

Topics on Anomaly Detection, High Dimensional Testing and Spectral Inference for Functional Data

by

Rafail Kartsioukas

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Statistics)
in The University of Michigan
2023

Doctoral Committee:

Professor Stilian Stoev, Co-Chair
Professor Tailen Hsing, Co-Chair
Dr. Michael Kallitsis, Merit Network
Professor Ambuj Tewari



Ἐπιστήμη ποιητική εὐδαιμονίας.

Knowlegde is the creator of bliss.

Rafail Kartsioukas

rkarts@umich.edu

ORCID iD: 0000-0002-2408-2851

© Rafail Kartsioukas 2023

For Thomas, Aristeidis and Eleni.

ACKNOWLEDGEMENTS

First and foremost, I would like to thank my advisor Stilian Stoev for his constant encouragement, his precious advices and his valuable guidance through all of my academic endeavors. I also warmly thank professors Tailen Hsing and Ambuj Tewari and doctor Michael Kallitsis for serving on my dissertation committee. In particular, I would like to thank Tailen for putting tremendous effort into my work on spectral inference; this work would have never reached such an excellent shape without his much appreciated input. Special thanks are due to Michael for bringing me in contact with the Greek-speaking community of Ann Arbor, making the distance from home seem much smaller than it actually is.

I want to also acknowledge the following people for the vital roles they have played during my time as a student at the University of Michigan: Ya'acov Ritov, Enes Diliber, Emil Lakkis, Jessica Liu, Wyliena Guan, Michael Law, Debarghya Mukherjee, Daniel Kessler, Andrew Yarger and Jordan Hensley. A special mention should be made to Zheng Gao as Chapter II is based on a joint work with him, while his ideas on exact support recovery have motivated the theoretical research in Chapter III. Generous funding was provided by the Onassis Foundation and the Horace H. Rackham Graduate School throughout my studies in the University of Michigan. Special thanks are additionally due to Judith McDonald, Jean McKee and the administrative staff of the University of Michigan Department of Statistics for all their assistance.

Many thanks to my parents, Michael and Pelagia, my brothers, Thomas and Aristeidis, and my sister Eleni. Despite the physical distance, their support was

decisive during the most critical moments of this effort. Finally, the day to day emotional support I received from my girlfriend Tara on the last year of my PhD program was of utmost importance. They all deserve a lot of credit for this work; the fruition of such a long and challenging task is inextricably linked to the people, with whom one chooses to most closely surround themselves.

TABLE OF CONTENTS

DEDICATION	ii
ACKNOWLEDGEMENTS	iii
LIST OF FIGURES	vii
LIST OF TABLES	ix
LIST OF ABBREVIATIONS	x
ABSTRACT	xii
CHAPTER	
I. Introduction	1
1.1 Part I	1
1.2 Part II	6
Part I Anomaly Identification and High Dimensional Testing	11
II. Anomaly Detection	11
2.1 Darknet and cyberthreats	12
2.2 Methodology	15
2.3 Performance evaluation	27
2.4 Theoretical Results	35
III. Concentration Rates	46
3.1 Relative stability and dependence	47
3.2 Concentration of maxima and high-dimensional inference	50
3.3 On the optimal rate of concentration	55

3.4	Rates of uniform relative stability	65
3.5	Technical proofs	79
Part II Spectral Inference for Functional Spatial Data		89
IV. The Lag-Window Estimator		89
4.1	The spectral density estimation problem	90
4.2	Covariance and spectral density of a stationary process in a Hilbert space	92
4.3	Spectral density estimation based on irregularly sampled data	99
4.4	Asymptotic properties	104
4.5	Data observed on a regular grid	113
4.6	An RKHS formulation based on discretely-observed functional data	123
4.7	Related work and discussions	126
4.8	Proofs	130
V. Minimax Rates		180
5.1	Minimax rates	180
5.2	Proof for discrete time	184
5.3	Proof for continuous time	190
5.4	Samarov-type lemmas	194
VI. Asymptotic Normality		199
6.1	Asymptotic distribution	199
6.2	Overview of Central Limit Theorem proof	204
6.3	Flat concentration	208
6.4	Convergence of moments	212
6.5	Cumulants and Isserlis' formulas	227
VII. Future Directions		238
7.1	Anomaly Detection	238
7.2	Concentration Rates	239
7.3	Spectral Inference	242
BIBLIOGRAPHY		245

LIST OF FIGURES

Figure

1.1	Log-transformed raw Internet traffic obtained from the ORION Network Telescope (<i>Merit Network, Inc.</i> , 2022). The top 50 ports are displayed on the heatmap, ordered with respect to the total volume of traffic. The dataset spans the entire month of September 2016.	3
1.2	Profiles of Argo data for January 15, 2019 in the area of the Aegean Sea of Greece. The plots contain measurements at different pressure levels (depth levels) for the temperature, the salinity level, the potential density and the density respectively. Each line in the plots corresponds to an individual float.	7
1.3	Locations of the fleet of operational floats of the Argo project, as of January 2023. The floats cover extensively the world’s ocean. Copyrights for this image belong to <i>ARGO</i> (2020).	8
2.1	(a) Incoming unique source IP traffic coming to port 5353, from 16 September, 0639 UTC to 23 September, 0719 UTC in 2016. (b) Incoming traffic for the same port after stabilization through log-transform. Another event is now visible despite diurnal trends in traffic denoted by a spike around 17 September 14:00 UTC. (c) Residuals for the same port from the traffic model. The large event is flagged by our algorithm. Our method adapts to the change of regime after the second event and does not flag the whole remaining time series as anomaly. Moreover, the periodic trend until 21 September is clearly filtered out after the application of the algorithm.	16
2.2	The largest principal angle between estimated and true subspace for incremental PCA (iPCA) (cf <i>Bjorck et al.</i> (1973); <i>Yu et al.</i> (2015)). The horizontal line (in red) shows the largest principal angle between the true subspace and the batch Principal Component Analysis (PCA) estimated one.	30
2.3	ROC curves of Algorithm 1 (iPCA-based) and Q-statistic for long duration anomaly (6 hour) of low magnitude (snr=2) in synthetic data.	33
2.4	Residuals and detection boundaries (top) and raw traffic (bottom) for ports 0 and 2323. Red color in bottom plots shows the detected anomalies.	36

2.5	Alert matrix for the Darknet dataset for September 2016 (Mirai onset). Red color indicates detection.	37
3.1	Phase transition boundary on the exact support recovery problem. If $r > g(\beta)$, and so in the green area, then we can find a thresholding estimator, so that exact support recovery is attainable, whether \mathcal{E} is URS or not. If $r < g(\beta)$, in the brown area, and \mathcal{E} is URS, then no matter the thresholding estimator, the probability of exactly estimating S_p vanishes as p grows to infinity.	53
4.1	Example of the Voronoi tessellation. Every cell includes a single sampling time point as a representative. A cell is defined as the part of the sampling domain containing all the time points that are closer to its representatives compared to all other representatives, based on the Euclidean metric on \mathbb{R}^d . The figure also shows the sampling framework imposed by Assumption S(a). Indeed, the sampling domain inflates, while the sample time points are denser and denser as the sample size grows; equivalently δ_n becomes smaller with the increase of the sample size.	101

LIST OF TABLES

Table

2.1	Recommendations for the choice of tuning parameters	31
2.2	Performance of Algorithm 1 on Synthetic Data. Observe the high True Positive Rate (TPR) and low False Positive Rate (FPR) for carefully chosen thresholds L	34

LIST OF ABBREVIATIONS

PCA Principal Component Analysis

GWAS genome-wide association studies

IP Internet Protocol

PCA Principal Component Analysis

URS Uniform Relative Stability

DDoS distributed denial of service

IoT Internet-of-Things

TAP Traffic Analysis Profile

EWMA Exponentially Weighted Moving Average

iPCA incremental PCA

REG robust guard estimation

fGn fractional Gaussian noise

snr signal to noise ratio

AUC area under curve

UDD Uniformly Decreasing Dependence

CONS complete orthonormal system

RKHS reproducing kernel Hilbert space

ABSTRACT

This dissertation consists of two major parts. The first part is concerned with high-dimensional testing and involves both a methodological and theoretical results. The methodology portion is centered around the detection of anomalous Internet traffic. It is motivated by and applied to network telescopes or “Darknet” type data, which is Internet traffic obtained by monitoring a large number of streams corresponding to “unused” Internet address space. We propose an algorithm for the synchronous on-line detection of abnormal Internet traffic, based on recent theoretical developments, and evaluate its performance both in the detection and the identification aspects. The remainder of the first part involves theoretical contributions which solve an open problem in probability on the rates of convergence of maxima for dependent Gaussian triangular arrays. These technical results on the rates allow us to establish that the concentration of maxima phenomenon holds in more general, not necessarily Gaussian, models. The latter phenomenon is the key to an important phase transition result that characterizes the statistical limits in the exact support recovery problem for a sparse high-dimensional signal observed in additive, light-tailed noise. Thus, our theoretical results make direct contributions to high-dimensional statistics. The second part of this dissertation is focused on the non-parametric estimation of the spectral density of space-time random field processes taking values in a separable Hilbert space. The estimator relies on kernel smoothing and is applicable to spatial sampling schemes where data are not necessarily observed at regular spatial locations. In a mixed-domain asymptotic setting and under general conditions, rates for the bias and variance of the estimator are obtained which lead to rates for its consis-

tency. Considering practical applications, where complete functional data are usually unavailable, our asymptotic results are specialized to the case of discretely-sampled functional data taking values in a reproducing kernel Hilbert space. Further, it is shown that when the data are observed on a regular spatial grid, the optimal rate of the estimator matches the minimax rate for the class of covariance functions that decay according to a power law. Finally, the asymptotic normality of the spectral density estimator is also established under general conditions for Gaussian Hilbert-space valued processes.

CHAPTER I

Introduction

The rapid development of information technology over the recent years has enabled researchers to collect and manipulate immense volumes of data at unparalleled speeds and at a fraction of the cost, compared to the past. As cliché as it might sound, we live in the era of Big Data. In this era, the biggest challenge in data analysis does not lie in the traditional scarcity of data to collect, but in the efficient and flexible manipulation of the overabundant data. The volume of collected data and its analysis is limited only by the constraints of the modern computational power and shall only grow with the latest developments in technology.

1.1 Part I

The unprecedented computational power, leading to the plethora of available data, has radically transformed the way that the vast majority of the scientific community approaches research problems. Traditionally, in the hypothesis-experiment-analysis cycle, a hypothesis is made, based on which experiments are being conducted to collect the necessary data to conduct an analysis. In direct contrast to this model, nowadays scientists are granted access to ample data, often before even a specific scientific question has been formulated. The data themselves contribute to the formulation of a large number of questions. Checking the simultaneous plausibility of these hypotheses

is known as multiple testing and is a procedure that arises in a great variety of fields. For example, one might look into genome-wide association studies (GWAS) (*Sun et al.*, 2006; *Mieth et al.*, 2016) or engineering applications, including voice activity recognition (*Ramírez et al.*, 2007) and spectrum misuse detection (*Zhang et al.*, 2017).

One main example of multiple testing problems emerge in cybersecurity. The synchronous and online monitoring of millions of Internet Protocol (IP) addresses is a central task in this sector. A typical example of what the traffic looks like in this domain is presented in Figure 1.1. This heatmap depicts the number of unique sources that target each monitored TCP/UDP port per minute, for the full month of September 2006. We order these ports by the volume of total traffic and show the top-50 ports. These data are obtained through Merit Network’s ORION network telescope (*Merit Network, Inc.*, 2022) and consist of communication requests from valid source IP addresses destined to unoccupied (dark) IP addresses controlled by Merit Network. Therefore such traffic is referred to as Darknet traffic. The observed Darknet traffic can be due to misconfigured computers, but in its majority is due to malicious scanning activity originating from ill-purposed actors. Detecting anomalies in the Darknet can thus help identify new types of cyber attacks (see e.g. (*Antonakakis et al.*, 2017)).

Commonly in cybersecurity, many different ports/streams of data are being monitored concurrently, with the main task being the discovery of possible anomalies in the Internet traffic. The smaller the time window allowed for the discovery, the more challenging the problem becomes, but also the more useful the results are. Many algorithms have been developed (*Xie et al.*, 2017; *Lin et al.*, 2020) for the detection of these anomalies, with often two goals in mind. Firstly, spotting whether there is an anomaly in the total traffic under consideration, and secondly, locating the streams where this deviation from the non-anomalous standard traffic occurs.

In Chapter II we focus on a methodological problem pertaining to the discussion

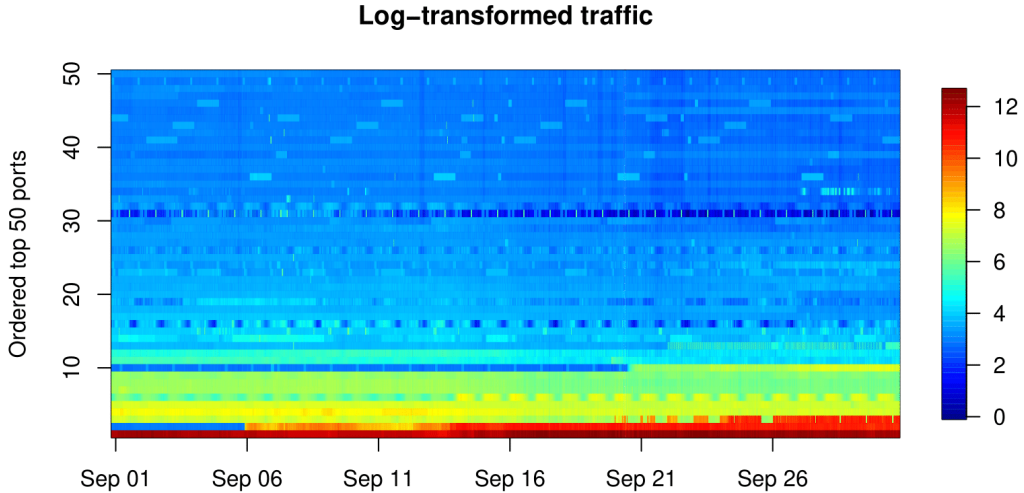


Figure 1.1: Log-transformed raw Internet traffic obtained from the ORION Network Telescope (*Merit Network, Inc.*, 2022). The top 50 ports are displayed on the heatmap, ordered with respect to the total volume of traffic. The dataset spans the entire month of September 2016.

above. The main goal is to detect sparse anomalies embedded in a high-dimensional low-rank background, i.e., our observations are of the form

$$x_t = Bf_t + u_t + \epsilon_t. \quad (1.1)$$

Here, $x_t = \{x_t(i)\}_{i=1}^p$ is a vector-time series denoting a collection of p data streams observed at time-bin t , while $\epsilon_t = \{\epsilon_t(i)\}_{i=1}^p$ models the “benign” background noise. The vector $f_t = \{f_t(j)\}_{j=1}^k$ denotes the $k(\ll p)$ periodic over time (diurnal, weekly, etc) trends that are shared across some of the streams and the matrix $B \in \mathbb{R}^{p \times k}$ contains k linearly independent columns that indicate which factors f_t affect each stream. Finally, the vector $u_t = \{u_t(i)\}_{i=1}^p$ represents the anomalous “signal” at time t for the streams $1, \dots, p$. When there are no anomalies, $u_t = 0$.

Our methodology involves learning the matrix B over time, with focus on a sequential update of its estimate \hat{B} , every time that new data are received. To this end

we utilize sequential PCA techniques. This allows us to estimate the rank- k subspace spanned by the columns of B without having to store large amounts of historical data. At the same time, sequential PCA allows us to adapt this estimate with new information and to changes in the low-rank space modeling the background. After obtaining \hat{B} , we project the observed data onto the orthogonal complement of the subspace $\text{span}(\hat{B})$. Under some assumed incoherence conditions, the projected data will follow the canonical signal-plus-noise model

$$x_t = \mu_t + \epsilon_t, \tag{1.2}$$

where ϵ_t is the background noise as in (1.1) and the signal $\mu_t = (\mu_t(i))_{i=1}^p$ is a p -dimensional vector with s non-zero components. The non-zero components are called the support set $S = \{i : \mu(i) \neq 0\}$, while $\epsilon = \{\epsilon(i)\}_{i=1}^p$ is a random error vector. In high-dimensional statistics, as well as in Chapter II, the goal is twofold:

- (I) *Signal Detection*: The detection of non-zero components in μ . Namely, test the global hypothesis $\mu = 0$.
- (II) *Support Recovery*: Estimation of the support set S . It is otherwise known as *support estimation* or *signal identification problem*.

The algorithm we implement uses the aforementioned projections to perform detection and identification of the anomalies jointly. We evaluate the performance of the algorithm against classical anomaly detection methods on a synthetic dataset, while we also demonstrate our methodology on a Darknet data application. Finally, based on the synthetic data, we provide a guide to the practitioner for the use of our algorithm. This concludes the most applied part of this thesis.

In Chapter III we solve an open theoretical problem motivated by the *exact* support recovery problem. This problem is directly motivated by and applicable to the anomaly identification context of Chapter II. It is, however, of independent interest

in high-dimensional statistics. In the identification problem, there have been multiple approaches. Other works aim at estimating the support set S under many different criteria. For example, *Butucea et al. (2018)* use the Hamming loss, while *Arias-Castro and Ying (2018)* aim at approximate recovery of the support set using the $|\widehat{S} \Delta S|/|S|$ metric. Our goal is the exact recovery of the support set of the anomaly, when the signal $\mu = \mu_t$ in (1.2) is considered sparse, we have $n = 1$ observation and the dimension p grows to infinity. Namely, we want to find \widehat{S}_p such that

$$\mathbb{P}\left(\widehat{S}_p = S_p\right) \rightarrow 1, \quad \text{as } p \rightarrow \infty, \quad (1.3)$$

for the case where the signal μ in (1.2) is sparse. By sparse, we mean that

$$|S_p| =: s_p \sim \lfloor p^{1-\beta} \rfloor, \quad \text{as } p \rightarrow \infty,$$

where $\beta \in (0, 1]$ is a sparsity parameter. The larger the β , the fewer the non-zero signal entries in S , and the sparser the signal.

It was recently shown by *Gao and Stoev (2020)* that the exact support recovery problem obeys a phase transition phenomenon at the limit of $p \rightarrow \infty$, when thresholding estimators are utilized. This phase transition result shows that there is a certain boundary function of the sparsity parameter β such that (i) exact support recovery is possible for signal amplitudes above the boundary and (ii) exact support recovery is impossible otherwise (cf Section 3.2.1). In *Gao and Stoev (2020)*, the errors in (1.2) follow the Gaussian distribution, and although this problem had been studied extensively for independent errors (cf e.g. *Arias-Castro and Chen (2017)*; *Cai et al. (2007)*), the errors therein are studied under dependence. *Gao and Stoev (2020)* prove that the exact support recovery is feasible when the errors are assumed to be Uniformly Relative Stable. It is also shown that in the Gaussian case, Uniform Relative Stability (URS) is equivalent to the Uniformly Decreasing Dependence condition.

In Chapter III we examine the rate at which the URS occurs, namely at what rate a properly normalized maximum of a triangular array of Gaussian random variables converges in probability to the constant 1. We establish an upper bound on this rate, which curiously depends on the normalizing constants of the maximum. Moreover, through this rate we expand the range of models obeying URS, to a variety of models obtained through transformations of the Gaussian random variables. This leads to a much broader variety of models that undergo the phase transition phenomenon with regards to the exact recovery problem.

1.2 Part II

Apart from the multiple hypothesis testing, a field impacted by the growing availability of data is the domain of Spatial Statistics. Data related to Spatial Statistics are not only concerned with observations indexed by geographic (spatial) location. They are often indexed by time as well as space, and involve multiple attributes. Some scientific areas, where spatial data are of interest, include oceanography (see, e.g., the Argo Project(*ARGO*, 2020)), geology (*Tian et al.*, 2018) and real estate (*Pace et al.*, 1998). Figure 1.2 shows an example of such type of functional data arising in oceanography. The plot is obtained through the R shiny app of *Yarger* (2020a) (see also *Yarger et al.* (2022)). The data used therein come from the Argo program, an international program that measures water properties across the world’s ocean. To collect these measurements, robotic instruments, called floats, drift with the currents and move up and down in the water, periodically measuring temperature, salinity and other variables as a function of depth (pressure). The spatial distribution of these floats as of January 2023 is given in Figure 1.3.

In the analysis of stationary spatial data, one of the main interests lies in examining their covariance structure. In the function-valued setting (see e.g. *Rudin* (1991); *Hsing and Eubank* (2015)), this amounts to studying either their operator-

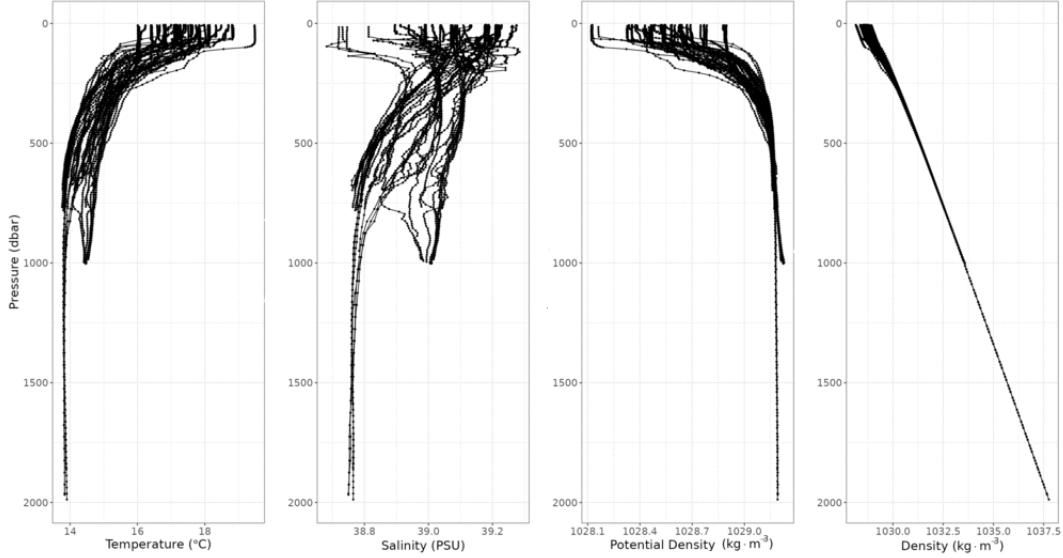


Figure 1.2: Profiles of Argo data for January 15, 2019 in the area of the Aegean Sea of Greece. The plots contain measurements at different pressure levels (depth levels) for the temperature, the salinity level, the potential density and the density respectively. Each line in the plots corresponds to an individual float.

valued auto-covariance or equivalently, the operator-valued spectral density. These two problems are duals of each other, meaning that estimation of either function suffices for the characterization of the second-order characteristics of the stochastic process under consideration.

In Chapter IV, we introduce a lag-window type estimator of the spectral density function. This estimator is defined in the very general case of stochastic processes taking values in separable Hilbert spaces, which can in particular be finite-dimensional or infinite-dimensional spaces of functions. Moreover, our estimator is defined on arbitrarily and irregularly sampled data in space, in contrast to existing periodogram-based estimators for functional time series (*Panaretos and Tavakoli, 2013; Politis, 2011*). In this chapter, we focus on establishing upper bounds on the rates of consistency of our estimator and also describing a potential modification with regards to discretely observed functional data.

Let $X = \{X(t), t \in \mathbb{R}^d\}$ be a process taking values in the separable Hilbert space

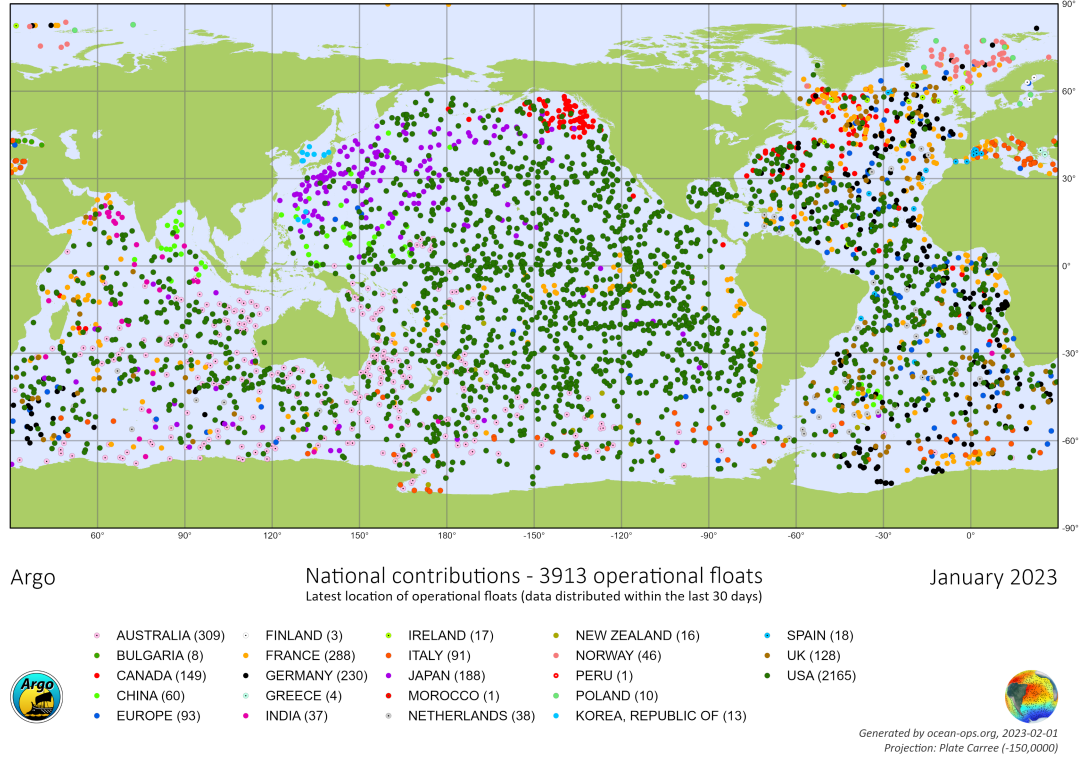


Figure 1.3: Locations of the fleet of operational floats of the Argo project, as of January 2023. The floats cover extensively the world's ocean. Copyrights for this image belong to *ARGO* (2020).

\mathbb{H} and let $f(\theta)$, $\hat{f}_n(\theta)$ denote the true spectral density and our estimator, respectively at the point $\theta \in \mathbb{R}^d$. The main goal of this chapter is to prove that $\hat{f}_n(\theta)$ is a consistent estimator of $f(\theta)$, and obtain rates on the consistency. Namely, using the Hilbert-Schmidt norm, we establish the rate at which the following quantity vanishes (uniformly for θ):

$$\mathbb{E} \|\hat{f}_n(\theta) - f(\theta)\|_{\text{HS}}.$$

To do so, we first use a bias-variance decomposition (using the Hilbert-Schmidt inner product)

$$\mathbb{E} \|\hat{f}_n(\theta) - f(\theta)\|_{\text{HS}}^2 = \|\mathbb{E} \hat{f}_n(\theta) - f(\theta)\|_{\text{HS}}^2 + \mathbb{E} \|\hat{f}_n(\theta) - \mathbb{E} \hat{f}_n(\theta)\|_{\text{HS}}^2$$

and explore the bias and variance vanish rates.

The rates we initially obtain for these quantities are quite general, as we work under a very broad framework of mixed-domain asymptotics with irregularly sampled data. In particular, the rate for the bias, under our assumptions, depends on the tail decay of the autocovariance operator of the stochastic process under consideration. In order to further study the optimality of the estimator, we focus on processes sampled over a rectangular grid and examine the class of the power-law decaying covariance operators. The known tail-decay leads to explicit consistency rates; we take advantage of this explicit form to optimally choose the bandwidth parameter of our estimator.

The abstract results when the observations take values in a Hilbert space are not directly useful in practice. Indeed, in applications one observes only discrete samples of functional data. In Section 4.6, we extend our estimator and results to this case, where the functional data are observed on discrete points and \mathbb{H} is a Reproducing Kernel Hilbert Space.

In Chapter V, we study the minimax optimal rates of the spectral density estimation problem for Hilbert space valued processes. Specifically, we focus on the aforementioned power-law decaying class of covariance operators. For this class, in both cases of processes sampled on a fixed grid and when the grid becomes denser, we derive minimax lower bounds for the rates of any estimator. Then, we show that our estimators $\hat{f}_n(\theta)$ are minimax rate-optimal. To the best of our knowledge, the minimax rate results obtained in this chapter are the first to be established for the pointwise inference of the spectral density for Hilbert space valued processes.

Finally, in Chapter VI we explore the asymptotic distribution of the estimator $\hat{f}_n(\theta)$. We impose the extra assumption that X is a Gaussian \mathbb{H} -valued process (cf Section 2 of *Rao (2014)*). Under this assumption, we establish a Central Limit Theorem type result for $\hat{f}_n(\theta)$ and obtain a stochastic representation of the limit. The proof is based on showing the convergence of all moments using a new type of Isserlis

(*Isserlis*, 1918) formula that we establish.

Part I

Anomaly Identification and High Dimensional Testing

CHAPTER II

Anomaly Detection

In this chapter, we are concerned with the multiple hypothesis testing problem in high dimensions, as it appears in the scope of cybersecurity. In this context, timely detection and identification of “anomalous” Internet traffic is of essence to the prompt termination of any cyberthreats. A unique window into observing Internet-wide scanners and other malicious entities is offered by network telescopes, commonly known as “Darknets”. However, monitoring Darknets for timely detection of coordinated and heavy scanning activities is a challenging task. The challenges mainly arise due to the non-stationarity and the dynamic nature of Internet traffic and, more importantly, the fact that one needs to monitor high-dimensional signals (e.g., all TCP/UDP ports) to search for “sparse” anomalies. We propose statistical methods to address both challenges in an efficient and “online” manner; our work is validated both with synthetic data and with real-world data from a large network telescope.

2.1 Darknet and cyberthreats

The Internet has evolved into a complex ecosystem, comprised of a plethora of network-accessible services and end-user devices that are frequently mismanaged, not properly maintained and secured, and outdated with untreated software vulnerabilities. Adversaries are increasingly becoming aware of this ill-secure Internet landscape and leverage it to their advantage for launching attacks against critical infrastructure. Examples abound: foraging for mismanaged NTP and DNS open resolvers (or other UDP-based services) is a well-known attack vector that can be exploited to incur volumetric reflection-and-amplification distributed denial of service (DDoS) attacks (*Czyz et al.*, 2014; *Rossow*, 2014; *Kühner et al.*, 2014); searching for and compromising insecure Internet-of-Things (IoT) devices (such as home routers, Web cameras, etc.) has led to the outset of the Mirai botnet back in 2016 that was responsible for some of the largest DDoS ever recorded (*Antonakakis et al.*, 2017; *Krebs*, 2021a; *Paganini*, 2016; *Perlroth*); variants of the Mirai epidemic still widely circulate (e.g., the Mozi (*Klopsch et al.*, 2020) and Meri (*Krebs*, 2021b) botnets) and assault services at a global scale; cybercriminals have been exploiting the COVID-19 pandemic to infiltrate networks via insecure VPN teleworking technologies that have been deployed to facilitate work-from-home opportunities (*DHS CISA and NCSC*, 2020).

Notably, the initial phase of the aforementioned attacks is *network scanning*, a step that is necessary to detect and afterwards exploit vulnerable services/hosts. Against this background, network operators are tasked with monitoring and protecting their networks and germane services utilized by their users. While many enterprises and large-scale networks operate sophisticated firewalls and intrusion detection systems (e.g., Zeek, Suricata or other non-open-source solutions), early signs of malicious network scanning activities may not be easily noticed from their vantage points. Large *Network Telescopes* or *Darknets* (*Moore et al.*, 2004; *Merit Network, Inc.*, 2022), however, can fill this gap and can provide *early warning notifications and insights*

for emergent network threats to security analysts. Network telescopes consist of monitoring infrastructure that receives and records unsolicited traffic destined to vast swaths of *unused* but *routed* Internet address spaces (i.e., millions of IPs). This traffic, coined as “Internet Background Radiation” (*Pang et al.*, 2004; *Wustrow et al.*, 2010), captures traffic from nefarious actors that perform Internet-wide scanning activities, malware and botnets that aim to infect other victims, “backscatter” activities that denote DoS attacks (*Wustrow et al.*, 2010), etc. Thus, Darknets offer a unique lens into macroscopic Internet activities and timely detection of new abnormal Darknet behaviors is extremely important.

In this thesis, we consider Darknet data from the ORION Network Telescope operated by Merit Network, Inc. (*Merit Network, Inc.*, 2022), and construct multi-variate signals for various TCP/UDP ports (as well as other types of traffic) that denote the amount of packets sent to the Darknet towards a particular port per monitoring interval (e.g., minutes) (see Figure 2.1.) Our goals are to *detect* when an “anomaly” occurs in the Darknet¹ and to also accurately *identify* the culprit port(s); such threat intelligence would be invaluable in diagnosing emerging new vulnerabilities (e.g., “zero-day” attacks). Our algorithms are based on the state-of-the-art theoretical results on the *sparse signal support recovery* problem in a high-dimensional setting (see, e.g., (*Gao and Stoev*, 2020) and the recent monograph (*Gao and Stoev*, 2021)). **Our main contributions are:** 1) we showcase, using simulated as well as real-world Darknet data, that signal trends (e.g., diurnal or weekly scanning patterns) can be filtered out from the multi-variate scanning signals using efficient *sequential PCA* techniques (*Arora et al.*, 2012a); 2) using recent theory (*Gao and Stoev*, 2020), we demonstrate that simple thresholding techniques applied individually on each univariate time-series exhibit better detection power than competing methods proposed

¹All traffic captured in the Darknet can be considered “anomalous” since Darknets serve no real services; however, henceforth we slightly abuse the terminology and refer to traffic “anomalies” in the statistical sense.

in the literature for diagnosing network anomalies (*Lakhina et al.*, 2004); 3) we propose and apply a non-parametric approach as a thresholding mechanism for the recovery of sparse anomalies; and 4) we illustrate our methods on real-world Darknet data using techniques amenable to online/streaming implementation.

2.1.1 Related Work

Network telescopes have been widely employed by the networking community to understand various macroscopic Internet events. E.g., Darknet data helped to shed light into botnets (*Antonakakis et al.*, 2017; *Dainotti et al.*, 2015), to obtain insights about network outages (*Benson et al.*, 2012; *Dainotti et al.*), to understand denial of service attacks (*Moore et al.*; *Jonker et al.*, 2017; *Czyz et al.*, 2014), for examining the behavior of IoT devices (*Shaikh et al.*, 2018), for observing Internet misconfigurations (*Czyz et al.*, 2013; *Wustrow et al.*, 2010), etc.

Mining of meaningful patterns in Darknet data is a challenging task due to the dimensionality of the data and the heterogeneity of the “Darknet features” that one could invoke. Several studies have resorted to unsupervised machine learning techniques, such as clustering, for the task at hand. *Niranjana et al.* (2020) propose using Mean Shift clustering algorithms on TCP features to cluster source IP addresses to find attack patterns in Darknet traffic. *Ban et al.* (2012) present a monitoring system that characterizes the behavior of long term cyber-attacks by mining Darknet traffic. In this system, machine learning techniques such as clustering, classification and function regression are applied to the analysis of Darknet traffic. *Bou-Harb et al.* (2014) propose a multidimensional monitoring method for source port 0 probing attacks by analyzing Darknet traffic. By performing unsupervised machine learning techniques on Darknet traffic, the activities by similar types of hosts are grouped by employing a set of statistical-based behavioral analytics. This approach is targeted only for source port 0 probing attacks. *Nishikaze et al.* (2015) present a machine learning approach

for large-scale monitoring of malicious activities on Internet using Darknet traffic. In the proposed system, network packets sent from a subnet to a Darknet are collected, and they are transformed into 27-dimensional Traffic Analysis Profile (TAP) feature vectors. Then, a hierarchical clustering is performed to obtain clusters for typical malicious behaviors. In the monitoring phase, the malicious activities in a subnet are estimated from the closest TAP feature cluster. Then, such TAP feature clusters for all subnets are visualized on the proposed monitoring system in real time to identify malicious activities. *Ban et al.* (2016) present a study on early detection of emerging novel attacks. The authors identify attack patterns on Darknet data using a clustering algorithm and perform nonlinear dimension reduction to provide visual hints about the relationship among different attacks.

Another family of methods rely on traffic prediction to detect anomalies. E.g., *Zakroum et al.* (2022) infer anomalies on network telescope traffic by predicting probing rates. They present a framework to monitor probing activities targeting network telescopes using Long Short-Term Memory deep learning networks to infer anomalous probing traffic and to raise early threat warnings.

2.2 Methodology

2.2.1 Problem Formulation

Consider a vector time series $\vec{x}_t = \{x_t(i)\}_{i=1}^p$ modelling a collection of p data streams (e.g., scanning activity against p ports). In network traffic monitoring applications, the data streams involve a highly non-stationary “baseline” traffic background signal $\vec{\theta}_t = \{\theta_t(i)\}_{i=1}^p$, which could be largely unpredictable and highly variable. Considering Internet traffic specifically, this “baseline” traffic often includes diurnal or weekly periodic trends that can be modeled by a small number of common factors. We encode these periodic phenomena through the classic linear factor model.

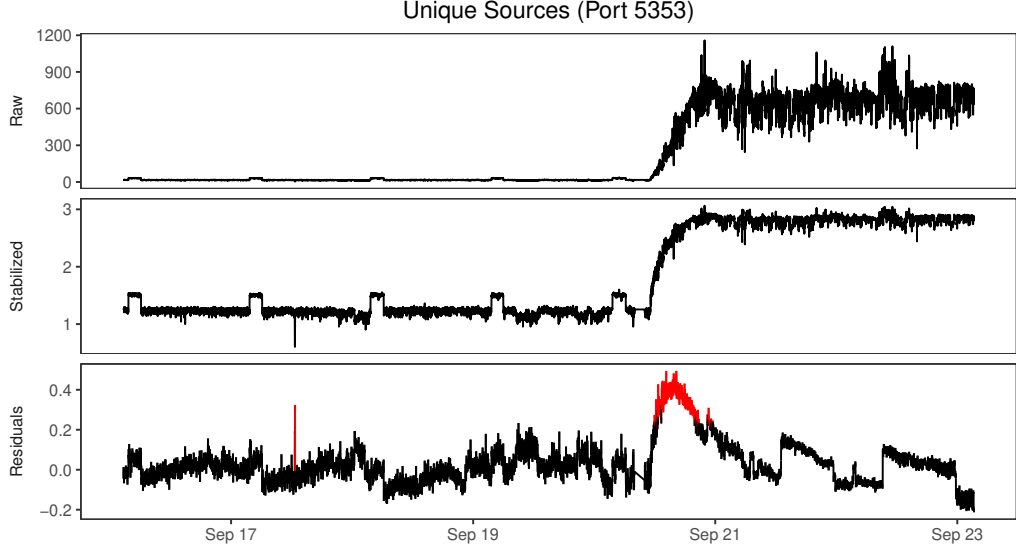


Figure 2.1: (a) Incoming unique source IP traffic coming to port 5353, from 16 September, 0639 UTC to 23 September, 0719 UTC in 2016. (b) Incoming traffic for the same port after stabilization through log-transform. Another event is now visible despite diurnal trends in traffic denoted by a spike around 17 September 14:00 UTC. (c) Residuals for the same port from the traffic model. The large event is flagged by our algorithm. Our method adapts to the change of regime after the second event and does not flag the whole remaining time series as anomaly. Moreover, the periodic trend until 21 September is clearly filtered out after the application of the algorithm.

Namely, we make the assumption that $\vec{\theta}_t = B\vec{f}_t$, where $B = (\vec{b}_1 \dots \vec{b}_k)_{p \times k}$ is a matrix of $k (\leq p)$ linearly independent columns that express the affected streams. Moreover, $\vec{f}_t = \{f_t(j)\}_{j=1}^k$ are the (non-stationary) factors, or periodic trends, that appear in these different streams. That is, the anomaly free regime can be modeled as

$$\vec{x}_t = B\vec{f}_t + \vec{\epsilon}_t, \quad (2.1)$$

where $\vec{\epsilon}_t = \{\epsilon_t(i)\}_{i=1}^p$ is a vector time-series modeling the “benign” noise. The time-series $\{\vec{\epsilon}_t\}$ may and typically does have a non-trivial (long-range) dependence structure *Willinger et al.* (2001), but may be assumed to be stationary. Additional assumptions on the dependence structure of ϵ_t may be used in Section 2.4 as needed.

Anomalies such as aggressive network scanning may be represented (in a suitable

feature space) by a mean-shift vector $\vec{u}_t = \{u_t(i)\}_{i=1}^p$, which is *sparse*, i.e., it affects only a relatively small and unknown subset of streams $S_t := \{i \in [p] : u_t(i) \neq 0\}$. Thus, in the anomalous regime, one observes

$$\vec{x}_t = B\vec{f}_t + \vec{u}_t + \vec{\epsilon}_t. \quad (2.2)$$

The general problems of interest are twofold:

- (i) *Detection*: Timely *detection* of the presence of the anomalous component \vec{u}_t .
- (ii) *Identification*: The estimation of the sets S_t of the streams containing anomalies.

2.2.2 An Algorithm for Joint Detection and Identification

We provide a high level summary of the algorithm we implement in our analysis. We examine high-dimensional data sourced from Darknet traffic observations and our goal is the joint detection and identification of any nefarious activity that might be present in the dataset.

In order to achieve our goals, we utilize sequential PCA techniques (i.e., incremental PCA or *iPCA*) attempting to estimate the factor subspace spanned by the common periodic trends. Once an estimate of this space is obtained, we project the new vector of data (one observation in each port signal) on this subspace and retain the residuals of the data minus their projections. Using an individual threshold for each stream, pertaining to the stream’s marginal variance and a control limit chosen by the operator (see Section 2.3.1.3 for guidance in system tuning), observations are flagged as anomalies individually.

The algorithm can be broken down in five phases.

1. **Input.** Tuning parameters include the length of the warm-up period (n_0) and smoothing parameters for some Exponentially Weighted Moving Average (EWMA) steps *Lucas et al.* (1990); $\lambda, \lambda_\mu, \lambda_\sigma$ denote the memory parameters

Algorithm 1 Online identification of sparse anomalies

Require: smoothing parameters $\lambda, \lambda_\mu, \lambda_\sigma$;
 control limit L ; effective subspace dimension k ;
 initial subspace estimates \hat{B} ; initial mean estimates $\hat{\nu}_x, \hat{\nu}_r$
 initial residual marginal variance estimates $\hat{\sigma}_r^2$,
 iPCA forget factor η , robust estimation guard R .
for new data $\vec{x}_t = (x_t(j))_{j=1}^p \in \mathbb{R}^p$ **do**
 for $j \notin \hat{S}_{t-1}$ **do**
 $\hat{\nu}_x(j) \leftarrow (1 - \lambda)\hat{\nu}_x(j) + \lambda x_t(j)$ ▷ Update mean estimates
 end for
 $\vec{r}_t \leftarrow (I - \hat{B}(\hat{B}^\top \hat{B})^{-1} \hat{B}^\top)(\vec{x}_t - \hat{\nu}_x)$ ▷ Projection step
 $\hat{B} \leftarrow \text{iPCA}(\vec{x}_t, \hat{B}, \hat{\nu}_x, k)$ ▷ Update subspace
 for $j: |r_t(j)| < R \cdot \hat{\sigma}_r(j)$ **do**
 $\hat{\nu}_r(j) \leftarrow (1 - \lambda_\mu)\hat{\nu}_r(j) + \lambda_\mu r_t(j)$
 end for
 for $j: |r_t(j) - \hat{\nu}_r(j)| < R \cdot \hat{\sigma}_r(j)$ **do**
 $\hat{\sigma}_r^2(j) \leftarrow (1 - \lambda_\sigma)\hat{\sigma}_r^2(j) + \lambda_\sigma (r_t(j) - \hat{\nu}_r(j))^2$ ▷ Update residual variances
 end for
 $\hat{S}_t \leftarrow \{j \mid |r_t(j) - \hat{\nu}_r(j)| > L \times \hat{\sigma}_r(j)\}$
 if $\hat{S}_t \neq \emptyset$ **then**
 Raise alert on \hat{S}_t
 end if
end for

for the EWMA of the mean of the data \vec{x}_t , the mean of the residuals \vec{r}_t and the marginal variance of the residuals ($\hat{\sigma}_r^2$). Additionally, the memory parameter of the *iPCA* (namely, η) can be chosen by the user. Moreover, one can choose the control limit L , used in the alert phase, the variance cut-off percentage π and the robust guard estimation (REG) parameter R , controlling the sequential update of the EWMA.

2. **Initialization.** The “warm-up” dataset, undergoes a batch PCA. Using the percentage π that denotes the fraction of variance “explained”, one can determine the effective dimension k of the factor subspace to be estimated, which dictates the trends that would be filtered out. The mean of the data ($\hat{\nu}_x$), the mean ($\hat{\nu}_r$) and the marginal variance ($\hat{\sigma}_r^2$) of the residuals are initialized based on the training data.

3. **Sequential Update.** A new vector of data (\vec{x}_t) is observed and passed through the algorithm. We update the estimated factor subspace \hat{B} using incremental PCA *Arora et al.* (2012a). The vector $\hat{\vec{v}}_x$ is updated using EWMA.
4. **Residuals.** Depending on whether there is an ongoing anomaly, we center \vec{x}_t as $\vec{x}_t^{(0)} = \vec{x}_t - \hat{\vec{v}}_x$. We project the centered data onto the orthogonal complement of \hat{B} and obtain the residual (\vec{r}_t). Thus, the residuals are defined as

$$\vec{r}_t = \text{Proj}_{\text{col}(\hat{B})^\perp}(\vec{x}_t^{(0)}) = (I - \hat{B}(\hat{B}^\top \hat{B})^{-1} \hat{B}^\top) (\vec{x}_t^{(0)}) \quad (2.3)$$

If any coordinate of \vec{r}_t is smaller in magnitude than R times the marginal variance, then the appropriate elements of both $\hat{\vec{v}}_r$ and $\hat{\sigma}_r^2$ are updated via EWMA.

5. **Alerts.** Finally, if any coordinate in the centered residuals exceeds in magnitude the corresponding element in $L \times \hat{\sigma}_r$, an alert is raised.

There are two important points that we would like to make regarding the effectiveness of the proposed Algorithm.

- First, non-stationarity in the form of e.g. diurnal cycles in the data is absorbed in the projection step, and not with a time-series model.
- The sparse anomalous signals remain largely unaffected by the projection step. This can be quantified with the help of an incoherence condition (see Proposition II.3 and its proof in Section 2.4) that is generally satisfied in network traffic monitoring problems. This then leaves us with the residuals approximating the signals

$$\vec{r}_t \approx \vec{u}_t + \vec{\epsilon}_t, \quad (2.4)$$

and enables us to detect and locate sparse anomalies.

Moreover, the sequentially and non-parametrically updated estimate of the variance $\widehat{\sigma}_r^2$ on Step (4) of the algorithm can be shown to be a consistent estimator in the special case that our residuals $\{\vec{r}_t\}_{t=1}^\infty$ form a Gaussian time series. (More general cases could be examined as future work.) Indeed, let $\{r_t(j), t \in \mathbb{Z}, j \in [p] := \{1, \dots, p\}\}$ be a zero-mean stationary Gaussian time series, with some correlation structure

$$\rho_t(j) = \text{Cor}(r_t(j), r_0(j)) = \text{Cor}(r_{t+h}(j), r_h(j)), \quad (2.5)$$

for $t \in \mathbb{Z}, j \in [p], \forall h \in \mathbb{Z}$. We propose the estimator

$$\widehat{\sigma}_t^2(j) = (1 - \lambda_\sigma)\widehat{\sigma}_{t-1}^2(j) + \lambda_\sigma r_t^2(j), \quad (2.6)$$

in order to estimate the unknown variance of \vec{r}_t non-parametrically. Namely for the estimation of this variance, we implement an EWMA on the squares of the zero-mean stationary series $\{\vec{r}_t, t \in \mathbb{Z}\}$, which the following Proposition II.1 proves consistent.

Proposition II.1. *Let $\{\widehat{\sigma}_t^2, t \in \mathbb{Z}\}$ be defined as in (2.6) and $\{\rho_t, t \in \mathbb{Z}\}$ as in (2.5) (we have dropped the index j to keep the notation uncluttered). If $\sum_{t=-\infty}^\infty \rho_t^2 < \infty$, then the estimator $\widehat{\sigma}_t^2$ is consistent.*

The proof of this Proposition is presented in Section 2.4.3.

Remark II.2. Note that at first glance the EWMA estimator for the variance of the residuals in (2.6) seems inconsistent with the one in Algorithm 1, due to the lack of the centering term $\widehat{\nu}_r(j)$. Recall that our method firstly projects the original data in the orthogonal complement of the estimated subspace \widehat{B} , obtaining the residuals $\widehat{\vec{r}}_t$, which are then centered. The centered residuals correspond to the residuals that we utilize in the proof and setting of Proposition II.1, i.e., we assume they are a zero-mean Gaussian stationary time series with square summable correlation structure. Thus, the update of $\widehat{\sigma}_r^2(j)$ in (2.6) does not require the existence of a centering factor.

We now proceed to elaborate on the incoherence condition, which ensures that the sparse anomalies are largely retained by the projection step.

2.2.3 Incoherence conditions

Disentangling sparse errors from low rank matrices has been a topic of extensive research since the last decade. While there are seminal works on recovering sparse contamination from a low-rank matrix *exactly*, they all resort to solving a convex program or a program with convex regularization terms (*Candès and Recht, 2012; Candès et al., 2011; Xu et al., 2010*). The form that most closely resembles the set-up of our problem is *Zhou et al. (2010)*, where the model has both a Gaussian error term and a sparse corruption; see also *Xu et al. (2010), Candès et al. (2011)* and references therein.

This line of literature often criticizes the classical PCA in the face of outliers and data corruptions. However, to the best of our knowledge, there has been no analysis documenting just how or when the classical PCA becomes brittle in recovering sparse errors when the training period itself is contaminated. Our empirical and theoretical findings point to the contrary.

Secondly, rarely can such convex optimization-based methods be made “online”. Our approach here is to separate sparse errors from streaming data rather than stacked vectors of observations (i.e., measurement matrices) collected over long stretches of time. The low-rank matrix completion line of work faces challenges when data comes in streams rather than in batches, because the matrix nuclear norm often used in such problems closely couples all data points. In a notable effort by *Feng et al. (2013)* an iterative algorithm was proposed to solve the optimization problem with streaming data. However, implementing the algorithm requires extensive calibration and tuning which makes it impractical for non-stationary data streams with ever-shifting structures.

We now make formal the following statement: under the suitable structural assumption on the factor loading matrix B stated below, PCA is still a resilient tool for recovery of the low-rank component and the sparse errors on observations \vec{x}_t . Our results provide theoretical guarantees on the error in recovering the locations and magnitudes of sparse anomalies.

We analyze the faithfulness of residuals in (2.4) in recovering sparse anomalies under the following assumptions.

$$\lambda_{\min}(B^\top B) \geq \phi(p), \text{ for some function } \phi \text{ of } p. \quad (2.7)$$

Entries of B are bounded by a constant C , uniformly in p :

$$|B(i, j)| \leq C, \quad \forall i \in [p] := \{1, \dots, p\}, j \in [k] \quad (2.8)$$

Conditions (2.7) & (2.8) are closely linked to the so-called incoherence conditions in high-dimensional statistics (*Candès and Recht, 2012*). Notice that for (2.7) and (2.8) to hold simultaneously, we need $\phi(p) \leq Cp$. These conditions are important for the theoretical guarantees in the high-dimensional asymptotics regime where $p \rightarrow \infty$ but in practice they are quite mild and natural.

Condition (2.7) ensures that the background signal in the factor model $B\vec{f}_t$ has enough energy or a “ground-clearance” relative to the dimension. In the context of Internet traffic monitoring, this is not a restrictive condition since otherwise the background signal can be modeled with a lower-dimensional factor model with smaller value of k . The second assumption (2.8), taken together with (2.7) ensures that the columns of B are not sparse and consequently the background traffic is not concentrated on one or a few ports. This is also natural. Indeed, if the background was limited to a sparse subset of ports, then they would always behave differently than the majority of the ports and one can simply analyze these ports separately using

a lower-dimensional model with smaller value of p , where we no longer have sparse background.

Let now $\Sigma = BB^\top$ and $\widehat{\Sigma}$ be an estimate of Σ obtained for example, by performing iPCA and taking the top- k principal components. (Alternatively, one can simply take $\widehat{\Sigma} := n^{-1} \sum_{t=1}^n \vec{x}_t \vec{x}_t^\top$, for a window of n past observations.)

Proposition II.3 (Resilience of PCA). *Assume (2.7) and (2.8) hold and let \vec{r}_t be defined as in (2.3). Then, for any k and p , and for each coordinate $i \in [p]$,*

$$\mathbb{E}(r_t(i) - (\epsilon_t(i) + u_t(i)))^2 \leq \left[\left(\mathbb{E} \|\widehat{\Sigma} - \Sigma\|^2 \right)^{1/2} \frac{2\sqrt{kp}}{\phi(p)} \left(c_f C k + 1 + \sqrt{\frac{\text{tr}(\Sigma_u)}{p}} \right) + \frac{\sqrt{k}C}{\phi(p)} \left(1 + \sqrt{kp} C \|\Sigma_u\|^{1/2} \right) \right]^2 \quad (2.9)$$

where $\Sigma_u = \mathbb{E}[\vec{u}_t \vec{u}_t^\top]$ and $(\mathbb{E} \|\vec{f}_t\|^2) \leq c_f^2 k$. Moreover, in ℓ_2 -norm, we have

$$\mathbb{E} \|\vec{r}_t - (\vec{\epsilon} + \vec{u})\|^2 \leq \left[\left(\mathbb{E} \|\widehat{\Sigma} - \Sigma\|^2 \right)^{1/2} \frac{2\sqrt{pk}}{\phi(p)} \left(c_f C k + 1 + \sqrt{\frac{\text{tr}(\Sigma_u)}{p}} \right) + \sqrt{k} + \sqrt{\min\{k \|\Sigma_u\|, \text{tr}(\Sigma_u)\}} \right]^2 \quad (2.10)$$

In Section 2.4.1, we provide a proof for this proposition.

This result shows that if $\sqrt{p} = \mathcal{O}(\phi(p))$, with k and Σ_u bounded, the upper bounds in (2.9) converge to zero, as the dimension p grows. **That is, the anomaly signal \vec{u}_t “passes through” to the residuals \vec{r}_t and the approximation in (2.4) can be quantified.** Indeed, assuming k is fixed for a moment, (2.9) entails that the point-wise bound on mean-squared difference between the unobserved anomaly-plus-noise signal $u_t(i) + \epsilon_t(i)$ and the residuals $r_t(i)$ obtained from our algorithm is $\mathcal{O}(\mathbb{E}(\|\widehat{\Sigma} - \Sigma\|^2)p/\phi^2(p))$. This means that in practice, provided $\Sigma = BB^\top$ is estimated well, anomalies $u_t(i)$ of magnitudes exceeding $\sqrt{p}/\phi(p)$ will be present in the residuals

$r_t(i)$. As seen in the lower bound in Eq. (2.14), this is a rather mild restriction. Thus, in view of Theorem II.4, provided $\phi(p) \gg \sqrt{p/\log(p)}$, the theoretically optimal exact identification of all sparse anomalies is unaffected by the background signal $B\vec{f}_t$.

In practice, however, the key caveat is the accurate estimation of Σ , which can be challenging if the noise $\vec{\epsilon}_t$ and/or the factor signals \vec{f}_t are long-range dependent. Further analysis, not presented here, establishes upper bounds on $\mathbb{E}[\|\hat{\Sigma} - \Sigma\|^2]$ via the Hurst long-range dependence parameter of the time-series $\{\epsilon_t(i)\}$, conditions on $\{f_t\}$, and the length of the training window n . It shows that even in the presence of long-range dependence, provided the training window and the memory of iPCA are sufficiently large, the background signal can be effectively filtered out without affecting the theoretical boundary for exact support recovery discussed in the next section.

2.2.4 Statistical Limits in Sparse Anomaly Identification

As argued in the previous section using for example PCA-based filtering methods, one can remove complex non-sparse spatio-temporal background/trend signal without affecting significant *sparse* anomalies. Therefore, the sparse anomaly detection and identification problems can be addressed transparently in the context of the signal-plus-noise model. Namely, assume that we have a way to filter out the background traffic $B\vec{f}$ induced by multivariate non-sparse trends² in (2.2). Thus, we suppose that we directly observe the term:

$$\vec{x} = \vec{\epsilon} + \vec{u},$$

where \vec{u} is a sparse vector with $s \ll p$ non-zero entries. In this context, we have two types of problems:

²We drop the time subscript to keep the notation uncluttered.

- **Detection problem.** Test the hypotheses

$$\mathcal{H}_0 : \vec{u} = 0 \text{ (no anomalies)} \text{ vs } \mathcal{H}_1 : \vec{u} \neq 0. \quad (2.11)$$

- **Identification problem.** Estimate the *sparse support set*

$$S = S_p := \{i \in [p] : u(i) \neq 0\} \quad (2.12)$$

of the locations of the anomalies in \vec{u} .

Starting from the seminal works of *Ingster* (1998b) and *Donoho and Jin* (2004b) the fundamental statistical limit of the *detection problem* has been studied extensively under the so-called high-dimensional asymptotic regime $p \rightarrow \infty$ (see, e.g., Theorem 3.1 in the recent monograph (*Gao and Stoev*, 2021) and the references therein). The identification problem and more precisely the exact recovery of the support S_p have only been addressed recently (*Butucea et al.*, 2018; *Gao and Stoev*, 2020). To illustrate, suppose that the errors have standard Gaussian distributions $\epsilon_t(i) \sim \mathcal{N}(0, 1)$ and are independent in i (more general distributional assumptions and results on *dependent* errors have been recently developed in (*Gao and Stoev*, 2021)). Consider the following parameterization of the anomalous signal support size and amplitude as a function of the dimension p :

- **Support sparsity.** For some $\beta \in (0, 1]$,

$$|S| \asymp p^{1-\beta}, \quad \text{as } p \rightarrow \infty. \quad (2.13)$$

The larger the β , the sparser the support.

- **Signal amplitude.** The non-zero signal amplitude satisfies

$$\sqrt{2\underline{r} \log(p)} \leq u(i) \leq \sqrt{2\bar{r} \log(p)}, \quad \text{for all } i \in S_p. \quad (2.14)$$

We have the following phase-transition result (see, e.g., Theorem 3.2 in *Gao and Stoev (2021)*).

Theorem II.4 (Exact support recovery). *Let the signal \vec{u} satisfy the sparsity and amplitude parameterization as in (2.13) and (2.14), and let*

$$g(\beta) = (1 + \sqrt{1 - \beta})^2, \quad \beta \in [0, 1].$$

1. *If $\bar{r} < g(\beta)$, then for any signal support estimator \hat{S}_p , we have*

$$\lim_{p \rightarrow \infty} \mathbb{P} \left[\hat{S}_p = S_p \right] = 0.$$

2. *If $g(\beta) < \underline{r}$, then for the thresholding support estimator $\hat{S} := \{j \in [p] : x(j) > t_p\}$ with $\mathbb{P}[\epsilon(i) > t_p] \sim \alpha(p)/p$, with $\alpha(p) \rightarrow 0$ such that $\alpha(p)p^\delta \rightarrow \infty, \forall \delta > 0$, we have*

$$\lim_{p \rightarrow \infty} \mathbb{P} \left[\hat{S}_p = S_p \right] = 1.$$

The above result establishes the statistical limits of the anomaly identification problem known as the *exact support recovery* problem as $p \rightarrow \infty$. It shows that for β -sparse signals with amplitudes below the boundary $g(\beta)$, there are no estimators that can fully recover the support. At the same time if the amplitude is above that boundary, suitably calibrated thresholding procedures are optimal and can recover the support exactly with probability converging to 1 as $p \rightarrow \infty$. Based on the concentration of maxima phenomenon, this phase transition phenomenon was shown to hold even for strongly dependent errors $\epsilon(i)$'s for the broad class of thresholding proce-

dures (Theorem 4.2 in (*Gao and Stoev, 2021*)). This recently developed theory shows that “simple” thresholding procedures are optimal in identifying sparse anomalies. It also shows the fundamental limits for the signal amplitudes as a function of sparsity where signals with insufficient amplitude cannot be fully identified in high dimension by *any* procedure. These results show that our thresholding-based algorithms for sparse anomaly *identification* are essentially optimal in high-dimensions provided the background signal can be successfully filtered out. More on the sub-optimality of certain popular *detection* procedures can be found in Section 2.3.2.

2.3 Performance evaluation

Next, we provide experimental assessments of our methodology using both *synthetic* and *real-world* traffic traces. We utilize two criteria to provide answers to the two-fold objective of our analysis. For the detection problem, we are only interested on whether any observation at a fixed time point is flagged as an anomaly. If so, the whole vector is treated as an anomalous vector and any F1-scores or ROC curves are “global” (i.e., for any port) and time-wise. On the other hand, for the identification problem, we are also interested in correctly flagging the positions of the anomalous data. Namely, looking at a vector of observations, representing the port space at a specific time point, we want to correctly find which individual ports/streams include the anomalous activity. We report a percentage of correct anomaly identifications per time unit.

2.3.1 Performance and Calibration using Synthetic Data

2.3.1.1 Linear Factor Model

We simulate “baseline” traffic data of 5 weeks via the linear factor model of Eq. 2.2, obtaining one observation every 2 minutes. This leads to a total of 25200 observa-

tions. The baseline traffic is created as a fractional Gaussian noise (fGn) with Hurst parameter $H = 0.9$ and variance set to 1, leading to a long range dependent sequence. An extensive analysis not shown here yields long-range dependence exponents around $H = 0.9$.

Together with the fGn, five extra sinusoidal curves—representing traffic trends—are blended in to create the final time-series. These sinusoidal curves reflect the diurnal and weekly trends that show up in real world Internet traffic. The first two inserted trends are daily, the third one is weekly and the last two of them have a period of 6 and 4.8 hours respectively. These trends do not affect every port in the same way; they all have a random offset and only influence the ports determined by our factor matrix B . In this matrix every row represents a port and the column specifies the trend; a presence of 1 in the position (i, j) means that the j -th trend is inserted in the i -th port. The matrix B is created by having all elements of the first column equal to 1, meaning that all ports are affected by the first trend. The number of ones in the rest of the columns is equal to $(1 - \lfloor (j - 1)/k \rfloor) * 100\%$, where j is the index of the trend and k the total number of them. The ones are distributed randomly in the columns.

We simulate traffic data that pertain to N distinct ports, leading to a matrix of observations of dimensions $N \times 25200$ (e.g., $N = 100$). Since we are interested in tuning the parameters of our model, we create 5 independent replications of this dataset. The first two weeks of observations, namely the first 10080 observations, are used as the warm-up period where the initial batch PCA is run in order to determine the number of principal components we will work with in the incremental PCA and proceed with initialization.

The anomaly is inserted in the start of the fourth week; the magnitude is determined by the *signal to noise ratio* (snr) and the number of anomalous observations depends on the given *duration*. We have three choices of snr and two choices of

duration in this simulation setup, leading to six combinations. The options for snr are 2, 3 and 7, leading to an additional traffic of size 2, 3 and 7 times the empirical standard deviation of the corresponding port the anomaly is inserted into. Moreover, the duration of the anomalies we use is 1 hour for our “short” anomalies, or 30 observations in this specific data, and 6 hours for our “long” anomalies; 180 observations respectively. The anomalies are always inserted in the first 3 simulated ports.

2.3.1.2 Demonstration of Sequential PCA

Figure 2.2 illustrates the performance of the incremental PCA algorithm as a function of its memory parameter—the reciprocal of the current time index parameter in *Arora et al. (2012b)*; *Degras and Cardot (2019)*. Since our methodology depends on how well we approximate the factor subspace, we need to measure distance between subspaces of \mathbb{R}^p . We do so in terms of the largest principal angle $\angle(\widehat{W}, W)$ (*Bjorck et al. (1973)*; *Yu et al. (2015)*) between the known true subspace W and the estimated one \widehat{W} , defined as:

$$\angle(\widehat{W}, W) := \text{acos}(\sigma_1), \quad \text{where } \sigma_1 = \max_{\substack{\hat{w} \in \widehat{W}, \bar{w} \in W \\ \|\hat{w}\| = \|\bar{w}\| = 1}} \bar{w}^\top \hat{w}.$$

As we can see in Figure 2.2, there is a “trade-off” with regards to the length of the memory and the performance. If this parameter is chosen to be too big, the angle of the two subspaces becomes the largest. This could be explained as relying too much on the latest observations, estimating a constantly changing subspace. If the memory is too long, then the impact of new observations is minimal, leading to very little adaptability of our estimator. In Figure 2.2, for the data used here, using 10 weeks worth of observations and 10 replications of the data, the best memory parameter seems to be 10^{-5} . Moreover, as expected a full PCA estimated subspace has smaller angle than the iPCA estimated one (see Figure 2.2). Some theoretical

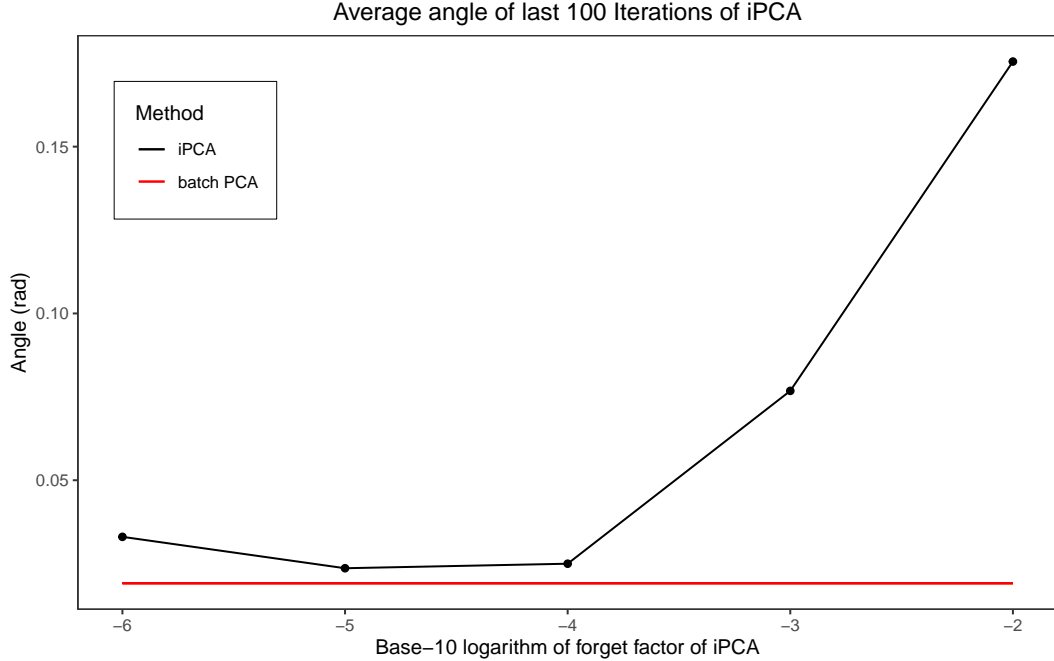


Figure 2.2: The largest principal angle between estimated and true subspace for iPCA (cf *Bjorck et al. (1973)*; *Yu et al. (2015)*). The horizontal line (in red) shows the largest principal angle between the true subspace and the batch PCA estimated one.

derivations regarding upper bounds on this angle between the space B (as in (2.2)) and the estimator \hat{B} , based on the expected distance between the covariance and the estimated covariance matrix of the data, can be found in Section 2.4.2.

2.3.1.3 Parameter Tuning: Operator’s Guide

Next, we proceed with hyper-parameter tuning for Algorithm 1. Our grid search for tuning involves the following parameters and values. We have an EWMA memory parameter for the data (λ), keeping track of their mean, so that this mean can be utilized in the incremental PCA. We also have an EWMA parameter keeping track of the mean of the residuals (λ_μ) in our algorithm. Both of these parameters are explored over the values 10^{-2} , 10^{-3} and 10^{-4} . Additionally, there is an EWMA parameter for the sequential updating of the marginal variance of the residuals (λ_σ); the possible values we examine are 10^{-4} , 10^{-5} and 10^{-6} . Finally, we explore the values of control

Table 2.1: Recommendations for the choice of tuning parameters

snr	duration	L	REG	ewma_data	ewma_mean	ewma_var
2	1	5	4	0.0010	0.0001	0.00010
5	1	7	5	0.0010	0.0010	0.00001
7	1	7	5	0.0001	0.0100	0.00010
2	6	5	3	0.0001	0.0010	0.00010
5	6	7	5	0.0010	0.0010	0.00001
7	6	7	3	0.0001	0.0100	0.00010

limit (L) that work best with the detection of anomalies in this data; we have a list of values including 0, 10^{-4} , 10^{-3} , 10^{-2} , 0.1, 0.5, 1, 2, 3, 4, 5, 6, 7, 20.

The best choices for our tuning parameters are evaluated based on a combination of $F1$ -score and area under curve (AUC) value and can be found on Table 2.1. We start by selecting a combination of snr and $duration$ and we find the AUC for each different combination of the rest of the tuning parameters with respect to L . We choose the parameters that lead to the largest AUC and then elect the control limit that corresponds to the largest $F1$ -score. Our definition of $F1$ -score is based on individual flagging of observations as discoveries or not; namely we look at the 15120×100 “test” observations as a vector of 1512000 observations and each one of them is categorized as true/false positive or true/false negative.

2.3.2 Sub-optimality of classic Chi-square statistic detection methods

The detection problem can be understood transparently in the signal-plus-noise setting. Namely, thanks to Proposition II.3, suppose that we observe a high dimensional vector as $\vec{x}_t = \vec{u}_t + \vec{\epsilon}_t$, where $\vec{u}_t = (u_t(i))_{i=1}^p$ is a (possibly zero) anomaly signal. Then, the detection problem can be cast as a (multiple) testing problem

$$\mathcal{H}_0 : (u_t(i))_{i=1}^p = \vec{0} \text{ vs } \mathcal{H}_a : u_t(i) \neq 0, \text{ for some } i \in [p].$$

Under the assumption that $\vec{\epsilon}_t \sim N(0, \Sigma_{p \times p})$, one popular statistic for detecting an

anomaly (on any port), i.e., testing \mathcal{H}_0 is:

$$Q = \bar{x}^\top \Sigma^{-1} \bar{x} \sim \chi_p^2, \text{ under } \mathcal{H}_0.$$

Note that the statistic is a function of time, but from now on, for simplicity we shall suppress the dependence on t . Using this statistic we have the following rule: We reject the H_0 at level $\alpha > 0$, if $Q > \chi_{p,1-\alpha}^2$.

Coming back to our algorithm, we have a control limit L , which we can use to raise alerts in our alert matrix. Namely, we raise an alert on port i at time t , if the i -th residual at time t is bigger than L times the marginal variance of these residuals. Note that at every time-point, Algorithm 1 (i.e., the iPCA-based method) gives us individual alerts for anomalous ports, while the Q-statistic only flags presence/absence of an anomaly over the port dimension. To perform a fair comparison of the two methods, we use the following definition of true positives and false negatives for Algorithm 1. We declare an alert at time t a true positive, if there is at least one anomaly in the port space at time t and Algorithm 1 raises at least one alarm at time t ; not necessarily at a port where the anomaly is taking place. A false negative takes place if an anomaly is present at the port space, but no alerts are being raised across the port space at time t .

To compare the two methods, we start with an initial warm-up period which we apply a batch PCA approach to. Then, we use the estimate of the variance for the Q-statistic method, while we also obtain the number of principal components to keep track of during the iPCA, using the number of principal components that explain 90% of the variability of this warm sample. We choose the tuning parameters for iPCA based on the recommendations provided in Section 2.3.1.3.

Figure 2.3 illustrates the performance of the two methods as the number of ports increases, while the sparsity of the anomalies in the simulated traffic remains the

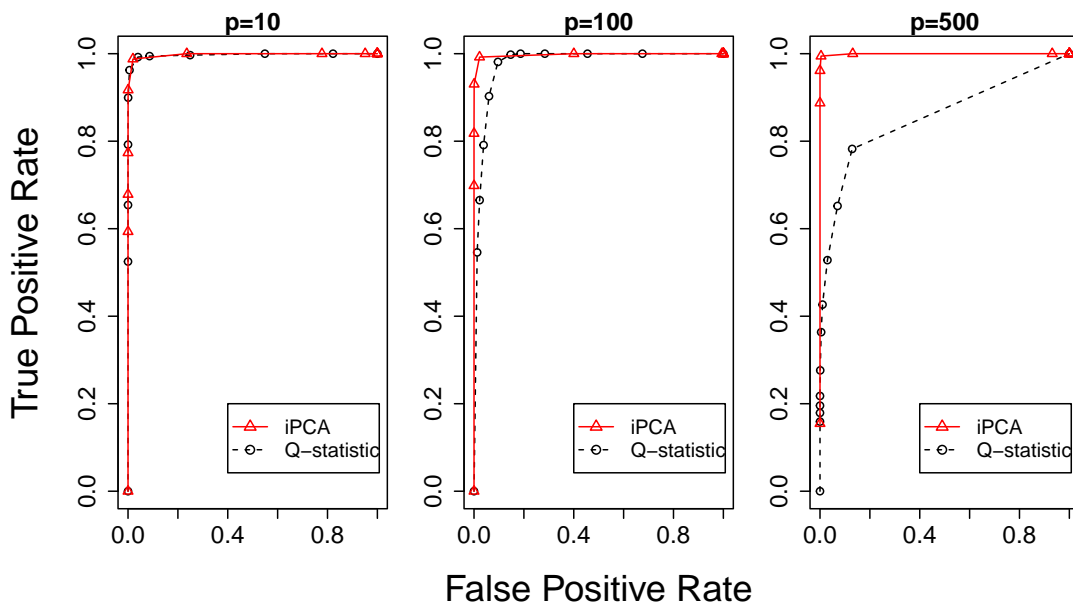


Figure 2.3: ROC curves of Algorithm 1 (iPCA-based) and Q-statistic for long duration anomaly (6 hour) of low magnitude ($\text{snr}=2$) in synthetic data.

same. The *sparsity parameter* β (see *Gao and Stoev (2021)*) is used to control the sparsity of the anomalies; for a number of ports p we insert anomalies in $\lfloor p^{1-\beta} \rfloor$ ports, where $\beta = 3/4$. As is evident from the plots, Algorithm 1 significantly outperforms the Q-statistic as the port dimension grows.

