's procedure may be applied for multiplicity corrections.

Power for narrow banded differences between the two spectral density operators: If we believe that the true difference are in a very narrow band of the spectra, e.g. H_ω is false only for $|\omega - \omega'| < \delta$, with $\omega' \in [0,\pi]$ and $\delta > 0$ small, then Γ should be chosen to be a dense grid over $[0,\pi]$. The largest gap between any two consecutive frequencies in Γ will indicate approximately smallest band-size δ for which the test would be able to detect departures from the global null H_G.

Frequencies near $\{0,\pi\}$**:** Although we expect $\tilde{\Delta}_K^{(T)}(\omega_j)$ to follow, for large T, approximately a $\chi^2_{K^2}$ distribution for any $\omega_j \notin \{0,\pi\}$, the approximation might not hold for frequencies ω_j very close to $\{0,\pi\}$. This happens because the asymptotic distribution of $\tilde{\Delta}_K^{(T)}(\omega)$ is $\chi^2_{K(K+1)/2}$ for $\omega \in \{0,\pi\}$, but $\chi^2_{K^2}$ for $\omega \in (0,\pi)$, and because $\tilde{\Delta}_K^{(T)}(\omega)$ is continuous in ω. Therefore, for ω_j close to $\{0,\pi\}$, the approximate distribution of $\tilde{\Delta}_K^{(T)}(\omega_j)$ is approximately a mixture of $\chi^2_{K(K+1)/2}$ and $\chi^2_{K^2}$ random variables, with unknown mixture proportion.

Another justification for this phenomenon comes from the fact that the sample spectral density operators, on which the test statistics $\tilde{\Delta}_K^{(T)}(\omega)$ are based, are constructed by smoothing locally the periodogram operators $\mathscr{P}_\omega^{(T)} := \widetilde{X}_\omega^{(T)} \otimes_2 \widetilde{X}_{-\omega}^{(T)}$. Therefore, if ω_j is close to 0 (say), the periodogram operators, upon which the sample spectral density operators $\mathscr{F}_0^{(T)}$ and $\mathscr{F}_{\omega_j}^{(T)}$ will be based, will intersect and the tests $\tilde{\Delta}_K^{(T)}(0)$ and $\tilde{\Delta}_K^{(T)}(\omega_j)$ will be correlated.

We therefore recommend that all the frequencies $\omega \in \Gamma$, with $\omega \notin \{0,\pi\}$, be at least at distance $2B_T$ of the frequencies $\{0,\pi\}$.

4.3.5 *Numerical Simulations*

In order to assess the finite sample performances of our testing procedure, we conducted some numerical simulations. The situation where the truncation level K is chosen using either AIC of AIC* is of particular interest, since our asymptotic framework requires having K fixed (deterministic) and $T \to \infty$, whereas K chosen with AIC/AIC* is random. Our simulation procedure is similar to those presented in Section 3.9: we simulate a stationary functional times series admitting the linear representation

$$X_t^{[\alpha]} = \sum_{s=0}^{2} \alpha_s A_s \varepsilon_{t-s},$$

where $\alpha_s \in \mathbb{R}$ are scaling parameters (described below), A_s are bounded operators, and the ε_t are i.i.d. random functions (the innovations), repre-

sented using a truncated Karhunen–Loève expansion:

$$\varepsilon_t(\tau) = \sum_{k=1}^{20} \xi_{k,t}\sqrt{\lambda_k} e_k(\tau),$$

and $e_k(\tau) = \sqrt{2}\sin[(k-1/2)\pi\tau]$ is orthonormal system in $L^2([0,1],\mathbb{R})$, see e.g. Adler (1990). The $\xi_{k,t}$ are i.i.d. random variables, whose choice governs the distribution of the random functions ε_t. The λ_ks are numbers that describe the roughness of random curves ε_t. If $\lambda_k \to 0$ very fast, then the curves are smooth. Conversely, if λ_k decays slowly, the curves are rough. We will consider the three following scenarios for $\xi_{k,t}$ and λ_k:

`Wiener`: the $\xi_{k,t}$ are independent standard Gaussian random variables, and

$$\lambda_k = 1/[(k-1/2)^2\pi^2].$$

The random curves ε_t therefore correspond to an approximation of the Wiener process, where the approximation is due to the truncation of the Karhunen–Loève expansion of ε_t.

`White-noise`: the $\xi'_{k,t}$ are independent standard Gaussian random variables, and $\lambda_k = 1$ for all $k \geq 1$. This process corresponds to a rougher version of the `Wiener` scenario, and is a projection of a true Gaussian white noise process.

`Student5`: The $\xi_{k,t}$ are i.i.d. distributed random variables, following Student's t distribution with $\nu = 5$ degrees of freedom, and $\lambda_k = 1$ for all $k \geq 1$. This process is similar to the `White-noise` process, except it is not Gaussian, and only its first 4 moments are finite.

We have constructed the operators A_s so that their image is contained within a 20-dimensional subspace of $L^2([0,1],\mathbb{R})$, spanned by an orthonormal basis $\psi_1,\ldots,\psi_{20}$. Representing ε_t in the $(e_k)_{k=1}^{20}$ basis, and A_s in the $(\psi_m \otimes_2 e_k)_{m,k=1}^{20}$ basis, we obtain a matrix representation of the process $X_t^{[\alpha]}$ as

$$\mathbf{X}_t^{[\alpha]} = \sum_{s=0}^{2} \alpha_s \mathbf{A}_s \boldsymbol{\varepsilon}_{t-s},$$

where $\mathbf{X}_t^{[\alpha]}$ is a 20×1 matrix, each $\mathbf{A}_s$ is a 20×20 matrix, and $\boldsymbol{\varepsilon}_t$ is a 20×1 matrix.

The matrices $\mathbf{A}_s$ are constructed by drawing, for each of their coordinates, i.i.d. Gaussian variables with mean 1 and standard deviation 0.5. In practice, their construction is done by fixing the random seed to a pre-chosen value and using the same generation scheme for each simulation run. The parameters $\alpha = (\alpha_0, \alpha_1, \alpha_2)$ are used to make the spectral density

operators of $X_t^{[\alpha]}$ less constant. We chose

$$\alpha = \alpha(\texttt{ma.diff}) = (-1.4, 2.3, -2 + \texttt{ma.diff}) \tag{4.3.11}$$

where `ma.diff` is a parameter that is allowed to change. The trace of the spectral density operators for the `Wiener` scenario are shown is in Figure 4.13 on page 185 for `ma.diff` $= 0, 0.1, \ldots, 0.5$. Visual appreciation of the roughness of each process can be obtained by plots of the percentage of explained variation at each frequency $\omega \in [0, \pi]$, that is, the proportion of the total variation of the infinitesimal increment process dZ_ω contained in its first k eigenspaces, i.e.

$$\frac{\sum_{j=1}^{k} \mu_j(\omega)}{\sum_{j=1}^{20} \mu_j(\omega)}, \quad k = 1, 2, \ldots, 20. \tag{4.3.12}$$

The percentages of explained variation per frequency are shown in Figure 4.11 on page 184 for the `Wiener` scenario, and in Figure 4.12 on page 184 for the `White-noise` and `Student5` scenarios. Notice that $k = 1$ already explains at least 80% of the variation at each frequency for the `Wiener` scenario, whereas we need to take $k = 8$ in the `White-noise` scenario to explain 80% of the variation at frequencies near 0.9.

For each `ma.diff` $\in \{0, 0.1, \ldots, 0.5\}$, and each $T \in \{2^6, 2^7, \ldots, 2^{10}\}$, we simulated for $b = 1, \ldots, B$ stretches of length T of the time series $X_t^{[\alpha(0)]}$ and $X_t^{[\alpha(\texttt{ma.diff})]}$. Denoting these observed times series by $X^{b,1}$ and $X^{b,2}$, we computed their spectral density operators using the bandwidth $B_T = T^{-1/5}$, (e.g. Grenander & Rosenblatt (1957, Par. 4.7), Brillinger (2001, Par. 7.4)) and took the weight function $W(x)$ to be the *Epanechnikov kernel* (e.g. Wand & Jones 1995), $W(x) = \frac{3}{4}(1 - x^2)$ if $|x| < 1$, and zero otherwise. We then compute and store the p-values

$$p_{j,k}^b = \mathbb{P}\left(\tilde{\Delta}_k^{(T)}(\omega_j) > \chi_\nu^2(\omega_j)\right), \quad j = 1, \ldots, J; k = 1, \ldots, 10,$$

where $\tilde{\Delta}_k^{(T)}(\cdot)$ is defined in Theorem 4.3.1, $\nu(\omega_j) = K(\omega_j)[K(\omega_j) - 1]/2$ if $\omega_j \in \{0, \pi\}$, and $\nu(\omega_j) = K(\omega_j)^2$ otherwise, and

$$\Gamma = \{\omega_1, \ldots, \omega_J\} \subset [0, \pi]$$

is a grid of frequencies. Following the discussion of Section 4.3.4, we choose

$$\omega_1 = 0; \omega_2 = 2B_T; \omega_{J-1} = \pi - 2B_T; \omega_J = \pi;$$

and $\omega_3, \ldots, \omega_{J-2}$ equispaced with spacing $\Delta\omega = \frac{2B_T}{10\pi}$. The reason behind this choice is that the spectral density operators are computed by weighted averaging of the periodogram operator on frequencies belonging to an interval of length $2B_T$. Therefore the spectral density operators will be very variable (approximately independent) for two frequencies separated

by $2B_T$, but not very variable if the frequencies are at distance less than $2B_T$. Our choice of $\Delta\omega$ corresponds to a coverage of each interval of length $2B_T$ with at least $9 = 10 - 1$ points of the grid Γ, except of course near the extremities $\{0, \pi\}$.

Boxplots of the raw p-values for $k = 1, \ldots, 4$ and $T \in \{128, 1024\}$ are shown for the `Wiener` scenario in Figures 4.14 to 4.19 (pages 186 to 191), for the `White-noise` scenario in Figures 4.20 to 4.25 (pages 193 to 198), and for the `Student5` scenario in Figures 4.26 to 4.31 (pages 200 to 205). In each of these plots, we drew a solid curve representing the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts for computing (4.3.1). This curve shows at which frequencies we would be able to detect the differences between the two spectral density operators, within each fixed value of k. We notice that the p-values $(p^b_{j,k})_{b=1,\ldots,B}$ seem to be uniformly distributed for `ma.diff` $= 0$, and their distribution is skewed towards zero at frequencies/k regions where the difference between the two spectral densities can indeed be detected; the skewness is accentuated for larger values of T, and also for larger values of `ma.diff`.

For each depth level k, and each b, we adjust the p-values $(p^b_{j,k})_{j=1,\ldots,J}$ using the Holm procedure. We then estimate the probability of rejecting the global null hypothesis $H_G := \bigcap_{\omega\in\Gamma} H_\omega$ at level α, within the depth k, by

$$B^{-1} \sum_{b=1}^{B} \mathbf{1}\left(\min_{j=1,\ldots,J} p^b_{j,k} \leq \alpha \right). \tag{4.3.13}$$

The results of this procedure—with $\alpha = 5\%$, and $k \in \{1, \ldots, 5\}$—are shown in Figures 4.32 on page 207, 4.33 on page 208 and 4.34 on page 209 for the `Wiener`, `White-noise` and `Student5` scenarios, respectively. We notice that the level is respected in each scenario, and for each sample size $T = 128, \ldots, 1024$. For the `Wiener` scenario, choosing $k = 1$ does not yield a powerful test, even with a sample size of $T = 1024$. However, taking $k = 2$ seems to already be reasonably powerful, even at sample size $T = 256$, at which it is almost as competitive as $k = 3, 4, 5$. For larger sample sizes, choosing k larger yields big differences in power: for instance in the case $T = 1024$, at `ma.diff` $= 0.2$, the power increases by roughly 0.2 if we increase k by one. The `White-noise` and `Student5` scenarios are more surprising: quite often, the case $k = 1$ yields the most (or near to the most) powerful test amongst $k = 1, 2, \ldots, 5$. At sample size $T = 1024$, and `ma.diff` $= 0.3$, the difference in power between $k = 1$ and $k = 5$ is about 40%, in favour of $k = 1$.

We also show in Figure 4.9 on page 177 the estimated probabilities of rejection for the global null, when the truncation level k is chosen either by AIC or by AIC*. That is, if we denote by $K(\omega)$, respectively $K^*(\omega)$, the truncation level K that minimizes the AIC, respectively the AIC* criteria,

we plot the values of

$$B^{-1}\sum_{b=1}^{B}\mathbf{1}\left(\min_{j=1,\dots,J}p^{b}_{j,K(\omega)}\leq\alpha\right) \tag{4.3.14}$$

and

$$B^{-1}\sum_{b=1}^{B}\mathbf{1}\left(\min_{j=1,\dots,J}p^{b}_{j,K^{*}(\omega)}\leq\alpha\right) \tag{4.3.15}$$

for each `ma.diff` $\in\{0,0.1,\dots,0.5\}$ and $T=64,128,\dots,1024$, for $\alpha=5\%$. We see that in the `Wiener` scenario, criterion AIC is much better than AIC*. However, in the `White-noise` or `Student5` scenarios, the opposite holds. Moreover, for the low sample sizes $T=64$, AIC seems to fail. Except for that, in every scenario and every other sample size, both criteria seem to respect the level $\alpha=5\%$. The power using either AIC criterion is not necessarily larger than that obtained using a fixed pre-chosen value for the truncation level k (that is, without correcting for multiplicity in the choice of k). However, it is not clear how to choose k a priori. Therefore, for most power in rejecting the global hypothesis H_G, we recommend using the AIC criterion in settings that are quite smooth (when the eigenvalues of the spectral density operators decay quickly), and the AIC* criterion in rougher settings (when the eigenvalues of the spectral density operators decay slowly).

4.3.6 Further Comparison at the Level of the Minicircles?

We have shown in this section how to compare the dynamics of DNA minicircles, or more generally two functional time series, by comparing their spectral density operators, first marginally at each frequency, and then by adjusting the marginal p-values for multiplicities in order to locate where *in frequency* these differences occur. It turned out for our minicircles that their differences are significant at *all* frequencies. The goal of the next section is to look at finer differences between the dynamics of the two minicircles, and detect *where on the minicircles*, and *within each frequency*, the differences occur.

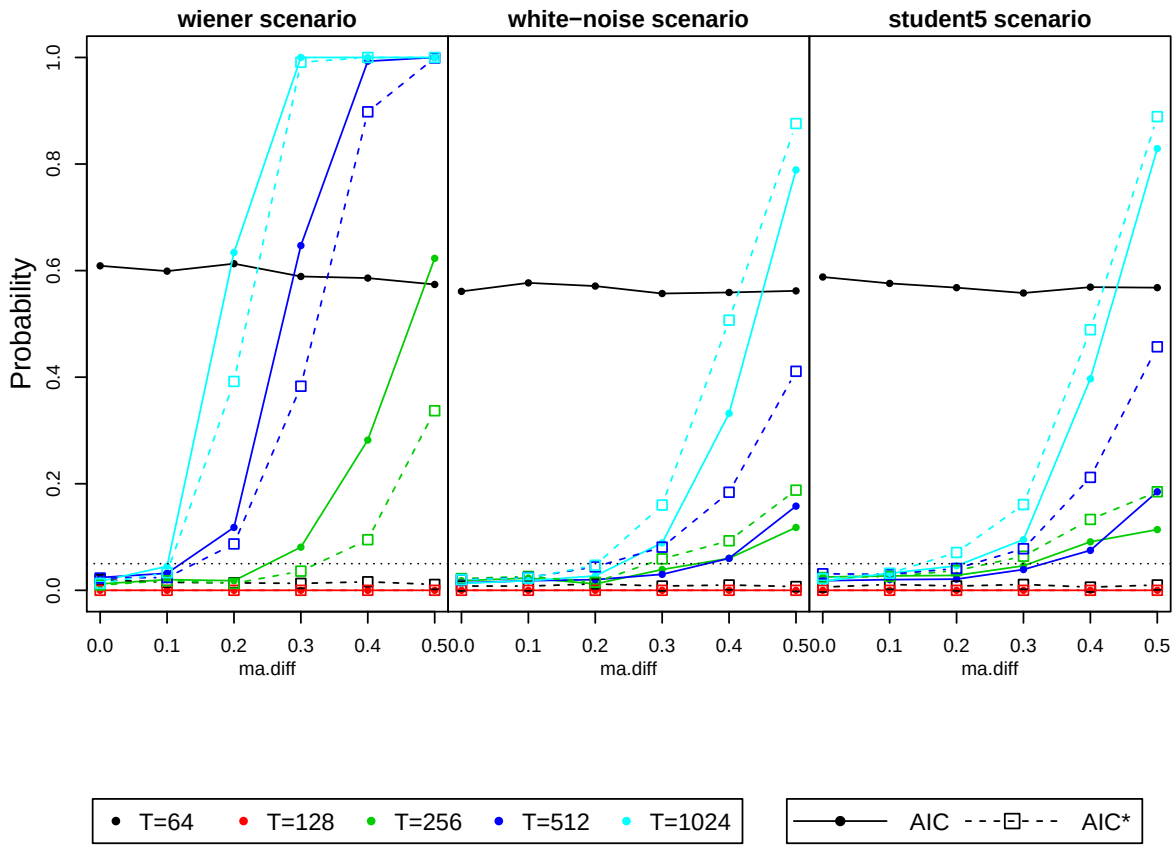


Figure 4.9 – The estimated probability of rejecting the global null hypothesis—see (4.3.14) and (4.3.15)—when the truncation levels K are chosen either by the AIC or the AIC* criterion. The standard deviation, based on a normal approximation, is at most $sd = 0.016$.

4.4 *Significant Frequencies, and Localization of Differences on the Minicircles*

We now wish to detect more detailed differences between CAP and TATA: we wish to first select frequencies at which CAP and TATA are significantly different (the *significant frequencies*), and then detect and localize on the minicircles, within each significant frequency, where the differences between the spectral density operators of CAP and TATA occur.

The basic idea is to base our procedure on the differences in the (i,j)-th basis coefficient between the spectral density operator of CAP and TATA, at a given frequency ω. Let us denote by $\mathbf{f}_\omega^a$, respectively $\mathbf{f}_\omega^{a,(T)}$, the 80×80 coefficient matrices with respect to the periodic B-spline basis (King et al. 2010) of the true spectral density operator, respectively the sample spectral density operator, at frequency ω, for the time series X_t^a, $a=1,2$. We shall call $\mathbf{f}_\omega^a$ the projected spectral density operator, and $\mathbf{f}_\omega^{a,(T)}$ the projected *sample* spectral density operator, and denote by $\mathbf{f}_\omega(i,j)$ the (i,j)-th entry of the matrix $\mathbf{f}_\omega$. The local null hypotheses we wish to test for are of the form

$$H_\omega(i,j):\mathbf{f}_\omega^1(i,j)=\mathbf{f}_\omega^2(i,j),\quad i,j=1,\ldots,80;\omega\in[0,\pi].$$

By symmetry of the projected spectral density operator, we restrict ourselves to the indices $1\le i\le j\le 80$. We point out that this approach is different from a classical multivariate approach, as discussed in Remark 4.4.1.

Remark 4.4.1 (Differences with multivariate analysis)**.** *Although the idea of comparing at the level of basis coefficients seems like a multivariate approach, it differs from it in that the choice of the basis functions will influence the qualitative conclusions that can be drawn from the analysis. Our choice of a periodic B-spline basis allows one to distinguish differences between CAP and TATA that are very localized on the minicircles. Another choice could be that of a wavelet basis, which would allow one to detect differences between CAP and TATA across multiple scales. The choice of the basis is therefore intimately related the directions (in function space) in which the test is most powerful.*

For each frequency ω and each $1\le i\le j\le 80$, assuming $\mathbf{f}_\omega(i,i)\mathbf{f}_\omega(j,j)\neq 0$, we can use the projected sample spectral density operator to construct a p-value $p(\omega;i,j)$ for the null hypothesis $H_\omega(i,j)$, as described in the following section.

4.4.1 Approximate p-values for Localizing Differences Within Frequencies

Theorem 3.6.5 tells us that under $\bigcap_{i,j=1}^{80} H_\omega(i,j)$, $\sqrt{TB_T}(\mathbf{f}_\omega^{1,(T)} - \mathbf{f}_\omega^{2,(T)})$ will be asymptotically distributed as a random matrix $\check{\mathbf{f}}_\omega$, which follows a complex Gaussian distribution (which is not necessarily circular; see Picinbono (1996), Schreier & Scharf (2010)) with mean zero, and second-order structure given by

Circular Gaussian random variable are defined in Definition 3.12.1 on page 134

$$\mathbb{E}\left[\check{\mathbf{f}}_{\omega_1}(i,j)\check{\mathbf{f}}_{\omega_2}(k,l)\right] = 4\pi\kappa^2 \cdot \left[\eta(\omega_1+\omega_2)\mathbf{f}_{\omega_1}(i,k)\mathbf{f}_{-\omega_1}(j,l) + \eta(\omega_1-\omega_2)\mathbf{f}_{\omega_1}(i,l)\mathbf{f}_{-\omega_1}(j,k)\right], \tag{4.4.1}$$

where $\kappa^2 = \int_{\mathbb{R}} W(x)^2 dx$ and $\eta(\omega) = 1$ if $\omega \in \{0, \pm\pi, \pm 2\pi\}$, and zero otherwise. We shall use this asymptotic distribution to obtain, for each (i,j), an approximate p-value for the null hypothesis $H_\omega(i,j)$, by renormalizing properly the difference $\mathbf{D}_\omega^{(T)} = \sqrt{TB_T}(\mathbf{f}_\omega^{1,(T)} - \mathbf{f}_\omega^{2,(T)})$. In the following, we shall use the notation $\tilde{\mathbf{f}}_\omega = (\mathbf{f}_\omega^{1,(T)} + \mathbf{f}_\omega^{2,(T)})/2$. The test statistic we use is

$$L^{(T)}(\omega,i,j) = \begin{cases} \left|\mathbf{D}_\omega^{(T)}(i,j)\right|^2 / \left\{4\pi\kappa^2\left[|\tilde{\mathbf{f}}_\omega(i,j)|^2 + \tilde{\mathbf{f}}_\omega(i,i)\tilde{\mathbf{f}}_\omega(j,j)\right]\right\} & \text{if } \omega \in \{0,\pi\} \\ \left|\mathbf{D}_\omega^{(T)}(i,i)\right|^2 / \left\{4\pi\kappa^2|\tilde{\mathbf{f}}_\omega(i,i)|^2\right\} & \text{if } \omega \in (0,\pi),\ i=j, \\ \left\{|\mathbf{D}_\omega^{(T)}(i,j)|^2/P - \Re\left[(\mathbf{D}_\omega^{(T)}(i,j))^2 R\right]/P\right\}/(2\pi\kappa^2) & \text{if } \omega \in (0,\pi),\ i\neq j, \end{cases}$$

where $\Re(\cdot)$ denotes the real part of a complex number,

$$P = P(\omega,i,j) = \tilde{\mathbf{f}}_\omega(i,i)\tilde{\mathbf{f}}_\omega(j,j) - |\tilde{\mathbf{f}}_\omega(i,j)|^4/[\tilde{\mathbf{f}}_\omega(i,i)\tilde{\mathbf{f}}_\omega(j,j)]$$

and $R = R(\omega,i,j) = [\overline{\tilde{\mathbf{f}}_\omega(i,j)}]^2/[\tilde{\mathbf{f}}_\omega(i,i)\tilde{\mathbf{f}}_\omega(j,j)]$.
The following proposition gives the asymptotic distribution of $L^{(T)}$, and its proof follows easily from results of Chapter 3 and Picinbono (1996):

Proposition 4.4.2. *Assume conditions of Theorem 3.6.5 hold, and*

$$B_T \to 0 \quad \& \quad TB_T \to \infty \quad as \quad T \to \infty.$$

Under $H_\omega(i,j)$, if $\mathbf{f}_\omega(i,i)\mathbf{f}_\omega(j,j) \neq 0$, the asymptotic distribution of the test statistic $L^{(T)}(\omega,i,j)$ is χ_1^2 if $\omega \in \{0,\pi\}$ or $i=j$, and χ_2^2 if $i \neq j$ and $\omega \in (0,\pi)$.

The reason the form of the test statistic $L^{(T)}(\omega,i,j)$ is quite complicated in the case $i \neq j \,\&\, \omega \in (0,\pi)$ is that in this case, $\mathbf{D}_\omega^{(T)}(i,j)$ follows a complex distribution that is not circular, and its renormalization cannot be done via the "usual" formula for Gaussian random vectors (Picinbono 1996, Schreier & Scharf 2010). Using Proposition 4.4.2, we can compute the approximate p-values $p(\omega,i,j), \omega \in [0,\pi], i \leq j$. Notice that within a frequency, the p-values $\{p(\omega,i,j) : i \leq j\}$ are correlated, with a complicated correlation structure. For instance, for $\omega \in \{0,\pi\}$ or $i=j \,\&\, k=l$, the

asymptotic covariance is given by

$$\lim_{T\to\infty} \operatorname{cov}\left(L^{(T)}(\omega,i,j), L^{(T)}(\omega,k,l)\right) = 2\operatorname{corr}\left(\check{\mathbf{f}}_\omega(i,j), \check{\mathbf{f}}_\omega(k,l)\right)^2, \qquad (4.4.2)$$

which is in general non-zero, see (4.4.1).

4.4.2 Back to the computation of the P-values within a frequency

The p-values are only computed on a subgrid $\Gamma = \{\omega_1,\dots,\omega_L\} \subset [0,\pi]$, which is chosen such that $|\omega_i - \omega_j| \geq 2B_T$, so that the p-values across different ω_js are approximately independent (see discussion in Section 4.3.4). We chose to select significant frequencies and localize the differences between CAP and TATA in a way that controls the expected average of the false discovery proportion over the significant frequencies (Benjamini & Bogomolov 2014). To make this statement precise, let $\mathbf{p}_l = \{p(\omega_l; i, j) : 1 \leq i \leq j \leq 80\}$ be the set of p-values at frequency ω_l, and $\mathbf{P} = \{\mathbf{p}_1,\dots,\mathbf{p}_L\}$ be the set of all p-values over the grid Γ. Let $S(\mathbf{P})$ be the selection procedure for the significant frequencies, based on all the p-values $\mathbf{P}$, that is $S(\mathbf{P}) \subset \Gamma$, and $|S(\mathbf{p})|$ denote the number of significant frequencies. Let $\mathrm{FDP}(\omega) = V(\omega)/R(\omega)$ be the false discovery proportion at frequency ω, where $V(\omega)$ denotes the (unknown) number of wrong rejections within frequency ω, and $R(\omega)$ denotes the total number of rejections at frequency ω. The error criterion we will seek to control is

$$\mathbb{E}\left[\sum_{l\in S(\mathbf{P})} \mathrm{FDP}(\omega_l)/\max\{|S(\mathbf{P})|,1\}\right]. \qquad (4.4.3)$$

Notice that if the selection procedure S is implemented without relying on the data, (4.4.3) simplifies to $\sum_{l\in S}\mathrm{FDR}(\omega_l)/|S|$, the average FDR over the selected frequencies, where $\mathrm{FDR}(\omega_l) = \mathbb{E}[V(\omega_l)/R(\omega_l)]$.
To select the significant frequencies and select the null hypotheses to reject within each significant frequency while controlling the expected average false discovery proportion (FDP) (4.4.3) at the level α, we use the following procedure (see Benjamini & Bogomolov 2014, Theorem 1 and Section 5):

1. Adjust, within each frequency ω_l, the p-values $\mathbf{p}_l$ for the control of the FDR, and denote the result by $\mathbf{q}_l$, also called *q-values* (see Remark 4.4.3).

2. Select the significant frequencies S by applying the BH procedure to the set of minimum q-values $\{\min\mathbf{q}_1, \min\mathbf{q}_2,\dots,\min\mathbf{q}_L\}$.

3. Within each significant frequency ω_l, $l \in S$, reject the null hypotheses whose corresponding q-value are smaller than $|S|\alpha/L$.

In addition to controlling the error criterion (4.4.3), this procedure also has the additional property that it controls the FDR at the level of the frequencies.

Remark 4.4.3 (P-value adjustment within each frequency)**.** *Since the p-values*
$\mathbf{p}_l = \{p(\omega_l; i, j) : 1 \le i \le j \le 80\}$ *are correlated with a non-trivial correlation structure (see e.g.* (4.4.2)*), we cannot use the BH procedure to control the FDR, nor the more recent procedures (which require e.g. dependence in finite blocks, see Storey et al. (2004) and Schwartzman et al. (2008) for instance). We therefore use the conservative version of FDR which works under arbitrary dependence structure of the p-values (Benjamini & Yekutieli 2001, Theorem 1.3) to obtain the q-values* $\mathbf{q}_l$*. Nevertheless, numerical simulation (not shown here) that we carried out to assess the validity of our procedure suggested that the BH procedure seems to control the FDR within each frequency. Further work along this line would be of interest.*

The result of applying this procedure to our minicircle data with $\alpha = 0.05$, and on the grid of frequencies

$$\Gamma = \{0, 0.15, 0.32, 0.49, 0.64, 0.81, 0.98, 1.15, 1.3, 1.47,$$
$$1.64, 1.81, 1.96, 2.13, 2.3, 2.47, 2.62, 2.79, 2.96, 3.14\},$$

are shown in Figure 4.10 on page 183 in the form of *zero-one plots*, which show graphically the regions where the spectral density operators of CAP and TATA differ significantly. We first notice that all the tested frequencies are significant, which is not surprising since the frequency tests (Section 4.3) suggested that the null hypothesis H_ω for each fixed frequency was rejected with a very small p-value. We also see that the rejected hypotheses are mostly situated on the diagonal of the spectral density operators, i.e., the rejected nulls are mostly of the form $H_\omega(i, j)$ with $|i - j|$ small. This means that most of the differences in the dynamics of CAP and TATA are about how their local interactions (between $X_t(\tau)$ and $X_0(\sigma)$ for $|\tau - \sigma|$ small) differ. This is not surprising since we have already seen (in Section 4.2.2) that most of the energy of the minicircles is in their local interactions. Nevertheless, some differences are detected farther away from the diagonal, at frequencies $\omega = 0.15$ or $\omega = 0.64$ for instance. This is potentially interesting, since a difference far from the diagonal would mean that the dynamics of the DNA minicircles, although being mostly local, seem to have different long-range ($|\tau - \sigma|$ large) effects, which are visibly varying throughout distinct frequencies. Further interpretation of these off-diagonal differences is difficult; notice also that since they are rare and we are controlling a FDR-type error, these might be false discoveries.

Another interesting conclusion is that the differences between the two minicircles do not only reside in the region where their base-pairs sequence is different (see Table 4.1), but extends to other regions of the minicircles. This phenomenon could be explained intuitively as a propagation effect by the following thought experiment, which is of course informal: imagine that a DNA minicircle is vibrating in a solution, and

that we introduce an impulse (or shock) on a localized region. If the impulse is strong enough (stronger than the "average" vibrations of the DNA minicircle that are due to its interaction with water molecules), it will propagate as a wave along the minicircle, with decreasing amplitude as it moves further along. The distance that the wave will travel will depend on the strength of the impulse, on the vibrations of the DNA minicircle, and on its mechanical properties. Nevertheless, we can imagine that this distance will be approximately an increasing function of the impulse strength. Therefore, a localized but strong enough vibration on a DNA minicircle could propagate and affect other regions of the minicircle. Applying this idea to CAP and TATA, their differences in base pairs (that is very localized) might create different kinds of vibrations that are also very localized, but might propagate along the minicircles, and have an impact on the vibrations of other regions of the minicircle.

4.5 Outlook

We have presented in this chapter a methodology for comparing the dynamics of DNA minicircles (or more generally any pair of functional time series), by comparing their spectral density operators either at the level of frequencies, or jointly at the level of frequencies and on the minicircles. Our technique was based on a multiple testing approach, thus allowing to localize the differences between the two spectral density operators with overall significance. Extensions and ameliorations of the methodology presented in this chapter can be done in several directions. One could use the asymptotic Gaussian structure of the sample spectral density operator at each frequency to increase the detecting power within each frequency, by looking at excursion sets of the sample spectral density kernel, and using the theory of excursion sets of Gaussian processes (Adler 1990, 2000, Vanmarcke 2010). The multiplicity correction could also be improved, by estimating and taking into account the local dependency (in frequencies) of the sample spectral density operators that are present in finite samples.