This sub-optimality phenomenon is well-understood in the high-dimensional inference literature (see (*Fan, 1996*) and Theorem 3.1 in (*Gao and Stoev, 2021*)). Namely, as $p \rightarrow \infty$, the Q-statistic based detection method will have vanishing power in detecting sparse anomalies in comparison with the optimal methods such as Tukey’s higher-criticism statistic. As shown in Theorem 3.1 in (*Gao and Stoev, 2021*), the thresholding methods like Algorithm 1 are also optimal in the very sparse regime $\beta \in [3/4, 1]$ in the sense that they can discover anomalous signals with magnitudes down to the theoretically possible *detection boundary* (cf *Ingster (1998b)*). **This explains the growing superiority of our method as the port dimension increases, as depicted in Figure 2.3.**

Table 2.2: Performance of Algorithm 1 on Synthetic Data. Observe the high True Positive Rate (TPR) and low False Positive Rate (FPR) for carefully chosen thresholds L .

L	tpr_rows	fpr_rows	tpr_indiv	fpr_indiv
2	1.00	1.00	1.00	1.00
3	1.00	0.92	1.00	0.92
4	1.00	0.11	0.99	0.11
5	1.00	0.00	0.97	0.00
6	1.00	0.00	0.93	0.00
7	1.00	0.00	0.87	0.00
20	0.08	0.00	0.02	0.00

At the same time, the proposed iPCA-based method also has the advantage of providing us with extra information in comparison to the Q-statistic. Indeed, using the individual time/space alerts, one can identify what percentage of individual true positives/false negatives exist in the time/port space. An example of this information is shown on Table 2.2, where the “rows” variables refer to the detection problem and “indiv” variables refer to the identification one. Namely, the “rows” variables are only concerned with temporal events, while the “indiv” variables take into account the spatial structure as well.

2.3.3 Application to real-world Darknet data

Finally, we demonstrate the performance of our method on real-world data obtained from the ORION Network Telescope *Merit Network, Inc.* (2022). The dataset spans the entire month of September 2016 and includes the early days of the infamous Mirai botnet *Antonakakis et al.* (2017). We constructed minute-wise time series for all TCP/UDP ports present in our data that represent the number of unique sources targeting a port at a given time. For the analysis here, we focused on the top-50 ports based on their frequency in the duration of a month. We implement the algorithm on the data by utilizing the calibrated tuning parameters suggested in 2.3.1.3. We use the first 5000 observations, roughly three and a half days, as our burn-in period to initialize the algorithm. Thus, we have a matrix of 43200×50 Darknet data

observations.

The selected dataset includes important security incidents. **Indeed, as Figure 2.4 shows, we detect the onset of scanning activities against TCP/2323 (Telnet for IoT) around September 6th that can be attributed to the Mirai botnet.** We also detect some interesting ICMP-related activities (that we represent as port-0) towards the end of the month. We are unsure of exactly what malicious acts these activities represent; however, upon payload inspection we found them to be related with some heavy DNS-related scans (payloads with DNS queries were encapsulated in the ICMP payloads).

In Figures 2.4 and 2.5, we are looking for large anomalies (snr=7) of long duration (6 hours). Moreover, inspecting the full alert matrix in Figure 2.5, we observe a few more alerts for anomalies that deserve further investigation. Some of them might be false positives, although there is no definitive “ground truth” in real-world data and all alerts merit some further forensics analysis. The encouraging observations from Figures 2.4 and 2.5 are that the incidents knownst to us are revealed, and that the elected hyper-parameters avoid causing the so-called “alert-fatigue” to the analysts. At the same time, the analysts could tune Algorithm 1 to their preferences, and prioritize which alerts to further investigate based on attack severity, duration, number of concurrently affected ports, etc. All this information is readily available from our methodology.

2.4 Theoretical Results

2.4.1 Proof for Proposition II.3

We introduce some further notations before we present the proof.

Let $b(\vec{j})$ be the j -th column of B , $j \in [k]$; $b_0(\vec{j}) = b(\vec{j})/\|b(\vec{j})\|$ the factor loadings normalized; $B_0 := [b_0(\vec{1}), \dots, b_0(\vec{k})]$. We shall denote by \hat{B} and \hat{B}_0 the corresponding

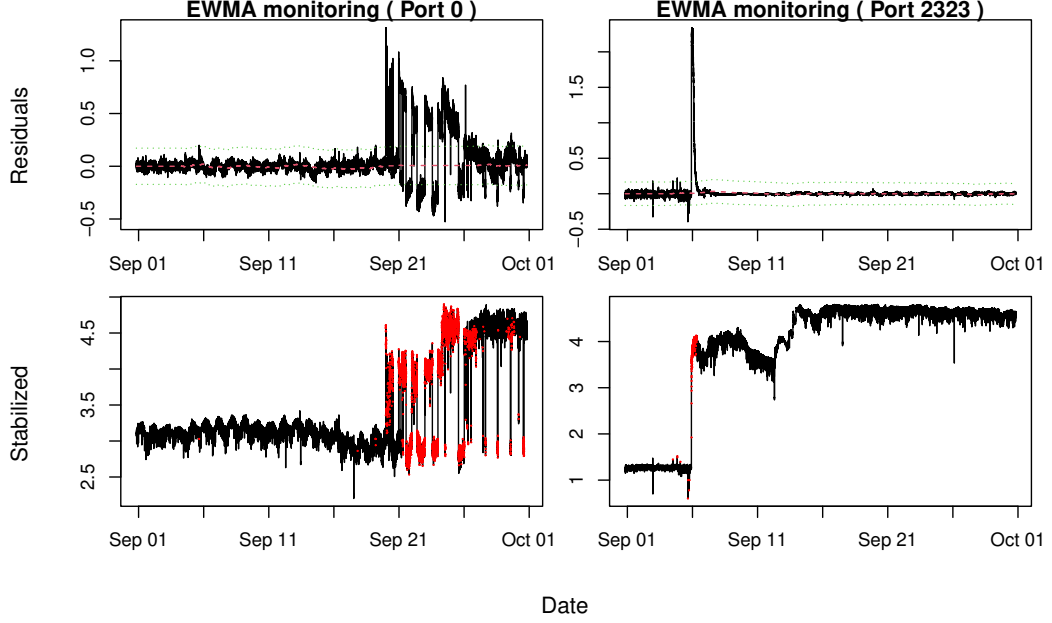


Figure 2.4: Residuals and detection boundaries (top) and raw traffic (bottom) for ports 0 and 2323. Red color in bottom plots shows the detected anomalies.

estimates of B and B_0 obtained for example by the iPCA algorithm.

For the purpose of the following theoretical results, we shall assume that \hat{B} and \hat{B}_0 are obtained using singular value decomposition of the sample covariance matrix

$$\hat{\Sigma} := \frac{1}{n} \sum_{t=1}^n \vec{x}_t \vec{x}_t^\top. \quad (2.15)$$

Proof. Residuals \vec{r} from projecting observations onto the space spanned by the first k eigenvectors, are

$$\begin{aligned} \vec{r} &= (\mathcal{I}_p - \mathcal{P}_{\hat{B}_0})(B\vec{f} + \vec{\epsilon} + \vec{u}) \\ &= (\mathcal{I}_p - \mathcal{P}_{\hat{B}_0})B\vec{f} + (\mathcal{I}_p - \mathcal{P}_{\hat{B}_0})\vec{\epsilon} + (\mathcal{I}_p - \mathcal{P}_{\hat{B}_0})\vec{u} \end{aligned}$$

or equivalently, we can write

$$\vec{r} - \vec{\epsilon} - \vec{u} = \underbrace{(\mathcal{I}_p - \mathcal{P}_{\hat{B}_0})B\vec{f}}_I - \underbrace{\mathcal{P}_{\hat{B}_0}\vec{\epsilon}}_{II} - \underbrace{\mathcal{P}_{\hat{B}_0}\vec{u}}_{III}$$

Alert Matrix For Select Ports, p=15

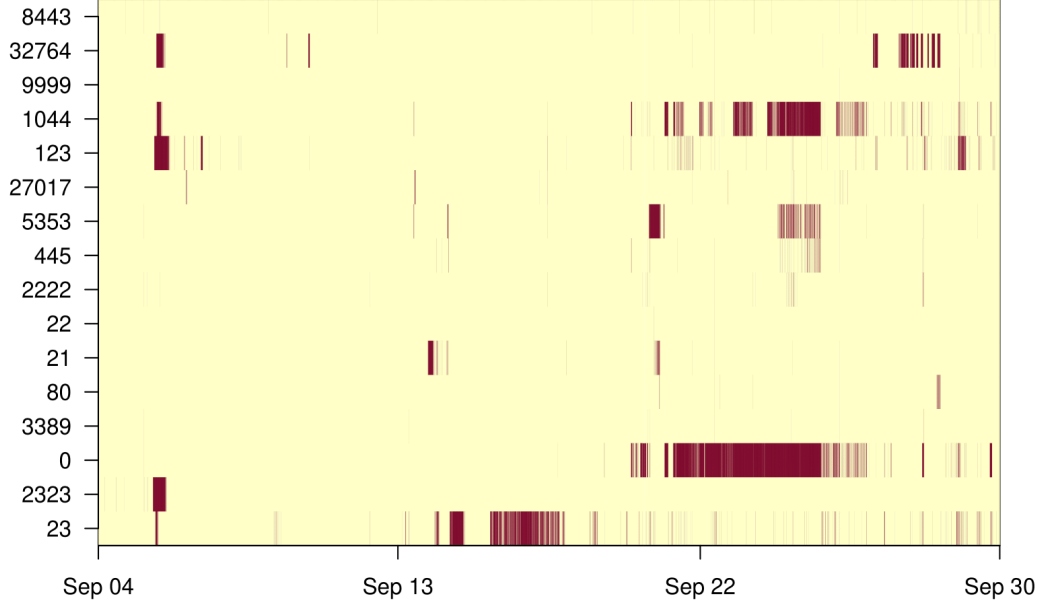


Figure 2.5: Alert matrix for the Darknet dataset for September 2016 (Mirai onset). Red color indicates detection.

We will establish upper bounds on all three terms. The first term,

$$\begin{aligned}
 I &= (\mathcal{I}_p - \mathcal{P}_{B_0} + \mathcal{P}_{B_0} - \mathcal{P}_{\widehat{B}_0})B\vec{f} \\
 &= (\mathcal{I}_p - \mathcal{P}_{B_0})B\vec{f} + (\mathcal{P}_{B_0} - \mathcal{P}_{\widehat{B}_0})B\vec{f} \\
 &= -(\mathcal{P}_{\widehat{B}_0} - \mathcal{P}_{B_0})B\vec{f}.
 \end{aligned}$$

Using the relationship

$$\|B\| \leq \sqrt{p} \max_{i=1, \dots, p} \left(\sum_{j=1}^k B_{i,j}^2 \right)^{1/2}$$

and (2.8) we obtain that $\|B\| \leq \sqrt{pk}C$, and also $(\mathbb{E}\|\vec{f}\|)^2 \leq \mathbb{E}[\|\vec{f}\|^2] \leq c_f^2 k$. By

Corollary II.6 discussed below,

$$\begin{aligned}
\mathbb{E}\|(\mathcal{P}_{\widehat{B}_0} - \mathcal{P}_{B_0})B\vec{f}\| &\leq \mathbb{E}\left[\|(\mathcal{P}_{\widehat{B}_0} - \mathcal{P}_{B_0})\|\|B\|\|\vec{f}\|\right] \\
&\leq \|B\|\left(\mathbb{E}\|\mathcal{P}_{\widehat{B}_0} - \mathcal{P}_{B_0}\|^2\right)^{1/2}\left(\mathbb{E}\|\vec{f}\|^2\right)^{1/2} \\
&\leq \sqrt{pk}C\frac{c_f\sqrt{k}}{\lambda_{\min}}2\sqrt{k}\left(\mathbb{E}\|\widehat{\Sigma} - \Sigma\|^2\right)^{1/2} \\
&\leq \frac{2c_fCk\sqrt{pk}}{\phi(p)}\left(\mathbb{E}\|\widehat{\Sigma} - \Sigma\|^2\right)^{1/2}
\end{aligned} \tag{2.16}$$

where we used (2.7) in the last inequality and which also provides a bound on each coordinate.

For the second term, II , decomposed as $\vec{\epsilon} - \mathcal{P}_{B_0}\vec{\epsilon} - (\mathcal{P}_{\widehat{B}_0} - \mathcal{P}_{B_0})\vec{\epsilon}$, notice that $\mathcal{P}_{B_0}\vec{\epsilon} = B_0B_0^\top\vec{\epsilon} \stackrel{d}{\sim} B_0\vec{\epsilon}_k$ where $\mathbb{E}\vec{\epsilon}_k = 0$ and $\mathbb{E}\vec{\epsilon}_k\vec{\epsilon}_k^\top = I_k$. The second moment of the i th coordinate is bounded by

$$\mathbb{E}(\mathcal{P}_{B_0}\vec{\epsilon})_i^2 = \mathbb{E}\left(\sum_{j=1}^k B_0(i, j)\epsilon_k(j)\right)^2 \leq \frac{kC^2}{\lambda_{\min}}$$

because $|B_0(i, j)| \leq C/\sqrt{\lambda_{\min}}$. While in ℓ_2 norms, $\mathbb{E}\|\mathcal{P}_{B_0}\vec{\epsilon}\|^2 = \mathbb{E}\text{tr}(B_0B_0^\top\vec{\epsilon}\vec{\epsilon}^\top B_0B_0^\top) = k$. The last part is controlled similarly as in I . Indeed, by the Cauchy-Schwartz inequality, Corollary II.6 and (2.7), we obtain

$$\begin{aligned}
\left(\mathbb{E}\|(\mathcal{P}_{\widehat{B}_0} - \mathcal{P}_{B_0})\vec{\epsilon}\|\right)^2 &\leq \mathbb{E}\|\mathcal{P}_{\widehat{B}_0} - \mathcal{P}_{B_0}\|^2\mathbb{E}\|\vec{\epsilon}\|^2 \\
&\leq \frac{4k\mathbb{E}\|\widehat{\Sigma} - \Sigma\|^2p}{\phi(p)^2}.
\end{aligned}$$

For the third term, we need the von Neumann's trace inequality: for two matrices X and Y , $|\text{tr}(XY)| \leq \langle \vec{\sigma}(X), \vec{\sigma}(Y) \rangle$, where $\vec{\sigma}(X)$, $\vec{\sigma}(Y)$ are vectors of the singular values of X and Y . By Holder inequality, $|\text{tr}(XY)| \leq \|\vec{\sigma}(X)\|_1\|\vec{\sigma}(Y)\|_\infty = \text{tr}(X)\|Y\|$, for real symmetric positive definite matrices.

Decompose III as $\vec{u} - \mathcal{P}_{B_0}\vec{u} - (\mathcal{P}_{\widehat{B}_0} - \mathcal{P}_{B_0})\vec{u}$. The second moment of the l -th

coordinate

$$\begin{aligned}
\mathbb{E}(\mathcal{P}_{B_0}u)_l^2 &= \mathbb{E}\left(\sum_{i=1}^p u_i \underbrace{\left(\sum_{j=1}^k B_0(lj)B_0(ij)\right)}_{a_{il}}\right)^2 \\
&= \mathbb{E}\left(\sum_{i=1}^p a_{il}u_i\right)^2 = \text{tr}(\vec{a}_l\vec{a}_l^\top \mathbb{E}\vec{u}\vec{u}^\top) \\
&\leq \text{tr}(\vec{a}_l\vec{a}_l^\top)\|\Sigma_u\| \leq \frac{k^2 C^4 p \|\Sigma_u\|}{\lambda_{\min}^2},
\end{aligned}$$

where $\vec{a}_l = (a_{1l}, \dots, a_{pl})^\top$, and the last inequality because $|a_{il}| \leq kC^2/\lambda_{\min}$. In l_2 norm, $\mathbb{E}\|\mathcal{P}_{B_0}\vec{u}\|^2 = \text{tr}(B_0B_0^\top \Sigma_u) \leq \min\{k\|\Sigma_u\|, \text{tr}(\Sigma_u)\}$. The last part in *III*, again using Corollary II.6 and (2.7), is bounded by

$$\begin{aligned}
(\mathbb{E}\|(\mathcal{P}_{\hat{B}_0} - \mathcal{P}_{B_0})\vec{u}\|)^2 &\leq \mathbb{E}\|\mathcal{P}_{\hat{B}_0} - \mathcal{P}_{B_0}\|^2 \mathbb{E}\text{tr}(\vec{u}\vec{u}^\top) \\
&\leq \frac{4k\mathbb{E}\|\hat{\Sigma} - \Sigma\|^2 \text{tr}(\Sigma_u)}{\phi(p)^2}
\end{aligned}$$

Taken together,

$$\begin{aligned}
\mathbb{E}(\vec{r} - \vec{c} - \vec{u})_i^2 &\leq \left[\sqrt{\mathbb{E}(I_i)^2} + \sqrt{\mathbb{E}(II_i)^2} + \sqrt{\mathbb{E}(III_i)^2} \right]^2 \\
&\leq \left[\frac{2c_f C k \sqrt{pk}}{\phi(p)} \left(\mathbb{E}\|\hat{\Sigma} - \Sigma\|^2\right)^{1/2} + \frac{2\sqrt{kp}}{\phi(p)} \left(\mathbb{E}\|\hat{\Sigma} - \Sigma\|^2\right)^{1/2} \right. \\
&\quad \left. + \frac{\sqrt{k}C}{\phi(p)} + \frac{kC^2\sqrt{p}\|\Sigma_u\|^{1/2}}{\phi(p)} + \frac{2\sqrt{k} \left(\mathbb{E}\|\hat{\Sigma} - \Sigma\|^2\right)^{1/2} \sqrt{\text{tr}(\Sigma_u)}}{\phi(p)} \right]^2 \\
&= \left[\left(\mathbb{E}\|\hat{\Sigma} - \Sigma\|^2\right)^{1/2} \frac{2\sqrt{kp}}{\phi(p)} \left(c_f C k + 1 + \sqrt{\frac{\text{tr}(\Sigma_u)}{p}}\right) \right. \\
&\quad \left. + \frac{\sqrt{k}C}{\phi(p)} \left(1 + \sqrt{kp}C\|\Sigma_u\|^{1/2}\right) \right]^2
\end{aligned}$$

and

$$\begin{aligned}
\mathbb{E}\|\vec{r} - \vec{\epsilon} - \vec{u}\|^2 &\leq \left[\sqrt{\mathbb{E}\|I\|^2} + \sqrt{\mathbb{E}\|II\|^2} + \sqrt{\mathbb{E}\|III\|^2} \right]^2 \\
&\leq \left[\frac{2c_f Ck\sqrt{kp}(\mathbb{E}\|\hat{\Sigma} - \Sigma\|^2)^{1/2}}{\phi(p)} + \frac{2\sqrt{kp}(\mathbb{E}\|\hat{\Sigma} - \Sigma\|^2)^{1/2}}{\phi(p)} + \sqrt{k} \right. \\
&\quad \left. + \sqrt{\min\{k\|\Sigma_u\|, \text{tr}(\Sigma_u)\}} + \frac{2\sqrt{k}(\mathbb{E}\|\hat{\Sigma} - \Sigma\|^2)^{1/2}\sqrt{\text{tr}(\Sigma_u)}}{\phi(p)} \right]^2 \\
&= \left[(\mathbb{E}\|\hat{\Sigma} - \Sigma\|^2)^{1/2} \frac{2\sqrt{pk}}{\phi(p)} \left(c_f Ck + 1 + \sqrt{\frac{\text{tr}(\Sigma_u)}{p}} \right) \right. \\
&\quad \left. + \sqrt{k} + \sqrt{\min\{k\|\Sigma_u\|, \text{tr}(\Sigma_u)\}} \right]^2
\end{aligned}$$

□

2.4.2 Bounds on difference of projection operators

Let $\sigma_1 \geq \dots \geq \sigma_k$ be the singular values of $\widehat{B}_0^\top B_0$. The k dimensional vector $(\cos^{-1}(\sigma_1), \dots, \cos^{-1}(\sigma_k))$, called the principal angles, are a generalization of the acute angles between two vectors. Let $\Theta(\widehat{B}_0, B_0)$ be the $k \times k$ diagonal matrix with j -th diagonal entry the j -th principal angle, then a measure of distance between the space spanned by \widehat{B}_0 and B_0 is $\sin \Theta(\widehat{B}_0, B_0)$, where \sin is taken entry-wise.

The following variant of Davis-Kahan theorem *Yu et al. (2015)* provides a bound for the distances between eigenspaces.

Theorem II.5 (Variant of Davis-Kahan $\sin \theta$ theorem). *Let $\Sigma, \hat{\Sigma} \in \mathbb{R}^{p \times p}$ be symmetric, with eigenvalues $\lambda_1 \geq \dots \geq \lambda_p$ and $\hat{\lambda}_1 \geq \dots \geq \hat{\lambda}_p$ respectively. Fix $1 \leq k \leq p$ and assume that $\lambda_k - \lambda_{k+1} > 0$, where we define $\lambda_{p+1} = -\infty$. Let $V = (\vec{v}_1, \vec{v}_2, \dots, \vec{v}_k) \in \mathbb{R}^{p \times k}$ and $\widehat{V} = (\widehat{\vec{v}}_1, \widehat{\vec{v}}_2, \dots, \widehat{\vec{v}}_k) \in \mathbb{R}^{p \times k}$ have orthonormal columns satisfying $\Sigma \vec{v}_j = \lambda_j \vec{v}_j$ and $\widehat{\Sigma} \widehat{\vec{v}}_j = \hat{\lambda}_j \widehat{\vec{v}}_j$ for $j = 1, 2, \dots, k$. Then*

$$\|\sin \Theta(\widehat{V}, V)\|_{\text{F}} \leq \frac{2k^{1/2}\|\widehat{\Sigma} - \Sigma\|}{\lambda_k - \lambda_{k+1}}$$

For our purpose, a more meaningful measure of distances between subspaces is the operator norm of the difference of the projections. Let $\mathcal{P}_{\widehat{B}_0}, \mathcal{P}_{B_0}$ be projections onto the column spaces of \widehat{B}_0 and B_0 respectively, we have the following

Corollary II.6. *Under assumptions (2.7) and (2.8) with $\widehat{B}_0, B_0, \widehat{\Sigma}$ and Σ as defined in (2.15), we have*

$$\mathbb{E}\|\mathcal{P}_{\widehat{B}_0} - \mathcal{P}_{B_0}\|^2 \leq \frac{4k\mathbb{E}\|\widehat{\Sigma} - \Sigma\|^2}{\lambda_k^2}. \quad (2.17)$$

Proof of Corollary II.6 . It can be shown that the l_2 operator norm of the difference between two projection matrices $\mathcal{P}_{\widehat{B}_0}$ and \mathcal{P}_{B_0} is determined by the maximum principal angle between the two subspaces (see, e.g. *Meyer (2000)* Chapter 5.15). That is,

$$\|\mathcal{P}_{\widehat{B}_0} - \mathcal{P}_{B_0}\| = \sin(\arccos(\sigma_k)) = \sqrt{1 - \sigma_k^2}.$$

Consequently, we have

$$\|\mathcal{P}_{\widehat{B}_0} - \mathcal{P}_{B_0}\|^2 \leq \left(\sum_{j=1}^k 1 - \sigma_j^2 \right) = \|\sin \Theta(\widehat{B}_0, B_0)\|_F^2 \quad (2.18)$$

Then applying Theorem II.5 and using the fact that $\lambda_{k+1} = 0$ completes the proof. \square

2.4.3 Consistency of EWMA updated marginal variance

In this section we present the proof of Proposition II.1.

Let $\{r_t, t \in \mathbb{Z}\}$ be a zero-mean stationary time series, with some correlation structure

$$\rho_t = \text{Cor}(r_t, r_0) = \text{Cor}(r_{t+h}, r_h), \quad t \in \mathbb{Z}, \quad \forall h \in \mathbb{Z}.$$

We want to handle the following problem: we want to estimate the unknown variance of X_k non-parametrically and show that our estimator is consistent. For the estimation of this variance, we implement an additional EWMA on the squares of the

zero-mean stationary series $\{X_k, k \in \mathbb{Z}\}$. Our proposed estimator is

$$\hat{\sigma}_t^2 = (1 - \lambda_\sigma)\hat{\sigma}_{t-1}^2 + \lambda_\sigma r_t^2. \quad (2.19)$$

To prove the consistency of this estimator, it suffices to show that the following two properties hold

$$\begin{aligned} \mathbb{E}[\hat{\sigma}_t^2] &\xrightarrow[\lambda_\sigma \rightarrow 0]{t \rightarrow \infty} \sigma_r^2 := \text{Var}(r_t) \\ \text{Var}(\hat{\sigma}_t^2) &\xrightarrow[\lambda_\sigma \rightarrow 0]{t \rightarrow \infty} 0. \end{aligned}$$

Starting with the expectation, we have that

$$\begin{aligned} \mathbb{E}[\hat{\sigma}_t^2] &= \sum_{j=0}^t \lambda_\sigma (1 - \lambda_\sigma)^j \mathbb{E}[r_{t-j}^2] = \sigma_r^2 \sum_{j=0}^t \lambda_\sigma (1 - \lambda_\sigma)^j \\ &= \sigma_r^2 \lambda_\sigma \frac{1 - (1 - \lambda_\sigma)^{t+1}}{1 - (1 - \lambda_\sigma)} = \sigma_r^2 (1 - (1 - \lambda_\sigma)^{t+1}) \\ &\xrightarrow{t \rightarrow \infty} \sigma_r^2, \end{aligned}$$

since $0 < \lambda_\sigma < 1$. Here we have used the expression

$$\hat{\sigma}_t^2 = \sum_{j=0}^t \lambda_\sigma (1 - \lambda_\sigma)^j r_{t-j}^2,$$

since we have a finite horizon on this EWMA.

Our second goal is to show that the variance of $\hat{\sigma}_t^2$ vanishes. We start by finding an explicit expression of this variance. We have in general that

$$\begin{aligned} \text{Var}(\hat{\sigma}_t^2) &= \text{Var}\left(\sum_{j=0}^t \lambda_\sigma (1 - \lambda_\sigma)^j r_{t-j}^2\right) \\ &= \sum_{j=0}^t \lambda_\sigma^2 (1 - \lambda_\sigma)^{2j} \text{Var}(r_{t-j}^2) \end{aligned}$$

$$+ \sum_{i=0}^t \sum_{\substack{j=0 \\ i \neq j}}^t \lambda_\sigma^2 (1 - \lambda_\sigma)^{i+j} \text{Cov}(r_{t-j}^2, r_{t-i}^2).$$

We make the following core assumption to continue our calculations.

Assumption II.7. *The process $\{r_t, t \in \mathbb{Z}\}$ is Gaussian.*

Using the fact that $\mathbb{E}(r_t) = 0$, one can immediately obtain that

$$\text{Var}(r_{t-j}^2) = \mathbb{E}[r_{t-j}^4] - \mathbb{E}[r_{t-j}^2]^2 = 3\sigma_r^4 - \sigma_r^4 = 2\sigma_r^4.$$

Now, we need to explore the covariance in the second summand of the above expression. Let

$$\begin{pmatrix} Z_1 \\ Z_2 \end{pmatrix} \sim N\left(0, \begin{pmatrix} 1 & \rho \\ \rho & 1 \end{pmatrix}\right).$$

We know that

$$Z_1 \stackrel{d}{=} \rho Z_2 + \sqrt{1 - \rho^2} Z_3,$$

where Z_2 and Z_3 are independent standard Normal random variables.. Then, we have that

$$\begin{aligned} \text{Cov}(Z_1^2, Z_2^2) &= \mathbb{E}[Z_1^2 Z_2^2] - \mathbb{E}[Z_1^2] \cdot \mathbb{E}[Z_2^2] = \mathbb{E}[Z_1^2 Z_2^2] - 1 \\ &= \mathbb{E}\left[\left(\rho^2 Z_2^2 + (1 - \rho^2) Z_3^2 + 2\rho\sqrt{1 - \rho^2} Z_2 Z_3\right) Z_2^2\right] - 1 \\ &= 3\rho^2 + (1 - \rho^2) - 1 = 2\rho^2. \end{aligned}$$

Using the above one immediately has that

$$\text{Var}(\hat{\sigma}_t^2) = 2\sigma_t^4 \sum_{j=0}^t \lambda_\sigma^2 (1 - \lambda_\sigma)^{2j} + 2\sigma_r^4 \sum_{i=0}^t \sum_{\substack{j=0 \\ i \neq j}}^t \lambda_\sigma^2 (1 - \lambda_\sigma)^{i+j} [\text{Cor}(r_{t-j}, r_{t-i})]^2$$

$$\begin{aligned}
&= 2\sigma_r^4 \lambda_\sigma^2 \frac{1 - [(1 - \lambda_\sigma)^2]^{t+1}}{1 - (1 - \lambda_\sigma)^2} + 2\sigma_r^4 \lambda_\sigma^2 \sum_{\substack{i=0 \\ i \neq j}}^t \sum_{j=0}^t (1 - \lambda_\sigma)^{i+j} \rho_{i-j}^2 \\
&= 2\sigma_r^4 \lambda_\sigma \cdot \frac{1 - [(1 - \lambda_\sigma)^2]^{t+1}}{2 - \lambda_\sigma} + 2\sigma_r^4 \lambda_\sigma^2 \sum_{\substack{i=0 \\ i \neq j}}^t \sum_{j=0}^t (1 - \lambda_\sigma)^{i+j} \rho_{i-j}^2,
\end{aligned}$$

where in the second equality we used the stationarity of $\{r_t, t \in \mathbb{Z}\}$.

We take the limit as $t \rightarrow \infty$ in the above expression, and we have that

$$\lim_{t \rightarrow \infty} \text{Var}(\hat{\sigma}_t^2) = \frac{2\sigma_r^4 \lambda_\sigma}{2 - \lambda_\sigma} + 2\lambda_\sigma^2 \sigma_r^4 \sum_{\substack{i=0 \\ i \neq j}}^{\infty} \sum_{j=0}^{\infty} (1 - \lambda_\sigma)^{i+j} \rho_{i-j}^2.$$

We need a condition on $\{\rho_t, t \in \mathbb{Z}\}$ in order to proceed.

We propose the following condition in order to secure the consistency of the variance.

Condition II.8. Assume that the sequence $\{\rho_t, t \in \mathbb{Z}\}$ is square summable, namely that

$$\sum_{t=-\infty}^{\infty} \rho_t^2 < \infty.$$

Under this condition, we have the following proposition.

Proposition II.9. *Let $\{r_t, t \in \mathbb{Z}\}$ and $\{\hat{\sigma}_t^2, t \in \mathbb{Z}\}$ be defined as in the start of Section 2.4.3 and (2.19) respectively. If Condition II.8 holds, then the estimator $\hat{\sigma}_t^2$ is consistent.*

Proof. The proof is based on the discussion preceding Proposition II.9 and the following. Let

$$A_t := 2\lambda_\sigma^2 \sigma_t^4 \sum_{\substack{i=0 \\ i \neq j}}^t \sum_{j=0}^t (1 - \lambda_\sigma)^{i+j} \rho_{i-j}^2.$$

Then

$$\begin{aligned}
A_t &\leq 2\lambda_\sigma^2 \sigma_r^4 \sum_{i=0}^t \sum_{j=0}^t (1 - \lambda_\sigma)^{i+j} \rho_{i-j}^2 \\
&= 2\lambda_\sigma^2 \sigma_r^4 \sum_{v=0}^{2t} (1 - \lambda_\sigma)^v \sum_{n=\max\{-t, v-2t\}}^{\min\{t, v\}} \rho_n^2 \\
&\leq 2\lambda_\sigma^2 \sigma_r^4 \sum_{v=0}^{2t} (1 - \lambda_\sigma)^v \sum_{n=-\infty}^{\infty} \rho_n^2 \\
&\leq 2\lambda_\sigma^2 \sigma_r^4 \sum_{v=0}^{\infty} (1 - \lambda_\sigma)^v \sum_{n=-\infty}^{\infty} \rho_n^2 = 2\lambda_\sigma \sigma_r^4 \sum_{n=-\infty}^{\infty} \rho_n^2,
\end{aligned}$$

where we have used the change of variables $v = i + j$, $n = i - j$ in the second equality.

Using the relationship

$$\begin{aligned}
\text{Var}(\hat{\sigma}_t^2) &= 2\sigma_r^4 \lambda_\sigma \cdot \frac{1 - [(1 - \lambda_\sigma)^2]^{t+1}}{2 - \lambda_\sigma} \\
&\quad + 2\lambda_\sigma^2 \sum_{i=0}^t \sum_{\substack{j=0 \\ i \neq j}}^t (1 - \lambda_\sigma)^{i+j} \rho_{i-j}^2,
\end{aligned}$$

we have that

$$\text{Var}(\hat{\sigma}_t^2) \leq \frac{2\sigma_r^4 \lambda_\sigma}{2 - \lambda_\sigma} + 2\lambda_\sigma \sigma_r^4 \sum_{n=-\infty}^{\infty} \rho_n^2.$$

Because of Condition II.8, letting $\lambda_\sigma \rightarrow 0$, we obtain the desired property, namely that $\text{Var}(\hat{\sigma}_k^2)$ vanishes. \square

CHAPTER III

Concentration Rates

Consider the maxima of a sequence of random variables and assume that there exists a sequence of constants, so that the maxima divided by these constants converge in probability to one. This property, known as relative stability of maxima, has been thoroughly studied in the case of independence. However, little is known for this otherwise referred to as concentration of maxima phenomenon, when the maxima are taken over dependent random variables. Only recently, *Gao and Stoev* (2020) established that this phenomenon is the key to solving the exact sparse support recovery problem in high dimensions. In this chapter, we obtain bounds on the rate of concentration of maxima in Gaussian triangular arrays in the context of dependence over space. These results are used to establish sufficient conditions for the uniform relative stability of functions of Gaussian arrays, leading to a variety of new models that exhibit phase transitions in the exact support recovery problem. Recall that the latter problem is closely related to the identification problem studied in Chapter II, as the canonical signal-plus-noise model, that is assumed for the residuals therein, is also utilized here. Finally, the optimal rate of concentration for Gaussian arrays is studied under general assumptions implied by the classic condition of *Berman* (1964). This chapter is based on an already published paper in the *Extremes* (*Kartsioukas et al.*, 2021).

3.1 Relative stability and dependence

Let Z_i , $i = 1, 2, \dots$ be independent and identically distributed (iid) standard Normal random variables. It is well known that their maxima under affine normalization converge to the Gumbel extreme value distribution. If, however, one chooses to standardize the maxima by only dividing by a sequence of positive numbers, then the only possible limits are constants. Specifically, for all $a_p \sim \sqrt{2 \log(p)}$, we have

$$\frac{1}{a_p} \max_{i \in [p]} Z_i \xrightarrow{\mathbb{P}} 1, \quad \text{as } p \rightarrow \infty, \quad (3.1)$$

where $[p] := \{1, \dots, p\}$ and in fact the convergence is valid almost surely. This property, known as *relative stability*, dates back to the seminal work of *Gnedenko* (1943) who has characterized it in terms of rapid variation of the law of the Z_i 's (see Section 3.2.2 below, as well as *Barndorff-Nielsen* (1963); *Resnick and Tomkins* (1973); *Kimoshita and Resnick* (1991)).

In contrast, if the Z_i 's are iid and heavy-tailed, i.e., $\mathbb{P}[Z_i > x] \propto x^{-\alpha}$, for some $\alpha > 0$, with $a_p \propto p^{1/\alpha}$, we have

$$\frac{1}{a_p} \max_{i \in [p]} Z_i \xrightarrow{d} \xi, \quad (3.2)$$

where ξ is a random variable with the α -Fréchet distribution.

Comparing (3.1) and (3.2), we see that the maxima have fundamentally different asymptotic behavior relative to rescaling with constant sequences. In the light-tailed regime, they *concentrate* around a constant in the sense of (3.1), whereas in the heavy-tailed regime they *disperse* according to a probability distribution viz (3.2).

Although this concentration of maxima phenomenon may be well-known under independence, we found that it is virtually unexplored under *dependence*. In this chapter, we will focus on Gaussian sequences, and in fact, more generally, Gaussian triangular arrays $\mathcal{E} = \{\epsilon_p(i), i \in [p], p \in \mathbb{N}\}$, where the $\epsilon_p(i)$'s are marginally standard Normal but possibly dependent. Let u_p be the $(1 - 1/p)$ -th quantile of the standard

Normal distribution, i.e., $p\bar{\Phi}(u_p) := p(1 - \Phi(u_p)) = 1$, where Φ is the corresponding cumulative distribution function. We say that the array \mathcal{E} is uniformly relatively stable (URS), if

$$\frac{1}{u_{|S_p|}} \max_{i \in S_p} \epsilon_p(i) \xrightarrow{\mathbb{P}} 1, \quad \text{as } |S_p| \rightarrow \infty, \quad (3.3)$$

for every choice of growing subsets $S_p \subset \{1, \dots, p\}$. Note that $u_p \sim \sqrt{2 \log(p)}$ (see *e.g.* Lemma III.6). Certainly, the relative stability property shows that all iid Gaussian arrays are trivially URS. The notion of uniform relative stability, however, is far from automatic or trivial under dependence. In the recent work of *Gao and Stoev (2020)*, it was found that URS is the key to establishing the fundamental limits in sparse-signal support estimation in high-dimensions. Specifically, under URS, a phase-transition phenomenon was shown to take place in the support recovery problem. For more details, see Section 3.2.1 below.

Theorem 3.1 in *Gao and Stoev (2020)* gives a surprisingly simple necessary and sufficient condition for a Gaussian array \mathcal{E} to be URS. As an illustration, in the special case where $\epsilon_p(i) \equiv Z_i$, $i \in \mathbb{N}$ form a stationary Gaussian time series, the array \mathcal{E} is URS if and only if the auto-covariance vanishes, i.e.,

$$\text{Cov}(Z_k, Z_0) \longrightarrow 0, \quad \text{as } k \rightarrow \infty. \quad (3.4)$$

That is, (3.1) holds (with $a_p \sim \sqrt{2 \log(p)}$), for *any* stationary Gaussian time series $Z = \{Z_i\}$ with vanishing auto-covariance, no matter the rate of decay. The “if” part of (3.4) appeared in Theorem 4.1 in *Berman (1964)*.

Condition (3.4) should be contrasted with the classic Berman condition,

$$\text{Cov}(Z_k, Z_0) = o\left(\frac{1}{\log(k)}\right), \quad \text{as } k \rightarrow \infty,$$

which entails *distributional* convergence under affine normalization. Here, our focus

is not on distributional limits but on merely the concentration of maxima under rescaling, which can take place under much more severe dependence. In fact, unlike Berman, here we are not limited to the time-series setting. For a complete statement of the characterization of URS, see Section 3.2.2, below.

While *Gao and Stoev* (2020) characterized the conditions under which the convergence (3.3) takes place, the rate of this convergence remained an open question. In this chapter, our goal is to establish bounds on the *rate of concentration* for maxima of Gaussian arrays. Specifically, we establish results of the type

$$\mathbb{P} \left[\left| \frac{1}{u_p} \max_{i \in [p]} \epsilon_p(i) - 1 \right| > \delta_p \right] \longrightarrow 0, \quad (3.5)$$

where $\delta_p \rightarrow 0$ decays at a certain rate. The rate of the sequence δ_p is quantified explicitly in terms of the covariance structure of the array. More precisely, the packing numbers $N(\tau)$ associated with the UDD condition introduced in *Gao and Stoev* (2020) will play a key role. These packing numbers arise from a Sudakov-Fernique type construction, which appear to be close to optimal, although at this point we do not know if the so obtained bounds on the rates can be improved (cf Conjecture III.11, below). After concluding the paper (*Kartsioukas et al.*, 2021), we became aware of the important results of *Tanguy* (2015), which are closely related to ours in the special case of stationary time series. Our approach, however, is technically different and yields explicit rates for the general case of Gaussian triangular arrays. For more details, see Remark III.33, below.

Our general results are illustrated with several models, where explicit bounds on the rates of concentration are derived. In Section 3.3, we study the *optimal rate* of concentration and show that under rather broad dependence conditions (including the iid setting), (3.5) holds if and only if $\delta_p \gg 1/\log(p)$. Somewhat curiously, the constant u_p matters and the popular choice of $u_p := \sqrt{2 \log(p)}$ leads to the slower

rates of $\log(\log(p))/\log(p)$.

Our bounds on the rate of concentration find important application in the study of uniform relative stability for functions of Gaussian arrays. Specifically, let $\eta_p(i) = f(\epsilon_p(i))$, where $\mathcal{E} = \{\epsilon_p(i), i \in [p], p \in \mathbb{N}\}$ is a Gaussian triangular array and f is a given deterministic function. In Section 3.4.2, using our results on the rate of concentration for the array \mathcal{E} , we establish conditions which imply the uniform relative stability of the array $\mathcal{H} = \{\eta_p(i), i \in [p], p \in \mathbb{N}\}$. Consequently, we establish that many dependent log-normal and χ^2 -arrays are URS, and hence obey the phase-transition result of *Gao and Stoev (2020)*.

This chapter is structured as follows. In Section 3.2, we review the statistical inference problem motivating the study of the concentration of maxima phenomenon. Recalled is the notion of uniform decreasing dependence involved in the characterization of uniform relative stability for Gaussian arrays. A brief discussion on the optimal rate of concentration is given in Section 3.3. Section 3.4 contains the statement of the main result as well as some examples and applications. Section 3.5 contains proofs and technical results, which may be of independent interest.

3.2 Concentration of maxima and high-dimensional inference

In this section, we start with the statistical inference problem that motivated us to study the concentration of maxima phenomenon. Readers who are convinced that this is a phenomenon of independent interest can skip to Section 3.2.2, where concrete definitions and notions are reviewed.

3.2.1 Fundamental limits of support recovery in high dimensions

Our main motivation to study the relative stability or *concentration of maxima* under dependence is the fundamental role it plays in recent developments on high-dimensional statistical inference, which we briefly review next. Consider the classic

signal plus noise model

$$x_p(i) = \mu_p(i) + \epsilon_p(i), \quad i \in [p], \quad (3.6)$$

where $\mu_p = (\mu_p(i)) \in \mathbb{R}^p$ is an unknown high-dimensional “signal” observed with additive noise. The noise is modeled with a triangular array $\mathcal{E} = \{\epsilon_p(i), i \in [p], p \in \mathbb{N}\}$, where for concreteness, all $\epsilon_p(i)$ ’s are standardized to have the same marginal distribution F . However, this noise can have arbitrary dependence structure, in principle.

One popular and important high-dimensional inference context, is the one where the dimension p grows to infinity and the signal is sparse. Namely, the signal support set $S_p := \{i \in [p] : \mu_p(i) \neq 0\}$ is of smaller order than its dimension:

$$|S_p| \sim p^{1-\beta}, \text{ for some } \beta \in (0, 1).$$

The parameter β controls the degree of sparsity; if β is larger, the signal is more sparse, i.e., has fewer non-zero components. In this context, many natural questions arise such as the detection of the presence of non-zero signal or the estimation of its support set (see, e.g., *Ingster* (1998a); *Donoho and Jin* (2004a); *Ji and Jin* (2012); *Arias-Castro and Chen* (2017)). Here, as in *Gao and Stoev* (2020), we focus on the fundamental *support recovery* problem. Particularly, under what conditions on the signal magnitude we can have *exact support recovery* in the sense that

$$\mathbb{P}[\widehat{S}_p = S_p] \longrightarrow 1, \quad \text{as } p \rightarrow \infty.$$

Gao and Stoev (2020) showed that a natural solution to this problem can be obtained using the concentration of maxima phenomenon. Specifically, consider the class of all

thresholding support estimators:

$$\widehat{S}_p := \{j \in [p] : x_p(j) > t_p(x)\}, \quad (3.7)$$

where $t_p(x)$ is possibly data-dependent threshold. For simplicity of exposition, suppose also that the signal magnitude is parametrized as follows

$$\mu_p(i) = \sqrt{2r \log(p)}, \quad i \in S_p,$$

where $r > 0$. Consider also the function

$$g(\beta) := (1 + \sqrt{1 - \beta})^2.$$

Theorems 2.1 and 2.2 of *Gao and Stoev* (2018) entail that if \mathcal{E} is URS (see Definition III.2 below), then we have the phase-transition:

$$\mathbb{P}[\widehat{S}_p = S_p] \longrightarrow \begin{cases} 1, & \text{if } r > g(\beta) \text{ for suitable } \widehat{S}_p \text{ as in (3.7)} \\ 0, & \text{if } r < g(\beta) \text{ for all } \widehat{S}_p \text{ as in (3.7)} \end{cases}, \quad \text{as } p \rightarrow \infty.$$

That is, for signal magnitudes above the boundary, thresholding (Bonferonni-type) estimators recover the support perfectly, as $p \rightarrow \infty$; whereas for signals below the boundary, no thresholding estimators can recover the support with positive probability. Further, as shown in *Gao and Stoev* (2020), thresholding estimators are optimal in the iid Gaussian setting and hence the above phase-transition applies to *all possible* support estimators leading to minimax-type results. Interestingly, both Gaussian and non-Gaussian noise arrays are addressed equally well, provided that they satisfy the uniform relative stability property. While URS is a very mild condition, except for the Gaussian case addressed in *Gao and Stoev* (2020), little is known in general. Here, we will fill this gap for a class of functions of Gaussian arrays (see

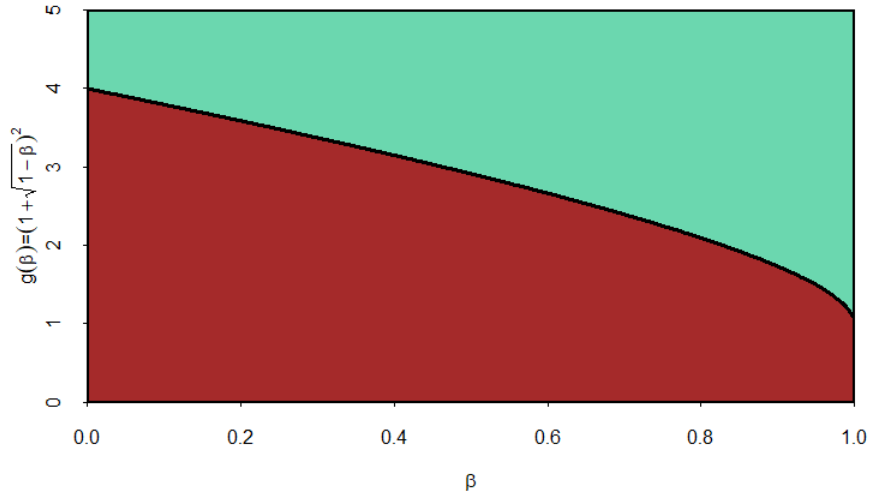


Figure 3.1: Phase transition boundary on the exact support recovery problem. If $r > g(\beta)$, and so in the green area, then we can find a thresholding estimator, so that exact support recovery is attainable, whether \mathcal{E} is URS or not. If $r < g(\beta)$, in the brown area, and \mathcal{E} is URS, then no matter the thresholding estimator, the probability of exactly estimating S_p vanishes as p grows to infinity.

Section 3.4.2), using our new results on the rates of concentration.

3.2.2 Concentration of maxima

In this section, we recall some definitions and a characterization of URS in *Gao and Stoev (2020)*. We start by presenting the notion of *relative stability*.

Definition III.1. (Relative stability). Let $\epsilon_p = (\epsilon_p(j))_{j=1}^p$ be a sequence of random variables with identical marginal distributions F . Define the sequence $(u_p)_{p=1}^\infty$ to be the $(1 - 1/p)$ -th quantile of F , i.e.,

$$u_p = F^{\leftarrow}(1 - 1/p). \quad (3.8)$$

The triangular array $\mathcal{E} = \{\epsilon_p, p \in \mathbb{N}\}$ is said to have relatively stable (RS) maxima if

$$\frac{1}{u_p} M_p := \frac{1}{u_p} \max_{i=1, \dots, p} \epsilon_p(i) \xrightarrow{\mathbb{P}} 1, \quad (3.9)$$

as $p \rightarrow \infty$.

Note that by Proposition 1.1 of *Gao and Stoev (2020)*, we have for the standard Normal distribution, that

$$u_p = \Phi^{-1}(1 - 1/p) \sim \sqrt{2 \log(p)}. \quad (3.10)$$

While relative stability is not directly used in this chapter, it is a natural prerequisite to introducing the following generalization.

Definition III.2. (Uniform Relative Stability (URS)). Under the notations established in Definition III.1, the triangular array $\mathcal{E} = \{\epsilon_p(i), i \in [p]\}$ is said to have uniform relatively stable (URS) maxima if for *every* sequence of subsets $S_p \subseteq \{1, \dots, p\}$ such that $|S_p| \rightarrow \infty$, we have

$$\frac{1}{u_{|S_p|}} M_{S_p} := \frac{1}{u_{|S_p|}} \max_{i \in S_p} \epsilon_p(i) \xrightarrow{\mathbb{P}} 1, \quad \text{as } p \rightarrow \infty. \quad (3.11)$$

Definition III.3. (Uniformly Decreasing Dependence (UDD)). A Gaussian triangular array \mathcal{E} with standard normal marginals is said to be uniformly decreasingly dependent (UDD) if for every $\tau > 0$ there exists a finite $N_{\mathcal{E}}(\tau) < \infty$, such that for every $i \in \{1, \dots, p\}$, and $p \in \mathbb{N}$, we have

$$\left| \{k \in \{1, \dots, p\} : \text{Cov}(\epsilon_p(k), \epsilon_p(i)) > \tau\} \right| \leq N_{\mathcal{E}}(\tau), \quad \text{for all } \tau > 0. \quad (3.12)$$

That is, for any coordinate j , the number of coordinates which are more than τ -correlated with $\epsilon_p(j)$ does not exceed $N_{\mathcal{E}}(\tau)$.

The next result provides the equivalence between uniform relative stability and uniformly decreasing dependence.

Theorem III.4 (Theorem 3.1 in *Gao and Stoev (2020)*). *Let \mathcal{E} be a Gaussian triangular array with standard Normal marginals. The array \mathcal{E} is URS if and only if it is UDD.*

Theorem III.4 is the starting point of the rate investigations in this dissertation. Our main result, Theorem III.13, below, extends the former by providing upper bounds on the rate of concentration. Before that, though, in Section 3.3 we study cases where the optimal rate can be formally established.

Remark III.5 (On the use of the term “upper bound”). Fix a positive sequence $\delta_p^* \downarrow 0$. We refer to δ_p^* as an upper bound on the rate of concentration when (3.5) holds for any sequence $\delta_p \gg \delta_p^*$. Further, for two positive sequences α_p and β_p we write $\alpha_p \asymp \beta_p$ if

$$0 < c_1 \leq \liminf_{p \rightarrow \infty} \left| \frac{\alpha_p}{\beta_p} \right| \leq \limsup_{p \rightarrow \infty} \left| \frac{\alpha_p}{\beta_p} \right| \leq c_2 < \infty.$$

Let δ_p^* be an upper bound on the rate of concentration and $\delta_p \gg \delta_p^*$. Then, naturally, (3.5) holds with δ_p replaced by $\tilde{\delta}_p$, for any $\tilde{\delta}_p \asymp \delta_p$.

3.3 On the optimal rate of concentration

In this section, we provide some general comments on the fastest possible rates of concentration for maxima of Gaussian variables. Somewhat surprisingly, the rate depends on the choice of the normalizing sequence u_p . As it turns out poor choices of normalizing sequences can lead to arbitrarily slow rates. On the other hand, for a wide range of dependence structures (including the iid case), the best possible rate will be shown to be $1/\log(p)$. The question of whether the maxima of dependent Gaussian arrays can concentrate faster than that rate, however unlikely this may be, is open, to the best of our knowledge (cf Conjecture III.11, below).

Consider a Gaussian array $\mathcal{E} = \{\epsilon_p(i), i \in [p]\}$ with standard Normal marginal.

We shall assume that \mathcal{E} is (uniformly) relatively stable, so that in particular,

$$\frac{1}{u_p} \max_{i \in [p]} \epsilon_p(i) =: \frac{M_p}{u_p} \xrightarrow{P} 1,$$

as $p \rightarrow \infty$, where $u_p := \Phi^{-1}(1 - 1/p)$ is the $(1/p)$ -th tail quantile of the standard Normal distribution.

We consider the iid case first and, for clarity, let M_p^* denote the maximum of p independent standard Normal random variables. Suppose that for some $a_p > 0$ and $a_p, b_p \in \mathbb{R}$, we have

$$\Phi(a_p^{-1}x + b_p)^p \rightarrow \Lambda(x) := \exp\{-e^{-x}\}, \quad \text{as } p \rightarrow \infty,$$

for all $x \in \mathbb{R}$. That is, we have

$$a_p(M_p^* - b_p) \xrightarrow{d} \zeta, \quad \text{as } p \rightarrow \infty, \quad (3.13)$$

where ζ has the standard Gumbel distribution Λ . The next result is well-known. We give it here since it summarizes and clarifies the possible choices of the normalizing constants a_p and b_p for (3.13) to hold.

Lemma III.6. (i) *We have that*

$$\tilde{u}_p(M_p^* - \tilde{u}_p) \xrightarrow{d} \zeta \quad \text{if and only if} \quad p\bar{\Phi}(\tilde{u}_p) \rightarrow 1, \quad (3.14)$$

as $p \rightarrow \infty$. In this case, $\tilde{u}_p \sim \sqrt{2 \log(p)}$ and more precisely

$$\sqrt{2 \log(p)}(\tilde{u}_p - u_p^*) \rightarrow 0, \quad \text{as } p \rightarrow \infty, \quad (3.15)$$

where

$$u_p^* := \sqrt{2 \log(p)} \left(1 - \frac{\log(\log(p)) + \log(4\pi)}{4 \log(p)} \right). \quad (3.16)$$

(ii) Relation (3.13) holds if and only if

$$a_p \sim \sqrt{2 \log(p)} \quad \text{and} \quad p\bar{\Phi}(b_p) \rightarrow 1.$$

In particular, by part (i), we have that (3.13) holds with $a_p := b_p$ and (3.15) holds with $\tilde{u}_p := b_p$.

Proof. Part (i). Observe that by the Mill's ratio (cf Lemma III.28), $p\bar{\Phi}(\tilde{u}_p) \rightarrow 1$ is equivalently expressed as follows:

$$p\bar{\Phi}(\tilde{u}_p) \sim p \frac{\phi(\tilde{u}_p)}{\tilde{u}_p} \rightarrow 1, \quad \text{as } p \rightarrow \infty,$$

where $\phi(x) = \exp\{-x^2/2\}/\sqrt{2\pi}$ is the standard Normal density. By taking logarithms, the above asymptotic relation is equivalent to having

$$\log(p) - \frac{\tilde{u}_p^2}{2} - \log(\tilde{u}_p) - \frac{1}{2} \log(2\pi) \rightarrow 0. \quad (3.17)$$

We first prove the “if” direction of part (i). Suppose that $p\bar{\Phi}(\tilde{u}_p) \rightarrow 1$, or equivalently, (3.17) holds. Then, one necessarily has $\tilde{u}_p \rightarrow \infty$. It is easy to see that (3.13) holds with $a_p := \tilde{u}_p$ and $b_p := \tilde{u}_p$, provided that, for all $x \in \mathbb{R}$,

$$\Phi \left(\tilde{u}_p + \frac{x}{\tilde{u}_p} \right)^p \rightarrow \Lambda(x), \quad \text{as } p \rightarrow \infty. \quad (3.18)$$

The latter, upon taking logarithms and using the fact that $\log(1+z) \simeq z$, as $z \rightarrow 0$, is equivalent to having

$$p\bar{\Phi} \left(\tilde{u}_p + \frac{x}{\tilde{u}_p} \right) \rightarrow -\log(\Lambda(x)) = e^{-x}. \quad (3.19)$$

To prove that (3.19) holds, as argued above, using the Mill's ratio, it is equivalent to verify that

$$A_p := \log(p) - \frac{1}{2}(\tilde{u}_p + x/\tilde{u}_p)^2 - \log(\tilde{u}_p + x/\tilde{u}_p) - \frac{1}{2}\log(2\pi) \rightarrow -x,$$

as $p \rightarrow \infty$. Note that, upon expanding the square and manipulating the logarithm, we obtain

$$A_p = \log(p) - \frac{\tilde{u}_p^2}{2} - \log(\tilde{u}_p) - \frac{1}{2}\log(2\pi) - x - x^2/(2\tilde{u}_p^2) - \log(1 + x/\tilde{u}_p^2).$$

In view of (3.17) and the fact that $\tilde{u}_p \rightarrow \infty$, we obtain that $A_p \rightarrow -x$, which yields (3.19) and completes the proof of the “if” direction of part (i).

Now, to show the “only if” direction of part (i), suppose that (3.13) holds with $a_p = b_p := \tilde{u}_p$, or, equivalently (3.18) holds. By letting $x = 0$ in Relation (3.18), we see that $\tilde{u}_p \rightarrow \infty$, and then, upon taking logarithms, necessarily $p\bar{\Phi}(\tilde{u}_p) \rightarrow 1$, which completes the proof of (3.14).

We now show (3.15). First, one can directly verify that (3.17) holds with \tilde{u}_p replaced by u_p^* in (3.16). This, as argued above, is equivalent to $p\bar{\Phi}(u_p^*) \rightarrow 1$. Suppose now that, for another sequence \tilde{u}_p , we have $p\bar{\Phi}(\tilde{u}_p) \rightarrow 1$. Then, by the shown equivalence in (3.14),

$$u_p^*(M_n^* - u_p^*) \xrightarrow{d} \zeta \quad \text{and} \quad \tilde{u}_p(M_n^* - \tilde{u}_p) \xrightarrow{d} \zeta.$$

Thus, the convergence of types theorem (see, *e.g.*, Theorem 14.2 in *Billingsley* (1995)) yields

$$u_p^* \sim \tilde{u}_p \quad \text{and} \quad u_p^*(u_p^* - \tilde{u}_p) \rightarrow 0.$$

The last convergence implies the claim of part (ii) since in view of (3.16), we have $u_p^* \sim \sqrt{2\log(p)}$.

Part (ii) is a direct consequence of the convergence to types theorem, as argued in the proof of part (i). \square

The following result characterizes the optimal rate of concentration under an additional distributional convergence assumption, which holds under the Berman condition for e.g. the case of stationary time series.

Proposition III.7. *Suppose that \mathcal{E} is a dependent triangular Gaussian array, such that*

$$\zeta_p := a_p(M_p - b_p) \xrightarrow{d} \zeta, \quad \text{as } p \rightarrow \infty, \quad (3.20)$$

for some non-degenerate random variable ζ , with the same constants as in the iid case (3.13). Suppose also that $\mathbb{P}(\zeta < x) > 0$ and $\mathbb{P}(\zeta > x) > 0$ for all $x \in \mathbb{R}$.

Let now the sequence $\delta_p \rightarrow 0$, be an upper bound on the rate of concentration, i.e., we have

$$\mathbb{P}\left(\left|\frac{M_p}{a_p} - 1\right| > \delta_p\right) \rightarrow 0, \quad p \rightarrow \infty. \quad (3.21)$$

The following two statements hold.

(a) When $\limsup_{p \rightarrow \infty} a_p |b_p - a_p| < \infty$, Relation (3.21) holds if and only if

$$\delta_p \gg \frac{1}{a_p^2} + \left| \frac{b_p}{a_p} - 1 \right| =: \delta_p^{opt}. \quad (3.22)$$

(b) When $\limsup_{p \rightarrow \infty} a_p |b_p - a_p| = \infty$, Relation (3.21) holds if and only if

$$\liminf_{p \rightarrow \infty} \left[\frac{\delta_p}{\delta_p^{opt}} - 1 \right] (1 + a_p |b_p - a_p|) = \infty. \quad (3.23)$$

Proof. (a) We will start with the “if” direction. Relation (3.20) implies that

$$\frac{1}{a_p} M_p = \frac{\zeta_p}{a_p^2} + \frac{b_p}{a_p}.$$

Since by assumption the constants a_p and b_p are the same as in the iid case (3.13), Lemma III.6 entails that $b_p \sim a_p \sim \sqrt{2 \log(p)}$. Hence

$$\frac{1}{a_p} M_p - 1 = \frac{\zeta_p}{a_p^2} + \left(\frac{b_p}{a_p} - 1 \right) \rightarrow 0, \quad (3.24)$$

which shows that the distributional limit in (3.20) entails concentration of the maxima M_p/a_p to 1. Relations (3.22) and (3.24), however imply that

$$\left| \frac{M_p}{a_p} - 1 \right| = o_P(\delta_p),$$

which entails (3.21) by Slutsky (or also Lemma III.9, below.)

Now, for the converse direction, suppose that (3.21) holds for some $\delta_p \not\asymp \delta_p^{opt}$. This means that we can find a subsequence $p(n)$ so that $\delta_{p(n)} \leq c \cdot \delta_{p(n)}^{opt}$, $\forall n \in \mathbb{N}$, for a positive constant c that does not depend on n . In view of (3.21), this would mean that

$$\theta_n := \mathbb{P} \left(\left| \frac{M_{p(n)}}{a_{p(n)}} - 1 \right| > c \delta_{p(n)}^{opt} \right) \rightarrow 0, \quad n \rightarrow \infty.$$

Moreover, since $\limsup_{p \rightarrow \infty} a_p |b_p - a_p| < \infty$, and $a_p > 0$, the sequence $(a_p |b_p - a_p|)_{p=1}^{\infty}$ is bounded. Namely, there exists $M > 0$, such that $0 \leq a_p |b_p - a_p| \leq M$, for all $p \in \mathbb{N}$.

However, we have that

$$\begin{aligned} \theta_n &\geq \mathbb{P} \left(\frac{M_{p(n)}}{a_{p(n)}} - 1 > c \delta_{p(n)}^{opt} \right) = \mathbb{P} \left(\frac{\zeta_{p(n)}}{a_{p(n)}^2} + \frac{b_{p(n)}}{a_{p(n)}} - 1 > \frac{c}{a_{p(n)}^2} + c \left| \frac{b_{p(n)}}{a_{p(n)}} - 1 \right| \right) \\ &= \mathbb{P} \left(\zeta_{p(n)} + a_{p(n)}(b_{p(n)} - a_{p(n)}) - c |a_{p(n)}(b_{p(n)} - a_{p(n)})| > c \right) \\ &\geq \mathbb{P} \left(\zeta_{p(n)} - (c+1)a_{p(n)} |b_{p(n)} - a_{p(n)}| > c \right) \\ &\geq \mathbb{P} \left(\zeta_{p(n)} > c + (c+1)a_{p(n)} |b_{p(n)} - a_{p(n)}| \right) \\ &\geq \mathbb{P}(\zeta_{p(n)} > c + (c+1)M) \\ &\rightarrow \mathbb{P}(\zeta > c + (c+1)M) > 0, \end{aligned}$$

where the last convergence holds because $\zeta_{p(n)} \xrightarrow{d} \zeta$. This is a contradiction and the proof is complete.

(b) We have that

$$\begin{aligned} \mathbb{P}\left(\left|\frac{M_p}{a_p} - 1\right| > \delta_p\right) &= \mathbb{P}(a_p |M_p - a_p| > \delta_p a_p^2) = \mathbb{P}(|\zeta_p + a_p(b_p - a_p)| > \delta_p a_p^2) \\ &= \mathbb{P}(\zeta_p < -\delta_p a_p^2 - a_p(b_p - a_p)) + \mathbb{P}(\zeta_p > \delta_p a_p^2 - a_p(b_p - a_p)) \\ &=: A(p) + B(p). \end{aligned}$$

Note, however, that (3.21) entails that both $A(p)$ and $B(p)$ vanish to 0, as $p \rightarrow \infty$.

This in turn means that

$$\liminf_{p \rightarrow \infty} (\delta_p a_p^2 - a_p(b_p - a_p)) = \infty \quad \text{and} \quad \liminf_{p \rightarrow \infty} (\delta_p a_p^2 + a_p(b_p - a_p)) = \infty, \quad (3.25)$$

because of the distributional convergence (3.20). We will work with $B(p)$. The result for $A(p)$ can be obtained by similar arguments. At first, for $B(p)$ to vanish to 0, we do need $\delta_p a_p^2 > a_p(b_p - a_p)$ eventually. Suppose that $\liminf_{p \rightarrow \infty} (\delta_p a_p^2 - a_p(b_p - a_p)) = c < \infty$, where $c \geq 0$. This would mean that there is a subsequence $p(n)$ such that

$$\delta_{p(n)} a_{p(n)}^2 - a_{p(n)}(b_{p(n)} - a_{p(n)}) \rightarrow c, \quad p \rightarrow \infty.$$

But then,

$$B(p(n)) = \mathbb{P}(\zeta_{p(n)} > \delta_{p(n)} a_{p(n)}^2 - a_{p(n)}(b_{p(n)} - a_{p(n)})) \rightarrow \mathbb{P}(\zeta > c) > 0,$$

which contradicts the fact that $B(p) \rightarrow 0$, as $p \rightarrow \infty$.

Finally, note that (3.25) is equivalent to $\liminf_{p \rightarrow \infty} (\delta_p a_p^2 - a_p |b_p - a_p|) = \infty$, which

with straightforward algebra can be expressed as (3.23). Indeed,

$$\begin{aligned} \delta_p a_p^2 - a_p |b_p - a_p| &= a_p^2 \left[\delta_p - \left| \frac{b_p}{a_p} - 1 \right| \right] = a_p^2 [\delta_p - \delta_p^{opt}] + 1 \\ &= a_p^2 \delta_p^{opt} \left[\frac{\delta_p}{\delta_p^{opt}} - 1 \right] + 1 \\ &= \left[\frac{\delta_p}{\delta_p^{opt}} - 1 \right] (1 + a_p |b_p - a_p|) + 1, \end{aligned}$$

which completes the proof. \square

Remark III.8 (On the optimality of the rate δ_p^{opt}). The rate δ_p^{opt} can be viewed as “the” optimal rate of concentration in (3.21) in the sense of (3.22) and (3.23). The distributional convergence in (3.20) (whenever it takes place) is much more informative than a simple concentration of maxima type convergence. Specifically, by Lemma III.6 (ii), one can take $u_p = a_p = b_p$, and in this case Relation (3.24) implies that $1/a_p^2 \asymp 1/\log(p)$ is both an upper and lower bound on the rate of concentration. That is, the rate $\delta_p^{opt} = 1/a_p^2 \asymp 1/\log(p)$ cannot be improved and in this sense is the optimal rate at which the maxima can concentrate. The rate of concentration, though, does depend on the choice of the normalization sequence u_p . We elaborate on this point next.

The role of the sequence u_p . It is well-known that under quite substantial dependence, the convergence in distribution (3.20) holds, with the *same constants* as in the independent case. For example, suppose that $\epsilon_p(i) = Z(i)$, $i \in \mathbb{Z}$ come from a stationary Gaussian time series, which satisfies the so-called *Berman condition* (*Berman, 1964*):

$$\text{Cov}(Z(k), Z(0)) = o\left(\frac{1}{\log(k)}\right), \quad \text{as } k \rightarrow \infty.$$

Notice, by Lemma III.6 (ii), however, we also have $\tilde{\zeta}_p := b_p(M_p - b_p) \xrightarrow{d} \zeta$, and

$$\frac{1}{b_p} M_p - 1 = \frac{\tilde{\zeta}_p}{b_p^2} = \mathcal{O}_P \left(\frac{1}{\log(p)} \right). \quad (3.26)$$

Compare Relations (3.24) and (3.26). Since $a_p \sim b_p \sim \sqrt{2\log(p)}$, from (3.26), we have that the rate of concentration of M_p relative to the sequence b_p is $1/\log(p)$. On the other hand, while the first term in the right-hand side of (3.24) is of order $1/\log(p)$ the presence of the second term can only make the rate of concentration therein *slower*. Indeed, this is formally established in Lemma III.9. To gain some more intuition that the poor choice of a sequence a_p can lead to a slower rate of concentration, suppose that $a_p = b_p/(1 + g(p))$, for an arbitrary sequence $g(p) > -1$, such that $g(p) \rightarrow 0$. Then, by (3.24),

$$\frac{1}{a_p} M_p - 1 = \frac{\zeta_p}{a_p^2} + g(p).$$

One can take $g(p) \rightarrow 0$ arbitrarily slow. Finally, as a more concrete example, one typically uses $b_p := u_p^* = \sqrt{2\log(p)}(1 - (\log(\log(p)) + \log(4\pi))/4\log(p))$ and $a_p := \sqrt{2\log(p)}$. It is easily seen that $b_p = a_p(1 + g(p))$, where

$$g(p) = -\frac{\log(\log(p)) + \log(4\pi)}{4\log(p)} \propto \frac{\log(\log(p))}{\log(p)}.$$

This shows that, in particular, in the case of iid maxima (as well as in the general case where (3.20) holds) the normalization $\sqrt{2\log(p)}$ *does not lead* to the optimal rate, since

$$\frac{1}{\sqrt{2\log(p)}} M_p^* - 1 \propto_P \frac{\log(\log(p))}{\log(p)},$$

where $\xi_p \propto_P \eta_p$ means that $\xi_p/\eta_p \rightarrow c$ in probability, for some positive constant c .

The optimal rate is $1/\log(p)$ and it is obtained by normalizing with any sequence b_p such that $p\bar{\Phi}(b_p) \rightarrow 1$. This follows from the next simple result, which shows that

the rate of concentration in (3.24) is the slower of the rates $1/a_p^2$ and $(b_p - a_p)/a_p$.

Lemma III.9. *Suppose that for some random variables ζ_p , we have $\zeta_p \xrightarrow{d} \zeta$, as $p \rightarrow \infty$, where ζ is a non-constant random variable. Then, for all sequences α_p and β_p , we have*

$$\alpha_p \zeta_p + \beta_p \xrightarrow{P} 0 \iff |\alpha_p| + |\beta_p| \longrightarrow 0.$$

That is, the rate of $\alpha_p \zeta_p + \beta_p$ is always the slower of the rates of $\{\alpha_p\}$ and $\{\beta_p\}$.

Proof. The “ \Leftarrow ” direction follows from Slutsky. To prove “ \Rightarrow ”, it is enough to show that for every $p(n) \rightarrow \infty$, there is a further sub-sequence $q(n) \rightarrow \infty$, $\{q(n)\} \subset \{p(n)\}$, such that

$$|\alpha_{q(n)}| + |\beta_{q(n)}| \longrightarrow 0.$$

In view of Skorokhod’s representation theorem (Theorem 6.7, page 70 in *Billingsley* (2013)), we may suppose that $\zeta_p^* \rightarrow \zeta^*$, with probability one, where $\zeta_p^* \stackrel{d}{=} \zeta_p$ and $\zeta^* \stackrel{d}{=} \zeta$. Also, assuming that $\alpha_{p(n)} \zeta_{p(n)}^* + \beta_{p(n)} \rightarrow 0$, in probability, implies that there is a further sub-sequence $q(n) \rightarrow \infty$, such that

$$\alpha_{q(n)} \zeta_{q(n)}^*(\omega) + \beta_{q(n)} \rightarrow 0, \quad \text{as } q(n) \rightarrow \infty, \quad (3.27)$$

for P -almost all ω . Since also $\zeta_{q(n)}^*(\omega) \rightarrow \zeta^*(\omega)$, for P -almost all ω , and since ζ^* is non-constant, we have $\zeta_{q(n)}^*(\omega_i) \rightarrow \zeta^*(\omega_i)$, $i = 1, 2$ for some $\zeta^*(\omega_1) \neq \zeta^*(\omega_2)$.

Thus, by subtracting two instances of Relation (3.27) corresponding to $\omega = \omega_1$ and $\omega = \omega_2$, we obtain

$$\alpha_{q(n)} (\zeta_{q(n)}^*(\omega_1) - \zeta_{q(n)}^*(\omega_2)) \rightarrow 0,$$

which since $(\zeta_{q(n)}^*(\omega_1) - \zeta_{q(n)}^*(\omega_2)) \rightarrow \zeta^*(\omega_1) - \zeta^*(\omega_2) \neq 0$, implies $\alpha_{q(n)} \rightarrow 0$. This, in view of (3.27) yields $\beta_{q(n)} \rightarrow 0$, and completes the proof. \square

Remark III.10. The above considerations establish the optimal rate of concentration of the maxima $M_p = \max_{i \in [p]} \epsilon_p(i)$, whenever the limit in distribution (3.20) holds. We have shown that this optimal rate is $1/\log(p)$ and is in fact obtained, when considering M_p/u_p , for $p\bar{\Phi}(u_p) \sim 1$. The rate of concentration of $M_p/\sqrt{2\log(p)}$ is $\log(\log(p))/\log(p)$, which is only slightly sub-optimal.

On the other hand, as we know by Theorem III.4, uniform relative stability is equivalent to UDD and hence the concentration of maxima phenomenon takes place even if (3.20) fails to hold. At this point, we do not know what is the optimal rate in general. In Section 3.4, we provide upper bounds on this rate. We conjecture, however, the presence of more severe dependence can only lead to slower rates of concentration and in particular the optimal rate of concentration for UDD arrays cannot be faster than $1/\log(p)$ – the one for independent maxima.

Conjecture III.11. *Let \mathcal{E} be a Gaussian URS array. Relation (3.30) implies $\delta_p \gg 1/\log(p)$.*

3.4 Rates of uniform relative stability

3.4.1 Gaussian arrays

Throughout Sections 3.4 and 3.5, $\mathcal{E} = \{\epsilon_p(i), i \in [p]\}$ will be a Gaussian array with standard Normal marginals, unless stated otherwise. We shall also assume that \mathcal{E} is URS. For simplicity of notation and without loss of generality we will work with $S_p = [p]$ (see Remark III.15). We will obtain upper bounds on the rate, i.e., sufficient conditions on the dependence structure of \mathcal{E} , which ensure certain rates. These results are of independent interest and will find concrete applications in Section 3.4.2, where conditions ensuring the URS of functions of Gaussian arrays are established.

The following definition is an ancillary tool for the comparison of the rates of two vanishing sequences and introduces some notation for this purpose.

Definition III.12. Let $(\alpha_p)_{p=1}^\infty$ and $(\beta_p)_{p=1}^\infty$ be two positive sequences converging to 0. We will say that α_p is of lower order than β_p (or slower than β_p), denoted by $\alpha_p \gg \beta_p$, if $\beta_p/\alpha_p \rightarrow 0$, as $p \rightarrow \infty$, i.e., $\beta_p = o(\alpha_p)$.

The next theorem constitutes the main result of this chapter.

Theorem III.13. Consider a UDD Gaussian triangular array $\mathcal{E} = \{\epsilon_p(i), i \in [p]\}$ with standard Normal marginals and let $N_{\mathcal{E}}(\tau)$ be as in Definition III.3. Let $\tau(p) \rightarrow 0$ be such that

$$\alpha(p) := \log N_{\mathcal{E}}(\tau(p))/\log(p) \rightarrow 0, \quad \text{as } p \rightarrow \infty. \quad (3.28)$$

Then, for all $\delta_p > 0$ such that

$$\delta_p \gg \alpha(p) + \tau(p) + \frac{1}{\log(p)}, \quad (3.29)$$

we have

$$\mathbb{P} \left(\left| \frac{\max_{i \in [p]} \epsilon_p(i)}{u_p} - 1 \right| > \delta_p \right) \rightarrow 0, \quad \text{as } p \rightarrow \infty. \quad (3.30)$$

Here u_p is defined as in (3.8) taking $F = \Phi$, the cumulative distribution function of standard Normal distribution.

The proof of Theorem III.13 depends on a number of technical results, which will be presented and proved in Section 3.5. In order to make the proof easier for the reader to follow, we postpone its demonstration until Section 3.5. We proceed next with several comments and examples.

Remark III.14. Note that in Theorem III.13 the covariance structure of \mathcal{E} appears only through $N_{\mathcal{E}}(\tau)$. The collection $\{N_{\mathcal{E}}(\tau), \tau \in (0, 1)\}$ constitutes a collection of uniform upper bounds on the number of covariances in each row of the triangular array \mathcal{E} that exceed the threshold τ . This means that the ordering of the p random variables in each row of \mathcal{E} is irrelevant.

Remark III.15. The support recovery results of *Gao and Stoev (2020)* require URS in the sense of (3.11) for a subsequence $S_p \subset [p]$, with $|S_p| \rightarrow \infty$. By the previous remark, upon relabelling the triangular array \mathcal{E} , Theorem III.13 applies in this setting with p replaced by $|S_p|$, and entails rates on the convergence in (3.11).

The preceding Theorem III.13 gives us an upper bound on the rate at which the convergence in (3.11) takes place for a UDD Gaussian array \mathcal{E} . Observe that this bound depends crucially on the covariance structure of \mathcal{E} through $N_{\mathcal{E}}(\tau)$. This dependence will be illustrated in the following examples, where the upper bound stated in (3.29) is obtained for three specific covariance structures.

Example III.16. *The iid case and optimality of the rate bounds.*

Suppose that all $\epsilon_p(j)$'s are iid. Then, we can pick $\tau(p) = 0$ or $\tau < 1$ vanishing to 0 arbitrarily fast, and we would have that $N_{\mathcal{E}}(\tau) = 1$, because of the strict inequality in (3.12). This implies that $\alpha(p) = \log(N_{\mathcal{E}}(\tau))/\log(p) = 0$. Thus, in this case, the upper bound in (3.29) becomes $1/\log(p)$. Observe that this rate matches the optimal rate in Conjecture III.11.

Example III.17. *Power-law covariance decay.*

Consider, first, the simple case where \mathcal{E} comes from a stationary Gaussian time series, $\epsilon_p(\kappa) = \epsilon(\kappa)$, with auto-covariance

$$\rho(\kappa) = \text{Cov}(\epsilon(\kappa), \epsilon(0)) \propto \kappa^{-\gamma}, \quad \gamma > 0. \quad (3.31)$$

Then, the classic Berman condition $\rho(\kappa) = o(1/\log(\kappa))$ holds and as shown in the discussion after Proposition III.7, the optimal rate in (3.11) is $1/\log(p)$.

In this example, we will demonstrate that our result [Theorem III.13] leads to the nearly optimal rate $\log(\log(p))/\log(p)$. As in the previous remark, we see that this is in fact the optimal rate if u_p in (3.11) is replaced by $\sqrt{2\log(p)}$. (See Section 3.3). Note, however, that our arguments apply in greater generality and do not depend on

the stationarity assumption. Indeed, assume that \mathcal{E} is a general Gaussian triangular array such that (UDD') of *Gao and Stoev (2020)* holds, i.e.,

$$|\text{Cov}(\epsilon_p(i), \epsilon_p(j))| \leq c |\pi_p(i) - \pi_p(j)|^{-\gamma} \quad (3.32)$$

for suitable permutations π_p of $\{1, \dots, p\}$, where c does not depend on p . (Note that (3.32) entails (3.31) for $\pi_p = id$, where id is the identity permutation.) Then, one can readily show that $N_{\mathcal{E}}(\tau) = \mathcal{O}(\tau^{-1/\gamma})$, as $\tau \rightarrow 0$. Thus,

$$\alpha(p) = \frac{\log(N_{\mathcal{E}}(\tau))}{\log(p)} \asymp \frac{\log(\tau^{-1/\gamma})}{\log(p)} = -\frac{\log(\tau)}{\gamma \log(p)}.$$

Using this $\alpha(p)$, the upper bound on the rate in Theorem III.13 becomes

$$\alpha(p) + \tau(p) + \frac{\log(\log(p))}{\log(p)} \asymp -\frac{\log(\tau)}{\log(p)} + \tau(p) + \frac{1}{\log(p)} \asymp -\frac{\log(\tau)}{\log(p)} + \tau(p). \quad (3.33)$$

This is minimized by taking $\tau(p) = 1/\log(p)$ in (3.33) and the upper bound on the rate becomes

$$\alpha(p) + \tau(p) + \frac{1}{\log(p)} \asymp \frac{\log(\log(p))}{\log(p)} + \frac{1}{\log(p)} \asymp \frac{\log(\log(p))}{\log(p)}.$$

Recall that in the case when \mathcal{E} has iid components, the optimal rate of concentration of the maxima is $1/\log(p)$ and in fact it becomes $\log(\log(p))/\log(p)$ when one uses the normalization $\sqrt{2\log(p)}$ in place of u_p . Therefore, this example shows that under mild power-law type covariance decay conditions, Gaussian triangular arrays continue to concentrate at the nearly optimal rates for the iid setting.

Example III.18. *Logarithmic covariance decay.*

Following suit from Example III.17, we consider first the case where the errors come

from a stationary time series with auto-covariance

$$\rho(\kappa) = \text{Cov}(\epsilon(\kappa), \epsilon(0)) \propto (\log(\kappa))^{-\nu}, \quad \text{as } \kappa \rightarrow \infty, \quad (3.34)$$

for some $\nu > 0$. Note that for $0 < \nu < 1$, the Berman condition $\rho(\kappa) = o(1/\log(\kappa))$ is no longer satisfied and the results from Section 3.3 cannot be applied to establish the optimal rate in (3.11). Using Theorem III.13, we will see that an upper bound on this rate is $\delta_p^* := (\log(p))^{-\frac{\nu}{\nu+1}}$.

Indeed, consider the more general case where \mathcal{E} is a Gaussian triangular array, such that (UDD') of *Gao and Stoev (2020)* holds, i.e.,

$$|\text{Cov}(\epsilon_p(i), \epsilon_p(j))| \leq c (\log(|\pi_p(i) - \pi_p(j)|))^{-\nu}, \quad (3.35)$$

for suitable permutations π_p of $\{1, \dots, p\}$ and c does not depend on p . Again, note that (3.35) implies (3.34) for the identity permutation. One can show that in this case $N_{\mathcal{E}}(\tau) = \mathcal{O}(e^{\tau^{-1/\nu}})$, as $\tau \rightarrow 0$ and thus,

$$\alpha(p) = \frac{\log(N_{\mathcal{E}}(\tau))}{\log(p)} \propto \frac{\log(e^{\tau^{-1/\nu}})}{\log(p)} = \frac{1}{\tau^{1/\nu} \log(p)}, \quad \text{as } p \rightarrow \infty.$$

To find the best bound on the rate in the context of (3.29) we minimize

$$\alpha(p) + \tau(p) + \frac{1}{\log(p)} \propto \frac{1}{\tau^{1/\nu} \log(p)} + \tau + \frac{1}{\log(p)},$$

with respect to τ . Considering p fixed, basic calculus gives us that the r.h.s. is minimized for $\tau(p) = (\nu \log(p))^{-\frac{\nu}{\nu+1}}$. With this choice of τ the fastest upper bound from Theorem III.13 becomes

$$\left[\nu^{-\frac{\nu}{\nu+1}} + \nu^{-\frac{1}{\nu+1}} \right] \cdot (\log(p))^{-\frac{\nu}{\nu+1}} + \frac{1}{\log(p)} \propto (\log(p))^{-\frac{\nu}{\nu+1}}.$$

It only remains to show that the choice of τ actually allows us to pick $N_{\mathcal{E}}(\tau) = \mathcal{O}\left(e^{\tau^{-1/\nu}}\right)$. A sufficient condition would be $p \geq \tilde{c} \cdot e^{\tau^{-1/\nu}}$ for a suitably chosen constant \tilde{c} not depending on either p or τ . Substituting $\tau = (\nu \log(p))^{-\frac{\nu}{\nu+1}}$, we equivalently need

$$p \geq \tilde{c} \cdot e^{(\nu \log(p))^{\frac{1}{\nu+1}}}.$$

It is readily checked, by taking logarithms in both sides, that this holds for p sufficiently large and thus, the fastest upper bound for this kind of dependence structure is $(\log(p))^{-\frac{\nu}{\nu+1}}$.

Observe that as $\nu \rightarrow \infty$ this upper bound approaches asymptotically the optimal rate $1/\log(p)$ achieved under the Berman condition (see Section 3.3). Our results yield, however, an upper bound on the rate of concentration in (3.11) for the case $0 < \nu < 1$, where the Berman condition does not hold.

3.4.2 Functions of Gaussian arrays

The main motivation behind the work in this section is to determine when the concentration of maxima property is preserved under transformations. Specifically, consider the triangular array

$$\mathcal{H} = \{\eta_p(j) = f(\epsilon_p(j)), j \in [p], p \in \mathbb{N}\}, \quad (3.36)$$

where $\mathcal{E} = \{\epsilon_p(j), j \in [p], p \in \mathbb{N}\}$ is a Gaussian triangular array with standard Normal marginals.

Given that (3.30) holds, our goal is to find bounds on a sequence $d_p \downarrow 0$, such that

$$\mathbb{P}\left(\left|\frac{\max_{j \in [p]} \eta_p(j)}{v_p} - 1\right| > d_p\right) \rightarrow 0, \quad \text{as } p \rightarrow \infty, \quad (3.37)$$

where $v_p = f(u_p)$ and u_p is as in (3.8). We first address the case of monotone non-

decreasing transformations.

Proposition III.19. *Assume that f is a non-decreasing differentiable and eventually strictly increasing function, with $\lim_{x \rightarrow \infty} f(x) \neq 0$ and the derivative $f'(x)$ is either eventually increasing or eventually decreasing as $x \rightarrow \infty$. If (3.30) holds with some $\delta_p > 0$, then (3.37) holds provided that*

$$d_p \geq d_p^* := \frac{u_p \delta_p \max \{|f'(u_p(1 - \delta_p))|, |f'(u_p(1 + \delta_p))|\}}{|f(u_p)|}. \quad (3.38)$$

Proof. Since $u_p \uparrow \infty$, by the monotonicity of f and the fact that it is eventually strictly increasing, one can show that $f(u_p) = v_p = F_\eta^{\leftarrow}(1 - 1/p)$, for p large enough.

We start by noticing that

$$\left| \frac{\max_{j \in [p]} \eta_p(j)}{v_p} - 1 \right| = \left| \frac{\max_{j \in [p]} f(\epsilon_p(j)) - f(u_p)}{f(u_p)} \right| = \left| \frac{f(\max_{j \in [p]} \epsilon_p(j)) - f(u_p)}{f(u_p)} \right|, \quad (3.39)$$

where the second equality follows by the monotonicity of f .

Now recall that f is differentiable. By the Mean Value Theorem, there exists a possibly random θ_p between u_p and $\max_{j \in [p]} \epsilon_p(j)$, such that

$$\left| \frac{f(\max_{j \in [p]} \epsilon_p(j)) - f(u_p)}{f(u_p)} \right| = \left| \frac{1}{f(u_p)} f'(\theta_p) \left(\max_{j \in [p]} \epsilon_p(j) - u_p \right) \right|. \quad (3.40)$$

Combining (3.39) and (3.40), we obtain

$$\begin{aligned} \mathbb{P} \left(\left| \frac{\max_{j \in [p]} \eta_p(j)}{v_p} - 1 \right| > d_p \right) &= \mathbb{P} \left(\left| \frac{u_p f'(\theta_p)}{f(u_p)} \right| \cdot \left| \frac{\max_{j \in [p]} \epsilon_p(j)}{u_p} - 1 \right| > d_p \right) \\ &= \mathbb{P} \left(\left| \frac{\max_{j \in [p]} \epsilon_p(j)}{u_p} - 1 \right| > \frac{d_p |f(u_p)|}{u_p |f'(\theta_p)|} \right), \end{aligned}$$

where the second equality follows from the fact that $f'(\theta_p) \neq 0$ over the event of interest, since $d_p > 0$. This shows that for any non-negative sequence δ_p vanishing to

0, such that (3.30) holds, we have that

$$\mathbb{P} \left(\left| \frac{\max_{j \in [p]} \eta_p(j)}{v_p} - 1 \right| > \tilde{d}_p \right) \rightarrow 0, \quad \text{as } p \rightarrow \infty, \quad (3.41)$$

where

$$\tilde{d}_p := \frac{u_p \delta_p |f'(\theta_p)|}{|f(u_p)|}. \quad (3.42)$$

Now, we know by (3.30) that

$$|\theta_p - u_p| \leq \left| \max_{j \in [p]} \epsilon_p(j) - u_p \right| \leq u_p \delta_p$$

with probability going to 1, as $p \rightarrow \infty$. This implies that

$$\mathbb{P}(u_p(1 - \delta_p) \leq \theta_p \leq u_p(1 + \delta_p)) \rightarrow 1, \quad \text{as } p \rightarrow \infty,$$

In turn, by the eventual monotonicity of f' , the last convergence implies that

$$\mathbb{P}(|f'(\theta_p)| \leq \max\{|f'(u_p(1 - \delta_p))|, |f'(u_p(1 + \delta_p))|\}) \rightarrow 1, \quad \text{as } p \rightarrow \infty,$$

and equivalently

$$\mathbb{P}(\tilde{d}_p \leq d_p^*) \rightarrow 1, \quad \text{as } p \rightarrow \infty. \quad (3.43)$$

By (3.42) and (3.43) we conclude that (3.41) holds with \tilde{d}_p substituted by d_p^* . This shows that d_p^* is an upper bound of the optimal rate of concentration, i.e., (3.38) implies (3.37). \square

A typical and very important case where Proposition III.19 applies is when the array \mathcal{E} undergoes an exponential transformation, illustrated in the following example.

Example III.20. Let \mathcal{E} be as in Proposition III.19 and consider

$$\mathcal{H}_{\mathcal{E}} = \{\eta_p(j) := e^{\epsilon_p(j)}, j \in [p], p \in \mathbb{N}\}, \quad (3.44)$$

which is a triangular array with lognormal marginal distributions. This is sometimes referred to as the multivariate lognormal model (*Halliwell, 2015*). Let δ_p be such that (3.30) holds. Then, an immediate application of Proposition III.19 shows that as long as $u_p \delta_p \rightarrow 0$, an upper bound on the rate of convergence in (3.37) is

$$d_p^* = u_p \delta_p e^{u_p \delta_p} \sim u_p \delta_p \sim \delta_p \sqrt{2 \log(p)}.$$

That is, lognormal arrays can have relatively stable maxima, provided that the underlying maxima of the Gaussian array concentrate at a rate $\delta_p = o\left(1/\sqrt{\log(p)}\right)$.

Popular models like the ones with χ_1^2 marginals can be obtained from Proposition III.19 with the monotone transformation $f(x) := F^{-1}(\Phi(x))$, where F is the cdf of the desired distribution. The classic multivariate χ_1^2 -models, however, are obtained by squaring the elements of the Gaussian array, i.e., via the non-monotone transformation $f(x) = x^2$. Such models are addressed in the next result.

Corollary III.21. *Let all the assumptions of Proposition III.19 hold and let d_p^* be defined as before. Assume now that f is an even ($f(x) = f(-x)$) differentiable and eventually strictly increasing function, with $\lim_{x \rightarrow \infty} f(x) \neq 0$. Assume also that f is monotone non-decreasing on $(0, \infty)$. Then, the conclusion (3.38) still holds.*

Proof. We start by observing that

$$\begin{aligned} \mathbb{P} \left(\left| \frac{\max_{j \in [p]} \eta_p(j)}{f(u_p)} - 1 \right| > d_p \right) &= \mathbb{P} \left(\left| \frac{\max_{j \in [p]} f(\epsilon_p(j)) - f(u_p)}{f(u_p)} \right| > d_p \right) \\ &\leq \mathbb{P} \left(\left| \frac{f(\min_{j \in [p]} \epsilon_p(j)) - f(u_p)}{f(u_p)} \right| > d_p \right) + \mathbb{P} \left(\left| \frac{f(\max_{j \in [p]} \epsilon_p(j)) - f(u_p)}{f(u_p)} \right| > d_p \right), \end{aligned} \quad (3.45)$$

because the symmetry and monotonicity of f on $(0, \infty)$ imply that $\max_{j \in [p]} f(\epsilon_p(j))$ equals either $f(\max_{j \in [p]} \epsilon_p(j))$ or $f(\min_{j \in [p]} \epsilon_p(j))$.

By Proposition III.19 we can readily obtain that for $d_p \geq d_p^*$ the second term of (3.45) converges to 0. Now, we handle the first term of (3.45). By the symmetry of f we have that

$$f(\min_{j \in [p]} \epsilon_p(j)) = f(-\min_{j \in [p]} \epsilon_p(j)) = f(\max_{j \in [p]} (-\epsilon_p(j))).$$

Notice that by verifying the equality of the covariance structures, we have

$$\{-\epsilon_p(j), j \in [p]\} \stackrel{d}{=} \{\epsilon_p(j), j \in [p]\}.$$

Hence $\max_{j \in [p]} (-\epsilon_p(j)) \stackrel{d}{=} \max_{j \in [p]} \epsilon_p(j)$, and again by Proposition III.19 we get that for $d_p \geq d_p^*$ the first term of (3.45) also converges to 0. This completes the proof. \square

Using Corollary III.21 we can now treat the multivariate χ^2 model introduced in *Dasgupta and Spurrier (1997)*.

Example III.22. Let \mathcal{E} be as in Proposition III.19 and consider

$$\mathcal{H}_{\mathcal{E}} = \{\eta_p(j) := \epsilon_p^2(j), j \in [p], p \in \mathbb{N}\},$$

a triangular array with χ_1^2 marginal distributions. Let δ_p be as in (3.30). Then, a simple application of Corollary III.21 implies (3.37), provided

$$d_p \geq d_p^* = 2\delta_p(1 + \delta_p) \sim 2\delta_p.$$

In contrast to Example III.20, taking squares does not lead to a slower rate of convergence. Indeed, in Example III.20 our estimate of the rate is slowed down by a factor of $\sqrt{\log(p)}$, while in the χ^2 case it remains δ_p .

We shall now see that the rate of convergence is not slowed down by any power transformation $x \mapsto x^\lambda$, for any $\lambda > 0$.

Example III.23. *Power-Law Transformations.*

Let once again \mathcal{E} be as in Proposition III.19 and consider the power transformations $f(x) = x^\lambda$, $\lambda > 0$. In the cases where $\lambda \notin \mathbb{N}$, we use the functions $f_1^\lambda(x) = |x|^\lambda$ or $f_2^\lambda(x) = x^{<\lambda>} = \text{sign}(x) \cdot |x|^\lambda$. Note that differentiability at 0 is not needed in any of the proofs, so using f_1^λ does not violate any of the assumptions. Let also δ_p be as in (3.30), i.e., a rate sequence for the convergence in (3.11). Then, a suitable application of Proposition III.19 or Corollary III.21, shows that an upper bound on the rate of convergence in (3.37) is

$$d_p^* = \lambda \delta_p (1 + \delta_p)^{\lambda-1} \sim \lambda \delta_p \quad \text{or} \quad d_p^* = \lambda \delta_p (1 - \delta_p)^{\lambda-1} \sim \lambda \delta_p.$$

In view of Examples III.16, III.17 and III.18, we now show how the rate $d_p^* \sim \lambda \delta_p$ is affected under different correlation structures of the underlying Gaussian array \mathcal{E} . Recall that in the iid case of Example III.16 we have that the optimal rate is $\delta_p \gg \delta_p^{\text{opt}} = 1/\log(p)$. This implies that an upper bound on the rate of concentration is

$$d_p^* \sim \lambda \delta_p \gg \frac{\lambda}{\log(p)}.$$

Moreover, for the power-law covariance decay covariance structure (Example III.17), we observe that compared to the iid case, the rate of concentration δ_p is scaled by a factor of $\log(\log(p))$. Namely, for the power-law transformations we get that the upper bound is

$$d_p^* \sim \lambda \delta_p \sim \frac{\lambda \log(\log(p))}{\log(p)}.$$

Finally, we examine the logarithmic covariance decay (Example III.18). Remember that in this case the rate we have for \mathcal{E} is $\delta_p = (\log(p))^{-\frac{\nu}{\nu+1}}$. This implies that

the upper bound of the rate of concentration for the power-law transformations is

$$d_p^* \sim \lambda \delta_p \sim \frac{\lambda}{(\log(p))^{\frac{\nu}{\nu+1}}}.$$

Observe that in this case, d_p^* is a valid upper bound aside from the value of ν . We will see in the following Example III.24, that the same is not true for the exponential power-law transformations.

In the last example of this section, we explore exponential power transformations and how they affect our bounds on the rate of convergence.

Example III.24. *Exponential Power-Law Transformations.*

Let \mathcal{E} be as in Proposition III.19 and consider the exponential power transformations $f(x) = e^{x^\lambda}$, $\lambda > 0$, $\lambda \neq 1$. (Note that $\lambda = 1$ is the lognormal case which we have already seen in Example III.20). In the cases where $\lambda \notin \mathbb{N}$, we use the functions $f_1^\lambda(x) = e^{|x|^\lambda}$ or $f_2^\lambda(x) = e^{x^{<\lambda>}} = e^{\text{sign}(x) \cdot |x|^\lambda}$. Similarly to Example III.23, differentiability at 0 is not needed in any of the proofs, so using f_1^λ does not violate any of the assumptions. Let also δ_p be as in (3.30). Then, suitable applications of Proposition III.19 or Corollary III.21 show that as long as $u_p^\lambda \delta_p \rightarrow 0$, an upper bound on the rate of convergence in (3.37) is

$$d_p^* = \lambda u_p^\lambda \delta_p (1 + \delta_p)^{\lambda-1} e^{u_p^\lambda [(1+\delta_p)^\lambda - 1]}, \quad \text{if } \lambda \geq 1$$

and

$$d_p^* = \lambda u_p^\lambda \delta_p (1 - \delta_p)^{\lambda-1} e^{u_p^\lambda [(1-\delta_p)^\lambda - 1]}, \quad \text{if } 0 < \lambda < 1.$$

In both cases we have $d_p^* \sim \lambda \delta_p (2 \log(p))^{\lambda/2}$, as $p \rightarrow \infty$. As a generalization of the lognormal case ($\lambda = 1$), we see that the iid rate δ_p is scaled by a factor of $(\sqrt{\log(p)})^\lambda$. This means that this kind of arrays would still have relatively stable maxima, provided that the underlying maxima of the Gaussian array concentrate at a rate $\delta_p = o(1/(\log(p))^{\lambda/2})$.

At this point, we examine how the rate $d_p^* \sim \lambda \delta_p (2 \log(p))^{\lambda/2}$ adjusts under the varying covariance structures of \mathcal{E} in Examples III.16, III.17 and III.18. In an analogous manner to Example III.23, we get that for the iid case, an upper bound on the rate of concentration is

$$d_p^* \sim \lambda \delta_p u_p^\lambda \gg 2^{\frac{\lambda}{2}} \lambda (\log(p))^{\frac{\lambda}{2}-1},$$

while for the power-law covariance decay covariance structure we obtain

$$d_p^* \sim \lambda \delta_p u_p^\lambda \sim 2^{\frac{\lambda}{2}} \lambda (\log(p))^{\frac{\lambda}{2}-1} \log(\log(p)).$$

In the previous two instances we notice that the covariance structure does not impose any restrictions on the values of λ , in order to guarantee concentration of maxima for the transformed triangular array. This is not the case for the logarithmic covariance decay, since the upper bound becomes

$$d_p^* \sim \lambda \delta_p u_p^\lambda \sim 2^{\frac{\lambda}{2}} \lambda (\log(p))^{\frac{\lambda}{2} - \frac{\nu}{\nu+1}}.$$

The aforementioned d_p^* is a sensible upper bound for the rate of concentration in this case, only if $d_p^* \rightarrow 0$, as $p \rightarrow \infty$. This is so, when $\nu > \frac{\lambda}{2+\lambda}$. Thus, our results imply that in the lognormal case ($\lambda = 1$), $\nu > \frac{1}{3}$ guarantees that the transformed array is relatively stable.

Remark III.25. In Conjecture III.11, we posit that the fastest rate of convergence for a UDD Gaussian array is bounded above by $1/\log(p)$. Nevertheless, from Example III.16 for the iid case, our bound in (3.29) is again $1/\log(p)$. Since $u_p \sim \sqrt{2 \log(p)}$, we see that we can get an upper bound on the rate of $f(x) = e^{x^\lambda}$ only for $0 < \lambda < 2$. The range $\lambda \in (0, 2)$ is also natural, because one can show that the transformation $f(x) = e^{x^\lambda}$, for $\lambda \geq 2$, leads to heavy power-law distributed variables $\eta_p(j)$. Heavy-

tailed random variables no longer have relatively stable maxima, which makes the question about the rate of concentration of maxima meaningless.

We will end this section with a corollary, readily obtained by the discussion in the end of Example III.24.

Corollary III.26. *Suppose that $\mathcal{H} := \{\eta_p(j), j \in [p], p \in \mathbb{N}\}$ is a multivariate log-normal array as in (3.44). Suppose that*

$$|\text{Cov}(\eta_p(j), \eta_p(k))| \leq c \cdot \frac{1}{(\log(|\pi_p(j) - \pi_p(k)|))^\nu}, \quad (3.46)$$

for some $\nu > 1/3$, permutations π_p of $\{1, \dots, p\}$ and a constant c independent of p . Then the array \mathcal{H} is URS.

Proof. Let $\mathcal{E} = \{\epsilon_p(j), j \in [p], p \in \mathbb{N}\}$ be the underlying Gaussian array. Then, we have that $\eta_p(j) = e^{\epsilon_p(j)}$ for every $j \in [p]$. Thus,

$$\begin{aligned} \text{Cov}(\eta_p(j), \eta_p(k)) &= \text{Cov}(e^{\epsilon_p(j)}, e^{\epsilon_p(k)}) \\ &= \mathbb{E}(e^{\epsilon_p(j) + \epsilon_p(k)}) - \mathbb{E}(e^{\epsilon_p(j)}) \mathbb{E}(e^{\epsilon_p(k)}). \end{aligned} \quad (3.47)$$

Recall that the moment generating function for a Normal random variable $X \sim N(\mu, \sigma^2)$ is $M(t) = \mathbb{E}(e^{tX}) = e^{\mu t + \sigma^2 t^2 / 2}$. Since $\epsilon_p(i)$ follow the standard Normal distribution, we have $\epsilon_p(j) + \epsilon_p(k) \sim N(0, 2 + 2\text{Cov}(\epsilon_p(j), \epsilon_p(k)))$, and hence (3.47) becomes

$$\text{Cov}(\eta_p(j), \eta_p(k)) = e \cdot (e^{\text{Cov}(\epsilon_p(j), \epsilon_p(k))} - 1). \quad (3.48)$$

In turn, (3.48) along with (3.46) implies that

$$|e \cdot (e^{\text{Cov}(\epsilon_p(j), \epsilon_p(k))} - 1)| \leq \frac{c}{e} \cdot \frac{1}{(\log(|\pi_p(j) - \pi_p(k)|))^\nu}. \quad (3.49)$$

Using the inequality $|x| \leq e|e^x - 1|$, $x \in [-1, 1]$ in (3.49), since $|\text{Cov}(\epsilon_p(j), \epsilon_p(k))| \leq 1$,

we finally obtain that

$$|\text{Cov}(\epsilon_p(j), \epsilon_p(k))| \leq \frac{c}{e} \cdot \frac{1}{(\log(|\pi_p(j) - \pi_p(k)|))^\nu}.$$

The last relation implies that \mathcal{E} has a logarithmic covariance decay covariance structure (see Example III.18). Combined with the discussion in the end of Example III.24, the proof is complete. \square

3.5 Technical proofs

In this section we present the proof of the capstone Theorem III.13. Recall that we desire to find an upper bound on the rate of positive vanishing sequences δ_p , such that

$$\mathbb{P} \left(\left| \frac{\max_{i \in [p]} \epsilon_p(i)}{u_p} - 1 \right| > \delta_p \right) \rightarrow 0, \quad \text{as } p \rightarrow \infty.$$

To this end, let

$$\xi_p := \frac{1}{u_p} \max_{i \in [p]} \epsilon_p(i), \tag{3.50}$$

where $\mathcal{E} = \{\epsilon_p(i), i \in [p]\}$ is a URS Gaussian array with standard Normal marginals. Observe that

$$\begin{aligned} \mathbb{P}(|\xi_p - 1| > \delta_p) &= \mathbb{P}(\xi_p > 1 + \delta_p) + \mathbb{P}(\xi_p < 1 - \delta_p) \\ &=: \text{I}(\delta_p) + \text{II}(\delta_p). \end{aligned} \tag{3.51}$$

Thus, to obtain the desired rate we need to recover a bound on the rate of $\text{I}(\delta_p)$ and $\text{II}(\delta_p)$. Note that in our endeavor to secure upper bound on the term $\text{II}(\delta_p)$ we will use the expectation of ξ_p . The integrability of ξ_p is ensured by Appendix A.2 of *Chatterjee* (2014), or *Pickands III* (1968) in conjunction with (3.3).

Term $\text{I}(\delta_p)$. In the following proposition, we find an upper bound on the rate of δ_p

in $I(\delta_p)$ of (3.51). Interestingly, the following result does not involve the dependence structure of the array \mathcal{E} .

Proposition III.27. *Let $\mathcal{E} = \{\epsilon_p(i), i \in [p]\}$ be an arbitrary Gaussian triangular array, where the marginal distributions are standard Normal and let ξ_p be defined as in (3.50). If $\delta_p \rightarrow 0$ is a positive sequence such that*

$$\delta_p \gg \frac{1}{\log(p)} \tag{3.52}$$

then, regardless of the dependence structure of \mathcal{E} , we have

$$\lim_{p \rightarrow \infty} (\delta_p^{-1} \mathbb{E}(\xi_p - 1)_+) = 0, \tag{3.53}$$

and consequently $\mathbb{P}(\xi_p > 1 + \delta_p) \rightarrow 0$, as $p \rightarrow \infty$.

We need the following simple bound for the Mill's ratio (see also (1.2.2) or (2.1.1) in *Adler and Taylor (2009)*).

Lemma III.28. *For all $u > 0$, we have*

$$1 - \frac{1}{1 \vee u^2} \leq \frac{\bar{\Phi}(u)}{\phi(u)/u} \leq 1,$$

where $\phi(u) = e^{-u^2/2}/\sqrt{2\pi}$ and $\bar{\Phi}(u) = \int_u^\infty \phi(x)dx$.

Proof. We have

$$\begin{aligned} \frac{\bar{\Phi}(u)}{\phi(u)/u} &= \frac{u}{\phi(u)} \int_u^\infty \phi(x)dx = u \int_u^\infty e^{-\frac{x^2-u^2}{2}} dx \\ &= u \int_0^\infty e^{-\frac{(z+u)^2-u^2}{2}} dz = \int_0^\infty e^{-\frac{z^2}{2}} u e^{-uz} dz = \mathbb{E}[e^{-E^2/(2u^2)}], \end{aligned}$$

where E is an exponentially distributed random variable with unit mean, and we used the change of variables $z := x - u$. Observing that $1 - x \leq e^{-x} \leq 1$, for all $x \geq 0$, we

get

$$1 - \frac{E^2}{2u^2} \leq e^{-E^2/(2u^2)} \leq 1.$$

The result follows upon taking expectation and recalling that $\mathbb{E}[E^2] = 2$. \square

Proposition III.27. Note first that (3.53) implies $\mathbb{P}(\xi_p > 1 + \delta_p) \rightarrow 0$. Indeed, this follows from the Markov inequality:

$$\mathbb{P}(\xi_p - 1 > \delta_p) = \mathbb{P}((\xi_p - 1)_+ > \delta_p) \leq \delta_p^{-1} \mathbb{E}(\xi_p - 1)_+.$$

Now, we focus on proving (3.53). We can write

$$\begin{aligned} \frac{1}{\delta_p} \mathbb{E}(\xi_p - 1)_+ &= \frac{1}{\delta_p} \int_0^\infty \mathbb{P}(\xi_p - 1 > z) dz \\ &= \int_0^\infty \mathbb{P}(\xi_p > 1 + \delta_p x) dx =: J(\delta_p), \end{aligned} \quad (3.54)$$

where in the last integral we used the change of variables $z = \delta_p x$.

Recalling that $\xi_p = u_p^{-1} \max_{i \in [p]} \epsilon_p(i)$, by the union bound, for the last integrand we have that

$$\mathbb{P}(\xi_p > 1 + \delta_p x) \leq p \bar{\Phi}(u_p(1 + \delta_p x)) = \frac{\bar{\Phi}(u_p(1 + \delta_p x))}{\bar{\Phi}(u_p)}. \quad (3.55)$$

By Lemma III.28, we further obtain that

$$\begin{aligned} \frac{\bar{\Phi}(u_p(1 + \delta_p x))}{\bar{\Phi}(u_p)} &\leq \frac{1}{1 - 1/(1 \vee u_p^2)} \cdot \frac{\phi(u_p(1 + \delta_p x))}{(1 + \delta_p x)\phi(u_p)} \\ &\leq \frac{1}{1 - 1/(1 \vee u_p^2)} \exp \left\{ -\frac{u_p^2}{2} \left((1 + \delta_p x)^2 - 1 \right) \right\} \\ &\leq B_p \exp \{ -u_p^2 \delta_p x \}, \end{aligned} \quad (3.56)$$

where $B_p := (1 - 1/(1 \vee u_p^2))^{-1} \rightarrow 1$, as $p \rightarrow \infty$, is a constant independent of $x \geq 0$ and in the last inequality we also used the simple bound $(1 + \delta_p x)^2 - 1 \geq 2\delta_p x$.

Condition (3.52) means that there is a sequence $\gamma(p)$ diverging to infinity slower than $\log(p)$ such that

$$\delta(p) = \frac{\gamma(p)}{\log(p)}.$$

Thus, by Relation (3.56) and the facts that $u_p^2 \sim 2 \log(p)$ and $B_p \sim 1$, as $p \rightarrow \infty$, we obtain

$$\frac{\bar{\Phi}(u_p(1 + \delta_p x))}{\bar{\Phi}(u_p)} \leq 2 \cdot e^{-2\gamma(p)x},$$

for all sufficiently large p . Since $\gamma(p) \rightarrow \infty$, Relation (3.55) and the Dominated Convergence Theorem applied to (3.54), implies

$$\lim_{p \rightarrow \infty} J(\delta_p) \leq \lim_{p \rightarrow \infty} \int_0^\infty 2e^{-2\gamma(p)x} dx = 0.$$

This completes the proof of (3.56). □

Term $\text{II}(\delta_p)$. Handling term II of (3.51) is more involved and this is where the dependence structure of the array plays a role. We start by presenting a more careful reformulation of Lemma B.1 in *Gao and Stoev* (2018).

Lemma III.29. *Let $(X_i)_{i=1}^p$ be p iid random variables with distribution F and density f , such that*

$$\mathbb{E}(X_i)_- \equiv \mathbb{E}(\max\{-X_i, 0\}) < \infty.$$

Denote the maximum of the X_i 's as $M_p := \max_{i=1, \dots, p} X_i$. Suppose that f is eventually decreasing, i.e., there exists a C_0 such that $f(x_1) \geq f(x_2)$ whenever $C_0 \leq x_1 \leq x_2$, then

$$\frac{\mathbb{E}M_p}{u_{p+1}} \geq (1 - F^p(C_0)) + \frac{\mathbb{E}[X_1 | X_1 < C_0]}{u_{p+1}} F^p(C_0),$$

where $u_{p+1} = F^{\leftarrow}(1 - 1/(p + 1))$.

Proof. For the proof, refer to the proof of Lemma B.1 in *Gao and Stoev* (2018). □

Recall that a Gaussian triangular array $\mathcal{E} = \{\epsilon_p(j)\}_{j=1}^p$ with standard Normal marginals is said to be UDD if for every $\tau > 0$,

$$N_{\mathcal{E}}(\tau) := \sup_{p \in \mathbb{N}} \max_{i=1, \dots, p} |\{\kappa \in [p] : \text{Cov}(\epsilon_p(i), \epsilon_p(\kappa)) > \tau\}| < \infty. \quad (3.57)$$

That is, for every p and $i \in [p]$, there are at most $N_{\mathcal{E}}(\tau)$ indices κ , such that the covariance between $\epsilon_p(i)$ and $\epsilon_p(\kappa)$ exceeds τ .

The function $N_{\mathcal{E}}(\tau)$ encodes certain aspects of the dependence structure of the array \mathcal{E} . It will play a key role in the derivation of the upper bound on the rate of concentration of maxima. The next result is an extension of Proposition A.1 in *Gao and Stoev* (2018) tailored to our needs. For the benefit of the reader, we reproduce the key argument involving a packing construction and the Sudakov-Fernique bounds, which may be of independent interest.

Proposition III.30. *For every UDD Gaussian array \mathcal{E} , and any subset $S_p \subseteq \{1, \dots, p\}$ with $q = |S_p|$, and $\tau \in (0, 1)$, we have that*

$$\mathbb{E} \left[\frac{\max_{j \in S_p} \epsilon_p(j)}{u_q} \right] \geq \frac{u_{q/N_{\mathcal{E}}(\tau)+1}}{u_q} \sqrt{1-\tau} \left(1 - \frac{1}{2^{q/N_{\mathcal{E}}(\tau)}} - \frac{\sqrt{2/\pi}}{u_{q/N_{\mathcal{E}}(\tau)+1}} \cdot \frac{1}{2^{q/N_{\mathcal{E}}(\tau)}} \right) \quad (3.58)$$

$$:= 1 - R_q, \quad (3.59)$$

where $N_{\mathcal{E}}(\tau)$ is given in (3.57).

Remark III.31. Note that without loss of generality we can assume $S_p = \{1, \dots, p\}$. We prove a slightly more general result, but the only application in this chapter will be for $q = p$.

Proof. Define the canonical (pseudo) metric on S_p ,

$$d(i, j) = \sqrt{\mathbb{E}(\epsilon(i) - \epsilon(j))^2}.$$

This metric takes values between 0 and 2, since $\epsilon_p(i)$, $i = 1, \dots, p$, have zero means and unit variances. Fix $\tau \in (0, 1)$, take $\gamma = \sqrt{2(1 - \tau)}$ and let Γ be a γ -packing of S_p . That is, let Γ be a subset of S_p , such that for any $i, j \in \Gamma$, $i \neq j$, we have $d(i, j) > \gamma$, i.e.,

$$d(i, j) = \sqrt{2(1 - \Sigma_p(i, j))} \geq \gamma = \sqrt{2(1 - \tau)},$$

or equivalently, $\Sigma_p(i, j) \leq \tau$. We claim that we can find a γ -packing Γ whose number of elements is at least

$$|\Gamma| \geq \frac{q}{N_{\mathcal{E}}(\tau)}. \quad (3.60)$$

Indeed, Γ can be constructed iteratively as follows:

Step 1: Set $S_p^{(0)} := S_p$ and $\Gamma := \{j_1\}$, where $j_1 \in S_p^{(0)}$ is an arbitrary element. Set $k := 1$.

Step 2: Set $S_p^{(k)} := S_p^{(k-1)} \setminus B_{\gamma}(j_k)$, where

$$B_{\gamma}(j_k) := \{i \in S_p : d(i, j_k) < \gamma\}.$$

Step 3: If $S_p^{(k)} \neq \emptyset$, pick an arbitrary $j_{k+1} \in S_p^{(k)}$, set $\Gamma := \Gamma \cup \{j_{k+1}\}$, and $k := k + 1$, go to Step 2; otherwise stop.

By the definition of UDD, there are at most $N_{\mathcal{E}}(\tau)$ coordinates whose covariance with $\epsilon_p(j)$ exceed τ . Therefore, at each iteration, $|B_{\gamma}(j_k)| \leq N_{\mathcal{E}}(\tau)$, and hence

$$|S_p^{(k)}| \geq |S_p^{(k-1)}| - |B_{\gamma}(j_k)| \geq q - kN_{\mathcal{E}}(\tau).$$

The construction can continue for at least $q/N_{\mathcal{E}}(\tau)$ iterations, which implies (3.60).

Now, we define on this γ -packing Γ an independent Gaussian process $\{\eta(j)\}_{j \in \Gamma}$,

$$\eta(j) = \frac{\gamma}{\sqrt{2}}Z(j), \quad j \in \Gamma,$$

where the $Z(j)$'s are iid standard Normal random variables. The increments of the new process are smaller than that of the original in the following sense,

$$\mathbb{E}(\eta(i) - \eta(j))^2 = \gamma^2 \leq d^2(i, j) = \mathbb{E}(\epsilon_p(i) - \epsilon_p(j))^2,$$

for all $i \neq j$, $i, j \in \Gamma$. Applying the Sudakov-Fernique inequality (see, e.g., Theorem 2.2.3 in *Adler and Taylor (2009)*) to $\{\eta(j)\}_{j \in \Gamma}$ and $\{\epsilon_p(j)\}_{j \in \Gamma}$, we have

$$\mathbb{E} \left[\max_{j \in \Gamma} (\eta(j)) \right] \leq \mathbb{E} \left[\max_{j \in \Gamma} (\epsilon_p(j)) \right] \leq \mathbb{E} \left[\max_{j \in S_p} (\epsilon_p(j)) \right].$$

This implies

$$\mathbb{E} \left[\frac{1}{u_q} \max_{j \in S_p} \epsilon_p(j) \right] \geq \mathbb{E} \left[\frac{1}{u_{|\Gamma|+1}} \max_{j \in \Gamma} \eta(j) \right] \cdot \frac{u_{|\Gamma|+1}}{u_q} \geq \frac{u_{|\Gamma|+1}}{u_q} \cdot \sqrt{1 - \tau} \cdot \mathbb{E} \left[\frac{1}{u_{|\Gamma|+1}} \max_{j \in \Gamma} Z(j) \right].$$

Now, the application of Lemma III.29 to the standard Normal distribution for $C_0 = 0$ entails that,

$$\frac{\mathbb{E} [\max_{j \in \Gamma} Z(j)]}{u_{|\Gamma|+1}} \geq 1 - \frac{1}{2^{|\Gamma|}} - \frac{\sqrt{2/\pi}}{u_{|\Gamma|+1}} \cdot \frac{1}{2^{|\Gamma|}}.$$

Since $|\Gamma| \geq q/N_{\mathcal{E}}(\tau)$ the desired lower bound in (3.58) is obtained. \square

We are now interested in the rate at which the lower bound in (3.58) converges to 1. Equivalently, we desire to find the rate of decay of R_q . This rate is obtained in the following Lemma.

Lemma III.32. *Let R_q , $\alpha(q)$ be defined as in (3.59) and (3.28) respectively. Then*

$$R_q \asymp \alpha(q) + \tau(q) + 2^{-q^{1-\alpha(q)}}, \quad \text{as } q \rightarrow \infty. \quad (3.61)$$

Proof. Note that by definition $R_q \rightarrow 0$, as $q \rightarrow \infty$. This implies that $R_q \sim \log(1 - R_q)$,

as $q \rightarrow \infty$, so we just need the rate of

$$\begin{aligned} \log(1 - R_q) &= \log \left(\frac{u_{q/N_{\mathcal{E}}(\tau)+1}}{u_q} \cdot \sqrt{1 - \tau(q)} \cdot \left(1 - \frac{1}{2^{q/N_{\mathcal{E}}(\tau)}} - \frac{\sqrt{2/\pi}}{u_{q/N_{\mathcal{E}}(\tau)+1}} \cdot \frac{1}{2^{q/N_{\mathcal{E}}(\tau)}} \right) \right) \\ &= \log \left(\frac{u_{q/N_{\mathcal{E}}(\tau)+1}}{u_q} \right) + \frac{1}{2} \log(1 - \tau(q)) + \log \left(1 - \frac{1}{2^{q/N_{\mathcal{E}}(\tau)}} - \frac{\sqrt{2/\pi}}{u_{q/N_{\mathcal{E}}(\tau)+1}} \cdot \frac{1}{2^{q/N_{\mathcal{E}}(\tau)}} \right). \end{aligned}$$

Now, the facts that $\alpha(q) = \log(N_{\mathcal{E}}(\tau))/\log(q)$ and $u_q \sim \sqrt{2 \log(q)}$ imply that

$$\frac{u_{q/N_{\mathcal{E}}(\tau)+1}}{u_q} \sim \sqrt{\frac{2 \log(1 + q^{1-\alpha(q)})}{2 \log(q)}} \sim \sqrt{\frac{\log(q^{1-\alpha(q)})}{\log(q)}} = \sqrt{1 - \alpha(q)},$$

where we used the relation

$$q^{1-\alpha(q)} = e^{\log(q) - \log(N_{\mathcal{E}}(\tau))} = \frac{q}{N_{\mathcal{E}}(\tau)}. \quad (3.62)$$

However, since $\alpha(q) = \log(N_{\mathcal{E}}(\tau(q)))/\log(q) \rightarrow 0$ and $\tau(q) \rightarrow 0$, we have

$$\log(1 - \alpha(q)) = -\alpha(q) + o(\alpha(q)),$$

$$\log(1 - \tau(q)) = -\tau(q) + o(\tau(q)),$$

and by (3.62)

$$\begin{aligned} \log \left(1 - \frac{1}{2^{q/N_{\mathcal{E}}(\tau)}} - \frac{\sqrt{2/\pi}}{u_{q/N_{\mathcal{E}}(\tau)+1}} \cdot \frac{1}{2^{q/N_{\mathcal{E}}(\tau)}} \right) &= \log \left(1 - 2^{-q^{1-\alpha(q)}} - \frac{\sqrt{2/\pi}}{u_{q/N_{\mathcal{E}}(\tau)+1}} \cdot 2^{-q^{1-\alpha(q)}} \right) \\ &= 2^{-q^{1-\alpha(q)}} + o \left(2^{-q^{1-\alpha(q)}} \right). \end{aligned}$$

As a result, we have

$$R_q \asymp \alpha(q) + \tau(q) + 2^{-q^{1-\alpha(q)}} + o \left(\max \left\{ \alpha(q), \tau(q), 2^{-q^{1-\alpha(q)}} \right\} \right), \quad (3.63)$$

which completes the proof. \square

Proof of Theorem III.13. We are now in position to complete the proof of Theorem III.13, which consists of a combination of the results that have already been established in Section 3.5.

Proof. Recall the definition of ξ_p in (3.50) and that

$$\mathbb{P}(|\xi_p - 1| > \delta_p) = \text{I}(\delta_p) + \text{II}(\delta_p),$$

where $\text{I}(\delta_p)$ and $\text{II}(\delta_p)$ are defined as in (3.51). We shall show that both terms vanish.

Proposition III.27, along with (3.29), imply that $\text{I}(\delta_p) = \mathbb{P}(\xi_p > 1 + \delta_p) \rightarrow 0$, as $p \rightarrow \infty$. Observe that the term $\text{I}(\delta_p) = \mathbb{P}(\xi_p > 1 + \delta_p)$ vanishes, regardless of the dependence structure of the array \mathcal{E} . The dependence plays a key role in the rate of the term $\text{II}(\delta_p)$.

We now steer our focus towards term $\text{II}(\delta_p)$. The Markov inequality yields

$$\text{II}(\delta_p) = \mathbb{P}(\xi_p < 1 - \delta_p) \leq \frac{\mathbb{E}(\xi_p - 1)_-}{\delta_p}.$$

Since $\mathbb{E}(\xi_p - 1)_- \leq \mathbb{E}(\xi_p - 1)_+ + |\mathbb{E}(\xi_p - 1)|$, we have

$$\begin{aligned} \text{II}(\delta_p) &\leq \frac{1}{\delta_p} (\mathbb{E}(\xi_p - 1)_+ + |\mathbb{E}(\xi_p - 1)|) \\ &= \frac{1}{\delta_p} (\mathbb{E}(\xi_p - 1)_+ + [\mathbb{E}(\xi_p - 1)]_+ + [\mathbb{E}(\xi_p - 1)]_-) \\ &\leq \frac{1}{\delta_p} (2\mathbb{E}(\xi_p - 1)_+ + [\mathbb{E}(\xi_p - 1)]_-), \end{aligned} \tag{3.64}$$

where the last inequality follows from the fact that $[\mathbb{E}(\xi_p - 1)]_+ \leq \mathbb{E}(\xi_p - 1)_+$.

Proposition III.27 and (3.29) imply that the term $\delta_p^{-1}\mathbb{E}(\xi_p - 1)_+$ in (3.64) vanishes. Moreover, Proposition III.30 entails

$$[\mathbb{E}(\xi_p - 1)]_- = \max\{0, -\mathbb{E}(\xi_p - 1)\} \leq |R_p|.$$

Thus, the term $\text{II}(\delta_p)$ vanishes, provided that $R_p/\delta_p \rightarrow 0$. This follows, however, from Lemma III.32 and (3.29), since for $\alpha(p) \rightarrow 0$, we have

$$\frac{1}{\log(p)} \gg 2^{-p^{1-\alpha(p)}}, \quad \text{as } p \rightarrow \infty$$

and the proof is complete. □

Remark III.33. After submitting the paper this chapter is based on to *Extremes*, we became aware of the important work of *Tanguy* (2015). According to their paper, in the stationary case, the upper bound of Theorem III.13 above partially follows from their Theorem 3. However, our work is in the general setting of triangular arrays and does not require stationarity. The result in Theorem 5 of *Tanguy* (2015), could in principle, be used to derive bounds on rates of concentration of maxima for non-stationary arrays. This, however, requires verifying two technical conditions. Our approach, based on the UDD condition yields rates that can be explicitly related to the covariance structure of the array. The in-depth comparison of the two approaches merits an independent study beyond the scope of the present work.

Part II

Spectral Inference for Functional Spatial Data

CHAPTER IV

The Lag-Window Estimator

The spectral density function describes the second-order properties of a stationary stochastic process on \mathbb{R}^d . In this chapter, we are interested in the nonparametric estimation of the spectral density of a continuous-time stochastic process taking values in a separable Hilbert space. Our estimator is based on kernel smoothing and can be applied to a wide variety of spatial sampling schemes including those in which data are observed at irregular spatial locations. Thus, it finds immediate applications in Spatial Statistics, where irregularly sampled data naturally arise. The rates for the bias and variance of the estimator are obtained under general conditions in a mixed-domain asymptotic setting. Finally, with a view towards practical applications the asymptotic results are specialized to the case of discretely-sampled functional data in a reproducing kernel Hilbert space.

4.1 The spectral density estimation problem

Historically, the study of signals, such as electromagnetic or acoustic waves, in physics naturally led to the investigation of the spectral density. The current literature on the inference problem of the spectral density contains an abundance of well-established estimators and algorithms (see, e.g., *Hannan*, 1970; *Brillinger*, 2001; *Brockwell and Davis*, 2006; *Percival and Walden*, 2020, and the references therein). The most classical approach is based on the periodogram (*Schuster*, 1898), which is at the core of the majority of the procedures that are known today. However, alternative approaches that involve, for instance, the inversion of the empirical covariance (see, e.g., the review paper of *Robinson*, 1983) and wavelets (*Percival and Walden*, 2006; *Bardet and Bertrand*, 2010) have also been extensively considered.

The traditional statistical research on spectral density estimation considers scalar-valued processes. Modern scientific applications involve, however, high-dimensional or even function-type data, which are typically indexed by space and/or time. Recently, there has been a growing interest in functional time series in general, where data are observed at times $1, 2, \dots, T$; see *Hörmann and Kokoszka* (2012), *Panaretos and Tavakoli* (2013), *Horváth et al.* (2014), *Li et al.* (2020), *Zhu and Politis* (2020), to mention a few. In particular, *Panaretos and Tavakoli* (2013) and *Zhu and Politis* (2020) both address the inference of the spectral density of functional time series. *Panaretos and Tavakoli* (2013) considers the smoothed periodogram estimator where the notion of periodogram kernel is introduced for functional data taking values in $L^2[0, 1]$. *Zhu and Politis* (2020) considers the same estimator, but focuses on a particular type of kernel, called flat-top kernel, in performing nonparametric smoothing. A flat kernel is a higher-order kernel that annihilates polynomials up to a prescribed degree and therefore leads to better rate of the bias in nonparametric estimation (at the expense of potentially more stringent assumptions on the process).

This thesis studies the nonparametric estimation of the spectral density for a

continuous-time stationary process $X = \{X(t), t \in \mathbb{R}^d\}$ taking values in some Hilbert space \mathbb{H} . More information will be given in Section 4.2 regarding \mathbb{H} and the definition of second-order stationarity. One of the novelties of the thesis is the consideration of functional data $X(t)$ sampled at irregular spatial locations $t_1, \dots, t_n \in \mathbb{R}^d$ as opposed to at regular grid points, e.g., $t = 1, 2, \dots$ as in functional time series. In general, spatial data are not gridded data. An excellent example is provided by the Argo dataset which has recently become an important resource for oceanography and climate research (cf. *Roemmich et al.*, 2012) and has inspired new approaches in spatial statistics (see, e.g., *Kuusela and Stein*, 2018; *Yarger et al.*, 2022).

For spatial data observed at irregular locations, periodogram-based approaches do not easily generalize. We consider in this thesis a so-called lag-window estimator (cf. *Brockwell and Davis*, 2006; *Zhu and Politis*, 2020) based on estimating the covariance, which can accommodate rather general observational schemes. The performance of the estimator will be evaluated by asymptotic theory. In doing so, we will assume the framework of the so-called mixed-domain asymptotics, which means that the sampling locations become increasingly dense and the sampling region becomes increasingly large as the number of observations increases; see, e.g., *Hall and Patil* (1994); *Fazekas and Kukush* (2000); *Matsuda and Yajima* (2009); *Chang et al.* (2017); *Maitra and Bhattacharya* (2020). The rate bound of the mean squared error of our estimator will be developed for a rather general mixed-domain setting. However, when data are observed on a regular grid assuming a specific covariance model, the rate bound calculations can be made precise, paving the way for assessing the optimality of the estimator. In particular, we establish the minimax rate optimality of our estimator based on gridded data if the decay of the covariance function is bounded by a power law.

We now provide a summary of each of the sections below. In Section 4.2, we describe the general notion of second-order stationarity for a process taking values

in a complex Hilbert space \mathbb{H} . Despite the prevalence of multidimensional spatial data, this notion is understood much less well than the corresponding notion in the one-dimensional case. In particular, we will explain the subtlety of why the scalar field of \mathbb{H} must be taken as complex in order to conduct the spectral analysis of the process. We will also review Bochner’s Theorem which facilitates the definition of spectral density. Section 4.3 introduces the key assumptions and defines the lag-window estimator that is the main focus of this chapter. In Section 4.4, we establish upper bounds on the rate of decay of the bias and variance, and hence the mean squared error of the spectral density estimator under general conditions. These rates are made more precise in Section 4.5 for the setting of gridded data, where the grid size either stays fixed or shrinks to zero with sample size and we focus mainly on a class of covariance functions that are dominated by a power law. A class of covariance functions dominated by an exponential power law is also examined. In Section 4.6, we consider the issue of incomplete functional data in the reproducing kernel Hilbert space (RKHS) setting. Finally, Section 4.7 briefly summarizes the results in *Panaretos and Tavakoli (2013)* and *Zhu and Politis (2020)*, and provides some comparisons with the ones in this thesis.

Whenever feasible, we will provide an outlined proof immediately after stating a result. However, all the detailed proofs are included in Section 4.8.

4.2 Covariance and spectral density of a stationary process in a Hilbert space

Throughout this thesis, let \mathbb{H} be a separable Hilbert space over the field of complex numbers \mathbb{C} . Common examples of \mathbb{H} in functional-data applications include L^2 spaces of functions and RKHS’s (see for example *Bosq (2000)*; *Ferraty and Vieu (2006)*; *Horváth and Kokoszka (2012)* and *Ramsay and Silverman (2005)*). However, with

the exception of Section 4.6, no additional assumptions will be made on \mathbb{H} .

The inner product and norm of \mathbb{H} are denoted by $\langle \cdot, \cdot \rangle$ and $\| \cdot \|$, respectively. In a small number of instances, we will denote these by $\langle \cdot, \cdot \rangle_{\mathbb{H}}$ and $\| \cdot \|_{\mathbb{H}}$ for clarity. The main purpose of this section is to recall some fundamental results for the spectral analysis of stochastic processes $X = \{X(t), t \in \mathbb{R}^d\}$ taking values in \mathbb{H} .

4.2.1 Second-order stationary

We first address the notion of second-order stationarity or covariance stationarity for a process taking values in a complex Hilbert space. We begin by considering a zero-mean Gaussian process X with $\mathbb{H} = \mathbb{C}$. Let $\Re(X(t))$ and $\Im(X(t))$ denote the real and imaginary parts of $X(t)$, respectively. Recall that X is strictly stationary if and only if the two-dimensional real Gaussian process

$$Y(t) = (X_R(t), X_I(t))^T := (\Re(X(t)), \Im(X(t)))^T \in \mathbb{R}^2,$$

is second-order stationary, i.e., the covariance function $C_Y(t, s) := \mathbb{E}[Y(t)Y(s)^T]$ is a function of $t - s$. Let

$$C(t, s) = \mathbb{E}[X(t)\overline{X}(s)] \quad \text{and} \quad \check{C}(t, s) = \mathbb{E}[X(t)X(s)].$$

It follows that

$$\begin{aligned} C(t, s) &= \mathbb{E}[X_R(t)X_R(s)] + \mathbb{E}[X_I(t)X_I(s)] - \mathfrak{i}(\mathbb{E}[X_R(t)X_I(s)] - \mathbb{E}[X_I(t)X_R(s)]), \\ \check{C}(t, s) &= \mathbb{E}[X_R(t)X_R(s)] - \mathbb{E}[X_I(t)X_I(s)] + \mathfrak{i}(\mathbb{E}[X_R(t)X_I(s)] + \mathbb{E}[X_I(t)X_R(s)]). \end{aligned}$$

Observe that $\{C(t, s), \check{C}(t, s), t, s \in \mathbb{R}^d\}$ contains the same information as that in $\{C_Y(t, s), t, s \in \mathbb{R}^d\}$. In particular, Y is second-order stationary if and only if both $C(t, s)$ and $\check{C}(t, s)$ are functions of $t - s$. The functions, $C(t, s)$ and $\check{C}(t, s)$, are

commonly referred to as the covariance function and pseudo-covariance function, respectively, which are equal if and only if X is real valued. Going beyond the Gaussian setting, we shall take this as the definition of second-order stationarity for a general complex-valued process X with finite second moments, where the inference on the covariance of X can be conducted on C_Y or C and \check{C} combined. While the stationary covariance C is positive definite, which provides a basis for inference in the spectral domain, it is not the case for \check{C} . Thus, the spectral inference on X must be carried out on the real process Y unless X itself is real, in which case we can simply focus on C . The discussion above extends in a straightforward manner to the finite-dimensional case $\mathbb{H} = \mathbb{C}^p$ for any finite p , for which the outer-product is $x\bar{y}^\top$, $x, y \in \mathbb{C}^p$ (cf. *Hannan*, 1970; *Brillinger*, 2001; *Tsay*, 2013).

If \mathbb{H} is an infinite-dimensional Hilbert space over \mathbb{C} , then the cross-product (or outer-product) of $x, y \in \mathbb{H}$ is the linear operator defined as

$$[x \otimes y](z) = x \cdot \langle z, y \rangle, \quad z \in \mathbb{H}, \quad (4.1)$$

and, provided that $\mathbb{E}[\|X(t)\|^2] < \infty$ for all t , we can define the covariance operator of X as

$$C(t, s) := \mathbb{E}[X(t) \otimes X(s)]. \quad (4.2)$$

Note that $C(t, s)$ takes values in the space of trace-class operators \mathbb{T} and is well-defined in the sense of Bochner in the Banach space $(\mathbb{T}, \|\cdot\|_{\text{tr}})$. More information on \mathbb{T} will be given below in Section 4.2.2. However, the discussion on stationarity for the finite-dimensional case and especially the notion of pseudo-covariance requires modification since an immediate notion of “complex conjugate” does not exist. Following *Shen et al.* (2022), we fix a complete orthonormal system (CONS) $\{e_j\}$ of \mathbb{H} and refer to it as the *real* CONS. Then, for each $x \in \mathbb{H}$ such that $x = \sum_j \langle x, e_j \rangle e_j$, define the

complex conjugate $\text{conj}(x) \equiv \bar{x}$ as

$$\bar{x} := \sum_j \overline{\langle x, e_j \rangle} e_j.$$

Thus, $\text{conj} : \mathbb{H} \mapsto \mathbb{H}$ is an anti-linear operator, i.e.,

$$\text{conj}(\alpha x + \beta y) = \bar{\alpha} \text{conj}(x) + \bar{\beta} \text{conj}(y), \quad x, y \in \mathbb{H}, \quad \alpha, \beta \in \mathbb{C}.$$

Also, for $x \in \mathbb{H}$, define its real and imaginary parts:

$$\Re(x) := \frac{\bar{x} + x}{2}, \quad \Im(x) := \frac{x - \bar{x}}{2i}.$$

This construction allows us to view the complex Hilbert space \mathbb{H} as

$$\mathbb{H} = \mathbb{H}_{\mathbb{R}} + i\mathbb{H}_{\mathbb{R}}, \tag{4.3}$$

where $\mathbb{H}_{\mathbb{R}} := \{x \in \mathbb{H} : \Im(x) = 0\}$ is the real Hilbert space of real elements of \mathbb{H} (see, e.g., *Cerovecki and Hörmann, 2017*). Consequently, x will be called real if $x \in \mathbb{H}_{\mathbb{R}}$.

Define the pseudo-covariance operator for a second-order process $\{X(t)\}$ as

$$\check{C}(t, s) := \mathbb{E}[X(t) \otimes \bar{X}(s)]. \tag{4.4}$$

The definition of second-order stationarity for a process in \mathbb{H} can now be stated as follows.

Definition IV.1. A zero-mean stochastic process $X = \{X(t), t \in \mathbb{R}^d\}$ taking values in \mathbb{H} is said to be an L^2 - or second-order process if $\mathbb{E}[\|X(t)\|^2] < \infty$. The process X will be referred to as second-order stationary or covariance stationary if both $C(t, s)$ and $\check{C}(t, s)$ depend only on the lag $t - s$. In this case, we write $C(h) := C(t + h, t)$ and $\check{C}(h) := \check{C}(t + h, t)$, which are referred to as the stationary covariance operator

and stationary pseudo-covariance operator, respectively.

It is important to note that while the definition of $\check{C}(t, s)$ depends on the designated real basis, whether $\check{C}(t, s)$ is a function of the lag is basis independent; this can be seen using a change-of-basis formula (cf Section 4.8.3).

We end this section with the following two remarks.

Remark IV.2. As in the one-dimensional case, one can equivalently define stationarity in terms of the real process $Y(t) := (\Re(X(t)), \Im(X(t)))$ taking values in the product Hilbert space $\mathbb{H}_{\mathbb{R}} \times \mathbb{H}_{\mathbb{R}}$ over \mathbb{R} . It follows that X is second-order stationary if and only if Y is. For much of the rest of this thesis, we shall assume for simplicity that the process X is real (based on some CONS), i.e., it takes values in $\mathbb{H}_{\mathbb{R}} \subset \mathbb{H}$ (cf. (4.3)), in which case, $C(h) = \check{C}(h)$. This simplification does not lead to less generality since all the results apply to Y . Two exceptions are Section 4.2.2 and Section 6.1 where we present more general results by considering a complex X .

Remark IV.3. In view of the last remark, a careful reader might wonder why we choose to work with the framework of complex Hilbert space in the first place. An important reason for that is because the spectral density of a process X , real or complex, in \mathbb{H} will in general take values in \mathbb{T}_+ , the space of positive trace-class operators over the *complex* Hilbert space \mathbb{H} . To demonstrate the point, consider the following simple example. Let $\{Z(t), t \in \mathbb{R}\}$ be a real, scalar-valued zero-mean Gaussian process with auto-covariance $\gamma(t) = \mathbb{E}[Z(t)Z(0)]$. Let $a > 0$, and define $X_a(t) := (Z(t), Z(t+a))^\top$. Then, $X_a = \{X_a(t), t \in \mathbb{R}\}$ is a stationary process in \mathbb{R}^2 , with auto-covariance

$$C_a(t) := \mathbb{E}[X_a(t)X_a(0)^\top] = \begin{pmatrix} \gamma(t) & \gamma(t-a) \\ \gamma(t+a) & \gamma(t) \end{pmatrix}.$$

This shows that so long as $\gamma(t+a) \neq \gamma(t-a)$, for all t , i.e., the auto-covariance does not vanish on $(-a/2, a/2)$, we have that $C_a(t) \neq C_a(-t) \equiv C_a(t)^\top$, namely, the

process X_a is not time-reversible. Remark IV.5 below then shows that the spectral density cannot be real-valued. The simple example illustrates that a complex spectral density is a norm rather than an exception if $d \neq 1$.

4.2.2 Bochner's Theorem

This subsection discusses the notion of spectral density for a second-order stationary process X in \mathbb{H} . First, we briefly review some basic facts on trace-class operators. The reader is referred to the standard texts on linear operators (e.g., *Simon, 2015*) for details. Denote by \mathbb{T} the collection of trace-class operators on \mathbb{H} , namely, linear operators $\mathcal{T} : \mathbb{H} \rightarrow \mathbb{H}$, with finite *trace norm*:

$$\|\mathcal{T}\|_{\text{tr}} := \sum_{j=1}^{\infty} \langle (\mathcal{T}^* \mathcal{T})^{1/2} e_j, e_j \rangle < \infty, \quad (4.5)$$

where $\{e_j\}$ is an arbitrary CONS on \mathbb{H} , and \mathcal{T}^* denotes the adjoint operator of \mathcal{T} , i.e., defined by $\langle \mathcal{T}^* f, g \rangle = \langle f, \mathcal{T} g \rangle, f, g \in \mathbb{H}$. The trace norm does not depend on the choice of the CONS, and the space \mathbb{T} equipped with the trace norm is a Banach space. By the definitions of the outer product (4.1) and trace norm (4.5), we have $\|X(t) \otimes X(s)\|_{\text{tr}} = \|X(t)\| \|X(s)\|$. The fact that X is second order then implies that

$$\mathbb{E}[\|X(t) \otimes X(s)\|_{\text{tr}}] \leq \sqrt{\mathbb{E}[\|X(t)\|^2] \mathbb{E}[\|X(s)\|^2]} < \infty.$$

Consequently, the covariance operator $C(t, s)$ in (4.2) is well defined in \mathbb{T} in the sense of Bochner; see, e.g., Lemma S.2.2 of *Shen et al. (2022)*.

Recall that \mathcal{T} is self-adjoint if $\mathcal{T} = \mathcal{T}^*$. Also \mathcal{T} is positive definite (or just positive), denoted $\mathcal{T} \geq 0$, if \mathcal{T} is self-adjoint and $\langle f, \mathcal{T} f \rangle \geq 0$, for all $f \in \mathbb{H}$. The class of positive, trace-class operators will be denoted by \mathbb{T}_+ .

The classical Bochner's Theorem (cf. *Bochner, 1948; Khintchine, 1934*), which characterizes positive-definite functions, has provided a fundamental tool for con-

structuring useful models for stationary processes. Below we state an extension of that classical result for our infinite-dimensional setting. To do so, we need the notion of integration with respect to a \mathbb{T}_+ -valued measure which we now briefly describe. Let $\mathcal{B}(\mathbb{R}^d)$ denote the σ -field of Borel sets in \mathbb{R}^d . We say that $\mu : \mathcal{B}(\mathbb{R}^d) \mapsto \mathbb{T}_+$ is a \mathbb{T}_+ -valued measure on $\mathcal{B}(\mathbb{R}^d)$ if μ is σ -additive. Note that, *a fortiori*, $\mu(\emptyset) = 0$ and μ is finite in the sense that $0 \leq \mu(B) \leq \mu(\mathbb{R}^d) \in \mathbb{T}_+$, $B \in \mathcal{B}(\mathbb{R}^d)$, where for $\mathcal{T}_1, \mathcal{T}_2 \in \mathbb{T}_+$, $\mathcal{T}_1 \leq \mathcal{T}_2$ means that $\mathcal{T}_2 - \mathcal{T}_1 \in \mathbb{T}_+$. Integration of a \mathbb{C} -valued measurable function on \mathbb{R}^d with respect to such μ can be defined along the line of Lebesgue integral (see, e.g., *Shen et al.*, 2022, for details).

Theorem IV.4. *Let X be a second-order stationary process taking values in \mathbb{H} , and let $C(h)$, $h \in \mathbb{R}^d$, be its \mathbb{T} -valued stationary covariance function defined in Definition IV.1. Assume that C is continuous at 0 in trace norm. Then, there exists a unique \mathbb{T}_+ -valued measure ν such that*

$$C(h) = \int_{\mathbb{R}^d} e^{-ih^\top \theta} \nu(d\theta), \quad h \in \mathbb{R}^d.$$

In particular, we have that $\|\nu(\mathbb{R}^d)\|_{\text{tr}} = \text{trace}(\nu(\mathbb{R}^d)) < \infty$.

If, moreover, $\int_{h \in \mathbb{R}^d} \|C(h)\|_{\text{tr}} dh < \infty$, then the measure ν has a density with respect to the Lebesgue measure given by

$$f(\theta) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{ih^\top \theta} C(h) dh, \quad \theta \in \mathbb{R}^d, \quad (4.6)$$

where the last integral is understood in the sense of Bochner.

The density function f in (4.6) is referred to as the spectral density of the stationary process. The detailed proof of Theorem IV.4 can be found in *Shen et al.* (2022), where the role of separability and complex scalar field are made clear.

The following is a follow-up remark to Remark IV.3.

Remark IV.5. Theorem IV.4 holds for a general second-order stationary process X in \mathbb{H} . Let us consider an interesting property of the spectral density if the process is real (defined according to some fixed CONS). To do that, define the conjugate $\overline{\mathcal{A}}$ of an operator $\mathcal{A} : \mathbb{H} \rightarrow \mathbb{H}$ by $\overline{\mathcal{A}} : x \mapsto \overline{\mathcal{A}(\overline{x})}$, $x \in \mathbb{H}$; accordingly, define

$$\Re(\mathcal{A}) := \frac{\overline{\mathcal{A}} + \mathcal{A}}{2}, \quad \Im(\mathcal{A}) := \frac{\mathcal{A} - \overline{\mathcal{A}}}{2i}.$$

Thus, \mathcal{A} will be called real if $\Im(\mathcal{A}) = 0$. Suppose now X is real (cf. Remark (IV.2)). By the simple fact that $x \otimes y$ is real if both x and y are real, we have $C(h) = \overline{C(h)}$. It then follows from (4.6) that the *time-reversed* process $Y = \{X(-t), t \in \mathbb{R}^d\}$ has the spectral density

$$f_Y(\theta) = \overline{f_X(\theta)}, \quad \theta \in \mathbb{R}^d.$$

The uniqueness of the spectral density entails that X and Y have the same auto-covariance if and only if $f_Y(\theta) = f_X(\theta) = \overline{f_X(\theta)}$, that is, $f_X(\theta)$ is a real operator, for all $\theta \in \mathbb{R}^d$. This is a special property that is automatically true only when \mathbb{H} is one-dimensional. For further discussions, see Section 4.3 of *Shen et al. (2022)*.

4.3 Spectral density estimation based on irregularly sampled data

Our inference problem focuses on a second-order real process $X = \{X(t), t \in \mathbb{R}^d\}$ taking values in \mathbb{H} . Following Definition IV.1, we define the stationary covariance operator C and assume that the following holds.

Assumption C. Let $C = \{C(h), h \in \mathbb{R}^d\}$ be the \mathbb{T} -valued stationary covariance operator of the second-order stationary real process $X = \{X(t), t \in \mathbb{R}^d\}$ taking values in \mathbb{H} . Assume that

(a) $\int_{h \in \mathbb{R}^d} \|C(h)\|_{\text{tr}} dh < \infty$, and

(b) $C(h)$ is L^1 - γ -Hölder in the following sense:

$$\int_{x \in \mathbb{R}^d} \left(\sup_{y: \|x-y\| \leq \delta} \|C(y) - C(x)\|_{\text{tr}} \right) dx \leq \|C\|_{\gamma} \cdot \delta^{\gamma}, \quad (4.7)$$

for some $0 < \gamma \leq 1$ and some (and hence all) $\delta > 0$, where $\|C\|_{\gamma} < \infty$ is a fixed constant.