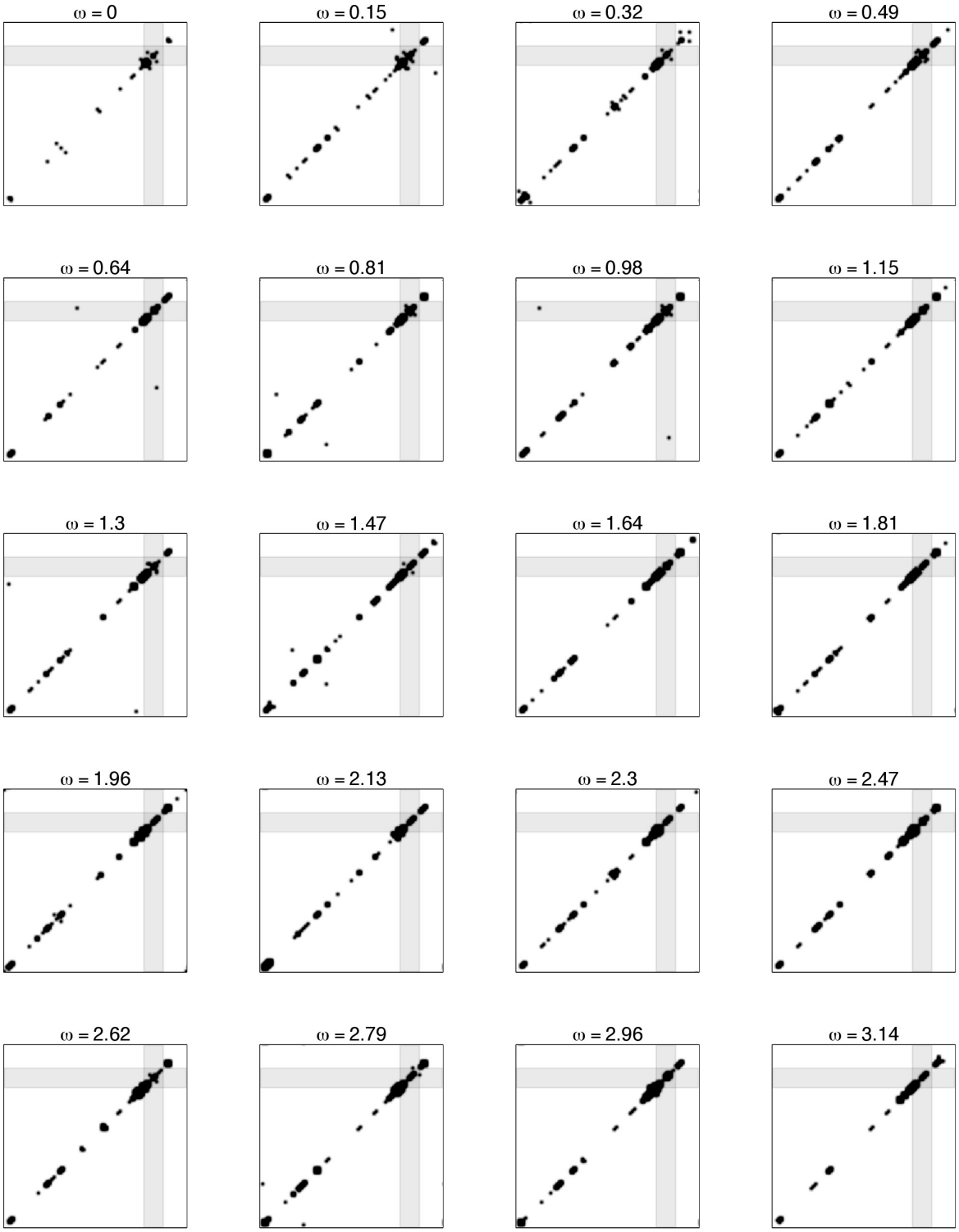


Figure 4.10 – The plots show the regions on the minicircles, for each frequency, where the spectral density operators of CAP and TATA are overall significantly different at a 5% level (with respect to the error criterion (4.4.3)). Each plot represents the regions of differences (in black) between the spectral density kernels of the two minicircles. The two grey vertical and horizontal bands correspond to the region where the base-pair sequences of the two DNA minicircles are different.

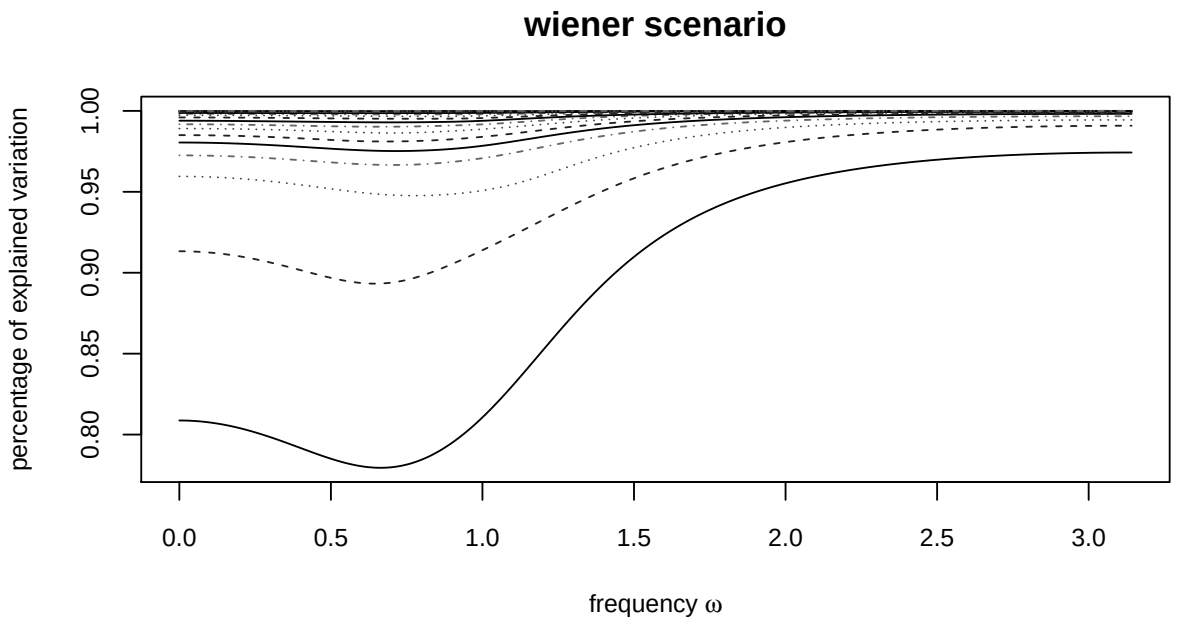


Figure 4.11 – The percentage of variation per frequency (computed using (4.3.12)), for the `Wiener` scenario. The bottom curve corresponds to the percentage of variation explained by $k = 1$, the one just above corresponds to $k = 2$, as so on. Notice the scale of the y-axis.

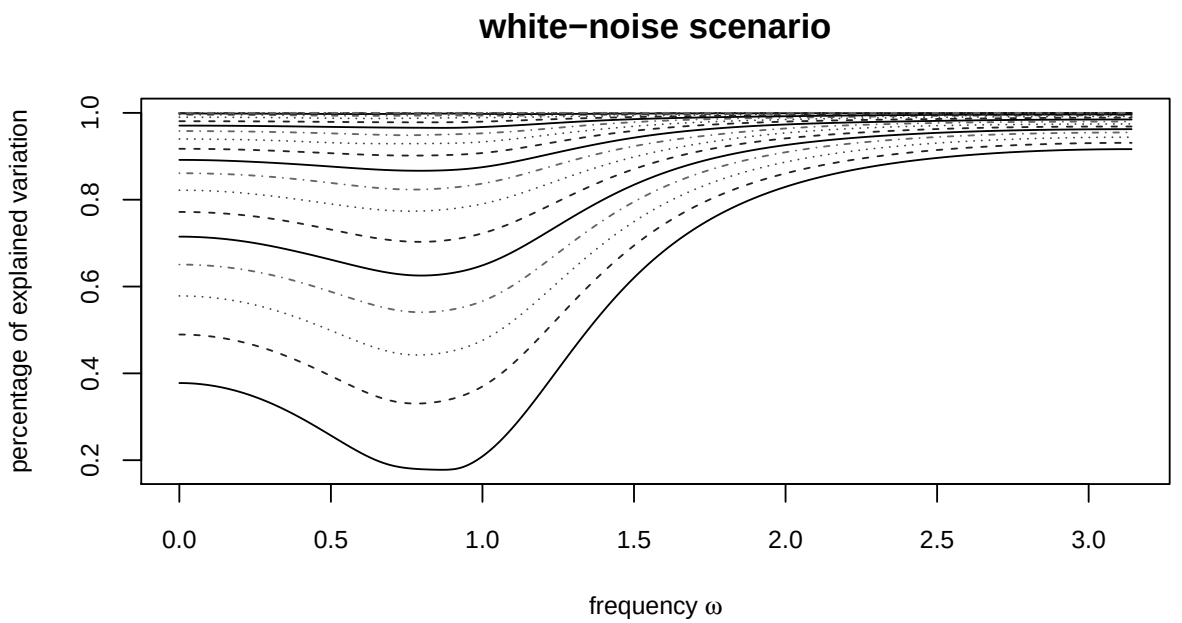


Figure 4.12 – The percentage of variation per frequency, for the `White-noise` and `Student5` scenarios. These are computed using (4.3.12). The bottom curve corresponds to the percentage of variation explained by $k = 1$, the one just above corresponds to $k = 2$, as so on. Notice the scale of the y-axis.

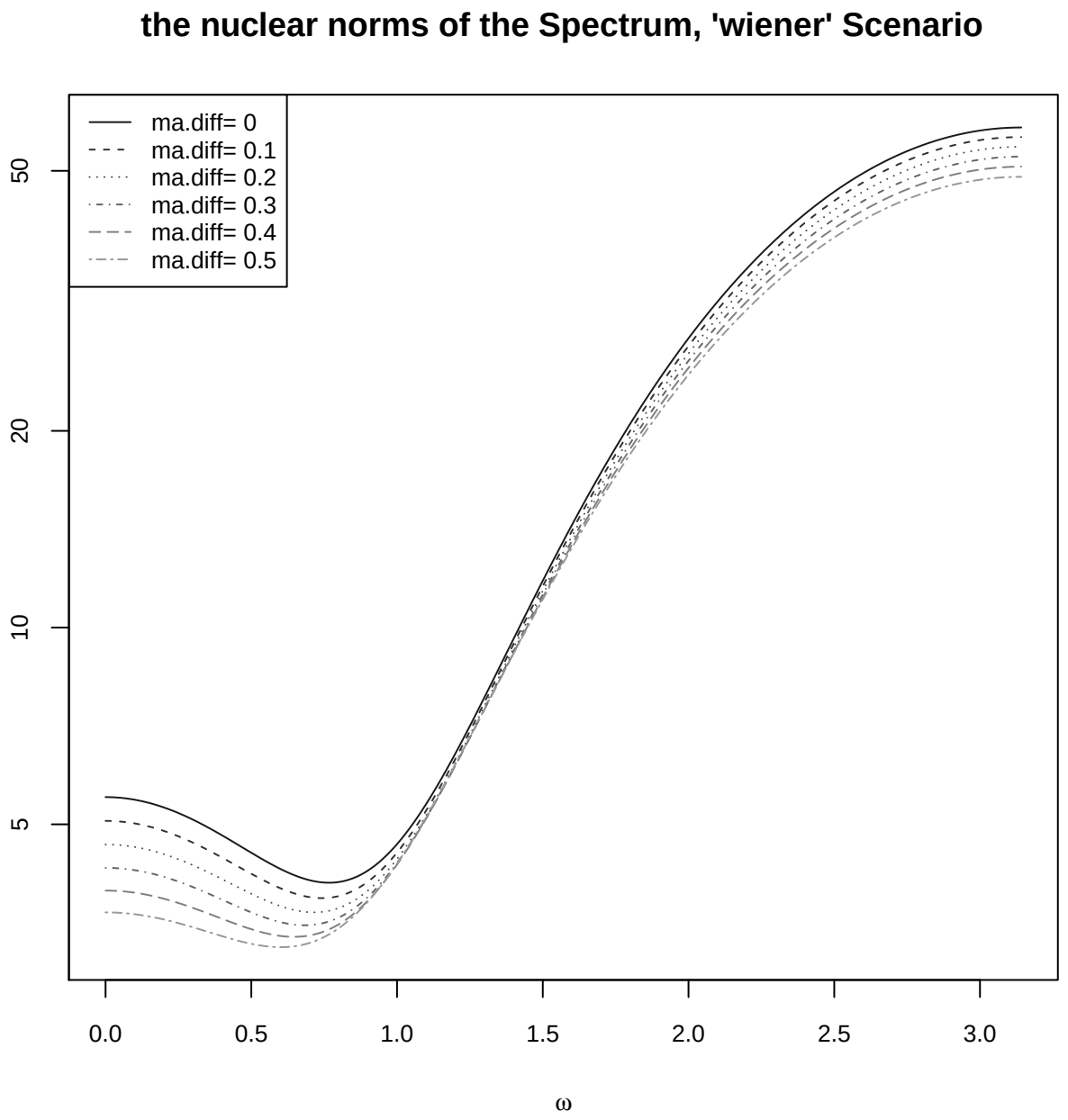


Figure 4.13 – The trace of the spectral density operators of $X_t^{[\alpha(\texttt{ma.diff})]}$ for different values of `ma.diff`.

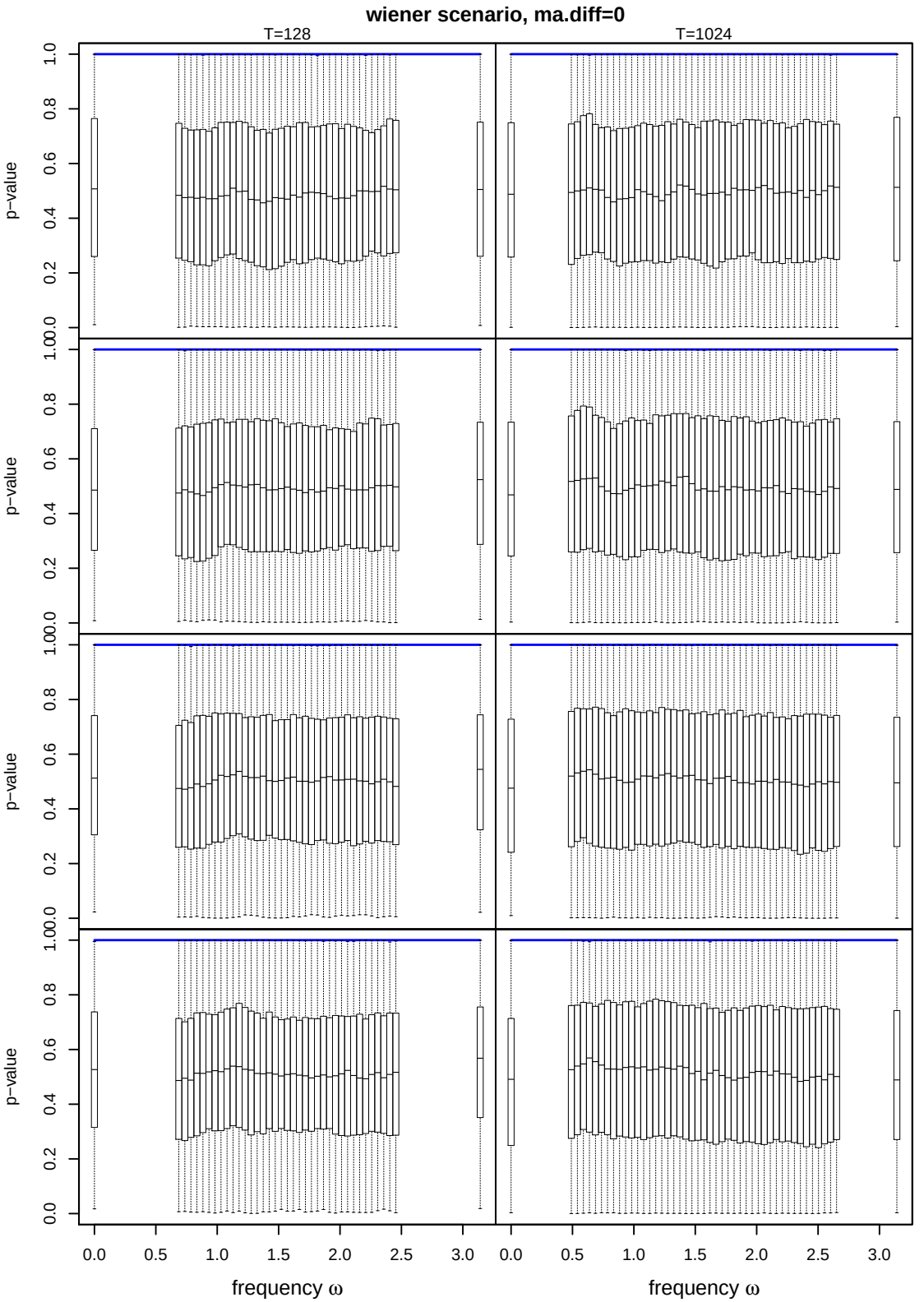


Figure 4.14 – Box plot of the computed p-values $p^b_{j,k}$ for the `Wiener` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

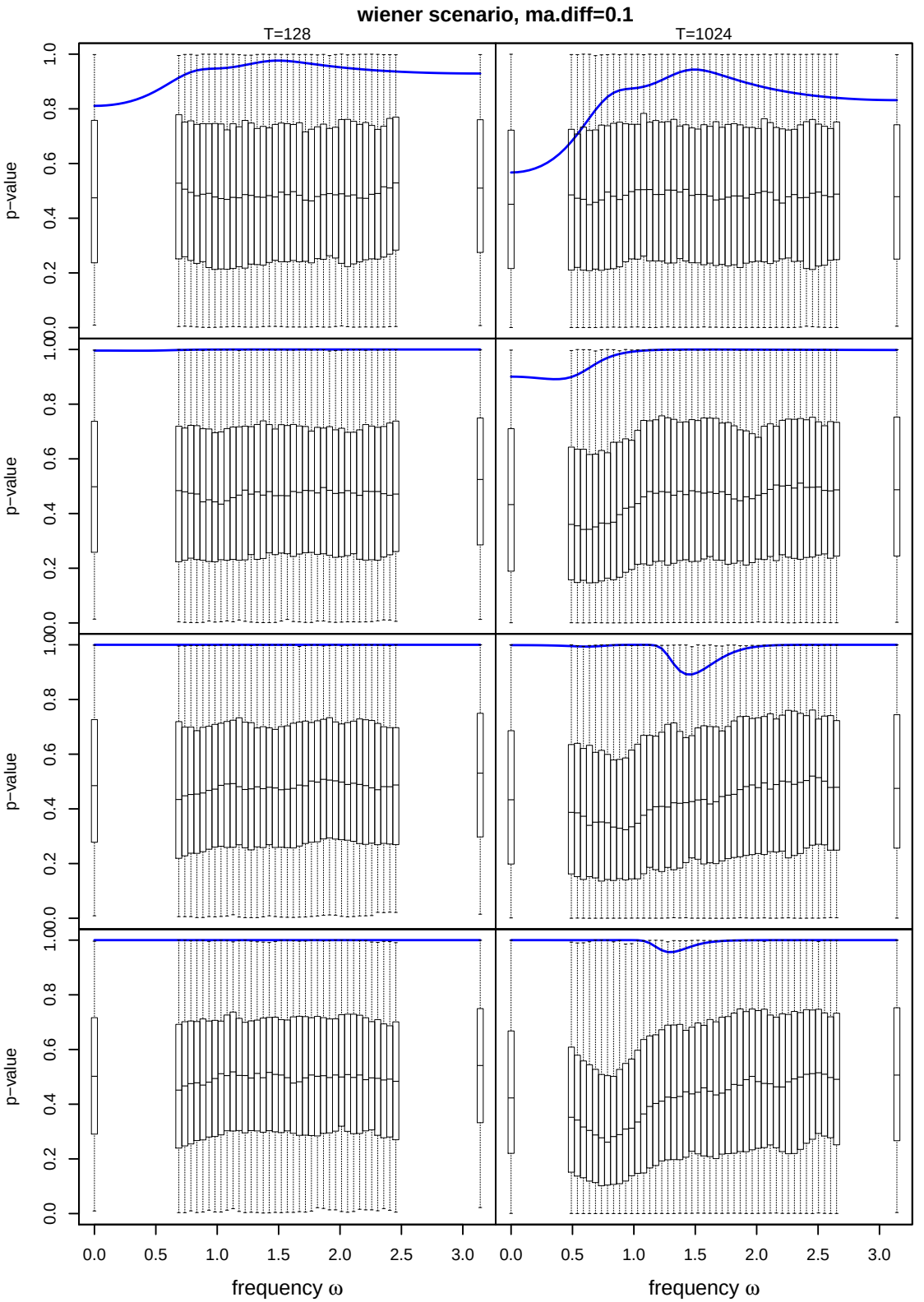


Figure 4.15 – Box plot of the computed p-values $p^b_{j,k}$ for the `Wiener` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

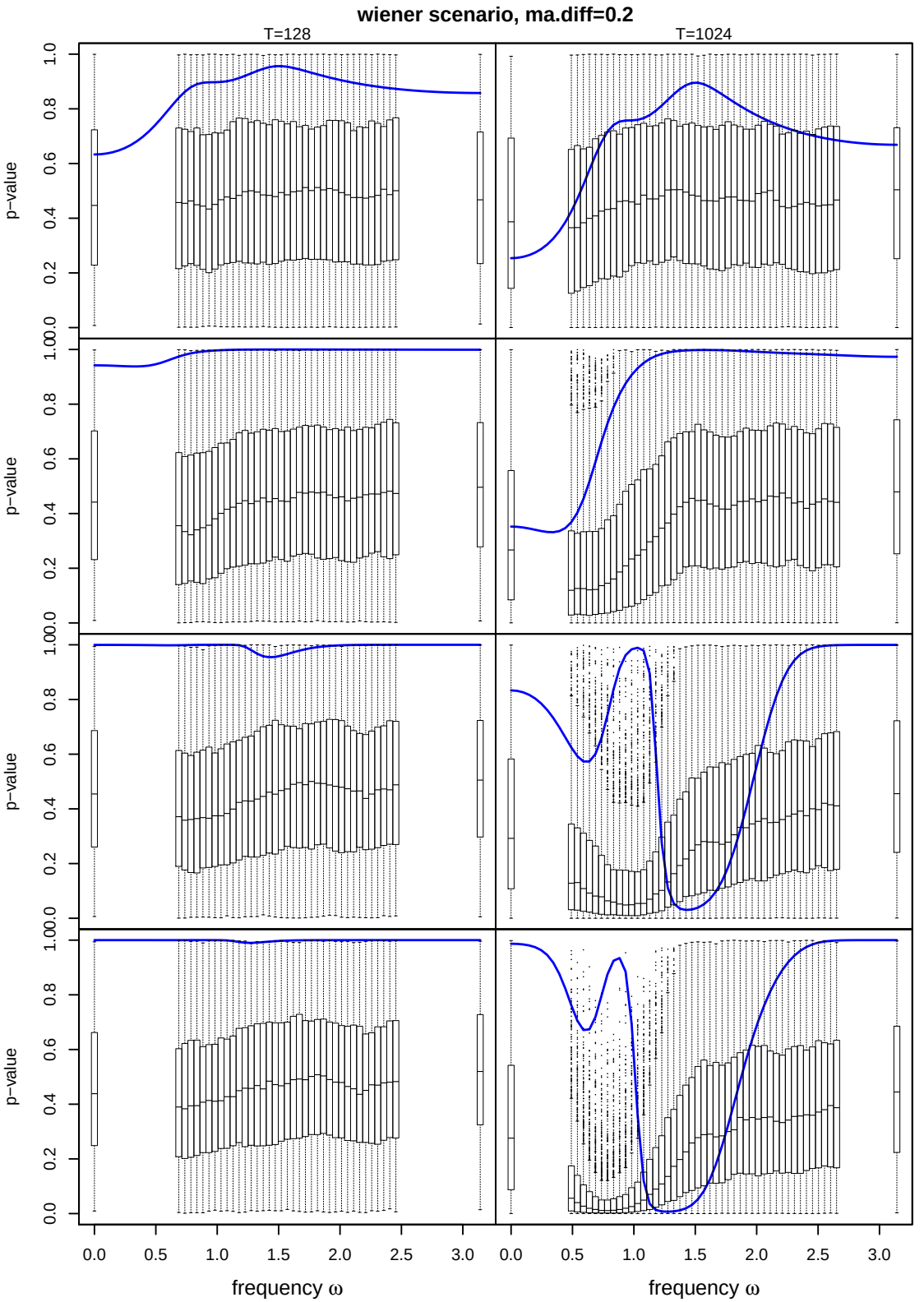


Figure 4.16 – Box plot of the computed p-values $p^b_{j,k}$ for the `Wiener` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

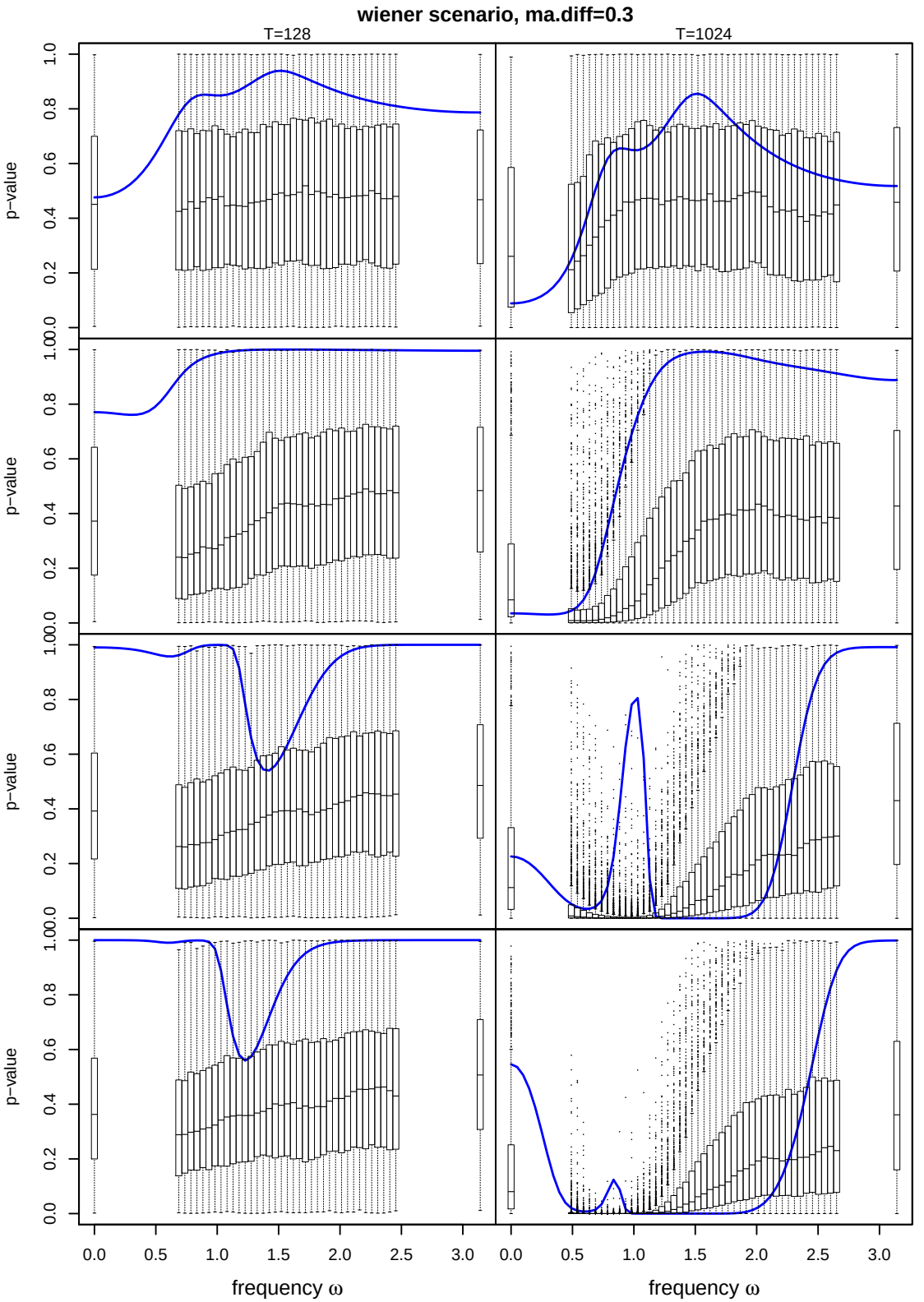


Figure 4.17 – Box plot of the computed p-values $p^b_{j,k}$ for the `Wiener` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

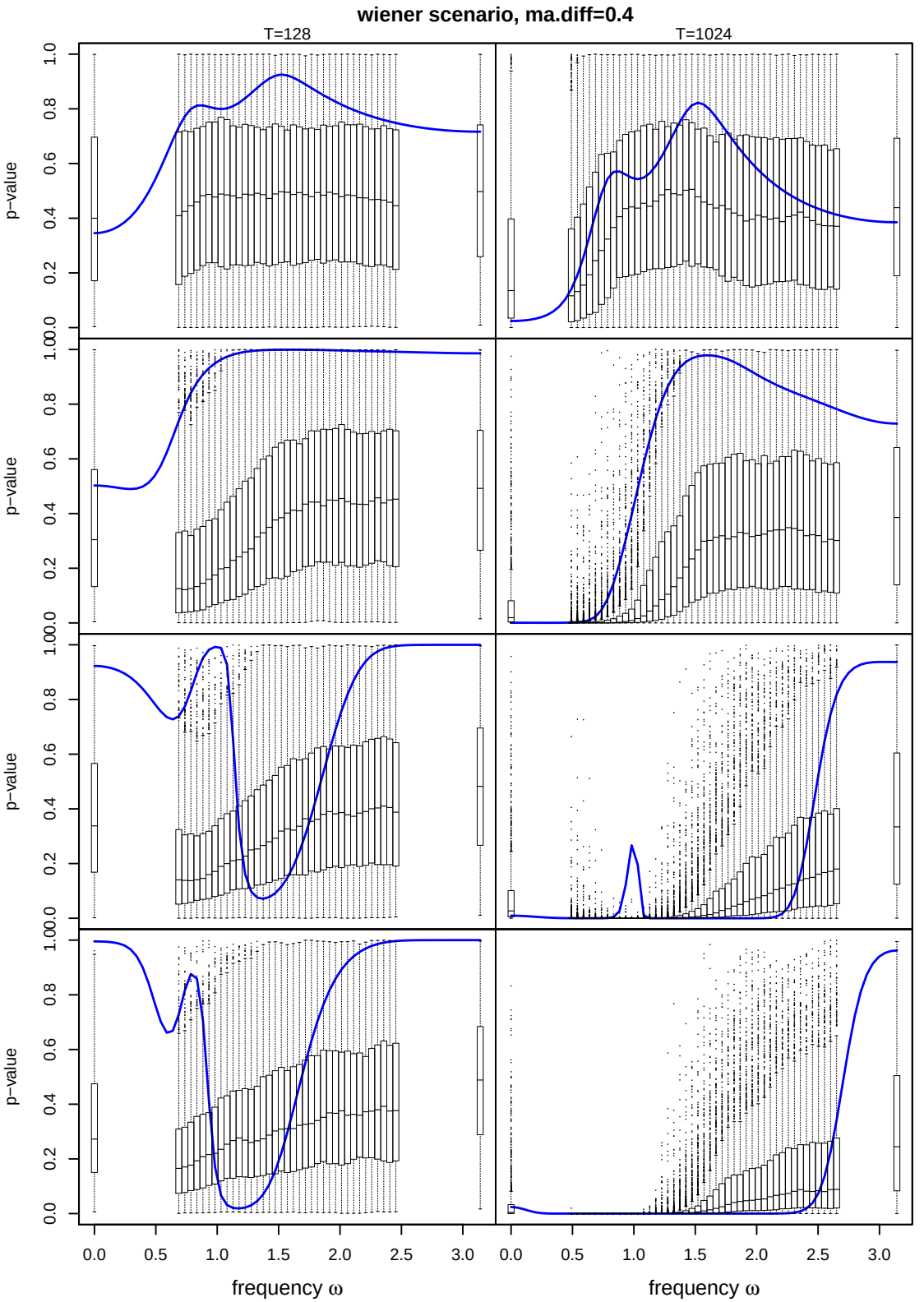


Figure 4.18 – Box plot of the computed p-values $p^b_{j,k}$ for the `Wiener` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