Property (a) in Assumption C guarantees the existence of the spectral density f given by (4.6). Property (b) will be needed to compute the bias of our estimator which is based on discretely observed data. It can be seen that Condition (b) holds with $\gamma = 1$ if C has an integrable and smoothly varying derivative.

We next introduce our sampling framework. As mentioned above, we adopt the mixed-domain asymptotics framework, which means that both the domain and the density of the data increase with sample size. Assume that the process $\{X(t), t \in \mathbb{R}^d\}$ is observed at distinct locations $t_{n,i}, i = 1, \dots, n$. Let $\mathbb{T}_n := \{t_{n,i}\}_{i=1}^n$, and T_n denote the closed convex hull of \mathbb{T}_n . We refer to T_n as the *sampling region*, which contains points where $X(t)$ could potentially be observed. However, as seen in our proofs, other contiguous regions may also be used for T_n . For our purpose, it is convenient to view T_n as a tessellation comprising disjoint cells, $V(t_{n,i})$, that are “centered” at the $t_{n,i}$:

$$T_n = \bigcup_{i=1}^n V(t_{n,i}), \quad \text{where } t_{n,i} \in V(t_{n,i}) \quad \text{and} \quad |V(t_{n,i}) \cap V(t_{n,j})| = 0, \quad i \neq j.$$

Here and elsewhere, $|A|$ denotes the Lebesgue measure of a measurable set $A \subset \mathbb{R}^d$. Denote $\mathbb{V} = \{V(t_{n,i}), i = 1, \dots, n\}$. The Voronoi tessellation (Voronoi, 1908) is a natural example of such tessellation and can also be efficiently constructed (Yan *et al.*, 2013). While our results hold for a wide class of tessellations, to fix ideas we will adopt the Voronoi tessellation in the sequel.

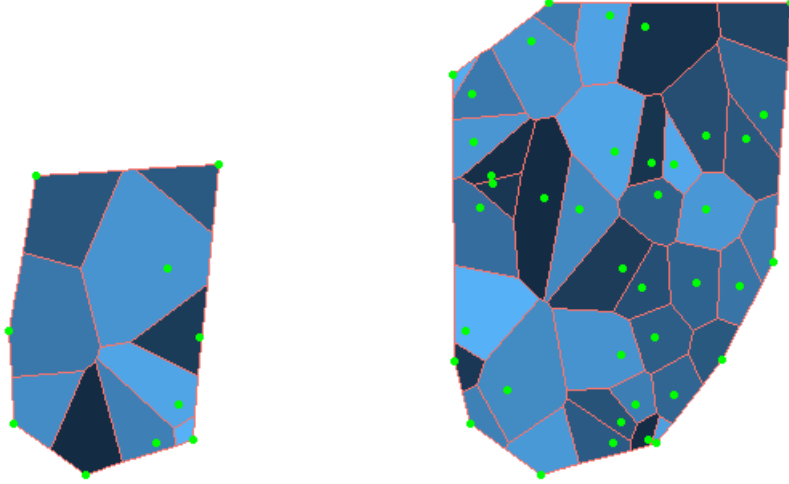


Figure 4.1: Example of the Voronoi tessellation. Every cell includes a single sampling time point as a representative. A cell is defined as the part of the sampling domain containing all the time points that are closer to its representatives compared to all other representatives, based on the Euclidean metric on \mathbb{R}^d . The figure also shows the sampling framework imposed by Assumption S(a). Indeed, the sampling domain inflates, while the sample time points are denser and denser as the sample size grows; equivalently δ_n becomes smaller with the increase of the sample size.

Define the diameter of the Voronoi tessellation as

$$\delta_n := \text{diam}_{T_n}(\{t_{n,i}\}) = \max_{i=1,\dots,n} \sup_{t \in V(t_{n,i})} \|t - t_{n,i}\|_2, \quad (4.8)$$

where $\|\cdot\|_2$ denotes the Euclidean norm in \mathbb{R}^d . The parameter δ_n can be thought of as a measure of the maximal size of the tessellation cells, and can be equivalently written as

$$\delta_n = \sup_{t \in T_n} \min_{i=1,\dots,n} \|t - t_{n,i}\|_2.$$

Throughout, we will assume the following rather general sampling framework.

Assumption S.

- (a) The sequence δ_n defined in (4.8) tends to zero as $n \rightarrow \infty$. Moreover, $|T_n| \rightarrow \infty$ as $n \rightarrow \infty$.

(b) The sample design is such that

$$\frac{T_n}{|T_n|^{1/d}} \rightarrow T, \text{ as } n \rightarrow \infty, \quad (4.9)$$

holds in probability, in the Hausdorff metric, for some fixed bounded convex set T with non-empty interior.

The condition (a) above describes the mixed-domain asymptotics framework alluded to earlier. Relation (4.9) in (b) essentially imposes a regularity condition on the boundary points of \mathbb{T}_n ; for instance, if $\mathbb{T}_n = \{1, 2, \dots, n\}^d$ then $T = [0, 1]^d$.

The definition of our proposed estimator involves a kernel function, which satisfies the following standard conditions.

Assumption K. The kernel K is a continuous function from \mathbb{R}^d to \mathbb{R}_+ satisfying

(a) The support $S_K := \{t \in \mathbb{R}^d : K(t) > 0\}$ of K is a bounded set containing 0;

(b) $\|K\|_\infty := \sup_{u \in S_K} K(u) = K(0) = 1$;

(c) K is differentiable in an ϵ -neighborhood of 0 for some fixed $\epsilon > 0$, with

$$\|\nabla K\|_\infty^{(\epsilon)} := \sup_{\|u\|_2 < \epsilon} \|\nabla K(u)\|_2 < \infty,$$

where ∇ stands for the gradient operator.

The estimator. In this thesis, we focus on the following non-parametric estimator of the spectral density $f(\theta)$:

$$\begin{aligned} \hat{f}_n(\theta) = & \frac{1}{(2\pi)^d} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} e^{i(t-s)^\top \theta} \frac{X(t) \otimes X(s)}{|T_n \cap (T_n - (t-s))|} \\ & \cdot K\left(\frac{t-s}{\Delta_n}\right) \cdot |V(t)| \cdot |V(s)|, \end{aligned} \quad (4.10)$$

where $\Delta_n > 0$ is a bandwidth parameter, the purpose of which is providing weighted averaging over observations that are at most $\Delta_n \cdot |S_K|$ apart. The choice of Δ_n that will lead to satisfactory estimation results depends on both δ_n and $|T_n|$.

The estimator in (4.10) can be applied to the general setting of functional data sampled irregularly over space and time, which is frequently encountered in applications (see, e.g., *Yarger et al.*, 2022). In the special case where \mathbb{T}_n is a regular grid, which includes the time-series setting, the terms $V(t)$ are constant for any $t \in \mathbb{T}_n$ and hence $|V(s)|$ and $|V(t)|$ can be factored out of the summation in \hat{f}_n (see Section 4.5). In this case, the estimator in (4.10) is related to the so-called lag window estimator in time-series analysis; see *Robinson* (1983), *Zhu and Politis* (2020) and the discussions in Section 4.7.1 below.

To gain some insight into the definition (4.10), consider the idealized setting where the full sample path of $\{X(t), t \in T_n\}$ is available. In view of (4.6), one would naturally use the estimator

$$g_n(\theta) = \frac{1}{(2\pi)^d} \int_{t \in T_n} \int_{s \in T_n} e^{i(t-s)^\top \theta} \frac{X(t) \otimes X(s)}{|T_n \cap (T_n - (t-s))|} K\left(\frac{t-s}{\Delta_n}\right) dt ds. \quad (4.11)$$

Since the full sample path is not available in practice one must consider approximations such as $\hat{f}_n(\theta)$, which can be viewed as a Riemann sum for the integral defining $g_n(\theta)$. The function $g_n(\theta)$ motivates the definition of $\hat{f}_n(\theta)$ and in fact arises in the proofs of the asymptotic theory.

We end the section with the following remarks.

Remark IV.6. In our data scheme, we assume a fixed design where the observation points $t_{n,i}$ are nonrandom. Our results can be modified in a straightforward manner to include the case of a random design that is independently generated from the process $\{X(t)\}$. In this case, the definition of the estimator in (4.10) needs to be modified slightly to incorporate the probability densities of the sample design in place of the

volume elements (cf., for example, *Matsuda and Yajima, 2009*).

Remark IV.7. The normalization $|T_n \cap (T_n - (t - s))|$ in (4.10) and (4.11) might seem unusual at first glance, whereas the simpler normalization by $|T_n|$ would seem more natural. It turns out that the use of the latter normalization leads to a bias with a higher order in the spatial context $d \geq 2$. Similar phenomenon arises for periodogram-based estimators in time series when data are observed over a regular lattice (*Guyon, 1982*).

Remark IV.8. The estimator \hat{f}_n is defined assuming that we have fully observed functional data $X(t), t \in \mathbb{T}_n$. If \mathbb{H} is infinite dimensional, then the functional data $X(t)$ can never be observed in its entirety. In that case, we need to approximate $X(t) \otimes X(s)$ in some manner based on what is actually observed for the functional data, which may affect the performance of the estimator. We will discuss this point in more detail in Section 4.6.

4.4 Asymptotic properties

We start our investigation of $\hat{f}_n(\theta)$ defined in Section 3 by first developing the asymptotic bounds for its bias and variance. This will yield results on the consistency and rate of convergence of the estimator. Although $f(\theta)$ and $\hat{f}_n(\theta)$ are trace-class operators on \mathbb{H} , in order to facilitate the variance calculation, it is more natural to work with the Hilbert-Schmidt (HS) norm. Let \mathbb{X} denote the class of Hilbert-Schmidt operators on \mathbb{H} . The Hilbert-Schmidt inner product of the linear operators $\mathcal{A}, \mathcal{B} \in \mathbb{X}$ is defined as

$$\langle \mathcal{A}, \mathcal{B} \rangle_{\text{HS}} = \text{trace}(\mathcal{A}^* \mathcal{B})$$

and $\|\mathcal{A}\|_{\text{HS}} := \sqrt{\|\mathcal{A}^* \mathcal{A}\|_{\text{tr}}}$ (see, e.g., *Simon, 2015*).

It is straightforward to establish the following bias-variance decomposition

$$\begin{aligned} \mathbb{E} \left\| \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}}^2 &= \mathbb{E} \left\| \hat{f}_n(\theta) - \mathbb{E} \hat{f}_n(\theta) \right\|_{\text{HS}}^2 + \left\| \mathbb{E} \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}}^2 \\ &=: \text{Var} \left(\hat{f}_n(\theta) \right) + \text{Bias} \left(\hat{f}_n(\theta) \right)^2. \end{aligned} \quad (4.12)$$

4.4.1 Asymptotic bias

In this subsection, we evaluate the rate of the bias of $\hat{f}_n(\theta)$ for large n . We start with a general bound, which is made more informative in the sequel. Consistent with (4.12), the bounds in the following Theorem IV.9 are stated in the Hilbert-Schmidt norm. However, we note that the result remains valid if the stronger trace norm is used throughout.

Theorem IV.9. *Let Assumptions C, K, and S hold. Choose $\Delta_n \rightarrow \infty$ such that*

$$\Delta_n \cdot S_K \subset T_n - T_n \quad , \text{ for all } n$$

where $A - B := \{a - b : a \in A, b \in B\}$ for sets $A, B \subset \mathbb{R}^d$. Then, for any bounded set Θ

$$\sup_{\theta \in \Theta} \left\| \mathbb{E} \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}} = \mathcal{O} \left(\delta_n^\gamma + B_1(\Delta_n) + B_2(\Delta_n) \right), \quad (4.13)$$

where

$$\begin{aligned} B_1(\Delta_n) &:= \left\| \int_{h \in \Delta_n \cdot S_K} e^{ih^\top \theta} C(h) \left(1 - K \left(\frac{h}{\Delta_n} \right) \right) dh \right\|_{\text{HS}}, \\ B_2(\Delta_n) &:= \left\| \int_{h \notin \Delta_n \cdot S_K} e^{ih^\top \theta} C(h) dh \right\|_{\text{HS}}. \end{aligned} \quad (4.14)$$

Proof (Outline). The complete proof of Theorem IV.9 is given in Section 4.8.1.1. Here, we provide a brief outline. Let $g_n(\theta)$ be defined by (4.11). By the triangle

inequality,

$$\left\| \mathbb{E} \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}} \leq \left\| \mathbb{E} \hat{f}_n(\theta) - \mathbb{E} g_n(\theta) \right\|_{\text{HS}} + \left\| \mathbb{E} g_n(\theta) - f(\theta) \right\|_{\text{HS}}.$$

It is immediate from the representation (4.6) for f and the inclusion $\Delta_n \cdot S_K \subset T_n - T_n$, that

$$\left\| \mathbb{E} g_n(\theta) - f(\theta) \right\|_{\text{HS}} \leq B_1(\Delta_n) + B_2(\Delta_n).$$

To complete the proof one needs to show that

$$\left\| \mathbb{E} \hat{f}_n(\theta) - \mathbb{E} g_n(\theta) \right\|_{\text{HS}} = \mathcal{O}(\delta_n^\gamma). \quad (4.15)$$

To evaluate $\left\| \mathbb{E} \hat{f}_n(\theta) - \mathbb{E} g_n(\theta) \right\|_{\text{HS}}$, let

$$h_n(t, s; \theta) := e^{i(t-s)^\top \theta} \frac{X(t) \otimes X(s)}{|T_n \cap (T_n - (t-s))|} K \left(\frac{t-s}{\Delta_n} \right),$$

and write

$$\begin{aligned} & g_n(\theta) - \hat{f}_n(\theta) \\ &= \frac{1}{(2\pi)^d} \sum_{w \in \mathbb{T}_n} \sum_{v \in \mathbb{T}_n} \int_{t \in V(w)} \int_{s \in V(v)} (h_n(t, s; \theta) - h_n(w, v; \theta)) \mathbb{1}_{(t \in V(w), s \in V(v))} dt ds. \end{aligned} \quad (4.16)$$

This implies that

$$\begin{aligned} & \left\| \mathbb{E} g_n(\theta) - \mathbb{E} \hat{f}_n(\theta) \right\|_{\text{HS}} \\ & \leq \frac{1}{(2\pi)^d} \sum_{w \in \mathbb{T}_n} \sum_{v \in \mathbb{T}_n} \int_{t \in V(w)} \int_{s \in V(v)} \left\| \mathbb{E} h_n(t, s; \theta) - \mathbb{E} h_n(w, v; \theta) \right\|_{\text{HS}} dt ds. \end{aligned}$$

Then, using the regularity conditions on K and C , routine but technical analysis shows that the last sum is of order $\mathcal{O}(\delta_n^\gamma)$. This yields (4.15) and completes the proof of (4.13). \square

Several remarks are in order.

Remark IV.10. Theorem IV.9 provides a general bound on the bias. Under the assumptions of the theorem, the bias vanishes as $n \rightarrow \infty$. We briefly discuss the terms δ_n^γ and $B_1(\Delta_n) + B_2(\Delta_n)$ which arise for different reasons.

1. As can be seen from the above sketch of the proof, the terms $B_1(\Delta_n)$ and $B_2(\Delta_n)$ in (4.13) control the bias of the idealized estimator $g_n(\theta)$ based on the idealized data. A more specific but crude bound of $B_1(\Delta_n)$ and $B_2(\Delta_n)$ is the following:

$$\begin{aligned} B_1(\Delta_n) &\leq \int_{|h| \leq \epsilon \Delta_n} \|C(h)\|_{\text{tr}} \left| 1 - K\left(\frac{h}{\Delta_n}\right) \right| dh + \int_{|h| > \epsilon \Delta_n} \|C(h)\|_{\text{tr}} dh \\ &\leq \|\nabla K\|_\infty^{(\epsilon)} \epsilon \int \|C(h)\|_{\text{tr}} dh + \int_{|h| > \epsilon \Delta_n} \|C(h)\|_{\text{tr}} dh. \end{aligned} \quad (4.17)$$

The first term on the rhs depends only on the kernel, whereas the second term, which dominates $B_2(\Delta)$ for any small $\epsilon < 1$, depends on the decay rate of $\|C(h)\|_{\text{tr}}$. Thus, the rate of $B_1(\Delta_n) + B_2(\Delta_n)$ is bounded by

$$\inf_{\epsilon} (\epsilon \vee \psi(\epsilon \Delta_n)) \quad \text{where} \quad \psi(u) := \int_{|h| > u} \|C(h)\|_{\text{tr}} dh.$$

More explicit bounds can be obtained by imposing specific assumptions on the behavior of $\psi(u)$ for large u , as will be demonstrated in Section 4.5.

2. In view of (4.15), the term $\mathcal{O}(\delta_n^\gamma)$ controls the bias due to discretization, which arises from sampling the process at the discrete set $\mathbb{T}_n \subset T_n$. In settings such as time series where the data are sampled on a regular grid, this term will be eliminated from the bias (cf. Theorem IV.16).

4.4.2 Asymptotic variance

In view of the form of $\hat{f}_n(\theta)$, a “fourth-moment” condition of X is needed to evaluate the variance of \hat{f}_n .

Recall the definition of cumulant for random variables: For real-valued random variables $Y_j, j = 1, \dots, k$,

$$\text{cum}(Y_1, \dots, Y_k) := \sum_{\nu=(\nu_1, \dots, \nu_q)} (-1)^{q-1} (q-1)! \prod_{l=1}^q \mathbb{E} \left(\prod_{j \in \nu_l} Y_j \right), \quad (4.18)$$

provided all the expectations on the rhs are well defined, where the sum is taken over all unordered partitions ν of $\{1, \dots, k\}$.

We now define a notion of fourth-order cumulant for complex Hilbert space valued random variables Y_1, Y_2, Y_3, Y_4 with mean zero.

Definition IV.11. Let Y_1, Y_2, Y_3, Y_4 take values in \mathbb{H} . Then the fourth-order cumulant is defined as

$$\begin{aligned} \text{cum}(Y_1, Y_2, Y_3, Y_4) &:= \mathbb{E} \langle Y_1 \otimes Y_2, Y_3 \otimes Y_4 \rangle_{\text{HS}} - \langle \mathbb{E}(Y_1 \otimes Y_2), \mathbb{E}(Y_3 \otimes Y_4) \rangle_{\text{HS}} \\ &\quad - \mathbb{E} \langle Y_1, Y_3 \rangle \cdot \mathbb{E} \langle Y_4, Y_2 \rangle - \langle \mathbb{E}(Y_1 \otimes \bar{Y}_4), \mathbb{E}(Y_3 \otimes \bar{Y}_2) \rangle_{\text{HS}}, \end{aligned}$$

whenever the expression is well defined and finite.

Note that $\text{cum}(Y_1, Y_2, Y_3, Y_4)$ is well defined and finite if $\mathbb{E} \|Y_i\|^4 < \infty$ for each i (cf. Proposition VI.10). It is easy to check that this definition reduces to (4.18) with $k = 4$ if $\mathbb{H} = \mathbb{R}$.

Some properties immediately follow from Proposition VI.10. First,

$$\langle Y_1 \otimes Y_2, Y_3 \otimes Y_4 \rangle_{\text{HS}} = \langle Y_1, Y_3 \rangle \langle Y_4, Y_2 \rangle, \quad (4.19)$$

and hence we can express the fourth-order cumulant as

$$\begin{aligned} \text{cum}(Y_1, Y_2, Y_3, Y_4) &= \text{Cov}(\langle Y_1, Y_3 \rangle, \langle Y_2, Y_4 \rangle) - \langle \mathbb{E}(Y_1 \otimes Y_2), \mathbb{E}(Y_3 \otimes Y_4) \rangle_{\text{HS}} \\ &\quad - \langle \mathbb{E}(Y_1 \otimes \bar{Y}_4), \mathbb{E}(Y_3 \otimes \bar{Y}_2) \rangle_{\text{HS}}. \end{aligned}$$

Next, for any CONS $\{e_j\}$ of \mathbb{H} , and with $Y_{i,j} := \langle Y_i, e_j \rangle$,

$$\text{cum}(Y_1, Y_2, Y_3, Y_4) = \sum_i \sum_j \text{cum}(Y_{1,i}, \overline{Y_{2,j}}, \overline{Y_{3,i}}, Y_{4,j}). \quad (4.20)$$

Observe that, unless \mathbb{H} is one dimensional, $\text{cum}(Y_1, Y_2, Y_3, Y_4)$ generally depends on the order in which the Y_i 's appear in the arguments.

For the real process X that we consider in our inference problem, assuming $\mathbb{E}\|X(t)\|^4 < \infty$ for all t , we have

$$\begin{aligned} & \text{cum}(X(t), X(s), X(w), X(v)) \\ & := \mathbb{E} \langle X(t) \otimes X(s), X(w) \otimes X(v) \rangle_{\text{HS}} - \langle C(t, s), C(w, v) \rangle_{\text{HS}} \\ & \quad - \mathbb{E} \langle X(t), X(w) \rangle_{\mathbb{H}} \cdot \mathbb{E} \langle X(v), X(s) \rangle_{\mathbb{H}} - \langle C(t, v), C(w, s) \rangle_{\text{HS}}. \end{aligned} \quad (4.21)$$

The following assumption will be needed to evaluate the variance of $\hat{f}_n(\theta)$.

Assumption V. Suppose that the process X is real and such that:

- (a) $\mathbb{E}\|X(t)\|^4 < \infty$ for all t ;
- (b) $\text{cum}(X(t + \tau), X(s + \tau), X(w + \tau), X(v + \tau)) = \text{cum}(X(t), X(s), X(w), X(v))$
for all t, s, w, v, τ ;
- (c) for some small enough $\delta > 0$,

$$\sup_{w \in \mathbb{R}^d} \int_{u \in \mathbb{R}^d} \int_{v \in \mathbb{R}^d} \sup_{\substack{\lambda_i \in B(0, \delta) \\ i=1,2,3}} |\text{cum}(X(\lambda_1 + u), X(\lambda_2 + v), X(\lambda_3 + w), X(0))| dv du < \infty.$$

The following are a few remarks regarding Assumption V.

Remark IV.12. 1. Part (b) of this assumption can be thought of as “fourth-order cumulant stationarity”, which is implied by but more general than strict stationarity. For a second-order stationary process X , by (4.19) and (4.21), part

(b) amounts to

$$\begin{aligned} & \mathbb{E}(\langle X(t), X(s) \rangle \langle X(w), X(v) \rangle) \\ &= \mathbb{E}(\langle X(t + \tau), X(s + \tau) \rangle \langle X(w + \tau), X(v + \tau) \rangle) \quad \text{for all } t, s, w, v, \tau. \end{aligned}$$

2. Part (c) of Assumption V is a variant of the cumulant condition “ $C(0, 4)$ ” of *Panaretos and Tavakoli* (2013) for functional time series (see Remark VI.11 for more details).
3. For Gaussian processes, by (4.20), the fourth-order cumulants vanish and hence Assumption V is trivially satisfied under stationarity.

The variance bound of $\hat{f}_n(\theta)$ is provided by the following result.

Theorem IV.13. *Let $X = \{X(t), t \in \mathbb{R}^d\}$ be a real process taking values in \mathbb{H} , which has mean zero and is second-order stationary. Suppose that Assumptions C, K, S, and V hold. Also, assume that Δ_n satisfies*

$$\Delta_n \cdot S_K \subset T_n - T_n \quad \text{for all } n, \text{ and } \Delta_n^d / |T_n| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Then

$$\sup_{\theta \in \Theta} \mathbb{E} \left(\left\| \hat{f}_n(\theta) - \mathbb{E} \hat{f}_n(\theta) \right\|_{\text{HS}}^2 \right) = \mathcal{O} \left(\frac{\Delta_n^d}{|T_n|} \right), \quad \text{as } n \rightarrow \infty. \quad (4.22)$$

Proof (Outline). The complete proof of Theorem IV.13 is presented in Section 4.8.1.2.

Here, we sketch the main steps. First,

$$\begin{aligned} & \mathbb{E} \left\| \hat{f}_n(\theta) - \mathbb{E} \hat{f}_n(\theta) \right\|_{\text{HS}}^2 \\ &= \frac{1}{(2\pi)^{2d}} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} \sum_{\substack{h \in [\Delta \cdot S_K] \cap (\mathbb{T}_n - t) \\ h' \in [\Delta \cdot S_K] \cap (\mathbb{T}_n - s)}} e^{i(h-h')^\top \theta} K \left(\frac{h}{\Delta} \right) K \left(\frac{h'}{\Delta} \right) \\ & \quad \cdot |V(t+h)| \cdot |V(t)| \cdot |V(s+h')| \cdot |V(s)| \\ & \quad \cdot \frac{\text{Cov}(X(t+h) \otimes X(t), X(s+h') \otimes X(s))}{|T \cap (T-h)| |T \cap (T-h')|} \end{aligned} \quad (4.23)$$

where

$$\begin{aligned} & \text{Cov}(X(t+h) \otimes X(t), X(s+h') \otimes X(s)) \\ & := \mathbb{E} \langle X(t+h) \otimes X(t) - C(h), X(s+h') \otimes X(s) - C(h') \rangle_{\text{HS}}. \end{aligned}$$

By (4.21),

$$\begin{aligned} & \text{Cov}(X(t+h) \otimes X(t), X(s+h') \otimes X(s)) \\ & = \mathbb{E} \langle X(t+h), X(s+h') \rangle_{\mathbb{H}} \cdot \mathbb{E} \langle X(s), X(t) \rangle_{\mathbb{H}} \\ & \quad + \langle C(t-s+h), C(s+h'-t) \rangle_{\text{HS}} \\ & \quad + \text{cum}(X(t+h), X(t), X(s+h'), X(s)). \end{aligned} \tag{4.24}$$

In our detailed proof (presented in Section 4.8), the components of the variance involving the cumulants will be evaluated using Assumption V, while the other two terms are handled using the integrability condition of the covariance of Assumption C. □

4.4.3 Rates of convergence

The results in Sections 4.4.1 and 4.4.2 allow us to obtain bounds on the rate of consistency of the estimator $\hat{f}_n(\theta)$. The following result is immediate from the bias-variance decomposition (4.12).

Theorem IV.14. *Let the assumptions of Theorem IV.13 hold. Then, for any bounded $\Theta \subset \mathbb{R}^d$, we have*

$$\sup_{\theta \in \Theta} \left(\mathbb{E} \left\| \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}}^2 \right)^{1/2} = \mathcal{O} \left(\delta_n^\gamma + B_1(\Delta_n) + B_2(\Delta_n) + \sqrt{\frac{\Delta_n^d}{|T_n|}} \right), \tag{4.25}$$

as $n \rightarrow \infty$, where $B_1(\Delta_n)$ and $B_2(\Delta_n)$ are as defined in Theorem IV.9.

Theorem IV.14 provides general bounds on the rate of consistency of the estimator

$\hat{f}_n(\theta)$. More explicit rates and their minimax optimality can be obtained under further conditions on the dependence structure of the process. We conclude with several comments.

- Remark IV.15.*
1. The bound on the rate of consistency for the estimator $\hat{f}_n(\theta)$ in (4.25) depends on the quantities δ_n, Δ_n and $|T_n|$. Among them, δ_n and T_n consist of artifacts of the sample design, while Δ_n is a tuning parameter which can be controlled. Under the assumptions of the theorem, any choice of the bandwidth with $\Delta_n \rightarrow \infty$ and $\Delta_n^d/|T_n| \rightarrow 0$ yields a consistent estimator $\hat{f}_n(\theta)$.
 2. As discussed in Remark IV.10, $B_1(\Delta_n)$ and $B_2(\Delta_n)$ in the rate mostly reflect the tail-decay of the covariance. They are not present in the bound on the variance (4.22), where smaller values of Δ_n lead to smaller variances of the estimator. The bound in (4.25) reflects a natural bias-variance trade-off, where the optimal bound is obtained by picking Δ_n that balances the contribution of the bias and the variance.
 3. Establishing rate-optimal choices of Δ_n depends on both the sampling design and the stochastic process under consideration. Indeed, the choice of Δ_n optimizing the bounds in (4.25) depends both on δ_n and T_n , as well as on the covariance structure of the process. In Section 4.5, we will compute $B_1(\Delta_n)$ and $B_2(\Delta_n)$ and consider the choice of Δ_n for certain classes of covariance structures.

4.5 Data observed on a regular grid

In this section, we focus on data observed on a regular grid, namely, the sampling set is

$$\mathbb{T}_n = \bigtimes_{\ell=1}^d \{\delta_n, \dots, n_\ell \delta_n\}, \quad (4.26)$$

where δ_n is the grid size. In our asymptotic theory in the next two subsections, we let $n_\ell \rightarrow \infty, \ell = 1, \dots, d$, and consider both cases of fixed δ_n and $\delta_n \rightarrow 0$.

In this setting, for convenience, we slightly modify our general estimator $\hat{f}_n(\theta)$ defined in (4.10) and consider

$$\hat{f}_n(\theta) := \frac{\delta_n^{2d}}{(2\pi)^d} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} e^{i(t-s)^\top \theta} \frac{X(t) \otimes X(s)}{|T_n \cap (T_n - (t-s))|} K\left(\frac{t-s}{\Delta_n}\right). \quad (4.27)$$

Under the condition

$$\sum_{k \in \mathbb{Z}^d} \|C(k\delta_n)\|_{\text{tr}} < \infty,$$

we have

$$C(k\delta_n) = \int_{\theta \in [-\pi/\delta_n, \pi/\delta_n]^d} e^{-ik^\top \theta \delta_n} f(\theta; \delta_n) d\theta, \quad k \in \mathbb{Z}^d, \quad (4.28)$$

where

$$f(\theta; \delta_n) := \frac{\delta_n^d}{(2\pi)^d} \sum_{k \in \mathbb{Z}^d} e^{ik^\top \theta \delta_n} C(k\delta_n), \quad \theta \in [-\pi/\delta_n, \pi/\delta_n]^d, \quad (4.29)$$

which is a positive trace-class operator since $\{C(k\delta_n), k \in \mathbb{Z}^d\}$ is positive definite. The

proof of (4.28) follows easily using the fact that the complex exponentials

$$\phi_k(\theta) := e^{ik^\top \theta \delta_n} (\delta_n/2\pi)^{d/2} \mathbb{1}_{[-\pi/\delta_n, \pi/\delta_n]^d}(\theta), \quad k \in \mathbb{Z}^d$$

constitute a CONS of $L^2([-\pi/\delta_n, \pi/\delta_n]^d)$. By Theorem IV.4, (4.28) also holds if $f(\theta; \delta_n)$ is replaced by the folded spectral density

$$f_{\text{fold}}(\theta) := \mathbb{1}_{[-\pi/\delta_n, \pi/\delta_n]^d}(\theta) \sum_{k \in \mathbb{Z}^d} f(\theta + 2\pi k/\delta_n).$$

Utilizing again the fact that $\{\phi_k(\theta), k \in \mathbb{Z}^d\}$ is a CONS of $L^2([-\pi/\delta_n, \pi/\delta_n]^d)$, $f(\theta; \delta_n)$ is equal to the folded spectral density. Thus, the knowledge of $C(k\delta_n), k \in \mathbb{Z}^d$, only allows us to identify the folded spectral density. In fact, this is reflected by our estimator \hat{f}_n since

$$\hat{f}_n(\theta + 2\pi k/\delta_n) = \hat{f}_n(\theta), \quad \theta \in [-\pi/\delta_n, \pi/\delta_n]^d,$$

for any vector $k \in \mathbb{Z}^d$.

For the purpose of estimating the folded spectral density, we define the following analogs of Assumptions C and V.

Assumption C'. The trace-norm of the operator auto-covariance is summable:

$$\sup_n \left\{ \delta_n^d \sum_{k \in \mathbb{Z}^d} \|C(\delta_n k)\|_{\text{tr}} \right\} < \infty.$$

Assumption V'. The process $\{X(\delta_n t), t \in \mathbb{Z}^d\}$ satisfies

- (a) $\sup_n \mathbb{E}\|X(\delta_n t)\|^4 < \infty$ for all t ;

(b) for all t, s, w, v, τ ,

$$\begin{aligned} & \text{cum}(X(\delta_n(t + \tau)), X(\delta_n(s + \tau)), X(\delta_n(w + \tau)), X(\delta_n(v + \tau))) \\ &= \text{cum}(X(\delta_n t), X(\delta_n s), X(\delta_n w), X(\delta_n v)); \end{aligned}$$

$$(c) \sup_n \left\{ \delta_n^{2d} \sup_{w \in \mathbb{Z}^d} \sum_{u \in \mathbb{Z}^d} \sum_{v \in \mathbb{Z}^d} |\text{cum}(X(\delta_n u), X(\delta_n v), X(\delta_n w), X(0))| \right\} < \infty.$$

Comparing with Assumptions C and V, the modifications in Assumptions C' and V' are motivated by the fact that discrete approximations of integrals is no longer an issue if our target of inference is the folded spectral density. We will apply these conditions in the time series context in Section 4.5.1.

As before, Assumption V' holds trivially for Gaussian processes since the 4th order cumulants vanish. More generally, it holds for a wide class of short-memory \mathbb{H} -valued processes (see Example IV.41 in Section 4.8).

Note that our assumptions on the cumulants in Assumption V' are different from but related to the assumption based cumulant kernels employed on page 571 in *Panaretos and Tavakoli (2013)*. For more details, see Remark VI.11.

4.5.1 The case of fixed grid

Consider the case where δ_n in (4.26) is fixed. Without loss of generality, let $\delta_n \equiv 1$. The discussion in the previous section shows that we can only identify the folded spectral density on $[-\pi, \pi]^d$. As such, without loss of generality, focus on a stochastic processes $\{X(t)\}$ indexed by $t \in \mathbb{Z}^d$. This framework includes time series (for $d = 1$), and, more generally, many random fields observed at discrete locations/times. The spectral density f in this case is defined by (4.29). With the normalization $|T_n \cap (T_n - (t - s))|$ replaced by $|T_n|$, $\hat{f}_n(\theta)$ recovers the classical lag-window estimator (cf. *Robinson, 1983*).

The following result on the rate of $\hat{f}_n(\theta)$ is the analog to Theorem IV.14 for the gridded setting.

Theorem IV.16. *Let $\{X(t), t \in \mathbb{Z}^d\}$ be a real process taking values in \mathbb{H} , which has mean zero and is second-order stationary. Suppose that Assumption K holds, Assumptions C' and V' hold with $\delta_n \equiv 1$, and*

$$\Delta_n \cdot S_K \cap \mathbb{Z}^d \subset \mathbb{T}_n - \mathbb{T}_n \text{ for all } n, \text{ and } \Delta_n^d/|\mathbb{T}_n| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Then,

$$\sup_{\theta \in [-\pi, \pi]^d} \left\| \mathbb{E} \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}} \leq B_1(\Delta_n) + B_2(\Delta_n) \quad (4.30)$$

$$\sup_{\theta \in [-\pi, \pi]^d} \mathbb{E} \left\| \hat{f}_n(\theta) - \mathbb{E}[\hat{f}_n(\theta)] \right\|_{\text{HS}}^2 = \mathcal{O} \left(\frac{\Delta_n^d}{|\mathbb{T}_n|} \right), \quad (4.31)$$

where

$$B_1(\Delta_n) := \left\| \sum_{k \in (\Delta_n \cdot S_K) \cap \mathbb{Z}^d} e^{ik^\top \theta} C(k) \left(1 - K \left(\frac{k}{\Delta_n} \right) \right) \right\|_{\text{HS}},$$

$$B_2(\Delta_n) := \left\| \sum_{k \in \mathbb{Z}^d \setminus (\Delta_n \cdot S_K)} e^{ik^\top \theta} C(k) \right\|_{\text{HS}}.$$

Consequently,

$$\sup_{\theta \in [-\pi, \pi]^d} \left(\mathbb{E} \left\| \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}}^2 \right)^{1/2} = \mathcal{O} \left(B_1(\Delta_n) + B_2(\Delta_n) + \frac{\Delta_n^{d/2}}{\sqrt{|\mathbb{T}_n|}} \right), \text{ as } n \rightarrow \infty.$$

In this result, the derivation of the bias bound (4.30) is more straightforward than that for the general case since it does not involve a Riemann approximation as in (4.16). Here, the first term on the rhs of (4.13) is no longer present and the other two terms, $B_1(\Delta_n)$ and $B_2(\Delta_n)$, are similar to (4.14), with sums replacing integrals. The derivation of the variance bound (4.31) is also simpler than that of (4.22), where

the term involving δ_n is no longer needed. For completeness, the variance bound is established in Proposition IV.35 of Section 4.8.

The bias bounds $B_1(\Delta_n), B_2(\Delta_n)$ in Theorem IV.16 hold for a very general class of models. However, more precise expressions of the bias can be obtained for specific models. We illustrate this next by considering a class of covariances that decay like the power law. The power-law decay class, $\mathcal{P}_D(\beta, L)$, for the discrete-time processes is defined as

$$\mathcal{P}_D(\beta, L) := \left\{ f(\theta) = (2\pi)^{-d} \sum_{k \in \mathbb{Z}^d} C(k) e^{i\theta^\top k} : \sum_{k \in \mathbb{Z}^d} \|C(k)\|_{\text{tr}} (1 + \|k\|_2^\beta) \leq L \right\}, \quad (4.32)$$

for $\beta, L > 0$. By the theory of the Fourier transform, larger values of β in this condition correspond to a higher order of smoothness of the spectral density at $\theta = 0$; see, e.g., *Bingham et al.* (1989).

Below we establish an explicit upper bound on the rate of $\hat{f}_n(\theta)$ for this class by focusing on the bias terms $B_1(\Delta_n)$ and $B_2(\Delta_n)$ of Theorem IV.16. First, we introduce an additional smoothness condition on the kernel K that is compatible with the covariance model in $\mathcal{P}_D(\beta, L)$. Let $\boldsymbol{\alpha} = (\alpha_1, \dots, \alpha_d) \in \mathbb{Z}_+^d$ and define the partial derivative

$$\partial^{\boldsymbol{\alpha}} K(h) = \frac{\partial^{\boldsymbol{\alpha}} K(h)}{\partial h_1^{\alpha_1} \dots \partial h_d^{\alpha_d}}.$$

Then, for an integer $\lambda \geq 1$, define the condition

$$\begin{aligned} \partial^{\boldsymbol{\alpha}} K(0) &= 0 \text{ for all } \boldsymbol{\alpha} \text{ with } 1 \leq |\boldsymbol{\alpha}| := \sum_{i=1}^d \alpha_i \leq \lambda, \text{ and} \\ \sup_h |\partial^{\boldsymbol{\alpha}} K(h)| &< \infty \text{ for all } \boldsymbol{\alpha} \text{ with } |\boldsymbol{\alpha}| = \lambda + 1. \end{aligned} \quad (4.33)$$

Theorem IV.17. *Let all the conditions of Theorem IV.16 hold. Moreover, assume that the spectral density f belongs to $\mathcal{P}_D(\beta, L)$ for some $\beta > 0$ and $L > 0$, and that (4.33) holds for some integer $\lambda > 0 \vee (\beta - 1)$. Then, the following is a uniform bound*

on the rate of the bias of $\hat{f}_n(\theta)$:

$$\sup_{f \in \mathcal{P}_D(\beta, L)} \sup_{\theta \in [-\pi, \pi]^d} \left\| \mathbb{E} \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}} = \mathcal{O}(\Delta_n^{-\beta}), \quad \text{as } n \rightarrow \infty. \quad (4.34)$$

Combining this with the variance bound $\Delta_n^d/|\mathbb{T}_n|$ in (4.31) and choosing bandwidth $\Delta_n = |\mathbb{T}_n|^{\frac{1}{2\beta+d}}$, the following uniform bound on the mean squared error of $\hat{f}_n(\theta)$ holds:

$$\sup_{f \in \mathcal{P}_D(\beta, L)} \sup_{\theta \in [-\pi, \pi]^d} \left(\mathbb{E} \left\| \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}}^2 \right)^{1/2} = \mathcal{O} \left(|\mathbb{T}_n|^{-\frac{\beta}{2\beta+d}} \right). \quad (4.35)$$

The proof of this result is given in Section 4.8.2. An important motivation for singling out the class $\mathcal{P}_D(\beta, L)$ is that it covers a broad range of realistic covariance models whose tail-decay can be controlled by the parameter β . Moreover, in Section 5.1 we establish a minimax lower bound for this class which matches the upper bound on the rate in (4.35). In this sense, our estimator with the oracle choice of the bandwidth is minimax rate-optimal.

4.5.2 Dense gridded data

We now turn to the setting (4.26) where we assume $\delta_n \rightarrow 0$. In doing so, we continue to focus on the estimator $\hat{f}_n(\theta)$ in (4.27) for gridded data. However, unlike the $\delta_n = 1$ case, here we are in a position to estimate the full spectral density as opposed to the folded spectral density. As in the previous subsection, we also study a similar power law decay class. However, some slight modifications are necessary. The continuous time power law decay class $\mathcal{P}_C(\beta, L)$ where $\beta, L > 0$, contains spectral densities for the continuous-time process, defined by

$$\mathcal{P}_C(\beta, L) := \left\{ f(\theta) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{ix^\top \theta} C(x) dx : \int_{\mathbb{R}^d} (1 + \|x\|_2^\beta) \|C(x)\|_{\text{tr}} dx \leq L \right\}. \quad (4.36)$$

Mimicking the approach in Section 4.5.1, the following result can be stated for this class.

Theorem IV.18. *Let all the assumptions of Theorem IV.14 hold and assume that the spectral density $f(\theta)$ belongs in $\mathcal{P}_C(\beta, L)$ for some $\beta, L > 0$. Suppose that (4.33) holds for some integer $\lambda > 0 \vee (\beta - 1)$. Then, for every $f \in \mathcal{P}_C(\beta, L)$ and bounded $\Theta \subset \mathbb{R}^d$, the rate of the bias is*

$$\sup_{\theta \in \Theta} \left\| \mathbb{E} \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}} = \mathcal{O} \left(\delta_n^\gamma + \Delta_n^{-\beta} \right), \quad \text{as } n \rightarrow \infty.$$

In conjunction with Theorem IV.14, with the rate-optimal choice of $\Delta_n := |T_n|^{1/(2\beta+d)}$, we obtain the overall rate bound:

$$\sup_{\theta \in \Theta} \left(\mathbb{E} \left\| \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}}^2 \right)^{1/2} = \mathcal{O} \left(\delta_n^\gamma \vee |T_n|^{-\beta/(2\beta+d)} \right). \quad (4.37)$$

The proof of Theorem IV.18 is given in Section 4.8.2.

Remark IV.19. Observe that, in contrast to Theorem IV.17, the rate bounds in Theorem IV.18 are not uniform over the class $\mathcal{P}_C(\beta, L)$. This is mainly because the constant $\|C\|_\gamma$ in (4.7) of Assumption C (b) cannot be bounded uniformly in $\mathcal{P}_C(\beta, L)$, since the tail behavior of $\|C(x)\|_{\text{tr}}$ does not regulate the smoothness of $C(x)$. At this point, we do not know whether there is an adaptive estimator for which the rate could be shown to be uniform.

Remark IV.20. To interpret the bound on the rate in (4.37), suppose, for example, that $\delta_n := n^{-\alpha}$ for some $\alpha \in (0, 1)$, which controls the sampling frequency relative to the sample size. The greater the value of α , the finer the grid. Also, assume that the grid is square with $n_\ell = n$, for all ℓ , so that $|T_n| \sim (n\delta)^d$. Let

$$\alpha_{\beta, \gamma} = \left(1 + \left(\frac{2}{d} + \frac{1}{\beta} \right) \gamma \right)^{-1} \quad (4.38)$$

and consider the following two regimes:

- (*fine sampling*) When $\alpha \geq \alpha_{\beta,\gamma}$, then $\delta_n^\gamma = \mathcal{O}(|T_n|^{-\beta/(2\beta+d)})$, and the rate bound in (4.37) is

$$\mathcal{O}\left((n\delta_n)^{-\beta d/(2\beta+d)}\right) = \mathcal{O}\left(n^{-\beta d(1-\alpha)/(2\beta+d)}\right).$$

- (*coarse sampling*) When $0 < \alpha < \alpha_{\beta,\gamma}$, then $|T_n|^{-\beta/(2\beta+d)} = \mathcal{O}(\delta_n^\gamma)$ and the rate bound becomes

$$\mathcal{O}(\delta_n^\gamma) = \mathcal{O}(n^{-\alpha\gamma/2}).$$

In the fine-sampling regime, the rate is the same as the minimax lower bound established in Theorem V.3 below. By (4.38), a larger γ (i.e., a smoother C) leads to a wider range of sampling rates under which the minimax rate can be achieved by $\hat{f}_n(\theta)$. Similarly, a larger d or larger β (i.e., faster tail decay of C) leads to a narrower range of sampling rates in order to achieve the minimax rate.

4.5.3 One more covariance class

In this subsection, we present explicit rate bounds on the bias for one more class of covariances. We focus on a class of covariances that decay like the exponential power law. For $\eta, L > 0$, the exponential power-law decay class, $\mathcal{EP}_D(\eta, L)$, for the discrete-time processes is defined as

$$\mathcal{EP}_D(\eta, L) := \left\{ f(\theta) = (2\pi)^{-d} \sum_{k \in \mathbb{Z}^d} C(k) e^{i\theta^\top k} : \|C(k)\|_{\text{tr}} \leq L \cdot e^{-\|k\|_2^\eta} \right\}. \quad (4.39)$$

In the following theorem we establish an explicit upper bound on the rate of $\hat{f}_n(\theta)$ for this class. As in the case of $\mathcal{P}_D(\beta, L)$, the proof concentrates on the bias terms $B_1(\Delta_n)$ and $B_2(\Delta_n)$ of Theorem IV.16. We need an extra smoothness condition,

different from (4.40) this time. In this case, we have the condition

$$K(h) = \mathcal{O}\left(e^{-1/\|h\|_2}\right), \quad (4.40)$$

that we will use in this subsection.

Theorem IV.21. *Let all the conditions of Theorem IV.16 hold. Moreover, assume that the spectral density f belongs to $\mathcal{EP}_D(\eta, L)$ for some $\eta > 0$ and $L > 0$, and that (4.40) holds. Then, the following is a uniform bound on the rate of the bias of $\hat{f}_n(\theta)$:*

$$\sup_{f \in \mathcal{P}_D(\beta, L)} \sup_{\theta \in [-\pi, \pi]^d} \left\| \mathbb{E} \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}} = \mathcal{O}\left(\Delta_n^{\frac{|d-\eta|_+}{\eta+1}} e^{-\Delta_n^{\frac{\eta}{\eta+1}}}\right), \quad \text{as } n \rightarrow \infty. \quad (4.41)$$

Combining this with the variance bound $\Delta_n^d/|\mathbb{T}_n|$ in (4.31) and choosing bandwidth $\Delta_n = (\log |\mathbb{T}_n|)^{\frac{\eta+1}{\eta}}$, the following uniform bound on the mean squared error of $\hat{f}_n(\theta)$ holds:

$$\sup_{f \in \mathcal{P}_D(\beta, L)} \sup_{\theta \in [-\pi, \pi]^d} \left(\mathbb{E} \left\| \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}}^2 \right)^{1/2} = \mathcal{O}\left(\frac{(\log |T_n|)^{\frac{d(\eta+1)}{2\eta}}}{\sqrt{|T_n|}}\right). \quad (4.42)$$

We present the proof of this result in Section 4.8.2.

Continuing along the lines of Section 4.5.2, we focus on the setting (4.26) where we assume $\delta_n \rightarrow 0$. The continuous time exponential power law decay class $\mathcal{EP}_C(\eta, L)$ where $\eta, L > 0$, contains spectral densities for the continuous-time process, defined by

$$\mathcal{EP}_C(\eta, L) := \left\{ f(\theta) = (2\pi)^{-d} \int_{\mathbb{R}^d} e^{ix^\top \theta} C(x) dx : \|C(h)\|_{\text{tr}} \leq L \cdot e^{-\|h\|_2^\eta} \right\}. \quad (4.43)$$

Similarly to $\mathcal{EP}_D(\eta, L)$, the following result can be stated for this class.

Theorem IV.22. *Let all the assumptions of Theorem IV.14 hold and assume that the spectral density $f(\theta)$ belongs in $\mathcal{EP}_C(\eta, L)$ for some $\eta, L > 0$. Suppose also that*

(4.40) holds. Then, for every $f \in \mathcal{EP}_C(\beta, L)$ and bounded $\Theta \subset \mathbb{R}^d$, the rate of the bias is

$$\sup_{\theta \in \Theta} \left\| \mathbb{E} \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}} = \mathcal{O} \left(\delta_n^\gamma + \Delta_n^{\frac{|d-\eta|_+}{\eta+1}} e^{-\Delta_n^{\frac{\eta}{\eta+1}}} \right), \quad \text{as } n \rightarrow \infty.$$

In conjunction with Theorem IV.14, and choosing $\Delta_n := (\log |\mathbb{T}_n|)^{\frac{\eta+1}{\eta}}$, we obtain the overall rate bound:

$$\sup_{\theta \in \Theta} \left(\mathbb{E} \left\| \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}}^2 \right)^{1/2} = \mathcal{O} \left(\delta_n^\gamma \vee \frac{(\log |T_n|)^{\frac{d(\eta+1)}{2\eta}}}{\sqrt{|T_n|}} \right). \quad (4.44)$$

The proof of this theorem is almost the same as the one of Theorem IV.21 and will be skipped.

Remark IV.23. Like in the power-law decay class, the rate bounds in Theorem IV.22 are not uniform over the class $\mathcal{EP}_C(\eta, L)$, as opposed to Theorem IV.21. The reasoning is the same here (cf Remark IV.19 for details).

Remark IV.24. Kernel Choice. Observe that in Proposition IV.21 we require the kernel to satisfy (4.40). By careful examination of the proof of Theorem IV.21, if we instead use kernels satisfying (4.33), one can see that the order of the rate would be slowed down, becoming

$$\sup_{\theta \in \Theta} \left\| \mathbb{E} g_n(\theta) - f(\theta) \right\|_{\text{HS}} = \mathcal{O} \left(\frac{1}{\Delta_n^{\lambda+1}} + \Delta_n^{d-\eta} e^{-(m_K \Delta_n)^\eta} \right).$$

This emphasizes the importance of the right choice of the kernel with regard to the covariance structure. The consistency rates are directly affected by the smoothness/differentiability of the kernel at zero. As a result, we can see that there is not a single choice of kernels that fits every model. As expected, unfortunate choices of kernels could slow down the consistency rates substantially.

4.6 An RKHS formulation based on discretely-observed functional data

In this section, we specialize the obtained results for an abstract Hilbert space to the case where \mathbb{H} is a space of functions. In real-data applications complete functions are not available and instead each of the functional data $X(t_i)$ is observed on a finite set of points. A natural space for this setting may be when \mathbb{H} is a RKHS. Unlike the more commonly considered space $L^2[a, b]$, an RKHS \mathbb{H} allows us to view \mathbb{H} -valued random elements as bona fide functions, since the point-evaluation functionals are well-defined and continuous. This enables a seamless interface between the theory that we have developed up to this point and applications based on discretely observed data. The literature on RKHS is extremely rich. For a quick overview on the role of RKHS in functional data analysis, the reader is referred to *Hsing and Eubank (2015)*.

Let \mathbb{H} be an RKHS containing functions on a compact set E , where the kernel $R(\cdot, \cdot)$ is continuous on $E \times E$. The reproducing property states that

$$g(u) = \langle g, R(u, \cdot) \rangle_{\mathbb{H}}, \quad u \in E.$$

Now, let $\{X(t), t \in \mathbb{R}^d\}$ be a stationary \mathbb{H} -valued process with covariance function C and spectral density f . Then, it can be viewed as a bivariate stochastic process $\{X(u, t) := \langle X(t), R(u, \cdot) \rangle_{\mathbb{H}}, u \in E, t \in \mathbb{R}^d\}$. We have

$$\begin{aligned} \text{Cov}(X(u, t+h), X(v, t)) &= \langle C(h)R(u, \cdot), R(v, \cdot) \rangle_{\mathbb{H}} \\ &= \int_{\mathbb{R}^d} e^{-ih^\top \theta} f_{u,v}(\theta) d\theta, \end{aligned} \tag{4.45}$$

where

$$f_{u,v}(\theta) = \langle f(\theta)R(u, \cdot), R(v, \cdot) \rangle_{\mathbb{H}}.$$

In view of (4.45), it may be convenient to refer to $f_{u,v}(\theta)$ as a spectral density. However, there is no guarantee that it is nonnegative for $u \neq v$. By the Cauchy-Schwartz inequality, our estimation rates on the operator $f(\theta)$ translate immediately to $f_{u,v}(\theta)$ for all u, v .

Assume that the process is observed on a common discrete set of points $D_n = \{u_{n,j}, j = 1, \dots, m_n\}$ for all $t \in \mathbb{T}_n$. To relate the partially observed functional data to complete functional data in \mathbb{H} , a possible approach is the following. Assume that the matrix

$$\mathbf{R}_n := \{R(u_{n,i}, u_{n,j})\}_{i,j=1}^{m_n} \quad (4.46)$$

is invertible for each n . Let \mathbb{H}_n be the subspace of \mathbb{H} spanned by $\{R(u, \cdot), u \in D_n\}$ and Π_n is the projection operator onto \mathbb{H}_n . Then, for any $g \in \mathbb{H}$,

$$\tilde{g} := \Pi_n g$$

interpolates g at the points in D_n and is in fact the minimum norm interpolant of g on D_n ; see *Wahba* (1990) or Proposition IV.36.

The covariance of the stationary process $\{\tilde{X}(t)\}$ is $\tilde{C}(h) := \Pi_n C(h) \Pi_n$. First note that

$$\|\tilde{C}(h)\|_{\text{tr}} \leq \|C(h)\|_{\text{tr}}.$$

This follows from Lemma IV.37 (i), since $\langle \tilde{C}(h), \mathcal{W} \rangle_{\text{HS}} = \langle C(h), \tilde{\mathcal{W}} \rangle_{\text{HS}}$, where $\tilde{\mathcal{W}} = \Pi_n \mathcal{W} \Pi_n$ is unitary for every unitary \mathcal{W} . Thus, the condition $\int \|C(h)\|_{\text{tr}} dh < \infty$ ensures that the spectral density \tilde{f} of $\{\tilde{X}(t)\}$ is well defined, and satisfies

$$\tilde{f}(\theta) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{ih^\top \theta} \tilde{C}(h) dh.$$

Following the approach in (4.10) based on the data $\tilde{X}(t_i)$, define

$$\tilde{f}_n(\theta) = \Pi_n \hat{f}_n(\theta) \Pi_n.$$

Consider the estimation of \tilde{f} by \tilde{f}_n . To keep the presentation simple we focus on the Gaussian case. The following result follows readily from Theorem IV.14.

Theorem IV.25. *Let the process $\{X(t), t \in \mathbb{R}^d\}$ be a zero-mean stationary Gaussian process taking values in \mathbb{H} . Suppose that Assumptions C, K, and S hold. If, additionally we have*

$$\Delta_n \cdot S_K \subset T_n - T_n \text{ for all } n.$$

Then, for any bounded set Θ ,

$$\sup_{\theta \in \Theta} \left(\mathbb{E} \left\| \tilde{f}_n(\theta) - \tilde{f}(\theta) \right\|_{\text{HS}}^2 \right)^{1/2} = \mathcal{O} \left(\delta_n^\gamma + B_1(\Delta_n) + B_2(\Delta_n) + \sqrt{\frac{\Delta_n^d}{|T_n|}} \right), \quad (4.47)$$

as $n \rightarrow \infty$.

Note that, in (4.47), we bounded $\tilde{B}_1(\Delta_n), \tilde{B}_2(\Delta_n)$, the counterparts of $B_1(\Delta_n), B_2(\Delta_n)$ where $C(h)$ therein is replaced by $\tilde{C}(h)$, by $B_1(\Delta_n), B_2(\Delta_n)$, respectively. This is achieved using the simple fact that $\|\mathcal{T}_1 \mathcal{T}_2\|_{\text{HS}} \leq \|\mathcal{T}_1\| \|\mathcal{T}_2\|_{\text{HS}}$ where $\|\mathcal{T}_1\|$ stands for the operator norm of \mathcal{T}_1 . In view of Theorem IV.25, to find the rate of $\mathbb{E} \|\tilde{f}_n(\theta) - \tilde{f}(\theta)\|_{\text{HS}}^2$, it is sufficient to consider the bias $\|\tilde{f}(\theta) - f(\theta)\|_{\text{HS}}$, which must be evaluated case by case, depending on the type of a RKHS being considered. Below, we consider an example that leads to a specific rate.

Consider the Sobolev space $\mathbb{H} = W_1[0, 1]$ which consists of functions on the interval $[0, 1]$ of the form $c + \int_0^1 (t \wedge u) h(u) du$, $c \in \mathbb{R}$ and h integrable (cf. Wahba, 1990). The inner-product in this space is $\langle f, g \rangle_{\mathbb{H}} := f(0)g(0) + \int_0^1 f'(t)g'(t)dt$, yielding the

norm

$$\|g\|_{\mathbb{H}}^2 = g(0)^2 + \int_0^1 (g'(t))^2 dt.$$

In context, we can state the following result for $\mathbb{E}\|\tilde{f}_n(\theta) - f(\theta)\|_{\text{HS}}^2$.

Theorem IV.26. *Let the positive trace-class operator $f(\theta)$ have the eigen decomposition:*

$$f(\theta) = \sum_{j=1}^{\infty} \nu_j \phi_j \otimes \phi_j,$$

where the eigenvalues ν_j are summable (since $f(\theta) \in \mathbb{T}_+$). Assume that, for each j , the derivative ϕ_j' is Lipschitz continuous with $|\phi_j'(s) - \phi_j'(t)| \leq C_j |s - t|$ for some finite constant C_j where $\sum_{j=1}^{\infty} C_j \nu_j^2 < \infty$. Also, assume that the sampling design is $u_{n,i} = i/m_n, 0 \leq i \leq m_n$. Then,

$$\|\tilde{f}(\theta) - f(\theta)\|_{\text{HS}} = \mathcal{O}(m_n^{-1/2}).$$

The proof of Theorem IV.26 is given in Section 4.8.4.

4.7 Related work and discussions

In this section we highlight the approaches in *Panaretos and Tavakoli (2013)* and *Zhu and Politis (2020)* focusing on the time-series setting and we explain how they relate to our approach.

4.7.1 Relation to flat-top kernel estimators

The flat-top kernel estimators have been advocated in the works of *Politis (2011)*; *Zhu and Politis (2020)*, among others. According to Relation (15) of *Zhu and Politis (2020)* the alternate estimator proposed in Section 3.1 therein takes the form

$$\frac{1}{2\pi} \sum_{|u| < T} \lambda(B_T u) \hat{r}_u(\tau, \sigma) e^{-i\omega u},$$

where

$$\hat{r}_u(\tau, \sigma) = \frac{1}{T} \sum_{0 \leq t, t+u \leq T-1} X_{t+u}(\tau) X_t(\sigma) \quad \text{and} \quad \lambda(s) = \int_{-\infty}^{\infty} \Lambda(x) e^{-isx} dx$$

for some $\Lambda(x)$. In the time-series setting with $d = 1$, an asymptotically equivalent adaptation of our estimator in (4.27) is given by:

$$\hat{f}_T(\theta) = \frac{1}{2\pi} \cdot \frac{1}{T} \sum_{|u| < T} K\left(\frac{u}{\Delta_T}\right) \cdot e^{-iu\theta} \sum_{0 \leq t, t+u \leq T-1} X_{t+u} \otimes X_t. \quad (4.48)$$

See Remark IV.7. Thus, the two estimators are essentially the same, with ω corresponds to θ , B_T to $1/\Delta_T$, and λ to K . *Zhu and Politis (2020)* focuses on p -times differentiable flat-top kernels λ with $\lambda(t) = 1$, for all $\|t\| \leq \epsilon$, for some $\epsilon > 0$, where p is adapted to the tail decay of the covariance function. Such kernels reduce the bias of the kernel spectral density estimator in essentially the same way as do the kernels K satisfying (4.33) in the present thesis. One can get a rough idea about that by the crude calculations in (4.17).

Moreover, in Section 5 of *Zhu and Politis (2020)*, an effective data-dependent choice of the bandwidth parameter B_T is developed. The authors base their selection on the functional version of correlogram/cross-correlogram. Using this quantity, an empirical rule is proposed for the choice of B_T . In practice, we recommend using flat-top kernels and a similar methodology for the selection of $\Delta_T = 1/B_T$. The thorough investigation of the data-driven, adaptive choice of Δ_T in our setting of irregularly sampled data, however, merits further theoretical and methodological investigation.

4.7.2 Periodogram-based estimators for functional time series

The seminal work of *Panaretos and Tavakoli (2013)* considers function-valued time series, taking values in $(L^2[0, 1], \mathbb{R})$. They develop comprehensive theory and methodology for inference of the spectral density operator extending the classic periodogram-

based approach to the functional time series setting. The proposed estimator therein is:

$$f_{\omega}^{(T)}(\tau, \sigma) = \frac{2\pi}{T} \sum_{s=1}^{T-1} W^{(T)}\left(\omega - \frac{2\pi s}{T}\right) p_{2\pi s/T}^{(T)}(\tau, \sigma), \quad (4.49)$$

where

$$W^{(T)}(x) = \sum_{j \in \mathbb{Z}} \frac{1}{B_T} W\left(\frac{x + 2\pi j}{B_T}\right),$$

with W being a taper weight function of bounded support. Here,

$$p_{\omega}^{(T)}(\tau, \sigma) = \tilde{X}_{\omega}^{(T)}(\tau) \tilde{X}_{-\omega}^{(T)}(\sigma)$$

is the periodogram, where

$$\tilde{X}_{\omega}^{(T)} = \frac{1}{\sqrt{2\pi T}} \sum_{t=0}^{T-1} X_t(\tau) e^{-i\omega t},$$

is the discrete Fourier transform (DFT). This is referred to as the smoothed periodogram estimator (cf. *Robinson*, 1983).

The asymptotic properties of these periodogram-based estimators are studied using the following general cumulant-based assumptions:

Condition $C(\ell, k)$. For each $j = 1, \dots, k-1$,

$$\sum_{t_1, \dots, t_{k-1} = -\infty}^{\infty} (1 + |t_j|^{\ell}) \|\text{cum}(X_{t_1}, \dots, X_{t_{k-1}}, X_0)\|_2 < \infty.$$

For example, by Theorem 3.6 in *Panaretos and Tavakoli* (2013), if $C(1, 2)$ and $C(1, 4)$ hold, the mean squared error of $f_{\omega}^{(T)}(\cdot, \cdot)$ for $\omega \neq 0, \pm\pi$ is:

$$\mathbb{E} \|\mathcal{F}_{\omega}^{(T)} - \mathcal{F}_{\omega}\|_{\text{HS}}^2 = \mathcal{O}\left(B_T^2 + B_T^{-1} T^{-1}\right),$$

where $\mathcal{F}_\omega^{(T)}$ and \mathcal{F}_ω are the operators with kernels $f_\omega^{(T)}$ and f_ω , respectively. The rate-optimal choice of B_T is $T^{-1/3}$, which yields the bound on the rate of consistency of the estimator $\mathcal{O}(T^{-1/3})$.

Our results provide more detailed estimates on the rates under simple structural assumptions on the covariances. Indeed, observe that the condition $C(1, 2)$ corresponds to our condition $\mathcal{P}_D(\beta, L)$ with $\beta = 1$ in (4.32). Our Theorem IV.17 (see Relation (4.35) with $d = 1$) yields the rate of consistency bound of $\mathcal{O}(T^{-\beta/(2\beta+1)})$, which for $\beta = 1$ matches the rate-optimal bound in *Panaretos and Tavakoli (2013)*. Our condition (4.32), however, allows for a wider range of covariance structures than Condition $C(1, 2)$, where we allow for $\beta > 0$ to be less than 1. As discussed in Section 5.1, the rate $\mathcal{O}(T^{-\beta/(2\beta+1)})$ is minimax optimal in the class $\mathcal{P}_D(\beta, L)$.

As observed in Section 3 of *Zhu and Politis (2020)*, one can relate the time-domain and frequency-domain (periodogram-based) estimators. Indeed, one can argue that our estimator in (4.48) corresponds asymptotically to the periodogram-based estimator in (4.49) with taper

$$W(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} K(t) dt,$$

where $\Delta_T \sim 1/B_T$. In this case, we have $2\pi\|W\|_2^2 = \|K\|_2^2$ and the asymptotic covariances of the estimators in (4.48) and (4.49) are identical (compare, e.g., Theorem 3.7 of *Panaretos and Tavakoli, 2013*, and our Corollary VI.3). Theorem 3.7 in *Panaretos and Tavakoli (2013)* establishes the asymptotic normality of the periodogram-based estimators under conditions $C(1, 2)$ and $C(1, 4)$, as well as $C(0, k)$, for all $k \geq 2$. In Theorem VI.1 we adopt the stronger assumption that the underlying process is Gaussian. We establish, however, the asymptotic normality of our estimators under milder tail-decay conditions on the operator covariance and pseudo-covariance functions.

4.8 Proofs

4.8.1 Proofs for Section 4.4

We begin by recalling some key notation. The spectral density of the \mathbb{H} -valued second order stationary process $X = \{X(t), t \in \mathbb{R}^d\}$ is:

$$f(\theta) := \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} e^{ih^\top \theta} C(h) dh, \quad \theta \in \mathbb{R}^d, \quad (4.50)$$

where the last integral is understood in the sense of Bochner, and where $C(t) = \mathbb{E}[X(t) \otimes X(0)]$ is the operator auto-covariance function of X .

The estimator of the spectral density is defined as:

$$\begin{aligned} \hat{f}_n(\theta) = \frac{1}{(2\pi)^d} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} e^{i(t-s)^\top \theta} \frac{X(t) \otimes X(s)}{|T_n \cap (T_n - (t-s))|} \\ \cdot K\left(\frac{t-s}{\Delta_n}\right) \cdot |V(t)| \cdot |V(s)|. \end{aligned} \quad (4.51)$$

Introduce also the auxiliary, idealized estimator based on the continuously sampled path $\{X(t), t \in T_n\}$:

$$g_n(\theta) = \frac{1}{(2\pi)^d} \int_{t \in T_n} \int_{s \in T_n} e^{i(t-s)^\top \theta} \frac{X(t) \otimes X(s)}{|T_n \cap (T_n - (t-s))|} K\left(\frac{t-s}{\Delta_n}\right) dt ds. \quad (4.52)$$

4.8.1.1 Proof of Theorem IV.9

We begin by recalling the statement.

Theorem IV.27 (Theorem IV.9). *Let Assumptions C, K, and S hold and suppose $\delta_n \vee |T_n|^{-1} \rightarrow 0$. Choose $\Delta_n \rightarrow \infty$ such that*

$$\Delta_n \cdot S_K \subset T_n - T_n \text{ for all } n,$$

where $A - B := \{a - b : a \in A, b \in B\}$ for sets $A, B \subset \mathbb{R}^d$. Then, for any bounded set

$\Theta \subset \mathbb{R}^d$, we have

$$\sup_{\theta \in \Theta} \left\| \mathbb{E} \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}} = \mathcal{O}(\delta_n^\gamma + B_1(\Delta_n) + B_2(\Delta_n)), \quad (4.53)$$

where

$$B_1(\Delta_n) := \left\| \int_{h \in \Delta_n \cdot S_K} e^{ih^\top \theta} C(h) \left(1 - K \left(\frac{h}{\Delta_n} \right) \right) dh \right\|_{\text{HS}},$$

$$B_2(\Delta_n) := \left\| \int_{h \notin \Delta_n \cdot S_K} e^{ih^\top \theta} C(h) dh \right\|_{\text{HS}}.$$

Proof. By the triangle inequality,

$$\left\| \mathbb{E} \hat{f}_n(\theta) - f(\theta) \right\|_{\text{HS}} \leq \left\| \mathbb{E} \hat{f}_n(\theta) - \mathbb{E} g_n(\theta) \right\|_{\text{HS}} + \left\| \mathbb{E} g_n(\theta) - f(\theta) \right\|_{\text{HS}}.$$

It is immediate from (4.50) for f and the inclusion $\Delta_n \cdot S_K \subset T_n - T_n$, that

$$\left\| \mathbb{E} g_n(\theta) - f(\theta) \right\|_{\text{HS}} \leq B_1(\Delta_n) + B_2(\Delta_n).$$

To complete the proof one needs to show that

$$\left\| \mathbb{E} \hat{f}_n(\theta) - \mathbb{E} g_n(\theta) \right\|_{\text{HS}} = \mathcal{O}(\delta_n^\gamma). \quad (4.54)$$

To evaluate $\left\| \mathbb{E} \hat{f}_n(\theta) - \mathbb{E} g_n(\theta) \right\|_{\text{HS}}$, first denote the integrand in (4.52) by

$$h_n(t, s; \theta) := e^{i(t-s)^\top \theta} \frac{X(t) \otimes X(s)}{|T_n \cap (T_n - (t-s))|} K \left(\frac{t-s}{\Delta_n} \right).$$

In view of (4.51) and since $|V(w)| \cdot |V(v)| = \int_{t \in V(w)} \int_{s \in V(v)} \mathbb{1}_{(t \in V(w), s \in V(v))} dt ds$, this

allows us to write:

$$\begin{aligned} & g_n(\theta) - \hat{f}_n(\theta) \\ &= \frac{1}{(2\pi)^d} \sum_{w \in \mathbb{T}_n} \sum_{v \in \mathbb{T}_n} \int_{t \in V(w)} \int_{s \in V(v)} (h_n(t, s; \theta) - h_n(w, v; \theta)) dt ds. \end{aligned}$$

This implies that

$$\begin{aligned} & \left\| \mathbb{E}g_n(\theta) - \mathbb{E}\hat{f}_n(\theta) \right\|_{\text{HS}} \\ & \leq \frac{1}{(2\pi)^d} \sum_{w \in \mathbb{T}_n} \sum_{v \in \mathbb{T}_n} \int_{t \in V(w)} \int_{s \in V(v)} \left\| \mathbb{E}h_n(t, s; \theta) - \mathbb{E}h_n(w, v; \theta) \right\|_{\text{HS}} dt ds. \end{aligned} \quad (4.55)$$

In the rest of the proof we will make use of the smoothness of K and C , and routine but technical analysis to show that the last sum is of order $\mathcal{O}(\delta_n^\gamma)$. This will yield (4.54) and complete the proof of (4.53).

Recall that S_K denotes the bounded support of the kernel function K . This means that

$$K\left(\frac{t-s}{\Delta_n}\right) = 0, \text{ whenever } t-s \notin \Delta_n \cdot S_K.$$

In each integral in the sums of (4.55) we have that $t \in V(w)$ and $s \in V(v)$. Thus,

$$t-s = t-w + w-v + v-s \in w-v + B(0, 2\delta_n),$$

where we used that $\max\{\|t-w\|, \|s-v\|\} \leq \delta_n$, by the definition of δ_n (4.8) and $B(0, r) = \{x \in \mathbb{R}^d : \|x\|_2 < r\}$.

By (4.55), we have

$$\begin{aligned} \left\| \mathbb{E}g_n(\theta) - \mathbb{E}\hat{f}_n(\theta) \right\|_{\text{HS}} & \leq \frac{1}{(2\pi)^d} \sum_{w \in \mathbb{T}_n} \sum_{v \in \mathbb{T}_n} \iint_{t, s \in \mathbb{T}_n} \|C(t-s) - C(w-v)\|_{\text{HS}} |L_n(w-v)| dt ds \\ & \quad + \frac{1}{(2\pi)^d} \sum_{w \in \mathbb{T}_n} \sum_{v \in \mathbb{T}_n} \iint_{t, s \in \mathbb{T}_n} \|C(t-s)\|_{\text{HS}} |L_n(t-s) - L_n(w-v)| ds dt \end{aligned}$$

$$=: I_n + J_n, \quad (4.56)$$

where

$$L_n(x) := \frac{e^{ix^\top \theta}}{|T_n \cap (T_n - x)|} K\left(\frac{x}{\Delta_n}\right), \quad x \in \mathbb{R}^d.$$

Observe that since $K(x/\Delta_n) = 0$ for $x \notin \Delta_n \cdot S_K$, and $\Delta_n^d/|T_n| \rightarrow 0$, Lemma IV.30 implies that $|T_n \cap (T_n - x)| \sim |T_n|$ uniformly in $x \in \Delta_n \cdot S_K$. This and the boundedness of the kernel K imply

$$\sup_{x \in \mathbb{R}^d} |L_n(x)| = \mathcal{O}\left(\frac{1}{|T_n|}\right). \quad (4.57)$$

Recall that by (4.7) in Assumption C, we have

$$\int_{\mathbb{R}^d} \sup_{y \in B(x, \delta)} \|C(y) - C(x)\|_{\text{tr}} dx \leq \|C\|_\gamma \cdot \delta^\gamma, \quad \delta \in (0, 1) \quad (4.58)$$

Thus, for the term I_n in (4.56), using Relations (4.57) and (4.58), and the change of variables $x := t - s$, we obtain

$$\begin{aligned} I_n &\leq \frac{1}{(2\pi)^d} \int_{t \in T_n} \int_{s \in T_n} \sup_{\substack{\|t' - t\| \leq \delta_n \\ \|s' - s\| \leq \delta_n}} \|C(t - s) - C(t' - s')\|_{\text{HS}} \cdot \sup_{\tau \in \mathbb{R}^d} |L_n(\tau)| dt ds \\ &\leq c \frac{1}{|T_n|} \int_{s \in T_n} \left(\int_{x \in s + T_n} \sup_{y: \|x - y\| \leq 2\delta_n} \|C(x) - C(y)\|_{\text{HS}} dx \right) ds \\ &\leq c \frac{1}{|T_n|} \int_{s \in T_n} \|C\|_\gamma \cdot \delta_n^\gamma ds = \mathcal{O}(\delta_n^\gamma). \end{aligned}$$

Next, focus on the term J_n in (4.56). We will show below that

$$\sup_{\|x - y\| \leq 2\delta_n} |L_n(x) - L_n(y)| = \mathcal{O}\left(\frac{\delta_n}{|T_n|}\right). \quad (4.59)$$

Thus, recalling that $\|t - s - (w - v)\| \leq 2\delta_n$, whenever $t \in V(w)$ and $s \in V(v)$, Relation

(4.59) for the term J_n in (4.56) implies

$$J_n \leq_c \frac{\delta_n}{|T_n|} \int_{t \in T_n} \int_{s \in T_n} \|C(t-s)\|_{\text{HS}} dt ds = \mathcal{O}(\delta_n),$$

where the last relation follows from a change of variables $x := t - s$ and Assumption C (a).