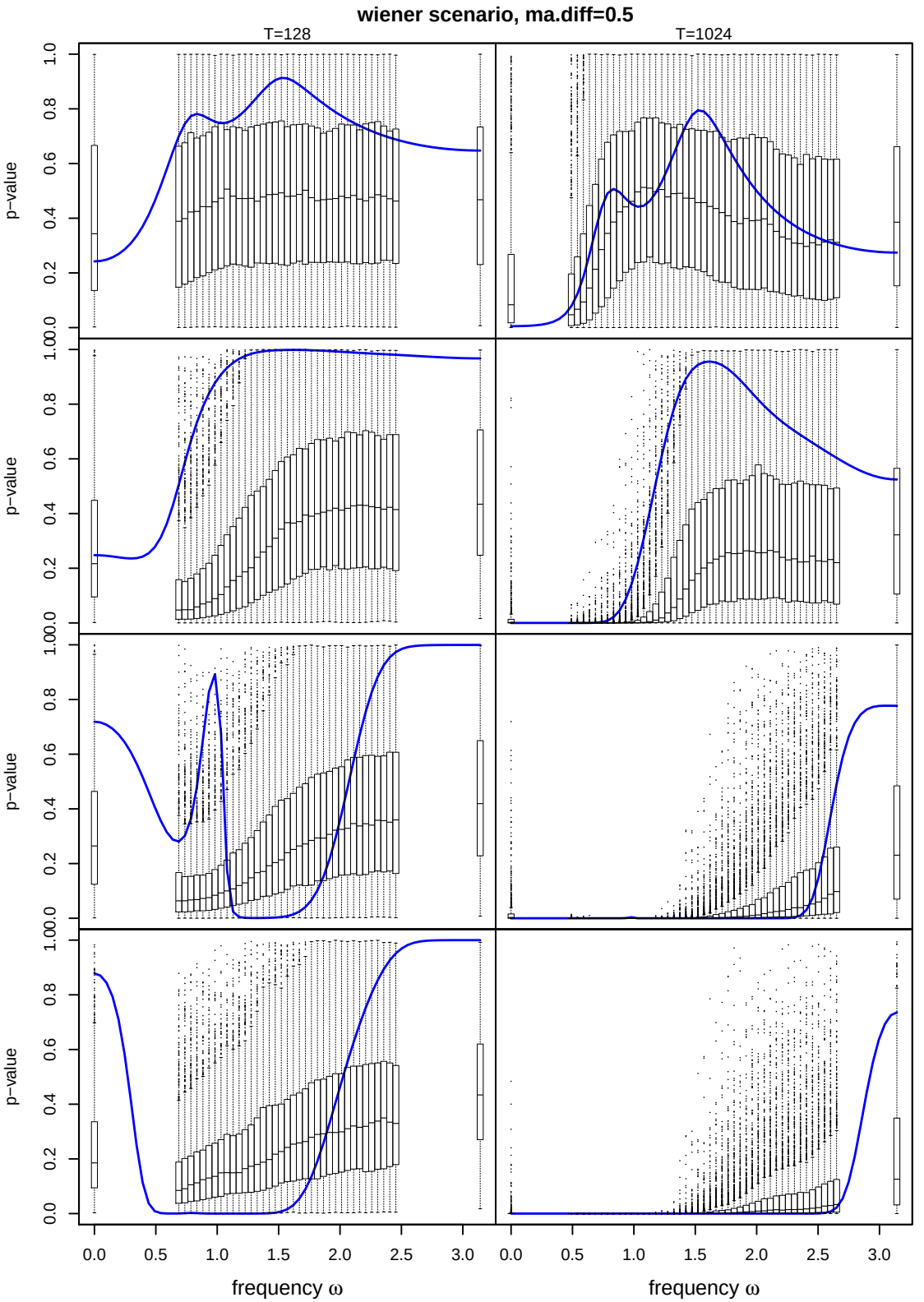


Figure 4.19 – Box plot of the computed p-values $p^b_{j,k}$ for the `Wiener` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

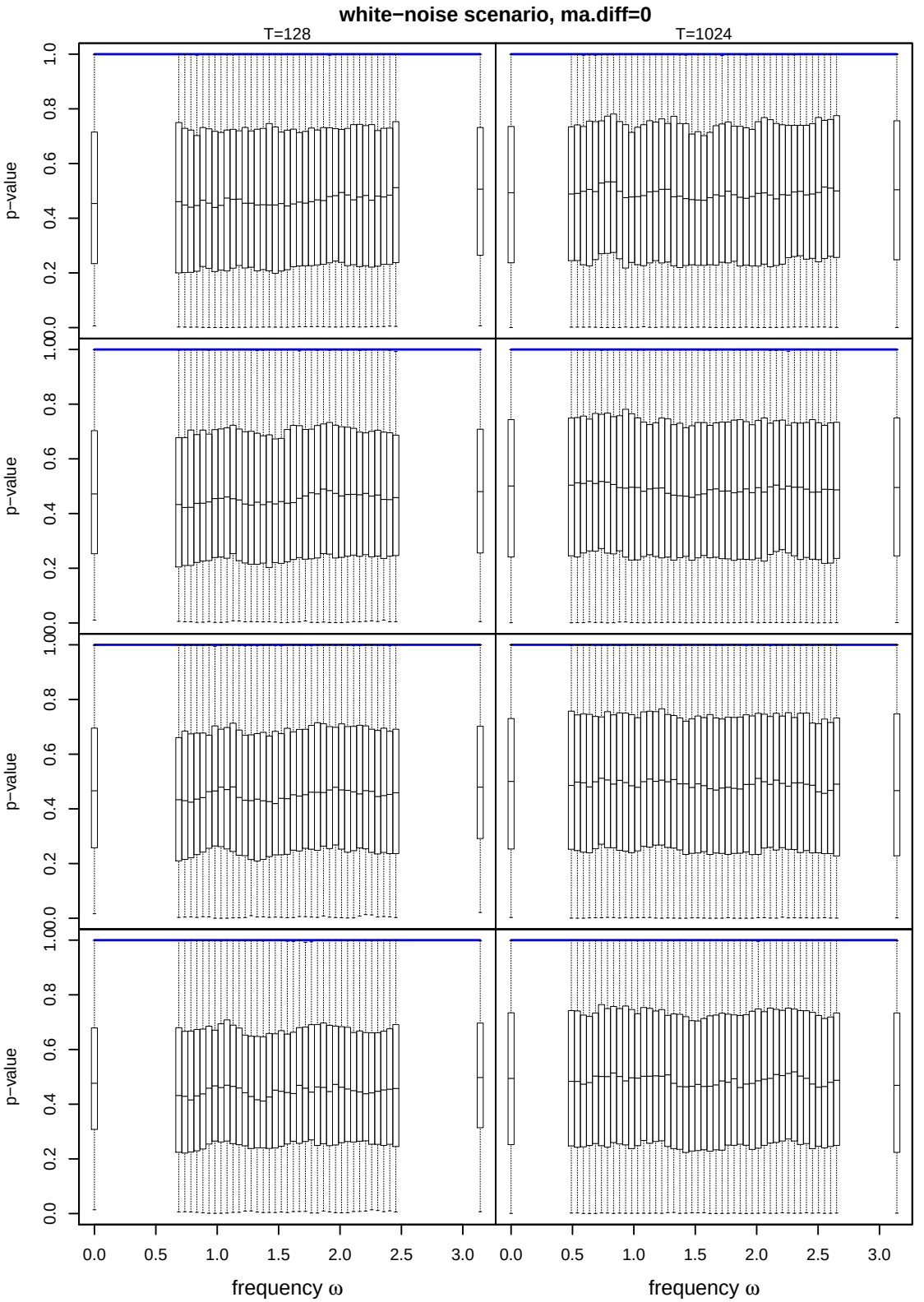


Figure 4.20 – Box plot of the computed p-values $p^b_{j,k}$ for the `White-noise` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

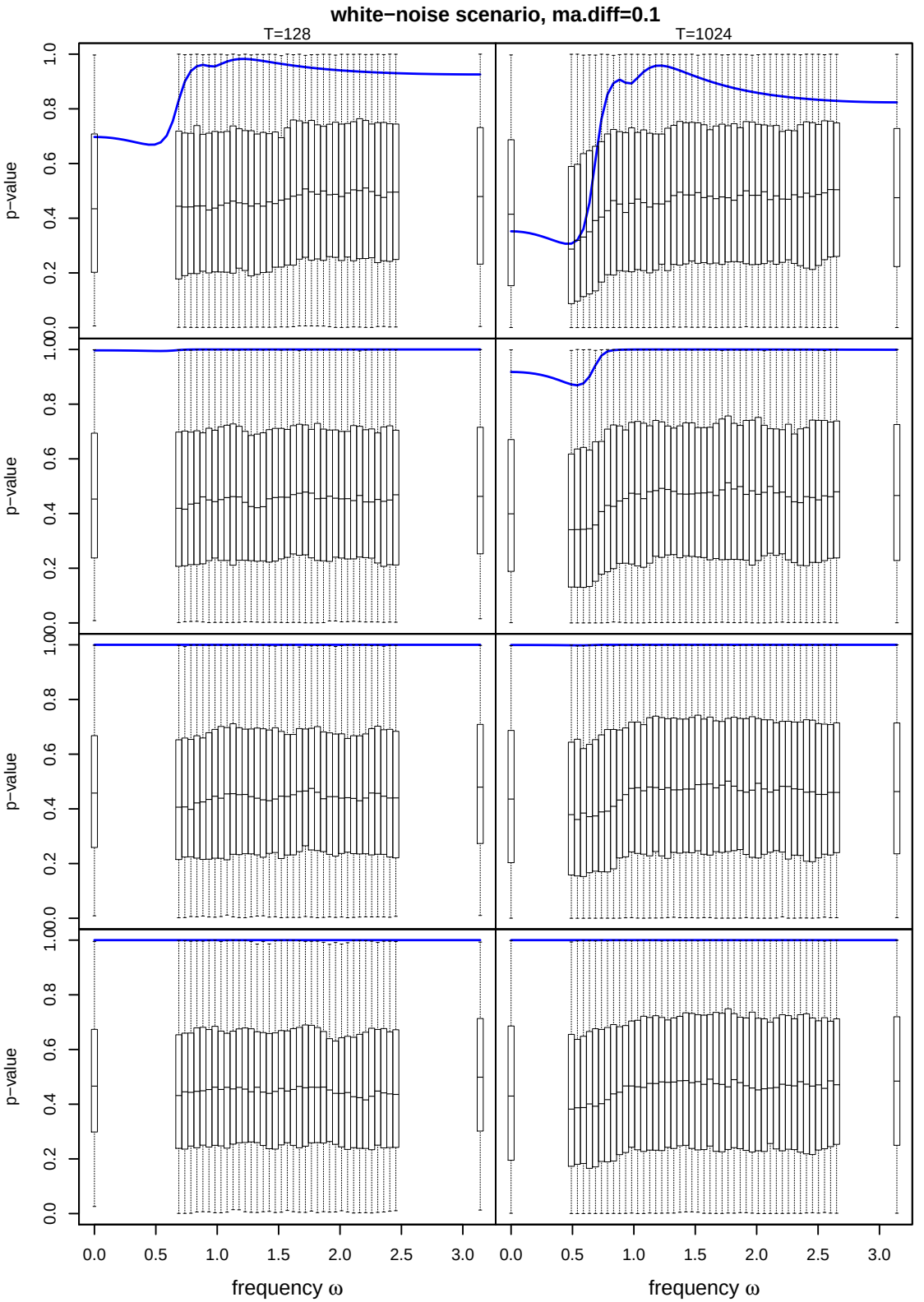


Figure 4.21 – Box plot of the computed p-values $p^b_{j,k}$ for the `White-noise` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

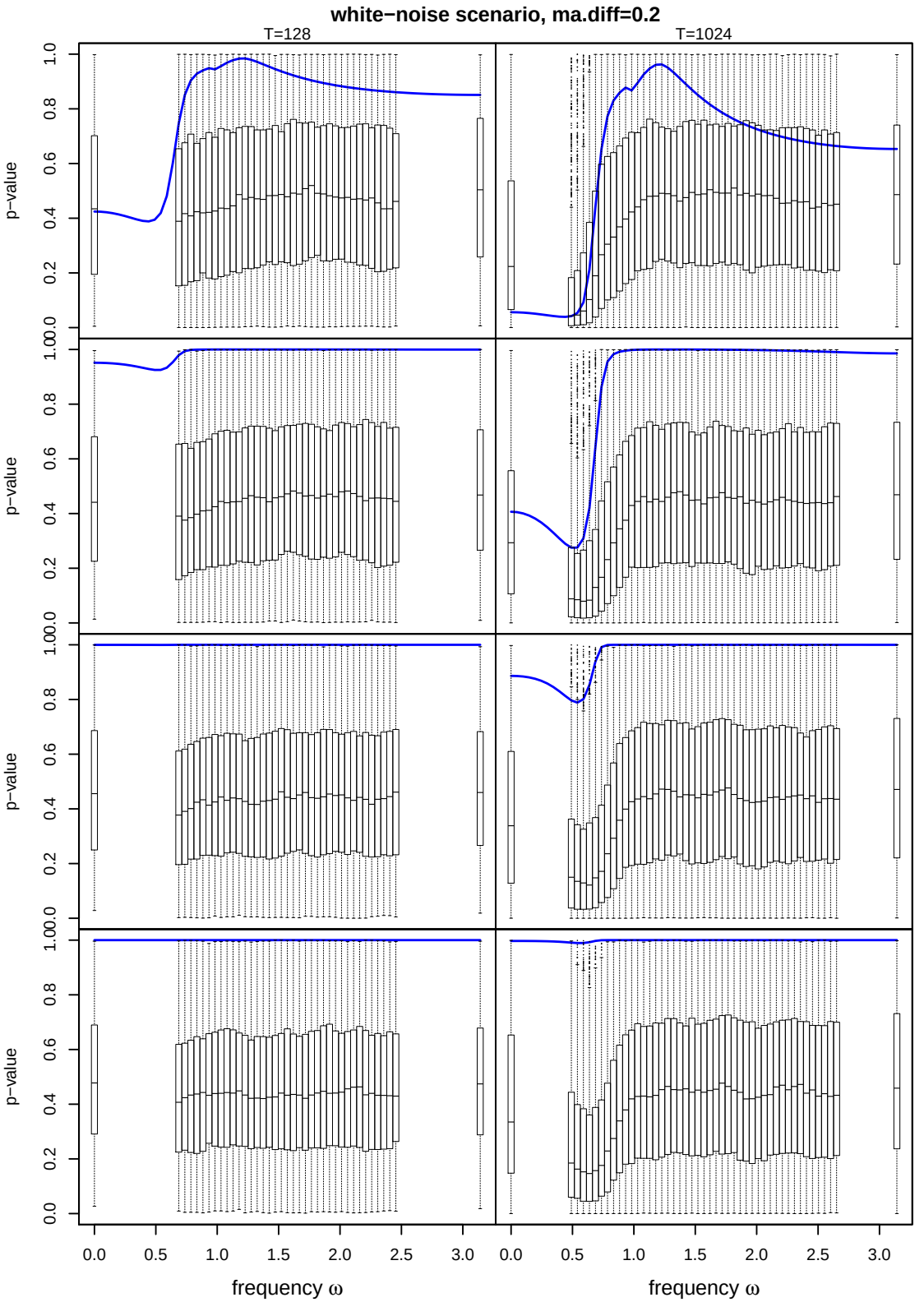


Figure 4.22 – Box plot of the computed p-values $p^b_{j,k}$ for the `White-noise` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

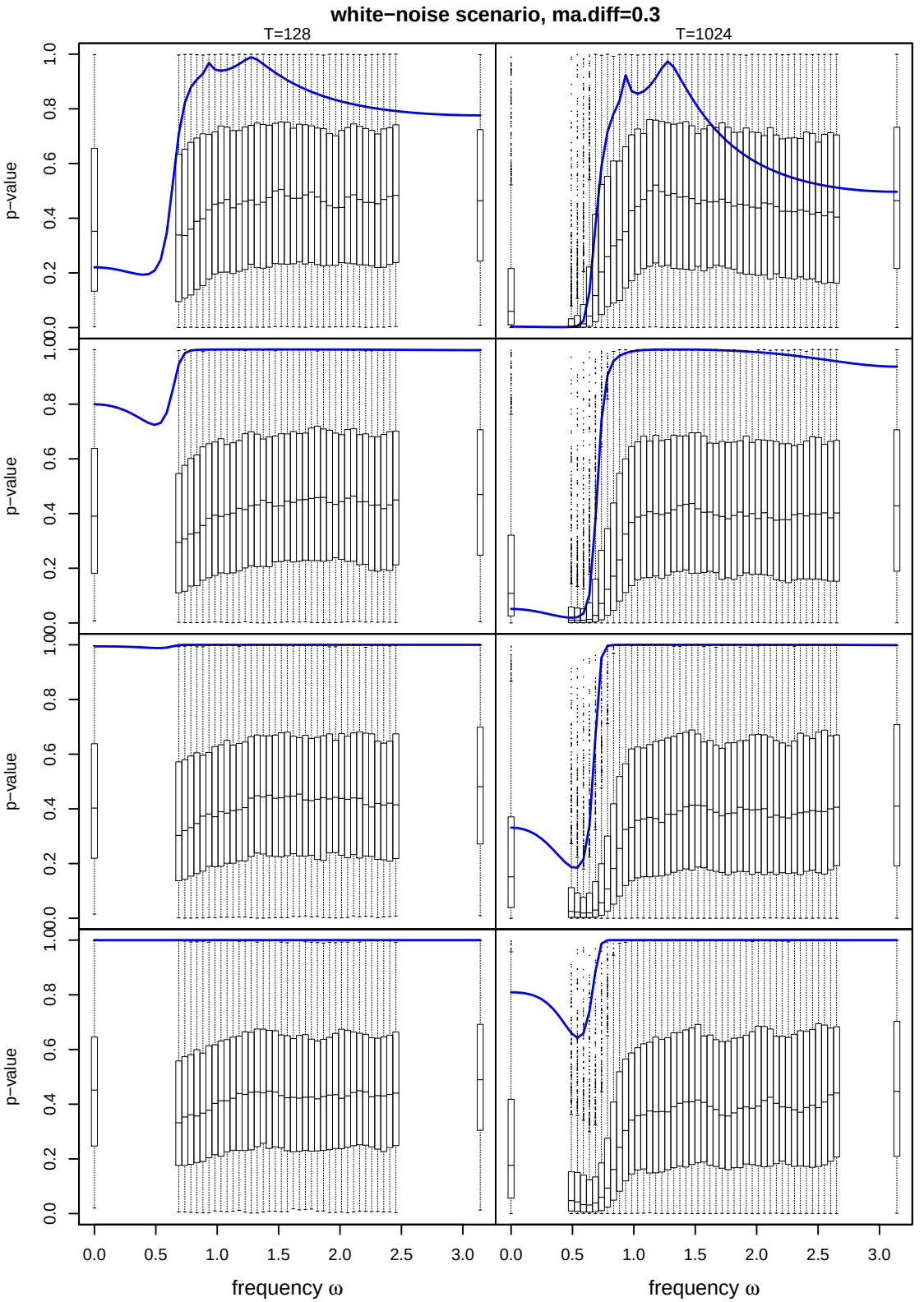


Figure 4.23 – Box plot of the computed p-values $p^b_{j,k}$ for the `White-noise` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

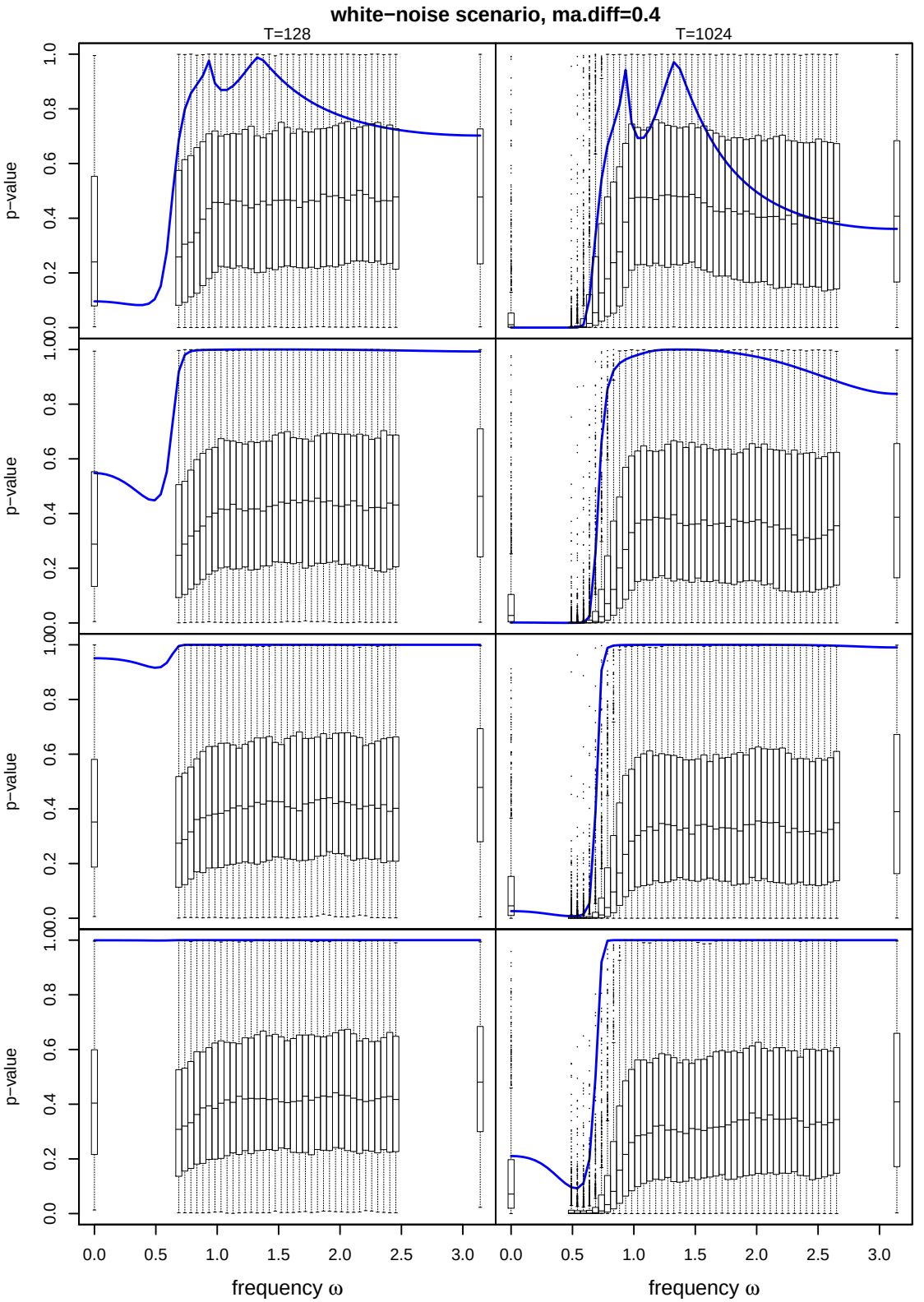


Figure 4.24 – Box plot of the computed p-values $p^b_{j,k}$ for the `White-noise` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

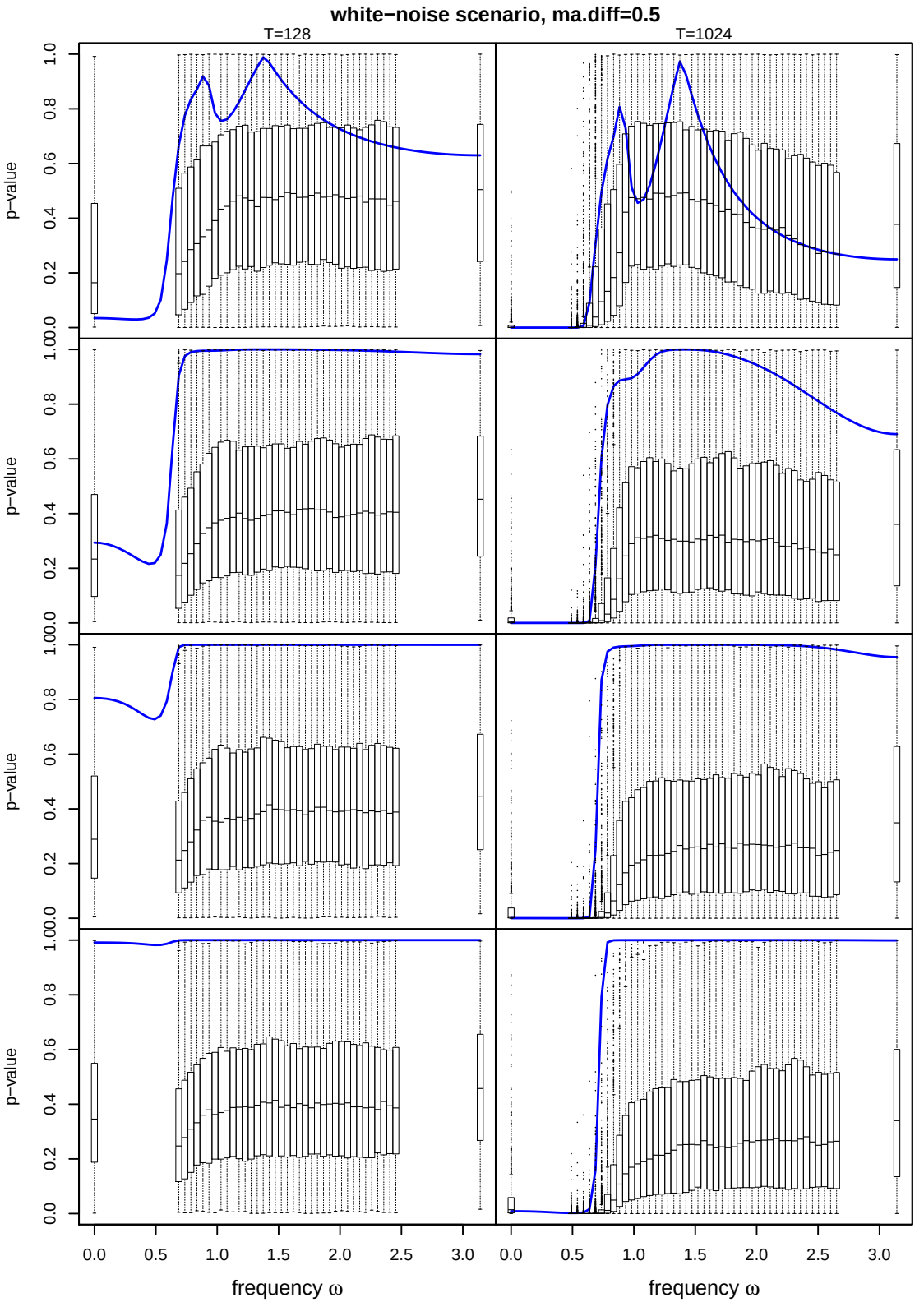


Figure 4.25 – Box plot of the computed p-values $p^b_{j,k}$ for the `White-noise` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

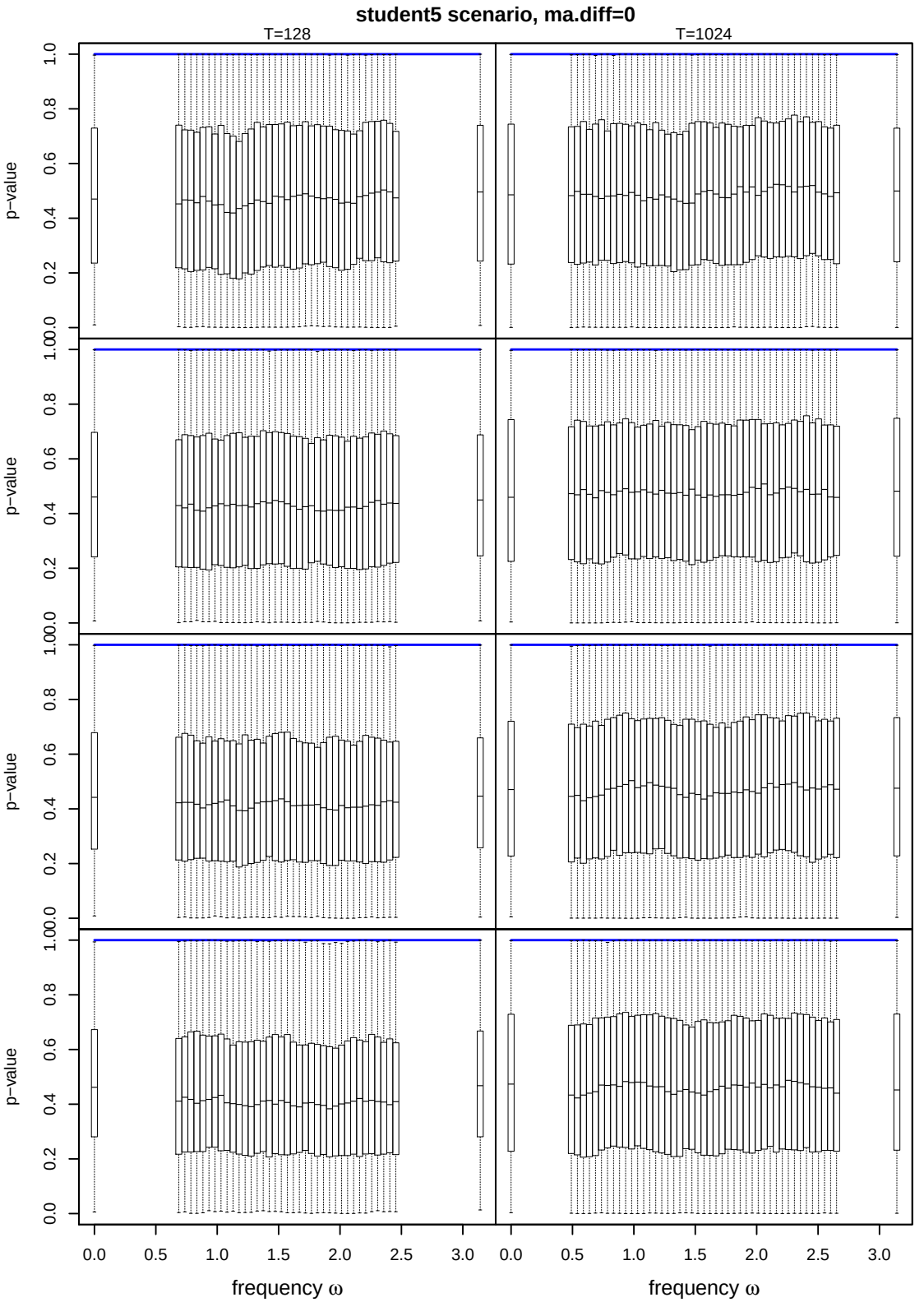


Figure 4.26 – Box plot of the computed p-values $p^b_{j,k}$ for the `Student5` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