To complete the proof, it remains to establish (4.59). By adding and subtracting terms, we obtain

$$|L_n(x) - L_n(y)| \leq |K(x/\Delta_n)| \left| \frac{1}{|T_n \cap (T_n - x)|} - \frac{1}{|T_n \cap (T_n - y)|} \right| \quad (4.60)$$

$$+ \frac{|K(x/\Delta_n) - K(y/\Delta_n)|}{|T_n \cap (T_n - y)|} + \frac{|K(y/\Delta_n)|}{|T_n \cap (T_n - y)|} \left| e^{ix^\top \theta} - e^{iy^\top \theta} \right| \quad (4.61)$$

$$=: A + B + C. \quad (4.62)$$

Note that $K(y/\Delta_n)$ and $K(x/\Delta_n)$ vanish whenever x and y are outside $\Delta_n \cdot S_K$. Therefore, since $\delta_n \rightarrow 0$ and $\Delta_n \rightarrow \infty$, the right-hand side of (4.60) vanishes for all $\|x - y\| \leq 2\delta_n$ such that $\|y\| \geq \text{const} \cdot \Delta_n$. Therefore, the supremum in (4.59) does not change if it is taken over the set

$$\mathcal{I}_n := \{(x, y) : \|x - y\| \leq 2\delta_n, \|y\| \leq \text{const} \cdot \Delta_n\}.$$

Thus, we restrict our attention to $(x, y) \in \mathcal{I}_n$. By Lemma IV.30, we have $|T_n \cap (T_n - y)| \sim |T_n|$, uniformly in $(x, y) \in \mathcal{I}_n$. This fact and the Lipschitz property of the complex exponentials and the kernel K (by (c) of Assumption K), immediately imply that

$$B \leq_c \frac{\delta_n/\Delta_n}{|T_n|} \quad \text{and} \quad C \leq_c \frac{\delta_n}{|T_n|},$$

uniformly in $(x, y) \in \mathcal{I}_n$.

Now, for term A , exploiting the boundedness of the kernel and the fact that

$\|x - y\| \leq 2\delta_n$, we obtain

$$\begin{aligned}
A &\leq \|K\|_\infty \frac{\left| |T_n \cap (T_n - x)| - |T_n \cap (T_n - y)| \right|}{|T_n \cap (T_n - x)| \cdot |T_n \cap (T_n - y)|} \\
&\leq \|K\|_\infty \frac{|T_n + B(0, 2\delta_n)| - |T_n|}{|T_n \cap (T_n - x)| \cdot |T_n \cap (T_n - y)|} \\
&\leq_c \frac{|T_n + B(0, 2\delta_n)| - |T_n|}{|T_n|^2} = \mathcal{O}\left(\frac{\delta_n}{|T_n|^{1+1/d}}\right),
\end{aligned}$$

where the last inequality follows from Lemma IV.30 and Assumption S. Note that the last bound is uniform in $(x, y) \in \mathcal{I}_n$. Combining the above bounds on the terms A, B , and C , we obtain (4.59). This completes the proof. \square

4.8.1.2 Proof of Theorem IV.13

For easy referencing, we begin by recalling the statement of Theorem IV.13.

Theorem IV.28 (Theorem IV.13). *Let $X = \{X(t), t \in \mathbb{R}^d\}$ be a zero-mean, strictly stationary real \mathbb{H} -valued process. Suppose that Assumptions C, K, S, and V hold. Also, assume that Δ_n satisfies*

$$\Delta_n \cdot S_K \subset T_n - T_n \quad \text{for all } n, \text{ and} \quad \delta_n + \Delta_n^d/|T_n| \rightarrow 0 \text{ as } n \rightarrow \infty.$$

Then

$$\sup_{\theta \in \Theta} \mathbb{E} \|\hat{f}_n(\theta) - \mathbb{E} \hat{f}_n(\theta)\|_{\text{HS}}^2 = \mathcal{O}\left(\frac{\Delta_n^d}{|T_n|}\right), \text{ as } n \rightarrow \infty.$$

Proof. In what follows we will use Δ and T instead of Δ_n and T_n respectively. Recall

(4.23) and (4.24). Namely, we have

$$\begin{aligned}
\mathbb{E}\|\hat{f}_n(\theta) - \mathbb{E}\hat{f}_n(\theta)\|_{\text{HS}}^2 &= \frac{1}{(2\pi)^{2d}} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} \sum_{\substack{h \in [\Delta \cdot S_K] \cap (\mathbb{T}_n - t) \\ h' \in [\Delta \cdot S_K] \cap (\mathbb{T}_n - s)}} e^{i(h-h')^\top \theta} K\left(\frac{h}{\Delta}\right) K\left(\frac{h'}{\Delta}\right) \\
&\quad \cdot |V(t+h)| \cdot |V(t)| \cdot |V(s+h')| \cdot |V(s)| \\
&\quad \cdot \frac{\text{Cov}(X(t+h) \otimes X(t), X(s+h') \otimes X(s))}{|T \cap (T-h)| |T \cap (T-h')|}
\end{aligned} \tag{4.63}$$

where

$$\begin{aligned}
&\text{Cov}(X(t+h) \otimes X(t), X(s+h') \otimes X(s)) \\
&:= \mathbb{E}\langle X(t+h) \otimes X(t) - C(h), X(s+h') \otimes X(s) - C(h') \rangle_{\text{HS}}.
\end{aligned}$$

By Definition IV.11, since X is real $C = \check{C}$, and

$$\begin{aligned}
&\text{Cov}(X(t+h) \otimes X(t), X(s+h') \otimes X(s)) \\
&= \mathbb{E}\langle X(t+h), X(s+h') \rangle_{\mathbb{H}} \cdot \mathbb{E}\langle X(s), X(t) \rangle_{\mathbb{H}} \\
&\quad + \langle C(t-s+h), C(s+h'-t) \rangle_{\text{HS}} \\
&\quad + \text{cum}(X(t+h), X(t), X(s+h'), X(s)).
\end{aligned} \tag{4.64}$$

For simplicity of notation, write $\text{cum}(s, t, u, v) = \text{cum}(X(s), X(t), X(u), X(v))$. We fix a real CONS $\{e_j\}$ and use the representation in Proposition VI.10 (see also (4.20)). Next, we split the sum on the right-hand side of (4.63) into three terms corresponding to the decomposition (4.64). Namely, we define

$$\begin{aligned}
A &:= \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} \sum_{\substack{h \in [\Delta \cdot S_K] \cap (\mathbb{T}_n - t) \\ h' \in [\Delta \cdot S_K] \cap (\mathbb{T}_n - s)}} e^{i(h-h')^\top \theta} K\left(\frac{h}{\Delta}\right) K\left(\frac{h'}{\Delta}\right) \cdot |V(t+h)| \cdot |V(t)| \\
&\quad \cdot |V(s+h')| \cdot |V(s)| \cdot \frac{\mathbb{E}\langle X(t+h), X(s+h') \rangle_{\mathbb{H}} \cdot \mathbb{E}\langle X(s), X(t) \rangle_{\mathbb{H}}}{|T \cap (T-h)| |T \cap (T-h')|},
\end{aligned}$$

$$\begin{aligned}
B := & \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} \sum_{\substack{h \in [\Delta \cdot S_K] \cap (\mathbb{T}_n - t) \\ h' \in [\Delta \cdot S_K] \cap (\mathbb{T}_n - s)}} e^{i(h-h')^\top \theta} K\left(\frac{h}{\Delta}\right) K\left(\frac{h'}{\Delta}\right) \cdot |V(t+h)| \cdot |V(t)| \\
& \cdot |V(s+h')| \cdot |V(s)| \cdot \frac{\langle C(t-s+h), C(s-t+h') \rangle_{\text{HS}}}{|T \cap (T-h)| |T \cap (T-h')|}
\end{aligned}$$

and

$$\begin{aligned}
C := & \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} \sum_{\substack{h \in [\Delta \cdot S_K] \cap (\mathbb{T}_n - t) \\ h' \in [\Delta \cdot S_K] \cap (\mathbb{T}_n - s)}} e^{i(h-h')^\top \theta} K\left(\frac{h}{\Delta}\right) K\left(\frac{h'}{\Delta}\right) \cdot |V(t+h)| \cdot |V(t)| \\
& \cdot |V(s+h')| \cdot |V(s)| \frac{\text{cum}(X(t+h), X(t), X(s+h'), X(s))}{|T \cap (T-h)| |T \cap (T-h')|}.
\end{aligned}$$

Thus,

$$(2\pi)^d \mathbb{E} \|\hat{f}_n(\theta) - \mathbb{E} \hat{f}_n(\theta)\|_{\text{HS}}^2 = A + B + C. \quad (4.65)$$

In the sequel, the bounds we shall obtain are based on the summation of the absolute values of the summands. Therefore, in view of Lemma IV.30 (below) and the assumption $\Delta^d = o(|T|)$, the denominators in A, B, C can be replaced by $|T|^{-2}$.

We start with the term C . Lemma IV.34 entails that

$$C = \mathcal{O}\left(\frac{\mathcal{N}(\Delta_n \cdot S_K, \mathbb{T}_n)}{|T_n|}\right) = \mathcal{O}\left(\frac{\Delta_n^d}{|T_n|}\right), \quad (4.66)$$

where the last relation follows from Lemma IV.29.

The term B is bounded above by

$$|B| \leq \frac{1}{|T_n|^2} \sum_{\substack{t, s \in \mathbb{T}_n \\ u \in \mathbb{T}_n \cap (t + \Delta \cdot S_K) \\ v \in \mathbb{T}_n \cap (s + \Delta \cdot S_K)}} \|C(u-s)\|_{\text{HS}} \|C(v-t)\|_{\text{HS}} \cdot |V(t)| \cdot |V(s)| \cdot |V(u)| \cdot |V(v)|,$$

where we have implemented the change of variables $u = t+h$ and $v = s+h'$. Applying Lemma IV.33, we immediately obtain that

$$B = \mathcal{O}\left(\frac{\mathcal{N}(\Delta_n \cdot S_K, \mathbb{T}_n)^2}{|T_n|^2}\right) = \mathcal{O}\left(\frac{\Delta_n^{2d}}{|T_n|^2}\right), \quad (4.67)$$

where we applied Lemma IV.29.

Finally, we steer our attention to the term A . Observe that

$$|\mathbb{E}\langle X, Y \rangle| = |\mathbb{E} \operatorname{trace}(X \otimes Y)| = |\operatorname{trace}(\mathbb{E}[X \otimes Y])| \leq \|\mathbb{E}[X \otimes Y]\|_{\operatorname{tr}} \quad (4.68)$$

by (iii) of Lemma IV.37. Thus,

$$\begin{aligned} |A| &\leq_c \frac{\|K\|_{\infty}^2}{|T|^2} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} \sum_{\substack{h \in [\Delta \cdot S_K] \cap (\mathbb{T}_n - t) \\ h' \in [\Delta \cdot S_K] \cap (\mathbb{T}_n - s)}} \|\mathbb{E}[X(t+h) \otimes X(s+h')]\|_{\operatorname{tr}} \\ &\quad \cdot \|\mathbb{E}[X(s) \otimes X(t)]\|_{\operatorname{tr}} \cdot |V(t+h)| \cdot |V(t)| \cdot |V(s+h')| \cdot |V(s)| \\ &= \frac{\|K\|_{\infty}^2 \mathcal{N}(\Delta \cdot S_K, \mathbb{T}_n)}{|T|} \cdot A_1 \cdot A_2, \end{aligned}$$

where

$$A_1 = \frac{1}{|T|} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} \|C(s-t)\|_{\operatorname{tr}} \cdot |V(t)| \cdot |V(s)|$$

and

$$A_2 = \frac{1}{\mathcal{N}(\Delta \cdot S_K, \mathbb{T}_n)} \sum_{\substack{h \in [\Delta \cdot S_K] \cap (\mathbb{T}_n - t) \\ h' \in [\Delta \cdot S_K] \cap (\mathbb{T}_n - s)}} \|C(t-s+h-h')\|_{\operatorname{tr}} \cdot |V(t+h)| \cdot |V(s+h')|.$$

Now

$$A_2 \leq \frac{1}{\mathcal{N}(\Delta \cdot S_K, \mathbb{T}_n)} \max_{t, s \in \mathbb{T}_n} \sum_{\substack{u \in \mathbb{T}_n \cap (t + \Delta \cdot S_K) \\ v \in \mathbb{T}_n \cap (s + \Delta \cdot S_K)}} \|C(u-v)\|_{\operatorname{tr}} \cdot |V(u)| \cdot |V(v)|.$$

By Lemma IV.32 we obtain that $A_2 = \mathcal{O}(1)$. Moreover, a close inspection of the proof of Lemma IV.32 shows that A_1 is also of the order $\mathcal{O}(1)$. Keeping only the

dominating bounds for A , we have that

$$A = \mathcal{O} \left(\frac{\mathcal{N}(\Delta \cdot S_K, \mathbb{T}_n)}{|T|} \right) = \mathcal{O} \left(\frac{\Delta_n^d}{|T_n|} \right), \quad (4.69)$$

by Lemma IV.29. In view of (4.65), gathering all the bounds in (4.66), (4.67), and (4.69), we complete the proof of the theorem. \square

4.8.1.3 Lemmas used in the proofs of Theorems IV.9 and IV.13

For the next lemmas, we need to define the quantity:

$$\mathcal{N}(A, \mathbb{T}_n) := \max_{w \in \mathbb{T}_n} \left| \cup_{v \in \mathbb{T}_n, w-v \in A+B(0,2\delta_n)} V(v) \right|. \quad (4.70)$$

This is the maximum volume over w of the unions of all tessellation cells for which the representatives v 's are in the $2\delta_n$ inflated A neighborhood of w .

Lemma IV.29. *Let $\mathcal{N}(\Delta_n \cdot S_K, \mathbb{T}_n)$ be defined as in (4.70) and suppose that (a) of Assumption S holds. Then*

$$\mathcal{N}(\Delta_n \cdot S_K, \mathbb{T}_n) = \mathcal{O}(\Delta_n^d), \quad \text{as } \Delta_n \rightarrow \infty.$$

Proof. Indeed, let $Z(t) := \cup_{s \in \mathbb{T}_n, t-s \in \Delta_n \cdot S_K + B(0,2\delta_n)} V(s)$ and $t_0 \in \mathbb{T}_n$. We will show that

$$Z(t_0) \subseteq B(t_0, 3\delta_n + \Delta_n \cdot M_K), \quad (4.71)$$

where $M_k = \sup_{h \in S_K} \|h\|_2$. Suppose $u \in Z(t_0)$. Then, there is $s_u \in \mathbb{T}_n$ such that

$$\|t_0 - s_u\|_2 \leq \Delta_n \cdot M_K + 2\delta_n$$

with $u \in V(s_u)$. Thus,

$$\|u - t_0\|_2 \leq \|u - s_u\|_2 + \|s_u - t_0\|_2 \leq \delta_n + \Delta_n M_K + 2\delta_n = 3\delta_n + \Delta_n M_K,$$

which implies (4.71). This entails that,

$$\begin{aligned} \mathcal{N}(\Delta_n \cdot S_K, \mathbb{T}_n) &= \max_{t \in \mathbb{T}_n} |Z(t)| \leq |B(0, 3\delta_n + \Delta_n M_K)| \\ &= \mathcal{O}((\Delta_n + \delta_n)^d) = \mathcal{O}(\Delta_n^d), \end{aligned}$$

as $\Delta_n \rightarrow \infty$, where the last relation follows from (a) of Assumption S. □

Lemma IV.30. *Under Assumption S, for $\|h\|_2 \leq |T_n|^{1/d}$ we have that*

$$\frac{|T_n| - |T_n \cap (T_n - h)|}{|T_n|} = \mathcal{O}\left(\frac{\|h\|_2}{|T_n|^{1/d}}\right), \text{ as } n \rightarrow \infty.$$

Consequently, if $\sup_{h \in A_n} \|h\|^d = o(|T_n|)$, we have that

$$\sup_{h \in A_n} \frac{|T_n \cap (T_n - h)|}{|T_n|} = \mathcal{O}(1).$$

Proof. We will make critical use of the Steiner formula from convex analysis (see, e.g. Gruber, 2007). We have that

$$\frac{|T_n| - |T_n \cap (T_n - h)|}{|T_n|} \leq \frac{|T_n + B(0, \|h\|_2)| - |T_n|}{|T_n|}.$$

An application of Steiner formula to the convex set T_n entails that

$$|T_n + B(0, \|h\|_2)| = \sum_{j=0}^d \mu_j(T_n) \|h\|_2^{d-j},$$

where $\mu_j(\cdot)$ denote the intrinsic volumes of order j . Note that $\mu_d(T_n) = |T_n|$. Thus,

$$\begin{aligned} \frac{|T_n + B(0, \|h\|_2)| - |T_n|}{|T_n|} &= \sum_{j=0}^{d-1} \frac{\mu_j(T_n) \|h\|_2^{d-j}}{|T_n|} \\ &= \sum_{j=0}^{d-1} \mu_j \left(\frac{T_n}{|T_n|^{1/d}} \right) \cdot \left(\frac{\|h\|_2}{|T_n|^{1/d}} \right)^{d-j}, \end{aligned}$$

where the last equality follows from the homogeneity of the intrinsic volumes.

Assumption S, part (b), along with the continuity of the intrinsic volumes in the Hausdorff metric on the set of convex bodies see, e.g., Section 1.2.2 in *Lotz et al. (2018)* or Theorem 6.13(iii) in *Gruber (2007)* and the fact that $\|h\|_2 \leq |T_n|^{1/d}$ complete the proof. \square

The following remark shows that the order of the bounds in Lemma IV.30 obtained using the Steiner formula cannot be improved.

Remark IV.31. For the case $d = 2$ and $d = 3$, when T_n is a circle and a sphere respectively, we can evaluate the desired volume exactly. Indeed, for $d = 2$, we have that

$$\begin{aligned} \frac{|T_n| - |T_n \cap (T_n - h)|}{|T_n|} &= 1 - \frac{2}{\pi} \arccos \left(\frac{\|h\|_2}{2|T_n|^{1/2}} \right) + \frac{1}{2|T_n|^{1/2}} \|h\|_2 \sqrt{4 - \left(\frac{\|h\|_2}{|T_n|^{1/2}} \right)^2} \\ &= \mathcal{O} \left(\frac{\|h\|_2}{|T_n|^{1/2}} \right), \end{aligned}$$

and for $d = 3$ we have that

$$\frac{|T_n| - |T_n \cap (T_n - h)|}{|T_n|} = \frac{3}{4} \frac{\|h\|_2}{|T_n|^{1/3}} - \frac{1}{16} \left(\frac{\|h\|_2}{|T_n|^{1/3}} \right)^3 = \mathcal{O} \left(\frac{\|h\|_2}{|T_n|^{1/3}} \right).$$

These two cases provide evidence that the application of Steiner formula is not giving us a loose upper bound, at least when T_n is an n -dimensional ball.

Now, when T_n is a square, and assuming that S_n is the side of the square, we have

that

$$\max \frac{|T_n| - |T_n \cap (T_n - h)|}{|T_n|} = \frac{S_n \cdot \|h\|_2 \cdot \sqrt{2} - \|h\|_2^2/2}{S_n^2} = \mathcal{O}\left(\frac{\|h\|_2}{|T_n|^{1/2}}\right).$$

Finally, when T_n is a cube of side S_n , we have

$$\max \frac{|T_n| - |T_n \cap (T_n - h)|}{|T_n|} = \frac{\frac{3}{2} \cdot S_n^2 \cdot \|h\|_2 - \frac{\sqrt{2}}{2} \cdot \|h\|_2^3}{S_n^3} = \mathcal{O}\left(\frac{\|h\|_2}{|T_n|^{1/3}}\right).$$

This suggests that using n -dimensional cubes leads indeed to the same rates as for the n -dimensional balls.

Lemma IV.32. *Let $\{A_n\}$ be a growing sequence of open sets such that $A_n \uparrow \mathbb{R}^d$, as $n \rightarrow \infty$. Moreover, let \mathbb{T}_n be the set of representatives of a tessellation of T_n , with the diameter $\delta_n \rightarrow 0$, as $n \rightarrow \infty$, where T_n is as in (4.10). Also, let Assumptions C, K and S hold. Then*

$$\frac{1}{\mathcal{N}(A_n, \mathbb{T}_n)} \max_{t, s \in \mathbb{T}_n} \sum_{\substack{u \in \mathbb{T}_n \cap (t + A_n) \\ v \in \mathbb{T}_n \cap (s + A_n)}} \|C(u - v)\|_{\text{tr}} \cdot |V(u)| \cdot |V(v)| = \mathcal{O}(\delta_n^\gamma + 1) = \mathcal{O}(1),$$

as $n \rightarrow \infty$, where $\mathcal{N}(\cdot, \cdot)$ is defined in (4.70).

Proof. Using the inequality

$$\left| \max_{i=1, \dots, m} a_i - \max_{j=1, \dots, m} b_j \right| \leq \max_{i=1, \dots, m} |a_i - b_i|,$$

valid for all $a_i, b_i \in \mathbb{R}$, $i = 1, \dots, m$, we obtain

$$\left| \max_{t, s \in \mathbb{T}_n} \sum_{\substack{u \in \mathbb{T}_n \cap (t + A_n) \\ v \in \mathbb{T}_n \cap (s + A_n)}} \|C(u - v)\|_{\text{tr}} \cdot |V(u)| \cdot |V(v)| \right. \\ \left. - \max_{t, s \in \mathbb{T}_n} \iint_{\substack{h \in \cup_{u \in \mathbb{T}_n \cap (t + A_n)} V(u) \\ h' \in \cup_{v \in \mathbb{T}_n \cap (s + A_n)} V(v)}} \|C(h - h')\|_{\text{tr}} dh' dh \right|$$

$$\begin{aligned}
&\leq \max_{t,s \in \mathbb{T}_n} \left| \sum_{\substack{u \in \mathbb{T}_n \cap (t+A_n) \\ v \in \mathbb{T}_n \cap (s+A_n)}} \|C(u-v)\|_{\text{tr}} \cdot |V(u)| \cdot |V(v)| \right. \\
&\quad \left. - \iint_{\substack{h \in \cup_{u \in \mathbb{T}_n \cap (t+A_n)} V(u) \\ h' \in \cup_{v \in \mathbb{T}_n \cap (s+A_n)} V(v)}} \|C(h-h')\|_{\text{tr}} dh' dh \right| \\
&= \max_{t,s \in \mathbb{T}_n} \left| \sum_{\substack{u \in \mathbb{T}_n \cap (t+A_n) \\ v \in \mathbb{T}_n \cap (s+A_n)}} \iint_{\substack{h \in V(u) \\ h' \in V(v)}} \|C(u-v)\|_{\text{tr}} - \|C(h-h')\|_{\text{tr}} dh' dh \right| \\
&\leq \max_{s \in \mathbb{T}_n} \sum_{v \in \mathbb{T}_n \cap (s+A_n)} \int_{h' \in V(v)} \int_{x \in \mathbb{R}^d} \left(\sup_{y: \|x-y\| \leq 2\delta_n} \|C(y) - C(x)\|_{\text{tr}} \right) dx dh' \\
&\leq \|C\|_{\gamma} (2\delta_n)^{\gamma} \cdot \max_{s \in \mathbb{T}_n} \sum_{v \in \mathbb{T}_n \cap (s+A_n)} \int_{h' \in V(v)} dh' \\
&\leq \|C\|_{\gamma} (2\delta_n)^{\gamma} \mathcal{N}(A_n, \mathbb{T}_n),
\end{aligned}$$

where we made the change of variables $x := h - h'$ and enlarged the domain of integration over $x \in \mathbb{R}^d$. The last two inequalities follow from (4.7) and definition of $\mathcal{N}(\cdot, \cdot)$ in (4.70).

To complete the proof, we show that

$$\frac{1}{\mathcal{N}(A_n, \mathbb{T}_n)} \max_{t,s \in \mathbb{T}_n} \iint_{\substack{h \in \cup_{u \in \mathbb{T}_n \cap (t+A_n)} V(u) \\ h' \in \cup_{v \in \mathbb{T}_n \cap (s+A_n)} V(v)}} \|C(h-h')\|_{\text{tr}} dh' dh = \mathcal{O}(1).$$

With the change of variables $x = h - h'$, we have that the aforementioned term is equal to

$$\begin{aligned}
&\frac{1}{\mathcal{N}(A_n, \mathbb{T}_n)} \max_{t,s \in \mathbb{T}_n} \iint_{\substack{x \in [\cup_{u \in \mathbb{T}_n \cap (t+A_n)} V(u) - \cup_{v \in \mathbb{T}_n \cap (s+A_n)} V(v)] \\ h' \in [\cup_{v \in \mathbb{T}_n \cap (s+A_n)} V(v)] \cap [\cup_{u \in \mathbb{T}_n \cap (t+A_n)} V(u) - x]}} \|C(x)\|_{\text{tr}} dh' dx \\
&\leq \max_{t,s \in \mathbb{T}_n} \int_{\substack{x \in [\cup_{u \in \mathbb{T}_n \cap (t+A_n)} V(u) \\ - \cup_{v \in \mathbb{T}_n \cap (s+A_n)} V(v)]}} \|C(x)\|_{\text{tr}} \frac{|[\cup_{v \in \mathbb{T}_n \cap (s+A_n)} V(v)] \cap [\cup_{u \in \mathbb{T}_n \cap (t+A_n)} V(u) - x]|}{\mathcal{N}(A_n, \mathbb{T}_n)} dx
\end{aligned}$$

$$\leq \int_{w \in \mathbb{R}^d} \|C(x)\|_{\text{tr}} dx = \mathcal{O}(1),$$

by Assumption C (a). The proof is complete. \square

The next lemma is similar to Lemma IV.32. It is used for the term B in the proof of Theorem IV.28.

Lemma IV.33. *Let all the assumptions of Lemma IV.32 hold. Then*

$$\begin{aligned} & \frac{1}{|T_n|^2} \sum_{\substack{t, s \in \mathbb{T}_n \\ u \in \mathbb{T}_n \cap (t + A_n) \\ v \in \mathbb{T}_n \cap (s + A_n)}} \|C(u - s)\|_{\text{HS}} \|C(t - v)\|_{\text{HS}} \cdot |V(t)| \cdot |V(s)| \cdot |V(u)| \cdot |V(v)| \\ &= \mathcal{O} \left(\frac{\mathcal{N}(A_n, \mathbb{T}_n)^2}{|T_n|^2} \right), \end{aligned} \quad (4.72)$$

as $n \rightarrow \infty$, where $\mathcal{N}(\cdot, \cdot)$ is defined in (4.70).

Proof. For any w in T_n , let τ_w denote the point $t_{n,i} \in \mathbb{T}_n$ that is in the same cell as w ; if w is on the boundary of a cell, then let τ_w be any of the $t_{n,i} \in \mathbb{T}_n$ in adjacent cells. Thus, $\|w - \tau_w\|_2 \leq \delta_n$. It follows that

$$\begin{aligned} & \sum_{\substack{t, s \in \mathbb{T}_n \\ u \in \mathbb{T}_n \cap (t + A_n) \\ v \in \mathbb{T}_n \cap (s + A_n)}} \|C(u - s)\|_{\text{HS}} \|C(t - v)\|_{\text{HS}} \cdot |V(t)| \cdot |V(s)| \cdot |V(u)| \cdot |V(v)| \\ &= \iiint_{\substack{w, x \in T_n \\ h \in \cup_{u \in \mathbb{T}_n \cap (\tau_w + A_n)} V(u) \\ h' \in \cup_{v \in \mathbb{T}_n \cap (\tau_x + A_n)} V(v)}} \|C(\tau_h - \tau_x)\|_{\text{HS}} \|C(\tau_w - \tau_{h'})\|_{\text{HS}} dh' dh dx dw \\ &\leq \iiint_{\substack{w, x \in T_n \\ h \in w + A_n + B(0, 2\delta_n) \\ h' \in x + A_n + B(0, 2\delta_n)}} \|C(\tau_h - \tau_x)\|_{\text{HS}} \|C(\tau_w - \tau_{h'})\|_{\text{HS}} dh' dh dx dw \\ &\leq \iiint_{\substack{w, x \in T_n \\ h \in w + A_n + B(0, 2\delta_n) \\ h' \in x + A_n + B(0, 2\delta_n)}} \sup_{\substack{\lambda_i \in B(0, 2\delta_n), \\ i=1,2}} \|C(\lambda_1 + h - x)\|_{\text{HS}} \|C(\lambda_2 + w - h')\|_{\text{HS}} dh' dh dx dw \end{aligned}$$

$$\begin{aligned}
&= \iiint_{\substack{w, x \in T_n \\ \tilde{h} \in A_n + B(0, 2\delta_n) \\ \tilde{h}' \in A_n + B(0, 2\delta_n)}} \sup_{\substack{\lambda_i \in B(0, 2\delta_n), \\ i=1,2}} \|C(\lambda_1 + \tilde{h} + w - x)\|_{\text{HS}} \|C(\lambda_2 + w - \tilde{h}' - x)\|_{\text{HS}} d\tilde{h}' d\tilde{h} dx dw \\
&\leq \iint_{\substack{\tilde{h} \in A_n + B(0, 2\delta_n) \\ \tilde{h}' \in A_n + B(0, 2\delta_n)}} \left(\int_{x \in \mathbb{R}^d} \sup_{\lambda \in B(0, 2\delta_n)} \|C(\lambda + x)\|_{\text{HS}} dx \right)^2 d\tilde{h}' d\tilde{h} \\
&= |A_n + B(0, 2\delta_n)|^2 \left(\int_{x \in \mathbb{R}^d} \sup_{\lambda \in B(0, 2\delta_n)} \|C(\lambda + x)\|_{\text{HS}} dx \right)^2
\end{aligned}$$

which in view of Assumption C implies (4.72) and completes the proof, since

$$|A_n + B(0, 2\delta_n)| \asymp \mathcal{N}(A_n, \mathbb{T}_n), \quad (4.73)$$

because $\delta_n/|A_n| \rightarrow 0$ (recall (4.70)). □

Finally, we state a lemma to handle term C in the proof of Theorem IV.28.

Lemma IV.34. *Let the assumptions of Lemma IV.32 and Assumption V hold. Moreover, assume that the process $\{X(t)\}$ is strictly stationary. Then,*

$$\frac{1}{|T_n|^2} \sum_{\substack{t, s \in \mathbb{T}_n \\ u \in \mathbb{T}_n \cap (t + A_n) \\ v \in \mathbb{T}_n \cap (s + A_n)}} \left| \text{cum}(X(u), X(t), X(v), X(s)) \right| \cdot |V(t)| \cdot |V(s)| \cdot |V(u)| \cdot |V(v)|$$

is of the order $\mathcal{O}(\mathcal{N}(A_n, \mathbb{T}_n)/|T_n|)$, where $\mathcal{N}(\cdot, \cdot)$ is defined in (4.70).

Proof. Proceeding as in Lemma IV.33, It follows that

$$\begin{aligned}
&\sum_{\substack{t, s \in \mathbb{T}_n \\ u \in \mathbb{T}_n \cap (t + A_n) \\ v \in \mathbb{T}_n \cap (s + A_n)}} |\text{cum}(u, t, v, s)| \prod_{\tau \in \{t, s, v, u\}} |V(\tau)| \\
&= \iiint_{\substack{w, x \in T_n \\ h \in \cup_{u \in \mathbb{T}_n \cap (\tau_w + A_n)} V(u) \\ h' \in \cup_{v \in \mathbb{T}_n \cap (\tau_x + A_n)} V(v)}} |\text{cum}(\tau_h, \tau_w, \tau_{h'}, \tau_x)| dh' dh dx dw
\end{aligned}$$

$$\begin{aligned}
&\leq \iiint_{\substack{w, x \in T_n \\ h \in \tau_w + A_n + B(0, \delta_n) \\ h' \in \tau_x + A_n + B(0, \delta_n)}} |\text{cum}(\tau_h, \tau_w, \tau_{h'}, \tau_x)| dh' dh dx dw \\
&\leq \iiint_{\substack{w, x \in T_n \\ h \in w + A_n + B(0, 2\delta_n) \\ h' \in x + A_n + B(0, 2\delta_n)}} |\text{cum}(\tau_h, \tau_w, \tau_{h'}, \tau_x)| dh' dh dx dw.
\end{aligned}$$

Applying (b) of Assumption V, the last expression becomes

$$\begin{aligned}
&\iiint_{\substack{w, x \in T_n \\ h \in w + A_n + B(0, 2\delta_n) \\ h' \in x + A_n + B(0, 2\delta_n)}} |\text{cum}(\tau_h - \tau_x, \tau_w - \tau_x, \tau_{h'} - \tau_x, 0)| dh' dh dx dw \\
&\leq \iiint_{\substack{w, x \in T_n \\ h \in w + A_n + B(0, 2\delta_n) \\ h' \in x + A_n + B(0, 2\delta_n)}} \sup_{\substack{\lambda_i \in B(0, 2\delta_n), \\ i=1,2,3}} |\text{cum}(\lambda_1 + h - x, \lambda_2 + w - x, \lambda_3 + h' - x, 0)| dh' dh dx dw \\
&= \iiint_{\substack{w, x \in T_n \\ \tilde{h}, \tilde{h}' \in A_n + B(0, 2\delta_n)}} \sup_{\substack{\lambda_i \in B(0, 2\delta_n), \\ i=1,2,3}} |\text{cum}(\lambda_1 + \tilde{h} + w - x, \lambda_2 + w - x, \lambda_3 + \tilde{h}', 0)| d\tilde{h}' d\tilde{h} dx dw \\
&= \iiint_{\substack{y \in T_n - T_n \\ \tilde{h}, \tilde{h}' \in A_n + B(0, 2\delta_n)}} |T_n \cap (T_n - y)| \sup_{\substack{\lambda_i \in B(0, 2\delta_n), \\ i=1,2,3}} |\text{cum}(\lambda_1 + \tilde{h} + y, \lambda_2 + y, \lambda_3 + \tilde{h}', 0)| d\tilde{h}' d\tilde{h} dy \\
&\leq |T_n| \int_{\substack{z \in \mathbb{R}^d \\ \tilde{h}' \in A_n + B(0, 2\delta_n)}} \sup_{\substack{y \in T_n - T_n \\ \tilde{h} \in A_n + B(0, 2\delta_n)}} \iiint_{\substack{\lambda_i \in B(0, 2\delta_n), \\ i=1,2,3}} |\text{cum}(\lambda_1 + \tilde{h} + y, \lambda_2 + y, \lambda_3 + z, 0)| d\tilde{h}' d\tilde{h} dy \\
&= \mathcal{O}(\mathcal{N}(A_n, \mathbb{T}_n) \cdot |T_n|),
\end{aligned}$$

where the last relation is justified by Assumption V (b) and (4.73). Note that we have applied two changes of variables; first $\tilde{h} = h - \tau_w$ and $\tilde{h}' = h' - x$, and second $v = \tilde{w} - x$. This completes the proof of the lemma. \square

4.8.2 Proofs for Section 4.5

We start this section obtaining rates on the variance of $\hat{f}_n(\theta)$ in Section 4.5. We first establish a result that is more general than what is needed for the proofs of Section 4.5. We will use it to evaluate the variance of $\hat{f}_n(\theta)$ in the time-series setting where $\delta_n \equiv 1$.

Proposition IV.35. *Let the process $\{X(t)\}_{t \in \delta_n \cdot \mathbb{Z}^d}$ be strictly stationary and suppose that Assumptions C, S, and V hold. Then, for the estimator $\hat{f}_n(\theta)$ defined in (4.27), we have the following upper bound on the rate of the variance*

$$\sup_{\theta \in \Theta} \mathbb{E} \|\hat{f}_n(\theta) - \mathbb{E} \hat{f}_n(\theta)\|_{\text{HS}}^2 = \mathcal{O} \left(\frac{\Delta_n^d}{|T_n|} \right), \quad \text{as } n \rightarrow \infty,$$

where $T_n = \delta_n \cdot [0, n]^d$ and $|T_n| = (n\delta_n)^d$.

Proof. As before, we will use that $\|\mathcal{A}\|_{\text{HS}} \leq \|\mathcal{A}\|_{\text{tr}}$ throughout. Recall that $\mathbb{T}_n = \delta_n \cdot \{1, \dots, n\}^d$ is a discrete set of n^d samples, while $T_n = \delta_n \cdot [0, n]^d$ is a hypercube of side $n\delta_n$.

We start with

$$\begin{aligned} & \hat{f}_n(\theta) - \mathbb{E} \hat{f}_n(\theta) \\ &= \frac{\delta_n^{2d}}{(2\pi)^d} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} e^{i(t-s)^\top \theta} \frac{X(t) \otimes X(s) - C(t-s)}{|T_n \cap (T_n - (t-s))|} K \left(\frac{t-s}{\Delta_n} \right) \\ &= \frac{\delta_n^{2d}}{(2\pi)^d} \sum_{t \in \mathbb{T}_n} \sum_{h \in \Delta_n S_K \cap \delta_n \cdot \mathbb{Z}^d} e^{ih^\top \theta} \frac{X(t+h) \otimes X(t) - C(h)}{|T_n \cap (T_n - h)|} K \left(\frac{h}{\Delta_n} \right) \cdot \mathbb{1}(h+t \in \delta_n \cdot \mathbb{Z}^d). \end{aligned}$$

This means that

$$\begin{aligned} & \hat{f}_n(\theta) - \mathbb{E} \hat{f}_n(\theta) \\ &= \frac{\delta_n^{2d}}{(2\pi)^d} \sum_{t \in \mathbb{T}_n} \sum_{h \in \Delta_n S_K \cap (\mathbb{T}_n - t)} e^{ih^\top \theta} \frac{X(t+h) \otimes X(t) - C(h)}{|T_n \cap (T_n - h)|} K \left(\frac{h}{\Delta_n} \right). \end{aligned}$$

Then, the variance becomes

$$\begin{aligned} & \mathbb{E} \|\hat{f}_n(\theta) - \mathbb{E} \hat{f}_n(\theta)\|_{\text{HS}}^2 \\ &= \frac{\delta_n^{4d}}{(2\pi)^{2d}} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} \sum_{\substack{h \in [\Delta_n \cdot S_K] \cap (\mathbb{T}_n - t) \\ h' \in [\Delta_n \cdot S_K] \cap (\mathbb{T}_n - s)}} e^{i(h-h')^\top \theta} K\left(\frac{h}{\Delta_n}\right) K\left(\frac{h'}{\Delta_n}\right) \\ & \quad \cdot \frac{\text{Cov}(X(t+h) \otimes X(t), X(s+h') \otimes X(s))}{|T_n \cap (T_n - h)| \cdot |T_n \cap (T_n - h')|}. \end{aligned}$$

By Proposition VI.10 we obtain that $\text{Cov}(X(t+h) \otimes X(t), X(s+h') \otimes X(s))$ is equal to

$$\begin{aligned} & \sum_{i \in I} \sum_{j \in I} \text{cum}(X_i(t+h), X_j(t), X_i(s+h'), X_j(s)) \\ & + \mathbb{E} \langle X(t+h), X(s+h') \rangle_{\mathbb{H}} \cdot \mathbb{E} \langle X(t), X(s) \rangle_{\mathbb{H}} + \langle C(t-s+h), C(s-t+h') \rangle_{\text{HS}}. \end{aligned}$$

In an analogous manner to the proof of Theorem IV.13, we define the quantities

$$\begin{aligned} A &:= \delta_n^{4d} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} \sum_{\substack{h \in [\Delta_n \cdot S_K] \cap (\mathbb{T}_n - t) \\ h' \in [\Delta_n \cdot S_K] \cap (\mathbb{T}_n - s)}} e^{i(h-h')^\top \theta} K\left(\frac{h}{\Delta_n}\right) K\left(\frac{h'}{\Delta_n}\right) \\ & \quad \cdot \frac{\mathbb{E} \langle X(t+h), X(s+h') \rangle_{\mathbb{H}} \cdot \mathbb{E} \langle X(s), X(t) \rangle_{\mathbb{H}}}{|T_n \cap (T_n - h)| \cdot |T_n \cap (T_n - h')|}, \\ B &:= \delta_n^{4d} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} \sum_{\substack{h \in [\Delta_n \cdot S_K] \cap (\mathbb{T}_n - t) \\ h' \in [\Delta_n \cdot S_K] \cap (\mathbb{T}_n - s)}} e^{i(h-h')^\top \theta} K\left(\frac{h}{\Delta_n}\right) K\left(\frac{h'}{\Delta_n}\right) \\ & \quad \cdot \frac{\langle C(t-s+h), C(s-t+h') \rangle_{\text{HS}}}{|T_n \cap (T_n - h)| \cdot |T_n \cap (T_n - h')|} \end{aligned}$$

and

$$\begin{aligned} C &:= \delta_n^{4d} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} \sum_{\substack{h \in [\Delta_n \cdot S_K] \cap (\mathbb{T}_n - t) \\ h' \in [\Delta_n \cdot S_K] \cap (\mathbb{T}_n - s)}} e^{i(h-h')^\top \theta} K\left(\frac{h}{\Delta_n}\right) K\left(\frac{h'}{\Delta_n}\right) \\ & \quad \cdot \sum_{i \in I} \sum_{j \in I} \frac{\text{cum}(X_i(t+h), X_j(t), X_i(s+h'), X_j(s))}{|T_n \cap (T_n - h)| \cdot |T_n \cap (T_n - h')|}. \end{aligned}$$

We start with term A . Bounding terms by their norm (using (4.68)) and changing variables, we obtain

$$\begin{aligned}
|A| &\leq \delta_n^{4d} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} \sum_{\substack{h \in [\Delta_n \cdot S_K] \cap (\mathbb{T}_n - t) \\ h' \in [\Delta_n \cdot S_K] \cap (\mathbb{T}_n - s)}} \frac{|\mathbb{E} \langle X(t+h), X(s+h') \rangle_{\mathbb{H}} \cdot \mathbb{E} \langle X(s), X(t) \rangle_{\mathbb{H}}|}{|T_n \cap (T_n - h)| \cdot |T_n \cap (T_n - h')|} \\
&\leq \delta_n^{4d} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} \sum_{\substack{h \in [\Delta_n \cdot S_K] \cap (\mathbb{T}_n - t) \\ h' \in [\Delta_n \cdot S_K] \cap (\mathbb{T}_n - s)}} \frac{\|C(h - h' + t - s)\|_{\text{tr}} \cdot \|C(s - t)\|_{\text{tr}}}{|T_n \cap (T_n - h)| \cdot |T_n \cap (T_n - h')|} \\
&= \delta_n^{4d} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} \|C(s - t)\|_{\text{tr}} \sum_{\substack{h \in [\Delta_n \cdot S_K] \cap (\mathbb{T}_n - t) \\ h' \in [\Delta_n \cdot S_K] \cap (\mathbb{T}_n - s)}} \frac{\|C(h - h' + t - s)\|_{\text{tr}}}{|T_n \cap (T_n - h)| \cdot |T_n \cap (T_n - h')|} \\
&= \delta_n^{4d} \sum_{w \in \mathbb{T}_n - \mathbb{T}_n} \|C(w)\|_{\text{tr}} \sum_{x \in \mathbb{T}_n \cap (\mathbb{T}_n - w)} \sum_{\substack{u \in [\Delta_n \cdot S_K \cap (\mathbb{T}_n - (w+x))] - [\Delta_n \cdot S_K \cap (\mathbb{T}_n - x)] \\ v \in [\Delta_n \cdot S_K \cap (\mathbb{T}_n - x)] \cap \{[\Delta_n \cdot S_K \cap (\mathbb{T}_n - (w+x))] - u\}}} \frac{\|C(u + w)\|_{\text{tr}}}{|T_n \cap (T_n - (u + v))| \cdot |T_n \cap (T_n - v)|},
\end{aligned}$$

where the last equality is obtained through the change of variables $w = s - t, x = s, u = h - h', v = h'$.

By Lemma IV.30, in view of Assumption S(ii), and inflating slightly the sums by dropping the intersections in the summations of u, v we obtain that

$$\begin{aligned}
|A| &\leq_c \frac{\delta_n^{2d} \times \delta_n^{2d}}{|T_n|^2} \sum_{w \in \mathbb{T}_n - \mathbb{T}_n} \|C(w)\|_{\text{tr}} \sum_{x \in \mathbb{T}_n \cap (\mathbb{T}_n - w)} \sum_{u \in \Delta_n(S_K - S_K)} \|C(u + w)\|_{\text{tr}} \\
&\quad \sum_{v \in (\Delta_n S_K) \cap (\Delta_n S_K - u) \cap \delta_n \cdot \mathbb{Z}^d} 1 \\
&\leq_c \frac{\delta_n^{2d}}{|T_n|^2} \sum_{w \in \mathbb{T}_n - \mathbb{T}_n} \|C(w)\|_{\text{tr}} \cdot |T_n \cap (T_n - w)| \\
&\quad \times \sum_{u \in \Delta_n(S_K - S_K) \cap \delta_n \cdot \mathbb{Z}^d} \|C(u + w)\|_{\text{tr}} \cdot |(\Delta_n S_K) \cap (\Delta_n S_K - u)| \\
&\leq_c \frac{\Delta_n^d}{|T_n|} \times \delta_n^d \sum_{w \in \delta_n \cdot \mathbb{Z}^d} \|C(w)\|_{\text{tr}} \times \delta_n^d \sum_{u \in \delta_n \cdot \mathbb{Z}^d} \|C(u)\|_{\text{tr}} = \mathcal{O}\left(\frac{\Delta_n^d}{|T_n|}\right),
\end{aligned}$$

by Assumption C', where we used that $\delta_n^d \sum_t 1_{\{t \in \mathbb{T}_n\}} \sim |\mathbb{T}_n|$ and

$$\delta_n^d \sum_t 1_{\{t \in (\Delta_n \cdot S_K) \cap (\Delta_n \cdot S_K - u) \cap \delta_n \cdot \mathbb{Z}^d\}} \leq 2|\Delta_n S_K| = \mathcal{O}(\Delta_n^d).$$

Now, we shift to term B . Using the change of variables $w := t - s$, the Cauchy-Schwartz inequality, we obtain

$$\begin{aligned} |B| &\leq \delta_n^{4d} \sum_{w \in \mathbb{T}_n - \mathbb{T}_n} \sum_s 1_{\{\mathbb{T}_n \cap (\mathbb{T}_n - w)\}}(s) \sum_{\substack{h \in [\Delta_n \cdot S_K] \cap \delta_n \cdot \mathbb{Z}^d \\ h' \in [\Delta_n \cdot S_K] \cap \delta_n \cdot \mathbb{Z}^d}} K\left(\frac{h}{\Delta_n}\right) K\left(\frac{h'}{\Delta_n}\right) \\ &\quad \cdot \frac{\|C(h+w)\|_{\text{HS}} \|C(h'-w)\|_{\text{HS}}}{|T_n \cap (T_n - h)| \cdot |T_n \cap (T_n - h')|} \\ &\leq c \frac{\delta_n^{3d}}{|\mathbb{T}_n|} \sum_{w \in \mathbb{T}_n - \mathbb{T}_n} \sum_{\substack{h \in [\Delta_n \cdot S_K] \cap \delta_n \cdot \mathbb{Z}^d \\ h' \in [\Delta_n \cdot S_K] \cap \delta_n \cdot \mathbb{Z}^d}} \|C(h+w)\|_{\text{HS}} \|C(h'-w)\|_{\text{HS}}, \end{aligned}$$

where we used that $\delta_n^d \sum_s 1_{\{\mathbb{T}_n \cap (\mathbb{T}_n - w)\}}(s) = \mathcal{O}(|\mathbb{T}_n|)$, and Lemma IV.30 to conclude that $|T_n \cap (T_n - h)| \sim |T_n \cap (T_n - h')| \sim |\mathbb{T}_n|$, uniformly in $h, h' \in \Delta_n \cdot S_K$. Now, with the change of variables $u := h + w$, and expanding the range of summation, we further obtain

$$\begin{aligned} |B| &\leq c \frac{\delta_n^{3d}}{|\mathbb{T}_n|} \sum_{u \in \delta_n \cdot \mathbb{Z}^d} \|C(u)\|_{\text{HS}} \sum_{\substack{h' \in \Delta_n S_K \cap \delta_n \cdot \mathbb{Z}^d \\ h \in \Delta_n S_K \cap (u - (\mathbb{T}_n - \mathbb{T}_n))}} \|C(h' - u + h)\|_{\text{HS}} \\ &\leq \frac{1}{|\mathbb{T}_n|} \left(\sum_{u \in \delta_n \cdot \mathbb{Z}^d} \delta_n^d \|C(u)\|_{\text{HS}} \right) \left(\delta_n^d \sum_{h' \in \Delta_n S_K \cap \delta_n \cdot \mathbb{Z}^d} 1 \right) \left(\sum_{h \in \delta_n \cdot \mathbb{Z}^d} \delta_n^d \|C(h)\|_{\text{HS}} \right) \\ &= \mathcal{O}\left(\frac{\Delta_n^d}{|\mathbb{T}_n|}\right), \end{aligned}$$

in view of Assumption C'.

Finally, we look at term C . An application of Lemma IV.30, again gives us that

$$|C| \leq c \frac{\delta_n^{4d}}{|\mathbb{T}_n|^2} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} \sum_{\substack{h \in [\Delta_n \cdot S_K] \cap (\mathbb{T}_n - t) \\ h' \in [\Delta_n \cdot S_K] \cap (\mathbb{T}_n - s)}} \sum_{i \in I} \sum_{j \in I} \left| \text{cum}(X_i(t+h), X_j(t), X_i(s+h'), X_j(s)) \right|$$

$$\begin{aligned}
&\leq \frac{\delta_n^{4d}}{|T_n|^2} \sum_{t \in \mathbb{T}_n} \sum_{s \in \mathbb{T}_n} \sum_{\substack{h \in \Delta_n \cdot S_K \cap \delta_n \cdot \mathbb{Z}^d \\ h' \in \Delta_n \cdot S_K \cap \delta_n \cdot \mathbb{Z}^d}} \left| \sum_{i \in I} \sum_{j \in I} \text{cum}(X_i(h+t-s), X_j(t-s), X_i(h'), X_j(0)) \right| \\
&\leq \frac{\delta_n^{3d}}{|T_n|} \sum_{w \in \mathbb{T}_n - \mathbb{T}_n} \sum_{\substack{h \in \Delta_n \cdot S_K \cap \delta_n \cdot \mathbb{Z}^d \\ h' \in \Delta_n \cdot S_K \cap \delta_n \cdot \mathbb{Z}^d}} \left| \sum_{i \in I} \sum_{j \in I} \text{cum}(X_i(h+w), X_j(w), X_i(h'), X_j(0)) \right| \\
&\leq_c \frac{\Delta_n^d}{|T_n|} \sup_{h' \in \delta_n \cdot \mathbb{Z}^d} \delta_n^{2d} \sum_{w \in \delta_n \cdot \mathbb{Z}^d} \sum_{h \in \delta_n \cdot \mathbb{Z}^d} \left| \sum_{i \in I} \sum_{j \in I} \text{cum}(X_i(h+w), X_j(w), X_i(h'), X_j(0)) \right| \\
&= \mathcal{O}\left(\frac{\Delta_n^d}{|T_n|}\right),
\end{aligned}$$

where we used that $\delta_n^d \sum_{h' \in \Delta_n \cdot S_K \cap \delta_n \cdot \mathbb{Z}^d} 1 = \mathcal{O}(|\Delta_n|^d)$, the fact that $\delta_n^d \sum_{s \in \mathbb{T}_n \cap \mathbb{T}_n} = \mathcal{O}(|T_n|)$, and Assumption V' . This completes the proof. \square

Proof of Theorem IV.17: As indicated, by following the proof of Proposition IV.35, we see that the variance bound $\mathcal{O}(\Delta_n^d/|T_n|)$ is uniform in $f \in \mathcal{P}_D(\beta, L)$. Therefore, to prove (4.34), by Relation (4.30) in Theorem IV.16, it is enough to bound terms $B_1(\Delta_n)$ and $B_2(\Delta_n)$, uniformly in $f \in \mathcal{P}_D(\beta, L)$. Let M_k and m_k be the radii of the smallest ball that contains S_K and the largest ball contained in S_K respectively. Starting with term B_2 we have,

$$\begin{aligned}
\sup_{f \in \mathcal{P}_D(\beta, L)} B_2(\Delta_n) &\leq \sum_{\|h\|_2 \geq \Delta_n m_k} \sup_{f \in \mathcal{P}_D(\beta, L)} \|C(h)\|_{\text{HS}} \|h\|_2^\beta \|h\|_2^{-\beta} \\
&\leq (\Delta_n m_k)^{-\beta} \sum_{\|h\|_2 \geq \Delta_n m_k} \sup_{f \in \mathcal{P}_D(\beta, L)} \|C(h)\|_{\text{HS}} \|h\|_2^\beta \\
&\leq (\Delta_n m_k)^{-\beta} \cdot L = \mathcal{O}(\Delta_n^{-\beta}),
\end{aligned}$$

in view of (4.32). Recalling that $K(0) = 1$, we have that for $\lambda + 1 > \beta$,

$$\begin{aligned}
\sup_{f \in \mathcal{P}_D(\beta, L)} B_1(\Delta_n) &\leq \sum_{0 < \|h\|_2 \leq \Delta_n M_K} \sup_{f \in \mathcal{P}_D(\beta, L)} \|C(h)\|_{\text{HS}} \left[1 - K\left(\frac{h}{\Delta_n}\right) \right] \\
&\leq \tilde{c} \sum_{0 < \|h\|_2 \leq \Delta_n M_K} \sup_{f \in \mathcal{P}_D(\beta, L)} \|C(h)\|_{\text{HS}} \frac{\|h\|_2^{\lambda+1}}{\Delta_n^{\lambda+1}}
\end{aligned}$$

$$\leq \tilde{c} \Delta_n^{-\beta} M_K^{\lambda+1-\beta} \sum_{0 < \|h\|_2 \leq \Delta_n M_K} \sup_{f \in \mathcal{P}_D(\beta, L)} \|C(h)\|_{\text{HS}} \|h\|_2^\beta = \mathcal{O}(\Delta_n^{-\beta}),$$

where the second inequality follows from the multivariate Taylor Theorem since (4.33) holds. Indeed, under this condition and using that $K(0) = 1$, we obtain

$$K\left(\frac{h}{\Delta_n}\right) = 1 + R_{0,\lambda}\left(\frac{h}{\Delta_n}\right),$$

where $|R_{0,\lambda}(h)| \leq \frac{L_1}{(\lambda+1)!} \|h\|_2^{\lambda+1}$.

Collecting the bounds for $B_1(\Delta_n)$ and $B_2(\Delta_n)$, we obtain that the bias is of order $\mathcal{O}(\Delta_n^\beta)$, uniformly over the class $\mathcal{P}_D(\beta, L)$. Now, by Theorem IV.16, the variance is of order $\mathcal{O}(\Delta_n^d/|T_n|)$ and picking $\Delta_n = |T_n|^{1/(2\beta+d)}$, we obtain the rate-optimal bound in (4.35). \square

Proof of Theorem IV.18. In view of Theorem IV.27, one only needs to bound the terms $B_1(\Delta_n)$ and $B_2(\Delta_n)$ appropriately. Starting with term B_2 , if m_K denotes the radius of the largest ball contained in S_K , we have

$$\begin{aligned} B_2(\Delta_n) &\leq \int_{x \notin \Delta_n \cdot S_K} \|C(x)\|_{\text{HS}} dx \\ &\leq (\Delta_n \cdot m_K)^{-\beta} \cdot \int_{\|x\| > \Delta_n \cdot m_K} \|x\|_2^\beta \cdot \|C(x)\|_{\text{HS}} dx \\ &\leq (\Delta_n \cdot m_K)^{-\beta} \cdot L = \mathcal{O}(\Delta_n^{-\beta}). \end{aligned}$$

Next, recall that

$$B_1(\Delta_n) := \left\| \int_{h \in \Delta_n \cdot S_K} e^{-ih^\top \theta} C(h) \left(1 - K\left(\frac{h}{\Delta_n}\right)\right) dh \right\|_{\text{HS}}.$$

Since $K(0) = 1$ and (4.33) holds, by the Taylor theorem, we have that

$$K\left(\frac{h}{\Delta_n}\right) = 1 + R_{0,\lambda}\left(\frac{h}{\Delta_n}\right),$$

where $|R_{0,\lambda}(h)| \leq \frac{L_1}{(\lambda+1)!} \|h\|_2^{\lambda+1}$. Thus, with M_K denoting the radius of the smallest ball centered at the origin that contains S_K , the term B_1 is bounded by

$$\begin{aligned} B_1(\Delta_n) &\leq \frac{L_1}{(\lambda+1)!} \int_{h \in \Delta_n \cdot S_K} \|C(h)\|_{\text{HS}} \left(\frac{\|h\|_2}{\Delta_n}\right)^{\lambda+1} dh \\ &\leq \frac{L_1 \cdot (M_K \cdot \Delta_n)^{\lambda+1-\beta}}{(\lambda+1)! \cdot \Delta_n^{\lambda+1}} \int_{h \in \Delta_n \cdot S_K} \|C(h)\|_{\text{HS}} \|h\|_2^\beta dh = \mathcal{O}(\Delta_n^{-\beta}), \end{aligned}$$

since $\lambda + 1 > \beta$ and in view of (4.36).

Collecting the bounds for B_1 and B_2 , we get $B_1(\Delta_n) + B_2(\Delta_n) = \mathcal{O}(\Delta_n^{-\beta})$. Now, the optimal choice of Δ_n is the one which balances the last bound with the rate of the variance, that is, $\Delta_n^d/|T_n| \sim \Delta_n^{-2\beta}$. This is achieved with

$$\Delta_n := |T_n|^{1/(2\beta+d)} \equiv (n\delta_n)^{d/(2\beta+d)},$$

which upon substitution yields the rate $\delta_n^\gamma \vee \Delta_n^{-\beta} = \delta_n^\gamma \vee (n\delta_n)^{-\beta d/(2\beta+d)}$ in (4.37). \square

Proof of Theorem IV.21: Following the proof of Theorem IV.17, we only need bound the terms $B_1(\Delta_n)$ and $B_2(\Delta_n)$ (in Theorem IV.16) uniformly in $f \in \mathcal{EP}_D(\eta, L)$. Let M_k and m_k be the radii of the smallest ball that contains S_K and the largest ball contained in S_K respectively. Starting with term B_2 we have,

$$\begin{aligned} \sup_{f \in \mathcal{EP}_D(\eta, L)} B_2(\Delta_n) &\leq \sum_{\|h\|_2 \geq \Delta_n m_k} \sup_{f \in \mathcal{EP}_D(\eta, L)} \|C(h)\|_{\text{HS}} \\ &\leq \sum_{\|h\|_2 \geq \Delta_n m_k} \sup_{f \in \mathcal{EP}_D(\eta, L)} L \cdot e^{-\|h\|_2^2} \\ &\leq L \cdot \int_{\|h\|_2 \geq (\Delta_n - 1)m_k} e^{-\|h\|_2^2} dh, \end{aligned}$$

where using hyperspherical coordinates, we obtain that

$$\sup_{f \in \mathcal{EP}_D(\eta, L)} B_2(\Delta_n) \leq L \cdot \frac{\pi^{d/2}}{\Gamma(d/2 + 1)} \cdot \frac{1}{\eta} \cdot \Gamma\left(\frac{d}{\eta}, [m_k(\Delta_n - 1)]^\eta\right),$$

with $\Gamma(s, x)$ being the upper incomplete Gamma function. Now, for term $B_1(\Delta_n)$,

$$\begin{aligned} \sup_{f \in \mathcal{EP}_D(\eta, L)} B_1(\Delta_n) &= \sup_{f \in \mathcal{EP}_D(\eta, L)} \left\| \sum_{h \in \Delta_n S_K \cap \mathbb{Z}^d} e^{ih^\top \theta} C(h) \left(1 - K\left(\frac{h}{\Delta_n}\right)\right) \right\|_{\text{HS}} \\ &\leq \sup_{f \in \mathcal{EP}_D(\eta, L)} \sum_{h \in \Delta_n S_K \cap \mathbb{Z}^d} \|C(h)\|_{\text{HS}} \left|1 - K\left(\frac{h}{\Delta_n}\right)\right| \\ &\leq \sup_{f \in \mathcal{EP}_D(\eta, L)} \left(\sum_{\|h\|_2 \leq \alpha} \|C(0)\|_{\text{HS}} \left|1 - K\left(\frac{h}{\Delta_n}\right)\right| \right. \\ &\quad \left. + \sum_{\alpha \leq \|h\|_2 \leq \Delta_n M_k} L \cdot e^{-\|h\|_2^\eta} \left|1 - K\left(\frac{h}{\Delta_n}\right)\right| \right) \\ &\leq L \cdot \sum_{\|h\|_2 \leq \alpha} \left|1 - K\left(\frac{h}{\Delta_n}\right)\right| + L \cdot \sum_{\alpha \leq \|h\|_2 \leq \Delta_n M_k} e^{-\|h\|_2^\eta} \left|1 - K\left(\frac{h}{\Delta_n}\right)\right|, \end{aligned}$$

for some positive fixed constant $0 < \alpha < 1$. Recalling that $K(0) = 1$ and since we are using a kernel satisfying Condition (4.40), there is a positive constant $\tilde{c}_K > 0$ such that

$$\begin{aligned} \sup_{f \in \mathcal{EP}_D(\eta, L)} B_1(\Delta_n) &\leq \tilde{c}_K \cdot L \cdot \sum_{\|h\|_2 \leq \alpha} e^{-\frac{\Delta_n}{\|h\|_2}} + \tilde{c}_K \cdot L \cdot \sum_{\alpha \leq \|h\|_2 \leq \Delta_n M_k} e^{-\|h\|_2^\eta} e^{-\frac{\Delta_n}{\|h\|_2}} \\ &\leq \tilde{c}_K \cdot L \cdot \int_{\|h\|_2 \leq \alpha} e^{-\frac{\Delta_n}{\|h\|_2}} dh + \tilde{c}_K \cdot L \cdot \int_{\alpha \leq \|h\|_2 \leq \Delta_n M_k} e^{-\|h\|_2^\eta - \frac{\Delta_n}{\|h\|_2}} dh \\ &\leq \pi^{d-1} \cdot L \cdot \tilde{c}_K \cdot \left[\Delta_n^d \cdot \Gamma\left(-d, \frac{\Delta_n}{\alpha}\right) + \int_{r=\alpha}^{\Delta_n \cdot M_K} r^{d-1} e^{-r^\eta - \frac{\Delta_n}{r}} dr \right], \end{aligned}$$

where we used hyperspherical coordinates in the last integral. We only need to bound the remaining integral of the upper bound. We discern cases for η . Let $\eta > d$ and

$X \sim Exp(1)$. We split the integral in two regions. We have,

$$\int_{r=\alpha}^1 r^{d-1} e^{-r^\eta - \frac{\Delta_n}{r}} dr \leq \int_{r=\alpha}^1 r^{d-1} e^{-\frac{\Delta_n}{r}} dr = \Delta_n^d \left[\Gamma(-d, \Delta_n) - \Gamma\left(-d, \frac{\Delta_n}{\alpha}\right) \right].$$

Also, for $v = \eta/(\eta + 1)$

$$\begin{aligned} \int_{r=1}^{\Delta_n \cdot M_K} r^{d-1} e^{-r^\eta - \frac{\Delta_n}{r}} dr &= \int_{x=1}^{(\Delta_n \cdot M_K)^\eta} \frac{1}{\eta} x^{\frac{d}{\eta}-1} e^{-x - \frac{\Delta_n}{x^{1/\eta}}} dx \leq \int_{x=1}^{(\Delta_n \cdot M_K)^\eta} \frac{1}{\eta} e^{-x - \frac{\Delta_n}{x^{1/\eta}}} dx \\ &\leq \int_{x=1}^{\Delta_n^v} \frac{1}{\eta} e^{-x - \frac{\Delta_n}{x^{1/\eta}}} dx + \int_{x=\Delta_n^v}^{(\Delta_n \cdot M_K)^\eta} \frac{1}{\eta} e^{-x - \frac{\Delta_n}{x^{1/\eta}}} dx \\ &\leq \frac{1}{\eta} e^{-\Delta_n^{1-\frac{v}{\eta}}} \mathbb{P}[X \leq \Delta_n^v] + \frac{1}{\eta} \int_{x=\Delta_n^v}^{(\Delta_n \cdot M_K)^\eta} e^{-x} dx \\ &\leq \frac{1}{\eta} e^{-\Delta_n^{1-\frac{v}{\eta}}} \mathbb{P}[X \leq \Delta_n^v] + \frac{1}{\eta} \left[e^{-\Delta_n^v} - e^{-(\Delta_n \cdot M_K)^\eta} \right] \\ &= \mathcal{O}\left(e^{-\Delta_n^{\frac{\eta}{\eta+1}}}\right). \end{aligned}$$

Consider now $0 < \eta \leq d$. Then, we have that

$$\begin{aligned} \int_{r=\alpha}^{\Delta_n \cdot M_K} r^{d-1} e^{-r^\eta - \frac{\Delta_n}{r}} dr &= \int_{x=\alpha^\eta}^{(\Delta_n \cdot M_K)^\eta} \frac{1}{\eta} x^{\frac{d}{\eta}-1} e^{-x - \frac{\Delta_n}{x^{1/\eta}}} dx \leq \frac{1}{\eta} \mathbb{E}\left[X^{\frac{d}{\eta}-1} e^{-\frac{\Delta_n}{X^{1/\eta}}}\right] \\ &\leq \frac{1}{\eta} \Delta_n^{v \cdot (\frac{d}{\eta}-1)} e^{-\Delta_n^{1-\frac{v}{\eta}}} \mathbb{P}[X \leq \Delta_n^v] + \frac{1}{\eta} \mathbb{E}\left[X^{\frac{d}{\eta}-1} \mathbb{1}(X \geq \Delta_n^v)\right] \\ &\leq \frac{1}{\eta} \Delta_n^{v \cdot (\frac{d}{\eta}-1)} e^{-\Delta_n^{1-\frac{v}{\eta}}} + \frac{1}{\eta} \int_{x=\Delta_n^v}^{\infty} x^{\frac{d}{\eta}-1} e^{-x} dx \\ &= \frac{1}{\eta} \Delta_n^{v \cdot (\frac{d}{\eta}-1)} e^{-\Delta_n^{1-\frac{v}{\eta}}} + \frac{1}{\eta} \cdot \Gamma\left(\frac{d}{\eta}, \Delta_n^v\right) \\ &\sim \frac{1}{\eta} \Delta_n^{v \cdot (\frac{d}{\eta}-1)} e^{-\Delta_n^{1-\frac{v}{\eta}}} + \frac{1}{\eta} (\Delta_n^v)^{\frac{d}{\eta}-1} e^{-\Delta_n^v} \end{aligned}$$

The asymptotic relationship

$$\frac{\Gamma(s, x)}{x^{s-1} e^{-x}} \rightarrow 1, \quad \text{as } x \rightarrow \infty,$$

entails that

$$\begin{aligned} B_1(\Delta_n) + B_2(\Delta_n) &= \mathcal{O} \left(\Delta_n^{\frac{[d-\eta]_+}{\eta+1}} e^{-\Delta_n^{\frac{\eta}{\eta+1}}} + \Delta_n^{d-\eta} e^{-(m_K \Delta_n)^\eta} \right) \\ &= \mathcal{O} \left(\Delta_n^{\frac{[d-\eta]_+}{\eta+1}} e^{-\Delta_n^{\frac{\eta}{\eta+1}}} \right), \end{aligned}$$

and the proof is complete. \square

4.8.3 Basis Independence of Second-order Stationarity

In this section, we demonstrate that the definition of second-order stationarity for a process in \mathbb{H} (recall Definition IV.1) is independent of the CONS under consideration. Namely, we show that whether $C(t, s)$ and $\check{C}(t, s)$, as defined in (4.2) and (4.4) respectively, are a function of the lag $t - s$, is independent of the basis we consider.