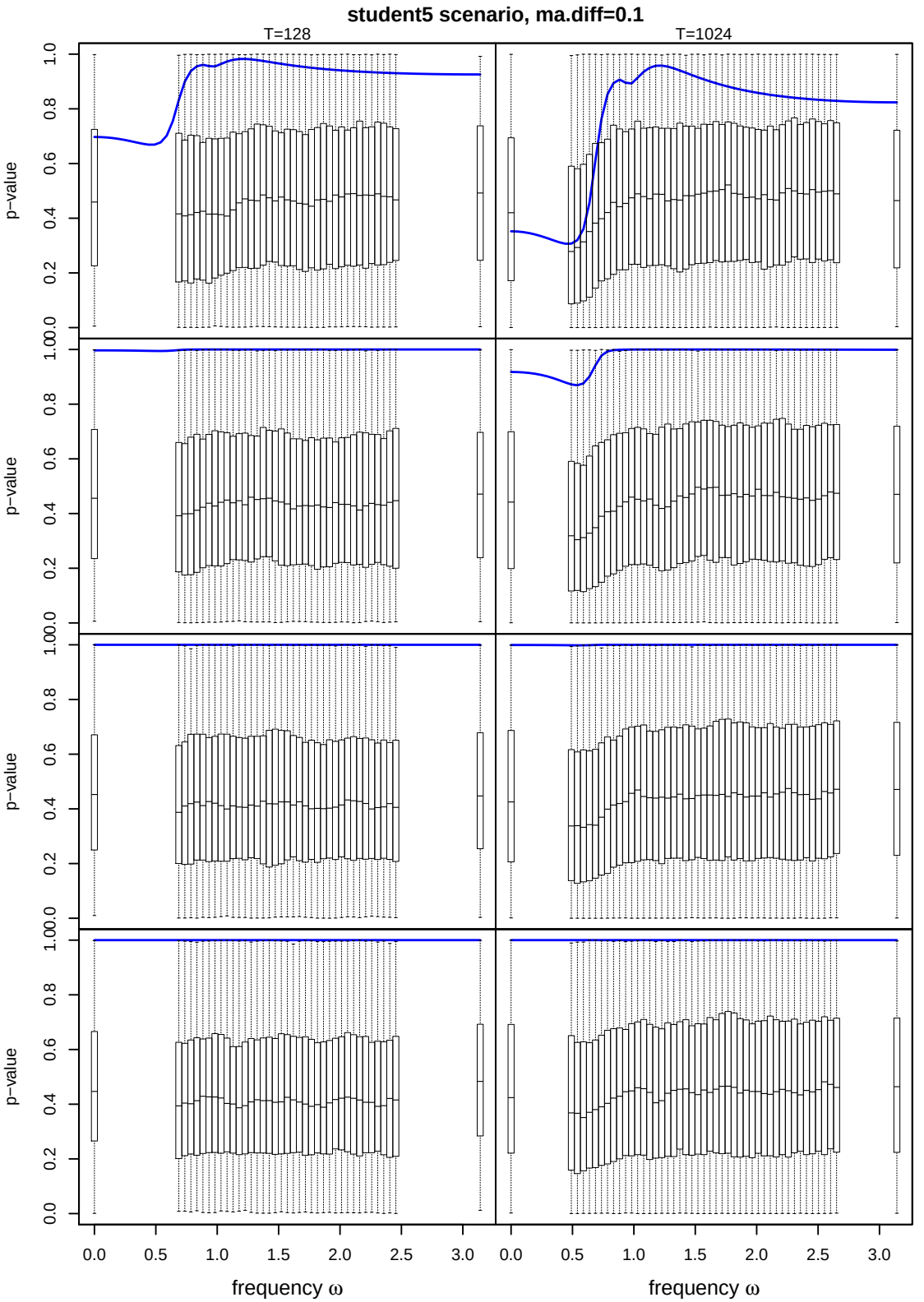


Figure 4.27 – Box plot of the computed p-values $p^b_{j,k}$ for the `Student5` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

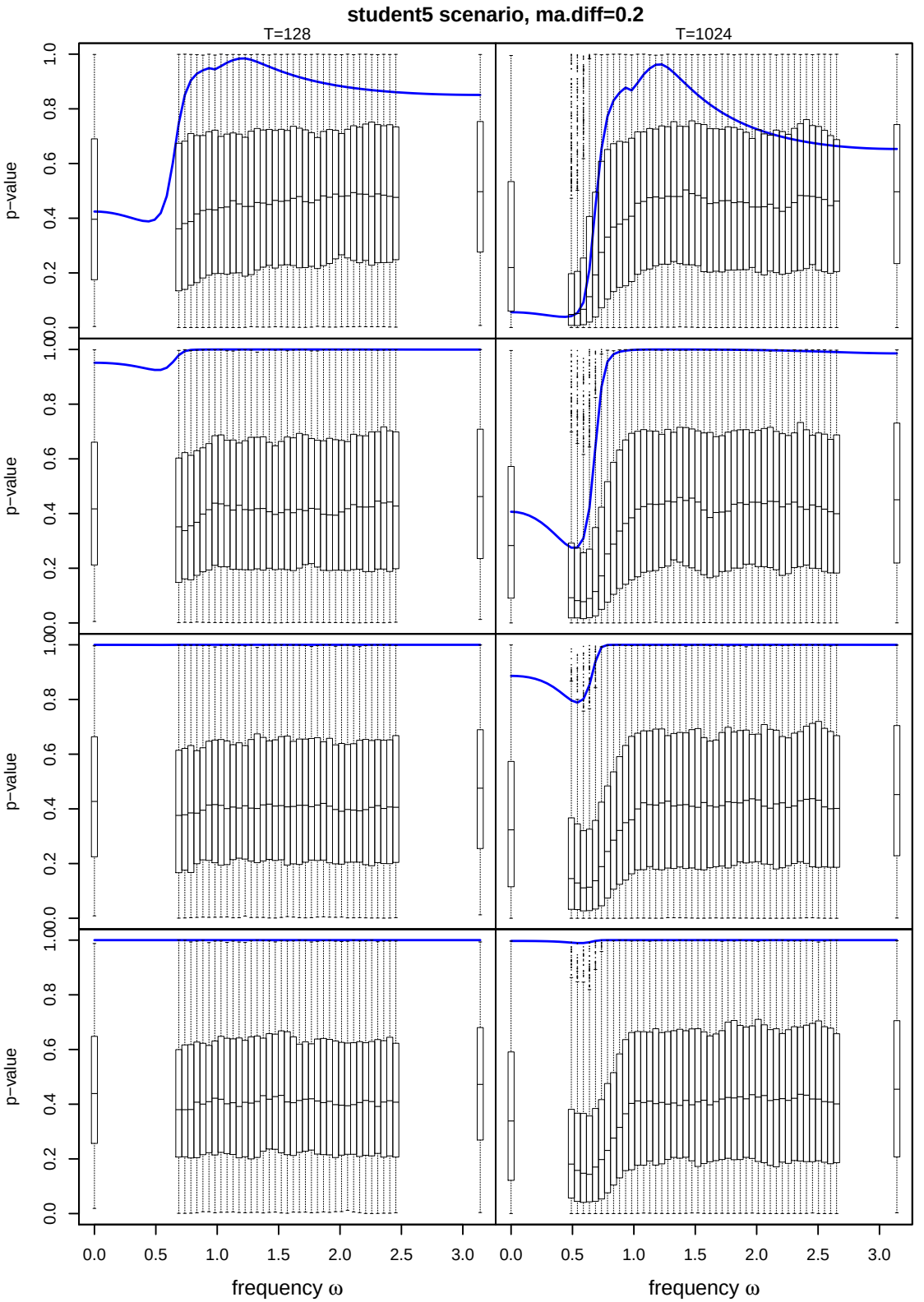


Figure 4.28 – Box plot of the computed p-values $p^b_{j,k}$ for the `Student5` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

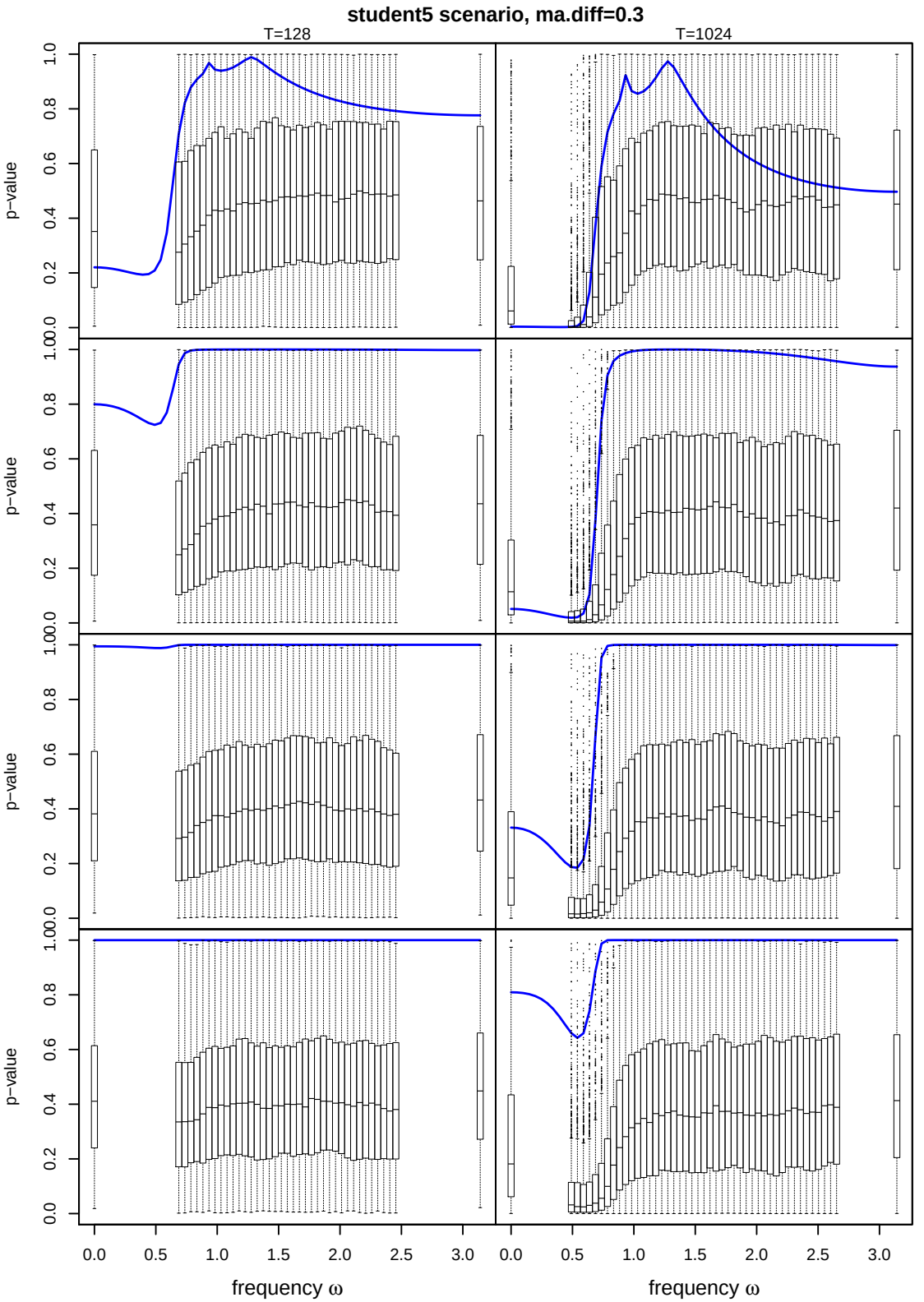


Figure 4.29 – Box plot of the computed p-values $p^b_{j,k}$ for the `Student5` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

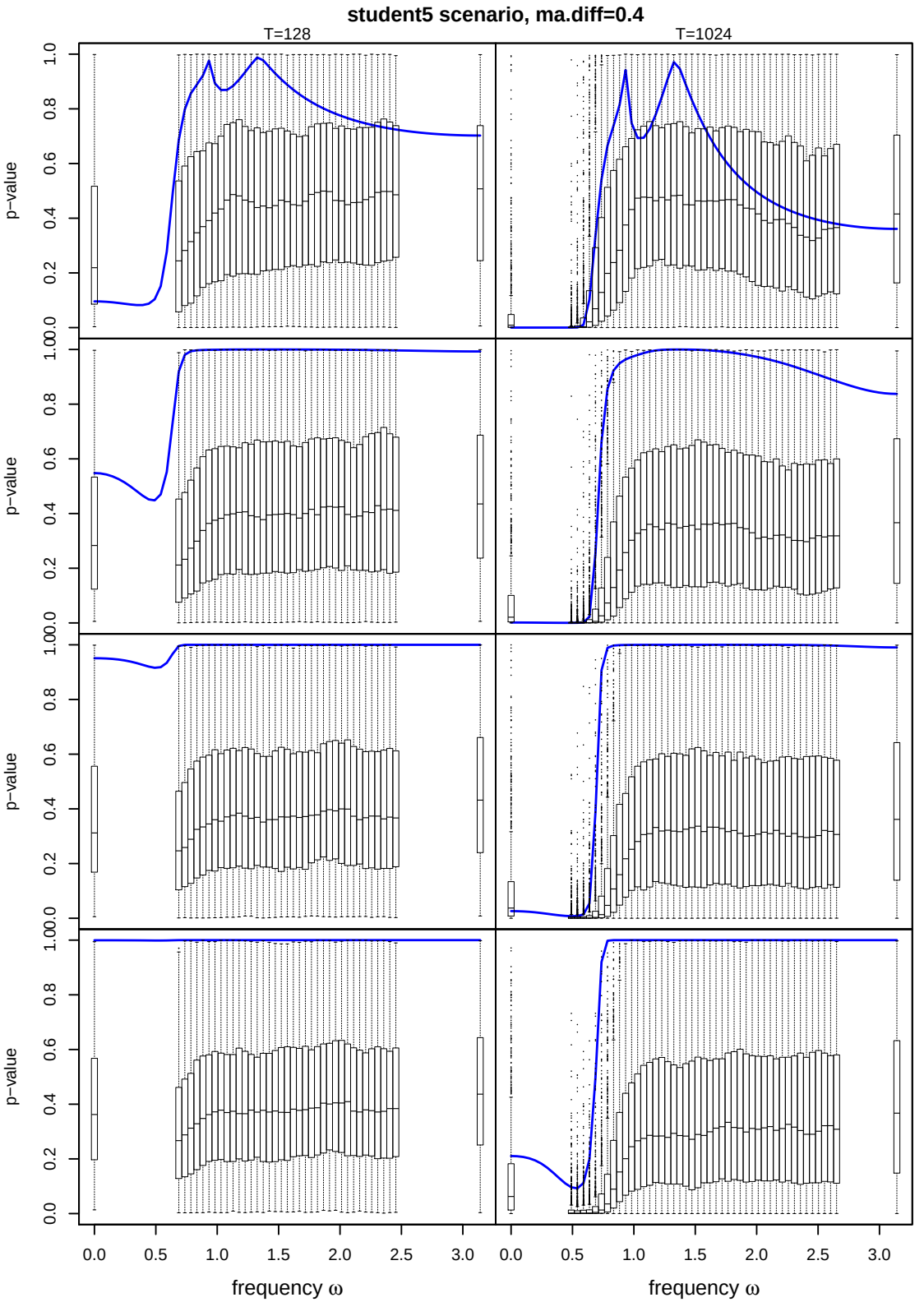


Figure 4.30 – Box plot of the computed p-values $p^b_{j,k}$ for the `Student5` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

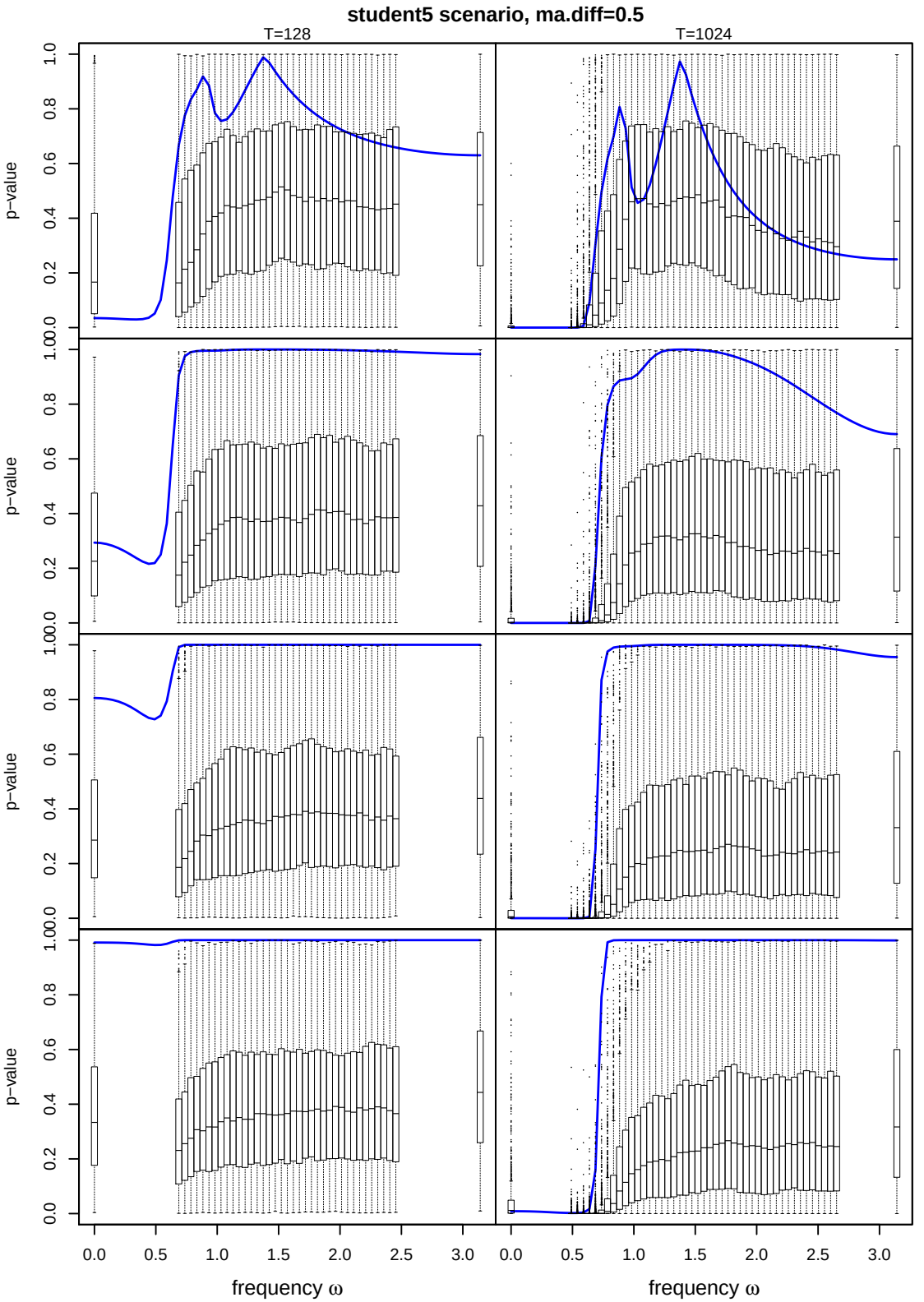


Figure 4.31 – Box plot of the computed p-values $p^b_{j,k}$ for the `Student5` scenario. The x-axis represents the frequencies ω, the left and right columns represent respectively the cases $T = 128$ and $T = 1024$. Each row corresponds to a different value of k, ranging from $k = 1$ (top row) to $k = 4$ (bottom row). From each frequency $\omega_j \in \Gamma$, $j = 1, \ldots, J$, a box plot of the p-values $(p^b_{j,k})_{b=1,\ldots,1000}$ is plotted at $\omega = \omega_j$. The solid curve represents the p-value that would have been observed if we replaced the sample version of the spectral density operator and its eigenstructure by their true counterparts.

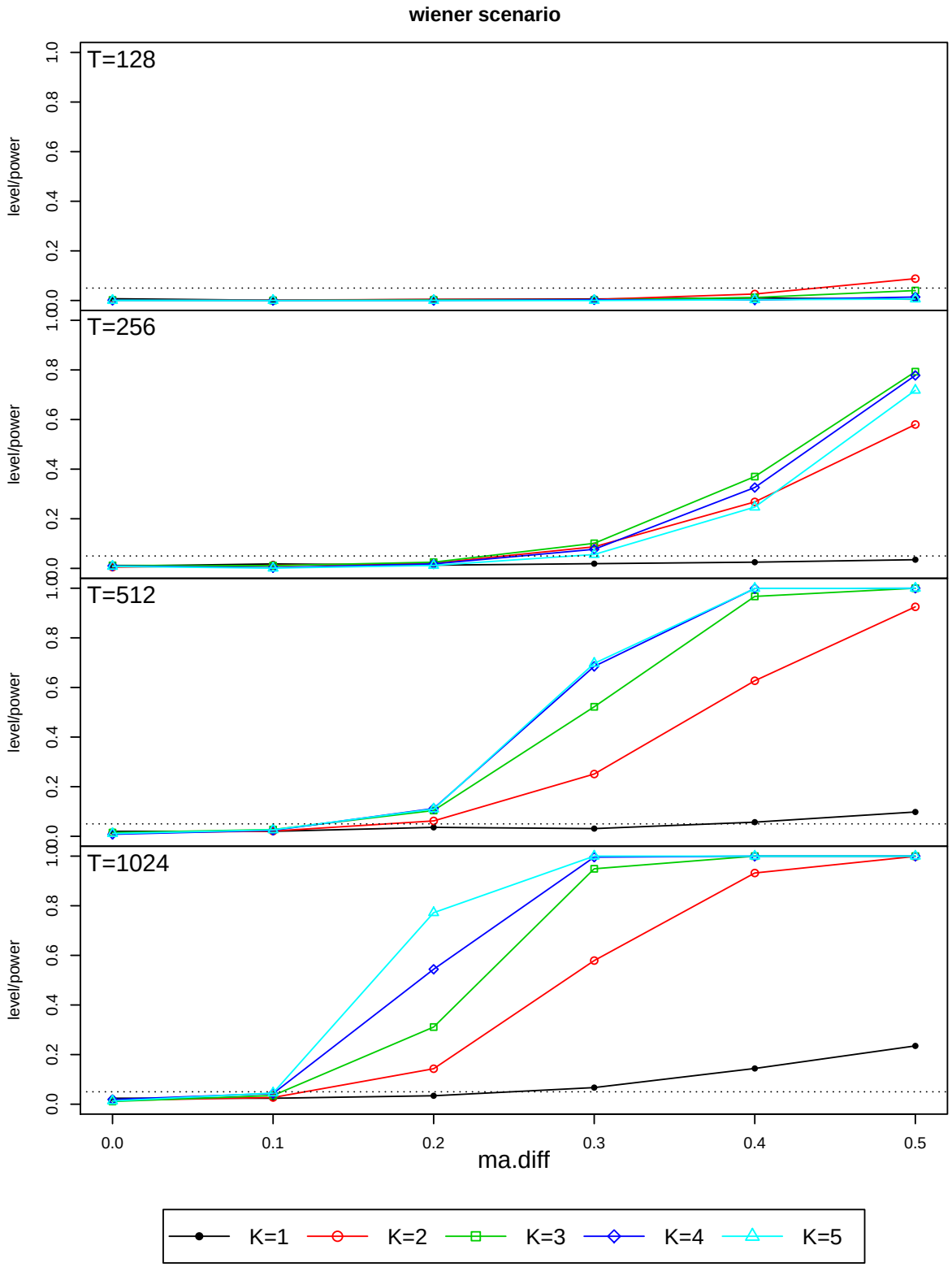


Figure 4.32 – The estimated probability of rejection of the global null hypothesis at level 5% in the `Wiener` scenario for each depth $k \in \{1,\dots,5\}$, see (4.3.13). The x-axis represents the coefficient `ma.diff`, and each plot corresponds to a different sample size, ranging from $T = 128$ (top) to $T = 1024$ (bottom). The horizontal dotted line corresponds to the 5% level, and the standard deviation, based on a normal approximation, is at most 0.016.

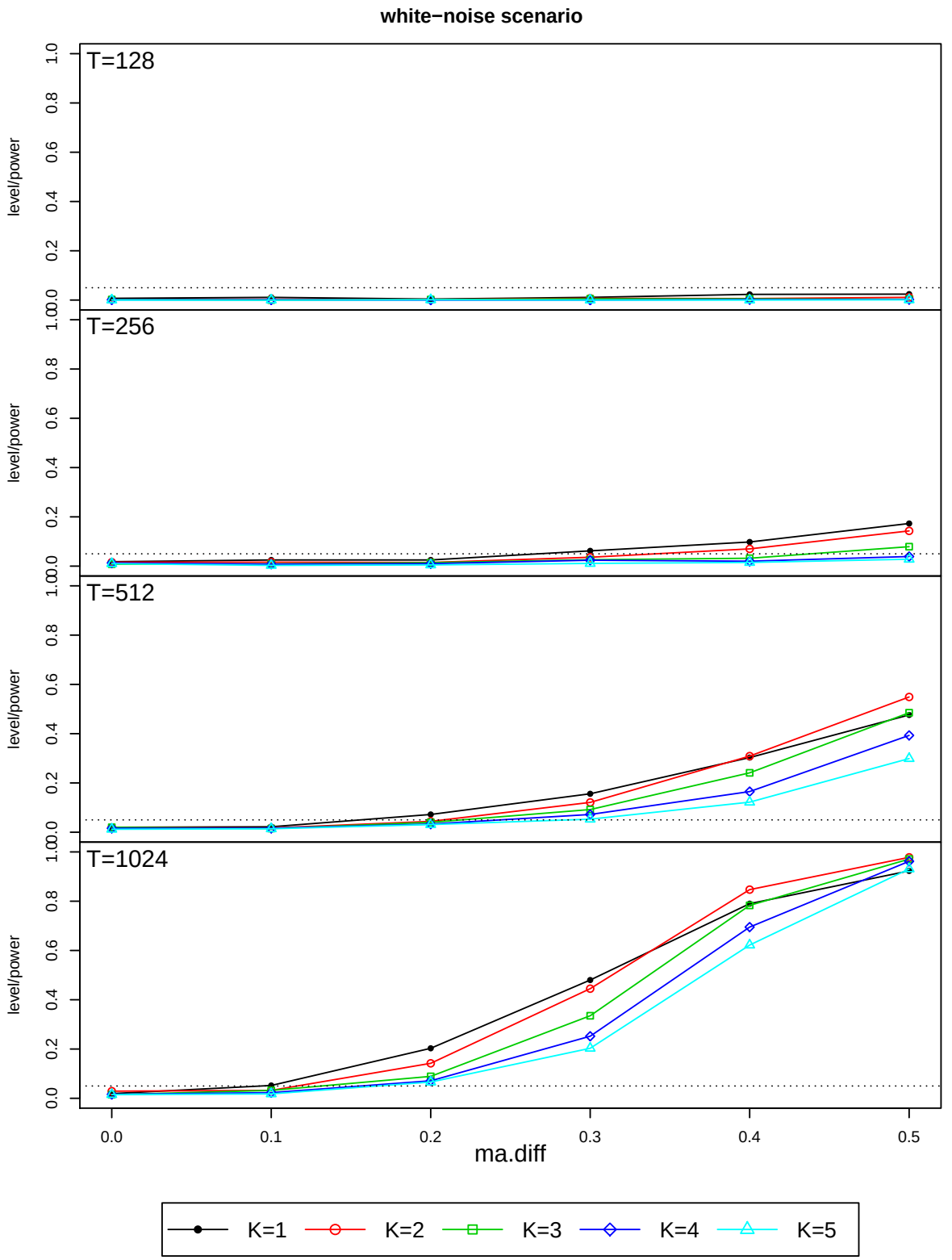


Figure 4.33 – The estimated probability of rejection of the global null hypothesis at level 5% in the `White-noise` scenario for each depth $k \in \{1,\dots,5\}$, see (4.3.13). The x-axis represents the coefficient `ma.diff`, and each plot corresponds to a different sample size, ranging from $T = 128$ (top) to $T = 1024$ (bottom). The horizontal dotted line corresponds to the 5% level, and the standard deviation, based on a normal approximation, is at most 0.016.

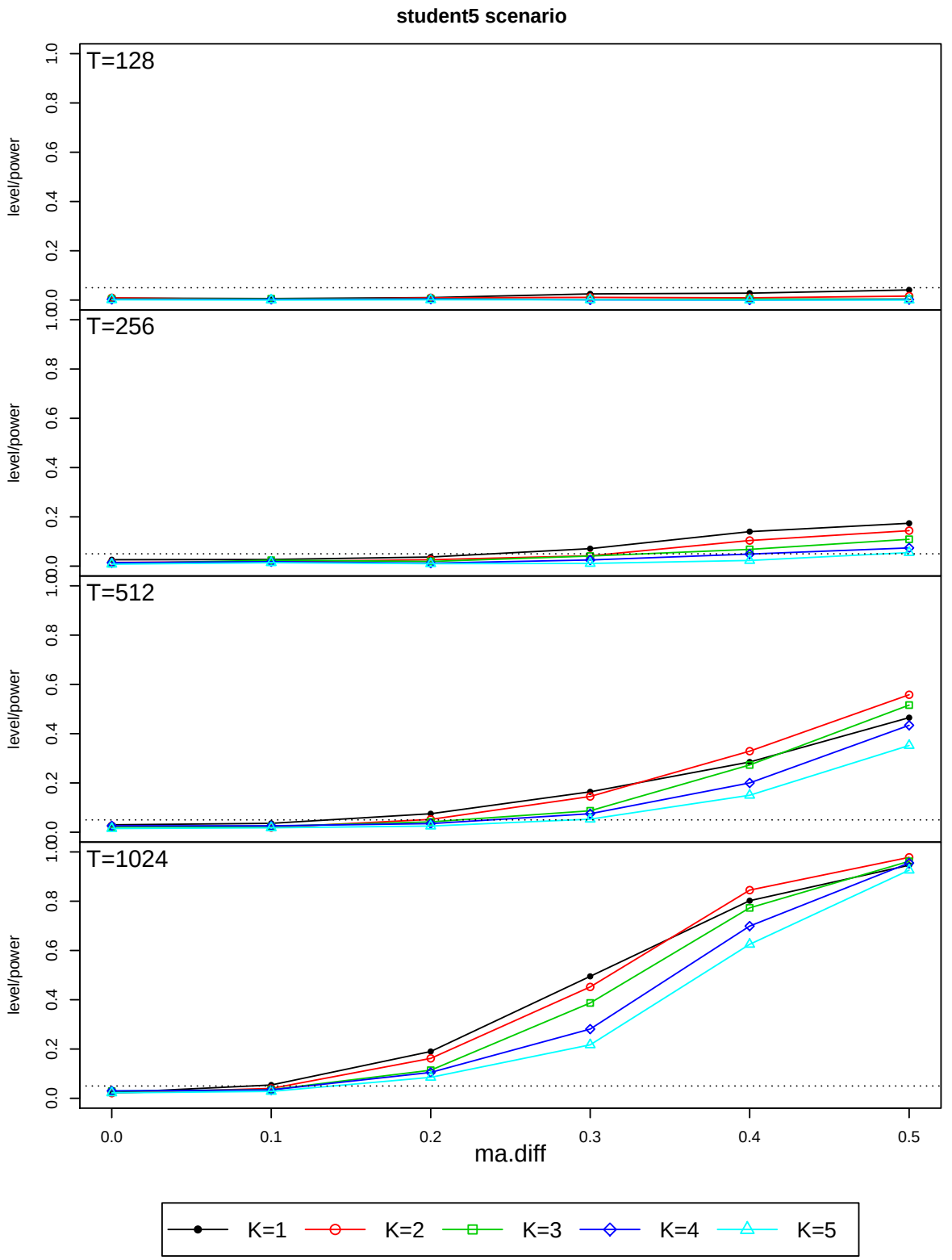


Figure 4.34 – The estimated probability of rejection of the global null hypothesis at level 5% in the `Student5` scenario for each depth $k \in \{1,\dots,5\}$, see (4.3.13). The x-axis represents the coefficient `ma.diff`, and each plot corresponds to a different sample size, ranging from $T = 128$ (top) to $T = 1024$ (bottom). The horizontal dotted line corresponds to the 5% level, and the standard deviation, based on a normal approximation, is at most 0.016.

APPENDIX A

Selected results for Hilbert and Banach Spaces

We first present Banach spaces, which are (possibly infinite) dimensional spaces with a topology given by a norm. Some references are (Dunford & Schwartz 1988*a*, Rudin 1991, Ciarlet 2013)

A.1 Banach Spaces

A (complex) *Banach space* is a couple $(B, \|\cdot\|)$, where B is a vector space, $\|\cdot\| : B \to \mathbb{C}$ is a norm, under which the space B is complete (every Cauchy sequence converges to an element in B).
Let $(B_1, \|\cdot\|_1)$ and $(B_2, \|\cdot\|_2)$ be complex Banach spaces. A function $T : B_1 \to B_2$ is called a *bounded linear operator*, or just *bounded operator*, if it is linear and there exists a positive constant C such that $\|Tx\|_2 \leq C\|x\|_1$ for all $x \in B_1$. The smallest such C is

$$|||T|||_\infty := \sup_{x \neq 0} \frac{\|Tx\|}{\|x\|}.$$

We denote the set of bounded operators $T : B_1 \to B_2$ by $\mathscr{S}_\infty(B_1, B_2)$. The special case $\mathscr{S}_\infty(B_1, \mathbb{C})$ is of particular interest, and is called the *dual*, or the *topological dual*, of B_1. It is denoted by B_1^*, and consists of all bounded linear *functionals* on B_1.
$\mathscr{S}_\infty(B_1, B_2)$ is actually a Banach space if we define $(\lambda T)x = \lambda(Tx)$ and $(S+T)x = (Sx) + (Tx)$ for all $x \in B_1$, $\lambda \in \mathbb{C}$ and $S, T \in \mathscr{S}_\infty(B_1, B_2)$.
If $B_1 = B_2 = B$, we use the shorthand notation $\mathscr{S}_\infty(B) = \mathscr{S}_\infty(B, B)$. For

$S, T \in \mathscr{S}_\infty(B)$, the product ST defined by $STx = S(Tx), x \in B$ is a bounded operator, with $|||ST|||_\infty \leq |||S|||_\infty |||T|||_\infty$, and therefore $ST \in \mathscr{S}_\infty(B)$. In particular, $\mathscr{S}_\infty(B)$ is an algebra over $\mathbb{C}$.

A.1.1 Unique Linear Extension

Proposition A.1.1 (Ciarlet (2013, Theorem 3.1-1))**.** *Let B_1, B_2 be Banach spaces, and $W \subset B_1$ be a dense subspace. Any continuous linear operator $T : W \to B_2$ admits a unique continuous linear extension $T' : B_1 \to B_2$, defined by*

$$T'(x) = \lim_{n\to\infty} T(x_n) \tag{A.1.1}$$

where $(x_n)_{n\geq 1} \subset W$ is any sequence converging to $x \in B_1$. Furthermore, T' is continuous, and

$$|||T'|||_\infty = |||T|||_\infty \tag{A.1.2}$$

A.1.2 Bounded Multilinear Mappings

Let $B_1, \ldots, B_k, B$ be complex Banach spaces. A mapping $T : B_1 \times B_k \to B$ is called a bounded (or continuous) multilinear mapping if it is linear in each coordinate, and if there exists a constant $c \geq 0$ such that

$$\|T(x_1, \ldots, x_k)\| \leq c\|x_1\| \cdots \|x_k\|, \quad x_i \in B_i, i = 1, \ldots, k. \tag{A.1.3}$$

The infimum of all $c \geq 0$ satisfying (A.1.3) is denoted $|||T|||_\infty$, and satisfies

$$|||T|||_\infty = \sup\{\|T(x_1, \ldots, x_k)\| : \|x_1\|, \ldots, \|x_k\| \leq 1\}.$$

The space of all multilinear bounded mappings $T : B_1 \times B_k \to B$ is denoted by $\mathscr{L}_k(B_1, \ldots, B_k; B)$, and is a Banach space when equipped with the norm $|||\cdot|||_\infty$. Furthermore there are canonical linear bijective isometries

$$\mathscr{L}_k(B_1, \ldots, B_k; B) \to \mathscr{L}(B_1; \mathscr{L}(B_2; \mathscr{L}(\ldots; B))) \tag{A.1.4}$$

A.2 Hilbert spaces

We now talk about Hilbert spaces, which are special kinds of Banach spaces. Some references are Gohberg & Krejn (1971), Ringrose (1971), Dunford & Schwartz (1988*a*,*b*), Gohberg et al. (1990), Kadison & Ringrose (1997), Zhu (2007).
A *real Hilbert space* is a couple $(H, \langle\cdot,\cdot\rangle)$ where

1. H is a real vector space,
2. $\langle\cdot,\cdot\rangle$ is a function $H \times H \to \mathbb{R}$ satisfying for all $u, v, w \in H$ and $\lambda \in \mathbb{R}$,
   (a) $\langle u, u\rangle \geq 0$ and $\langle u, u\rangle = 0 \iff u = 0$
   (b) $\langle u, v\rangle = \langle v, u\rangle$

(c) $\langle u + \lambda v, w\rangle = \langle u, w\rangle + \lambda\langle v, w\rangle$

3. H is complete metric space under the norm $\|u\| := \sqrt{\langle u, u\rangle}$.

A *complex Hilbert space* $(H, \langle\cdot,\cdot\rangle)$ is slightly different because of richer structure of the complex field $\mathbb{C}$:

1. H is a vector space,

2. $\langle\cdot,\cdot\rangle$ is a function $H \times H \to \mathbb{C}$ satisfying for all $u, v, w \in H$ and $\lambda \in \mathbb{C}$,

   (a) $\langle u, u\rangle \geq 0$ and $\langle u, u\rangle = 0 \iff u = 0$

   (b) $\langle u, v\rangle = \overline{\langle v, u\rangle}$

   (c) $\langle u + \lambda v, w\rangle = \langle u, w\rangle + \lambda\langle v, w\rangle$

3. H is complete metric space under the norm $\|u\| := \sqrt{\langle u, u\rangle}$.

Here $\overline{\alpha}$ denote the conjugate transpose of $\alpha \in \mathbb{C}$. The function $\langle\cdot,\cdot\rangle$ is called the *inner product,* or *scalar product* of H.
A Hilbert space is called *separable* if it contains a dense countable subset. Every separable Hilbert space admits a (non-unique) orthonormal basis $(e_n)_{n\geq 1}$ which satisfies $\langle e_n, e_m\rangle = \delta_{n,m}$, and

$$u = \sum_{n\geq 1} \langle u, e_n\rangle e_n, \quad \forall u \in H.$$

From now on, H will denote a separable Hilbert space.

A.2.1 Operators on Hilbert Spaces

Let H be a (complex) separable Hilbert space with scalar product $\langle\cdot,\cdot\rangle$ and norm $\|\cdot\|$. Since a Hilbert space is particular type of Banach space, the definition of bounded operator (and its norm) extends naturally to Hilbert spaces. We will denote Id_H the identity operator on H. If a bounded operator $T : H \to H'$ between two Hilbert spaces preserves the inner-product, i.e.

$$\langle Tx, Ty\rangle = \langle x, y\rangle, \quad \forall x, y \in H, \tag{A.2.1}$$

then it is called an *isometry*. Notice that an isometry is necessarily injective, and therefore allows one to view the Hilbert space H as a subspace of H' (with same norm) via the identification $H \cong T(H)$. An operator $T : H \to H'$ that admits an inverse $S : H' \to H$ (i.e. $TS = \mathrm{Id}_{H'}$ and $ST = \mathrm{Id}_H$) is called an *isomorphism*. If it is also an isometry, it is called an *isometric isomorphism*, or a *unitary operator*.
To any bounded operator T on H corresponds another bounded operator $T^\dagger$, called the *adjoint* of T, and which satisfies the property

$$\langle Tx, y\rangle = \left\langle x, T^\dagger y\right\rangle, \quad \forall x, y \in H.$$

In particular, $|||T^\dagger|||_\infty = |||T|||_\infty$. An operator which satisfies $T = T^\dagger$ is called *self-adjoint*, or *hermitian*, and

$$T \text{ hermitian} \implies |||T|||_\infty = \sup\{|\langle Tx, x\rangle| : \|x\| = 1\}.$$

The space of bounded operators is very large. A important subspace of it are the compact operators.

Definition A.2.1. *A bounded operator $T \in \mathscr{S}_\infty(H)$ is* compact *if every bounded sequence $\{x_n : n \geq 1\}$ admits a subsequence $\{x_{n_k} : k \geq 1\}$ such that Tx_{n_k} converges in H.*

Compactness of an operator is a strong condition: for instance, the identity operator is *not* compact. However, compact operators have nice properties. Let us denote by $\mathscr{S}_c(H)$ the set of compact operators. This set is actually a closed subspace of $\mathscr{S}_\infty(H)$, and admits the following characterization:

Proposition A.2.2 (Singular Value Decomposition)**.** *A bounded operator T on H is compact if, and only if there exists orthonormal bases $(e_n : n \geq 1)$ and $(f_n : n \geq 1)$ of H and a sequence of decreasing positive numbers $(\lambda_n : n \geq 1) \subset [0, +\infty)$ with $\lambda_n \to 0$ such that*

$$Tx = \sum_{n\geq 1} \lambda_n \langle x, f_n\rangle e_n, \quad x \in H, \tag{A.2.2}$$

The λ_ns are called the singular values *of T, and we write*

$$\text{sing}(T) = (\lambda_n : n \geq 1).$$

We point out that the sequence $(\lambda_n : n \geq 1)$ does not need to be strictly decreasing. The following Theorem gives an important characteristic of the singular values of a compact operator:

Theorem A.2.3 (Gohberg et al. (1990, Chapter VI, Theorem 3.1))**.** *Let T be a compact operator on H, with singular values $\lambda_1(T) \geq \lambda_2(T) \geq \cdots \geq 0$. Then for any $n = 1, 2, \ldots \leq dim H$, we have*

$$\sum_{j=1}^{n} \lambda_j(T) = \max_{U, \psi_1, \ldots, \psi_n} \left| \sum_{j=1}^{n} \langle UT\psi_j, \psi_j\rangle \right|,$$

where the maximum is taken over all unitary operators U on H, and all orthonormal systems $\{\psi_1, \ldots, \psi_n\} \subset H$.

The following Lemma tells us about the behaviour of the singular values of compact operators under compact perturbations

Lemma A.2.4 (Gohberg & Krejn (1971, Corollaire 2.3, p.31), Bosq (2000, Lemma 4.2), Gohberg et al. (1990, Chapter VI, Corollary 1.6))**.**

Let T, S be compact operators on H, with singular value decompositions

$$T = \sum_{n\geq 1} \lambda_n \langle ., f_n \rangle e_n,$$
$$S = \sum_{n\geq 1} \mu_n \langle ., v_n \rangle u_n.$$

Then,

$$|\lambda_n - \mu_n| \leq |||T - S|||_\infty, \quad \forall n \geq 1. \tag{A.2.3}$$

Notice that the singular values of either S or T need not be strictly decreasing.

If $(T_m)_{m\geq 1}$ is a sequence of compact operators on H converging to $T \in \mathscr{S}_c(H)$ in the operator norm, this lemma tells us the singular values of T_m (arranged in decreasing order, and repeated according to their multiplicity) converge to the singular values of T (also arranged in decreasing order) uniformly in n as $m \to \infty$. Now if $(T_m)_{m\geq 1}$ and T are also self-adjoint, there is a similar perturbation result for their eigenspaces. We need to introduce some notation: Let $u \otimes_2 v$ denote the operator on $\mathscr{S}_\infty(H)$ defined by $(u \otimes_2 v) f = \langle f, v \rangle u$, for $u, v, f \in H$. Using this notation, we write the singular value decompositions of T_m and T:

$$T = \sum_{n\geq 1} \mu_n e_n \otimes_2 e_n$$
$$T_m = \sum_{n\geq 1} \mu_{n,m} e_{n,m} \otimes_2 e_{n,m}$$

$\{\mu_n : n \geq 1\}$ is a non-increasing positive sequences tending to zero. We denote by $\{\lambda_n\}_{n\geq 1}$ the decreasing sequence of distinct elements of $\{\mu_n\}_{n\geq 1}$, and define the set

$$I_k = \{i \geq 1 : \mu_i = \lambda_k\},$$

and we denote its cardinality by $m_k = |I_k|$. We will also write

$$I = \{i \geq 1 : \mu_i > 0\} = \bigcup_{k\geq 1 \& \lambda_k > 0} I_k \tag{A.2.4}$$

for the set of indices of the repeated non-zero eigenvalues of T, and

$$J = \{j \geq 1 : \lambda_j > 0\}, \tag{A.2.5}$$

the set of indices of the non-repeated non-zero eigenvalues of T. We define

$$\Pi_k = \sum_{i \in I_k} e_i \otimes_2 e_i,$$

which is the projection onto the k^{th} eigenspace of T, also called the *k-th*

eigenprojector of T. This way, we have

$$T = \sum_{j\in J} \lambda_j \Pi_j,$$
$$T_m = \sum_{j\in J} \left(\sum_{i\in I_j} \mu_{i,m} e_{i,m} \otimes_2 e_{i,m} \right)$$

It turns out that the term in parentheses converges to $\lambda_j \Pi_j$ as $m \to \infty$. This is rigorously stated in the following perturbation result for the eigenspaces of compact self-adjoint operators.

Theorem A.2.5. *Let $\{T_m\}_{m\geq 1}$ be a sequence of compact self-adjoint operators on H converging to the compact operator T in the operator norm $|||\cdot|||_\infty$. Then, using the notation introduced above, we have*

$$\left|\left|\left| \left(\sum_{i\in I_k} e_{i,m} \otimes_2 e_{i,m} \right) - \Pi_k \right|\right|\right|_\infty \to 0, \qquad m \to \infty,$$
$$\left|\left|\left| \left(\sum_{i\in I_k} \mu_{i,m} e_{i,m} \otimes_2 e_{i,m} \right) - \lambda_k \Pi_k \right|\right|\right|_\infty \to 0, \qquad m \to \infty.$$

Proof. The result is a consequence of Riesz & Sz.-Nagy (1968, Theorem, §135, p.369), since

$$\sup_{x\in H\setminus\{0\}} \frac{\|(T_m - T)x\|}{\|x\| + \|Tx\|} \leq |||T_m - T|||_\infty \to 0, \quad m \to \infty.$$

□

A.2.2 Schatten Spaces

Notice that $\mathrm{sing} : \mathscr{S}_c(H) \to c_0$, where c_0 is the space of sequences $a = (a_1, a_2, \ldots)$ such that $\lim_n a_n \to 0$. Recall that ℓ_p, $p \in [1, +\infty)$, is the space of sequences $a = (a_1, a_2, \ldots) \subset \mathbb{R}^{\mathbb{N}}$ such that

$$\|a\|_{\ell_p} := \left(\sum_{n\geq 1} |a_n|^p \right)^{1/p} < \infty,$$

and that $1 \leq p \leq q < \infty$ implies $\ell_p \subset \ell_q \subset c_0$. The set of compact operators whose singular values are in ℓ_p are called *Schatten spaces*:

Proposition A.2.6 (Definition)**.** *For $p \in [1, \infty)$, we define the* Schatten space *$\mathscr{S}_p(H)$ as the space of operators*

$$\mathscr{S}_p(H) = \left\{ T \in \mathscr{S}_c(H) : \mathrm{sing}(T) \in \ell_p \right\}. \tag{A.2.6}$$

It is a Banach space with the norm

$$|||T|||_p = \left\| \mathrm{sing}(T) \right\|_{\ell_p}, \quad T \in \mathscr{S}_p(H).$$

Equivalently, it can be shown that a bounded operator T on H is in $\mathscr{S}_p(H)$ if, and only if

$$\sum_{n\geq 1} |\langle Te_n, e_n\rangle|^p < \infty$$

for every orthonormal basis (e_n) of H. As a consequence, the Schatten classes are closed under taking adjoints. The Schatten classes have the same inclusion properties of the ℓ_p spaces, namely $1 \leq p \leq q < \infty$ implies

$$\mathscr{S}_1(H) \subset \mathscr{S}_p(H) \subset \mathscr{S}_q(H) \subset \mathscr{S}_c(H) \subset \mathscr{S}_\infty(H),$$

where we recall that $\mathscr{S}_\infty(H)$ is the space of bounded linear operators on H.
For any $u, v \in H$, we can define an bounded operator $u \otimes_2 v$ on H by

$$(u \otimes_2 v)h = \langle h, v\rangle u, \quad h \in H. \tag{A.2.7}$$

The operator $u \otimes_2 v$ is called a *rank-one* operator (because its image is a subspace of H of dimension one), and any finite linear combination of rank-one operators is called a *finite rank operator*. The rank-one operators have the property that

$$|||u \otimes_2 v|||_p = \|u\|\|v\|, \quad p \in [1,\infty], \tag{A.2.8}$$

and therefore the finite rank operators belong to all the Schatten spaces. An important fact is that the space of finite-rank operators is actually dense in each of the Schatten spaces $\mathscr{S}_p(H)$, $p \in [1,\infty)$, with their respective norm, as well as in the space of compact operators, with the operator norm. Similarly to the ℓ_p spaces, the Schatten spaces satisfy a Hölder-type inequality:

Proposition A.2.7 (Hölder inequality for Schatten spaces)**.** *Let $1 \leq r, s, t \leq \infty$, such that $r^{-1} = s^{-1} + t^{-1}$, and $S \in \mathscr{S}_s(H)$, $T \in \mathscr{S}_t(H)$. Then $ST \in \mathscr{S}_r(H)$ and*

$$|||ST|||_r \leq |||S|||_s |||T|||_t. \tag{A.2.9}$$

A.2.2.1 Trace-class Operators

We now look at the Schatten space $\mathscr{S}_1(H)$, which is called the space of *trace-class* or *nuclear* operators on H, and for which we can define the trace.

Definition A.2.8. *We define the* trace *of an operator $T \in \mathscr{S}_1(H)$ by*

$$\mathrm{Tr}(T) = \sum_{n\geq 1} \langle Te_n, e_n\rangle, \tag{A.2.10}$$

where (e_n) is an orthonormal basis of H.

The definition of the trace is independent of the choice of the orthonormal basis (e_n), and we have the following properties

Proposition A.2.9. *For all $S \in \mathscr{S}_1(H)$, $u, v \in H$,*

1. $\operatorname{Tr}(u \otimes_2 v) = \langle u, v\rangle$.
2. *For any orthonormal basis $(e_n)_{n\geq 1}$,*
$$|\operatorname{Tr}(S)| \leq \sum_{n\geq 1} |\langle Se_n, e_n\rangle| \leq |||S|||_1, \quad S \in \mathscr{S}_1(H).$$
3. $\operatorname{Tr}\left(S^\dagger\right) = \overline{\operatorname{Tr}(S)}$,
4. *if $\langle Sx, x\rangle \geq 0$ for all $x \in H$ (i.e. S is positive), then*
$$\operatorname{Tr}(S) = |||S|||_1 \geq 0.$$

In particular, the trace defines a continuous linear functional $\mathscr{S}_1(H) \to \mathbb{C}$. Moreover, if $1 \leq s, t \leq \infty$ such that $s^{-1} + t^{-1} = 1$, then for any $S \in \mathscr{S}_s(H)$ and $T \in \mathscr{S}_t(H)$,
$$\operatorname{Tr}(ST) = \operatorname{Tr}(TS). \tag{A.2.11}$$

A.2.2.2 Hilbert–Schmidt Operators

Another important Schatten space is $\mathscr{S}_2(H)$, called the space of *Hilbert–Schmidt* operators. It is particular, because it is actually a Hilbert space, if we equip it with the scalar product

$$\langle T, S\rangle_{\mathscr{S}_2} = \operatorname{Tr}\left(T^\dagger S\right) = \sum_{n\geq 1} \langle Se_n, Te_n\rangle, \quad S, T \in \mathscr{S}_2(H), \tag{A.2.12}$$

where (e_n) is any orthonormal basis of H. In particular, $|||T|||_2 = \sum_{n\geq 1} \|Te_n\|^2$, and $\{e_i \otimes_2 e_j : i, j \geq 1\}$ is an orthonormal basis of $\mathscr{S}_2(H)$. Here are some properties of the tensor product $\cdot \otimes_2 \cdot$ (recall the definition of $\otimes_2$; see (A.2.7))

Proposition A.2.10. *For any $u, v, f, g \in H$, $A, B \in \mathscr{S}_2(H)$*

1. *$\cdot \otimes_2 \cdot$ is linear on the left, and conjugate-linear on the right,*
2. $\langle u \otimes_2 v, f \otimes_2 g\rangle_{\mathscr{S}_2} = \langle u, f\rangle\langle g, v\rangle = \langle (u \otimes_2 v)g, f\rangle$,
3. $\langle A, u \otimes_2 v\rangle_{\mathscr{S}_2} = \langle Av, u\rangle = \langle v \otimes_2 u, A^\dagger\rangle_{\mathscr{S}_2}$,
4. $|||u \otimes_2 v|||_2 = \|u\|\,\|v\|$,
5. $(u \otimes_2 v)(f \otimes_2 g) = \langle f, v\rangle u \otimes_2 g$,
6. $(u \otimes_2 v)^\dagger = v \otimes_2 u$,
7. $(A \bigotimes_2 B)^\dagger = B \bigotimes_2 A$.

Proof. The proof follows from the definition and the properties of the inner product, and is therefore omitted. □

The following result characterizes the Hilbert–Schmidt operators on L^2 spaces:

Proposition A.2.11 (Weidmann 1980, Theorem 6.11)**.** *Let $H = L^2(M, \mathbb{C})$, with M a measurable subset of $\mathbb{R}^n$. An operator $T : H \to H$ is Hilbert–Schmidt if and only if $\exists k \in L^2(M \times M, \mathbb{C})$ such that for all $f \in H$,*

$$Tf(x) = \int_M k(x,y) f(y)\, dy \quad \text{a.s. for } x \in M.$$

We have $|||T|||_2 = \iint_{M\times M} |k(x,y)|^2\, dx\, dy$, and the adjoint $T^\dagger$ is induced by the kernel $k^\dagger(x,y) = \overline{k(y,x)}$.

The following result follows from Fubini's theorem:

Lemma A.2.12. *Let $K \subset \mathbb{R}^n$ be measurable and compact. Let $a, b, c \in L^2(K \times K, \mathbb{R})$, with induced operators A, B, C on $L^2(K, \mathbb{R})$. That is,*

$$Af(\tau) = \int_K a(\tau,\sigma) f(\sigma) d\sigma, \quad \text{in } L^2, \quad \forall f \in L^2(K, \mathbb{R}).$$

We can define the product operator AB of two operators by composition, i.e. the operator AB is defined by $(AB)f = A(Bf)$ for $f \in L^2(K, \mathbb{R})$, and has kernel

$$r(\tau,\sigma) = \int_K a(\tau,\mu) b(\mu,\sigma) d\sigma, \quad \text{in } L^2.$$

This operation is associative.

A.2.3 Complexification of a Real Hilbert Space

Given a real Hilbert space $H_{\mathbb{R}}$, we can construct a new complex Hilbert space H which contains a subspace isometrically isomorphic to $H_{\mathbb{R}}$. The process is called the complexification of $H_{\mathbb{R}}$ (Halmos 1974, Conway 1990). H consists of all pairs (u, v), where $u, v \in H_{\mathbb{R}}$, and we define the sum

$$(u, v) + (u', v') = \left(u + u', v + v'\right), \quad u, u', v, v' \in H_{\mathbb{R}},$$

and the multiplication by complex numbers $\alpha + \mathbf{i}\beta$ by

$$(\alpha + \mathbf{i}\beta)(u, v) = (\alpha u - \beta v, \beta u + \alpha v).$$

Furthermore, we define the inner-product on H by

$$\left\langle u + \mathbf{i}v, u' + \mathbf{i}v' \right\rangle = \left\langle u, u' \right\rangle + \left\langle v, v' \right\rangle - \mathbf{i}\left[\left\langle u, v' \right\rangle - \left\langle v, u' \right\rangle\right]. \tag{A.2.13}$$

Notice that the induced norm satisfies

$$\|u + \mathbf{i}v\|^2 = \|u\|^2 + \|v\|^2,$$

for all $u, v \in H_{\mathbb{R}}$. In other words, the formulas behave as if $(u, v) = u + \mathbf{i}v$, and we will write the generic element of H in this way. We notice that the mapping $\iota : H_{\mathbb{R}} \to H$ defined by $\iota(u) = u + \mathbf{i}v$ is an $\mathbb{R}$-linear isomorphism, which allows us to identify $H_{\mathbb{R}}$ with $\iota(H_{\mathbb{R}}) \subset H$. Furthermore, we define the *real part* and *imaginary part* mappings $\Re, \Im : H \to H_{\mathbb{R}}$ by

$$\Re(u + \mathbf{i}v) = u, \quad \Im(u + \mathbf{i}v) = v.$$

These maps are $\mathbb{R}$-linear, but not $\mathbb{C}$-linear. The space H with $\langle \cdot, \cdot \rangle$ is a Hilbert space, and is called the *complexification of* $H_{\mathbb{R}}$. Notice that the complexification of a complex Hilbert space is not necessarily isometrically isomorphic to itself. We shall call *Complexified Hilbert Space* a Hilbert space H that is the complexification of a real Hilbert space $H_{\mathbb{R}}$.
If $(e_n)_{n\geq 1}$ is an orthonormal basis of $H_{\mathbb{R}}$, then it is also an orthonormal basis of H, and any bounded linear mapping T on $H_{\mathbb{R}}$ can be extended to an operator T' on H by the formula

$$T'(u + \mathbf{i}v) = Tu + \mathbf{i}Tv.$$

The extension has the following properties: for all bounded operators S, T on $H_{\mathbb{R}}$,

1. T' is bounded with $\left|\left|\left| T' \right|\right|\right|_\infty = |||T|||_\infty$
2. $(\alpha T)' = \alpha T', \quad \alpha \in \mathbb{R}$
3. $(T + S)' = T' + S'$
4. $(T^\dagger)' = (T')^\dagger$

We also define the conjugation $\bar{\cdot} : H \to H$ by

$$\overline{u + \mathbf{i}v} = u - \mathbf{i}v.$$

Notice that it is not linear with respect to $\mathbb{C}$-multiplication, but linear with respect to $\mathbb{R}$-multiplication. The conjugation has the following properties:

1. It is an involution, i.e. $\overline{\overline{u + \mathbf{i}v}} = u + \mathbf{i}v$, and
2. $\overline{\langle x, y \rangle} = \langle \overline{x}, \overline{y} \rangle, \quad x, y \in H$

We can also define the conjugate of an operator T' on H by the formula

$$\overline{T'}x = \overline{\left(T'\overline{x}\right)} \tag{A.2.14}$$

and the transpose of an operator by

$$T'^{\mathsf{T}} = \overline{T'}^\dagger \tag{A.2.15}$$

These have the following properties:

1. $T'^{\mathsf{T}} = \overline{\overline{T'}^{\dagger}} = \overline{T'}^{\dagger}$
2. T'^{T} restricted to corresponds to the usual adjoint operator on $H_{\mathbb{R}}$, i.e.
$$\langle Tu, v\rangle = \left\langle u, T'^{\mathsf{T}} v\right\rangle, \quad \forall u, v \in H_{\mathbb{R}}, T \in \mathscr{S}_{\infty}(H_{\mathbb{R}}).$$

Example A.2.13. *Here are some examples of complexified Hilbert spaces:*

1. $\mathbb{C}^d$ *is the complexification of* $\mathbb{R}^d$.
2. $L^2([0,1],\mathbb{C})$ *is the complexification of* $L^2([0,1],\mathbb{R})$.
3. $\ell_2(\mathbb{C})$ *is the complexification of* $\ell_2(\mathbb{R})$.

A.3 Tensor Products

A.3.1 Hilbert Tensor Products

Given k Hilbert spaces $(H_j, \langle\cdot,\cdot\rangle_j)$, $j = 1,\dots,k$, on can construct their algebraic tensor product, denoted $H_1 \odot \cdots \odot H_k$, whose elements are linear combinations of *simple tensors* $\otimes_{j=1}^k u_j = u_1 \otimes \cdots \otimes u_k$, where $u_j \in H_j$. The simple tensors are linear in each coordinate, that is,

$$\begin{aligned} u_1 \otimes \cdots \otimes u_{l-1} \otimes \left(u_l + \lambda u_l'\right) \otimes u_{l+1} \otimes \cdots \otimes u_k = \\ u_1 \otimes \cdots \otimes u_{l-1} \otimes u_l \otimes u_{l+1} \otimes \cdots \otimes u_k \\ + \lambda \left(u_1 \otimes \cdots \otimes u_{l-1} \otimes u_l' \otimes u_{l+1} \otimes \cdots \otimes u_k\right), \end{aligned}$$

for all $u_j \in H_j$, $u_l, u_l' \in H_l, \lambda \in \mathbb{C}$. H_0 is a linear space, which we equip with the following inner product, that we call

$$\left\langle \bigotimes_{j=1}^k u_j, \bigotimes_{j=1}^k v_j \right\rangle := \prod_{j=1}^k \langle u_j, v_j\rangle_j.$$

The completion of H_0 under the norm generated by this scalar product,

$$\left\| \bigotimes_{j=1}^k u_j \right\| = \prod_{j=1}^k \|u_j\|_j,$$

is a Hilbert space, called the *Hilbert tensor product*, whose inner product is called the *Hilbert tensor scalar product*, and whose norm is called the *Hilbert tensor norm*. The Hilbert tensor product is denoted

$$H = \bigotimes_{j=1}^k H_j = H_1 \otimes H_2 \otimes \cdots \otimes H_k, \tag{A.3.1}$$

If $(e_{j,n})_{n\geq 1}$ is a orthonormal basis of H_j, for $j = 1,\dots,k$, then

$$\left(e_{1,n_1} \otimes e_{2,n_2} \otimes \cdots \otimes e_{k,n_k}\right)_{n_1,\dots,n_k \geq 1}$$

is an orthonormal basis of H. Taking tensor products of Hilbert spaces is associative (provided we identify isomorphic spaces), that is

$$(H_1 \otimes H_2) \otimes H_3 = H_1 \otimes (H_2 \otimes H_3) = H_1 \otimes H_2 \otimes H_3,$$

and commutative,

$$H_1 \otimes H_2 = H_2 \otimes H_1.$$

In the particular case of the Hilbert space $H = L^2([0,1],\mathbb{C})$, we have that

$$(\varphi \otimes \psi)(\tau,\sigma) = \varphi(\tau)\psi(\omega), \quad \varphi,\psi \in L^2([0,1],\mathbb{C}); \tau,\sigma \in [0,1].$$

Given bounded operators $T_i : H_i \to H_i', i = 1,\dots,k$ between Hilbert spaces, we define the their tensor product

$$T_1 \mathbin{\widetilde{\bigotimes}} \cdots \mathbin{\widetilde{\bigotimes}} T_k : \overset{k}{\underset{i=1}{\otimes}} H_i \longrightarrow \overset{k}{\underset{i=1}{\otimes}} H_i'$$

as the unique bounded linear operator satisfying

$$\left(T_1 \mathbin{\widetilde{\bigotimes}} \cdots \mathbin{\widetilde{\bigotimes}} T_k\right)(x_1 \otimes \cdots x_k) = T_1 x_1 \otimes \cdots \otimes T_k x_k,$$

for all $x_i \in H_i, i = 1,\dots,k$. Amongst the various properties of the tensor product of bounded operators, we have that

$$\left(T_1 \mathbin{\widetilde{\bigotimes}} T_2\right)^\dagger = T_1^\dagger \mathbin{\widetilde{\bigotimes}} T_2^\dagger.$$

The k-fold tensor product of H with himself will be denoted $H^{\otimes k}$

A.3.1.1 2-fold Tensor products and Hilbert–Schmidt Operators

Provided H is a complexified Hilbert space, i.e. it has a mapping $\bar{\cdot} : H \to H$ that is anti-linear and whose square is the identity, the tensor product $H \otimes H$ and the space of Hilbert–Schmidt operators $\mathscr{S}_2(H)$ are isomorphic:

Proposition A.3.1. *The mapping $\Phi : H \otimes H \to \mathscr{S}_2(H)$ defined by the continuous linear extension of*

$$\Phi(u \otimes v) = u \otimes_2 \overline{v}, \quad \forall u, v \in H \tag{A.3.2}$$

is an isometric isomorphism. In particular,

$$\langle A, B\rangle = \langle \Phi(A), \Phi(B)\rangle_{\mathscr{S}_2}, \quad \forall A, B \in H \otimes H \tag{A.3.3}$$

Proof. Fix a basis $(e_n)_{n\geq 1}$ of H. The mapping is clearly linear, it's image contains all the elements $\{e_i \otimes_2 e_j : i, j \geq 1\}$ and property (A.3.3) is valid on the basis $\{e_i \otimes \overline{e_j} : i, j \geq 1\}$. Therefore the continuous linear extension is well defined, and the proof is complete. □

A.3.2 Projective Tensor Products of Banach Spaces

We present briefly in this Section the projective tensor products of Banach spaces, which are of interest for us because of their natural analogy to the nuclear norm of operators on Hilbert spaces (see Proposition A.3.2). Some references for this Section are Schatten (1950), Grothendieck (1953), Ryan (2002), and Diestel et al. (2008).