Let \mathbb{H} be a complex Hilbert space and $\{e_i, i = 1, 2, \dots\}, \{f_j, j = 1, 2, \dots\}$ be two CONS of \mathbb{H} . Using $\{e_i\}$ we create the “real” Hilbert space

$$\mathbb{H}_{\mathbb{R}} := \left\{ h = \sum h_i e_i, h_i \in \mathbb{R}, \sum h_i^2 < \infty \right\}.$$

Thus, $\{e_i\}$ consists a “real” basis of the space

$$\mathbb{H} = \mathbb{H}_{\mathbb{R}} + \mathbf{i}\mathbb{H}_{\mathbb{R}}.$$

In this setting, we can express the CONS $\{f_j\}$ as

$$\begin{aligned} f_j &= \sum_i [\operatorname{Re}(\langle f_j, e_i \rangle) + \mathbf{i}\operatorname{Im}(\langle f_j, e_i \rangle)] e_i \\ &= \sum_i [f_{j,i}^R + \mathbf{i}f_{j,i}^I] e_i, \end{aligned}$$

where $f_{j,i}^R, f_{j,i}^I \in \mathbb{R}$. We also have that

$$\begin{aligned} X(t) &= \sum_i [\operatorname{Re}(\langle X(t), e_i \rangle) + \mathfrak{i} \operatorname{Im}(\langle X(t), e_i \rangle)] e_i \\ &= \sum_j [\operatorname{Re}(\langle X(t), f_j \rangle) + \mathfrak{i} \operatorname{Im}(\langle X(t), f_j \rangle)] f_j. \end{aligned}$$

Now,

$$\begin{aligned} \langle X(t), f_j \rangle &= \langle X(t), \sum_i [f_{j,i}^R + \mathfrak{i} f_{j,i}^I] e_i \rangle = \sum_i [f_{j,i}^R - \mathfrak{i} f_{j,i}^I] \langle X(t), e_i \rangle \\ &= \sum_i [f_{j,i}^R - \mathfrak{i} f_{j,i}^I] [\operatorname{Re}(\langle X(t), e_i \rangle) + \mathfrak{i} \operatorname{Im}(\langle X(t), e_i \rangle)] \\ &= \sum_i f_{j,i}^R \operatorname{Re}(\langle X(t), e_i \rangle) + f_{j,i}^I \operatorname{Im}(\langle X(t), e_i \rangle) \\ &\quad + \mathfrak{i} [f_{j,i}^R \operatorname{Im}(\langle X(t), e_i \rangle) - f_{j,i}^I \operatorname{Re}(\langle X(t), e_i \rangle)]. \end{aligned}$$

So,

$$\begin{aligned} \operatorname{Re}(\langle X(t), f_j \rangle) &= \sum_i [f_{j,i}^R \operatorname{Re}(\langle X(t), e_i \rangle) + f_{j,i}^I \operatorname{Im}(\langle X(t), e_i \rangle)] \\ \operatorname{Im}(\langle X(t), f_j \rangle) &= \sum_i [-f_{j,i}^I \operatorname{Re}(\langle X(t), e_i \rangle) + f_{j,i}^R \operatorname{Im}(\langle X(t), e_i \rangle)]. \end{aligned}$$

Let

$$\begin{aligned} a_i^e &= (e_i, 0) & b_i^e &= (0, e_i) \\ a_j^f &= (f_j, 0) & b_j^f &= (0, f_j) \end{aligned}$$

and we have that

$$Y(t) = \sum_i \operatorname{Re}(\langle X(t), e_i \rangle) a_i^e + \sum_i \operatorname{Im}(\langle X(t), e_i \rangle) b_i^e.$$

Thus,

$$\begin{aligned}
C_Y^{(e)}(t) &= \mathbb{E}Y(t) \otimes Y(0) \\
&= \mathbb{E} \left(\sum_k \operatorname{Re}(\langle X(t), e_k \rangle) a_k^e + \sum_k \operatorname{Im}(\langle X(t), e_k \rangle) b_k^e \right) \\
&\quad \otimes \left(\sum_\ell \operatorname{Re}(\langle X(0), e_\ell \rangle) a_\ell^e + \sum_\ell \operatorname{Im}(\langle X(0), e_\ell \rangle) b_\ell^e \right) \\
&= \mathbb{E} \left[\sum_{k,\ell} \operatorname{Re}(\langle X(t), e_k \rangle) \operatorname{Re}(\langle X(0), e_\ell \rangle) a_k^e \otimes a_\ell^e \right. \\
&\quad + \sum_{k,\ell} \operatorname{Re}(\langle X(t), e_k \rangle) \operatorname{Im}(\langle X(0), e_\ell \rangle) a_k^e \otimes b_\ell^e \\
&\quad + \sum_{k,\ell} \operatorname{Im}(\langle X(t), e_k \rangle) \operatorname{Re}(\langle X(0), e_\ell \rangle) b_k^e \otimes a_\ell^e \\
&\quad \left. + \sum_{k,\ell} \operatorname{Im}(\langle X(t), e_k \rangle) \operatorname{Im}(\langle X(0), e_\ell \rangle) b_k^e \otimes b_\ell^e \right].
\end{aligned}$$

Using the CONS $\{f_j\}$, we similarly have that

$$\begin{aligned}
C_Y^{(f)}(t) &= \mathbb{E} \left[\sum_{k,\ell} \operatorname{Re}(\langle X(t), f_k \rangle) \operatorname{Re}(\langle X(0), f_\ell \rangle) a_k^f \otimes a_\ell^f \right. \\
&\quad + \sum_{k,\ell} \operatorname{Re}(\langle X(t), f_k \rangle) \operatorname{Im}(\langle X(0), f_\ell \rangle) a_k^f \otimes b_\ell^f \\
&\quad + \sum_{k,\ell} \operatorname{Im}(\langle X(t), f_k \rangle) \operatorname{Re}(\langle X(0), f_\ell \rangle) b_k^f \otimes a_\ell^f \\
&\quad \left. + \sum_{k,\ell} \operatorname{Im}(\langle X(t), f_k \rangle) \operatorname{Im}(\langle X(0), f_\ell \rangle) b_k^f \otimes b_\ell^f \right].
\end{aligned}$$

For the coordinates, we have

$$\begin{aligned}
\operatorname{Re}(\langle X(t), f_k \rangle) \operatorname{Re}(\langle X(0), f_\ell \rangle) &= \sum_{i,j} f_{k,i}^R f_{\ell,j}^R \operatorname{Re}(\langle X(t), e_i \rangle) \operatorname{Re}(\langle X(0), e_j \rangle) \\
&\quad + \sum_{i,j} f_{k,i}^R f_{\ell,j}^I \operatorname{Re}(\langle X(t), e_i \rangle) \operatorname{Im}(\langle X(0), e_j \rangle)
\end{aligned}$$

$$\begin{aligned}
& + \sum_{i,j} f_{k,i}^I f_{\ell,j}^R \operatorname{Im}(\langle X(t), e_i \rangle) \operatorname{Re}(\langle X(0), e_j \rangle) \\
& + \sum_{i,j} f_{k,i}^I f_{\ell,j}^I \operatorname{Im}(\langle X(t), e_i \rangle) \operatorname{Im}(\langle X(0), e_j \rangle) \\
\operatorname{Re}(\langle X(t), f_k \rangle) \operatorname{Im}(\langle X(0), f_\ell \rangle) & = - \sum_{i,j} f_{k,i}^R f_{\ell,j}^I \operatorname{Re}(\langle X(t), e_i \rangle) \operatorname{Re}(\langle X(0), e_j \rangle) \\
& + \sum_{i,j} f_{k,i}^R f_{\ell,j}^R \operatorname{Re}(\langle X(t), e_i \rangle) \operatorname{Im}(\langle X(0), e_j \rangle) \\
& - \sum_{i,j} f_{k,i}^I f_{\ell,j}^I \operatorname{Im}(\langle X(t), e_i \rangle) \operatorname{Re}(\langle X(0), e_j \rangle) \\
& + \sum_{i,j} f_{k,i}^I f_{\ell,j}^R \operatorname{Im}(\langle X(t), e_i \rangle) \operatorname{Im}(\langle X(0), e_j \rangle) \\
\operatorname{Im}(\langle X(t), f_k \rangle) \operatorname{Re}(\langle X(0), f_\ell \rangle) & = - \sum_{i,j} f_{k,i}^I f_{\ell,j}^R \operatorname{Re}(\langle X(t), e_i \rangle) \operatorname{Re}(\langle X(0), e_j \rangle) \\
& - \sum_{i,j} f_{k,i}^R f_{\ell,j}^I \operatorname{Re}(\langle X(t), e_i \rangle) \operatorname{Im}(\langle X(0), e_j \rangle) \\
& + \sum_{i,j} f_{k,i}^R f_{\ell,j}^R \operatorname{Im}(\langle X(t), e_i \rangle) \operatorname{Re}(\langle X(0), e_j \rangle) \\
& + \sum_{i,j} f_{k,i}^R f_{\ell,j}^I \operatorname{Im}(\langle X(t), e_i \rangle) \operatorname{Im}(\langle X(0), e_j \rangle) \\
\operatorname{Im}(\langle X(t), f_k \rangle) \operatorname{Im}(\langle X(0), f_\ell \rangle) & = \sum_{i,j} f_{k,i}^I f_{\ell,j}^I \operatorname{Re}(\langle X(t), e_i \rangle) \operatorname{Re}(\langle X(0), e_j \rangle) \\
& - \sum_{i,j} f_{k,i}^I f_{\ell,j}^R \operatorname{Re}(\langle X(t), e_i \rangle) \operatorname{Im}(\langle X(0), e_j \rangle) \\
& - \sum_{i,j} f_{k,i}^R f_{\ell,j}^I \operatorname{Im}(\langle X(t), e_i \rangle) \operatorname{Re}(\langle X(0), e_j \rangle) \\
& + \sum_{i,j} f_{k,i}^R f_{\ell,j}^R \operatorname{Im}(\langle X(t), e_i \rangle) \operatorname{Im}(\langle X(0), e_j \rangle).
\end{aligned}$$

Let now

$$\begin{aligned}
C_1^{(e)}(t) & = \left(\mathbb{E} \operatorname{Re}(\langle X(t), e_i \rangle) \operatorname{Re}(\langle X(0), e_j \rangle) \right)_{i,j \in \mathbb{N}} \\
C_2^{(e)}(t) & = \left(\mathbb{E} \operatorname{Re}(\langle X(t), e_i \rangle) \operatorname{Im}(\langle X(0), e_j \rangle) \right)_{i,j \in \mathbb{N}} \\
C_3^{(e)}(t) & = \left(\mathbb{E} \operatorname{Im}(\langle X(t), e_i \rangle) \operatorname{Re}(\langle X(0), e_j \rangle) \right)_{i,j \in \mathbb{N}}
\end{aligned}$$

$$C_4^{(e)}(t) = \left(\mathbb{E} \text{Im} (\langle X(t), e_i \rangle) \text{Im} (\langle X(0), e_j \rangle) \right)_{i,j \in \mathbb{N}}$$

and

$$\begin{aligned} f_k^R &= \left(f_{k,i}^R \right)_{i \in \mathbb{N}}, & f_k^I &= \left(f_{k,i}^I \right)_{i \in \mathbb{N}} \\ F_R &= (f_1^R, f_2^R, \dots), & F_I &= (f_1^I, f_2^I, \dots) \end{aligned}$$

Denoting

$$\tilde{C}_Y^{(e)}(t) = \begin{pmatrix} C_1^{(e)}(t) & C_2^{(e)}(t) \\ C_3^{(e)}(t) & C_4^{(e)}(t) \end{pmatrix}$$

one can see that all the coordinates taken into account in the Hilbert-Schmidt norm can be found in the matrix

$$\begin{pmatrix} F_R^\top & F_I^\top \\ -F_I^\top & F_R^\top \end{pmatrix} \tilde{C}_Y^{(e)}(t) \begin{pmatrix} F_R & -F_I \\ F_I & F_R \end{pmatrix}.$$

Thus, if $C_Y^{(e)}(t)$ (and of course $C_X^{(e)}(t)$) depends only on the lag t , then the same is true for $C_Y^{(f)}(t)$ as well. Similarly, one can show that this holds for the pseudo-covariance as well.

4.8.4 Proofs for Section 4.6

As in Section 4.6, \mathbb{H}_n denotes the space spanned by $\{R(u, \cdot), u \in D_n\}$ and Π_n is the projection operator onto \mathbb{H}_n . Although the following result is standard, we include it here for the sake of completeness.

Proposition IV.36. *Assume that the matrix $\mathbf{R}_n = \{R(u_{n,i}, u_{n,j})\}_{i,j=1}^{m_n}$ is invertible. Let $g \in \mathbb{H}$ and $\mathbf{g} = (g(u_{n,1}), \dots, g(u_{n,m_n}))^\top$. Then, the following hold.*

- (i) *The projection $\tilde{g} = \Pi_n g = \sum_i c_i R(u_{n,i}, \cdot)$ where $\mathbf{c} := (c_1, \dots, c_{m_n})^\top = \mathbf{R}_n^{-1} \mathbf{g}$,*

and $\tilde{g}(u_{n,i}) = g(u_{n,i})$ for all $u_{n,i} \in D_n$. Moreover, $\|g - \tilde{g}\|_{\mathbb{H}}^2 = \|g\|_{\mathbb{H}}^2 - \mathbf{g}^\top \mathbf{R}_n^{-1} \mathbf{g}$.

$$(ii) \quad |\tilde{g}(u) - g(u)| \leq \|g\|_{\mathbb{H}} \inf_{u' \in D_n} \sqrt{R(u, u) - 2R(u, u') + R(u', u')}.$$

Proof.

(i) By the property of projection,

$$\tilde{g} = \operatorname{argmin}_{h \in \mathbb{H}_n} \|g - h\|_{\mathbb{H}}.$$

For $h = \sum_i c_i R(u_{n,i}, \cdot)$, the reproducing property entails

$$\|g - h\|_{\mathbb{H}}^2 = \|g\|_{\mathbb{H}}^2 - 2\mathbf{c}^\top \mathbf{g} + \mathbf{c}^\top \mathbf{R}_n \mathbf{c},$$

from which we conclude the minimizer \mathbf{c} is $\mathbf{R}_n^{-1} \mathbf{g}$. It then follows that

$$\tilde{\mathbf{g}} := (\tilde{g}(u_{n,1}), \dots, \tilde{g}(u_{n,m_n}))^\top = \mathbf{R}_n \mathbf{c} = \mathbf{g}.$$

(ii) Applying again the fact that $g - \tilde{g} \perp R(u', \cdot)$ for all $u' \in D_n$, we have for any arbitrary $u \in E$,

$$\tilde{g}(u) - g(u) = \langle \tilde{g} - g, R(u, \cdot) \rangle_{\mathbb{H}} = \langle \tilde{g} - g, R(u, \cdot) - R(u', \cdot) \rangle_{\mathbb{H}}, \quad u' \in D_n.$$

By (i) and the Cauchy-Schwarz inequality

$$|\tilde{g}(u) - g(u)| \leq \|g\|_{\mathbb{H}} \inf_{u' \in D_n} \|R(u, \cdot) - R(u', \cdot)\|_{\mathbb{H}},$$

where

$$\|R(u, \cdot) - R(u', \cdot)\|_{\mathbb{H}}^2 = R(u, u) - 2R(u, u') + R(u', u').$$

□

Proof of Theorem IV.26: First,

$$\begin{aligned}
\|\tilde{f}(\theta) - f(\theta)\|_{\text{HS}} &= \|\Pi_n f(\theta)\Pi_n - f(\theta)\|_{\text{HS}} \\
&\leq \|\Pi_n f(\theta)\Pi_n - \Pi_n f(\theta)\|_{\text{HS}} + \|\Pi_n f(\theta) - f(\theta)\|_{\text{HS}} \\
&\leq \|f(\theta)(\Pi_n - \mathbf{I})\|_{\text{HS}} + \|(\Pi_n - \mathbf{I})f(\theta)\|_{\text{HS}} \\
&= 2\|(\Pi_n - \mathbf{I})f(\theta)\|_{\text{HS}} \\
&= 2\left(\sum_{j=1}^{\infty} \nu_j^2 \|(\Pi_n - \mathbf{I})\phi_j\|_{\mathbb{H}}^2\right)^{1/2}.
\end{aligned}$$

Next, we consider $\|(\Pi_n - \mathbf{I})g\|_{\mathbb{H}}^2 = \|\tilde{g} - g\|_{\mathbb{H}}^2$ for a function $g \in \mathbb{H}$ with a Lipschitz continuous derivative. The derivation of this depends little on the value of $g(0)$. To simplify notation, let us make the simplification that the Sobolev space contains functions g with $g(0) = 0$. Thus, we take the kernel as $R(s, t) = s \wedge t$, i.e., the covariance kernel of the standard Brownian motion. Then the matrix \mathbf{R}_n in (4.46) is indeed invertible. By Proposition IV.36,

$$\|\tilde{g} - g\|_{\mathbb{H}}^2 = \|g\|_{\mathbb{H}}^2 - \mathbf{g}^\top \mathbf{R}_n^{-1} \mathbf{g} \quad (4.74)$$

where $\mathbf{g} = (g(u_{n,i}))_{i=1}^{m_n}$ contains the values of g at the $u_{n,i}$. It follows that \mathbf{R}_n has the Cholesky decomposition

$$\mathbf{R}_n = m_n^{-1} \mathbf{L}_n \mathbf{L}_n^\top, \quad (4.75)$$

where \mathbf{L}_n is a lower triangular matrix of 1's and has inverse

$$\mathbf{L}_n^{-1} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 & 0 \\ -1 & 1 & 0 & \cdots & 0 & 0 \\ 0 & -1 & 1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix}.$$

Indeed, by the independence and the stationarity of the increments of the standard Brownian motion B , we have $\mathbf{Z} = \sqrt{m_n}(\mathbf{L}_n^{-1})^\top \mathbf{B}$, where $\mathbf{B} = \left(B(i/m_n) - B((i-1)/m_n) \right)_{i=1}^{m_n}$ and $\mathbf{Z} \sim \mathcal{N}(0, \mathbf{I}_{m_n})$ is a standard Normal random vector. Since $\mathbf{R}_n = \mathbb{E}[\mathbf{B}\mathbf{B}^\top]$, we obtain $\mathbf{I}_{m_n} = m_n(\mathbf{L}_n^{-1})^\top \mathbf{R}_n \mathbf{L}_n^{-1}$, which yields (4.75). Thus,

$$\mathbf{g}^\top \mathbf{R}_n^{-1} \mathbf{g} = m_n \mathbf{g}^\top (\mathbf{L}_n^{-1})^\top \mathbf{L}_n^{-1} \mathbf{g} = m_n \sum_{i=1}^{m_n} (g(i/m_n) - g((i-1)/m_n))^2, \quad (4.76)$$

which is a Riemann approximation of $\|g\|_{\mathbb{H}}^2 = \int_0^1 (g'(t))^2 dt$ (recall $g(0) = 0$). Since $|g'(s) - g'(t)| \leq C|s - t|$, it follows from (4.74) and (4.76) that

$$\|\tilde{g} - g\|_{\mathbb{H}}^2 \leq C m_n^{-1}.$$

Indeed, by the mean value theorem, we have $g(i/m_n) - g((i-1)/m_n) = g'(\xi_{n,i}) m_n^{-1}$, for some $\xi_{n,i} \in [(i-1)/m_n, i/m_n]$, and hence

$$\begin{aligned} \|g - \tilde{g}\|_{\mathbb{H}}^2 &= \int_0^1 (g'(t))^2 dt - \frac{1}{m_n} \sum_{i=1}^{m_n} (g'(\xi_{n,i}))^2 \\ &\leq \sum_{i=1}^{m_n} \int_{(i-1)/m_n}^{i/m_n} |g'(t) - g'(\xi_{n,i})| \cdot |g'(t) + g'(\xi_{n,i})| dt \\ &\leq \frac{C}{m_n} \left(\int_0^1 |g'(t)| dt + \frac{1}{m_n} \sum_{i=1}^{m_n} |g'(\xi_{n,i})| \right) = \mathcal{O}(m_n^{-1}), \end{aligned}$$

where in the last relation we used the fact that the Riemann sum converges to the integral $\int_0^1 |g'(t)| dt < \infty$, as $m_n \rightarrow \infty$. Applying this bound and by the assumption on the ϕ_j , we obtain

$$\sum_{j=1}^{\infty} \nu_j^2 \|(\Pi_n - \mathbf{I})\phi_j\|_{\mathbb{H}}^2 \leq m_n^{-1} \sum_{j=1}^{\infty} C_j \nu_j^2.$$

This completes the proof. □

4.8.5 Some properties of the trace norm

We collect some elementary facts of the trace norm in the following lemma.

Lemma IV.37. *Let \mathcal{A} be a trace class operator on the Hilbert space \mathbb{H} . Then*

- (i) $\|\mathcal{A}\|_{\text{tr}} = \sup_{\mathcal{W}: \text{unitary}} |\langle \mathcal{A}, \mathcal{W} \rangle_{\text{HS}}|$;
- (ii) $\sum_i |\langle \mathcal{A}f_i, g_i \rangle| \leq \|\mathcal{A}\|_{\text{tr}}$ for any CONS $\{f_i\}$ and $\{g_i\}$;
- (iii) $\sum_i |\langle \mathcal{A}e_i, e_i \rangle| \leq \|\mathcal{A}\|_{\text{tr}}$ for any CONS $\{e_i\}$.

Proof. (i) Suppose A has the SVD

$$\mathcal{A} = \sum_j \lambda_j v_j \otimes w_j, \tag{4.77}$$

where $\lambda_j \geq 0$ and $\{v_j\}, \{w_j\}$ are CONS of \mathbb{H} . Then, we can write

$$\mathcal{A} = \left(\sum_j \lambda_j v_j \otimes v_j \right) \left(\sum_k v_k \otimes w_k \right) =: \mathcal{P}\mathcal{U},$$

which is a polar decomposition of \mathcal{A} . It follows that

$$\|\mathcal{A}\|_{\text{tr}} = \text{trace}(\mathcal{P}) = \text{trace}(\mathcal{A}\mathcal{U}^*) = \langle \mathcal{A}, \mathcal{U} \rangle_{\text{HS}}.$$

Suppose \mathcal{W} is unitary and has the SVD $\mathcal{W} = \sum_k a_k \otimes b_k$. Then

$$\begin{aligned} |\langle \mathcal{A}, \mathcal{W} \rangle_{\text{HS}}| &= \left| \sum_k \langle \mathcal{A} b_k, \mathcal{W} b_k \rangle \right| \\ &= \left| \sum_j \sum_k \lambda_j \langle (v_j \otimes w_j) b_k, a_k \rangle \right| \\ &= \left| \sum_j \lambda_j \sum_k \langle v_j, a_k \rangle \langle b_k, w_j \rangle \right| \leq \sum_j \lambda_j \end{aligned}$$

by the Cauchy-Schwarz inequality.

(ii) By (4.77),

$$\begin{aligned} \sum_i |\langle \mathcal{A} f_i, g_i \rangle| &= \sum_i \left| \sum_j \lambda_j \langle v_j, g_i \rangle \langle w_j, f_i \rangle \right| \\ &\leq \sum_j \lambda_j \sum_i |\langle v_j, g_i \rangle \langle w_j, f_i \rangle|, \end{aligned}$$

and the result again follows from the Cauchy-Schwarz inequality.

(iii) This is a special case of (ii) with $f_i = g_i$.

□

4.8.6 Examples

In this section, we discuss several concrete examples that illustrate the breadth and scope of the conditions imposed in various results in this thesis.

4.8.6.1 An example of the class $\mathcal{P}_D(\beta, L)$

We consider in this section with an example of a class of covariance structures, where the rate of consistency nearly matches the optimal rate of $\mathcal{P}_D(\beta, L)$. This class consists of regularly varying covariance structures, as follows.

Example IV.38. Consider $d = 1$ and the scalar-valued case $\mathbb{H} = \mathbb{C}$. Let

$$C(k) = |k|^{-\beta-1}S(|h|), \beta > 0, k \in \mathbb{Z}$$

where S is a slowly varying function at infinity. It is not hard to see that the corresponding spectral densities $f \in \mathcal{P}_D(\beta + \epsilon, L)$ for any $\epsilon > 0$, depending on the value of L . Also, assume that the kernel function is of the form

$$K(h) = [1 - |h|^{\lambda+1}]_+, h \in \mathbb{R}$$

for some $\lambda > 0$. We work in the discrete time setting, so we are using the estimator $\hat{f}_n(\theta)$.

Thus, we have that

$$(2\pi)[f(\theta) - \mathbb{E}\hat{f}_n(\theta)] = \sum_{|k| \geq \Delta_n} e^{ik\theta} C(k) + \sum_{|k| < \Delta_n} e^{ik\theta} C(k) \left[1 - K\left(\frac{k}{\Delta_n}\right) \right]$$

Consider $\theta = 0$. Then, the previous expression is equal to

$$2 \sum_{k \geq \Delta_n} k^{-\beta-1} S(k) + 2 \sum_{k < \Delta_n} k^{-\beta-1} S(k) \cdot \frac{k^{\lambda+1}}{\Delta_n^{\lambda+1}}$$

Using the fact that for $p > -1$,

$$\int_{\alpha}^x t^p S(t) dt \sim (p+1)^{-1} x^{p+1} S(x), \text{ as } x \rightarrow \infty$$

and for $p < -1$,

$$\int_x^{\infty} t^p S(t) dt \sim |p+1|^{-1} x^{p+1} S(x), \text{ as } x \rightarrow \infty$$

we obtain that for $0 < \beta < \min\{1, \lambda + 1\}$:

$$(2\pi)[f(0) - \mathbb{E}\hat{f}_n(0)] \sim \frac{\Delta_n^{-\beta} S(\Delta_n)}{\beta} + \frac{\Delta_n^{-\beta} S(\Delta_n)}{\lambda + 1 - \beta} = \mathcal{O}(\Delta_n^{-\beta} \cdot S(\Delta_n))$$

Also, for $\beta > 1$, the same expression can be evaluated to be of the same order

$$\mathcal{O}(\Delta_n^{-\beta} \cdot S(\Delta_n)).$$

Compare this to the rate of Proposition IV.17 for $0 < \beta \neq 1$.

We shift our interest now to the variance. At first, using T and Δ in place of T_n and Δ_n respectively, we have

$$\begin{aligned} \hat{f}_n(0) - \mathbb{E}\hat{f}_n(0) &= \frac{1}{2\pi} \cdot \sum_{t=0}^T \sum_{s=0}^T K\left(\frac{t-s}{\Delta}\right) \frac{X(t) \cdot X(s) - C(t-s)}{|T-(t-s)|} \\ &= \frac{1}{2\pi} \cdot \sum_{v=0}^T \sum_{h=-\Delta}^{\Delta} K\left(\frac{h}{\Delta}\right) \frac{X(v+h) \cdot X(v) - C(h)}{|T-h|} \cdot \mathbb{1}(0 \leq h+v \leq T), \end{aligned}$$

by the change of variables $h = t - s, v = s$. Thus,

$$\begin{aligned} (2\pi)^2 \mathbb{E} \left| \hat{f}_n(0) - \mathbb{E}\hat{f}_n(0) \right|^2 &= \sum_{v=0}^T \sum_{w=0}^T \sum_{|h| \leq \Delta} \sum_{|\tilde{h}| \leq \Delta} K\left(\frac{h}{\Delta}\right) K\left(\frac{\tilde{h}}{\Delta}\right) \\ &\quad \frac{\mathbb{E} \left\{ [X(v+h) \cdot X(v) - C(h)] [X(w+\tilde{h}) \cdot X(w) - C(\tilde{h})] \right\}}{|T-h| |T-\tilde{h}|} \\ &\quad \cdot \mathbb{1}(0 \leq h+v \leq T) \cdot \mathbb{1}(0 \leq \tilde{h}+w \leq T) \end{aligned}$$

After using Isserlis' lemma for the expectation in the middle we finally obtain that

$$(2\pi)^2 \mathbb{E} \left| \hat{f}_n(0) - \mathbb{E}\hat{f}_n(0) \right|^2 = \frac{1}{T^2} \cdot \sum_{v=0}^T \sum_{w=0}^T \sum_{|h| \leq \Delta} \sum_{|\tilde{h}| \leq \Delta} K\left(\frac{h}{\Delta}\right) K\left(\frac{\tilde{h}}{\Delta}\right)$$

$$\frac{C(v-w+h-\tilde{h})C(v-w)+C(v-w+h)C(w-v+\tilde{h})}{|T-h||T-\tilde{h}|} \cdot \mathbb{1}(0 \leq h+v \leq T) \cdot \mathbb{1}(0 \leq \tilde{h}+w \leq T).$$

We have already shown that both of these terms are absolutely of the order $\mathcal{O}\left(\frac{\Delta}{T}\right)$. Hence, showing asymptotic equivalence of just one of these integrals with a term of order Δ/T is enough to show that the variance as a whole is of the same order. We focus on the first summand and have with the change of variables $x = v - w, y = w, z = h - \tilde{h}, u = \tilde{h}$, that

$$\begin{aligned} & \sum_{v=0}^T \sum_{w=0}^T \sum_{|h| \leq \Delta} \sum_{|\tilde{h}| \leq \Delta} K\left(\frac{h}{\Delta}\right) K\left(\frac{\tilde{h}}{\Delta}\right) \frac{C(v-w+h-\tilde{h})C(v-w)}{|T-h||T-\tilde{h}|} \\ & \quad \cdot \mathbb{1}(0 \leq h+v \leq T) \cdot \mathbb{1}(0 \leq \tilde{h}+w \leq T) \\ &= \sum_{x=-T}^T \sum_{z=-2\Delta}^{2\Delta} C(x+z)C(z) \cdot \sum_{u=-\Delta \vee (-\Delta-z)}^{\Delta \wedge (\Delta-z)} K\left(\frac{z+u}{\Delta}\right) K\left(\frac{u}{\Delta}\right) \\ & \quad \cdot \frac{|T-x|}{|T-(z+u)||T-u|} \sum_{y=0 \vee (-x)}^{T \wedge (T-x)} \mathbb{1}(x+y+z+u \in [0, T]) \mathbb{1}(u+y \in [0, T]). \end{aligned}$$

Observing that

$$\begin{aligned} \frac{1}{T} \sum_{y=0 \vee (-x)}^{T \wedge (T-x)} \mathbb{1}(x+y+z+u \in [0, T]) \mathbb{1}(u+y \in [0, T]) &\leq 1 - \frac{|x|}{T} \\ \frac{1}{2\Delta} \sum_{u=-\Delta \vee (-\Delta-z)}^{\Delta \wedge (\Delta-z)} K\left(\frac{z+u}{\Delta}\right) K\left(\frac{u}{\Delta}\right) &\leq 1 - \frac{|z|}{2\Delta} \\ \frac{|T-x|}{|T-(z+u)||T-u|} &\leq \frac{1}{|T|} \end{aligned}$$

and the fact that $-T \leq x \leq T$ and $-2\Delta \leq z \leq 2\Delta$, we see that the aforementioned

terms are bounded by 1. Also, by Assumption C', we have that

$$\sum_{x=-T}^T \sum_{z=-2\Delta}^{2\Delta} C(x+z)C(z)dzdx < \infty.$$

Using the Dominated Convergence Theorem, we obtain that the quadruple summation, divided by $2\Delta/T$ converges to a constant. Thus, it is asymptotically equivalent to Δ/T as desired.

This, for $\lambda + 1 > \beta > 0$, leads to the consistency rate

$$\mathcal{O} \left(\sqrt{\frac{\Delta_n}{|T_n|}} + \Delta_n^{-\beta} \cdot S(\Delta_n) \right).$$

Considering $0 < \beta \neq 1$, we see that the optimal consistency rate in this case essentially matches the one in Theorem IV.17.

Observe that the regular variation only played a role in establishing asymptotic equivalence of the bias vanish rate. Indeed, for the rate of the variance, we only needed the integrability of the Covariance operator and the regular variation was not used. Also, recall that here the spectral density $f \in \mathcal{P}_D(\beta + \epsilon, L)$. So the rate we should be comparing to is

$$|T_n|^{-\frac{\beta+\epsilon}{2(\beta+\epsilon)+1}}.$$

4.8.6.2 Examples on Assumptions V and V'

We present some examples of non-trivial processes that satisfy the Assumptions V and V', so as to demonstrate that the assumptions are not vacuous. We first consider an example that satisfies Assumption V.

Example IV.39. Consider the process

$$X(t) = Z(t)^2 - 1,$$

where $Z := \{Z(t), t \in \mathbb{R}^d\}$ is a zero-mean, real-valued stationary Gaussian process with standard normal marginals. Denote the stationary covariance of Z by $C_Z(\cdot)$ which we assume to satisfy $\int_{u \in \mathbb{R}^d} \sup_{\lambda \in B(0, 2\delta)} |C_Z(\lambda + u)| du < \infty$, for some small enough $\delta > 0$. This condition is quite mild and can be satisfied by covariances that are integrable and sufficiently smooth in the tail. We verify that Assumption V holds for X . We start by considering $\tilde{X}(t) = Z(t)^2$. It follows that

$$\begin{aligned}
\text{(a)} \quad & \mathbb{E}[\tilde{X}(t)] = 1 \\
\text{(b)} \quad & \mathbb{E}[\tilde{X}(t_1)\tilde{X}(t_2)] = 1 + 2C_Z(t_1 - t_2)^2 \\
\text{(c)} \quad & \mathbb{E}[\tilde{X}(t_1)\tilde{X}(t_2)\tilde{X}(t_3)] = 15a_2^2a_3^2 + 3a_2^2b_3^2 + 3a_2^2c_3^2 + 3b_2^2a_3^2 + 3b_2^2b_3^2 + b_2^2c_3^2 + 6a_3b_3a_2b_2, \\
\text{(d)} \quad & \mathbb{E}[\tilde{X}(t_1)\tilde{X}(t_2)\tilde{X}(t_3)\tilde{X}(t_4)] = 105a_2^2a_3^2a_4^2 + 15a_2^2a_3^2b_4^2 + 15a_2^2a_3^2c_4^2 + 15a_2^2a_3^2d_4^2 \\
& + 15a_2^2b_3^2a_4^2 + 9a_2^2b_3^2b_4^2 + 3a_2^2b_3^2c_4^2 + 3a_2^2b_3^2d_4^2 \\
& + 15a_2^2c_3^2a_4^2 + 3a_2^2c_3^2b_4^2 + 9a_2^2c_3^2c_4^2 + 3a_2^2c_3^2d_4^2 \\
& + 30a_2^2a_3b_3a_4b_4 + 30a_2^2a_3c_3a_4c_4 + 6a_2^2b_3c_3b_4c_4 \\
& + 15b_2^2a_3^2a_4^2 + 9b_2^2a_3^2b_4^2 + 3b_2^2a_3^2c_4^2 + 3b_2^2a_3^2d_4^2 \\
& + 9b_2^2b_3^2a_4^2 + 15b_2^2b_3^2b_4^2 + 3b_2^2b_3^2c_4^2 + 3b_2^2b_3^2d_4^2 \\
& + 3b_2^2c_3^2a_4^2 + 3b_2^2c_3^2b_4^2 + 3b_2^2c_3^2c_4^2 + b_2^2c_3^2d_4^2 \\
& + 36b_2^2a_3b_3a_4b_4 + 12b_2^2a_3c_3a_4c_4 + 12b_2^2b_3c_3b_4c_4 \\
& + 60a_2b_2a_3^2a_4b_4 + 36a_2b_2a_4b_4b_3^2 + 12a_2b_2c_3^2a_4b_4 \\
& + 60a_2b_2a_3b_3a_4^2 + 36a_2b_2a_3b_3b_4^2 + 12a_2b_2a_3b_3c_4^2 \\
& + 12a_2b_2a_3b_3d_4^2 + 24a_2b_2a_3b_3d_4^2 + 24a_2b_2a_3c_3b_4c_4 + 24a_2b_2b_3c_3a_4c_4,
\end{aligned}$$

where

$$a_2 = C_Z(t_1 - t_2)$$

$$a_3 = C_Z(t_1 - t_3)$$

$$\begin{aligned}
b_2 &= \sqrt{1 - C_Z(t_1 - t_2)^2} \\
b_3 &= \frac{C_Z(t_2 - t_3) - C_Z(t_1 - t_2) \cdot C_Z(t_1 - t_3)}{\sqrt{1 - C_Z(t_1 - t_2)^2}} \\
c_3 &= \sqrt{1 - C_Z(t_1 - t_3)^2 - \frac{[C_Z(t_2 - t_3) - C_Z(t_1 - t_2) \cdot C_Z(t_1 - t_3)]^2}{1 - C_Z(t_1 - t_2)^2}} \\
a_4 &= C_Z(t_1 - t_4) \\
b_4 &= \frac{C_Z(t_2 - t_4) - C_Z(t_1 - t_2)C_Z(t_1 - t_4)}{\sqrt{1 - C_Z(t_1 - t_2)^2}} \\
c_4 &= \frac{C_Z(t_3 - t_4) - C_Z(t_1 - t_3)C_Z(t_1 - t_4) - b_3b_4}{c_3} \\
d_4 &= \sqrt{1 - a_4^2 - b_4^2 - c_4^2}.
\end{aligned}$$

After centering the process as $X(t) = \tilde{X}(t) - 1$ and simplifying the aforementioned expressions, we end up with the following moments:

- (a) $\mathbb{E}[X(t)] = 0$
- (b) $\mathbb{E}[X(t_1)X(t_2)] = 2C_Z(t_1 - t_2)^2$
- (c) $\mathbb{E}[X(t_1)X(t_2)X(t_3)] = 8C_Z(t_1 - t_2)C_Z(t_1 - t_3)C_Z(t_2 - t_3),$
- (d) $\mathbb{E}[X(t_1)X(t_2)X(t_3)X(t_4)] = 4C_Z(t_1 - t_2)^2C_Z(t_3 - t_4)^2$
 $+ 4C_Z(t_1 - t_3)^2C_Z(t_2 - t_4)^2 + 4C_Z(t_1 - t_4)^2C_Z(t_2 - t_3)^2$
 $+ 16C_Z(t_1 - t_3)C_Z(t_1 - t_4)C_Z(t_2 - t_3)C_Z(t_2 - t_4)$
 $+ 16C_Z(t_1 - t_2)C_Z(t_1 - t_4)C_Z(t_2 - t_3)C_Z(t_3 - t_4)$
 $+ 16C_Z(t_1 - t_2)C_Z(t_1 - t_3)C_Z(t_2 - t_4)C_Z(t_3 - t_4)$.

Using (d) above, we obtain that

$$\mathbb{E}|X(t)|^4 = 60C_Z(0)^4 = 60 < \infty,$$

showing that (a) of Condition V holds.

By the definition of the cumulants in Definition VI.8 we obtain that

$$\begin{aligned} \text{cum}(X(t_1), X(t_2), X(t_3), X(t_4)) &= 16C_Z(t_1 - t_3)C_Z(t_1 - t_4)C_Z(t_2 - t_3)C_Z(t_2 - t_4) \\ &\quad + 16C_Z(t_1 - t_2)C_Z(t_1 - t_4)C_Z(t_2 - t_3)C_Z(t_3 - t_4) \\ &\quad + 16C_Z(t_1 - t_2)C_Z(t_1 - t_3)C_Z(t_2 - t_4)C_Z(t_3 - t_4). \end{aligned}$$

Then, Assumption V(b) is also satisfied, since

$$\begin{aligned} &\sup_{w \in \mathbb{R}^d} \int_{u \in \mathbb{R}^d} \int_{v \in \mathbb{R}^d} \sup_{\lambda_1, \lambda_2, \lambda_3 \in B(0, \delta)} |\text{cum}(X(\lambda_1 + u), X(\lambda_2 + v), X(\lambda_3 + w), X(0))| dv du \\ &\leq 16 \sup_{w \in \mathbb{R}^d} \int_{u \in \mathbb{R}^d} \int_{v \in \mathbb{R}^d} \sup_{\lambda_1, \lambda_2, \lambda_3 \in B(0, \delta)} \\ &\quad \left\{ |C_Z(\lambda_1 - \lambda_3 + u - w)C_Z(\lambda_1 + u)C_Z(\lambda_2 - \lambda_3 + v - w)C_Z(\lambda_2 + v)| \right. \\ &\quad + |C_Z(\lambda_1 - \lambda_2 + u - v)C_Z(\lambda_1 + u)C_Z(\lambda_2 - \lambda_3 + v - w)C_Z(\lambda_3 + w)| \\ &\quad \left. + |C_Z(\lambda_1 - \lambda_2 + u - v)C_Z(\lambda_1 - \lambda_3 + u - w)C_Z(\lambda_2 + v)C_Z(\lambda_3 + w)| \right\} dv du \\ &\leq 16|C_Z(0)|^2 \int_{u \in \mathbb{R}^d} \int_{v \in \mathbb{R}^d} \sup_{\lambda_1, \lambda_2 \in B(0, \delta)} \left\{ |C_Z(\lambda_1 + u)C_Z(\lambda_2 + v)| \right. \\ &\quad + |C_Z(\lambda_1 - \lambda_2 + u - v)C_Z(\lambda_1 + u)| \\ &\quad \left. + |C_Z(\lambda_1 - \lambda_2 + u - v)C_Z(\lambda_2 + v)| \right\} dv du \\ &\leq 48|C_Z(0)|^2 \left(\int_{u \in \mathbb{R}^d} \sup_{\lambda \in B(0, 2\delta)} |C_Z(\lambda + u)| du \right)^2 < \infty. \end{aligned}$$

□

Next, we present an example inspired by the linear processes in Proposition 4.1 of *Panaretos and Tavakoli (2013)*. Assume that \mathbb{H} is a separable (typically infinite-dimensional) Hilbert space. Let ϵ_t , $t \in \mathbb{Z}$ be iid random elements of \mathbb{H} such that $\mathbb{E}\|\epsilon_0\|^4 < \infty$ and consider a sequence of bounded linear operators $A_s : \mathbb{H} \rightarrow \mathbb{H}$, $s \in \mathbb{Z}$. Define

$$X(t) = \sum_{s \in \mathbb{Z}} A_s \epsilon_{t-s}, \quad t \in \mathbb{Z}. \quad (4.78)$$

In the following lemma, we show that the real process $X(t)$ is well defined under a mild square-summability condition on the operator norms of the coefficients. To this end, let $\mathcal{L}^2(\mathbb{H})$ denote the Hilbert space of \mathbb{H} -valued random elements equipped with the inner product $\langle A, B \rangle_{\mathcal{L}^2} = \mathbb{E}\langle A, B \rangle$ for all \mathbb{H} -valued random elements A and B such that $\mathbb{E}[\|A\|^2 + \|B\|^2] < \infty$. The resulting norm in $\mathcal{L}^2(\mathbb{H})$ will be denoted by $\|\cdot\|_{\mathcal{L}^2(\mathbb{H})}$.

Lemma IV.40. *Assume that the operator norms of $\{A_s, s \in \mathbb{Z}\}$ are square summable, namely that*

$$\sum_{s \in \mathbb{Z}} \|A_s\|_{\text{op}}^2 < \infty. \quad (4.79)$$

Then, the series in (4.78) converges in $\|\cdot\|_{\mathcal{L}^2(\mathbb{H})}$ and the process $\{X(t), t \in \mathbb{Z}\}$ defined.

Proof. Let $\Sigma_\epsilon = \mathbb{E}[\epsilon_0 \otimes \epsilon_0]$ be the covariance operator of every $\epsilon_t, t \in \mathbb{Z}$. We start by defining

$$X^{(N)}(t) = \sum_{|s| \leq N} A_s \epsilon_{t-s} \quad \text{and} \quad X^{-(N)}(t) = \sum_{|s| > N} A_s \epsilon_{t-s}. \quad (4.80)$$

We have that

$$\begin{aligned} \mathbb{E}\langle X^{(N)}(t), X^{(N)}(t) \rangle &= \sum_{|s_1| \leq N} \sum_{|s_2| \leq N} \mathbb{E}\langle A_{s_1} \epsilon_{t-s_1}, A_{s_2} \epsilon_{t-s_2} \rangle \\ &= \sum_{|s| \leq N} \mathbb{E}\langle A_s \epsilon_{t-s}, A_s \epsilon_{t-s} \rangle = \sum_{|s| \leq N} \mathbb{E}\langle \epsilon_{t-s}, A_s^* A_s \epsilon_{t-s} \rangle, \end{aligned}$$

where the second equality follows by the independence of the ϵ 's. Let now $\{e_j\}$ be a CONS of \mathbb{H} that diagonalizes Σ_ϵ . Then, we can express the ϵ 's as

$$\epsilon_{t-s} = \sum_{j=1}^{\infty} Z_{t-s,j} e_j,$$

where $Z_{s,j} := \langle \epsilon_s, e_j \rangle$, are independent in s because the ϵ_s 's are iid. Also, because of the choice of $\{e_j\}$ as the eigenvectors of the covariance operator Σ_ϵ , we have that for

each fixed s , the $Z_{s,j}$'s are uncorrelated in j :

$$\mathbb{E} [Z_{s,i} \overline{Z_{s,j}}] = \lambda_i \cdot \delta_{i-j}.$$

Using those, we obtain

$$\begin{aligned} \mathbb{E} \langle X^{(N)}(t), X^{(N)}(t) \rangle &= \sum_{|s| \leq N} \mathbb{E} \left\langle \sum_k e_k Z_{t-s,k}, A_s^* A_s \sum_\ell e_\ell Z_{t-s,\ell} \right\rangle \\ &= \sum_{|s| \leq N} \sum_k \sum_\ell \mathbb{E} [Z_{t-s,k} \overline{Z_{t-s,\ell}}] \cdot \langle e_k, A_s^* A_s e_\ell \rangle \\ &= \sum_{|s| \leq N} \sum_k \lambda_k \cdot \langle e_k, A_s^* A_s e_k \rangle \leq \sum_k \lambda_k \sum_{|s| \leq N} \langle e_k, A_s^* A_s e_k \rangle \\ &\leq \text{tr}(\Sigma_\epsilon) \cdot \sum_{|s| \leq N} \|A_s^* A_s\|_{\text{op}} \leq \text{tr}(\Sigma_\epsilon) \cdot \sum_{|s| \leq N} \|A_s\|_{\text{op}}^2 < \infty. \end{aligned} \tag{4.81}$$

With a similar argument to (4.81), one has that for $M < N$

$$\mathbb{E} \|X^{(N)}(t) - X^{(M)}(t)\|^2 \leq \text{tr}(\Sigma_\epsilon) \sum_{M < |s| \leq N} \|A_s\|_{\text{op}}^2 \rightarrow 0,$$

as $N, M \rightarrow \infty$. This shows that the sequence $\{X^{(N)}(t)\}_{N \in \mathbb{N}}$ is a Cauchy sequence in the Hilbert space $(\mathcal{L}^2(\mathbb{H}), \langle \cdot, \cdot \rangle_{\mathcal{L}^2})$, where $\langle A, B \rangle_{\mathcal{L}^2} = \mathbb{E} \langle A, B \rangle$ for A, B random elements of \mathbb{H} . Thus, the limit of this sequence exists and

$$X^{(N)}(t) \rightarrow X(t) \in \mathcal{L}^2(\mathbb{H}),$$

which completes the proof. □

Proposition IV.41. *Let $X(t)$ defined as in (4.78). Assume that $\{A_s, s \in \mathbb{Z}\}$ are Hilbert-Schmidt operators with $\sum_{s \in \mathbb{Z}} \|A_s\|_{\text{HS}} < \infty$. Moreover, letting $Z_{s,j} = \langle \epsilon_s, e_j \rangle$,*

where $\{e_j\}$ is a CONS diagonalizing $\Sigma_\epsilon := \mathbb{E}[\epsilon_0 \otimes \epsilon_0]$, assume that

$$\sum_{\ell_1, \ell_2, \ell_3, \ell_4} \text{cum}(Z_{0, \ell_1}, Z_{0, \ell_2}, Z_{0, \ell_3}, Z_{0, \ell_4})^2 \leq B < \infty.$$

Then, the process $\{X(t), t \in \mathbb{Z}\}$ satisfies Assumption V' .

Proof. Recall that part (a) of Assumption V' entails the finite fourth moment of $\|X(t)\|$.

Let $X^{(N)}(t)$ and $X^{-(N)}(t)$ be as defined in (4.80). Then, for every $k \in \mathbb{N}$ such that $\mathbb{E}\|\epsilon_t\|^k < \infty$, we have that

$$\begin{aligned} \mathbb{E}\|X^{-(N)}(t)\|^k &\leq \sum_{|s_1|, \dots, |s_k| > N} \|A_{s_1}\|_{\text{op}} \dots \|A_{s_k}\|_{\text{op}} \mathbb{E}(\|\epsilon_{t-s_1}\| \dots \|\epsilon_{t-s_k}\|) \\ &\leq \sum_{|s_1|, \dots, |s_k| > N} \|A_{s_1}\|_{\text{op}} \dots \|A_{s_k}\|_{\text{op}} \mathbb{E}(\|\epsilon_{t-s_1}\|^k)^{1/k} \dots \mathbb{E}(\|\epsilon_{t-s_k}\|^k)^{1/k} \\ &= \mathbb{E}\|\epsilon_0\|^k \cdot \sum_{|s_1|, \dots, |s_k| > N} \|A_{s_1}\|_{\text{op}} \dots \|A_{s_k}\|_{\text{op}} \\ &= \mathbb{E}\|\epsilon_0\|^k \cdot \left(\sum_{|s| > N} \|A_s\|_{\text{op}} \right)^k \rightarrow 0, \text{ as } N \rightarrow \infty, \end{aligned} \tag{4.82}$$

where the inequality in the second line follows from the generalized Hölder inequality (cf. Theorem 11 of *Hardy et al.*, 1952) and we used that $\|A_s\|_{\text{op}} \leq \|A_s\|_{\text{HS}}$. Hence, we have $\mathcal{L}^k(\mathbb{H})$ -convergence of $X_t^{(N)}$ to X_t , in the sense that

$$\lim_{N \rightarrow \infty} \left(\mathbb{E}\|X(t) - X^{(N)}(t)\|_{\mathbb{H}}^k \right)^{1/k} = 0.$$

These previous calculations also show directly that $\mathbb{E}\|X_t\|^k < \infty$. Specifically, for $k = 4$, part (a) is proved.

Now, for part (b), as in the proof of Lemma IV.40, letting $\{e_j\}$ be a CONS diag-

onalizing $\Sigma_\epsilon = \mathbb{E}[\epsilon_0 \otimes \epsilon_0]$, we write

$$A_s = \sum_{i,j} a_{ij}(s) e_i \otimes e_j \quad \text{and} \quad \epsilon_{t-s} = \sum_k Z_{t-s,k} e_k,$$

with $Z_{s,k} := \langle \epsilon_s, e_k \rangle$. Note that $\{e_i \otimes e_j\}$ is a CONS in the Hilbert space \mathbb{X} of Hilbert-Schmidt operators on \mathbb{H} equipped with $\langle \cdot, \cdot \rangle_{\text{HS}}$ and the above expression for A_s converges in $\| \cdot \|_{\text{HS}}$. Let also

$$\begin{aligned} A_{s,i} &:= \sum_j a_{ij}(s) e_i \otimes e_j \\ X_i(t) &:= \langle X(t), e_i \rangle = \sum_{s \in \mathbb{Z}} \sum_j a_{ij}(s) Z_{t-s,j} = \sum_{s \in \mathbb{Z}} A_{s,i} \epsilon_{t-s}, \end{aligned}$$

so that $X(t) = \sum_i X_i(t) e_i$. Recall the representation in Proposition VI.10 (see also (4.20)). For notational simplicity suppose that the process $X(t)$ is real relative to the CONS $\{e_i\}$, i.e., all the $X_i(t)$'s are real random variables.

We start by exploiting the multilinearity of the cumulants and the fact that ϵ_t 's are iid. We have by Proposition VI.10 that $\text{cum}(X(u), X(v), X(w), X(0))$ equals:

$$\begin{aligned} & \left| \sum_i \sum_j \text{cum}(X_i(u), X_j(v), X_i(w), X_j(0)) \right| \\ &= \left| \sum_i \sum_j \text{cum} \left(\sum_{s_1 \in \mathbb{Z}} A_{s_1,i} \epsilon_{u-s_1}, \sum_{s_2 \in \mathbb{Z}} A_{s_2,j} \epsilon_{v-s_2}, \sum_{s_3 \in \mathbb{Z}} A_{s_3,i} \epsilon_{w-s_3}, \sum_{s_4 \in \mathbb{Z}} A_{s_4,j} \epsilon_{-s_4} \right) \right| \\ &= \left| \sum_i \sum_j \text{cum} \left(\sum_{s_1 \in \mathbb{Z}} A_{u-s_1,i} \epsilon_{s_1}, \sum_{s_2 \in \mathbb{Z}} A_{v-s_2,j} \epsilon_{s_2}, \sum_{s_3 \in \mathbb{Z}} A_{w-s_3,i} \epsilon_{s_3}, \sum_{s_4 \in \mathbb{Z}} A_{-s_4,j} \epsilon_{s_4} \right) \right| \\ &= \left| \sum_i \sum_j \sum_{s_1 \in \mathbb{Z}} \sum_{s_2 \in \mathbb{Z}} \sum_{s_3 \in \mathbb{Z}} \sum_{s_4 \in \mathbb{Z}} \text{cum}(A_{u-s_1,i} \epsilon_{s_1}, A_{v-s_2,j} \epsilon_{s_2}, A_{w-s_3,i} \epsilon_{s_3}, A_{-s_4,j} \epsilon_{s_4}) \right| \quad (4.83) \end{aligned}$$

$$= \left| \sum_i \sum_j \sum_{s \in \mathbb{Z}} \text{cum}(A_{u-s,i} \epsilon_s, A_{v-s,j} \epsilon_s, A_{w-s,i} \epsilon_s, A_{-s,j} \epsilon_s) \right|, \quad (4.84)$$

where (4.84) follows from the fact that ϵ_t 's are iid and (4.83) will be justified in the

end of this proof.

Continuing, (4.84) is equal to

$$\left| \sum_i \sum_j \sum_{s \in \mathbb{Z}} \sum_{\ell_1, \ell_2, \ell_3, \ell_4} a_{i\ell_1}(u-s) a_{j\ell_2}(v-s) a_{i\ell_3}(w-s) a_{j\ell_4}(-s) \text{cum}(Z_{s,\ell_1}, Z_{s,\ell_2}, Z_{s,\ell_3}, Z_{s,\ell_4}) \right| \quad (4.85)$$

or, equivalently

$$\left| \sum_i \sum_j \sum_{s \in \mathbb{Z}} \sum_{\ell_1, \ell_2, \ell_3, \ell_4} a_{i\ell_1}(u-s) a_{j\ell_2}(v-s) a_{i\ell_3}(w-s) a_{j\ell_4}(-s) \text{cum}(Z_{0,\ell_1}, Z_{0,\ell_2}, Z_{0,\ell_3}, Z_{0,\ell_4}) \right|. \quad (4.86)$$

Changing the order of summation and applying the Cauchy-Schwarz inequality over $\sum_{\ell_1, \dots, \ell_4}$, we have that (4.86) is bounded above by

$$\begin{aligned} & \left| \sum_s \sqrt{\sum_{\ell_1, \dots, \ell_4} \text{cum}(Z_{0,\ell_1}, Z_{0,\ell_2}, Z_{0,\ell_3}, Z_{0,\ell_4})^2} \right. \\ & \quad \left. \cdot \sqrt{\sum_{\ell_1, \dots, \ell_4} \left(\sum_{i,j} \left[a_{i\ell_1}(u-s) a_{j\ell_2}(v-s) a_{i\ell_3}(w-s) a_{j\ell_4}(-s) \right] \right)^2} \right| \\ & \leq B \sum_{s \in \mathbb{Z}} \left[\sum_{\ell_1, \dots, \ell_4} \left(\sum_i a_{i\ell_1}(u-s)^2 \right) \left(\sum_i a_{i\ell_3}(w-s)^2 \right) \right. \\ & \quad \left. \cdot \left(\sum_j a_{j\ell_2}(v-s)^2 \right) \left(\sum_j a_{j\ell_4}(-s)^2 \right) \right]^{1/2}, \\ & = B \sum_{s \in \mathbb{Z}} \|A_{u-s}\|_{\text{HS}} \|A_{v-s}\|_{\text{HS}} \|A_{w-s}\|_{\text{HS}} \|A_{-s}\|_{\text{HS}}, \end{aligned}$$

where the above inequality follows by applying the Cauchy-Schwarz inequality twice – once over \sum_i and once over \sum_j . The last relation follows from the fact that $\|A_t\|_{\text{HS}}^2 = \sum_{\ell,i} a_{i\ell}(t)^2$.

Thus, we finally obtain:

$$\begin{aligned}
& \sup_{w \in \mathbb{Z}} \sum_{u \in \mathbb{Z}} \sum_{v \in \mathbb{Z}} \left| \sum_i \sum_j \text{cum} (X_i(u), X_j(v), X_i(w), X_j(0)) \right| \\
& \leq \sup_{w \in \mathbb{Z}} \sum_{u \in \mathbb{Z}} \sum_{v \in \mathbb{Z}} B \sum_{s \in \mathbb{Z}} \|A_{u-s}\|_{\text{HS}} \|A_{v-s}\|_{\text{HS}} \|A_{w-s}\|_{\text{HS}} \|A_{-s}\|_{\text{HS}} \\
& \leq B \sup_{w \in \mathbb{Z}} \|A_w\|_{\text{HS}} \cdot \left(\sum_{s \in \mathbb{Z}} \|A_{u-s}\|_{\text{HS}} \right)^3 < \infty.
\end{aligned}$$

Now, it only remains to justify the equality (4.83). We will use $X^{(N)}(t)$ and $X^{-(N)}(t)$ again. The calculations in (4.82) imply again by the generalized Hölder inequality and the Dominated Convergence Theorem that

$$\mathbb{E} \left[\lim_{N \rightarrow \infty} X_i^{(N)}(u) X_j^{(N)}(v) X_i^{(N)}(w) X_j^{(N)}(0) \right] = \lim_{N \rightarrow \infty} \mathbb{E} \left[X_i^{(N)}(u) X_j^{(N)}(v) X_i^{(N)}(w) X_j^{(N)}(0) \right].$$

To this end, we introduce some notation. For each pair $m = (m_1, m_2) \in \{(u, i), (v, j), (w, i), (0, j)\}$, we write X_m^N for $X_{m_2}^N(m_1)$. For example, for $m = (u, i)$ we have that $X_m^{(N)} = X_i^{(N)}(u)$. Thus, using the definition of cumulants, we obtain

$$\begin{aligned}
& \text{cum} (X_i(u), X_j(v), X_i(w), X_j(0)) \\
& = \sum_{\nu=(\nu_1, \dots, \nu_q)} (-1)^{q-1} (q-1)! \prod_{l=1}^q \mathbb{E} \left[\prod_{m \in \nu_l} \lim_{N \rightarrow \infty} X_m^{(N)} \right] \\
& = \sum_{\nu=(\nu_1, \dots, \nu_q)} (-1)^{q-1} (q-1)! \prod_{l=1}^q \mathbb{E} \left[\lim_{N \rightarrow \infty} \prod_{m \in \nu_l} X_m^{(N)} \right] \\
& = \lim_{N \rightarrow \infty} \sum_{\nu=(\nu_1, \dots, \nu_q)} (-1)^{q-1} (q-1)! \prod_{l=1}^q \mathbb{E} \left[\prod_{m \in \nu_l} X_m^{(N)} \right] \\
& = \lim_{N \rightarrow \infty} \text{cum} \left(\sum_{|s_1| \leq |N|} A_{s_1, i} \cdot \epsilon_{u-s_1}, \sum_{|s_2| \leq |N|} A_{s_2, j} \cdot \epsilon_{v-s_2}, \right. \\
& \qquad \qquad \qquad \left. \sum_{|s_3| \leq |N|} A_{s_3, i} \cdot \epsilon_{w-s_3}, \sum_{|s_4| \leq |N|} A_{s_4, i} \cdot \epsilon_{-s_4} \right)
\end{aligned}$$

$$\begin{aligned}
&= \lim_{N \rightarrow \infty} \sum_{|s_1|, |s_2|, |s_3|, |s_4| \leq |N|} \text{cum}(A_{s_1, i} \cdot \epsilon_{u-s_1}, A_{s_2, j} \cdot \epsilon_{v-s_2}, A_{s_3, i} \cdot \epsilon_{w-s_3}, A_{s_4, i} \cdot \epsilon_{-s_4}) \\
&= \sum_{|s_1|, |s_2|, |s_3|, |s_4| \in \mathbb{Z}} \text{cum}(A_{s_1, i} \cdot \epsilon_{u-s_1}, A_{s_2, j} \cdot \epsilon_{v-s_2}, A_{s_3, i} \cdot \epsilon_{w-s_3}, A_{s_4, i} \cdot \epsilon_{-s_4}),
\end{aligned}$$

where the sum is over all unordered partitions of $\{(u, i), (v, j), (w, i), (0, j)\}$. The proof is complete. □

CHAPTER V

Minimax Rates

In this chapter we continue examining the setting of gridded data, where the grid size either stays fixed or shrinks to zero as the sample size increases (cf Section 4.5). Moreover, we focus our interest on $\mathcal{P}_D(\beta, L)$, a class of covariance functions that are dominated by a power law as in (4.32). For this class of covariance functions, we were able to obtain explicit rates (upper bounds) on the consistency of our proposed lag-window estimator. By comparing with these carefully computed rates, we show that our spectral density estimator is minimax rate optimal for spectral densities in $\mathcal{P}_D(\beta, L)$. These minimax rate results, to the best of our knowledge, are the first to be established for the pointwise inference of the spectral density of functional time series or function-valued, continuous-time processes observed at discrete time points. The ideas behind the proofs of these results are heavily influenced by *Samarov* (1977).

5.1 Minimax rates

The minimax rates for the spectral density estimation problem have received some attention. A few examples of such studies for times series include *Samarov* (1977), *Bentkus* (1985), and *Efromovich* (1998), among others. The continuous-time setting, however, appears to have been less studied (see, e.g., *Ginovyan*, 2011, and the references therein). To the best of our knowledge, results on minimax rates for the

pointwise inference of the spectral density of functional time series or function-valued, continuous-time processes observed at discrete time points have not yet been established. Also, we are not aware of such results for random fields indexed by \mathbb{Z}^d or \mathbb{R}^d , $d > 1$.

Assuming $\{X(t)\}$ is Gaussian, below we extend the work of *Samarov* (1977) by focusing on the classes $\mathcal{P}_D(\beta, L)$ and $\mathcal{P}_C(\beta, L)$ considered in Section 4.5. As in Section 4.5, we assume the data are observed on a grid. Our first result is concerned with the case $\delta_n = 1$, where, in accordance with Section 4.5.1, we consider a discrete parameter process $\{X(t), t \in \mathbb{Z}^d\}$.

Theorem V.1. *Assume that $\{X(t), t \in \mathbb{Z}^d\}$ is a stationary Gaussian process with spectral density function f . Let \mathcal{M}_n be the class of all possible estimators f_n of f based on the observations $X(t), t \in \{1, \dots, n\}^d$. Then, for any interior point $\theta_0 \in (-\pi, \pi)^d$ and $\beta, L > 0$,*

$$\liminf_{n \rightarrow \infty} \inf_{f_n \in \mathcal{M}_n} \sup_{f \in \mathcal{P}_D(\beta, L)} \mathbb{P} \left(\|f_n(\theta_0) - f(\theta_0)\|_{\text{HS}} \geq n^{-\frac{d\beta}{2\beta+d}} \right) > 0, \quad (5.1)$$

where $\mathcal{P}_D(\beta, L)$ is defined in (4.32).

Remark V.2. Note that $|\mathbb{T}_n| = n^d$ for $\mathbb{T}_n = \{1, \dots, n\}^d$. Hence, by Theorem IV.17, the estimator $\hat{f}_n(\theta_0)$ achieves the minimax rate $|\mathbb{T}_n|^{-\beta/(2\beta+d)} = n^{-(d\beta)/(2\beta+d)}$ uniformly over the class $\mathcal{P}_D(\beta, L)$. Thus, in the setting of processes indexed by \mathbb{Z}^d , our estimators are rate-optimal in a uniform sense, for the power-law class and for all dimensions $d \geq 1$.

Proof of Theorem V.1 (Outline). The detailed proof of Theorem V.1 is given in Section 5.2. We describe the key elements of the proof here. First, for any member e_i of the real CONS, consider the (scalar) real-valued process $X_{e_i}(t) := \langle X(t), e_i \rangle_{\mathbb{H}}$ and let $C_{e_i}(x)$ and $f_{e_i}(\theta)$ be its stationary covariance and spectral density, respective. If

$f \in \mathcal{P}_D(\beta, L)$ then

$$\int_{\mathbb{R}^d} (1 + \|x\|_2^\beta) |C_{e_i}(x)| dx \leq L \quad \text{and} \quad |\hat{f}_{e_i}(\theta_0) - f_{e_i}(\theta_0)| \leq \|\hat{f}(\theta_0) - f(\theta_0)\|_{\text{HS}}.$$

These follow from the simple fact that $|\langle \mathcal{A}\phi, \phi \rangle_{\mathbb{H}}| \leq \|\mathcal{A}\|_{\text{op}}$ for any bounded linear operator \mathcal{A} and unitary $\phi \in \mathbb{H}$. Thus, it suffices to prove Theorem V.1 by focusing on scalar, real-valued processes. $\phi = e_i$, the real basis. The crucial step of the proof is constructing two functions $f_{0,n}, f_{1,n}$ in $\mathcal{P}_D(\beta, L)$ such that the distance between them accurately measures the complexity of the estimation problem. Let

$$f_{0,n}(\theta) = L/2 \cdot \mathbb{1}(\theta \in [-\pi, \pi]^d).$$

For $\theta = (\theta_i)_{i=1}^d \in \mathbb{R}^d$, define the function

$$g(\theta) = \epsilon \cdot \prod_{i=1}^d \varphi(\theta_i), \quad \text{where } \varphi(x) = \exp\left(-\frac{1}{1 - (x/\pi)^2}\right) \mathbb{1}(|x| < \pi), \quad x \in \mathbb{R},$$

for some $\epsilon > 0$. Note that the so-called ‘‘bump’’ function g is compactly supported and infinitely differentiable. Consider

$$g_n(\theta) = h_n^\beta g\left(\frac{\theta - \theta_0}{h_n}\right),$$

where $h_n = M \cdot n^{-d/(2\beta+d)}$ for some appropriate constant M . Now, let †

$$f_{1,n}(\theta) = f_{0,n}(\theta) + [g_n(\theta) + g_n(-\theta)].$$

Thus, the distance between $f_{0,n}(\theta)$ and $f_{1,n}(\theta)$ is $g_n(\theta) + g_n(-\theta) = \mathcal{O}(n^{-d\beta/(2\beta+d)})$. We then apply Theorem 2.5(iii) in *Tsybakov* (2008) to obtain the desired result by verifying the following:

- (1) $f_{0,n}, f_{1,n} \in \mathcal{P}_D(\beta, L)$;

(2) $f_{1n}(\theta_0) - f_{0n}(\theta_0) = c_{\theta_0} n^{-d\beta/(2\beta+d)}$ for n large enough, where $c_{\theta_0} = M^\beta(1 + \mathbb{1}(\theta_0 = 0)) > 0$;

(3) $\sup_n \text{KL}(\mathbb{P}_{1n}, \mathbb{P}_{0n}) < \infty$, where KL stands for the Kullback-Leibler divergence and $\mathbb{P}_{0,n}$ and $\mathbb{P}_{1,n}$ are probability distributions under $f_{0,n}$ and $f_{1,n}$ respectively.

The most technically challenging part of the proof is the computation of $\text{KL}(\mathbb{P}_{1n}, \mathbb{P}_{0n})$ in part (3), which is accomplished by following and extending an approach introduced in *Samarov* (1977). For details, see Sections 5.2 and 5.3. \square

The next result gives the minimax rate for the continuous-parameter Gaussian process whose spectral density belongs to $\mathcal{P}_C(\beta, L)$, defined in (4.36).

Theorem V.3. *Let $\{X(t), t \in \mathbb{R}^d\}$ be a stationary Gaussian process with spectral density f . Let \mathcal{M}_n be the class of all possible estimators f_n of f based on the observations $X(k\delta_n), k \in \{1, \dots, n\}^d$. Then, for each $\theta_0 \in \mathbb{R}^d$ and $\beta, L > 0$,*

$$\liminf_{n \rightarrow \infty} \inf_{f_n \in \mathcal{M}_n} \sup_{f \in \mathcal{P}_C(\beta, L)} \mathbb{P}(\|f_n(\theta_0) - f(\theta_0)\|_{\text{HS}} \geq (n\delta_n)^{-d\beta/(2\beta+d)}) > 0, \quad (5.2)$$

where $\mathcal{P}_C(\beta, L)$ is defined in (4.36).

The proof of Theorem V.3 is similar to that of Theorem V.1 and is presented in Section 5.3. We conclude this section with several remarks.

Remark V.4. Comparing the minimax lower bounds in (5.1) and (5.2), one can interpret $(n\delta_n)^d$ as the “effective” sample size in the case of mixed-domain asymptotics:

$$\delta_n \rightarrow 0 \quad \text{and} \quad n\delta_n \rightarrow \infty.$$

1. Recall Remark IV.20 and observe that, in the fine sampling regime, the rate of $\hat{f}_n(\theta)$ obtained in (4.37) matches the minimax lower bound in (5.2). To the best of our knowledge, this is the first result on the minimax rate for spectral density estimation in a mixed-domain setting.

2. An open problem is the construction of a narrower class \mathcal{P}_C , which reflects both the tail-decay of the auto-covariance (through β) and its smoothness (through γ) so that the upper- and lower-bounds on the rate of the estimators match in both the fine- and coarse-sampling regimes (cf. Remark IV.20).

5.2 Proof for discrete time

For any member e_i of the real CONS, consider the (scalar) real-valued process

$$X_{e_i}(t) := \langle X(t), e_i \rangle_{\mathbb{H}}$$

and let $C_{e_i}(x)$ and $f_{e_i}(\theta)$ be its stationary covariance and spectral density, respectively.

If $f \in \mathcal{P}_D(\beta, L)$ then

$$\int_{\mathbb{R}^d} (1 + \|x\|_2^\beta) |C_{e_i}(x)| dx \leq L \quad \text{and} \quad |\hat{f}_{e_i}(\theta_0) - f_{e_i}(\theta_0)| \leq \|\hat{f}(\theta_0) - f(\theta_0)\|_{\text{HS}}.$$

These follow from the simple fact that $|\langle \mathcal{A}\phi, \phi \rangle_{\mathbb{H}}| \leq \|\mathcal{A}\|_{\text{op}}$ for any bounded linear operator \mathcal{A} and unitary $\phi \in \mathbb{H}$. Thus, it suffices to prove Theorems V.1 and V.3 for scalar, real-valued processes, which we do below.

Proof of Theorem V.1. Let $\|\cdot\|$ denote the Euclidean norm in \mathbb{R}^d and C_g be the covariance that corresponds to the spectral density g . Fix an interior point $\theta_0 \in (-\pi, \pi)^d$ and let $f_{0,n}(\theta) = L/(2 \cdot (2\pi)^d) \cdot \mathbb{1}(\theta \in [-\pi, \pi]^d)$. Then,

$$C_{f_{0,n}}(k) = \int_{\theta \in [-\pi, \pi]^d} e^{-i\theta^\top k} \frac{L}{2 \cdot (2\pi)^d} d\theta = \mathbb{1}(k = 0)L/2, \quad (5.3)$$

and therefore

$$\sum_{k \in \mathbb{Z}^2} |C_{f_{0,n}}(k)|(1 + \|k\|^\beta) = C_{f_{0,n}}(0) = L/2 < L. \quad (5.4)$$

Let for $\theta = (\theta_i)_{i=1}^d \in \mathbb{R}^d$,

$$g(\theta) = \epsilon \cdot \prod_{i=1}^d \varphi(\theta_i), \quad \text{where } \varphi(x) = \exp\left(-\frac{1}{1 - (x/\pi)^2}\right) \mathbb{1}(|x| < \pi), \quad (x \in \mathbb{R}) \quad (5.5)$$

for some $\epsilon > 0$, which is to be determined. The function φ is a type of a “bump” function that belongs to $C_0^\infty(\mathbb{R})$ (the class of infinitely differentiable functions with compact support). The support of φ is the compact interval $[-\pi, \pi]$. Hence $g \in C_0^\infty(\mathbb{R}^d)$ and its support is $[-\pi, \pi]^d$. Consider the function

$$g_n(\theta) = h_n^\beta g\left(\frac{\theta - \theta_0}{h_n}\right), \quad (5.6)$$

where $0 < h_n \leq 1$ and tends to 0 at a rate to be determined later. Observe that since $\theta_0 \in (-\pi, \pi)^d$, the support of g_n is included in $\theta_0 + h_n \cdot [-\pi, \pi]^d \subset (-\pi, \pi)^d$, for all sufficiently small h_n .

Now, consider the “alternative” spectral density models:

$$f_{1,n}(\theta) = f_{0,n}(\theta) + [g_n(\theta) + g_n(-\theta)] =: \frac{L}{2} \cdot \mathbb{1}_{(-\pi, \pi)^d}(\theta) + r_n(\theta), \quad \theta \in [-\pi, \pi]^d.$$

We will choose the sequence h_n and the constant $\epsilon > 0$ such that the following three properties hold.

Properties:

- (1) $f_{0,n}, f_{1,n} \in \mathcal{P}_D(\beta, L)$, where the class $\mathcal{P}_D(\beta, L)$ is defined in (4.32).
- (2) For all n large enough, we have

$$f_{1,n}(\theta_0) - f_{0,n}(\theta_0) = h_n^\beta [g(0) + g(2\theta_0/h_n)] = g(0)(1 + \mathbb{1}(\theta_0 = 0)) \cdot h_n^\beta. \quad (5.7)$$

- (3) $\text{KL}(\mathbb{P}_{1n}, \mathbb{P}_{0n}) \leq C < \infty$, where KL stands for the Kullback-Leibler distance and \mathbb{P}_{0n} and \mathbb{P}_{1n} are probability distributions of the data $\{X(k), k \in \{1, \dots, n\}^d\}$

under $f_{0,n}$ and $f_{1,n}$ respectively.

Proof of Property (1). We have already shown that $f_{0,n} \in \mathcal{P}_D(\beta, L)$. Recalling (4.32), and in view of (5.3) and (5.4), to prove $f_{1,n} \in \mathcal{P}_D(\beta, L)$, it is enough to show that

$$\sum_{k \in \mathbb{Z}^d} |C_{g_n}(k)| (1 + \|k\|^\beta) < \frac{L}{4}, \quad (5.8)$$

where $C_{g_n}(k) = \int_{\theta \in [-\pi, \pi]^d} e^{-i\theta^\top k} g_n(\theta) d\theta$.

We have that for all $k = (k_i)_{i=1}^d \in \mathbb{Z}^d$,

$$\begin{aligned} C_{g_n}(k) &= \int_{\theta \in [-\pi, \pi]^d} e^{-i\theta^\top k} h_n^\beta g \left(\frac{\theta - \theta_0}{h_n} \right) d\theta \\ &= h_n^{\beta+d} \cdot e^{-i\theta_0^\top k} \int_{x \in [-\pi, \pi]^d} e^{-ix^\top k h_n} g(x) dx \\ &= \epsilon \cdot h_n^{\beta+d} \cdot e^{-i\theta_0^\top k} \prod_{i=1}^d \widehat{\varphi}(k_i h_n), \end{aligned} \quad (5.9)$$

where we used the change of variables $x = (\theta - \theta_0)/h_n$ and the fact that $\theta_0 + h_n \cdot [-\pi, \pi]^d \subset (-\pi, \pi)^d$, for all sufficiently small h_n . The last relation follows from (5.5), where $\widehat{\varphi}(x) := \int_{-\pi}^{\pi} e^{-ixu} \varphi(u) du$ denotes the Fourier transform of the bump function φ . Now using the fact that the derivatives of φ vanish at $\pm\pi$, i.e., $\varphi^{(\ell)}(\pm\pi) = 0$, for all $\ell = 0, 1, \dots$, integration by parts yields

$$\widehat{\varphi}(x) = \frac{1}{(-ix)^\ell} \int_{-\pi}^{\pi} e^{-ixu} \varphi^{(\ell)}(u) du, \quad \ell = 0, 1, \dots$$

Indeed, for all ℓ , the derivative $\varphi^{(\ell)}(x)$ is continuous and supported on $[-\pi, \pi]$, and thus

$$|\widehat{\varphi}(x)| \leq c_0 \wedge (|x|^{-\ell} c_\ell), \quad \text{where } c_\ell := \int_{-\pi}^{\pi} |\varphi^{(\ell)}(u)| du.$$

In view of (5.9), we have

$$|C_{g_n}(k)| \leq \epsilon \cdot h_n^{\beta+d} \prod_{i=1}^d \left(c_0 \wedge \frac{c_\ell}{|k_i h_n|^\ell} \right). \quad (5.10)$$

We will choose $\ell \geq 2$ and $\epsilon > 0$ to satisfy (5.8) for all sufficiently small h_n . Notice that $\|k\|^\beta \leq d^{(\beta-1) \vee 0} \sum_{i=1}^d |k_i|^\beta$. Hence, (5.8) follows from

$$d^{(\beta-1) \vee 0} \sum_{i=1}^d \sum_{k \in \mathbb{Z}^d} |C_{g_n}(k)| (1 + |k_i|^\beta) = d^{\beta \vee 1} \sum_{k \in \mathbb{Z}^d} |C_{g_n}(k)| (1 + |k_1|^\beta) < \frac{L}{4},$$

where $k = (k_i)_{i=1}^d$. Indeed, this follows from the observation that, by (5.9), we have

$$\sum_{i=1}^d \sum_{k \in \mathbb{Z}^d} |C_{g_n}(k_1, \dots, k_d)| |k_i|^\beta = d \sum_{k \in \mathbb{Z}^d} |C_{g_n}(k_1, \dots, k_d)| |k_1|^\beta.$$

Thus, it suffices to show that

$$\begin{aligned} \sum_{k \in \mathbb{Z}^d} |C_{g_n}(k)| (1 + |k_1|^\beta) &\leq |C_{g_n}(0)| + 2 \sum_{k=(k_i)_{i=1}^d \in \mathbb{Z}^d} |C_{g_n}(k)| |k_1|^\beta \\ &\leq \frac{L}{4(d^{\beta \vee 1})}. \end{aligned} \quad (5.11)$$

Since $h_n \in (0, 1)$, (5.10) readily implies that

$$|C_{g_n}(0)| \leq \epsilon \cdot c_0^d.$$

Also, applying (5.10),

$$\begin{aligned}
& \sum_{k=(k_i)_{i=1}^d \in \mathbb{Z}^d} |C_{g_n}(k)| |k_1|^\beta \\
& \leq \epsilon \cdot \sum_{k \in \mathbb{Z}^d} \left[h_n^{\beta+1} |k_1|^\beta \left(c_0 \wedge \frac{c_\ell}{|k_1 h_n|^\ell} \right) \cdot h_n^{d-1} \prod_{i=2}^d \left(c_0 \wedge \frac{c_\ell}{|k_i h_n|^\ell} \right) \right] \\
& = \epsilon \cdot \left[h_n \cdot \sum_{j \in \mathbb{Z}} |j h_n|^\beta \left(c_0 \wedge \frac{c_\ell}{|j h_n|^\ell} \right) \right] \times \left[h_n \sum_{j \in \mathbb{Z}} \left(c_0 \wedge \frac{c_\ell}{|j h_n|^\ell} \right) \right]^{d-1} \\
& =: \epsilon \times A_n \times (B_n)^{d-1}.
\end{aligned} \tag{5.12}$$

Observe that A_n and B_n are Riemann sums for the integrals

$$A := \int_{x \in \mathbb{R}} |x|^\beta \left(c_0 \wedge \frac{c_\ell}{|x|^\ell} \right) dx \quad \text{and} \quad B := \int_{x \in \mathbb{R}} \left(c_0 \wedge \frac{c_\ell}{|x|^\ell} \right) dx,$$

which are clearly finite for $\ell \geq \lfloor \beta \rfloor + 2$. Taking such a value of ℓ and using the fact that $A_n \rightarrow A$ and $B_n \rightarrow B$, as $h_n \rightarrow 0$, we obtain that the right hand side of (5.12) is bounded above by $2\epsilon \times A \times B^{d-1}$ for all sufficiently small h_n . Therefore, we can ensure that (5.11) holds by picking $\epsilon > 0$ such that

$$0 < \epsilon \cdot \left[c_0^d + 4A \times B^{d-1} \right] \leq \frac{L}{4(d^{\beta \vee 1})}.$$

This shows that $f_{1,n} \in \mathcal{P}_D(\beta, L)$ and completes the proof of Property (1).

Proof of Property (2). This is immediate. Relation (5.7) holds for all sufficiently large n since $g(\theta_0/h_n) \rightarrow g(0)\mathbb{1}(\theta_0 = 0)$, as $h_n \rightarrow 0$, by the fact that g is supported on $[-\pi, \pi]$.

Proof of Property (3). Let D_n and $B_{n,\xi}$ be the covariance matrices of the data $X(t)$, $t \in \{1, \dots, n\}^d$, that correspond to, respectively, the spectral densities $r_n(\theta) = g_n(\theta) +$

$g_n(-\theta)$ and $f_{0,n}(\theta) + \xi r_n(\theta)$, for some $\xi \in [0, 1]$. By Lemma V.6,

$$\text{KL}(\mathbb{P}_{1n}, \mathbb{P}_{0n}) \leq \frac{1}{2} \|D_n\|_{\mathbb{F}}^2 \|B_{n,\xi}^{-1}\|_{\text{op}}^2, \quad (5.13)$$

where $\|\cdot\|_{\mathbb{F}}$ is the Frobenius norm and $\|\cdot\|_{\text{op}}$ is the matrix operator norm induced by the Euclidian vector norm. It follows from part (iii) of Lemma V.5 applied to $A_n := D_n$ that

$$\frac{1}{n^d} \|D_n\|_{\mathbb{F}}^2 \leq (2\pi)^d \int_{\theta \in [-\pi, \pi]^d} r_n^2(\theta) d\theta.$$

Thus, recalling $r_n(\theta) = g_n(\theta) + g_n(-\theta)$, Relation (5.6), and using a change of variables, we obtain:

$$\frac{1}{n^d} \|D_n\|_{\mathbb{F}}^2 \leq (2\pi)^d h_n^{2\beta+d} \left\{ 4 \int_{\theta \in [-\pi, \pi]^d} g^2(\theta) d\theta \right\} = 4 \cdot (2\pi)^d \cdot \epsilon^2 h_n^{2\beta+d} \|\varphi\|_{L^2}^{2d}, \quad (5.14)$$

where we used (5.5). Applying (i) and (ii) of Lemma V.5, we obtain

$$\|B_{n,\xi}^{-1}\|_{\text{op}} \leq \frac{1}{(2\pi)^d} \sup_{\theta \in [-\pi, \pi]^d} [f_{0n}(\theta) + \xi r_n(\theta)]^{-1} \leq \frac{2}{L}, \quad (5.15)$$

since $r_n(\theta) \geq 0$ and $f_{0n}(\theta) = L/(2 \cdot (2\pi)^d)$, $\theta \in [-\pi, \pi]$. Combining (5.13) - (5.15),

$$\text{KL}(\mathbb{P}_{1n}, \mathbb{P}_{0n}) \leq \frac{1}{2} \|D_n\|_{\mathbb{F}}^2 \|B_{n,\xi}^{-1}\|_{\text{op}}^2 \leq \left(\frac{8 \cdot (2\pi)^d \epsilon^2 \|g\|_{L^2}^2}{L^2} \right) n^d h_n^{2\beta+d},$$

which is bounded, if we set

$$h_n = M \cdot n^{-d/(2\beta+d)}.$$

By picking $M = M_{\theta_0}$ so that $[g(0) + g(0)\mathbb{1}(\theta_0 = 0)] \cdot M^\beta = 1$, we have that for all sufficiently large n ,

$$|f_{1n}(\theta_0) - f_{0n}(\theta_0)| = [g(0) + g(0)\mathbb{1}(\theta_0 = 0)] \cdot M^\beta \cdot n^{-\frac{d\beta}{2\beta+d}} = n^{-\frac{d\beta}{2\beta+d}},$$

which is a lower bound in the estimation error. The proof is complete by appealing to Theorem 2.5(iii) of *Tsybakov* (2008). \square

5.3 Proof for continuous time

In this section, we present the proof of Theorem V.3. Recall that for this theorem we are interested in processes whose spectral density belongs in $\mathcal{P}_C(\beta, L)$.

Proof of Theorem V.3. As argued in the proof of Theorem V.1, it suffices to focus on the case of scalar-valued processes $\{X(t), t \in \mathbb{R}^d\}$. As for the discrete-time case, we will introduce two models with spectral densities $f_{0,n}(\theta)$ and $f_{1,n}(\theta)$, and corresponding auto-covariances $C_{0,n}(t)$ and $C_{1,n}(t)$. Consider the function:

$$f_{0,n}(\theta) := \epsilon \cdot \prod_{i=1}^d \phi(\delta_n \theta_i), \quad \theta = (\theta_i)_{i=1}^d \in \mathbb{R}^d, \quad (5.16)$$

where $\phi(z) = e^{-z^2/2}/\sqrt{2\pi}$, $z \in \mathbb{R}$ is the standard Normal density.

With a straightforward change of variables, we obtain:

$$C_{0,n}(x) = \int_{\mathbb{R}^d} e^{-ix^\top \theta} f_{0,n}(\theta) d\theta = \epsilon (2\pi)^{d/2} \cdot \delta_n^{-d} \prod_{i=1}^d \phi(x_i/\delta_n), \quad (5.17)$$

where we used the fact that $\int_{\mathbb{R}} e^{-ixu} \phi(u) du = \sqrt{2\pi} \phi(x)$.

As in the time-series setting, let

$$f_{1,n}(\theta) = f_{0,n}(\theta) + h_n^\beta \left[g\left(\frac{\theta - \theta_0}{h_n}\right) + g\left(\frac{\theta + \theta_0}{h_n}\right) \right], \quad (5.18)$$

where g is as in (5.5). Following the proof of Theorem V.1, we will verify the following.

Properties:

- (1) $f_{0,n}, f_{1,n} \in \mathcal{P}_C(\beta, L)$, where the class $\mathcal{P}_C(\beta, L)$ is defined in (4.36).

(2) The functions $f_{0,n}$ and $f_{1,n}$ satisfy Relation (5.7).

(3) The KL-divergence is bounded, i.e., $\sup_n \text{KL}(\mathbb{P}_{1n}, \mathbb{P}_{0n}) < \infty$, where $\mathbb{P}_{i,n}$ are the probability distributions of the data $\{X(\delta_n k), k \in \{1, \dots, n\}^d\}$ under the models $f_{i,n}$, $i = 0, 1$.

Property (2) above is immediate by definition since the difference $f_{0,n}(\theta) - f_{1,n}(\theta)$ is constructed as in the proof of Theorem V.1.

Proof of Property (1): The fact that $f_{0,n} \in \mathcal{P}_C(\beta, L)$ is straightforward. Indeed, by (5.17), we have

$$\begin{aligned} \int_{\mathbb{R}^d} (1 + \|x\|^\beta) |C_{0,n}(x)| dx &\leq \epsilon \cdot \delta_n^{-d} \int_{\mathbb{R}^d} (1 + \|x\|^\beta) e^{-\|x\|^2/2\delta_n^2} dx \\ &= \epsilon \int_{\mathbb{R}^d} (1 + \|\delta_n \cdot u\|^\beta) e^{-\|u\|^2/2} du \leq \epsilon \int_{\mathbb{R}^d} (1 + \|u\|^\beta) e^{-\|u\|^2/2} du \leq L/2, \end{aligned} \quad (5.19)$$

for all $\delta_n \in (0, 1)$ and for a sufficiently small $\epsilon > 0$. This follows from the fact that with $\delta_n \in (0, 1)$, we have $\|\delta_n u\|^\beta \leq \|u\|^\beta$ and the fact that $\int_{\mathbb{R}^d} (1 + \|u\|^\beta) e^{-\|u\|^2/2} du < \infty$. This ensures that (4.36) holds with C replaced by $C_{0,n}$ and L by $L/2$. That is, $f_{0,n} \in \mathcal{P}_C(\beta, L)$.

Now, we show that $f_{1,n}$ defined in (5.18) belongs to $\mathcal{P}_C(\beta, L)$, by perhaps lowering the value of $\epsilon > 0$. Let

$$C_{g_n}(x) := \int_{\mathbb{R}^d} e^{-i\theta^\top x} g_n(\theta) d\theta,$$

where g_n is as in (5.6). As argued in the proof of Theorem V.1, in view of (5.19), it suffices to show that

$$\int_{\mathbb{R}^d} (1 + \|x\|^\beta) |C_{g_n}(x)| dx \leq \frac{L}{4}. \quad (5.20)$$

Note that Relation (5.10) remains valid if $k \in \mathbb{Z}^d$ therein is replaced with $x \in \mathbb{R}^d$.

Therefore, (5.20) follows by picking a possibly smaller value of $\epsilon > 0$, provided

$$h_n^{\beta+d} \int_{\mathbb{R}^d} (1 + \|x\|^\beta) \prod_{i=1}^d \left(c_0 \wedge \frac{c_\ell}{|h_n x_i|^{\ell_i}} \right) dx < \infty,$$

for some $\ell \in \mathbb{N}$, $i = 1, \dots, d$. Notice that for $h_n \in (0, 1]$, we have $h_n^\beta(1 + \|x\|^\beta) \leq (1 + \|h_n x\|^\beta)$, and hence the last integral is bounded above by

$$h_n^d \int_{\mathbb{R}^d} (1 + \|h_n x\|^\beta) \prod_{i=1}^d \left(c_0 \wedge \frac{c_\ell}{|h_n x_i|^\ell} \right) dx = \int_{\mathbb{R}^d} (1 + \|u\|^\beta) \prod_{i=1}^d \left(c_0 \wedge \frac{c_\ell}{|u_i|^\ell} \right) du,$$

where we used the change of variables $u := h_n x$. Clearly, the last integral is finite provided $\ell \geq \lfloor \beta \rfloor + 2$. This implies that (5.20) holds with a suitably chosen $\epsilon > 0$, showing that $f_{1,n} \in \mathcal{P}_C(\beta, L)$ and completing the proof of Property (2).

Proof of Property (3): Now, as in the proof of Theorem V.1 we will bound the KL-divergence $\text{KL}(\mathbb{P}_{1,n}, \mathbb{P}_{0,n})$, where $\mathbb{P}_{i,n}$ is the law of the Gaussian vector $\{X_i(\delta_n k), k \in \{1, \dots, n\}^d\}$ under the model $f_{i,n}$, $i = 0, 1$.

Observe that the \mathbb{Z}^d -indexed stationary process $\{X_i(\delta_n k), k \in \mathbb{Z}^d\}$ has the so-called folded spectral density

$$\tilde{f}_{i,h}(\theta) := \delta_n^{-d} \sum_{\ell \in \mathbb{Z}^d} f_{i,n} \left(\frac{\theta + 2\pi\ell}{\delta_n} \right), \quad \theta \in [-\pi, \pi]^d, \quad i = 0, 1. \quad (5.21)$$

We shall apply the same argument as in the proof of Theorem V.1 based on Samarov's Lemmas V.6 and V.5 applied to the *folded spectral densities*.

For $\xi \in [0, 1]$, let D_n and $B_{n,\xi}$ be the covariance matrices of zero-mean Gaussian vectors having spectral densities $\tilde{r}_n(\theta) := \tilde{f}_{1,n}(\theta) - \tilde{f}_{0,n}(\theta)$ and $\tilde{f}_{0,n}(\theta) + \xi \tilde{r}_n(\theta)$, $\theta \in [-\pi, \pi]^d$, respectively, where

$$\tilde{r}_n(\theta) = h_n^\beta \delta_n^{-d} \sum_{\ell \in \mathbb{Z}^d} \left[g \left(\frac{\theta + 2\pi\ell}{h_n \delta_n} - \frac{\theta_0}{h_n} \right) + g \left(\frac{\theta + 2\pi\ell}{h_n \delta_n} + \frac{\theta_0}{h_n} \right) \right].$$

Then, by Lemma V.6, we have

$$\text{KL}(\mathbb{P}_{1,n}, \mathbb{P}_{0,n}) \leq \frac{1}{2} \|D_n\|_{\text{F}}^2 \|B_{n,\xi}^{-1}\|_{\text{op}}^2, \quad (5.22)$$

for some $\xi \in [0, 1]$. As in (5.14) from Lemma V.5(iii) applied to $A_n := D_n$, we obtain

$$\begin{aligned} \|D_n\|_{\text{F}}^2 &\leq (2\pi)^d \cdot n^d \int_{[-\pi, \pi]^d} \tilde{r}_n(\theta)^2 d\theta \\ &\leq 4 \cdot (2\pi)^d \cdot n^d h_n^{2\beta} \delta_n^{-2d} \int_{[-\pi, \pi]^d} \sum_{\ell \in \mathbb{Z}^d} g\left(\frac{\theta + 2\pi\ell}{h_n \delta_n} - \frac{\theta_0}{h_n}\right)^2 d\theta \\ &= 4 \cdot (2\pi)^d \cdot n^d h_n^{2\beta} \delta_n^{-d} \int_{\mathbb{R}^d} g\left(\frac{u}{h_n} - \frac{\theta_0}{h_n}\right)^2 du \\ &= 4 \cdot (2\pi)^d \cdot n^d h_n^{2\beta+d} \delta_n^{-d} \|g^2\|_{L^2}^2 = 4\epsilon \cdot (2\pi)^d \cdot n^d h_n^{2\beta+d} \delta_n^{-d} \|\varphi\|_{L^2}^2, \end{aligned} \quad (5.23)$$

where in the last two integrals we made changes of variables, and the last relation follows from the definition of g in (5.5).

Now, we deal with bounding $\|B_{n,\xi}^{-1}\|_{\text{op}}$. Notice that $B_{n,\xi}$ is the covariance matrix of a Gaussian vector $\{X_\xi(\delta_n k), k \in \{1, \dots, n\}^d\}$ coming from a stationary process $Y(k) = X_\xi(\delta_n k)$, $k \in \mathbb{Z}^d$ with spectral density $\tilde{f}_{0,n}(\theta) + \xi \tilde{r}_n(\theta)$, $\theta \in [-\pi, \pi]^d$, where $\tilde{r}_n(\theta) \geq 0$ and $\xi \in [0, 1]$. By Lemma V.5(ii), we then have that

$$\|B_{n,\xi}^{-1}\|_{\text{op}} \leq \sup_{\theta \in [-\pi, \pi]^d} \left[\tilde{f}_{0,n}(\theta) + \xi \tilde{r}_n(\theta) \right]^{-1} \leq \sup_{\theta \in [-\pi, \pi]^d} \left[\tilde{f}_{0,n}(\theta) \right]^{-1}.$$

Recalling the definition of $\tilde{f}_{0,n}$ in (5.16) and the *folded* spectral density in (5.21), we obtain $\tilde{f}_{0,n}(\theta) \geq \epsilon \delta_n^{-d} \prod_{i=1}^d \phi(\theta_i) \geq \epsilon \delta_n^{-d} e^{-d\pi^2/2} / (2\pi)^{d/2}$ for $\theta \in [-\pi, \pi]^d$. Hence

$$\|B_{n,\xi}^{-1}\|_{\text{op}} \leq \frac{1}{(2\pi)^d} \sup_{\theta \in [-\pi, \pi]^d} \left[\tilde{f}_{0,n}(\theta) \right]^{-1} \leq \frac{e^{d\pi^2/2}}{(2\pi)^{d/2} \cdot \epsilon} \cdot \delta_n^d \quad (5.24)$$

Finally, by (5.22), (5.23), and (5.24),

$$\text{KL}(\mathbb{P}_{1n}, \mathbb{P}_{0n}) \leq \frac{1}{2} \|D_n\|_{\text{F}}^2 \|B_{n,\xi}^{-1}\|_{\text{op}}^2 \leq c \cdot n^d h_n^{2\beta+d} \delta_n^{-d} \cdot \delta_n^{2d} = c \cdot (n\delta_n)^d h_n^{2\beta+d},$$

where $c = 2\epsilon^{-1}\|\varphi\|_{L^2}^{2d}e^{d\pi^2}$. Thus, the KL-divergence is uniformly bounded if we set

$$h_n = M \cdot (n\delta_n)^{-d/(2\beta+d)}.$$

Picking M so that $g(0)(1 + \mathbb{1}(\theta_0 = 0)) \cdot M^\beta = 1$, we have that

$$|f_{1n}(\theta_0) - f_{0n}(\theta_0)| = g(0)(1 + \mathbb{1}(\theta_0 = 0)) \cdot M^\beta \cdot (n\delta_n)^{-\frac{d\beta}{2\beta+d}} = (n\delta_n)^{-\frac{d\beta}{2\beta+d}},$$

which, by appealing to Theorem 2.5(iii) of *Tsybakov* (2008), yields the desired lower bound in estimation error in (5.2). \square

5.4 Samarov-type lemmas

The technical lemmas needed in the above proofs come from *Samarov* (1977). The first is a slight extension to the d -dimensional case. We provide proofs below for the sake of completeness.

Lemma V.5. *Let a_j , $j \in \mathbb{Z}^d$ be a sequence of numbers such that $\sum_{j \in \mathbb{Z}^d} |a_j|^2 < \infty$ and $a_j = a_{-j}$. Let also A_n be a matrix of dimensions $n^d \times n^d$, whose (j, k) -th element equals a_{j-k} , where j and k are multi-indices that belong to $[0 : n - 1]^d := \{0, 1, \dots, n - 1\}^d$ (i.e., the (j, k) -th element based on a natural ordering of the multi-indices of $[0 : n - 1]^d$). Finally, define $\alpha(\lambda) = (2\pi)^{-d} \sum_{j \in \mathbb{Z}^d} a_j e^{ij^\top \lambda}$, for $\lambda \in [-\pi, \pi]^d$. Then, for the norms of A_n , the following claims are true.*

$$(i) \quad \|A_n\|_{\text{op}} \leq (2\pi)^d \cdot \sup_{\lambda \in [-\pi, \pi]^d} |\alpha(\lambda)|.$$

$$(ii) \quad \text{If } A_n \text{ is positive definite, then } \|A_n^{-1}\|_{\text{op}} \leq (2\pi)^{-d} \cdot \sup_{\lambda \in [-\pi, \pi]^d} |1/\alpha(\lambda)|.$$

$$(iii) \quad n^{-d} \|A_n\|_{\text{F}}^2 \leq \sum_{j \in \mathbb{Z}^d} |a_j|^2 = (2\pi)^d \int_{[-\pi, \pi]^d} \alpha^2(\lambda) d\lambda.$$

Proof. Let $N := n^d$ and use the notation $[0 : n - 1]^d := \{0, 1, \dots, n - 1\}^d$. We will follow the arguments in *Samarov* (1977).

(i) Since A_n is a symmetric matrix, we have that

$$\|A_n\|_{\text{op}} = \sup \{ |x^\top A_n y| : \|x\|_2 = \|y\|_2 = 1, x, y \in \mathbb{R}^N \},$$

where is the matrix (operator) norm induced by the Euclidian vector norm. Now, let $x = (x_i)_{i \in [0:n-1]^d}$ and $y = (y_i)_{i \in [0:n-1]^d}$. By Fourier inversion, we have that

$$a_{j-k} = \int_{[-\pi, \pi]^d} e^{-i(j-k)^\top \lambda} \alpha(\lambda) d\lambda.$$

Therefore,

$$\begin{aligned} |x^\top A_n y| &= \left| \sum_{j \in [0:n-1]^d} \sum_{k \in [0:n-1]^d} \int_{[-\pi, \pi]^d} x_j e^{-i(j-k)^\top \lambda} y_k \alpha(\lambda) d\lambda \right| \\ &\leq \int_{[-\pi, \pi]^d} \left| \sum_{j \in [0:n-1]^d} x_j e^{ij^\top \lambda} \right| \cdot |\alpha(\lambda)| \cdot \left| \sum_{k \in [0:n-1]^d} e^{ik^\top \lambda} y_k \right| d\lambda \\ &\leq \sup_{\lambda \in [-\pi, \pi]^d} |\alpha(\lambda)| \cdot \left(\int_{[-\pi, \pi]^d} \left| \sum_{j \in [0:n-1]^d} x_j e^{ij^\top \lambda} \right|^2 d\lambda \right)^{1/2} \\ &\quad \times \left(\int_{[-\pi, \pi]^d} \left| \sum_{k \in [0:n-1]^d} y_k e^{ik^\top \lambda} \right|^2 d\lambda \right)^{1/2} \\ &= (2\pi)^d \cdot \sup_{\lambda \in [-\pi, \pi]^d} |\alpha(\lambda)| \cdot \|x\|_2 \cdot \|y\|_2, \end{aligned}$$

The second inequality follows from the Cauchy-Schwarz inequality and the last equality follows by Parseval's identity, since the functions $\varphi_k(\lambda) := (2\pi)^{-d/2} e^{ik^\top \lambda}$, $\lambda \in [-\pi, \pi]$, $k \in [0 : (n-1)]^d$, are orthonormal in $L^2([-\pi, \pi]^d; \mathbb{C})$.

(ii) If A_n is also positive definite, then A_n is invertible.

Since $\|\cdot\|_{\text{op}}$ is the spectral norm, we have that

$$\|A_n^{-1}\|_{\text{op}} = \max \sigma_i(A_n^{-1}) = \max \frac{1}{\sigma_i(A_n)} = \sup \left\{ \frac{1}{x^\top A_n x} : \|x\|_2 = 1, x \in \mathbb{R}^N \right\}$$

where the last equality follows from Rayleigh quotient optimization results, when A_n is positive definite.

As in part (1), with z^* denoting the complex conjugate of z , we have that

$$\begin{aligned}
|x^\top A_n x| &= x^\top A_n x = \int_{[-\pi, \pi]^d} \left(\sum_{j \in [0:n-1]^d} x_j e^{ij^\top \lambda} \right) \alpha(\lambda) \left(\sum_{k \in [0:n-1]^d} e^{ik^\top \lambda} y_k \right) d\lambda \\
&= \int_{[-\pi, \pi]^d} \left| \sum_{j \in [0:n-1]^d} x_j e^{-ij^\top \lambda} \right|^2 \alpha(\lambda) d\lambda \\
&\geq \inf_{\lambda \in [-\pi, \pi]^d} |\alpha(\lambda)| \int_{[-\pi, \pi]^d} \left| \sum_{j \in [0:n-1]^d} x_j e^{-ij^\top \lambda} \right|^2 d\lambda \\
&= \left(\sup_{\lambda \in [-\pi, \pi]^d} |\alpha^{-1}(\lambda)| \right)^{-1} \cdot \|x\|_2^2 \cdot (2\pi)^d,
\end{aligned}$$

and the result follows.

(iii) It follows that

$$\|A_n\|_{\text{F}}^2 = \sum_{i \in [0:n-1]^d} \sum_{j \in [0:n-1]^d} |a_{i-j}|^2 = \sum_{k \in [-(n-1):(n-1)]^d} \prod_{i=1}^d (n - |k_i|) |a_k|^2,$$

where $[-(n-1):(n-1)]^d := \{-(n-1), \dots, n-1\}^d$. Thus,

$$\frac{1}{n^d} \|A_n\|_{\text{F}}^2 = \sum_{k \in [-(n-1):(n-1)]^d} \prod_{i=1}^d \left(1 - \frac{|k_i|}{n}\right) |a_k|^2 \leq \sum_{j \in \mathbb{Z}^d} |a_j|^2 = (2\pi)^d \int_{[-\pi, \pi]^d} \alpha^2(\lambda) d\lambda,$$

by Parseval's identity.

□

Lemma V.6. *Let B_0 and B_1 be symmetric, positive definite $n \times n$ matrices such that $D := B_1 - B_0$ is non-negative definite. Let P_0 and P_1 be the probability distributions of zero-mean Gaussian vectors with covariance matrices B_0 and B_1 , respectively. Then,*

there is a $\xi \in [0, 1]$ such that

$$\text{KL}(P_1, P_0) \leq \frac{1}{2} \|D\|_{\text{F}}^2 \|B_\xi^{-1}\|_{\text{op}}^2,$$

where $B_\xi := B_0 + \xi D$, and where $\|\cdot\|_{\text{F}}$ stands for the matrix Frobenius norm and $\|\cdot\|_{\text{op}}$ stands for the matrix operator norm induced by the Euclidean vector norm.

Proof. Since the data is assumed Gaussian, we can immediately obtain that

$$\text{KL}(P_1, P_0) = \frac{1}{2} \left\{ \text{tr}(B_1 B_0^{-1} - E) + \log |B_0| - \log |B_1| \right\}, \quad (5.25)$$

where $\text{tr}(A)$ and $|A|$ are the trace and determinant of the matrix A and E is the identity matrix of dimensions $n \times n$. Notice that $B_\lambda := B_0 + \lambda D$, $\lambda \in [0, 1]$ is a positive definite covariance matrix. The expression in (5.25) can be rewritten as

$$\text{KL}(P_1, P_0) = \frac{1}{2} \left\{ \text{tr}[B_1(B_0^{-1} - B_1^{-1})] + \log |B_0| - \log |B_1| \right\}. \quad (5.26)$$

We define the function $\phi(\lambda) = \text{tr}(B_1 B_\lambda^{-1}) + \log |B_\lambda|$ and note that by the intermediate value theorem, we have

$$\text{KL}(P_1, P_0) = \frac{\phi(0) - \phi(1)}{2} = -\frac{1}{2} \cdot \phi'(\xi),$$

for some $\xi \in [0, 1]$. Using the following differentiation rules

$$\begin{aligned} \frac{d}{d\lambda} A^{-1}(\lambda) &= -A^{-1}(\lambda) \left(\frac{d}{d\lambda} A(\lambda) \right) A^{-1}(\lambda), \\ \frac{d}{d\lambda} \log |A(\lambda)| &= \text{tr} \left(A^{-1}(\lambda) \frac{d}{d\lambda} A(\lambda) \right), \end{aligned}$$

and the fact that $dB_\lambda/d\lambda = D$, we obtain that (5.26) becomes:

$$\begin{aligned} \text{KL}(P_1, P_0) &= \frac{1}{2} \cdot \text{tr} [B_1 B_\xi^{-1} D B_\xi^{-1} - D B_\xi^{-1}] \\ &= \frac{1}{2} \cdot \text{tr} [(B_1 - B_\xi) B_\xi^{-1} D B_\xi^{-1}], \end{aligned} \quad (5.27)$$

for some $\xi \in [0, 1]$. To estimate the rhs of (5.27) we will use the following inequalities (see, e.g., *Davies, 1973*):

$$|\text{tr}(AB)| \leq \|A\|_{\text{F}} \cdot \|B\|_{\text{F}}. \quad (5.28)$$

If the matrices A and B are symmetric, then

$$\|BA\|_{\text{F}} = \|AB\|_{\text{F}} \leq \|A\|_{\text{op}} \|B\|_{\text{F}}, \quad (5.29)$$

where

$$\|A\|_{\text{F}} = \left(\sum_{i,j} a_{ij}^2 \right)^{1/2}, \quad \|A\|_{\text{op}} = \sup\{\|Ax\|_2; \|x\|_2 = 1, x \in \mathbb{R}^n\}.$$

From (5.27) with the help of (5.28) and (5.29) we obtain:

$$\text{KL}(P_1, P_0) \leq \frac{1}{2} \|B_1 - B_\xi\|_{\text{F}} \|D\|_{\text{F}} \|B_\xi^{-1}\|_{\text{op}}^2 \leq \frac{1}{2} \|D\|_{\text{F}}^2 \|B_\xi^{-1}\|_{\text{op}}^2.$$

□

CHAPTER VI

Asymptotic Normality

In this chapter, we investigate the asymptotic distribution of the spectral density estimator introduced in Chapter IV. We adopt the extra assumption that the process X in the complex Hilbert space \mathbb{H} is Gaussian. With this assumption in place, we establish that the asymptotic distribution of our proposed estimator is also Gaussian, in what can be considered a Central Limit Theorem result. The proof is based on the computation of all the moments of the standardized estimator, which we achieve by establishing a novel Isserlis type formula.

6.1 Asymptotic distribution

In this chapter, we continue to consider the case of gridded data described by (4.26) and (4.27). The goal here is to present a central limit theorem for our spectral density estimator $\hat{f}_n(\theta)$ assuming that $\{X(t)\}$ is a stationary Gaussian process, where in this section we do not restrict X to be real in \mathbb{H} . However, due to the technical nature of this topic, we will focus on the case $d = 1$. As discussed in Remark IV.7, for $d = 1$ the normalization $|\mathbb{T}_n \cap (\mathbb{T}_n - (t - s))|$ in $\hat{f}_n(\theta)$ does not affect the rate. Thus, for convenience, we will eliminate that and consider instead

$$\hat{f}_n(\theta) = \frac{\delta_n}{2\pi n} \sum_{i,j=1}^n e^{i(i-j)\delta_n\theta} X(\delta_n i) \otimes X(\delta_n j) K\left(\frac{i-j}{\Delta_n} \cdot \delta_n\right). \quad (6.1)$$

We will prove a central limit theorem for $\hat{f}_n(\theta)$ assuming that $\delta_n \rightarrow$ some $\delta_\infty \in [0, \infty)$ as $n \rightarrow \infty$. The time-series and mixed-domain cases are covered by $\delta_\infty = 1$ and 0 , respectively.

Interestingly, the asymptotic distribution of $\hat{f}_n(\theta)$ involves the notion of pseudo-covariance. Recall that from (4.4) the pseudo-covariance function is defined as $\check{C}(h) = \mathbb{E}[X(t+h) \otimes \overline{X}(t)]$. In accordance with (4.6) and (4.29), define the *pseudo-spectral density*:

$$\check{f}(\theta) = \frac{1}{2\pi} \int_{\mathbb{R}} e^{-i\theta x} \check{C}(x) dx, \quad \theta \in \mathbb{R},$$

and, for $\delta > 0$, the *folded pseudo-spectral density*:

$$\check{f}(\theta; \delta) := \frac{\delta}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik^\top \theta \delta} \check{C}(k\delta), \quad \theta \in [-\pi/\delta, \pi/\delta].$$

Note that $\check{f}(\theta)$ and $\check{f}(\theta; \delta)$ are well defined assuming that $\int_{\mathbb{R}} \|\check{C}(x)\|_{\text{tr}} dx < \infty$ and $\sum_{k=-\infty}^{\infty} \|\check{C}(k\delta)\|_{\text{tr}} < \infty$, respectively. For convenience, also write $f(\theta; 0) = f(\theta)$ and $\check{f}(\theta; 0) = \check{f}(\theta)$.

Let now $\{e_j\}$ be an arbitrary fixed CONS of \mathbb{H} , and define

$$C_{k,\ell}(t) = \langle C(t)e_k, e_\ell \rangle = \mathbb{E}[\langle X(t), e_\ell \rangle \overline{\langle X(0), e_k \rangle}],$$

$$\check{C}_{k,\ell}(t) = \langle \check{C}(t)\overline{e}_k, e_\ell \rangle = \mathbb{E}[\langle X(t), e_\ell \rangle \langle X(0), e_k \rangle].$$

The following assumption will be needed for establishing the central limit theorem.

Assumption CLT. Let the grid size δ_n and bandwidth Δ_n satisfy $\delta_n \rightarrow$ some $\delta_\infty \in [0, \infty)$ and $(n\delta_n)/\Delta_n \rightarrow \infty$. Also, assume that there exist positive constants L_n such that

$$L_n \delta_n \rightarrow \infty, \quad L_n / \Delta_n \rightarrow 0,$$

and for which the following hold:

- (a) $\sup_n \delta_n \sum_{x=-\infty}^{\infty} \|C(\delta_n x)\|_{\text{tr}} < \infty$ and $\delta_n \sum_{|x| > L_n} \|C(\delta_n x)\|_{\text{tr}} \rightarrow 0$;

- (b) $\|f(\theta; \delta_n) - f(\theta; \delta_\infty)\|_{\text{tr}} \rightarrow 0$;
- (c) $\sup_n \delta_n \sum_{x=-\infty}^{\infty} \|\check{C}(\delta_n x)\|_{\text{tr}} < \infty$ and $\delta_n \sum_{|x|>L_n} \|\check{C}(\delta_n x)\|_{\text{tr}} \rightarrow 0$;
- (d) $\|\check{f}(\theta; \delta_n) - \check{f}(\theta; \delta_\infty)\|_{\text{tr}} \rightarrow 0$;
- (e) $\delta_n^2 \sum_{x_1, x_2=-\infty}^{\infty} |C_{k,k}(\delta_n x_1)| \cdot |C_{\ell,\ell}(\delta_n x_2)| \leq a_{k,\ell}$, such that $\sum_{k,\ell} a_{k,\ell} < \infty$;
- (f) $\delta_n^2 \sum_{x_1, x_2=-\infty}^{\infty} |\check{C}_{k,\ell}(\delta_n x_1)| \cdot |\check{C}_{k,\ell}(\delta_n x_2)| \leq b_{k,\ell}$, such that $\sum_{k,\ell} b_{k,\ell} < \infty$.

Note that if $\delta_n = \delta_\infty \in (0, \infty)$ for all n , then the conditions (a)-(f) follow from $\sum_{x=-\infty}^{\infty} \|C(\delta_\infty x)\|_{\text{tr}} < \infty$ and $\sum_{x=-\infty}^{\infty} \|\check{C}(\delta_\infty x)\|_{\text{tr}} < \infty$. For $\delta_\infty = 0$, the conditions (a) and (c) in the above assumption are related to the notion of directly Riemann integrability (dRi) (cf., e.g., *Feller*, 2008); if, in addition, $C(x)$ and $\check{C}(x)$ are functions in \mathbb{C} , then the dRi of $C(x)e^{ix\theta}$ and $\check{C}(x)e^{-ix\theta}$ also implies (b) and (d) respectively.

The following modified assumption on the kernel K is also needed.

Assumption K'. The nonnegative kernel K has compact support, is symmetric about 0, and is of bounded variation.

The following result is a central limit theorem for $\hat{f}_n(\theta)$, where the weak convergence is defined in the space \mathbb{X} of Hilbert-Schmidt operators on \mathbb{H} .

Theorem VI.1. *Consider the stationary zero-mean Gaussian process $\{X(t), t \in \mathbb{R}\}$ and assume that Assumptions CLT and K' hold. Define*

$$\mathcal{T}_n(\theta) := \sqrt{\frac{n\delta_n}{\Delta_n}} \left[\hat{f}_n(\theta) - \mathbb{E}\hat{f}_n(\theta) \right], \quad \theta \in \mathbb{R},$$

where $\hat{f}_n(\theta)$ is given in (6.1). Then, for any $\theta \in [-\pi/\delta_\infty, \pi/\delta_\infty]$, which is taken as \mathbb{R} if $\delta_\infty = 0$,

$$\mathcal{T}_n(\theta) \xrightarrow{d} \mathcal{T}(\theta) \quad \text{in } \mathbb{X},$$

where $\mathcal{T}(\theta)$ is a zero-mean Gaussian element of \mathbb{X} , such that for every finite collection $\{g_\ell, \ell = 1, \dots, m\}$, and positive numbers $\{a_\ell, \ell = 1, \dots, m\}$,

$$\begin{aligned} & \text{Var} \left(\sum_{\ell=1}^m a_\ell \langle \mathcal{T}(\theta) g_\ell, g_\ell \rangle \right) \\ &= \|K\|_2^2 \sum_{\ell_1, \ell_2=1}^m a_{\ell_1} a_{\ell_2} \left[|\langle f(\theta; \delta_\infty) g_{\ell_2}, g_{\ell_1} \rangle|^2 + c(\theta) |\langle \check{f}(\theta; \delta_\infty) \overline{g_{\ell_2}}, g_{\ell_1} \rangle|^2 \right], \end{aligned} \quad (6.2)$$

where $\|K\|_2^2 = \int K^2(x) dx$, and $c(\theta) = I_{(\theta=0)}$ if $\delta_\infty = 0$ and $I_{(\theta=0, \pm\pi/\delta_\infty)}$ if $\delta_\infty > 0$.

Remark VI.2. 1. Observe that the quantity $\sum_{\ell=1}^m a_\ell \langle \mathcal{T}(\theta) g_\ell, g_\ell \rangle$ in (6.2) is real since K is assumed symmetric.

2. The variances in (6.2) for all choices of $\{a_\ell\}$ and $\{g_\ell\}$ completely characterize the distribution of \mathcal{T} . The expression $\langle \check{f}(\theta) \overline{g_{\ell_2}}, g_{\ell_1} \rangle$ in (6.2) does not depend on the choice of real CONS, since

$$\langle \check{C}(t, s) \overline{g}, h \rangle = \mathbb{E} \langle X(t), h \rangle \langle \overline{g}, \overline{X(s)} \rangle = \mathbb{E} \langle X(t), h \rangle \langle X(s), g \rangle, \quad g, h \in \mathbb{H}.$$

The proof of this result, given in Section 6.2, is based on verifying the convergence of “all moments” of the estimator together with a tightness condition.

The previous result does not provide an explicit representation of the limit. In what follows, we obtain such an explicit, stochastic representation of $\mathcal{T}(\theta)$ for $c(\theta) = 0$, where $c(\theta)$ as in (6.2). Define the complex Gaussian random variables $Z_{i,j}$ ’s as follows:

$$Z_{i,j} = \xi_{i,j} + \mathfrak{i} \eta_{i,j}, \quad i < j, \quad (6.3)$$

where $\xi_{i,j}$ and $\eta_{i,j}$ are iid $N(0, 1/2)$ and $Z_{j,i} := \overline{Z_{i,j}}$. For $i = j$, we have the $Z_{i,i}$ ’s are real and $N(0, 1)$, independent from the $Z_{i,j}$ ’s, for $i \neq j$. Then, one obtains that the

$Z_{i,j}$'s are zero-mean complex Gaussian variables such that

$$Z_{i,j} = \overline{Z_{j,i}} \quad \text{and} \quad \mathbb{E}[Z_{i,j} \overline{Z_{i',j'}}] = \delta_{(i,j),(i',j')}. \quad (6.4)$$

Corollary VI.3. *Let $c(\theta) = 0$ in (6.2) and assume the conditions of Theorem VI.1. Let $\{e_i(\theta)\}$ be the (not necessarily real) CONS diagonalizing $f(\theta)$, i.e.,*

$$f(\theta) = \sum_i \lambda_i(\theta) e_i(\theta) \otimes e_i(\theta).$$

The random variable $\mathcal{T}(\theta)$ has the stochastic representation

$$\mathcal{T}(\theta) \stackrel{d}{=} \|K\|_2 \sum_{i,j} \sqrt{\lambda_i(\theta) \lambda_j(\theta)} Z_{i,j} e_i(\theta) \otimes e_j(\theta), \quad (6.5)$$

where $Z_{i,j}$ as defined in (6.3). In particular, the covariance operator of $\mathcal{T}(\theta)$ is

$$\mathbb{E}[\mathcal{T}(\theta) \otimes_{\text{HS}} \mathcal{T}(\theta)] = \|K\|_2^2 \sum_{i,j} \lambda_i(\theta) \lambda_j(\theta) (e_i(\theta) \otimes e_j(\theta)) \otimes_{\text{HS}} (e_i(\theta) \otimes e_j(\theta)).$$

Proof. Let g_ℓ , $\ell = 1, \dots, m$ be arbitrary in \mathbb{H} and suppose

$$g_\ell = \sum_i x_i(\ell) e_i, \quad x_i(\ell) \in \mathbb{C}.$$

Then, by Theorem VI.1, it is enough to verify that the representation of \mathcal{T} in (6.5) satisfies

$$\text{Var} \left(\sum_{\ell=1}^m a_\ell \langle \mathcal{T} g_\ell, g_\ell \rangle \right) = \|K\|_2^2 \sum_{\ell_1, \ell_2} a_{\ell_1} a_{\ell_2} |\langle f(\theta) g_{\ell_2}, g_{\ell_1} \rangle|^2, \quad (6.6)$$

for real constants $a_\ell \in \mathbb{R}$, $\ell = 1, \dots, m$. Observe that

$$\langle \mathcal{T} g_\ell, g_\ell \rangle = \|K\|_2 \sum_{i,j} \sqrt{\lambda_i \lambda_j} Z_{i,j} x_i(\ell) \overline{x_j(\ell)}.$$

Thus, in view of (6.4), the LHS of (6.6) equals

$$\begin{aligned} & \|K\|_2^2 \sum_{\ell_1, \ell_2} a_{\ell_1} a_{\ell_2} \sum_{i, i', j, j'} x_i(\ell_1) \overline{x_j(\ell_1) x_{i'}(\ell_2) x_{j'}(\ell_2)} \sqrt{\lambda_i \lambda_j \lambda_{i'} \lambda_{j'}} \mathbb{E}[Z_{i,j} \overline{Z_{i',j'}}] \\ & = \|K\|_2^2 \sum_{\ell_1, \ell_2, i, j} a_{\ell_1} a_{\ell_2} \lambda_i \lambda_j x_i(\ell_1) \overline{x_j(\ell_1) x_i(\ell_2)}. \end{aligned} \tag{6.7}$$

The latter expression is the RHS of (6.6). On the other hand,

$$\langle f(\theta) g_{\ell_2}, g_{\ell_1} \rangle = \sum_i \lambda_i x_i(\ell_1) \overline{x_i(\ell_2)}.$$

Thus, it is easy to see that that the right-hand sides of (6.6) and (6.7) are the same. \square

We end this section with the following remark.

Remark VI.4. Observe that $\mathcal{T}(\theta)$, for $c(\theta) = 0$ in (6.2), is a zero-mean random element in the Hilbert space \mathbb{X} of Hilbert-Schmidt operators. Therefore, Relation (6.5) provides its Karhunen-Loève type representation. That is, the covariance operator of $\mathcal{T}(\theta)$ is diagonalized in the basis $e_{i,j}(\theta) := e_i(\theta) \otimes e_j(\theta)$, $(i, j) \in \mathbb{N}^2$, where $\{e_i(\theta)\}$ is the CONS of \mathbb{H} diagonalizing the operator $f(\theta)$. The eigenvalues of the covariance operator $\mathbb{E}[\mathcal{T}(\theta) \otimes \mathcal{T}(\theta)]$ are precisely $\lambda_{i,j}(\theta) := \lambda_i(\theta) \lambda_j(\theta)$, where the $\lambda_i(\theta)$'s are the eigenvalues of $f(\theta)$.

6.2 Overview of Central Limit Theorem proof

This section presents the proof of Theorem VI.1. The proof relies on the standardized estimator $\{\mathcal{T}_n\}$ satisfying two properties; flat concentration and convergence of moments. Because of the complexity of the verification of these properties for $\{\mathcal{T}_n\}$, we establish them separately, in Sections 6.3 and 6.4 respectively.

As stated in Section 6.1, $\{X(t), t \in \mathbb{R}\}$ is a stationary Gaussian process in the

complex Hilbert space \mathbb{H} , we observe X at $t = k\delta_n, k = 1, \dots, n$, where $n\delta_n \rightarrow \infty$. We will focus on the case $\delta_n \equiv 1$, which contains the main ideas of the proof. The proof for the general case follows from a straightforward adaptation of the special case.

As mentioned, we focus on the case $\delta_n \equiv 1$, which is essentially the time-series setting (cf. Section 4.5.1). Thus, we consider a discrete-time stationary process $X = \{X(t), t \in \mathbb{Z}\}$, where $X(t)$ are Gaussian elements of the complex Hilbert space \mathbb{H} . We explained in Section 6.1 that the conditions (a)-(f) in Assumption CLT hold in this case if

$$\sum_{x=-\infty}^{\infty} [\|C(x)\|_{\text{tr}} + \|\check{C}(x)\|_{\text{tr}}] < \infty, \quad (6.8)$$

which we assume below. Note that C and \check{C} are defined in (4.2) and (4.4) respectively. Recall that \mathbb{X} denotes the Hilbert space of Hilbert-Schmidt operators $\mathcal{A} : \mathbb{H} \rightarrow \mathbb{H}$, equipped with the HS-inner product $\langle \mathcal{A}, \mathcal{B} \rangle_{\text{HS}} := \text{trace}(\mathcal{B}^* \mathcal{A})$, $\mathcal{A}, \mathcal{B} \in \mathbb{X}$, and corresponding norm $\|\mathcal{A}\|_{\text{HS}} = \langle \mathcal{A}, \mathcal{A} \rangle_{\text{HS}}^{1/2}$. The spectral and pseudo spectral density functions in this case are given by

$$f(\theta) = \frac{\delta}{2\pi} \sum_{k=-\infty}^{\infty} e^{ik^\top \theta \delta} C(k), \quad \check{f}(\theta) = \frac{\delta}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ik^\top \theta \delta} \check{C}(k), \quad \theta \in [-\pi, \pi].$$

Also,

$$\mathcal{T}_n(\theta) := \sqrt{\frac{n}{\Delta_n}} \left(\hat{f}_n(\theta) - \mathbb{E} \hat{f}_n(\theta) \right),$$

where

$$\hat{f}_n(\theta) = \frac{1}{2\pi n} \sum_{i,j=1}^n e^{i(i-j)\theta} X(i) \otimes X(j) K \left(\frac{i-j}{\Delta_n} \right), \quad \theta \in [-\pi, \pi].$$

For this special setting, we will establish the following theorem:

Theorem VI.5. Let $X = \{X(t), t \in \mathbb{Z}\}$ be a stationary Gaussian process of the complex Hilbert space \mathbb{H} . Let $\Delta_n \rightarrow \infty$, $\Delta_n/n \rightarrow 0$, and assume that (6.8) and Assumption K' hold. Define

$$\mathcal{T}_n(\theta) := \sqrt{\frac{n}{\Delta_n}} \left[\hat{f}_n(\theta) - \mathbb{E} \hat{f}_n(\theta) \right], \quad \theta \in \mathbb{R},$$

where $\hat{f}_n(\theta)$ is given in (6.1). Then, for any $\theta \in [-\pi, \pi]$,

$$\mathcal{T}_n(\theta) \xrightarrow{d} \mathcal{T}(\theta) \quad \text{in } \mathbb{X},$$

where $\mathcal{T}(\theta)$ is a zero-mean Gaussian element of \mathbb{X} , such that for every finite collection $\{g_\ell, \ell = 1, \dots, m\}$, and positive numbers $\{a_\ell, \ell = 1, \dots, m\}$,

$$\begin{aligned} & \text{Var} \left(\sum_{\ell=1}^m a_\ell \langle \mathcal{T}(\theta) g_\ell, g_\ell \rangle \right) \\ &= \|K\|_2^2 \sum_{\ell_1, \ell_2=1}^m a_{\ell_1} a_{\ell_2} \left[|\langle f(\theta) g_{\ell_2}, g_{\ell_1} \rangle|^2 + I_{(\theta=0, \pm\pi)} |\langle \check{f}(\theta) \overline{g_{\ell_2}}, g_{\ell_1} \rangle|^2 \right]. \end{aligned}$$

The following proposition describes the roadmap for proving this result. For simplicity of notation, we will henceforth suppress the argument θ in $\mathcal{T}_n(\theta)$ since it is fixed.

Proposition VI.6. Let the assumptions of Theorem VI.5 hold. Also, let $\{e_i, i \geq 1\}$ be a CONS of \mathbb{H} . Assume that

(i) for any $\epsilon, \delta > 0$, there exists $u \in \mathbb{Z}_+$ such that

$$\sup_{n \geq 1} \mathbb{P}(\|(I - \Pi_u) \mathcal{T}_n\|_{\text{HS}} > \epsilon) < \delta,$$

where $\Pi_u : \mathbb{X} \rightarrow \mathbb{X}$ is the orthogonal projection operator on $\mathbb{X}_u := \text{span}(e_i \otimes e_j, i, j \leq u)$, and

(ii) for all $a_\ell \in \mathbb{R}$ and $g_\ell \in \mathbb{H}$, we have

$$\mathbb{E} \left[\sum_{\ell=1}^m a_\ell \langle \mathcal{T}_n g_\ell, g_\ell \rangle \right]^k \sim \begin{cases} \mathcal{O} \left(\left(\frac{\Delta_n}{n} \right)^{\frac{1}{2}} \right) & , k \text{ odd,} \\ (k-1)!! [\sigma_{a,g}^2]^{\frac{k}{2}} & , k \text{ even,} \end{cases} \quad (6.9)$$

where

$$\sigma_{a,g}^2 := \|K\|_2^2 \sum_{\ell_1, \ell_2=1}^m a_{\ell_1} a_{\ell_2} \left(|\langle f(\theta) g_{\ell_1}, g_{\ell_2} \rangle|^2 + 1_{\{0, \pm\pi\}}(\theta) |\langle \check{f}(\theta) \overline{g_{\ell_1}}, g_{\ell_2} \rangle|^2 \right).$$

Then there exists a Gaussian process \mathcal{T} in \mathbb{X} that fulfills the description of Theorem VI.5.

Proof. First, we state a useful identity for a complex Hilbert space. Write

$$\langle \mathcal{T}_n, e_i \otimes e_j \rangle_{\text{HS}} = \langle \mathcal{T}_n e_j, e_i \rangle_{\mathbb{H}} =: \mathcal{T}_n(e_j, e_i),$$

By Lemma A.8 of *Shen et al.* (2022),

$$\begin{aligned} \mathcal{T}_n(e_j, e_i) &= \frac{\mathfrak{i} - 1}{2} (\mathcal{T}_n(e_j, e_j) + \mathcal{T}_n(e_i, e_i)) + \frac{1}{2} \mathcal{T}_n(e_j + e_i, e_j + e_i) \\ &\quad - \frac{\mathfrak{i}}{2} \mathcal{T}_n(\mathfrak{i}e_j + e_i, \mathfrak{i}e_j + e_i). \end{aligned} \quad (6.10)$$

Also, recall that $\{e_i \otimes e_j, i, j \geq 1\}$ is a CONS of \mathbb{X} . Thus, (i) implies the flat concentration condition of Condition 1 of Theorem 7.7.4 of *Hsing and Eubank* (2015). It follows from (ii), applying (6.10) plus Markov's inequality, that Condition 2 of Theorem 7.7.4 of *Hsing and Eubank* (2015) also holds. Thus, $\{\mathcal{T}_n, n \geq 1\}$ is tight and hence relatively compact. To show that \mathcal{T}_n converges in distribution to some \mathcal{T} , it suffices to show that if $\mathcal{T}_{n'} \xrightarrow{d} \mathcal{T}$ along some subsequence $\{n'\}$, then \mathcal{T} does not depend on the subsequence. Now, by the continuous mapping theorem, (6.9) and a

standard uniform integrability argument, we have

$$\mathbb{E} \left[\sum_{\ell=1}^m a_\ell \langle \mathcal{T}_n g_\ell, g_\ell \rangle \right]^k \rightarrow \mathbb{E} \left[\sum_{\ell=1}^m a_\ell \langle \mathcal{T} g_\ell, g_\ell \rangle \right]^k \quad \text{for all } k,$$

where the limiting moments entail that $\sum_{\ell=1}^m a_\ell \langle \mathcal{T} g_\ell, g_\ell \rangle$ is distributed as $N(0, \sigma_{a,g}^2)$ (cf. Theorem 30.1 of *Billingsley, 2012*). Relation (6.10) shows that the finite-dimensional distributions of the real Gaussian process $\{\langle \mathcal{T} g_\ell, g_\ell \rangle, g_\ell \in \mathbb{H}\}$ determine the finite-dimensional distributions of $\{\langle \mathcal{T} g, h \rangle, g, h \in \mathbb{H}\}$, which in turn characterize the law of the \mathbb{X} -valued random element \mathcal{T} . The result thus follows. \square

We will complete the proof of Theorem VI.5 by verifying the conditions (i) and (ii) of Proposition VI.6, which will be established in Sections 6.3 and 6.4, respectively.

6.3 Flat concentration

In this section, we establish that (i) of Proposition VI.6 holds for $\{\mathcal{T}_n\}$ under the assumptions of the Proposition. This property is known as the flat concentration of $\{\mathcal{T}_n\}$.

By Markov's inequality,

$$\mathbb{P}(\|(I - \Pi_u)\mathcal{T}_n\|_{\text{HS}} > \epsilon) \leq \epsilon^{-2} \mathbb{E}\|(I - \Pi_u)\mathcal{T}_n\|_{\text{HS}}^2.$$

Since

$$\mathbb{E}\|(I - \Pi_u)\mathcal{T}_n\|_{\text{HS}}^2 = \sum_{k \vee \ell > u} \mathbb{E}|\langle \mathcal{T}_n, e_k \otimes e_\ell \rangle_{\text{HS}}|^2, \quad (6.11)$$

it is sufficient to show that

$$\sum_{k, \ell} \sup_n \mathbb{E}|\langle \mathcal{T}_n, e_k \otimes e_\ell \rangle_{\text{HS}}|^2 < \infty, \quad (6.12)$$

which implies (i) of Proposition VI.6 by (6.11).

Without loss of generality suppose that the CONS $\{e_j, j \in \mathbb{N}\}$ of \mathbb{H} is *real* and thus $X(t) = \sum_{j \in \mathbb{N}} X_j(t)e_j$, where $X_j(t) = \langle X(t), e_j \rangle_{\mathbb{H}}$ are complex zero-mean Gaussian random variables. Also, let

$$C_{k,\ell}(x) = \langle C(x), e_k \otimes e_\ell \rangle_{\text{HS}} = \langle C(x)e_\ell, e_k \rangle. \quad (6.13)$$

It follows that

$$\begin{aligned} & \langle \mathcal{T}_n, e_k \otimes e_\ell \rangle_{\text{HS}} \\ &= \frac{(2\pi)^{-1}}{\sqrt{n\Delta_n}} \sum_{i,j=1}^n \langle X(i) \otimes X(j) - C(i-j), e_k \otimes e_\ell \rangle_{\text{HS}} e^{i\theta(i-j)} K\left(\frac{i-j}{\Delta_n}\right) \\ &= \frac{(2\pi)^{-1}}{\sqrt{n\Delta_n}} \sum_{i,j=1}^n \left[X_k(i)\overline{X_\ell(j)} - C_{k,\ell}(i-j) \right] e^{i\theta(i-j)} K\left(\frac{i-j}{\Delta_n}\right), \end{aligned}$$

where $C_{k,\ell}(i-j) = \mathbb{E}[X_k(i)\overline{X_\ell(j)}]$. By Lemma VI.12, we have

$$\begin{aligned} \mathbb{E}[X_k(i_1)\overline{X_\ell(j_1)}X_k(i_2)\overline{X_\ell(j_2)}] &= C_{k,\ell}(i_1-j_1)\overline{C_{k,\ell}(i_2-j_2)} + \check{C}_{k,\ell}(i_1-j_2)\overline{\check{C}_{k,\ell}(i_2-j_1)} \\ &\quad + C_{k,k}(i_1-i_2)\overline{C_{\ell,\ell}(j_1-j_2)}. \end{aligned}$$

Thus,

$$\begin{aligned} & \mathbb{E} |\langle \mathcal{T}_n, e_k \otimes e_\ell \rangle_{\text{HS}}|^2 \\ &= \frac{(2\pi)^{-2}}{n\Delta_n} \sum_{i_1,j_1} \sum_{i_2,j_2} e^{i\theta(i_1-j_1-i_2+j_2)} K\left(\frac{i_1-j_1}{\Delta_n}\right) K\left(\frac{i_2-j_2}{\Delta_n}\right) \\ &\quad \times \mathbb{E} \left\{ \left[X_k(i_1)\overline{X_\ell(j_1)} - C_{k,\ell}(i_1-j_1) \right] \left[X_k(i_2)\overline{X_\ell(j_2)} - \overline{C_{k,\ell}(i_2-j_2)} \right] \right\} \\ &= \frac{(2\pi)^{-2}}{n\Delta_n} \sum_{i_1,j_1} \sum_{i_2,j_2} e^{i\theta(i_1-j_1-i_2+j_2)} K\left(\frac{i_1-j_1}{\Delta_n}\right) K\left(\frac{i_2-j_2}{\Delta_n}\right) \\ &\quad \times C_{k,k}(i_1-i_2)\overline{C_{\ell,\ell}(j_1-j_2)} \end{aligned} \quad (6.14)$$

$$\begin{aligned}
& + \frac{(2\pi)^{-2}}{n\Delta_n} \sum_{i_1, j_1} \sum_{i_2, j_2} e^{i\theta(i_1 - j_1 - i_2 + j_2)} K\left(\frac{i_1 - j_1}{\Delta_n}\right) K\left(\frac{i_2 - j_2}{\Delta_n}\right) \\
& \quad \times \check{C}_{k,\ell}(i_1 - j_2) \overline{\check{C}_{k,\ell}(i_2 - j_1)} \\
& =: A_{k,\ell} + B_{k,\ell}.
\end{aligned}$$

We start with $A_{k,\ell}$. With the change of variables

$$\begin{aligned}
x_1 &= i_1 - i_2, & x_2 &= j_1 - j_2, \\
y_1 &= i_1 - j_1, & y_2 &= i_1,
\end{aligned}$$

we obtain

$$\begin{aligned}
A_{k,\ell} &= \frac{(2\pi)^{-2}}{n\Delta_n} \sum_{x_1, x_2=1-n}^{n-1} C_{k,k}(x_1) \overline{C_{\ell,\ell}(x_2)} e^{i\theta(x_1 - x_2)} \\
& \quad \times \sum_{y_1=(-\Delta_n) \vee (1-n+x_1-x_2)}^{\Delta_n \wedge (n-1+x_1-x_2)} K\left(\frac{y_1}{\Delta_n}\right) K\left(\frac{-x_1 + x_2 + y_1}{\Delta_n}\right) \sum_{y_2=1 \vee (1+y_1)}^{n \wedge (n+y_1)} 1 \\
&= \frac{(2\pi)^{-2}}{n\Delta_n} \sum_{x_1, x_2=1-n}^{n-1} C_{k,k}(x_1) \overline{C_{\ell,\ell}(x_2)} e^{i\theta(x_1 - x_2)} \\
& \quad \times \sum_{y_1=(-\Delta_n) \vee (1-n+x_1-x_2)}^{\Delta_n \wedge (n-1+x_1-x_2)} K\left(\frac{y_1}{\Delta_n}\right) K\left(\frac{-x_1 + x_2 + y_1}{\Delta_n}\right) (n - |y_1|).
\end{aligned}$$

Thus, with $\|K\|_\infty := \max_t |K(t)|$, we obtain

$$\begin{aligned}
|A_{k,\ell}| &\leq \frac{\|K\|_\infty^2}{2\pi^2} \sum_{x_1, x_2=1-n}^{n-1} |C_{k,k}(x_1)| |C_{\ell,\ell}(x_2)| \\
&\leq \frac{\|K\|_\infty^2}{2\pi^2} \sum_{x_1, x_2=-\infty}^{\infty} |C_{k,k}(x_1)| |C_{\ell,\ell}(x_2)| =: \alpha_{k,\ell}.
\end{aligned}$$

By (6.13) and (ii) of Lemma IV.37, we have

$$\sum_{k,\ell} \alpha_{k,\ell} \leq \frac{\|K\|_\infty^2}{2\pi^2} \left(\sum_{x=-\infty}^{\infty} \|C(x)\|_{\text{tr}} \right)^2 < \infty. \quad (6.15)$$

We now turn to $B_{k,\ell}$ in (6.14). With the change of variables

$$\begin{aligned}x_1 &= i_1 - j_2, & x_2 &= i_2 - j_1, \\ y_1 &= i_1 - j_1, & y_2 &= i_1,\end{aligned}$$

$$\begin{aligned}B_{k,\ell} &= \frac{(2\pi)^{-2}}{n\Delta_n} \sum_{x_1, x_2=1-n}^{n-1} \check{C}_{k,\ell}(x_1) \overline{\check{C}_{k,\ell}(x_2)} e^{i\theta(-x_1-x_2+2y_1)} \\ &\quad \times \sum_{y_1=(-\Delta_n) \vee (1-n+x_1+x_2)}^{\Delta_n \wedge (n-1+x_1+x_2)} K\left(\frac{y_1}{\Delta_n}\right) K\left(\frac{x_1+x_2-y_1}{\Delta_n}\right) \sum_{y_2=1 \vee (1+y_1)}^{n \wedge (n+y_1)} 1 \\ &= \frac{(2\pi)^{-2}}{n\Delta_n} \sum_{x_1, x_2=1-n}^{n-1} \check{C}_{k,\ell}(x_1) \overline{\check{C}_{k,\ell}(x_2)} e^{i\theta(-x_1-x_2+2y_1)} \\ &\quad \times \sum_{y_1=(-\Delta_n) \vee (1-n+x_1+x_2)}^{\Delta_n \wedge (n-1+x_1+x_2)} K\left(\frac{y_1}{\Delta_n}\right) K\left(\frac{x_1+x_2-y_1}{\Delta_n}\right) (n - |y_1|).\end{aligned}$$

Thus,

$$\begin{aligned}|B_{k,\ell}| &\leq \frac{\|K\|_\infty^2}{2\pi^2} \sum_{x_1, x_2=1-n}^{n-1} |\check{C}_{k,\ell}(x_1)| |\check{C}_{k,\ell}(x_2)| \\ &\leq \frac{\|K\|_\infty^2}{2\pi^2} \sum_{x_1, x_2=1-n}^{n-1} |\check{C}_{k,\ell}(x_1)| |\check{C}_{k,\ell}(x_2)| =: \beta_{k,\ell}.\end{aligned}$$

Applying the Cauchy-Schwarz inequality and in view of the definition of the Hilbert-Schmidt inner product,

$$\begin{aligned}\sum_{k,\ell} \beta_{k,\ell} &\leq \frac{\|K\|_\infty^2}{2\pi^2} \sum_{x_1, x_2=-\infty}^{\infty} \sum_{k,\ell} |\check{C}_{k,\ell}(x_1)| |\check{C}_{k,\ell}(x_2)| \\ &\leq \frac{\|K\|_\infty^2}{2\pi^2} \sum_{x_1, x_2=-\infty}^{\infty} \sqrt{\sum_{k,\ell} |\check{C}_{k,\ell}(x_1)|^2} \sqrt{\sum_{k,\ell} |\check{C}_{k,\ell}(x_2)|^2} \\ &= \frac{\|K\|_\infty^2}{2\pi^2} \left(\sum_{x=-\infty}^{\infty} \|\check{C}(x)\|_{\text{HS}} \right)^2 \leq \frac{\|K\|_\infty^2}{2\pi^2} \left(\sum_{x=-\infty}^{\infty} \|\check{C}(x)\|_{\text{tr}} \right)^2 < \infty.\end{aligned}\tag{6.16}$$

Since the upper bounds $\alpha_{k,\ell}$ and $\beta_{k,\ell}$ do not depend on n , we have

$$\sup_n \mathbb{E} |\langle \mathcal{T}_n, e_k \otimes e_\ell \rangle_{\text{HS}}|^2 \leq \alpha_{k,\ell} + \beta_{k,\ell},$$

and Relations (6.15) and (6.16) imply (6.12).

6.4 Convergence of moments

The proof of (ii) of Proposition VI.6 is quite lengthy, and constitutes the core of the central limit theorem proof. In Section 6.4.1, we will first focus on showing (ii) for the case that X is scalar, i.e., X take values in \mathbb{C} . There we will take advantage of this simple setting to explain the ideas of the proof. In Section 6.4.2, we will prove the proposition for the general case of $X \in \mathbb{H}$.

6.4.1 The scalar case

In this section, we focus on a zero-mean, stationary Gaussian time-series taking values in \mathbb{C} and compute the moments of \mathcal{T}_n . The purpose of this section is to develop technical tools for the general moment calculations needed to prove (ii).

In this setting, $C(t-s) = \mathbb{E}[X(t)\overline{X(s)}]$, $\check{C}(t-s) := \mathbb{E}[X(t)X(s)]$ and

$$f(\theta) = \frac{1}{2\pi} \sum_{x=-\infty}^{\infty} C(x)e^{ix\theta}. \quad (6.17)$$

Also, (6.8) becomes

$$\sum_{x=-\infty}^{\infty} \left[|C(x)| + |\check{C}(x)| \right] < \infty. \quad (6.18)$$

Observe that since $C(x) = \overline{C(-x)}$ in this case, we have that $f(\theta)$ is real, even though the process $\{X_t, t \in \mathbb{Z}\}$ is complex-valued. We shall also need the so-called pseudo-

spectral density, defined as:

$$\check{f}(\theta) = \frac{1}{2\pi} \sum_{x=-\infty}^{\infty} \check{C}(x) e^{-ix\theta}. \quad (6.19)$$

Recall also that the spectral density estimator is

$$\hat{f}_n(\theta) = \frac{1}{2\pi n} \sum_{i,j=1}^n X_i \overline{X_j} K_\theta(i-j),$$

where

$$K_\theta(y) = e^{iy\theta} K(y/\Delta_n).$$

The following proposition gives the asymptotic expression of $\mathbb{E}(\mathcal{T}_n^k)$ for the scalar case.

Proposition VI.7. *Assume that the conditions of Proposition VI.6 hold for the setting $\mathbb{H} = \mathbb{C}$. Then, as $n \rightarrow \infty$,*

$$\mathbb{E}(\mathcal{T}_n^k) = \begin{cases} \mathcal{O}\left(\left(\frac{\Delta_n}{n}\right)^{\frac{1}{2}}\right) & , k \text{ odd,} \\ (1 + o(1))(k-1)!! \left[(f(\theta)^2 + 1_{\{0, \pm\pi\}}(\theta) |\check{f}(\theta)|^2) \|K\|_2^2 \right]^{\frac{k}{2}} & , k \text{ even,} \end{cases}$$

where $f(\theta)$ and $\check{f}(\theta)$ are as in (6.17) and (6.19).

Proof. Assume without loss of generality that the support of K is $[-1, 1]$.

The case $k = 2$

By Lemma VI.13 (for $N = 0$, $M = 2$, i.e., see (6.27)),

$$\begin{aligned} (2\pi)^2 \mathbb{E}(\mathcal{T}_n^2) &= \frac{1}{n\Delta_n} \sum_{i_1, j_1=1}^n \sum_{i_2, j_2=1}^n K_\theta(i_1 - j_1) K_\theta(i_2 - j_2) \\ &\quad \times \mathbb{E} \left[(X_{i_1} \overline{X_{j_1}} - C(i_1 - j_1)) (X_{i_2} \overline{X_{j_2}} - C(i_2 - j_2)) \right] \\ &= \frac{1}{n\Delta_n} \sum_{i_1, j_1=1}^n \sum_{i_2, j_2=1}^n K_\theta(i_1 - j_1) K_\theta(i_2 - j_2) \end{aligned}$$

$$\begin{aligned}
& \times \left[C(i_1 - j_2)C(i_2 - j_1) + \check{C}(i_1 - i_2)\overline{\check{C}(j_1 - j_2)} \right] \\
& =: \tilde{A}_2 + \bar{A}_2,
\end{aligned}$$

where

$$\tilde{A}_2 := \frac{1}{n\Delta_n} \sum_{i_1, j_1=1}^n \sum_{i_2, j_2=1}^n K_\theta(i_1 - j_1) K_\theta(i_2 - j_2) \cdot C(i_1 - j_2)C(i_2 - j_1),$$

and

$$\bar{A}_2 := \frac{1}{n\Delta_n} \sum_{i_1, j_1=1}^n \sum_{i_2, j_2=1}^n K_\theta(i_1 - j_1) K_\theta(i_2 - j_2) \cdot \check{C}(i_1 - i_2)\overline{\check{C}(j_1 - j_2)}.$$

The two terms have somewhat different properties and we start with \tilde{A}_2 . With the change of variables

$$\begin{aligned}
x_1 &= i_1 - j_2, & x_2 &= i_2 - j_1, \\
y_1 &= i_1 - j_1, & y_2 &= i_1,
\end{aligned}$$

we have

$$\begin{aligned}
\tilde{A}_2 &= \frac{1}{n\Delta_n} \sum_{x_1, x_2=-n+1}^{n-1} C(x_1)C(x_2)e^{ix_1\theta}e^{ix_2\theta} \\
& \quad \times \sum_{y_1=(1-n)\vee(1-n+x_1+x_2)}^{(n-1)\wedge(n-1+x_1+x_2)} K\left(\frac{y_1}{\Delta_n}\right) K\left(\frac{x_1+x_2-y_1}{\Delta_n}\right) \sum_{y_2=1\vee(1+y_1)}^{n\wedge(n+y_1)} 1 \\
&= \frac{1}{n\Delta_n} \sum_{x_1, x_2=-n+1}^{n-1} C(x_1)C(x_2)e^{ix_1\theta}e^{ix_2\theta} \\
& \quad \times \sum_{y_1=(1-n)\vee(1-n+x_1+x_2)}^{(n-1)\wedge(n-1+x_1+x_2)} K\left(\frac{y_1}{\Delta_n}\right) K\left(\frac{x_1+x_2-y_1}{\Delta_n}\right) \cdot (n - |y_1|) \\
&= \frac{1}{\Delta_n} \sum_{|x_1|\vee|x_2|\leq L} C(x_1)C(x_2)e^{ix_1\theta}e^{ix_2\theta}
\end{aligned}$$

$$\begin{aligned}
& \times \sum_{y_1=(1-n) \vee (1-n+x_1+x_2)}^{(n-1) \wedge (n-1+x_1+x_2)} K\left(\frac{y_1}{\Delta_n}\right) K\left(\frac{x_1+x_2-y_1}{\Delta_n}\right) \frac{n-|y_1|}{n} \\
& + \frac{1}{\Delta_n} \sum_{|x_1| \vee |x_2| \geq L} C(x_1)C(x_2)e^{ix_1\theta}e^{ix_2\theta} \\
& \times \sum_{y_1=(1-n) \vee (1-n+x_1+x_2)}^{(n-1) \wedge (n-1+x_1+x_2)} K\left(\frac{y_1}{\Delta_n}\right) K\left(\frac{x_1+x_2-y_1}{\Delta_n}\right) \frac{n-|y_1|}{n} \\
& =: B_1 + B_2,
\end{aligned}$$

for some $L = L_n \rightarrow \infty$ and $L = o(\Delta_n)$. One can easily see that

$$\begin{aligned}
|B_2| & \leq \frac{\|K\|_\infty}{\Delta_n} \sum_{|x_1| \vee |x_2| \geq L} |C(x_1)||C(x_2)| \sum_{y_1=-n+1}^{n-1} K\left(\frac{y_1}{\Delta_n}\right) \\
& \leq 2\|K\|_\infty \sum_{|x_1| \vee |x_2| \geq L} |C(x_1)||C(x_2)| = o(1), \text{ as } L \rightarrow \infty,
\end{aligned}$$

by (6.18). Now, adding and subtracting the same term in B_1 , one obtains that $B_1 = C_1 + C_2$, where

$$C_1 := \frac{1}{\Delta_n} \sum_{|x_1| \vee |x_2| \leq L} C(x_1)C(x_2)e^{ix_1\theta}e^{ix_2\theta} \sum_{y_1=(1-n) \vee (1-n+x_1+x_2)}^{(n-1) \wedge (n-1+x_1+x_2)} K^2\left(\frac{y_1}{\Delta_n}\right) \frac{n-|y_1|}{n}$$

and

$$\begin{aligned}
C_2 & := \frac{1}{\Delta_n} \sum_{|x_1| \vee |x_2| \leq L} C(x_1)C(x_2)e^{ix_1\theta}e^{ix_2\theta} \\
& \times \sum_{y_1=(1-n) \vee (1-n+x_1+x_2)}^{(n-1) \wedge (n-1+x_1+x_2)} K\left(\frac{y_1}{\Delta_n}\right) \left[K\left(\frac{x_1+x_2-y_1}{\Delta_n}\right) - K\left(\frac{y_1}{\Delta_n}\right) \right] \times \frac{n-|y_1|}{n}.
\end{aligned}$$

We examine C_1 first.