Given Banach spaces $(B_i, \|\cdot\|_i)_{i=1,2}$, we define the norm *projective norm* $\|\cdot\|_\pi : B_1 \odot B_2 \to \mathbb{R}$ on its algebraic tensor by

$$\|x\|_\pi = \inf\left\{ \sum_{i=1}^n \|u_i\|_1 \|v_i\|_2 : x = \sum_{i=1}^n u_i \otimes v_i \right\}. \tag{A.3.4}$$

This is indeed a norm, with the property that

$$\|u \otimes v\|_\pi = \|u\|_1 \|v\|_2, \quad \forall u \in B_1, v \in B_2. \tag{A.3.5}$$

Furthermore, it is the largest norm amongst all the so-called *cross-norms* $\|\cdot\|_*$, satisfying

$$\|u \otimes v\| \leq \|u\|_1 \|v\|, \quad \forall u \in B_1, v \in B_2$$

We denote by $B_1 \otimes_\pi B_2$ the completion of $B_1 \odot B_2$ with respect to the projective norm, and call it the *projective tensor product* of B_1 and B_2. It is a Banach space. Furthermore, iterating the process, we can define the k-fold projective tensor product of Banach spaces $B_1, \ldots, B_k$,

$$B_1 \otimes_\pi \cdots \otimes_\pi B_k$$

and we have, up to isomorphisms,

$$B_1 \otimes_\pi B_2 = B_2 \otimes_\pi B_1, \quad (B_1 \otimes_\pi B_2) \otimes_\pi B_3 = B_1 \otimes_\pi (B_2 \otimes_\pi B_3) \tag{A.3.6}$$

In the special case where the Banach spaces are in fact Hilbert spaces, we notice that the projective norm is stronger than the Hilbert tensor norm, and up to isomorphism, we can say that

$$H_1 \otimes_\pi H_2 \subset H_1 \otimes H_2, \tag{A.3.7}$$

and the inclusion is continuous, since the Hilbert tensor norm is also a cross-norm.

If $T_i : B_i \to B_i'$, $i = 1, \ldots, k$, are bounded operators between Banach spaces, then

$$(T_1 \otimes_\pi \cdots \otimes_\pi T_k) : \bigotimes_{i=1}^k{}_\pi B_i \to \bigotimes_{i=1}^k{}_\pi B_i'$$

denotes the unique bounded operator satisfying

$$(T_1 \otimes_\pi \cdots \otimes_\pi T_k)(x_1 \otimes \cdots \otimes x_k) = T_1 x_1 \otimes \cdots \otimes T_k x_k,$$

for all $x_i \in B_i$, $i = 1, \ldots, k$. Furthermore,

$$|||T_1 \otimes_\pi \cdots \otimes_\pi T_k|||_\infty \leq |||T_1|||_\infty \cdots |||T_k|||_\infty.$$

Using this and (A.3.7), we can show that

$$H_1 \otimes_\pi \cdots \otimes_\pi H_k \subset H_1 \otimes \cdots \otimes H_k \tag{A.3.8}$$

up to isomorphism.
The following universal property characterizes the projective tensor product (Ryan 2002, Theorem 2.9): For any Banach space B, the spaces $\mathscr{L}_2(B_1, B_2; B)$ of bilinear continuous mappings $B_1 \times B_2 \to B$ and the space $\mathscr{L}(B_1 \otimes_\pi B_2; B)$ of bounded linear mapping $B_1 \otimes_\pi B_2 \to B$ are isometrically isomorphic. Furthermore, using (A.1.4), we can extend this statement to

$$\mathscr{L}(B_1 \otimes_\pi \cdots \otimes_\pi B_k; B) = \mathscr{L}_k(B_1, \ldots, B_k; B) \tag{A.3.9}$$

up to a isometric isomorphism.

A.3.2.1 Projective Tensor Products and Nuclear Operators

The is a very close relationship between projective tensor products of Hilbert spaces and nuclear operators.

Proposition A.3.2 (Schatten (1950))**.** *Let H be a separable Hilbert space. Then the mapping $\Psi : H \otimes_\pi H \to \mathscr{S}_1(H)$ defined by linear extension of*

$$\Psi(u \otimes v) = u \otimes_2 \overline{v}, \quad \forall u, v \in H \tag{A.3.10}$$

is a bijective isomorphism.

A.3.3 Permutation of Tensors

We describe here briefly permutations of tensor products. We will do it for the Banach space case, with the projective tensor product $. \otimes_\pi .$, but this holds also for the Hilbert space tensor product.
Given Banach spaces $B_1, \ldots, B_n$, and a permutation $\nu : \{1, \ldots, n\} \to \{1, \ldots, n\}$, we define

$$\mathrm{Perm}_\nu(\cdot) : B_1 \otimes_\pi B_2 \otimes_\pi \cdots \otimes_\pi B_n \to B_{\nu(1)} \otimes_\pi \cdots \otimes_\pi B_{\nu(n)}$$

as the unique linear operator satisfying

$$\mathrm{Perm}_\nu \left(\bigotimes_{i=1}^n x_i \right) = \bigotimes_{i=1}^n x_{\nu(i)},$$

for all simple tensors. It is clearly a bijective isomorphism. We can also extend the permutation operator $\mathrm{Perm}_\nu(\cdot)$ to tensors of operators. Given

bounded operators $T_i : B_i \to B_i'$, recall that we have defined their tensor as the bounded linear operator

$$T_1 \otimes_\pi \cdots \otimes_\pi T_n : B_1 \otimes_\pi \cdots \otimes_\pi B_n \to B_1' \otimes_\pi \cdots \otimes_\pi B_n$$

satisfying $\otimes_\pi{}_{i=1}^n T_i \left(\otimes_{i=1}^n x_i\right) = \otimes_{i=1}^n T_i x_i$ for all simple tensors. We define

$$\mathrm{Perm}_\nu \left(T_1 \otimes_\pi \cdots \otimes_\pi T_n\right) = T_{\nu(1)} \otimes_\pi \cdots \otimes_\pi T_{\nu(n)}.$$

We have the following property:

$$\mathrm{Perm}_\nu \left(\left[\underset{i=1}{\overset{n}{\otimes_\pi}} A_i \right] \left[\underset{i=1}{\overset{n}{\otimes}} x_i \right] \right) = \mathrm{Perm}_\nu \left(\underset{i=1}{\overset{n}{\otimes_\pi}} A_i \right) \mathrm{Perm}_\nu \left(\underset{i=1}{\overset{n}{\otimes}} x_i \right) \tag{A.3.11}$$

A.3.4 Kronecker Products

For two operators $A, B \in \mathscr{S}_\infty(H)$, we define their *Kronecker product* $A \mathbin{\widetilde{\otimes}_2} B \in \mathscr{S}_\infty(\mathscr{S}_\infty(H))$, by $A \mathbin{\widetilde{\otimes}_2} B(C) = ACB^\dagger$, for $C \in \mathscr{S}_\infty(H)$. It has the following properties:

Proposition A.3.3. *For any $A, B, C, D \in \mathscr{S}_\infty(H)$, $u, v, f, g \in H$,*

1. *$\cdot \mathbin{\widetilde{\otimes}_2} \cdot$ is linear on the left, and sesquilinear on the right.*
2. *$(A \mathbin{\widetilde{\otimes}_2} B)(u \otimes_2 v) = Au \otimes_2 Bv$*
3. *$\left|\!\left|\!\left| A \mathbin{\widetilde{\otimes}_2} B \right|\!\right|\!\right|_\infty \le |\!|\!| A |\!|\!|_\infty |\!|\!| B |\!|\!|_\infty$*
4. *$(A \mathbin{\widetilde{\otimes}_2} B)(C \mathbin{\widetilde{\otimes}_2} D) = AC \mathbin{\widetilde{\otimes}_2} BD$*
5. *$(u \otimes_2 v) \mathbin{\widetilde{\otimes}_2} (f \otimes_2 g) = (u \otimes_2 f) \mathbin{\widetilde{\otimes}_2} (v \otimes_2 g)$*
6. *$(A \mathbin{\widetilde{\otimes}_2} B)^\dagger = A^\dagger \mathbin{\widetilde{\otimes}_2} B^\dagger$, where $(A \mathbin{\widetilde{\otimes}_2} B)$ is viewed as bounded operator on the Hilbert space $\mathscr{S}_2(H)$.*

When $A, B \in \mathscr{S}_2(H)$, $A \mathbin{\widetilde{\otimes}_2} B \in \mathscr{S}_2(\mathscr{S}_2(H))$, and we have

Proposition A.3.4. *For any $A, B, C, D \in \mathscr{S}_2(H)$, $u, v, f, g \in H$,*

1. *$\left\langle A \mathbin{\widetilde{\otimes}_2} B, C \mathbin{\widetilde{\otimes}_2} D \right\rangle_{\mathscr{S}_2} = \langle A, C \rangle_{\mathscr{S}_2} \langle D, B \rangle_{\mathscr{S}_2}$*
2. *$\left|\!\left|\!\left| A \mathbin{\widetilde{\otimes}_2} B \right|\!\right|\!\right|_2 = |\!|\!| A |\!|\!|_2 |\!|\!| B |\!|\!|_2$*

In the case $H = L^2([0,1], \mathbb{C})$, if $A, B \in \mathscr{S}_2(H)$ are Hilbert–Schmidt operators, hence they are also kernel operators, with kernels $a(\tau, \sigma), b(\tau, \sigma)$, respectively. The operator $A \mathbin{\widetilde{\otimes}_2} B$ is then also a Hilbert–Schmidt operator on $\mathscr{S}_2(H)$, with kernel

$$k(\tau, \sigma, x, y) = a(\tau, x)\overline{b(\sigma, y)}.$$

For two operators $A, B \in \mathscr{S}_\infty(H)$, we also define their *transpose Kronecker product* $A \mathbin{\widetilde{\otimes}_\mathsf{T}} B \in \mathscr{S}_\infty(\mathscr{S}_\infty(H))$, by $A \mathbin{\widetilde{\otimes}_\mathsf{T}} B(C) = (A \mathbin{\widetilde{\otimes}_2} B)(C^\mathsf{T}) = AC^\mathsf{T}B^\mathsf{T}$, for $C \in \mathscr{S}_\infty(H)$.

Proposition A.3.5. *For any $A, B, C, D \in \mathscr{S}_\infty(H)$, $u, v, f, g \in H$,*

1. $\left(A \mathbin{\widetilde{\otimes}_{\mathsf{T}}} B\right) C = \left(A \mathbin{\widetilde{\otimes}_2} \overline{B}\right) \overline{C}^\dagger$
2. $\cdot \mathbin{\widetilde{\otimes}_{\mathsf{T}}} \cdot$ *is bilinear.*
3. $(A \mathbin{\widetilde{\otimes}_{\mathsf{T}}} B)(u \otimes_2 v) = A\overline{v} \otimes_2 \overline{Bu}$
4. $(u \otimes_2 v) \mathbin{\widetilde{\otimes}_{\mathsf{T}}} (f \otimes_2 g) = (u \otimes_2 g) \mathbin{\widetilde{\otimes}_2} \left(\overline{f} \otimes_2 \overline{v}\right) = \left(u \otimes_2 \overline{f}\right) \otimes_2 \left(g \otimes_2 \overline{v}\right)$
5. $\left|\!\left|\!\left| A \mathbin{\widetilde{\otimes}_{\mathsf{T}}} B \right|\!\right|\!\right|_\infty \le |\!|\!| A |\!|\!|_\infty |\!|\!| B |\!|\!|_\infty$
6. $(A \mathbin{\widetilde{\otimes}_{\mathsf{T}}} B)(C \mathbin{\widetilde{\otimes}_{\mathsf{T}}} D) = (AD \mathbin{\widetilde{\otimes}_2} \overline{BC})$
7. $(A \mathbin{\widetilde{\otimes}_{\mathsf{T}}} B)(C \mathbin{\widetilde{\otimes}_2} D) = A\overline{D} \mathbin{\widetilde{\otimes}_{\mathsf{T}}} BC$
8. $(A \mathbin{\widetilde{\otimes}_2} B)(C \mathbin{\widetilde{\otimes}_{\mathsf{T}}} D) = AC \mathbin{\widetilde{\otimes}_{\mathsf{T}}} \overline{B} D$
9. *Viewing $(A \mathbin{\widetilde{\otimes}_{\mathsf{T}}} B)$ as an operator on the Hilbert space $\mathscr{S}_2(H)$, we get*

$$(A \mathbin{\widetilde{\bigotimes}_{\mathsf{T}}} B)^\dagger = B^\dagger \mathbin{\widetilde{\bigotimes}_{\mathsf{T}}} A^\dagger, \quad (A \mathbin{\widetilde{\bigotimes}_{\mathsf{T}}} B)^{\mathsf{T}} = B^{\mathsf{T}} \mathbin{\widetilde{\bigotimes}_{\mathsf{T}}} A^{\mathsf{T}}$$

In the case $H = L^2([0,1], \mathbb{C})$, if $A, B \in \mathscr{S}_2(H)$ are Hilbert–Schmidt operators, they are also kernel operators, with kernels $a(\tau,\sigma), b(\tau,\sigma)$, respectively. The operator $A \mathbin{\widetilde{\otimes}_{\mathsf{T}}} B$ is then also a Hilbert–Schmidt operator on $\mathscr{S}_2(H)$, with kernel

$$k(\tau,\sigma,x,y) = a(\tau,y)\, b(\sigma,x).$$

Proposition A.3.6. *For any $A, B, C, D \in \mathscr{S}_2(H)$,*

1. $\left\langle A \mathbin{\widetilde{\otimes}_{\mathsf{T}}} B, C \mathbin{\widetilde{\otimes}_{\mathsf{T}}} D \right\rangle_{\mathscr{S}_2} = \langle A, C \rangle_{\mathscr{S}_2} \langle D, B \rangle_{\mathscr{S}_2}$
2. $\left|\!\left|\!\left| A \mathbin{\widetilde{\otimes}_{\mathsf{T}}} B \right|\!\right|\!\right|_2 = |\!|\!| A |\!|\!|_2 |\!|\!| B |\!|\!|_2$

The proofs of this subsection are omitted; they follow easily from the definition of the Kronecker products and tensor products, by evaluating (if needed) the left-hand and right-hand side of the equations on simple tensors.

APPENDIX B

The Bochner Integral

Since we are working with random elements of Hilbert spaces, or even Banach spaces, we will need to give a meaning to integrals of mappings taking values in such spaces. The Bochner integral gives a rigorous treatment of this. Some references are Hildebrandt (1953), Yosida (1995), Ryan (2002), Mikusinski & Weiss (2014).
Let $(\Omega, \mathscr{O}, \mu)$ be a complete finite measure space, i.e., $\mathscr{O}$ is a σ-algebra of subsets of Ω, and $\mu : \mathscr{O} \to [0, +\infty)$ be σ-additive, with $\mu(\Omega) < \infty$. For $A \in \mathscr{O}$, We call $\mu(A)$ the measure of A. We assume that $S \subset A \in \mathscr{O}$ and $\mu(A) = 0 \implies S \in \mathscr{O}$.

Definition B.0.7 (and Proposition)**.** *Let B be a complex Banach space, with norm $\|\cdot\|$, and Borel σ-algebra $\mathscr{B}$. A function $X : (\Omega, \mathscr{O}) \to (B, \mathscr{B})$ is called* strongly measurable *if*

$$X \text{ is measurable and } X(\Omega) \text{ is separable,}$$

and Bochner integrable *if*

$$X \text{ is strongly measurable and } \int_\Omega \|X\| \, d\mu < \infty.$$

Notice that if B is separable, then "measurable" and "strongly measurable" have the same meaning. Strong measurability has the following characterization:

Lemma B.0.8. *Let $X : \Omega \to B$ be a function. The following assertions are equivalent:*

1. *X is strongly measurable;*

2. *There is a sequence of simple functions* $(X_n)_{n\geq 1}$ *such that*

$$\|X_n(\omega) - X(\omega)\| \to 0, \quad \forall \omega \in \Omega, n \to \infty;$$

3. $X(\Omega)$ *is separable, and* $\phi(X) : \Omega \to \mathbb{C}$ *is measurable for all* $\phi \in B^*$ *(the dual of* B*).*

Moreover, if X *is strongly measurable, then one can choose the sequence of simple functions* (X_n) *such that* $\|X_n(\omega)\| \leq 2\|X(\omega)\|$ *for all* $\omega \in \Omega$.

Proposition B.0.9. *For a Bochner integrable function* X*, we define its integral* $\int_\Omega X d\mu \in B$ *to be the unique element satisfying*

$$\phi(\int_\Omega X d\mu) = \int_\Omega \phi(X) d\mu, \quad \forall \phi \in B^*, \tag{B.0.1}$$

where B^* *is the space of all linear and continuous functionals* $\phi : B \to \mathbb{C}$ *(the topological dual of* B*). Moreover, for* $A \in \mathcal{O}$*, we define*

$$\int_A X d\mu = \int_\Omega X \mathbf{1}_A d\mu.$$

We have the following properties:

1. *For every Bochner integrable function, there exists a unique element in* B *satisfying* (B.0.1).

2. *There exists a sequence of simple functions* $X_n = \sum_{i=1}^n b_i \mathbf{1}_{A_i}$ *converging pointwise to* X*, such that*

$$\int_\Omega \|X - X_n\| d\mu \to 0 \tag{B.0.2}$$

Moreover, for every such sequence of simple functions, we have

$$\int_\Omega X d\mu = \lim_{n\to\infty} \sum_{i=1}^n b_i \mu(A_i) \tag{B.0.3}$$

3. *Contraction property:* $\left\|\int_\Omega X d\mu\right\| \leq \int_\Omega \|X\| d\mu$

4. *if* $T : B \to B'$ *is a bounded linear operator onto another Banach space* B'*, then*

$$T \int_\Omega X d\mu = \int_\Omega TX d\mu \tag{B.0.4}$$

The space of Bochner integrable functions will be denoted $\mathcal{L}(\Omega, B, \mu)$.

Proposition B.0.10 (Dominated Convergence Theorem)**.** *Let* $X_n \in \mathcal{L}(\Omega, B, \mu)$ *for all* $n = 1, 2, \ldots$*, and let* $X : \Omega \to B$ *be strongly measurable such that* $X_n \to X$ *pointwise* μ*-almost everywhere. If there exists a function* $g : \Omega \to [0, +\infty)$ *such that* $\|X_n\| \leq g$ *for all* n *and* $\int_\Omega g d\mu < \infty$*, then*

1. $X \in \mathscr{L}(\Omega, B, \mu)$
2. $\lim_{n\to\infty} \int_\Omega X_n d\mu = \int_\Omega X d\mu$

We now wish to define L^p-Bochner spaces. Let $\mathscr{N} = \{N \in \mathscr{O} : \mu(N) = 0\}$ be the set of all measurable sets of measure zero. We first deal with the case $p \in [1, \infty)$.

Definition B.0.11. *1. For $p \in [1,\infty)$, we define $\mathscr{L}^p(\Omega, B, \mu)$ to be the set of all functions $X : \Omega \to B$ such that $X\mathbf{1}_{\Omega\setminus N}$ is strongly measurable for some $N \in \mathscr{N}$, and*

$$\|X\|_{\mathscr{L}^p} := \left(\int_{\Omega\setminus N} \|X\|^p d\mu\right)^{1/p} < \infty.$$

2. For $p = \infty$, we define $\mathscr{L}^\infty(\Omega, B, \mu)$ to be the set of all functions $X : \Omega \to B$ such that $X\mathbf{1}_{\Omega\setminus N}$ is strongly measurable for some $N \in \mathscr{N}$, and

$$\|X\|_{\mathscr{L}^\infty} := \inf\{c \in [0,\infty] : \mu(\|X\| > c) = 0\} < \infty.$$

Let now $p \in [1, +\infty]$. For $X \in \mathscr{L}^p(\Omega, B, \mu)$, we define its equivalence classes $[X]$ by

$$[X] = \left\{Y \in \mathscr{L}^p(\Omega, B, \mu) : Y\mathbf{1}_{\Omega\setminus N} = X\mathbf{1}_{\Omega\setminus N} \text{ for some } N \in \mathscr{N}\right\},$$

We have that the space

$$L^p(\Omega, B, \mu) = \left\{[X] : X \in \mathscr{L}^p(\Omega, B, \mu)\right\}, \tag{B.0.5}$$

equipped with the norm $\|[X]\|_{L^p} = \|X\|_{\mathscr{L}^p}$ is a Banach space. We shall often abuse of notation and write X instead of $[X]$.
When $\Omega \subset \mathbb{R}^d$ for some positive integer d, and $\mu = \lambda$ is the Lebesgue measure, we shall write $L^p(\Omega, B)$ instead of $L^p(\Omega, B, \lambda)$.
The following result, which follows from Proposition B.0.8 and the Dominated Convergence Theorem, will be used later:

Proposition B.0.12. *The space of simple functions is dense in $(L^p(\Omega, B, \mu), \|\cdot\|_{L^p})$, for each $p \in [1,\infty)$.*

B.0.4.1 Bochner L^2 Spaces on $[-\pi, \pi]$

The particular case of the space $L^2([-\pi,\pi], B)$, which consists of functions $X : [-\pi,\pi] \to B$ with

$$\|X\|_{L^2}^2 = \int_{-\pi}^{\pi} \|X(\omega)\|^2 d\omega < \infty,$$

will be of particular interest in the thesis. The following Lemma gives some of its dense subsets.

Lemma B.0.13. *The following subspaces are dense in $L^p([-\pi,\pi],B)$, $p \in [1,\infty)$:*

1. *The step functions*

$$\left\{ \sum_{j=1}^{J} b_j \mathbf{1}_{[\omega_j,\omega'_j]} : b_j \in B; \omega_j, \omega'_j \in [-\pi,\pi], j = 1,\dots,J; J = 1,2,\dots \right\} \tag{B.0.6}$$

2. *The càdlàg step functions*

$$\left\{ \sum_{j=1}^{J} b_j \mathbf{1}_{[\omega_j,\omega_{j+1})} : b_j \in B; -\pi = \omega_1 < \omega_2 < \cdots < \omega_{J+1} = \pi; J = 1,2,\dots \right\} \tag{B.0.7}$$

3. *The trigonometric polynomials*

$$\left\{ \sum_{n=-N}^{N} b_n e_n : b_n \in B, n = -N,-N+1,\dots,N; N \in \mathbb{N} \right\}, \tag{B.0.8}$$

where $e_n(\omega) = e^{\mathbf{i}\omega n}, \omega \in [-\pi,\pi]$.

Proof. Since the space of simple functions is dense in $L^2([-\pi,\pi],B)$, we only need to show that each of the subspaces is dense in the space of simple functions

$$\left\{ \sum_{j=1}^{J} b_j \mathbf{1}_{I_j} : b_j \in B; I_j \in \mathscr{O}; J < \infty \right\}.$$

By the triangle inequality, it is enough to show that $b\mathbf{1}_I$, $b \in B$, $I \in \mathscr{O}$, can be approximated arbitrarily well from each subspace.

(i) For the step functions, notice that

$$\left\| b\mathbf{1}_I - \sum_{j=1}^{J} b\mathbf{1}_{[\omega_j,\omega'_j]} \right\|_{L^p}^p = \|b\|^p \int_{-\pi}^{\pi} \left| \mathbf{1}_I(\omega) - \sum_{j=1}^{J} \mathbf{1}_{[\omega_j,\omega'_j]}(\omega) \right|^p d\omega, \tag{B.0.9}$$

therefore the result follows from classical Lebesgue integration theory.

(ii) For the càdlàg step functions, the result follows from (i) once we notice that each step function is equal to a càdlàg step function almost everywhere.

(iii) For the trigonometric polynomials, similarly to (B.0.9),

$$\left\| b\mathbf{1}_I - \sum_{|n|<N} b\alpha_n e_n \right\|_{L^p}^p = \|b\|^p \int_{-\pi}^{\pi} \left| \mathbf{1}_I(\omega) - \sum_{|n|<N} \alpha_n e^{\mathbf{i}\omega n} \right|^p d\omega, \tag{B.0.10}$$

where $\alpha_n \in \mathbb{C}$, and the result follows from the theory of Fourier series, since Fejér kernels is an approximate identity.

□

B.0.4.2 The Case $B = \mathscr{S}_p(H)$

Let us add a note about the case where B is a Schatten class of operators on a Hilbert space. We will take here $\Omega = [-\pi,\pi]$. First, notice that if $A : [-\pi,\pi] \to \mathscr{S}_p(H)$ is strongly measurable, then so is $A^\dagger(\omega) = (A(\omega))^\dagger$. Furthermore, if $A_1 : [-\pi,\pi] \to \mathscr{S}_p(H)$ and $A_2 : [-\pi,\pi] \to \mathscr{S}_q(H)$ are strongly measurable, with $p^{-1} + q^{-1} = 1$, then the function $A_1A_2(\omega) = A_1(\omega)A_2(\omega)$ is a strongly measurable mapping $[-\pi,\pi] \to \mathscr{S}_1(H)$. Indeed, let $(A_{1,n})$, respectively $(A_{2,n})$, be a sequence of simple functions in converging to A_1 in $|||\cdot|||_p$, respectively A_2 in $|||\cdot|||_q$. Then by Hölder's inequality,

$$\left|\left|\left| A_1A_2 - A_{1,n}A_{2,n} \right|\right|\right|_1 \leq \left|\left|\left| A_1 - A_{1,n} \right|\right|\right|_p |||A_2|||_q + \left|\left|\left| A_{1,n} \right|\right|\right|_p \left|\left|\left| A_2 - A_{2,n} \right|\right|\right|_q,$$

and therefore $A_{1,n}A_{2,n}$ converges pointwise to A_1A_2. Since a product of simple function is a simple function, A_1A_2 is strongly measurable.
The following result shall be useful at a later stage:

Proposition B.0.14. *Let $S \in L^p([-\pi,\pi],\mathscr{S}_1(H))$ and $A_1, A_2 \in L^{2q}([-\pi,\pi],\mathscr{S}_\infty(H))$, where $1 \leq p,q \leq \infty$ and $p^{-1} + q^{-1} = 1$. Then the function $A_1SA_2^\dagger(\omega) = A_1(\omega)S(\omega)A_2^\dagger(\omega)$ belongs to $L^1([-\pi,\pi],\mathscr{S}_1(H))$.*

Proof. Assume without loss of generality that both A_1 and A_2 are strongly measurable on $[-\pi,\pi]$. Then $T = A_1SA_2^\dagger$ is also strongly measurable, and using Hölder's inequality for the Schatten norms and for the L^p norms, we get

$$\begin{aligned}\int_{-\pi}^{\pi} \left|\left|\left| A_1(\omega)S(\omega)A_2^\dagger(\omega) \right|\right|\right|_1 d\omega &\leq \int_{-\pi}^{\pi} |||A_1(\omega)|||_\infty |||S(\omega)|||_1 \left|\left|\left| A_2^\dagger(\omega) \right|\right|\right|_\infty d\omega \\ &\leq \left(\int_{-\pi}^{\pi} |||\mathscr{F}_\omega|||_1^p d\omega \right)^{1/p} \left(\int_{-\pi}^{\pi} |||A_1(\omega)|||_\infty^{2q} d\omega \int_{-\pi}^{\pi} |||A_2(\omega)|||_\infty^{2q} d\omega \right)^{1/2q} \\ &< \infty,\end{aligned}$$

which finishes the proof. □

B.0.4.3 Approximate Identities for Bochner Spaces, and Cesaro-sums of Fourier Series

Let $\mathbb{T} = [-\pi,\pi]$ denote the unit circle, viewed as the quotient group $\mathbb{R}/2\pi\mathbb{Z}$, i.e. $\alpha + 2k\pi = \alpha \in \mathbb{T}$ for all $\alpha \in \mathbb{T}$, let B be a Banach space, and the measure $\mu = \lambda$, the Lebesgue measure. For a function $K \in L^1(\mathbb{T},\mathbb{C})$, and a function $f \in L^p(\mathbb{T},B)$, $p \in [1,\infty]$, we define the convolution $K * f : \mathbb{T} \to B$ by

$$(K * f)(\alpha) = \int_{-\pi}^{\pi} K(\omega) f(\alpha - \omega) d\omega, \quad \alpha \in \mathbb{T}. \tag{B.0.11}$$

Notice that $K * f$ is strongly measurable, and that by Young's inequality (Hunter & Nachtergaele 2001, Theorem 12.58),

$$\left(\int_{-\pi}^{\pi} \left\| K * f(\omega) \right\|^p d\omega\right)^{1/p} \leq \int_{-\pi}^{\pi} |K(\omega)| d\omega \left(\int_{-\pi}^{\pi} \left\| f \right\|^p d\omega\right)^{1/p},$$

in other words, $\left\| K * f \right\|_{L^p} \leq \|K\|_{L^1} \left\| f \right\|_{L^p}$, and the mapping $f \mapsto K * f$ is a continuous and linear mapping $L^p(\mathbb{T}, B) \to L^p(\mathbb{T}, B)$. Furthermore, a simple change of variable yields $K * f = f * K$.
A sequence of functions $(K_n)_{n\geq 1} \in L^1(\mathbb{T}, \mathbb{C})$ is called an *approximate identity* if

1. $\sup_n \|K_n\|_{L^1} < \infty$,
2. $\lim_{n\to\infty} \int_{-\pi}^{\pi} K_n(\omega) d\omega = 1$
3. $\lim_{n\to\infty} \int_{\delta<|x|<\pi} |K_n(\omega)| d\omega = 0$, for all $0 < \delta < \pi$.

Approximate identities are important in real Fourier analysis, but also for us, because their nice properties extend to Bochner spaces. Let us introduce the notation $C(\mathbb{T}, B)$ to denote continuous strongly measurable functions $f : \mathbb{T} \to B$. The following result is proved by following the proof of Edwards (1967, Theorem 3.2.2), by replacing the modulus by $\|\cdot\|$ when appropriate:

Proposition B.0.15. *Let $(K_n)_{n\geq 1}$ be an approximate identity. Then*

$$\lim_{n\to\infty} \sup_{\omega\in\mathbb{T}} \left\| K_n * f(\omega) - f(\omega) \right\| = 0, \quad f \in C(\mathbb{T}, B), \tag{B.0.12}$$

and

$$\lim_{n\to\infty} \left\| K_n * f - f \right\|_p = 0, \quad f \in L^p(\mathbb{T}, B), \tag{B.0.13}$$

provided $p \in [1, \infty)$.

A concrete example of a approximate identity is the *Fejér kernel*,

$$F_N(\omega) = \frac{1}{2\pi N} \left(\frac{\sin(N\omega/2)}{\sin(\omega/2)}\right)^2 = (2\pi)^{-1} \sum_{|n|<N} \left(1 - \frac{|n|}{N}\right) e^{\mathbf{i}\omega n}, \tag{B.0.14}$$

which is important because it is related to the Cesaro-sum of a Fourier series.

Proposition B.0.16. *Let $p \in [1, \infty)$. For $f \in L^p(\mathbb{T}, B)$, let*

$$\hat{f}(n) = (2\pi)^{-1} \int_{-\pi}^{\pi} f(\omega) e^{-\mathbf{i}\omega n} d\omega \in B,$$

and define

$$\sigma_N f(\omega) := \sum_{|n|<N} \left(1 - \frac{|n|}{N}\right) \hat{f}(n) e^{\mathbf{i}\omega n}.$$

Then,

$$\left\|\sigma_N f - f\right\|_{L^p} \to 0, \quad N \to \infty. \tag{B.0.15}$$

Furthermore, if $f, g \in L^p(\mathbb{T}, B)$*,* $p \in [1, \infty]$*, and they satisfy*

$$\int_{-\pi}^{\pi} f(\omega) e^{\mathbf{i}\omega n} d\omega = \int_{-\pi}^{\pi} g(\omega) e^{\mathbf{i}\omega n} d\omega, \quad \forall n \in \mathbb{Z}, \tag{B.0.16}$$

then $f = g$ *almost everywhere, and if they are both continuous,* $f = g$ *everywhere.*

Proof. Notice that

$$\sigma_N f(\omega) = \sum_{|n|<N} \left(1 - \frac{|n|}{N}\right) \hat{f}(n) e^{\mathbf{i}\omega n} = F_N * f(\omega),$$

and by Proposition B.0.15 implies (B.0.15).
For (B.0.16), since $L^p(\mathbb{T}, B) \subset L^1(\mathbb{T}, B)$ for any $1 < p \leq \infty$, we only need to prove the result for $p = 1$. Notice that $\hat{f}(n) - \hat{g}(n) = 0$ for all n. Therefore, applying (B.0.15) to the function $h = f - g$ yields $\|\sigma_N h - h\|_{L^1} \to 0$. But since $\sigma_N h = 0$, the left-hand side is $\|h\|_{L^1}$, and does not depend on N. Therefore $\|h\|_{L^1} = 0$ and therefore $f = g$ almost surely. □

B.0.4.4 Fourier Series for Hilbert Space Valued Functions

Let us now consider the Bochner space $L^2(\mathbb{T}, H)$, where $\mathbb{T} = \mathbb{R}/2\pi\mathbb{Z}$ is defined on page 233, and H is a complex separable Hilbert space. We define

$$\langle A_1, A_2 \rangle_{L^2} = \int_{-\pi}^{\pi} \langle A_1(\omega), A_2(\omega) \rangle d\omega,$$

for $A_1, A_2 \in L^2(\mathbb{T}, H)$. This turns the space $L^2(\mathbb{T}, H)$ into a separable Hilbert space.
The following Proposition tells us that Fourier series for the space $L^2(\mathbb{T}, H)$ enjoy the same L^2 properties as Fourier series for functions in $L^2(\mathbb{T}, \mathbb{C})$.

Proposition B.0.17. *For any* $A \in L^2(\mathbb{T}, H)$*, we have*

$$\lim_{N\to\infty} \int_{-\pi}^{\pi} \left\| A(\omega) - \sum_{|n|<N} e^{\mathbf{i}\omega n} \left(\frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-\mathbf{i}\alpha n} A(\alpha) d\alpha \right) \right\|^2 d\omega = 0$$

Remark B.0.18. *For any* $A \in L^2(\mathbb{T}, H)$*, we have Denoting the Fourier coefficients*

$$a_n = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{-\mathbf{i}\alpha n} A(\alpha) d\alpha \in H,$$

and the truncated Fourier sums

$$S_N = \sum_{|n|<N} a_n \mathfrak{e}_n,$$

where $\mathfrak{e}_n(\omega) = \exp(\mathbf{i}\omega n)$*, the statement of the Proposition can be rewritten in the following compact form:*

$$\|A - S_N\|_{L^2} \to 0, \quad \text{as } N \to \infty.$$

Proof of Proposition B.0.17. We shall use the notation of Remark B.0.18.

Let us first show that $(S_N)_{N\geq 1}$ is a Cauchy sequence in $L^2(\mathbb{T}, H)$. Notice that for all $N, M \geq 1$,

$$\begin{aligned}
\langle S_N, S_M\rangle_{L^2} &= \sum_{|n|<N}\sum_{|m|<M} \langle a_n\mathfrak{e}_n, a_m\mathfrak{e}_m\rangle_{L^2} \\
&= \sum_{|n|<N}\sum_{|m|<M} \int_{-\pi}^{\pi} \langle a_n, a_m\rangle \exp\left[\mathbf{i}(n-m)\omega\right] d\omega \\
&= \sum_{|n|<N}\sum_{|m|<M} \langle a_n, a_m\rangle \int_{-\pi}^{\pi} \exp\left[\mathbf{i}(n-m)\omega\right] d\omega \\
&= \sum_{|n|<N}\sum_{|m|<M} \langle a_n, a_m\rangle 2\pi\delta_{n,m} \\
&= \sum_{|n|<N}\sum_{|m|<M} \langle a_n, a_m\rangle 2\pi\delta_{n,m} \\
&= 2\pi \sum_{|n|<\min(N,M)} \|a_n\|^2.
\end{aligned}$$

Therefore, if $N > M \geq 1$,

$$\|S_N - S_M\|_{L^2}^2 = \sum_{M\leq|n|<N} \|a_n\|^2.$$

Let $(\varphi_n)_{n\geq 1}$ be an orthonormal basis of H. We have

$$\begin{aligned}
\sum_{n\in\mathbb{Z}} \|a_n\|^2 &= \sum_{n\in\mathbb{Z}}\sum_{k\geq 1} |\langle a_n, \varphi_k\rangle|^2 \\
&= \sum_{n\in\mathbb{Z}}\sum_{k\geq 1} \left|\left\langle \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-\mathbf{i}\alpha n} A(\alpha)d\alpha, \varphi_k\right\rangle\right|^2 \\
&= \sum_{k\geq 1}\sum_{n\in\mathbb{Z}} \left|\frac{1}{2\pi}\int_{-\pi}^{\pi} e^{-\mathbf{i}\alpha n}\langle A(\alpha), \varphi_k\rangle d\alpha\right|^2 \\
&= \sum_{k\geq 1} \int_{-\pi}^{\pi} |\langle A(\omega), \varphi_k\rangle|^2 d\omega \qquad \text{(by Parseval's identity)} \\
&= \int_{-\pi}^{\pi} \sum_{k\geq 1} |\langle A(\omega), \varphi_k\rangle|^2 d\omega \\
&= \int_{-\pi}^{\pi} \|A(\omega)\|^2 d\omega.
\end{aligned}$$

Therefore $(S_N)_{N\geq 1}$ is a Cauchy sequence, and converges to an element $S \in L^2(\mathbb{T}, H)$.

Now notice that for any $\varphi \in H$, we have

$$\int_{-\pi}^{\pi} |\langle A(\omega) - S_N(\omega), \varphi \rangle|^2 d\omega \to 0, \quad \text{as } N \to \infty.$$

Indeed, the function $\langle A(\cdot), \varphi \rangle$ is in $L^2([0,1], \mathbb{C})$, and

$$\langle S_N(\cdot), \varphi \rangle = \sum_{|n|<N} \mathfrak{e}_n \langle a_n, \varphi \rangle$$

is its truncated Fourier series, which converges in $L^2([0,1], \mathbb{C})$ to $\langle A(\cdot), \varphi \rangle$, from the classical theory of Fourier series (see e.g. Edwards (1967)). Therefore, we get

$$\int_{-\pi}^{\pi} |\langle A(\omega) - S(\omega), \varphi \rangle|^2 d\omega = \lim_{N\to\infty} \int_{-\pi}^{\pi} |\langle A(\omega) - S_n(\omega), \varphi \rangle|^2 d\omega = 0,$$

for all $\varphi \in H$. This implies that $S = A$ almost everywhere (indeed, replace φ by φ_m; for each $m \geq 1$, $A \neq S$ on a set of measure zero; the statement follows since a countable union of measure zero sets has measure zero) and finishes the proof. □

APPENDIX C

Random Elements in Banach and Hilbert Space

C.1 Generalities

In the special case where μ is a probability measure, denoted $\mu = \mathbb{P}$, the Bochner integral allows us to define rigorously the notion of *random elements* of a Banach space.

Definition C.1.1. *Let $(\Omega, \mathcal{O}, \mathbb{P})$ be a complete probability space and B be a Banach space. A random element of B is a strongly measurable function $X : \Omega \to B$. If $\int_\Omega \|X\| d\mathbb{P} < \infty$, we write*

$$\mathbb{E}\, X = \int_\Omega X d\mathbb{P}$$

In other words, if a random element X of B satisfies $\mathbb{E}\|X\| < \infty$, then its expectation is well defined, and satisfies $T\,\mathbb{E}\, X = \mathbb{E}\, TX$ for every bounded operator $T : B \to B'$, where B' is another Banach space. In the particular case where $B = H$ is a separable Hilbert space, if X, Y are random elements of H with mean zero and finite second moment ($\mathbb{E}\|X\|^2 + \mathbb{E}\|Y\|^2 < \infty$), then the cross-covariance operator of X and Y,

$$\mathscr{R}_{XY} = \mathbb{E}[X \otimes_2 Y]$$

is a well defined, and is a nuclear operator since

$$|||\mathscr{R}_{XY}|||_1 \leq \mathbb{E}|||X \otimes_2 Y|||_1 = \mathbb{E}\|X\|\|Y\| \leq \sqrt{\mathbb{E}\|X\|^2\, \mathbb{E}\|Y\|^2} < \infty.$$

Indeed, the first inequality comes from the contraction property of the Bochner integral, and the second inequality comes from the Cauchy-Schwarz inequality. In particular, since the trace is a linear operator,

$$\mathrm{Tr}\,(\mathscr{R}_{XY}) = \mathbb{E}\,\mathrm{Tr}\,(X \otimes_2 Y) = \mathbb{E}\,\langle X, Y\rangle.$$

Furthermore, if T, S are bounded operators on H, then

$$\left(T \,\widetilde{\bigotimes}_2\, S\right) \mathbb{E}\,[X \otimes_2 Y] = \mathbb{E}\,[TX \otimes_2 SY], \tag{C.1.1}$$

since $(T \,\widetilde{\otimes}_2\, S)$ is a bounded operator on $\mathscr{S}_1(H)$ (by Hölder's inequality), and therefore

$$\mathbb{E}\,\langle TX, SY\rangle = \mathrm{Tr}\left(T\mathscr{R}_{XY}S^{\dagger}\right) \tag{C.1.2}$$

If $X = Y$, then the cross-covariance operator is simply called the *covariance operator*, denoted $\mathscr{R} = \mathbb{E}\,[X \otimes_2 X]$. It is a hermitian positive operator, which is trace-class, with trace

$$\mathrm{Tr}\,(\mathscr{R}) = \mathbb{E}\,\mathrm{Tr}\,(X \otimes_2 X) = \mathbb{E}\,\|X\|^2.$$

The following technical result will be useful for computing the cross-covariance operator of mean square limits of random elements.

Lemma C.1.2. *Let $(X_n)_{n\geq 1}$, respectively $(Y_n)_{n\geq 1}$ be sequences of random elements of H with finite second moment, converging in mean square to X, respectively Y, i.e.*

$$\mathbb{E}\,\|X_n - X\|^2 \to 0, \quad \mathbb{E}\,\|Y_n - Y\|^2 \to 0, \text{ as } n \to \infty.$$

Then the cross covariance operator of X and Y is well defined and

$$\mathbb{E}\,[X \otimes_2 Y] = \lim_{n\to\infty} \mathbb{E}\,[X_n \otimes_2 Y_n], \quad \text{in } \mathscr{S}_1(H).$$

Proof. Recall that $\|X\|_{L^2} = \sqrt{\mathbb{E}\,\|X\|^2}$ defines a norm on $L^2(\Omega, H, \mathbb{P})$. Therefore X and Y have both finite moments, and their cross-covariance operator is well-defined. Now since

$$\mathbb{E}\,[X \otimes_2 Y] - \mathbb{E}\,[X_n \otimes_2 Y_n] = \mathbb{E}\,[(X - X_n) \otimes_2 Y] + \mathbb{E}\,[X_n \otimes_2 (Y - Y_n)],$$

the contraction property and (A.2.8) yield the result. □

The following well-known result relates the trace of the covariance operator of a random element of an L^2 space with an integral of its covariance kernel:

Lemma C.1.3. *Let $K = \prod_{j=1}^{n}[a_j, b_j] \subset \mathbb{R}^n$, where $-\infty < a_j < b_j < \infty$ for $j = 1, \ldots, n$. Let μ denote Lebesgue measure on $\mathbb{R}^n$. Let X be a random element of*

$$L^2(K, \mathbb{C}) = \left\{f : K \to \mathbb{C} : \left\|f\right\|_2 < \infty\right\},$$

where $\|f\|_2 = \sqrt{\langle f, f\rangle}$ *and*

$$\langle f, g\rangle = \int_K f\overline{g}\, d\mu, \quad f, g \in L^2(K, \mathbb{C}).$$

Let $r(t,s) = \operatorname{cov}(X(t), X(s)), \quad t, s \in K$ *be the covariance kernel of* X*, and* $\mathscr{R}$ *be the operator on* $L^2(K,\mathbb{C})$ *induced by the kernel. If* $\mathbb{E}\|X\|_2^2 < \infty$*, then* $\mathscr{R}$ *is trace-class and*

$$\int_K r(t,t)\, d\mu(t) = \operatorname{Tr}(\mathscr{R}) < \infty.$$

Proof. We already know that $\operatorname{Tr}(\mathscr{R}) = \mathbb{E}\|X - \mathbb{E}X\|^2 < \infty$. On the other hand, Tonelli's Theorem yields

$$\mathbb{E}\|X - \mathbb{E}X\|_2^2 = \int_K \mathbb{E}|X(t) - \mathbb{E}X(t)|^2\, dt = \int_K r(t,t)\, d\mu(t).$$

Hence $\int_K r(t,t)\, d\mu(t) = \operatorname{Tr}(\mathscr{R}) < \infty$. □

C.2 Convergence in Distribution

In this section, we briefly talk about convergence in distributions for random elements of Banach or Hilbert space. Some references are Vakhania et al. (1987), Ledoux & Talagrand (2011), Billingsley (1999), and Kallenberg (1997).

Recall that a sequence of random elements (X_n) of a separable Banach space B is said to converge in distribution to the random element $X \in B$ if $\mathbb{P} \circ X_n^{-1}$ converges weakly to $\mathbb{P} \circ X^{-1}$. The following characterization of convergence in distribution is very useful in practice:

Theorem C.2.1 (e.g. Ledoux & Talagrand (2011))**.** *A sequence of random elements* (X_n) *of a separable Banach space* B *converges in distribution to the random element* $X \in B$ *if and only if*

Convergence of Projections: $\phi(X_n) \xrightarrow{d} \phi(X)$ *for all* $\phi \in B^*$, and,

Tightness: (X_n) *is tight, i.e.*

$$\sup_K \liminf_{n\to\infty} \mathbb{P}X_n \in K = 1,$$

where the supremum is taken over all compact sets $K \subset B$*, and*

The tightness condition is *crucial* in proving convergence in distribution. It is needed because otherwise we could have $\mathbb{P}X \in B < 1$, a phenomenon that can be intuitively understood as that "mass escaped to infinity". The convergence of the projections ensures that the limiting distribution is indeed given by the law of X.

Since tightness is related to compact subsets, it is important to have a characterization of all compact subsets of a given space, or at least to

be able to construct "large" compact subset. For the Banach spaces of continuous real functions on compact Hausdorff spaces, such a characterization is given by the Arzelà-Ascoli Theorem. In the case where $B = H$ is a separable Hilbert space, I haven't found an explicit characterization of compact subset in the literature. If $(e_n)_{n\geq 1}$ is an orthonormal basis of a Hilbert space H, it can be shown that for every $b = (b_1, b_2, \ldots) \in \ell_2(\mathbb{R})$, the set

$$\left\{ \sum_{n\geq 1} a_n e_n : |a_n| \leq |b_n| \, \forall \, n \geq 1 \right\} \tag{C.2.1}$$

is a compact subset of H. This following tells us how to construct compact subsets that are larger that this one, because they will always contain the sets elements of the form (C.2.1), and also elements of the form $x = \alpha_n e_n$, where $\alpha_n \to 0$, but possibly with $\sum_{n\geq 1} |\alpha_n|^2 = \infty$.

Lemma C.2.2 (A class of compact sets for separable Hilbert spaces)**.** *Let H be a separable Hilbert space with scalar product $\langle \cdot, \cdot \rangle$ and norm $\|\cdot\|_2$. For any sequence of integers $1 = n_1 < n_2 < \cdots$, any sequence of positive numbers l_k such that $\lim_{k\to\infty} l_k = +\infty$, and any complete orthonormal sequence $(e_n)_{n=1,2,\ldots}$ of H, the set $K = \bigcap_{k=1}^{\infty} B_k$ is compact, where $B_k = \left\{ x \in H : \sum_{j=n_k}^{\infty} \langle x, e_j \rangle^2 \leq l_k^{-1} \right\}$.*

Proof. Suppose H is a real separable Hilbert space (all the following steps can be reproduced for a complex separable Hilbert space). The sequence (e_n) induces an isometric isomorphism $H \to \ell_2(\mathbb{R})$ via

$$x \in H \longmapsto (\langle x, e_1 \rangle, \langle x, e_2 \rangle, \ldots).$$

We can therefore write $x = (x_1, x_2, \ldots)$, where $x_j = \langle x, e_j \rangle$.
Since K is in particular a metric space, showing its compactness is equivalent to showing that it is complete and totally bounded (Munkres 2000, Theorem 45.1). Recall that a metric space (M, d) is totally bounded if, for every $\varepsilon > 0$, there exists a *finite* covering of M by ε-balls. This means that there exists a subset $F_\varepsilon \subset M$ such that

$$\text{for any point of } m \in M, \text{ there is a point } p \in F_\varepsilon \text{ with } d(m, p) < \varepsilon. \tag{C.2.2}$$

The set F_ε is called a (finite) ε-net of M.
First notice that K is complete since it is a closed subset of H. Let us show that is totally bounded. Fix $\varepsilon > 0$, and let k be the smallest integer such that $l_k^{-1} < \varepsilon^2/2$. Let

$$H_k = \left\{ (x_1, x_2, \ldots) \in H : x_j = 0 \text{ for } j > k \right\}.$$

Since the set

$$H_k(l_1^{-1/2}) = \left\{ x \in H_k : \|x\|_2 < l_1^{-1/2} \right\}$$

is compact, it is totally bounded, and there exists a finite $\frac{\varepsilon}{\sqrt{2}}$-net F of

$H_k(l_1^{-1/2})$. Let us show that F is a (finite) ε-net of K. Let $x = (x_1, x_2, \ldots) \in K$, $P_k : H \to H$ denote the orthogonal projection onto H_k, and $I : H \to I$ be the identity operator on H. Notice that

$$P_k(K) \subset P_k(B_1) = H_k(l_1^{-1/2}),$$

hence there exists a point $p \in F$ such that $\|P_k x - p\|_2^2 < \frac{\varepsilon^2}{2}$. Notice also that $\|(I - P_k)x\|_2^2 < l_k^{-1} < \frac{\varepsilon^2}{2}$ by definition of B_k. Hence

$$\|x - p\|_2^2 = \|P_k(x-p)\|_2^2 + \|(I-P_k)(x-p)\|_2^2 = \|P_k x - p\|_2^2 + \|(I-P_k)x\|_2^2,$$

since $p \in H_k$. We thus have $\|x - p\|_2 < \varepsilon$, and K is compact. □

The following Lemma gives necessary conditions for tightness of a sequence of random elements in a separable Hilbert space. It collects ideas found in Bosq (2000, Theorem 2.7). We note that it is slightly weaker than Panaretos & Tavakoli (2013*b*, Lemma 7.1).

Lemma C.2.3 (Criterion for tightness in Hilbert Space)**.** *Let H be a (real or complex) separable Hilbert Space, and $X_T : \omega \to H, \quad T = 1, 2, \ldots$ be a sequence of random variables. If for some fixed orthonormal basis $(e_n)_{n\geq 1}$ of H, and some $T' > 1$, we have*

1. $\sup_{T>T'} \mathbb{E}\|X_T\|^2 < \infty$
2. $\lim_{n\to\infty} \sup_{T>T'} \sum_{j\geq n} \mathbb{E}\left|\langle X_T, e_j\rangle\right|^2 = 0,$

then (X_T) is tight.

Proof. Fix $\varepsilon > 0$. We shall define a compact set $K \subset H$ such that $\mathbb{P}(X_T \notin K) \leq \varepsilon$, for all $T > T'$. This will show that X_T is tight.
Set $S_n = \sup_{T>T'} \sum_{j\geq n} \mathbb{E}|\langle X_T, e_j\rangle|^2$. By assumption, $S_1 < \infty$ and

$$\lim_{n\to\infty} S_n = 0.$$

Set $n_1 = 1$, and $l_1 = \varepsilon/(2S_1)$. We then define $l_k = kl_1$ and choose integers $1 < n_2 < n_3 < \cdots$ such that

$$S_{n_k} \leq \frac{S_1}{k2^{k-1}}.$$

Define $B_k = \left\{x \in H : \sum_{j=n_k}^{\infty} \langle x, \varphi_j\rangle^2 \leq l_k^{-1}\right\}$, and $K = \bigcap_{k=1}^{\infty} B_k$, which is compact by Lemma C.2.2. Using successively the union bound and

Markov's inequality, we obtain, for all $T > T'$,

$$\begin{aligned}\mathbb{P}[X_T \notin K] &\leq \sum_{k\geq 1} \mathbb{P}\left[\sum_{j\geq n_k} \left|\langle X_T, e_j\rangle\right|^2 \geq l_k^{-1}\right] \\ &\leq \sum_k l_k \sum_{j\geq n_k} \mathbb{E}\left|\langle X_T, e_j\rangle\right|^2 \\ &\leq \sum_k l_k S_{n_k} \leq \sum_k \frac{\varepsilon}{2^k} = \varepsilon.\end{aligned}$$

□

Bibliography

Adler, R. J. (1990), *An introduction to continuity, extrema, and related topics for general Gaussian processes*, Institute of Mathematical Statistics Lecture Notes—Monograph Series, 12, Institute of Mathematical Statistics, Hayward, CA. (see pages 124, 173 and 182)

Adler, R. J. (2000), 'On excursion sets, tube formulas and maxima of random fields', *Annals of Applied Probability* **10**(1), 1–74.
URL: *http://dx.doi.org/10.1214/aoap/1019737664* (see page 182)

Amzallag, A., Vaillant, C., Jacob, M., Unser, M., Bednar, J., Kahn, J. D., Dubochet, J., Stasiak, A. & Maddocks, J. H. (2006), '3d reconstruction and comparison of shapes of DNA minicircles observed by cryo-electron microscopy', *Nucleic Acids Research* **34**(18), e125.
URL: *http://nar.oxfordjournals.org/content/34/18/e125.abstract* (see pages 153 and 154)

Anderson, T. (1994), *The statistical analysis of time series*, Wiley, New York. (see page 130)

Andrews, D. W. K. (1984), 'Nonstrong mixing autoregressive processes', *J. Appl. Probab.* **21**(4), 930–934. (see page 18)

Antoniadis, A., Paparoditis, E. & Sapatinas, T. (2006), 'A functional wavelet-kernel approach for time series prediction', *Journal of the Royal Statistical Society series B* **68**(5), 837–857.
URL: *http://dx.doi.org/10.1111/j.1467-9868.2006.00569.x* (see page 18)

Antoniadis, A., Paparoditis, E. & Sapatinas, T. (2009), 'Bandwidth selection for functional time series prediction', *Statistics & Probability Letters* **79**(6), 733 – 740. (see page 17)

Antoniadis, A. & Sapatinas, T. (2003), 'Wavelet methods for continuous-time prediction using Hilbert-valued autoregressive processes', *Journal of Multivariate Analysis* **87**(1), 133–158.
URL: *http://dx.doi.org/10.1016/S0047-259X(03)00028-9* (see page 17)

Ash, R. B. & Gardner, M. F. (1975), *Topics in stochastic processes*, Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London. Probability and Mathematical Statistics, Vol. 27. (see pages 9 and 10)

Aston, J. A. D., Chiou, J.-M. & Evans, J. P. (2010), 'Linguistic pitch analysis using functional principal component mixed effect models', *J. R. Stat. Soc. Ser. C. Appl. Stat.* **59**(2), 297–317.
URL: *http://dx.doi.org/10.1111/j.1467-9876.2009.00689.x* (see page 14)

Aston, J. A. D. & Kirch, C. (2012*a*), 'Detecting and estimating changes in dependent functional data', *Journal of Multivariate Analysis* **109**, 204–220.
URL: *http://dx.doi.org/10.1016/j.jmva.2012.03.006* (see pages 18 and 19)

Aston, J. A. D. & Kirch, C. (2012*b*), 'Evaluating stationarity via change-point alternatives with applications to fMRI data', *Annals of Applied Statistics* **6**(4), 1906–1948.
URL: *http://dx.doi.org/10.1214/12-AOAS565* (see pages 17, 18 and 19)

Aue, A., Hörmann, S., Horváth, L. & Hušková, M. (2014, to appear), 'Dependent functional linear models with applications to monitoring structural change', *Statistica Sinica* . (see page 19)

Aue, A., Norinho, D. D. & Hörmann, S. (2014, to appear), 'On the prediction of stationary functional time series', *Journal of the American Statistical Association* .
URL: *http://dx.doi.org/10.1080/01621459.2014.909317* (see pages 19 and 66)

Benjamini, Y. & Bogomolov, M. (2014), 'Selective inference on multiple families of hypotheses', *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **76**(1), 297–318.
URL: *http://dx.doi.org/10.1111/rssb.12028* (see page 180)

Benjamini, Y. & Hochberg, Y. (1995), 'Controlling the false discovery rate: a practical and powerful approach to multiple testing', *J. Roy. Statist. Soc. Ser. B* **57**(1), 289–300. (see page 171)

Benjamini, Y. & Yekutieli, D. (2001), 'The control of the false discovery rate in multiple testing under dependency', *Annals of Statistics* **29**(4), 1165–1188.
URL: *http://dx.doi.org/10.1214/aos/1013699998* (see page 181)

Benko, M., Härdle, W. & Kneip, A. (2009), 'Common functional principal components', *Annals of Statistics* **37**(1), 1–34.
URL: *http://dx.doi.org/10.1214/07-AOS516* (see page 13)

Besse, P. C., Cardot, H. & Stephenson, D. B. (2000), 'Autoregressive forecasting of some functional climatic variations', *Scand. J. Statist.* **27**(4), 673–687.
URL: *http://dx.doi.org/10.1111/1467-9469.00215* (see page 17)

Billingsley, P. (1999), *Convergence of probability measures*, Wiley Series in Probability and Statistics: Probability and Statistics, second edn, John Wiley & Sons, Inc., New York. A Wiley-Interscience Publication.
URL: *http://dx.doi.org/10.1002/9780470316962* (see page 241)

Blackman, R. B. & Tukey, J. W. (1959), *The measurement of power spectra: From the point of view of communications engineering*, Dover Publications, Inc., New York. (see page 83)

Bloomfield, P. (2000), *Fourier analysis of time series*, Wiley Series in Probability and Statistics: Applied Probability and Statistics, second edn, Wiley-Interscience [John Wiley & Sons], New York. An introduction.
URL: *http://dx.doi.org/10.1002/0471722235* (see page 83)

Bosq, D. (2000), *Linear processes in function spaces*, Vol. 149 of *Lecture Notes in Statistics*, Springer-Verlag, New York. Theory and applications.
URL: *http://dx.doi.org/10.1007/978-1-4612-1154-9* (see pages 9, 12, 15, 17, 155, 169, 216 and 243)

Bosq, D. & Blanke, D. (2007), *Inference and prediction in large dimensions*, Wiley Series in Probability and Statistics, John Wiley & Sons, Ltd., Chichester; Dunod, Paris.
URL: *http://dx.doi.org/10.1002/9780470724033* (see page 17)

Bradley, R. C. (2007*a*), *Introduction to strong mixing conditions. Vol. 1*, Kendrick Press, Heber City, UT. (see page 18)

Bradley, R. C. (2007*b*), *Introduction to strong mixing conditions. Vol. 2*, Kendrick Press, Heber City, UT. (see page 18)

Bradley, R. C. (2007*c*), *Introduction to strong mixing conditions. Vol. 3*, Kendrick Press, Heber City, UT. (see page 18)

Brillinger, D. (2012), *Selected works of David Brillinger*, Selected Works in Probability and Statistics, Springer, New York. Edited by Peter Guttorp and Brillinger.
URL: *http://dx.doi.org/10.1007/978-1-4614-1344-8* (see page 1)

Brillinger, D. R. (2001), *Time Series: Data Analysis and Theory*, Vol. 36 of *Classics in Applied Mathematics*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA. Data analysis and theory, Reprint of the 1981 edition.
URL: *http://dx.doi.org/10.1137/1.9780898719246* (see pages 23, 56, 57, 76, 78, 83, 97, 125, 130, 162, 163, 169 and 174)

Cai, T. T. & Hall, P. (2006), 'Prediction in functional linear regression', *Annals of Statistics* **34**(5), 2159–2179.
URL: *http://dx.doi.org/10.1214/009053606000000830* (see page 14)

Cai, T. T. & Yuan, M. (2011), 'Optimal estimation of the mean function based on discretely sampled functional data: phase transition', *Annals of Statistics* **39**(5), 2330–2355.
URL: *http://dx.doi.org/10.1214/11-AOS898* (see page 13)

Cai, T. T. & Yuan, M. (2012), 'Minimax and adaptive prediction for functional linear regression', *Journal of the American Statistical Association* **107**(499), 1201–1216.
URL: *http://dx.doi.org/10.1080/01621459.2012.716337* (see page 14)

Cardot, H., Ferraty, F., Mas, A. & Sarda, P. (2003), 'Testing hypotheses in the functional linear model', *Scandinavian Journal of Statistics* **30**(1), 241–255.
URL: *http://dx.doi.org/10.1111/1467-9469.00329* (see page 14)

Cardot, H., Ferraty, F. & Sarda, P. (1999), 'Functional linear model', *Statistics & Probability Letters* **45**(1), 11–22. (see page 14)

Cardot, H., Ferraty, F. & Sarda, P. (2003), 'Spline estimators for the functional linear model', *Statistica Sinica* **13**(3), 571–591. (see page 14)

Chen, D. & Müller, H.-G. (2012), 'Nonlinear manifold representations for functional data.', **40**(1), 1–29. (see page 8)

Ciarlet, P. G. (2013), *Linear and nonlinear functional analysis with applications*, Society for Industrial and Applied Mathematics, Philadelphia, PA. (see pages 213 and 214)

Conway, J. B. (1990), *A course in functional analysis*, Vol. 96 of *Graduate Texts in Mathematics*, second edn, Springer-Verlag, New York. (see page 221)

Cooley, J. W. & Tukey, J. W. (1965), 'An algorithm for the machine calculation of complex Fourier series', *Math. Comp.* **19**, 297–301. (see page 83)

Crambes, C. & Mas, A. (2013), 'Asymptotics of prediction in functional linear regression with functional outputs', *Bernoulli* **19**(5B), 2627–2651.
URL: *http://dx.doi.org/10.3150/12-BEJ469* (see page 14)

Cugliari, J. (2013), 'Conditional Autoregressive Hilbertian processes', *ArXiv e-prints* . (see page 17)

Damon, J. & Guillas, S. (2005), 'Estimation and simulation of autoregressive Hilbertian processes with exogenous variables', *Stat. Inference Stoch. Process.* **8**(2), 185–204.
URL: *http://dx.doi.org/10.1007/s11203-004-1031-6* (see page 17)

Dauxois, J., Pousse, A. & Romain, Y. (1982), 'Asymptotic theory for the principal component analysis of a vector random function: Some applications to statistical inference.', *Journal of Multivariate Analysis* **12**, 136–154. (see page 13)

Diestel, J., Fourie, J. H. & Swart, J. (2008), *The metric theory of tensor products*, American Mathematical Society, Providence, RI. Grothendieck's résumé revisited. (see page 225)

Doukhan, P. (1994), *Mixing: Properties and examples.*, New York: Springer-Verlag. (see page 17)

Dryden, I., Kume, A., Le, H., Wood, A. & Laughton, C. (2002), Size-and-shape analysis of DNA molecular dynamics simulations, *in* R. Aykroyd, K. Mardia & P. McDonnell, eds, 'Statistics of large datasets', Department of Statistics, University of Leeds, pp. 23–26. (see page 154)

Dudoit, S., Shaffer, J. P. & Boldrick, J. C. (2003), 'Multiple hypothesis testing in microarray experiments', *Statistical Science* **18**(1), 71–103.
URL: *http://dx.doi.org/10.1214/ss/1056397487* (see page 171)

Duhamel, P. & Vetterli, M. (1990), 'Fast Fourier transforms: a tutorial review and a state of the art', *Signal Process.* **19**(4), 259–299.
URL: *http://dx.doi.org/10.1016/0165-1684(90)90158-U* (see page 83)

Dunford, N. & Schwartz, J. T. (1988*a*), *Linear operators. Part I*, Wiley Classics Library, John Wiley & Sons, Inc., New York. General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication. (see pages 213 and 214)

Dunford, N. & Schwartz, J. T. (1988*b*), *Linear operators. Part II*, Wiley Classics Library, John Wiley & Sons, Inc., New York. Spectral theory. Selfadjoint operators in Hilbert space, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1963 original, A Wiley-Interscience Publication. (see page 214)

Edwards, R. E. (1967), *Fourier series: A modern introduction. Vol. I*, Holt, Rinehard and Winston, Inc., New York-Montreal, Que.-London. (see pages 1, 87, 151, 152, 234 and 237)

Efromovich, S. (1999), *Nonparametric curve estimation*, Springer Series in Statistics, Springer-Verlag, New York. Methods, theory, and applications. (see page 11)

Einstein, A. (1914), 'Methode pour la determination de valeurs statistiques d'observations concernant des grandeurs soumises a des fluctuations irregulieres', *Archives des Sciences* **37**, 254–256. (see page 83)

Fan, J. & Gijbels, I. (1996), *Local polynomial modelling and its applications*, Vol. 66 of *Monographs on Statistics and Applied Probability*, Chapman & Hall, London. (see page 11)

Fan, J. & Yao, Q. (2003), *Nonlinear time series*, Springer Series in Statistics, Springer-Verlag, New York. Nonparametric and parametric methods.
URL: *http://dx.doi.org/10.1007/b97702* (see page 83)

Ferraty, F., ed. (2011), *Recent advances in functional data analysis and related topics*, Contributions to Statistics, Physica-Verlag/Springer, Heidelberg. Selected papers from the 2nd International Workshop on Functional and Operatorial Statistics (IWFOS'2011) held in Santander, June 16–18, 2011.
URL: *http://dx.doi.org/10.1007/978-3-7908-2736-1* (see pages 1 and 5)

Ferraty, F. & Romain, Y., eds (2011), *The Oxford handbook of functional data analysis*, Oxford University Press, Oxford. (see pages 5 and 155)

Ferraty, F. & Vieu, P. (2006), *Nonparametric Functional Data Analysis: Theory and Practice*, Springer. (see pages 5, 14 and 83)

Fourier, J. B. J. (1822), *Théorie analytique de la Chaleur*, chez Firmin Didot, Père et Fils, A Paris. (see page 1)

Gohberg, I. C. & Krejn, M. G. (1971), *Introduction à la théorie des opérateurs linéaires non auto-adjoints dans un espace hilbertien*, Dunod, Paris. Traduit du russe par Guy Roos, Monographies Universitaires de Mathématiques, No. 39. (see pages 214 and 216)

Gohberg, I., Goldberg, S. & Kaashoek, M. A. (1990), *Classes of linear operators. Vol. I*, Vol. 49 of *Operator Theory: Advances and Applications*, Birkhäuser Verlag, Basel.
URL: *http://dx.doi.org/10.1007/978-3-0348-7509-7* (see pages 214 and 216)

Gonzalez, O. & Maddocks, J. H. (2001), 'Extracting parameters for base-pair level models of DNA from molecular dynamics simulations', *Theoretical Chemistry Accounts* **106**(1-2), 76–82.
URL: *http://dx.doi.org/10.1007/s002140100256* (see page 154)

Gonzalez, O., Petkevičiūtė, D. & Maddocks, J. H. (2013), 'A sequence-dependent rigid-base model of dna', *The Journal of Chemical Physics* **138**(5), –.
URL: *http://scitation.aip.org/content/aip/journal/jcp/138/5/10.1063/1.4789411* (see page 153)

Grenander, U. (1981), *Abstract inference*, John Wiley & Sons, Inc., New York. Wiley Series in Probability and Mathematical Statistics. (see pages 5, 9 and 10)

Grenander, U. & Rosenblatt, M. (1957), *Statistical analysis of stationary time series*, Wiley, New York. (see pages 125 and 174)

Grothendieck, A. (1953), 'Résumé de la théorie métrique des produits tensoriels topologiques', *Bol. Soc. Mat. São Paulo* **8**, 1–79. (see page 225)

Hadjipantelis, P. Z., Aston, J. A. D. & Evans, J. P. (2012), 'Characterizing fundamental frequency in mandarin: A functional principal component approach utilizing mixed effect models', *The Journal of the Acoustical Society of America* **131**(6), 4651–4664.
URL: *http://scitation.aip.org/content/asa/journal/jasa/131/6/10.1121/1.4714345* (see page 14)

Hall, P. & Horowitz, J. L. (2007), 'Methodology and convergence rates for functional linear regression', *Annals of Statistics* **35**(1), 70–91.
URL: *http://dx.doi.org/10.1214/009053606000000957* (see page 14)

Hall, P. & Hosseini-Nasab, M. (2006), 'On properties of functional principal components analysis', *Journal of the Royal Statistical Society series B* **68**(1), 109–126.
URL: *http://dx.doi.org/10.1111/j.1467-9868.2005.00535.x* (see page 13)

Hall, P. & Vial, C. (2006), 'Assessing the finite dimensionality of functional data', *Journal of the Royal Statistical Society series B* **68**(4), 689–705.
URL: *http://dx.doi.org/10.1111/j.1467-9868.2006.00562.x* (see page 117)

Halmos, P. R. (1974), *Finite-dimensional vector spaces*, second edn, Springer-Verlag, New York-Heidelberg. Undergraduate Texts in Mathematics. (see page 221)

Hannan, E. J. (1970), *Multiple time series*, John Wiley and Sons, Inc., New York-London-Sydney. (see page 130)

He, G., Müller, H.-G., Wang, J.-L. & Yang, W. (2010), 'Functional linear regression via canonical analysis', *Bernoulli* **16**(3), 705–729.
URL: *http://dx.doi.org/10.3150/09-BEJ228* (see page 14)

He, G., Müller, H. & Wang, J. (2000), 'Extending correlation and regression from multivariate to functional data', *Asymptotics in statistics and probability* pp. 197–210. (see page 14)

Hildebrandt, T. H. (1953), 'Integration in abstract spaces', *Bull. Amer. Math. Soc.* **59**, 111–139. (see page 229)

Hilgert, N., Mas, A. & Verzelen, N. (2013), 'Minimax adaptive tests for the functional linear model', *Annals of Statistics* **41**(2), 838–869.
URL: *http://dx.doi.org/10.1214/13-AOS1093* (see page 14)

Hörmann, S. & Kokoszka, P. (2010), 'Weakly dependent functional data', *Annals of Statistics* **38**(3), 1845–1884. (see pages 18, 19, 99, 131 and 155)

Horváth, L., Hušková, M. & Kokoszka, P. (2010), 'Testing the stability of the functional autoregressive process', *Journal of Multivariate Analysis* **101**(2), 352–367.
URL: *http://dx.doi.org/10.1016/j.jmva.2008.12.008* (see page 17)

Horváth, L., Hušková, M. & Rice, G. (2013), 'Test of independence for functional data', *Journal of Multivariate Analysis* **117**, 100–119.
URL: *http://dx.doi.org/10.1016/j.jmva.2013.02.005* (see page 17)

Horváth, L. & Kokoszka, P. (2012), *Inference for functional data with applications*, Springer series in statistics, Springer, New York, NY. (see pages 5 and 17)

Horváth, L., Kokoszka, P. & Reeder, R. (2013), 'Estimation of the mean of functional time series and a two-sample problem', *J. R. Stat. Soc. Ser. B. Stat. Methodol.* **75**(1), 103–122.
URL: *http://dx.doi.org/10.1111/j.1467-9868.2012.01032.x* (see pages 18, 79, 99 and 155)

Hunter, J. K. & Nachtergaele, B. (2001), *Applied analysis*, World Scientific Publishing Co., Inc., River Edge, NJ.
URL: *http://dx.doi.org/10.1142/4319* (see pages 77 and 234)

James, G. M., Wang, J. & Zhu, J. (2009), 'Functional linear regression that's interpretable', *Annals of Statistics* **37**(5A), 2083–2108.
URL: *http://dx.doi.org/10.1214/08-AOS641* (see page 14)

Kadison, R. V. & Ringrose, J. R. (1997), *Fundamentals of the theory of operator algebras. Vol. II*, Vol. 16 of *Graduate Studies in Mathematics*, American Mathematical Society, Providence, RI. Advanced theory, Corrected reprint of the 1986 original. (see pages 91, 139 and 214)

Kahn, J. D. & Crothers, D. M. (1992), 'Protein-induced bending and DNA cyclization', *Proceedings of the National Academy of Science* **89**(14), 6343–6347. (see page 154)

Kallenberg, O. (1997), *Foundations of modern probability*, Probability and its Applications (New York), Springer-Verlag, New York. (see page 241)

Karhunen, K. (1947), 'Über lineare Methoden in der Wahrscheinlichkeitsrechnung', *Ann. Acad. Sci. Fennicae. Ser. A. I. Math.-Phys.* **1947**(37), 79. (see page 9)

King, A. A., Ionides, E. L., Bretó, C. M., Ellner, S., Kendall, B., Wearing, H., Ferrari, M. J., Lavine, M. & Reuman, D. C. (2010), *pomp: Statistical inference for partially observed Markov processes (R package)*. (see pages 159 and 178)

Kokoszka, P. (2012), 'Dependent functional data', *ISRN Probability and Statistics* **2012**. (see page 19)

Kokoszka, P. & Reimherr, M. (2013), 'Asymptotic normality of the principal components of functional time series', *Stochastic Processes and their Applications* **123**(5), 1546–1562.
URL: *http://dx.doi.org/10.1016/j.spa.2012.12.011* (see page 19)

Kuo, H.-H. (1996), *White noise distribution theory*, Probability and Stochastics Series, CRC Press, Boca Raton, FL. (see page 66)

Lankas, F., Gonzalez, O., Heffler, L. M., Stoll, G., Moakher, M. & Maddocks, J. H. (2009), 'On the parameterization of rigid base and basepair models of DNA from molecular dynamics simulations', *Phys. Chem. Chem. Phys.* **11**, 10565–10588.
URL: *http://dx.doi.org/10.1039/B919565N* (see page 154)

Leach, A. R. (2001), *Molecular modelling: principles and applications*, 2nd ed edn, Prentice Hall, Harlow, Essex. (see pages 154 and 155)

Lebesgue, H. (1904), 'Leçons sur l'intégration et la recherche de fonctions primitives.', Paris: Gauthier-Villars. VII u. 136 S. 8° (1904). (see page 4)

Ledoux, M. & Talagrand, M. (2011), *Probability in Banach spaces*, Classics in Mathematics, Springer-Verlag, Berlin. Isoperimetry and processes, Reprint of the 1991 edition. (see pages 78 and 241)

Lee, C.-K., Haralick, R. M. & Deguchi, K. (1993), Estimation of curvature from sampled noisy data, *in* 'Computer Vision and Pattern Recognition, 1993. Proceedings CVPR'93., 1993 IEEE Computer Society Conference on', IEEE, pp. 536–541. (see page 158)

Lee, T. C. M. (1997), 'A simple span selector for periodogram smoothing', *Biometrika* **84**(4), 965–969.
URL: *http://dx.doi.org/10.1093/biomet/84.4.965* (see page 162)

Lévy, P. (1948), *Processus Stochastiques et Mouvement Brownien. Suivi d'une note de M. Loève*, Gauthier-Villars, Paris. (see page 9)

Liu, W. & Wu, W. B. (2010), 'Asymptotics of spectral density estimates', *Econometric Theory* **26**(4), 1218–1245.
URL: *http://dx.doi.org/10.1017/S026646660999051X* (see page 131)

Liu, X. & Müller, H.-G. (2004), 'Functional convex averaging and synchronization for time-warped random curves', *Journal of the American Statistical Association* **99**(467), 687–699.
URL: *http://dx.doi.org/10.1198/016214504000000999* (see page 11)

Mas, A. (2002), 'Weak convergence for the covariance operators of a Hilbertian linear process', *Stochastic Processes and their Applications* **99**(1), 117–135.
URL: *http://dx.doi.org/10.1016/S0304-4149(02)00087-X* (see pages 15, 17 and 155)

Mas, A. (2007), 'Weak convergence in the functional autoregressive model', *Journal of Multivariate Analysis* **98**(6), 1231–1261.
URL: *http://dx.doi.org/10.1016/j.jmva.2006.05.010* (see page 15)

Mas, A. & Menneteau, L. (2003), Perturbation approach applied to the asymptotic study of random operators, *in* 'High dimensional probability, III (Sandjberg, 2002)', Vol. 55 of *Progr. Probab.*, Birkhäuser, Basel, pp. 127–134. (see pages 13, 17, 102 and 104)

Mas, A. & Pumo, B. (2011), Linear processes for functional data, *in* 'The Oxford handbook of functional data analysis', Oxford Univ. Press, Oxford, pp. 47–71. (see pages 15 and 17)

Mas, A. & Ruymgaart, F. (2014, to appear), 'High dimensional principal projections', *Complex Analysis and Operator Theory* . (see page 13)

Merlevède, F. (1996), 'Central limit theorem for linear processes with values in a Hilbert space', *Stochastic Processes and their Applications* **65**(1), 103–114.
URL: *http://dx.doi.org/10.1016/S0304-4149(96)00099-3* (see page 16)

Merlevède, F., Peligrad, M. & Utev, S. (1997), 'Sharp conditions for the CLT of linear processes in a Hilbert space', *J. Theoret. Probab.* **10**(3), 681–693.
URL: *http://dx.doi.org/10.1023/A:1022653728014* (see pages 15 and 17)

Mikusinski, P. & Weiss, E. (2014), 'The Bochner Integral', *ArXiv e-prints* . (see page 229)

Morris, J. S. & Carroll, R. J. (2006), 'Wavelet-based functional mixed models', *Journal of the Royal Statistical Society series B* **68**(2), 179–199.
URL: *http://dx.doi.org/10.1111/j.1467-9868.2006.00539.x* (see page 11)

Müller, H.-G. & Stadtmüller, U. (2005), 'Generalized functional linear models', *Annals of Statistics* **33**(2), 774–805.
URL: *http://dx.doi.org/10.1214/009053604000001156* (see page 14)

Munkres, J. (2000), *Topology*, Featured Titles for Topology Series, Prentice Hall, Incorporated.
URL: *http://books.google.ch/books?id=XjoZAQAAIAAJ* (see page 242)

Noble, D. (2010), 'Biophysics and systems biology', *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences* **368**(1914), 1125–1139.
URL: *http://rsta.royalsocietypublishing.org/content/368/1914/1125.abstract* (see page 153)

Panaretos, V. M., Kraus, D. & Maddocks, J. H. (2010), 'Second-order comparison of Gaussian random functions and the geometry of DNA minicircles', *Journal of the American Statistical Association* **105**(490), 670–682. Supplementary materials available online.
URL: *http://dx.doi.org/10.1198/jasa.2010.tm09239* (see pages 154, 165 and 169)

Panaretos, V. M. & Tavakoli, S. (2013*a*), 'Cramér–Karhunen–Loève representation and harmonic principal component analysis of functional time series', *Stochastic Processes and their Applications* **123**(7), 2779–2807.
URL: *http://dx.doi.org/10.1016/j.spa.2013.03.015* (see pages iii, 21, 65, 66 and 68)

Panaretos, V. M. & Tavakoli, S. (2013*b*), 'Fourier analysis of stationary time series in function space', *Annals of Statistics* **41**(2), 568–603.
URL: *http://dx.doi.org/10.1214/13-AOS1086* (see pages iii, 68, 155, 160, 162 and 243)

Peligrad, M. & Wu, W. B. (2010), 'Central limit theorem for Fourier transforms of stationary processes', *Annals of Probability* **38**(5), 2009–2022.
URL: *http://dx.doi.org/10.1214/10-AOP530* (see page 131)

Percival, D. B. & Walden, A. T. (1993), *Spectral analysis for physical applications: multitaper and conventional univariate techniques*, University Press, Cambridge. (see page 83)

Picinbono, B. (1996), 'Second-order complex random vectors and normal distributions', *IEEE Transactions on Signal Processing* **44**(10), 2637–2640. (see pages 134 and 179)

Pigoli, D. & Sangalli, L. M. (2012), 'Wavelets in functional data analysis: estimation of multidimensional curves and their derivatives', *Comput. Statist. Data Anal.* **56**(6), 1482–1498.
URL: *http://dx.doi.org/10.1016/j.csda.2011.12.016* (see page 11)

Pollard, D. (1984), *Convergence of stochastic processes*, Springer Series in Statistics, Springer-Verlag, New York.
URL: *http://dx.doi.org/10.1007/978-1-4612-5254-2* (see page 79)

Pólya, G. & Szegő, G. (1972), *Problems and theorems in analysis. Vol. I: Series, integral calculus, theory of functions*, Springer-Verlag, New York-Berlin. Translated from the German by D. Aeppli, Die Grundlehren der mathematischen Wissenschaften, Band 193. (see page 149)

Priestley, M. B. (2001), *Spectral analysis and time series, Vol. I and II*, Academic Press, London. (see pages 83 and 162)

Ramsay, J., Graves, S. & Hooker, G. (2009), *Functional Data Analysis with R and MATLAB*, Springer, New York. (see page 5)

Ramsay, J. O. & Silverman, B. W. (2002), *Applied functional data analysis*, Springer Series in Statistics, Springer-Verlag, New York. Methods and case studies.
URL: *http://dx.doi.org/10.1007/b98886* (see page 11)

Ramsay, J. O. & Silverman, B. W. (2005), *Functional data analysis*, Springer Series in Statistics, second edn, Springer, New York. (see pages 1, 5, 11, 154 and 159)

Riesz, F. & Sz.-Nagy, B. (1968), *Leçons d'analyse fonctionnelle*, Gauthier-Villars, Editeur-Imprimeur-Libraire, Paris; Akadémiai Kiadó, Budapest. Cinquième édition. Académie des Sciences de Hongrie. (see page 218)

Ringrose, J. R. (1971), *Compact non-self-adjoint operators*, Van Nostrand Reinhold Co., London. (see page 214)

Rosenblatt, M. (1984), 'Asymptotic normality, strong mixing and spectral density estimates', *Annals of Probability* **12**(4), 1167–1180. (see page 130)

Rosenblatt, M. (1985), *Stationary sequences and random fields*, Birkhäuser Boston, Inc., Boston, MA.
URL: *http://dx.doi.org/10.1007/978-1-4612-5156-9* (see pages 17, 130 and 136)

Rudin, W. (1976), *Principles of mathematical analysis*, third edn, McGraw-Hill Book Co., New York-Auckland-Düsseldorf. International Series in Pure and Applied Mathematics. (see pages 72 and 151)

Rudin, W. (1991), *Functional analysis. 2nd ed.*, 2nd ed. edn, New York, NY: McGraw-Hill. (see pages 66 and 213)

Ryan, R. A. (2002), *Introduction to tensor products of Banach spaces*, Springer Monographs in Mathematics, Springer-Verlag London, Ltd., London.
URL: *http://dx.doi.org/10.1007/978-1-4471-3903-4* (see pages 225, 226 and 229)

Sambriski, E. J., Schwartz, D. C. & de Pablo, J. J. (n.d.), 'A mesoscale model of DNA and its renaturation', *Biophysical Journal* **96**(5), 1675–1690.
URL: *http://www.cell.com/biophysj/abstract/S0006-3495(08)04006-X* (see page 153)

Sangalli, L. M., Secchi, P., Vantini, S. & Veneziani, A. (2009), 'Efficient estimation of three-dimensional curves and their derivatives by free-knot regression splines, applied to the analysis of inner carotid artery centrelines', *J. R. Stat. Soc. Ser. C. Appl. Stat.* **58**(3), 285–306.
URL: *http://dx.doi.org/10.1111/j.1467-9876.2008.00653.x* (see pages 11 and 158)

Schatten, R. (1950), *A Theory of Cross-Spaces*, Annals of Mathematics Studies, no. 26, Princeton University Press, Princeton, N. J. (see pages 225 and 226)

Schreier, P. J. & Scharf, L. L. (2010), *Statistical signal processing of complex-valued data*, Cambridge University Press, Cambridge. The theory of improper and noncircular signals.
URL: *http://dx.doi.org/10.1017/CBO9780511815911* (see pages 134 and 179)

Schrödinger, E. (1944), *What is Life?: The Physical Aspect of the Living Cell*, What is Life?: The Physical Aspect of the Living Cell, Cambridge Univ. Press, Cambridge. (see page 153)

Schwartzman, A., Dougherty, R. F. & Taylor, J. E. (2008), 'False discovery rate analysis of brain diffusion direction maps', *Annals of Applied Statistics* **2**(1), 153–175.
URL: *http://dx.doi.org/10.1214/07-AOAS133* (see page 181)

Shao, X. & Wu, W. B. (2007), 'Asymptotic spectral theory for nonlinear time series', *Annals of Statistics* **35**(4), 1773–1801.
URL: *http://dx.doi.org/10.1214/009053606000001479* (see pages 130 and 131)

Shore, D. & Baldwin, R. L. (1983), 'Energetics of DNA twisting: I. relation between twist and cyclization probability', *Journal of Molecular Biology* **170**(4), 957 – 981.
URL: *http://www.sciencedirect.com/science/article/pii/S0022283683801983* (see page 154)

Shore, D., Langowski, J. & Baldwin, R. L. (1981), 'DNA flexibility studied by covalent closure of short fragments into circles', *Proceedings of the National Academy of Science* **78**(8), 4833–4837. (see page 154)

Srivastava, A., Wu, W., Kurtek, S., Klassen, E. & Marron, J. S. (2011), 'Registration of Functional Data Using Fisher-Rao Metric', *ArXiv e-prints* . (see page 11)

Storey, J. D., Taylor, J. E. & Siegmund, D. (2004), 'Strong control, conservative point estimation and simultaneous conservative consistency of false discovery rates: a unified approach', *Journal of the Royal Statistical Society series B* **66**(1), 187–205.
URL: *http://dx.doi.org/10.1111/j.1467-9868.2004.00439.x* (see pages 171 and 181)

Sy, J., Taylor, J. & Cumberland, W. (1997), 'A stochastic model for the analysis of bivariate longitudinal aids data', *Biometrics* pp. 542–555. (see page 11)

Tao, T. (2011), *An introduction to measure theory.*, Providence, RI: American Mathematical Society (AMS). (see page 4)

Tibshirani, R. (1996), 'Regression shrinkage and selection via the lasso', *J. Roy. Statist. Soc. Ser. B* **58**(1), 267–288. (see page 14)

Tibshirani, R. (2011), 'Regression shrinkage and selection via the lasso: a retrospective', *Journal of the Royal Statistical Society series B* **73**(3), 273–282.
URL: *http://dx.doi.org/10.1111/j.1467-9868.2011.00771.x* (see page 14)

Vakhania, N. N., Tarieladze, V. I. & Chobanyan, S. A. (1987), *Probability distributions on Banach spaces*, Vol. 14 of *Mathematics and its Applications (Soviet Series)*, D. Reidel Publishing Co., Dordrecht. Translated from the Russian and with a preface by Wojbor A. Woyczynski.
URL: *http://dx.doi.org/10.1007/978-94-009-3873-1* (see page 241)

Vanmarcke, E. (2010), *Random fields*, new edn, World Scientific Publishing Co. Pte. Ltd., Hackensack, NJ. Analysis and synthesis.
URL: *http://dx.doi.org/10.1142/5807* (see page 182)

Walter, J., Gonzalez, O. & Maddocks, J. H. (2010), 'On the stochastic modeling of rigid body systems with application to polymer dynamics', *Multiscale Model. Simul.* **8**(3), 1018–1053.
URL: *http://dx.doi.org/10.1137/090765705* (see page 153)

Wand, M. P. & Jones, M. C. (1995), *Kernel smoothing*, Vol. 60 of *Monographs on Statistics and Applied Probability*, Chapman and Hall, Ltd., London.
URL: *http://dx.doi.org/10.1007/978-1-4899-4493-1* (see pages 11, 125 and 174)

Weidmann, J. (1980), *Linear operators in Hilbert spaces*, Vol. 68 of *Graduate Texts in Mathematics*, Springer-Verlag, New York-Berlin. Translated from the German by Joseph Szücs. (see page 221)

Wheeden, R. L. & Zygmund, A. (1977), *Measure and integral*, Marcel Dekker, Inc., New York-Basel. An introduction to real analysis, Pure and Applied Mathematics, Vol. 43. (see pages 145, 147 and 149)

Yao, F. & Lee, T. C. M. (2006), 'Penalized spline models for functional principal component analysis', *Journal of the Royal Statistical Society series B* **68**(1), 3–25.
URL: *http://dx.doi.org/10.1111/j.1467-9868.2005.00530.x* (see page 11)

Yao, F., Müller, H.-G. & Wang, J.-L. (2005), 'Functional data analysis for sparse longitudinal data', *Journal of the American Statistical Association* **100**(470), 577–590.
URL: *http://dx.doi.org/10.1198/016214504000001745* (see page 13)

Yosida, K. (1995), *Functional analysis*, Classics in Mathematics, Springer-Verlag, Berlin. Reprint of the sixth (1980) edition. (see page 229)

Zhang, X., Shao, X., Hayhoe, K. & Wuebbles, D. J. (2011), 'Testing the structural stability of temporally dependent functional observations and application to climate projections', *Electronic Journal of Statistics* **5**, 1765–1796.
URL: *http://dx.doi.org/10.1214/11-EJS655* (see page 19)

Zhu, K. (2007), *Operator theory in function spaces*, Vol. 138 of *Mathematical Surveys and Monographs*, second edn, American Mathematical Society, Providence, RI.
URL: *http://dx.doi.org/10.1090/surv/138* (see page 214)

Index

Shahin Tavakoli

Mathematics Department
EPFL, Station 8
1015 Lausanne, Switzerland
+41.(0)79.735.04.65
shahin.tavakoli@epfl.ch

Background

Date and Place of Birth 3rd May 1986, Epsom, Surrey, United Kingdom.

Nationality British and Swiss (dual citizenship).

Education

Ph. D. (expected 07.2014) **Statistics**, École Polytechnique Fédérale de Lausanne (EPFL), Switzerland.
– Topic : PCA for dependent functional data, Spectral analysis of functional time series, Applications to DNA molecular dynamics.
– Advisor : Prof. V. M. Panaretos.

M. Sc. 2010 **Applied Mathematics**, EPFL, Switzerland.
Extra semester in Management.

B. Sc. 2008 **Mathematics**, EPFL, Switzerland.
One year exchange at the University of Granada (Spain).

Research Interests

Functional Data Analysis, High-Dimensional Statistics, Stochastic Processes, Time Series Analysis, Multiple Testing, Shrinkage Estimation, Applications to Biophysics, Finance and Imaging.

Papers

2014 **Functional Data Analysis for Random Curves in Motion with Applications to DNA Molecular Dynamics**, *In preparation*, (with V. M. Panaretos).

2013 **Cramér-Karhunen-Loève Representation and Harmonic Principal Component Analysis of Functional Time Series**, *Stochastic Processes and their Applications*, Volume 123 (7), (with V. M. Panaretos).

2013 **Fourier Analysis of Stationary Time Series in Function Space**, *Annals of Statistics*, Volume 41 (2), (with V. M. Panaretos).

Honors and Awards

2013 **Dean's special prize for teaching**, *EPFL*,
The letter reads : "[...] rewarding your services of exceptional value".

2013 **Selected to participate in the 1st Heidelberg Laureate Forum**,
The website reads : "[...] Bringing together our laureates [Fields Medal, Abel Prize, ACM Turing Award] with the best of the students from around the world for a full week.".
(The 200 most qualified young researchers among 600 international applicants were selected)

2008 **"Bourse d'Excellence au niveau Master"**, *EPFL*.
Scholarship based on academic excellence, awarded each year to 20 Master's students of EPFL.

2006 **Cousin Prize**, *EPFL*.
Best exam marks for the first year, with special skills in Mathematics.

2004 **Marc Birkigt Prize**, *Sismondi High-School*, Geneva, Switzerland.
Best Physics & Mathematics grades.

2000 **Treuthardt Prize in mathematics**, *Collège des Colombières*, Versoix, Switzerland.
Best grades in Mathematics.

Undergraduate Academic Projects

Spring 2010 **Elements of Statistical Inference in Hilbert Spaces**, *EPFL*.
Master Thesis in Statistics.

Winter 2009 **Some Shape Theory, and an application to the Old Stones of Land's End**, *EPFL*.
Semester project in Statistics.

Winter 2008 **An Introduction to The Stein Phenomenon**, *EPFL*.
Semester project in Statistics.

Invited Talks

02.2014 **Optimal Finite Dimensional Reduction of Functional Time Series with Applications to DNA Molecular Dynamics**, *Statistical Laboratory*, University of Cambridge (UK).

06.2013 **Thematic Presentation on Functional Data Analysis**, *Young Researcher's Conference in Applied Probability and Statistics (CUSO)*, EPFL.

12.2012 **Invited talk in the "Functional data with spatial dependence" session**, *5th International Conference of the ERCIM*, Oviedo (Spain).

Contributed Talks

08.2013 Joint Statistical Meeting, Montreal (Canada).

06.2012 First Conference of the International Society for Non-Parametric Statistics, Halkidiki (Greece).

Posters

09.2012 High Dimensional and Dependent Functional Data Research Workshop, University of Bristol (U.K.) ; **received financial support for travel expenses**.

08.2011 ISI Young Statisticians Meeting, Dublin (Ireland).

Other Conferences/Workshops Attended

08.2013 SAMSI, Summer School of the Program on Low-dimensional Structure in High-dimensional Systems.

09.2010 SAMSI, Opening Workshop of the Program on Analysis of Object Data.

Academic Services

Served as Referee for : Biometrika, Statistics and Computing, Communications in Statistics.

Work Experience as Teaching Assistant at EPFL

Fall 2013 Principal Teaching Assistant for "Statistical Theory" (given by Prof. V. M. Panaretos)
- One semester MSc course equivalent to a first year Graduate course at the level of Bickel & Doksum.
- Course webpage : `http://smat.epfl.ch/courses/theory.php`

Spring 2013 Principal Teaching Assistant for "Probability and Statistics" (given by Prof. V. M. Panaretos)
- Supervised a group of 4 teaching assistants, with weekly meetings to prepare them for the exercise sessions.

Fall 2012 Principal Teaching Assistant for "Linear Models" (given by Prof. V. M. Panaretos).
- Course on Linear Regression based on Matrix Algebra.
- Course webpage : `http://smat.epfl.ch/courses/regression.php`

2010–2012 Teaching Assistant for "Mathematics for Chemists" (given by Dr. J. L. Zuleta Estrugo).

Fall 2008 Teaching Assistant for "Analyse III" (given by Prof. R. Dalang).
- Course in Vector Calculus and Complex Analysis.

Undergraduate Student Project Supervision

Fall 2013 Supervised an undergraduate semester projects on **multiple testing procedures**.

Spring 2013 Supervised an undergraduate semester project on **the spectral analysis of time series**.

Spring 2012 Supervised two undergraduate semester projects on **the bootstrap** and on **kernel density estimation**.

Associative/Professional Experience

2013– **Member of the founding committee of QED, the association of Mathematics' PhD students at EPFL**.

2013– **Member of the commission of the doctoral school of Mathematics, EPFL**, *Representative of Mathematics' PhD Students in Mathematics*.

2011–2013 **Festival de la Cité**, *Co-chief of the largest bar of the Festival*.

2009–2013 **Festival de la Cité, Cully Jazz Festival & Jazzonze+ festival**, *Bartender*.

2008–2010 **Junior Entreprise EPFL (JE EPFL)**, *Member of the Management Committee*.

2005–2006 **Member of the founding committee of CQFD, the association of Mathematics' students at EPFL**.

2005–2006 **Class representative**, *1^{st} year of Bachelor*.

Languages

English	**Excellent**	
French	**Excellent**	
German	**Very good level**	*4 months in Vienna, effectiveness (C1) level paper.*
Spanish	**Good level**	*University exchange year in Granada, Spain.*
Farsi	**Mother tongue**	*Spoken : fluent, Written : basic knowledge.*
Italian	**Basic understanding**	

Computer skills

Programming C++, R, Latex, Unix environment.

Office Word, Excel, Outlook, Powerpoint.

General Interests

Sports Badminton, Triathlon.

Dance Practice of Salsa for 3 years. Beginner in Argentine Tango.

General Improvisational theater, Cooking, Jazz music, Gardening, Photography.