Observe first that the inner sum over y_1 is confined to $-\Delta_n \leq y_1 \leq \Delta_n$, since K is supported on $[-1, 1]$. Moreover, since $L = o(\Delta_n)$ and $\Delta_n = o(n)$, for all $|x_1| \vee |x_2| \leq L$, and all sufficiently large n , we have that $(1-n) \vee (1-n+x_1+x_2) \leq -\Delta_n$

and $\Delta_n \leq (n-1) \wedge (n-1+x_1+x_2)$. This means, that the inner summation in the definitions of C_1 and C_2 is over the range $[-\Delta_n, \Delta_n]$ and it does not depend on x_1 and x_2 . That is, for all sufficiently large n ,

$$\begin{aligned} C_1 &= \sum_{|x_1| \vee |x_2| \leq L} C(x_1)C(x_2)e^{ix_1\theta}e^{ix_2\theta} \times \frac{1}{\Delta_n} \sum_{y_1=-\Delta_n}^{\Delta_n} K^2\left(\frac{y_1}{\Delta_n}\right) \\ &\sim 4\pi^2 f(\theta)^2 \int_{-1}^1 K^2(y)dy, \quad \text{as } n \rightarrow \infty, \end{aligned}$$

where the last relation follows from the Riemann integrability of K^2 and the fact that $\sum_{|x| \leq L} C(x)e^{ix\theta} \rightarrow 2\pi f(\theta)$, as $L \rightarrow \infty$. Now, focus on the term C_2 . Using the facts that K is an even function and

$$|K(x) - K(y)| \leq c|x - y|,$$

(since K' is bounded), we get $|K((x_1 + x_2 - y_1)/\Delta_n) - K(y_1/\Delta_n)| \leq 2cL/\Delta$, for all $|x_1| \vee |x_2| \leq L$. Thus, by Condition (6.18) and the Riemann integrability of K , we obtain

$$\begin{aligned} |C_2| &\leq \left(\sum_{|x| \leq L} |C(x)| \right)^2 \frac{1}{\Delta_n} \sum_{y=-\Delta_n}^{\Delta_n} K\left(\frac{y}{\Delta_n}\right) \frac{2Lc}{\Delta_n} \\ &\sim \frac{2Lc}{\Delta_n} \left(\sum_{x=-\infty}^{\infty} |C(x)| \right)^2 \int_{u=-1}^1 K(u)du = o(1), \end{aligned}$$

since $L = o(\Delta_n)$. Summarizing, we have that for all θ (including $\theta = 0$ and $\theta = \pm\pi$)

$$\tilde{A}_2 = C_1 + C_2 + B_2 \sim 4\pi^2 f(\theta)^2 \int_{-1}^1 K^2(u)du.$$

We next consider \bar{A}_2 . Similar to the derivation for \tilde{A}_2 , with the change of variables

$$x_1 = i_1 - i_2, \quad x_2 = j_1 - j_2,$$

$$y_1 = i_1 - j_1, \quad y_2 = i_1,$$

we get

$$\begin{aligned} \bar{A}_2 &= \frac{1}{n\Delta_n} \sum_{x_1, x_2 = -n+1}^{n-1} \check{C}(x_1) \overline{\check{C}(x_2)} e^{-ix_1\theta} e^{ix_2\theta} \\ &\quad \cdot \sum_{y_1 = (1-n) \vee (1-n+x_1-x_2)}^{(n-1) \wedge (n-1+x_1-x_2)} K\left(\frac{y_1}{\Delta_n}\right) K\left(\frac{-x_1+x_2+y_1}{\Delta_n}\right) e^{i2y_1\theta} \sum_{y_2 = 1 \vee (1+y_1)}^{n \wedge (n+y_1)} 1 \\ &= \frac{1}{\Delta_n} \sum_{x_1, x_2 = -n+1}^{n-1} \check{C}(x_1) \overline{\check{C}(x_2)} e^{-ix_1\theta} e^{ix_2\theta} \\ &\quad \cdot \sum_{y_1 = (1-n) \vee (1-n+x_1-x_2)}^{(n-1) \wedge (n-1+x_1-x_2)} K\left(\frac{y_1}{\Delta_n}\right) K\left(\frac{-x_1+x_2+y_1}{\Delta_n}\right) e^{i2y_1\theta} \cdot \frac{n-|y_1|}{n}. \end{aligned}$$

Observe first that for $\theta = \pm\pi$ or $\theta = 0$, we have $e^{i2y_1\theta} = 1$, $y_1 \in \mathbb{Z}$ and for the term \bar{A}_2 with the same argument as for the term \tilde{A}_2 , we obtain

$$\bar{A}_2 \sim 4\pi |\check{f}(\theta)|^2 \|K\|_2^2, \quad \text{where } \check{f}(\theta) = \frac{1}{2\pi} \sum_{x=-\infty}^{\infty} \check{C}(x) e^{-ix\theta}, \quad \theta \in \{0, \pm\pi\}.$$

Suppose now $\theta \neq 0$ and $\theta \neq \pm\pi$, so that the term $e^{i2y_1\theta}$ is present. By adding and subtracting a term, we have that $\bar{A}_2 = D_1 + D_2$, where D_2 is defined in (6.21) below and

$$\begin{aligned} D_1 &:= \frac{1}{\Delta_n} \sum_{x_1, x_2 = -n+1}^{n-1} \check{C}(x_1) \overline{\check{C}(x_2)} e^{-ix_1\theta} e^{ix_2\theta} \\ &\quad \cdot \sum_{y_1 = (1-n) \vee (1-n+x_1-x_2)}^{(n-1) \wedge (n-1+x_1-x_2)} K^2\left(\frac{y_1}{\Delta_n}\right) e^{i2y_1\theta} \frac{n-|y_1|}{n} \quad (6.20) \\ &= \mathcal{O}\left(\frac{1}{\Delta_n}\right). \end{aligned}$$

Indeed, write

$$w_n(y) = K^2\left(\frac{y}{\Delta_n}\right) \frac{n-|y|}{n}$$

and consider, for any $c_1, c_2 \in [1, \Delta_n]$,

$$\begin{aligned}
& \sum_{y=c_1}^{c_2} w_n(y) e^{i2y\theta} (e^{i2\theta} - 1) \\
&= \sum_{y=c_1+1}^{c_2+1} w_n(y-1) e^{i2y\theta} - \sum_{y=c_1}^{c_2} w_n(y) e^{i2y\theta} \\
&= w_n(c_2) e^{i2(c_2+1)\theta} - w_n(c_1) e^{i2c_1\theta} + \sum_{y=c_1+1}^{c_2} (w_n(y-1) - w_n(y)) e^{i2y\theta}.
\end{aligned}$$

Focusing on the second term,

$$\begin{aligned}
& \sum_{y=c_1+1}^{c_2} (w_n(y-1) - w_n(y)) e^{i2y\theta} \\
&= \sum_{y=c_1+1}^{c_2} \left(K^2 \left(\frac{y-1}{\Delta_n} \right) \frac{n-1-(y-1)}{n} - K^2 \left(\frac{y}{\Delta_n} \right) \frac{n-1-y}{n} \right) e^{i2y\theta} \\
&= \frac{1}{n} \sum_{y=c_1+1}^{c_2} K^2 \left(\frac{y-1}{\Delta_n} \right) e^{i2y\theta} + \sum_{y=c_1+1}^{c_2} \left(K^2 \left(\frac{y-1}{\Delta_n} \right) - K^2 \left(\frac{y}{\Delta_n} \right) \right) \frac{n-1-y}{n} e^{i2y\theta} \\
&=: E_1 + E_2.
\end{aligned}$$

Clearly, $E_1 = \mathcal{O}(\Delta_n/n) = o(1)$ uniformly in c_1, c_2 . Also, it follows that

$$|E_2| \leq \sum_{y=c_1+1}^{c_2} \left| K^2 \left(\frac{y-1}{\Delta_n} \right) - K^2 \left(\frac{y}{\Delta_n} \right) \right| < \infty \text{ uniformly in } c_1, c_2$$

since K^2 is of bounded variation (recall that K' is bounded and K is compactly supported). Thus,

$$\begin{aligned}
& \sum_{y=c_1}^{c_2} w_n(y) e^{i2y\theta} \\
&= (e^{i2\theta} - 1)^{-1} \left(w_n(c_2) e^{i(c_2+1)\theta} - w_n(c_1) e^{i2c_1\theta} + \sum_{y=c_1+1}^{c_2} (w_n(y-1) - w_n(y)) e^{iy\theta} \right),
\end{aligned}$$

which is uniformly bounded. (Note that here we used the fact that $e^{i2\theta} - 1 \neq 0$,

since $\pm\pi \neq \theta \neq 0$.) Applying this argument, we see that the inner sum in (6.20) is uniformly bounded, and hence by Condition (6.18), we obtain that $D_1 = o(1)$.

On the other hand, for the term D_2 , we obtain

$$\begin{aligned}
D_2 &:= \frac{1}{\Delta_n} \sum_{x_1, x_2 = -n+1}^{n-1} \check{C}(x_1) \overline{\check{C}(x_2)} e^{-ix_1\theta} e^{ix_2\theta} \\
&\quad \times \sum_{y_1 = (1-n) \vee (1-n+x_1-x_2)}^{(n-1) \wedge (n-1+x_1-x_2)} K\left(\frac{y_1}{\Delta_n}\right) \left[K\left(\frac{-x_1+x_2+y_1}{\Delta_n}\right) - K\left(\frac{y_1}{\Delta_n}\right) \right] \\
&\quad \times e^{i2y_1\theta} \cdot \frac{n - |y_1|}{n} \\
&= o(1),
\end{aligned} \tag{6.21}$$

using the same arguments as for B_2 and C_2 . Thus,

$$\overline{A}_2 \rightarrow 0 \quad \text{for } \theta \neq 0 \quad \text{and } \theta \neq \pm\pi.$$

This completes the derivation for $\mathbb{E}(\mathcal{T}_n^2)$.

The case $k \geq 3$

Fix some integer $k \geq 3$. Let \mathcal{P} be the set of all possible pairings of $\{i_\ell, j_\ell : \ell = 1, \dots, k\}$. Then a pairing $\mathbb{p} \in \mathcal{P}$ iff

$$\mathbb{p} = \left\{ \{I, J\}, \{I, \tilde{I}\}, \{J, \tilde{J}\} : I \neq \tilde{I} \in \{i_\ell, \ell = 1, \dots, k\}, J \neq \tilde{J} \in \{j_\ell, \ell = 1, \dots, k\} \right\},$$

where all symbols $i_\ell, j_\ell, \ell = 1, \dots, k$ can be used only once.

By Lemma VI.13,

$$\begin{aligned}
\mu_k &:= (2\pi)^k \mathbb{E}(\mathcal{T}_n^k) = \frac{1}{(n\Delta_n)^{k/2}} \sum_{\substack{i_\ell, j_\ell = 1 \\ \ell = 1, \dots, k}}^n \mathbb{E} \left[\prod_{\ell=1}^k [X_{i_\ell} \overline{X_{j_\ell}} - C(i_\ell - j_\ell)] \right] \prod_{\ell=1}^k K_\theta(i_\ell - j_\ell) \\
&= \frac{1}{(n\Delta_n)^{k/2}} \sum_{\substack{i_\ell, j_\ell = 1 \\ \ell = 1, \dots, k}}^n \prod_{\ell=1}^k K_\theta(i_\ell - j_\ell)
\end{aligned}$$

$$\times \sum_{\substack{\mathbb{P} \in \mathcal{P}: \\ \cup_{p=1}^k \{i_p, j_p\} \cap \mathbb{P} = \emptyset}} \prod_{\{i, j\} \in \mathbb{P}} \left[C(I - J) \mathbb{1}_{\{i, j\} = \{I, J\}} + \check{C}(I - \tilde{I}) \mathbb{1}_{\{i, j\} = \{I, \tilde{I}\}} \right. \\ \left. + \overline{\check{C}(J - \tilde{J})} \mathbb{1}_{\{i, j\} = \{J, \tilde{J}\}} \right],$$

Let now $r \leq k$ and fix a subset of $2r$ indices $\{i'_1, j'_1, \dots, i'_r, j'_r\} \subset \{i_1, j_1, \dots, i_k, j_k\}$, where $(i'_1, j'_1) = (i_{\sigma_1}, j_{\sigma_1}), \dots, (i'_r, j'_r) = (i_{\sigma_r}, j_{\sigma_r})$, for some $1 \leq \sigma_1 < \dots < \sigma_r \leq k$. A partition of $\{i'_1, j'_1, \dots, i'_r, j'_r\}$ into pairs will be called a sub-pairing of order r . Namely, it is a partition into pairs that involves r couples of i' and/or j' symbols taken only from the set $\{i'_1, j'_1, \dots, i'_r, j'_r\}$.

A (sub)pairing will be called *irreducible*, if does not have further sub-pairings, i.e., it cannot be broken up into a disjoint union of two or more sub-pairings of lower order. Let $C_{\mathcal{P}, r, k}$ denote the set of all *irreducible* sub-pairings of order r .

Looking at a single summand of the second sum in μ_k , one can see that every pairing $\mathbb{P} \in \mathcal{P}$ is the union of multiple irreducible pairings of the form $C_{\mathcal{P}, r, k}$ with $r \geq 2$. We will argue below that among all pairings in \mathcal{P} only the ones involving irreducible components of order $r = 2$ contribute asymptotically, and the remaining pairings are of lower order, as $n \rightarrow \infty$.

Let $\mathbb{P} \in \mathcal{P}$ denote a pairing that shows up in the second sum of $(2\pi)^d \mathbb{E}(\mathcal{T}_n^k)$, and suppose that

$$\mathbb{P} = \mathbb{P}_{r_1} \cup \dots \cup \mathbb{P}_{r_m},$$

where the $\mathbb{P}_{r_i} \in C_{\mathcal{P}, r_i, k}$, $r_i \geq 2$, $i = 1, \dots, m$, are the irreducible sub-pairings of \mathbb{P} .

Then,

$$\begin{aligned}
\mu_k &= (2\pi)^k \mathbb{E}(\mathcal{T}_n^k) = \frac{1}{(n\Delta_n)^{k/2}} \sum_{\substack{i_\ell, j_\ell=1 \\ \ell=1, \dots, k}}^n \left\{ \prod_{\ell=1}^k K_\theta(i_\ell - j_\ell) \right. \\
&\quad \times \sum_{\substack{\mathbb{P} \in \mathcal{P}: \\ \cup_{p=1}^k \{\{i_p, j_p\}\} \cap \mathbb{P} = \emptyset}} \prod_{\{i, j\} \in \mathbb{P}} \left[C(I - J) \mathbb{1}_{\{i, j\} = \{I, J\}} + \check{C}(I - \tilde{I}) \mathbb{1}_{\{i, j\} = \{I, \tilde{I}\}} \right. \\
&\quad \left. \left. + \overline{\check{C}(J - \tilde{J})} \mathbb{1}_{\{i, j\} = \{J, \tilde{J}\}} \right] \right\} \\
&=: \left(\frac{n}{\Delta_n} \right)^{k/2} \sum_{\substack{\mathbb{P} \in \mathcal{P} \\ \mathbb{P} = \mathbb{P}_{r_1} \cup \dots \cup \mathbb{P}_{r_m}, \ m \geq 1}} \prod_{t=1}^m A_{\mathcal{P}, \mathbb{P}_{r_t}, k},
\end{aligned} \tag{6.22}$$

where $A_{\mathcal{P}, \mathbb{P}_{r_t}, k}$ involves a product of the terms restricted to the irreducible sub-pairing \mathbb{P}_{r_t} , and where $r_1 + \dots + r_m = k$, with $r_i \geq 2$, $i = 1, \dots, m$. Namely, assuming that the subset of indices $\{i'_1, j'_1, \dots, i'_r, j'_r\} = \{i_{\sigma_2}, j_{\sigma_1}, \dots, i_{\sigma_r}, j_{\sigma_r}\}$, $r \leq k$, is involved in the irreducible pairing $A_{\mathcal{P}, \mathbb{P}_r, k}$ we have

$$\begin{aligned}
A_{\mathcal{P}, \mathbb{P}_r, k} &= \frac{1}{n^r} \sum_{\substack{i'_\ell, j'_\ell=1 \\ \ell=1, \dots, r}}^n \prod_{\ell=1}^r K_\theta(i'_\ell - j'_\ell) \times \\
&\quad \prod_{\{i, j\} \in \mathbb{P}_r} \left[C(I - J) \mathbb{1}_{\{i, j\} = \{I, J\}} + \check{C}(I - \tilde{I}) \mathbb{1}_{\{i, j\} = \{I, \tilde{I}\}} \right. \\
&\quad \left. + \overline{\check{C}(J - \tilde{J})} \mathbb{1}_{\{i, j\} = \{J, \tilde{J}\}} \right].
\end{aligned}$$

Let $r \geq 3$, $r \leq k$, and apply the change of variables

$$\begin{aligned}
x_\ell &= i - j, \quad i, j \in \mathbb{P}_r \\
y_\ell &= i'_\ell - j'_\ell, \quad \ell = 1, \dots, r-1, \\
y_r &= i'_r,
\end{aligned}$$

where the order of i, j for x_ℓ is determined by the order they appear in the C, \check{C} and

\bar{C} terms. Note that since the kernel K is non-negative and bounded,

$$\left| \prod_{\ell=1}^r K_{\theta}(i'_{\ell} - j'_{\ell}) \right| \leq \|K\|_{\infty} \prod_{\ell=1}^{r-1} K\left(\frac{y_{\ell}}{\Delta_n}\right).$$

Then, letting $D(x) := |C(x)| \vee |\check{C}(x)|$, we obtain

$$\begin{aligned} & |A_{\mathcal{P}, \mathbb{P}_r, k}| \\ & \leq \frac{\|K\|_{\infty}}{n^r} \sum_{\substack{x_{\ell} = -n+1 \\ \ell=1, \dots, r}}^{n-1} \prod_{\ell=1}^r D(x_{\ell}) \cdot \sum_{\substack{y_m = -n+1 \\ m=1, \dots, r-1}}^{n-1} \prod_{m=1}^{r-1} K\left(\frac{y_m}{\Delta_n}\right) \sum_{y_r=1}^n 1 \\ & \leq \frac{\|K\|_{\infty}}{n^{r-1}} \left(\sum_{x \in \mathbb{Z}} D(x) \right)^r \cdot \sum_{\substack{y_m = -n+1 \\ m=1, \dots, r-1}}^{n-1} \prod_{m=1}^{r-1} K\left(\frac{y_m}{\Delta_n}\right) \\ & = \mathcal{O}\left(\left(\frac{\Delta_n}{n}\right)^{r-1}\right). \end{aligned} \tag{6.23}$$

where we used that by Relation (6.18), $\sum_x D(x) < \infty$ and the compactness of the support of K .

Using (6.23), in view of (6.22), one immediately has that

$$\mathbb{E}(\mathcal{T}_n^k) = \mathcal{O}\left(\left(\frac{n}{\Delta_n}\right)^{k/2} \cdot \max_{\substack{m=1, \dots, \lfloor k/2 \rfloor \\ r_1 + \dots + r_m = k, \ r_t \geq 2}} \left(\frac{\Delta_n}{n}\right)^{\sum_{t=1}^m (r_t - 1)}\right) \equiv \mathcal{O}\left(\left(\frac{\Delta_n}{n}\right)^{k/2 - M}\right),$$

where

$$M := \max_{m=1, \dots, \lfloor k/2 \rfloor} \left\{ m : r_1 + \dots + r_m = k, \text{ where } r_t \in \{2, \dots, k\} \right\}.$$

Clearly, if k is odd, then $M = (k-1)/2$, we have $k/2 - M = 1/2$ and by the above bound, we obtain

$$\mathbb{E}[\mathcal{T}_n^k] = \mathcal{O}((\Delta_n/n)^{1/2}),$$

completing the proof of Proposition VI.7 in this case. Note that this moment vanishes

as $n \rightarrow \infty$.

If k is even, then $M = k/2$ and $k/2 - M = 0$. By the above argument, the only pairings that do not vanish asymptotically, as $n \rightarrow \infty$, correspond to $r_1 = \dots = r_{k/2} = 2$. That is, the indices $\{i_1, j_1, \dots, i_k, j_k\}$ are paired into $k/2$ irreducible sub-pairings of order 2 and this case algebraically reduces to the case $k = 2$.

Consider four indices $\{i_1, j_1, i_2, j_2\}$ and let $A_{\mathcal{P},k}^{\{\{i_1,j_2\},\{i_2,j_1\}\}}$ and $A_{\mathcal{P},k}^{\{\{i_1,i_2\},\{j_1,j_2\}\}}$ be the terms of (6.22) corresponding to the subpairings $\{\{i_1, j_2\}, \{i_2, j_1\}\}$ and $\{\{i_1, i_2\}, \{j_1, j_2\}\}$ respectively. Let also

$$A_{\mathcal{P},k}^{\{\{i_1,j_1,i_2,j_2\}\}} := A_{\mathcal{P},k}^{\{\{i_1,j_2\},\{i_2,j_1\}\}} + A_{\mathcal{P},k}^{\{\{i_1,i_2\},\{j_1,j_2\}\}}$$

By the first part of the proof, the sum of these two order-2 irreducible subpairings that correspond to the same indices $\{i_1, j_1, i_2, j_2\}$ contributes the following term to the rate of the expectation:

$$\left(\frac{n}{\Delta_n}\right)^1 A_{\mathcal{P},k}^{\{\{i_1,j_1,i_2,j_2\}\}} \rightarrow \sigma_f^2(\theta) := \left(f(\theta)^2 + 1_{\{0,\pm\pi\}}(\theta) |\check{f}(\theta)|^2\right) \|K\|_2^2,$$

as $n \rightarrow \infty$. Therefore, in view of (6.22),

$$\begin{aligned} \mu_k &= (2\pi)^k \mathbb{E}(\mathcal{T}_n^k) = \left(\frac{n}{\Delta_n}\right)^{k/2} \sum_{\substack{\mathbb{p} \in \mathcal{P} \\ \mathbb{p} = \mathbb{p}_{r_1} \cup \dots \cup \mathbb{p}_{r_m}, \ m \geq 1}} \prod_{t=1}^m A_{\mathcal{P}, \mathbb{p}_{r_t}, k} \\ &\asymp \left(\frac{n}{\Delta_n}\right)^{k/2} \sum_{\substack{\mathbb{p} \in \mathcal{P} \\ \mathbb{p} = \mathbb{p}_{r_1} \cup \dots \cup \mathbb{p}_{r_{k/2}} \\ r_1 = \dots = r_{k/2} = 2}} \prod_{t=1}^m A_{\mathcal{P}, \mathbb{p}_{r_t}, k} \\ &= \left(\frac{n}{\Delta_n}\right)^{k/2} \sum_{\substack{\mathfrak{q} \in \mathcal{Q}_2 \\ \mathfrak{q} = \{\{i'_1, j'_1, i'_2, j'_2\}, \dots, \\ \{i'_{k-1}, j'_{k-1}, i'_k, j'_k\}\}}} \prod_{\substack{\ell, m=1, \dots, k \\ \ell \neq m}} \left(A_{\mathcal{P},k}^{\{\{i'_\ell, j'_m\}, \{i'_m, j'_\ell\}\}} + A_{\mathcal{P},k}^{\{\{i'_\ell, i'_m\}, \{j'_\ell, j'_m\}\}} \right) \\ &= \left(\frac{n}{\Delta_n}\right)^{k/2} \sum_{\substack{\mathfrak{q} \in \mathcal{Q}_2 \\ \mathfrak{q} = \{\{i'_1, j'_1, i'_2, j'_2\}, \dots, \\ \{i'_{k-1}, j'_{k-1}, i'_k, j'_k\}\}}} \prod_{\substack{\ell, m=1, \dots, k \\ \ell \neq m}} A_{\mathcal{P},k}^{\{\{i'_\ell, j'_\ell, i'_m, j'_m\}\}} \end{aligned}$$

$$\rightarrow N_2(k) \times \left[\sigma_f^2(\theta) \right]^{k/2},$$

where $N_2(k)$ denotes the number of ways one can partition the set $\{i_1, j_1, \dots, i_k, j_k\}$ into $k/2$ sets of 4 members including both i and j of the same index, and \mathcal{Q}_2 denotes the collection of all those partitions.

To complete the proof of Proposition VI.7, it remains to argue that $N_2(k) = |\mathcal{Q}_2| = (k-1)!!$. Note that every $q \in \mathcal{Q}_2$ is determined by a partition into sets of 4 indices $\{i_{\ell_1}, j_{\ell_1}, i_{\ell_2}, j_{\ell_2}\}$ from the $2k$ symbols $\{i_1, j_1, \dots, i_k, j_k\}$. Thus, determining the number $N_2(k)$ is equivalent to counting the number of partitions of the set $\{i_1, \dots, i_k\}$ into 2-point subsets $\{i_{\ell_1}, i_{\ell_2}\}$. The number of ways to pick the first pair is $\binom{k}{2}$, the second pair $\binom{k-2}{2}$, and so on. Therefore $N_2(k)$ equals

$$\frac{1}{(k/2)!} \binom{k}{2} \cdot \binom{k-2}{2} \cdots \binom{2}{2} = (k-1)!!,$$

where we divide by $(k/2)!$ since the order of the subsets $\{i_{\ell_1}, i_{\ell_2}\}$ does not matter. \square

6.4.2 The general case

The purpose of this section is to finish the verification of (ii) of Proposition VI.6 for a general \mathbb{H} under the assumptions of the Proposition. Recall that we already verified (ii) for the spatial setting $\mathbb{H} = \mathbb{C}$ in the previous subsection. The extension from the scalar to the general case is actually quite straightforward. We illustrate this for the second moment.

Recall that $X_{g_\ell}(i) = \langle X(i), g_\ell \rangle$. Denote $(2\pi)^2 \mathbb{E} \left[\sum_{\ell=1}^m a_\ell \langle \mathcal{T}_n g_\ell, g_\ell \rangle \right]^2$ by $A_{n,2}$. By Isserlis' formula in Lemma VI.12,

$$\begin{aligned} A_{n,2} &= \sum_{\ell_1, \ell_2=1}^m a_{\ell_1} a_{\ell_2} \frac{1}{n \Delta_n} \sum_{i_1, j_1=1}^n \sum_{i_2, j_2=1}^n K_\theta(i_1 - j_1) K_\theta(i_2 - j_2) \\ &\quad \times \mathbb{E} \left\{ \left[X_{g_{\ell_1}}(i_1) \overline{X_{g_{\ell_1}}(j_1)} - \mathbb{E} X_{g_{\ell_1}}(i_1) \overline{\mathbb{E} X_{g_{\ell_1}}(j_1)} \right] \right\} \end{aligned}$$

$$\begin{aligned}
& \times \left[X_{g_{\ell_2}}(i_2) \overline{X_{g_{\ell_2}}(j_2)} - \mathbb{E} X_{g_{\ell_2}}(i_2) \overline{X_{g_{\ell_2}}(j_2)} \right] \Big\} \\
= & \sum_{\ell_1, \ell_2=1}^m a_{\ell_1} a_{\ell_2} \frac{1}{n \Delta_n} \sum_{i_1, j_1=1}^n \sum_{i_2, j_2=1}^n K_\theta(i_1 - j_1) K_\theta(i_2 - j_2) \\
& \times \left[\mathbb{E} X_{g_{\ell_1}}(i_1) \overline{X_{g_{\ell_1}}(j_1)} X_{g_{\ell_2}}(i_2) \overline{X_{g_{\ell_2}}(j_2)} \right. \\
& \quad \left. - \mathbb{E} X_{g_{\ell_1}}(i_1) \overline{X_{g_{\ell_1}}(j_1)} \mathbb{E} X_{g_{\ell_2}}(i_2) \overline{X_{g_{\ell_2}}(j_2)} \right] \\
= & \sum_{\ell_1, \ell_2=1}^m a_{\ell_1} a_{\ell_2} \frac{1}{n \Delta_n} \sum_{i_1, j_1=1}^n \sum_{i_2, j_2=1}^n K_\theta(i_1 - j_1) K_\theta(i_2 - j_2) \\
& \times \mathbb{E} X_{g_{\ell_1}}(i_1) \overline{X_{g_{\ell_2}}(j_2)} \mathbb{E} X_{g_{\ell_2}}(i_2) \overline{X_{g_{\ell_1}}(j_1)} \\
+ & \sum_{\ell_1, \ell_2=1}^m a_{\ell_1} a_{\ell_2} \frac{1}{n \Delta_n} \sum_{i_1, j_1=1}^n \sum_{i_2, j_2=1}^n K_\theta(i_1 - j_1) K_\theta(i_2 - j_2) \\
& \times \mathbb{E} X_{g_{\ell_1}}(i_1) X_{g_{\ell_2}}(i_2) \overline{\mathbb{E} X_{g_{\ell_1}}(j_1) X_{g_{\ell_2}}(j_2)} \\
= &: \tilde{A}_{n,2} + \bar{A}_{n,2}.
\end{aligned}$$

Define

$$C_{g_{\ell_1}, g_{\ell_2}}(t) := \mathbb{E} X_{g_{\ell_1}}(t) \overline{X_{g_{\ell_2}}(0)}. \quad (6.24)$$

Start with $\theta \notin \{0, \pm\pi\}$. By the same arguments as in Proposition VI.7, one can focus only on \tilde{A}_2 . By the change of variables

$$\begin{aligned}
x_1 &= i_1 - j_2, & x_2 &= i_2 - j_1, \\
y_1 &= i_1 - j_1, & y_2 &= i_1,
\end{aligned}$$

we have that

$$\begin{aligned}
\tilde{A}_{n,2} &= \sum_{\ell_1, \ell_2=1}^m a_{\ell_1} a_{\ell_2} \frac{1}{n \Delta_n} \sum_{x_1, x_2=-n+1}^{n-1} C_{g_{\ell_1}, g_{\ell_2}}(x_1) \overline{C_{g_{\ell_1}, g_{\ell_2}}(-x_2)} \\
& \times \sum_{y_1=(1-n) \vee (1-n+x_1+x_2)}^{(n-1) \wedge (n-1+x_1+x_2)} K_\theta(y_1) K_\theta(x_1 + x_2 - y_1) \sum_{y_2=1 \vee (1+y_1)}^{n \wedge (n+y_1)} 1
\end{aligned}$$

$$\begin{aligned}
&= \sum_{\ell_1, \ell_2=1}^m a_{\ell_1} a_{\ell_2} \sum_{x_1, x_2=-n+1}^{n-1} C_{g_{\ell_1}, g_{\ell_2}}(x_1) e^{ix_1 \theta} \overline{C_{g_{\ell_1}, g_{\ell_2}}(-x_2) e^{-ix_2 \theta}} \\
&\quad \times \frac{1}{\Delta_n} \sum_{y_1=(1-n) \vee (1-n+x_1+x_2)}^{(n-1) \wedge (n-1+x_1+x_2)} K(y_1) K(x_1 + x_2 - y_1) \frac{n - |y_1|}{n} \\
&\sim \sum_{\ell_1, \ell_2=1}^m a_{\ell_1} a_{\ell_2} 4\pi^2 \int_{y=-1}^1 K^2(y) dy \left| \sum_{x=-\infty}^{\infty} C_{g_{\ell_1}, g_{\ell_2}}(x) e^{ix\theta} \right|^2
\end{aligned}$$

based on the proof of Proposition VI.7. By (6.24), this is precisely

$$4\pi^2 \int_{y=-1}^1 K^2(y) dy \sum_{\ell_1, \ell_2=1}^m a_{\ell_1} a_{\ell_2} |\langle f(\theta) g_{\ell_1}, g_{\ell_2} \rangle|^2.$$

The derivations for $\bar{A}_{n,2}$ are similar. For instance, consider $\theta = 0$:

$$\begin{aligned}
\bar{A}_{n,2} &= \frac{1}{n\Delta_n} \sum_{i_1, j_1=1}^n \sum_{i_2, j_2=1}^n K_0(i_1 - j_1) K_0(i_2 - j_2) \\
&\quad \times \sum_{\ell_1, \ell_2=1}^m a_{\ell_1} a_{\ell_2} \check{C}_{g_{\ell_1}, g_{\ell_2}}(i_1 - i_2) \overline{\check{C}_{g_{\ell_1}, g_{\ell_2}}(j_1 - j_2)},
\end{aligned}$$

where

$$\check{C}_{g_{\ell_1}, g_{\ell_2}}(t) := \mathbb{E}[X_{g_{\ell_1}}(t) X_{g_{\ell_2}}(0)] = \left\langle \mathbb{E} \left[X(t) \otimes \overline{X(0)} \right] \overline{g_{\ell_2}}, g_{\ell_1} \right\rangle = \langle \check{C}(t) \overline{g_{\ell_2}}, g_{\ell_1} \rangle.$$

Making the change of variables

$$\begin{aligned}
x_1 &= i_1 - i_2, & x_2 &= j_1 - j_2, \\
y_1 &= i_1 - j_1, & y_2 &= i_1,
\end{aligned}$$

we have that

$$\bar{A}_{n,2} = \sum_{\ell_1, \ell_2=1}^m a_{\ell_1} a_{\ell_2} \frac{1}{\Delta_n} \sum_{x_1, x_2=1-n}^{n-1} \check{C}_{g_{\ell_1}, g_{\ell_2}}(x_1) \overline{\check{C}_{g_{\ell_1}, g_{\ell_2}}(x_2)}$$

$$\begin{aligned}
& \times \sum_{y_1=(1-n) \vee (1-n+x_1-x_2)}^{(n-1) \wedge (n-1)+x_1-x_2} K\left(\frac{y_1}{\Delta_n}\right) K\left(\frac{-x_1+x_2+y_1}{\Delta_n}\right) \frac{n-|y_1|}{n} \\
& \sim 4\pi^2 \int_{y=-1}^1 K^2(y) dy \sum_{\ell_1, \ell_2=1}^m a_{\ell_1} a_{\ell_2} |\langle \check{f}(0) \overline{g_{\ell_2}}, g_{\ell_1} \rangle|^2,
\end{aligned}$$

using again the arguments of the proof of Proposition VI.7. Thus, we have verified (ii) of Proposition VI.6 for $k = 2$.

By the definition of pseudo-covariance in (4.4),

$$\mathbb{E}[X_{g_{\ell_1}}(t)X_{g_{\ell_2}}(0)] = \left\langle \mathbb{E}\left[X(t) \otimes \overline{X(0)}\right] \overline{g_{\ell_2}}, g_{\ell_1} \right\rangle = \langle \check{C}(t) \overline{g_{\ell_2}}, g_{\ell_1} \rangle = \check{C}_{g_{\ell_1}, g_{\ell_2}}(t).$$

In a similar manner, the derivation of $\mathbb{E}[\sum_{\ell=1}^m a_{\ell} \langle \mathcal{T}_n g_{\ell}, g_{\ell} \rangle]^k$ for $k \geq 3$ for a general space \mathbb{H} can be extended from that for the scalar case, and the details are omitted.

6.5 Cumulants and Isserlis' formulas

6.5.1 Cumulants for functional data

This section provides an extension of Isserlis' theorem to the regime of Hilbert space valued Gaussian random variables. This extension is critically used in the verification of property (ii) of Theorem VI.5 in Section 6.4. We start by providing the definition of the cumulants for scalar random variables taking values in \mathbb{R} .

Definition VI.8. Let Y_1, \dots, Y_k be random variables taking values in \mathbb{R} such that $\mathbb{E}(\prod_{j \in B} Y_j)$ is well defined and finite for all subsets B of $\{1, \dots, k\}$. Then,

$$\text{cum}(Y_1, \dots, Y_k) := \sum_{\nu=(\nu_1, \dots, \nu_q)} (-1)^{q-1} (q-1)! \prod_{l=1}^q \mathbb{E} \left[\prod_{j \in \nu_l} Y_j \right],$$

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

The following lemma follows from the discussion on page 34 of *Rosenblatt (1985)*.

Lemma VI.9. Let $Y_i, i = 1, \dots, k$ be real random variables such that $\mathbb{E}(\prod_{j \in B} Y_j)$ is well defined and finite for all subsets B of $\{1, \dots, k\}$. Then

$$\mathbb{E}[Y_1 \cdot \dots \cdot Y_k] = \sum_{\nu=(\nu_1, \dots, \nu_p)} \prod_{l=1}^p \text{cum}(Y_i; i \in \nu_l),$$

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

Proposition VI.10. Let $\{X(t)\}$ be a stochastic process taking values in a Hilbert space \mathbb{H} , where $\mathbb{E}(\|X(t)\|^4) < \infty$ for all t . Note that we do not assume here X to be real. Fix an arbitrary real CONS $\{e_i, i \in I\}$ of \mathbb{H} and denote by $X_i(t) := \langle X(t), e_i \rangle$. Then for any $t, s, w, v \in \mathbb{R}^d$, we have that

$$\text{cum}(X(t), X(s), X(w), X(v)) = \sum_i \sum_j \text{cum}(X_j(t), \overline{X_i(s)}, \overline{X_j(w)}, X_i(v)).$$

Proof. Recall the definition of cumulant in (IV.11):

$$\begin{aligned} & \text{cum}(X(t), X(s), X(w), X(v)) \\ &= \mathbb{E} \langle X(t) \otimes X(s), X(w) \otimes X(v) \rangle_{\text{HS}} - \langle \mathbb{E}(X(t) \otimes X(s)), \mathbb{E}(X(w) \otimes X(v)) \rangle_{\text{HS}} \\ & \quad - \mathbb{E} \langle X(t), X(w) \rangle_{\mathbb{H}} \cdot \mathbb{E} \langle X(v), X(s) \rangle_{\mathbb{H}} \\ & \quad - \left\langle \mathbb{E}(X(t) \otimes \overline{X(v)}), \mathbb{E}(X(w) \otimes \overline{X(s)}) \right\rangle_{\text{HS}}. \end{aligned}$$

For any $x(1), \dots, x(4) \in \mathbb{H}$,

$$\begin{aligned} \langle x(1) \otimes x(2), x(3) \otimes x(4) \rangle_{\text{HS}} &= \sum_i \langle (x(1) \otimes x(2))e_i, (x(3) \otimes x(4))e_i \rangle_{\mathbb{H}} \\ &= \langle x(1), x(3) \rangle_{\mathbb{H}} \overline{\langle x(2), x(4) \rangle_{\mathbb{H}}} \\ &= \sum_i \sum_j x_i(1) \overline{x_i(3)} \overline{x_j(2)} x_j(4). \end{aligned}$$

It follows that

$$\langle X(t) \otimes X(s), X(w) \otimes X(v) \rangle_{\text{HS}} = \sum_i \sum_j \overline{X_i(s)} X_i(v) X_j(t) \overline{X_j(w)}.$$

It suffices to show that

$$\mathbb{E} \langle X(t) \otimes X(s), X(w) \otimes X(v) \rangle_{\text{HS}} = \sum_i \sum_j \mathbb{E}(\overline{X_i(s)} X_i(v) X_j(t) \overline{X_j(w)}),$$

where the interchange of the order of summation and expectation can be justified by the fourth-moment assumption on the $X(t)$ and Fubini's Theorem.

Similarly, we have that

$$\langle \mathbb{E}(X(t) \otimes X(s)), \mathbb{E}(X(w) \otimes X(v)) \rangle_{\text{HS}} = \sum_i \sum_j \mathbb{E}(\overline{X_i(s)} X_j(t)) \mathbb{E}(X_i(v) \overline{X_j(w)})$$

and

$$\langle \mathbb{E}(X(t) \otimes \overline{X(v)}), \mathbb{E}(X(w) \otimes \overline{X(s)}) \rangle_{\text{HS}} = \sum_i \sum_j \mathbb{E}(X_j(t) X_i(v)) \mathbb{E}(\overline{X_i(s)} \overline{X_j(w)}),$$

where we used the fact that the CONS $\{e_j\}$ is real in order to write $\overline{X(s)} = \sum_i \overline{X_i(s)} e_j$.

Finally,

$$\mathbb{E} \langle X(t), X(w) \rangle_{\mathbb{H}} \cdot \mathbb{E} \langle X(v), X(s) \rangle_{\mathbb{H}} = \sum_i \sum_j \mathbb{E} X_j(t) \overline{X_j(w)} \mathbb{E} X_i(v) \overline{X_i(s)}.$$

Gathering all four terms one can easily see that the cumulant sum

$$\sum_i \sum_j \text{cum}(X_j(t), \overline{X_i(s)}, \overline{X_j(w)}, X_i(v))$$

is reconstructed. □

We end this subsection with a remark on the connection with a related but different notion of cumulant employed in *Panaretos and Tavakoli* (2013).

Remark VI.11. *Panaretos and Tavakoli* (2013) defines a notion of cumulant on the bottom of page 571 of the paper. In this remark, we will attempt to explain the connection between the condition $C(0, 4)$ in *Panaretos and Tavakoli* (2013) with (c) of Assumption V'.

For simplicity, we shall work with real Hilbert spaces. Recall that in *Panaretos and Tavakoli* (2013), the authors consider $\mathbb{H} = L^2[0, 1]$ and define the so-called *cumulant kernel*:

$$\text{cum}_{\text{ker}}(X(t_1), \dots, X(t_k)) := \sum_{\nu=(\nu_1, \dots, \nu_p)} (-1)^{p-1} (p-1)! \prod_{\ell=1}^p \mathbb{E} \left[\prod_{j \in \nu_\ell} X(\tau_j; t_j) \right],$$

where $X(t) := (X(\tau; t), \tau \in [0, 1]) \in L^2([0, 1])$. For a kernel of order $2k$, one can define the so-called cumulant operator $\mathcal{R} : L^2([0, 1]^k) \rightarrow L^2([0, 1]^k)$, as

$$\mathcal{R}(h) := \int_{[0, 1]^{2k}} \text{cum}_{\text{ker}}(X(t_1), \dots, X(t_{2k}))(\tau_1, \dots, \tau_{2k}) h(\tau_{k+1}, \dots, \tau_{2k}) d\tau_{k+1} \cdots d\tau_{2k},$$

where the latter is understood as a function of (τ_1, \dots, τ_k) that can be shown to belong to $L^2([0, 1]^k)$.

Fixing a CONS $\{e_j\}$ of $L^2([0, 1])$, for $k = 2$, we obtain that

$$\text{cum}_{\text{ker}}(X(t_1), \dots, X(t_4)) = \sum_{i, j, k, \ell} \text{cum}(X_i(t_1), X_j(t_2), X_k(t_3), X_\ell(t_4)) e_i \otimes e_j \otimes e_k \otimes e_\ell,$$

where cum stands for the usual cumulant of random variables, and where $X_i(t) = \langle X(t), e_i \rangle$ are the coordinates of $X(t)$ in the basis $\{e_j\}$. Thus, in the basis $\{e_i \otimes e_j\}$ of $L^2([0, 1]^2) \equiv L^2([0, 1]) \otimes L^2([0, 1])$, one can view the cumulant operator $\mathcal{R} :$

$L^2([0, 1]) \otimes L^2([0, 1]) \rightarrow L^2([0, 1]) \otimes L^2([0, 1])$ as

$$\mathcal{R} = \sum_{i,j,k,\ell} r_{(i,j),(k,\ell)} (e_i \otimes e_j) \otimes (e_k \otimes e_\ell),$$

where $r_{(i,j),(k,\ell)} := \text{cum}(X_i(t_1), X_j(t_2), X_k(t_3), X_\ell(t_4))$.

From this perspective, by (4.20), we obtain that our notion of a cumulant coincides with the trace of the Hilbert-Schmidt cumulant operator \mathcal{R} :

$$\text{trace}(\mathcal{R}) = \text{cum}(X(t_1), X(t_2), X(t_3), X(t_4)) = \sum_{i,j} r_{(i,j),(i,j)}.$$

On the other hand, the norm of the cumulant kernel employed in the $C(0,4)$ condition of *Panaretos and Tavakoli* (2013) becomes:

$$\|\text{cum}_{\text{ker}}(X(t_1), \dots, X(t_4))\|_{L^2}^2 = \sum_{i,j,k,\ell} \text{cum}(X_i(t_1), X_j(t_2), X_k(t_3), X_\ell(t_4))^2.$$

Whereas, recall that

$$\text{cum}(X(t_1), \dots, X(t_4)) = \sum_{i,j,k,\ell} \text{cum}(X_i(t_1), X_j(t_2), X_k(t_3), X_\ell(t_4)).$$

Thus, the condition $C(0, 4)$ of *Panaretos and Tavakoli* (2013) that

$$\sum_{t_1, t_2, t_3} \|\text{cum}_{\text{ker}}(X(t_1), X(t_2), X(t_3), X(0))\|_{L^2} < \infty$$

is neither strictly weaker nor stronger than our condition (c) in Assumption V' .

6.5.2 Isserlis' formulas

The following is an extension of the classical Isserlis' formula to univariate complex Gaussian variables.

Lemma VI.12. Let $Z_j = X_j + \mathbf{i}Y_j$, $j = 1, 2, \dots$ be zero-mean, complex jointly Gaussian random variables. That is, X_j, Y_j , $j = 1, 2, \dots$ are zero-mean jointly Gaussian \mathbb{R} -valued random variables. Then, for all $m \in \mathbb{N}$, we have $\mathbb{E}[\prod_{i=1}^{2m-1} Z_j] = 0$, and

$$\mathbb{E} \left(\prod_{j=1}^{2m} Z_j \right) = \sum_{\pi} \prod_{i=1}^m \mathbb{E}(Z_{a_{\pi,i}} Z_{b_{\pi,i}}),$$

where a pairing π refers to a decomposition of $\{1, \dots, 2m\}$ into m pairs, which are denoted as $(a_{\pi,i}, b_{\pi,i}), i = 1, \dots, m$.

Proof. Recall that $Z_j = X_j + \mathbf{i}Y_j$, where X_i, Y_i are real. Let $\sigma_{a,b}^{(0,0)} = \mathbb{E}(X_a X_b)$, $\sigma_{a,b}^{(0,1)} = \mathbb{E}(X_a Y_b)$, $\sigma_{a,b}^{(1,0)} = \mathbb{E}(Y_a X_b)$, $\sigma_{a,b}^{(1,1)} = \mathbb{E}(Y_a Y_b)$. Write

$$\mathbb{E} \left(\prod_{j=1}^{2m} (X_j + \mathbf{i}Y_j) \right) = \sum_{S \subset \{1, \dots, 2m\}} \mathbf{i}^{|S|} \mathbb{E} \left(\prod_{j \notin S} X_j \prod_{k \in S} Y_k \right).$$

By the Isserlis formula for real Gaussian random variables introduced by *Isserlis* (1918), we have

$$\mathbb{E} \left(\prod_{j \notin S} X_j \prod_{k \in S} Y_k \right) = \sum_{\pi} \prod_{i=1}^m \sigma_{a_{\pi,i}, b_{\pi,i}}^{(\mathbb{1}(a_{\pi,i} \in S), \mathbb{1}(b_{\pi,i} \in S))},$$

and hence

$$\mathbb{E} \left(\prod_{j=1}^{2m} (X_j + \mathbf{i}Y_j) \right) = \sum_{\pi} \sum_{S \subset \{1, \dots, 2m\}} \mathbf{i}^{|S|} \prod_{i=1}^m \sigma_{a_{\pi,i}, b_{\pi,i}}^{(\mathbb{1}(a_{\pi,i} \in S), \mathbb{1}(b_{\pi,i} \in S))}.$$

For any given π and S , we let $\alpha_i = \mathbb{1}(a_{\pi,i} \in S)$, $\beta_i = \mathbb{1}(b_{\pi,i} \in S)$. Therefore,

$$\begin{aligned}
\mathbb{E} \left(\prod_{j=1}^{2m} (X_j + \mathfrak{i}Y_j) \right) &= \sum_{\pi} \sum_{\substack{\alpha_i, \beta_i=0,1 \\ i=1, \dots, m}} \mathfrak{i}^{\sum_i \alpha_i + \sum_i \beta_i} \prod_{i=1}^m \sigma_{a_{\pi,i}, b_{\pi,i}}^{(\alpha_i, \beta_i)} \\
&= \sum_{\pi} \prod_{i=1}^m \sum_{\alpha_i, \beta_i=0,1} \mathfrak{i}^{\alpha_i + \beta_i} \sigma_{a_{\pi,i}, b_{\pi,i}}^{(\alpha_i, \beta_i)} \\
&= \sum_{\pi} \prod_{i=1}^m (1, \mathfrak{i}) C(a_{\pi,i}, b_{\pi,i}) (1, \mathfrak{i})^{\top},
\end{aligned} \tag{6.25}$$

where

$$C(a, b) = \begin{pmatrix} \mathbb{E}(X_a X_b) & \mathbb{E}(X_a Y_b) \\ \mathbb{E}(X_b Y_a) & \mathbb{E}(Y_a Y_b) \end{pmatrix}.$$

Notice that $\mathbb{E}[Z_a Z_b] = (1, \mathfrak{i}) C(a, b) (1, \mathfrak{i})^{\top}$ and thus the right-hand side of (6.25) equals

$$\sum_{\pi} \prod_{i=1}^m \mathbb{E}(Z_{a_{\pi,i}} Z_{b_{\pi,i}}),$$

which shows that the Isserlis formula for complex-valued r.v.'s is exactly the same as that for real-valued random variables. \square

Lemma VI.13. *Let $\{X(t), t \in \mathbb{R}\}$ be a stationary Gaussian process in \mathbb{C} with $C(t-s) = \mathbb{E}X(t)\overline{X(s)}$ and $\check{C}(t-s) = \mathbb{E}X(t)X(s)$. Consider $X(t_i), X(s_i), i = 1, \dots, N+M$ for some $N, M \in \mathbb{Z}$, with $N \geq 0$ and $M \geq 0$. Denote by $\mathcal{P}_{N,M}$ the class of all pairings of the set*

$$\{t_i, s_i | i = 1, \dots, N+M\}$$

and by $\mathcal{P}_{N,M,-k}$ the class of all pairings of

$$\{t_i, s_i | i = 1, \dots, N+M\} \setminus \{t_k, s_k\}.$$

This means that $\tilde{u} \in \mathcal{P}_{N,M}$ iff

$$\tilde{u} = \left\{ \left\{ \tau, \sigma \right\}, \left\{ \tau, \tilde{\tau} \right\}, \left\{ \sigma, \tilde{\sigma} \right\} \mid \begin{aligned} &\tau \neq \tilde{\tau} \in \{t_i : i = 1, \dots, N + M\}, \\ &\sigma \neq \tilde{\sigma} \in \{s_i : i = 1, \dots, N + M\} \end{aligned} \right\}$$

and each symbol $t_i, s_i, i = 1, \dots, N + M$ can be used only once. Then,

$$\begin{aligned} &\mathbb{E} \left[\prod_{n=1}^N X(t_n) \overline{X(s_n)} \cdot \prod_{m=1}^M \left(X(t_{N+m}) \overline{X(s_{N+m})} - C(t_{N+m} - s_{N+m}) \right) \right] \\ &= \sum_{\substack{\tilde{u} \in \mathcal{P}_{N,M}: \\ \cup_{m=1}^M \{t_{N+m}, s_{N+m}\} \cap \tilde{u} = \emptyset}} \prod_{\{i,j\} \in \tilde{u}} \left[C(\tau - \sigma) \mathbb{1}_{\{i,j\} = \{\tau, \sigma\}} + \check{C}(\tau - \tilde{\tau}) \mathbb{1}_{\{i,j\} = \{\tau, \tilde{\tau}\}} \right. \\ &\quad \left. + \overline{C(\sigma - \tilde{\sigma})} \mathbb{1}_{\{i,j\} = \{\sigma, \tilde{\sigma}\}} \right]. \end{aligned} \quad (6.26)$$

This Isserlis-type result is used in the proof of Proposition VI.7 (see e.g. (6.22)), where the k th order moments of the spectral density estimators involve terms as in (6.26) where $N = 0$. The reason we formulate (6.26) for general $N \geq 0$ is to facilitate the proof of this relation by the method of induction.

Proof of Lemma VI.13. We will prove the desired equality by using induction on $N + M$. When $N + M = 1$, the equality holds trivially. For the basis of our induction, we use $N + M = 2$. We look at the three different cases.

(a) $N = 2$. The equality trivially holds by the Isserlis' formula.

(b) $N = M = 1$. We have

$$\begin{aligned} &\mathbb{E} \left[X(t_1) \overline{X(s_1)} \cdot (X(t_2) \overline{X(s_2)} - C(t_2 - s_2)) \right] \\ &= \mathbb{E} \left[X(t_1) \overline{X(s_1)} X(t_2) \overline{X(s_2)} \right] - C(t_1 - s_1) C(t_2 - s_2) \\ &= \mathbb{E} \left[X(t_1) \overline{X(s_1)} \right] \mathbb{E} \left[X(t_2) \overline{X(s_2)} \right] + \mathbb{E} [X(t_1) X(t_2)] \mathbb{E} \left[\overline{X(s_1)} X(s_2) \right] \\ &\quad + \mathbb{E} \left[X(t_1) \overline{X(s_2)} \right] \mathbb{E} \left[\overline{X(s_1)} X(t_2) \right] - C(t_1 - s_1) C(t_2 - s_2) \end{aligned}$$

$$= \check{C}(t_1 - t_2)\overline{\check{C}(s_1 - s_2)} + C(t_1 - s_2)C(t_2 - s_1)$$

where Isserlis' formula was used in the second equality.

(c) $M = 2$. We have that

$$\begin{aligned} & \mathbb{E} \left[(X(t_1)\overline{X(s_1)} - C(t_1 - s_1)) \cdot (X(t_2)\overline{X(s_2)} - C(t_2 - s_2)) \right] \\ &= \mathbb{E} \left[X(t_1)\overline{X(s_1)}X(t_2)\overline{X(s_2)} \right] - C(t_1 - s_1)C(t_2 - s_2) \\ &= \check{C}(t_1 - t_2)\overline{\check{C}(s_1 - s_2)} + C(t_1 - s_2)C(t_2 - s_1) \end{aligned} \quad (6.27)$$

similarly to case (b).

For the induction hypothesis, we assume that the desired equality holds when $N + M = r$. Let now $N + M = r + 1$. We discern two cases.

(a) $N = r + 1$. The result follows directly by Isserlis' formula.

(b) $N < r + 1$. The following holds

$$\begin{aligned} & \mathbb{E} \left[\prod_{n=1}^N X(t_n)\overline{X(s_n)} \cdot \prod_{m=1}^M \left(X(t_{N+m})\overline{X(s_{N+m})} - C(t_{N+m} - s_{N+m}) \right) \right] \\ &= \mathbb{E} \left[\prod_{n=1}^{N+1} X(t_n)\overline{X(s_n)} \prod_{m=2}^M \left(X(t_{N+m})\overline{X(s_{N+m})} - C(t_{N+m} - s_{N+m}) \right) \right] \\ & \quad - \mathbb{E} \left[\prod_{n=1}^N X(t_n)\overline{X(s_n)} \cdot \prod_{m=2}^M \left(X(t_{N+m})\overline{X(s_{N+m})} - C(t_{N+m} - s_{N+m}) \right) \right] \\ & \quad \times C(t_{N+1} - s_{N+1}) \\ &= \dots \\ &= \mathbb{E} \left[\prod_{n=1}^{N+M} X(t_n)\overline{X(s_n)} \right] \\ & \quad - \sum_{m=1}^M C(t_{N+m} - s_{N+m}) \mathbb{E} \left[\prod_{n=1}^{N+m-1} X(t_n)\overline{X(s_n)} \right. \\ & \quad \left. \cdot \prod_{k=m+1}^M \left(X(t_{N+k})\overline{X(s_{N+k})} - C(t_{N+k} - s_{N+k}) \right) \right] \end{aligned}$$

Applying Isserlis' formula the first summand is equal to

$$\sum_{\tilde{u} \in \mathcal{P}_{N,M}} \prod_{\{i,j\} \in \tilde{u}} \left[C(\tau - \sigma) \mathbb{1}_{\{i,j\}=\{\tau,\sigma\}} + \check{C}(\tau - \tilde{\tau}) \mathbb{1}_{\{i,j\}=\{\tau,\tilde{\tau}\}} + \overline{\check{C}(\sigma - \tilde{\sigma})} \mathbb{1}_{\{i,j\}=\{\sigma,\tilde{\sigma}\}} \right].$$

By applying the induction hypothesis in the second summand, since all terms involve r factors in total, we have that the second term is equal to

$$\sum_{m=1}^M C(t_{N+m} - s_{N+m}) \sum_{\substack{\tilde{u} \in \mathcal{P}_{N,M,-m}: \\ \cup_{i=N+m+1}^{N+M} \{\{t_i, s_i\}\} \cap \tilde{u} = \emptyset}} \prod_{\{i,j\} \in \tilde{u}} \left[C(\tau - \sigma) \mathbb{1}_{\{i,j\}=\{\tau,\sigma\}} + \check{C}(\tau - \tilde{\tau}) \mathbb{1}_{\{i,j\}=\{\tau,\tilde{\tau}\}} + \overline{\check{C}(\sigma - \tilde{\sigma})} \mathbb{1}_{\{i,j\}=\{\sigma,\tilde{\sigma}\}} \right].$$

Denote the first term of the previous sum as A , the second term as B , write $B := \sum_{m=1}^M B_m$. Also, let

$$C_{i,j}^{\tilde{u}} := C(\tau - \sigma) \mathbb{1}_{\{i,j\}=\{\tau,\sigma\}} + \check{C}(\tau - \tilde{\tau}) \mathbb{1}_{\{i,j\}=\{\tau,\tilde{\tau}\}} + \overline{\check{C}(\sigma - \tilde{\sigma})} \mathbb{1}_{\{i,j\}=\{\sigma,\tilde{\sigma}\}}.$$

Then we have that

$$\begin{aligned} A - B_M &= \sum_{\tilde{u} \in \mathcal{P}_{N,M}} \prod_{\{i,j\} \in \tilde{u}} C_{i,j}^{\tilde{u}} - C(t_{N+M} - s_{N+M}) \sum_{\tilde{u}' \in \mathcal{P}_{N,M,-M}} \prod_{\{i,j\} \in \tilde{u}'} C_{i,j}^{\tilde{u}'} \\ &= \sum_{\substack{\tilde{u} \in \mathcal{P}_{N,M}: \\ \{t_{N+M}, s_{N+M}\} \notin \tilde{u}}} \prod_{\{i,j\} \in \tilde{u}} C_{i,j}^{\tilde{u}}. \end{aligned}$$

Similarly

$$A - B_M - B_{M-1} = \sum_{\substack{\tilde{u} \in \mathcal{P}_{N,M}: \\ \cup_{i=M-1}^M \{\{t_{N+i}, s_{N+i}\}\} \cap \tilde{u} = \emptyset}} \prod_{\{i,j\} \in \tilde{u}} C_{i,j}^{\tilde{u}}.$$

Continuing this way for all terms B_j , $j = 1, \dots, M$, we have that the proof is complete.

□

CHAPTER VII

Future Directions

The purpose of this chapter is to succinctly present the core message of the two parts of this thesis. After summarizing the main contributions of each chapter, we discuss potential expansions of our work and questions that could be posed as future research problems, since they yet remain unanswered in this dissertation.

7.1 Anomaly Detection

In Chapter II, we studied the anomaly detection problem for high dimensional data in the context of Internet traffic. We assumed that the observed traffic \vec{x}_t follows the linear factor model

$$\vec{x}_t = B\vec{f}_t + \vec{u}_t + \vec{\epsilon}_t$$

and we developed Algorithm 1, in order to detect the anomalies \vec{u}_t . The effectiveness of this algorithm is based on the incoherence conditions of Section 2.2.3, under which we expect the anomalies to “pass-through” to the residuals, obtained after the projection step of the algorithm. Mathematically, we have that

$$\vec{r}_t = \text{Proj}_{\text{col}(\hat{B})^\perp}(\vec{x}_t) = \text{Proj}_{\text{col}(\hat{B})^\perp}(B\vec{f}_t + \vec{u}_t + \vec{\epsilon}_t) \approx \vec{u}_t + \vec{\epsilon}_t, \quad (7.1)$$

which means that if the estimation of the column subspace of B is sufficiently good, then the anomalies will not get filtered out after the projection on the orthogonal complement of $\text{col}(\hat{B})$. This follows from the incoherence conditions, because of which the anomalies do not belong on the subspace produced by the columns of B .

A theoretical result quantifying the approximation in (7.1) is stated in Proposition II.3. Note, however, that the inequalities (2.9) and (2.10) both depend on the quantity $\mathbb{E}\|\hat{\Sigma} - \Sigma\|^2$, where $\Sigma = BB^\top$ and $\hat{\Sigma}$ is an estimate of Σ . One challenge, that has not been addressed in Chapter II, is the estimation of bounds for $\|\hat{\Sigma} - \Sigma\|$. Specifically, in order to complete the result of Proposition II.3, one needs to find appropriate bounds on $\mathbb{E}\|\hat{\Sigma} - \Sigma\|^2$ under assumptions on the dependence structure of $\vec{\epsilon}_t$.

An additional challenge for this chapter pertains to the efficient application of Algorithm 1. We have demonstrated the effectiveness of our method in both synthetic and real-world Darknet data and have also shown its superiority against competing methods in the literature. In order to further improve the algorithm, we would like to implement it efficiently in an *online* fashion. In particular, apart from the initialization phase, where a big “warm-up” dataset has to be utilized, we would like to find a way to make the rest of the algorithm absolutely sequential. This would save resources, in both the aspect of storage space and of computation time, making the algorithm faster to implement and thus even more suitable for quick identification of anomalies at their onset.

7.2 Concentration Rates

Our focus in Chapter III is centered on studying the rates of Uniform Relative Stability (URS) for Gaussian triangular arrays in the context of dependence. As already discussed, our motivations for studying the concentration rates in the URS property are twofold. Firstly, as established by *Gao and Stoev* (2020), this property is the key to understanding the “0/1” phase-transition phenomenon in the exact support

recovery problem for the canonical signal-plus-noise model (see (3.6)), irrespective of the marginal distribution of the error terms $\epsilon_p(i)$. Secondly, utilizing the obtained upper bounds on the rates of concentration for Gaussian triangular arrays, we want to explore whether the URS property is preserved under transformations of Gaussian arrays.

For broad classes of transformations of Gaussian triangular arrays, we have shown that URS is indeed preserved, as described in Proposition III.19 and Corollary III.21. This leads to a plethora of models, for the error terms of the signal-plus-noise model, that obey a phase transition in the exact support recovery problem, namely whether one can find a thresholding estimator, \widehat{S}_p , that achieves perfect recovery of the sparse support set of the signal, S_p . Two characteristic models associated with these transformations are the χ^2 and the log-Normal model.

One interesting question arises by examining the rate upper bounds on the power law and on the exponential power law transformations. Indeed, the former transformations, presented in Example III.23, seem to be an “easy” case in the preservation of the URS property. The upper bound therein is

$$d_p^{*,\text{UB}} \sim \delta_p^{\text{UB}},$$

which means that if \mathcal{E} is URS, then the same holds for $\mathcal{H} = f(\mathcal{E})$, with the covariance structure of \mathcal{E} playing no role. However, this is not the case for the latter transformations, described in Example III.24. The order of the upper bound for these transformations is

$$d_p^{*,\text{UB}} \sim \delta_p^{\text{UB}}(2 \log(p))^{\lambda/2}.$$

Together with Conjecture III.11, this means that this rate bound is only valid for $0 < \lambda < 2$. Moreover, even if λ is in this range, Theorem III.13 cannot be used to secure that the URS property will be preserved irrespective of the dependence

structure of \mathcal{E} . As an example, let \mathcal{E} have logarithmic covariance decay as in Example III.18. Then, we have established that $\delta_p^{\text{UB}} \sim (\log(p))^{-\frac{\nu}{\nu+1}}$ and thus

$$d_p^{*,\text{UB}} \sim (\log(p))^{\frac{\lambda}{2} - \frac{\nu}{\nu+1}}.$$

If $\nu \leq \lambda/(2 - \lambda)$, the above rate does not vanish and thus cannot be used as an upper bound for the URS. With our results so far, whether the resulting array $\mathcal{H} = e^{\mathcal{E}^\lambda}$ is URS, if \mathcal{E} is URS, remains an open question. An even more general open question is the characterization of URS for non-Gaussian triangular arrays, that are more general than transformations of Gaussian noise, under general dependence structures.

Returning to the Gaussian regime for \mathcal{E} , an open problem would be to explore the relative stability of the maxima under an even broader sense of uniformity. In particular, let the following property hold

$$\max_{\substack{|S| \geq g \\ S \subset \{1, \dots, p\}}} \left| \frac{M_S}{u_{|S|}} - 1 \right| \xrightarrow{\mathbb{P}} 0, \text{ as } g \rightarrow \infty,$$

where $M_S := \max_{i \in S} \epsilon_p(i)$ and $u_{|S|}$ is as defined in (3.8). Relating this property to the exact support recovery phase transition, as well as examining if any improvements on the upper bounds of the optimal concentration rates are possible would be a problem of interest.

One more possible direction for future research would be focusing on applications. Specifically, we have already mentioned that for a broad class of error models, formed as transformations of the Gaussian model, a phase transition phenomenon holds in the exact support recovery problem. This means that for a suitable boundary function, there is always an appropriate thresholding estimator so that the support set of the signal can be recovered exactly, if the signal is larger in magnitude than the boundary. The choice of the threshold of the estimator in practice, ideally through a data-driven approach, is an unsolved methodological question.

Finally, all the derivations in Chapter III apply to thresholding estimators of the signal support set. A natural question to pose is whether results of similar type hold for support estimators not included in this category. At this time, whether a phase transition phenomenon holds under dependence for types of estimators than are not based on a threshold remains an open problem.

7.3 Spectral Inference

In the second part of the thesis, we focused on the estimation of the spectral density for functional spatial data. Starting with Chapter IV, we presented a unified approach on pointwise spectral density estimation, with results that are valid for any finite space/time dimension and any separable Hilbert space. However, a question that has been left unaddressed is a recommendation for the selection of the bandwidth parameter Δ_n in practice.

An answer to this problem is not trivial to obtain. *Zhu and Politis (2020)* provide an empirical recommendation for a data-dependent bandwidth choice for estimators very similar to ours. Their recommendation in Section 5 therein refers to the so-called flat-top kernel estimators and the approach is based on the correlogram of the observed process.

An alternative approach to deciding the bandwidth parameter could be through the derivation of a cross-validation methodology. Consider $f(\theta; \Delta)$ an estimate of the true spectral density $f(\theta)$. The key premises of this methodology would be the following:

1. $f(\theta; \Delta)$ and $f(\theta'; \Delta)$ are asymptotically independent for θ and θ' different!
2. $f(\theta)$ is “close” to $f(\theta')$, for θ close to θ' .
3. $f(\theta; \Delta_1)$ has a smaller-order bias for “large” Δ_1 .

The proposed methodology would be to define

$$\text{CV}(\Delta) := \|f(\theta, \Delta)\|^2 - 2\Re\langle f(\theta', \Delta_1), f(\theta, \Delta) \rangle,$$

where θ' is “nearby” θ and $0 < \Delta \ll \Delta_1$. The idea would be to hold Δ_1 fixed and optimize $\text{CV}(\Delta)$ over Δ . Alternatively, one could pick a sample of $\theta'(i), i = 1, \dots, k$ that are different but surround θ and define

$$\text{CV}_k(\Delta) = \|f(\theta, \Delta)\|^2 - 2\frac{1}{k} \sum_{i=1}^k \Re\langle f(\theta'(i), \Delta_1), f(\theta, \Delta) \rangle.$$

This way could provide a less-biased estimate of $f(\theta)$ by averaging $f(\theta'(i); \Delta_1)$ over the $\theta'(i)$'s. This methodology is relatively easy to implement and is rather “cheap” since it doesn't involve “leave out”. The idea is that we look at “nearby” $\theta'(i)$ to break the dependence between $f(\theta'; \Delta_1)$ and $f(\theta; \Delta)$. Also looking at $\Delta_1 \gg \Delta$ ensures that the bias of $f(\theta'; \Delta_1)$ is of smaller order. To the best of our knowledge, such an approach has not been developed yet.

The development of a data-driven selection method for the bandwidth would pave the road for the application of the estimator in practice. For this purpose, an efficient algorithmic implementation remains to be constructed. Once this construction is complete, the estimator $\hat{f}_n(\theta)$ could be used for the pointwise estimation of the spectral density for a wide variety of potentially irregularly sampled functional spatial stochastic processes. Moreover, the estimator could find application in testing the reversibility of stationary stochastic processes for $d \geq 1$. In this setting, it can be shown that a process is time reversible if and only if the spectral density is real. Thus an estimate of the spectral density with very small imaginary part would be a good indicator that the process under examination is time reversible.

In Chapter V we were able to establish a minimax result for the power law decaying covariance class, both in the discrete and the continuous time case. However, as stated

in Remark V.4, there is a “gap” on this minimax rate result in the continuous time case, for the coarse sampling regime (cf. Remark IV.20). The construction of a class narrower than \mathcal{P}_C (cf. (4.36)), still depending on β and γ , so that the upper- and lower-bounds on the rate of the estimators match in both the fine- and coarse-sampling regimes is an open problem.

Finally, in Chapter VI we secured a CLT type result for the estimator $\hat{f}_n(\theta)$, under the extra assumption of Gaussianity of the underlying stochastic process. An open question related to this chapter is to relax the Gaussianity assumption. In this direction, *Panaretos and Tavakoli* (2013) were able to obtain some relevant results using assumptions on the higher order cumulants of the process. Another idea to tackle this problem would be by using the fourth moment theorem in the context of Wiener chaos [see e.g. *Peccati and Taqqu* (2011)].

BIBLIOGRAPHY

BIBLIOGRAPHY

- Adler, R. J., and J. E. Taylor (2009), *Random fields and geometry*, Springer Science & Business Media.
- Antonakakis, M., et al. (2017), Understanding the Mirai Botnet, in *26th USENIX Security Symposium (USENIX Security 17)*.
- ARGO (2020), Argo float data and metadata from global data assembly centre (Argo GDAC). SEANOE.
- Arias-Castro, E., and S. Chen (2017), Distribution-free multiple testing, *Electronic Journal of Statistics*, 11(1), 1983–2001.
- Arias-Castro, E., and A. Ying (2018), Detection of Sparse Mixtures: Higher Criticism and Scan Statistic, *arXiv preprint arXiv:1802.08715*.
- Arora, R., A. Cotter, K. Livescu, and N. Srebro (2012a), Stochastic optimization for PCA and PLS, in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 861–868, doi:10.1109/Allerton.2012.6483308.
- Arora, R., A. Cotter, K. Livescu, and N. Srebro (2012b), Stochastic optimization for PCA and PLS, in *2012 50th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*, pp. 861–868, IEEE.
- Ban, T., L. Zhu, J. Shimamura, S. Pang, D. Inoue, and K. Nakao (2012), Behavior analysis of long-term cyber attacks in the darknet, in *International Conference on Neural Information Processing*.
- Ban, T., S. Pang, M. Eto, D. Inoue, K. Nakao, and R. Huang (2016), Towards early detection of novel attack patterns through the lens of a large-scale darknet, in *2016 Intl IEEE UIC/ATC/ScalCom/CBDCCom/IoP/SmartWorld*.
- Bardet, J.-M., and P. R. Bertrand (2010), A non-parametric estimator of the spectral density of a continuous-time Gaussian process observed at random times, *Scand. J. Stat.*, 37(3), 458–476, doi:10.1111/j.1467-9469.2009.00684.x.
- Barndorff-Nielsen, O. (1963), On the limit behaviour of extreme order statistics, *The Annals of Mathematical Statistics*, 34(3), 992–1002.

- Benson, K., A. Dainotti, K. Claffy, and E. Aben (2012), Gaining insight into AS-level outages through analysis of internet background radiation, in *Proceedings of the 2012 ACM conference on CoNEXT student workshop*, pp. 63–64.
- Bentkus, R. (1985), Rate of uniform convergence of statistical estimators of spectral density in spaces of differentiable functions, *Lithuanian Mathematical Journal*, 25(3), 209–219.
- Berman, S. M. (1964), Limit theorems for the maximum term in stationary sequences, *The Annals of Mathematical Statistics*, pp. 502–516.
- Billingsley, P. (1995), *Measure and Probability*, John Wiley & Sons: New York.
- Billingsley, P. (2012), *Probability and Measure, anniversary ed*, Wiley New York.
- Billingsley, P. (2013), *Convergence of Probability Measures*, John Wiley & Sons.
- Bingham, N. H., C. M. Goldie, J. L. Teugels, and J. Teugels (1989), *Regular Variation*, 27, Cambridge University Press.
- Bjorck, A., G. H. Golub, et al. (1973), Numerical methods for computing angles between linear subspaces, *Mathematics of computation*.
- Bochner, S. (1948), *Vorlesungen über Fouriersche integrale*, Chelsea Publishing Company.
- Bosq, D. (2000), *Linear processes in function spaces: theory and applications*, vol. 149, Springer Science & Business Media.
- Bou-Harb, E., N.-E. Lakhdari, H. Binsalleeh, and M. Debbabi (2014), Multidimensional investigation of source port 0 probing, *Digital Investigation*, 11, S114–S123.
- Brillinger, D. R. (2001), *Time series, Classics in Applied Mathematics*, vol. 36, xx+540 pp., Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, doi:10.1137/1.9780898719246, Data analysis and theory, Reprint of the 1981 edition.
- Brockwell, P. J., and R. A. Davis (2006), *Time Series: Theory and Methods*, Springer Series in Statistics, xvi+577 pp., Springer, New York, reprint of the second (1991) edition.
- Butucea, C., M. Ndaoud, N. A. Stepanova, and A. B. Tsybakov (2018), Variable selection with Hamming loss, *The Annals of Statistics*, 46(5), 1837–1875.
- Cai, T. T., J. Jin, M. G. Low, et al. (2007), Estimation and confidence sets for sparse normal mixtures, *The Annals of Statistics*, 35(6), 2421–2449.
- Candès, E., and B. Recht (2012), Exact matrix completion via convex optimization, *Communications of the ACM*.

- Candès, E. J., X. Li, Y. Ma, and J. Wright (2011), Robust principal component analysis?, *Journal of the ACM (JACM)*, 58(3), 11.
- Cerovecki, C., and S. Hörmann (2017), On the CLT for discrete fourier transforms of functional time series, *Journal of multivariate analysis*, 154, 282–295.
- Chang, C.-H., H.-C. Huang, and C.-K. Ing (2017), Mixed domain asymptotics for a stochastic process model with time trend and measurement error, *Bernoulli*, 23(1), 159–190.
- Chatterjee, S. (2014), *Superconcentration and related topics*, vol. 15, Springer.
- Czyz, J., K. Lady, S. G. Miller, M. Bailey, M. Kallitsis, and M. Karir (2013), Understanding IPv6 Internet Background Radiation, in *IMC 2013*.
- Czyz, J., M. Kallitsis, M. Gharaibeh, C. Papadopoulos, M. Bailey, and M. Karir (2014), Taming the 800 Pound Gorilla: The Rise and Decline of NTP DDoS attacks, in *Proceedings of the 2014 Conference on Internet Measurement Conference*.
- Dainotti, A., R. Amman, E. Aben, and K. C. Claffy (), Extracting Benefit from Harm: Using Malware Pollution to Analyze the Impact of Political and Geophysical Events on the Internet, *SIGCOMM CCR 2012*.
- Dainotti, A., A. King, K. Claffy, F. Papale, and A. Pescapé (2015), Analysis of a “/0” Stealth Scan From a Botnet, *IEEE/ACM Transactions on Networking*, 23(2), 341–354, doi:10.1109/TNET.2013.2297678.
- Dasgupta, N., and J. D. Spurrier (1997), A class of multivariate chi-square distributions with applications to comparison with a control, *Communications in Statistics - Theory and Methods*, 26(7), 1559–1573, doi:10.1080/03610929708832000.
- Davies, R. B. (1973), Asymptotic inference in stationary Gaussian time-series, *Advances in Applied Probability*, pp. 469–497.
- Degras, D., and H. Cardot (2019), onlinePCA: Online Principal Component Analysis, <https://rdr.io/cran/onlinePCA/>.
- DHS CISA, and NCSC (2020), COVID-19 Exploited by Malicious Cyber Actors, <https://www.cisa.gov/uscert/ncas/alerts/aa20-099a>.
- Donoho, D., and J. Jin (2004a), Higher Criticism for detecting sparse heterogeneous mixtures, *The Annals of Statistics*, 32(3), 962–994.
- Donoho, D., and J. Jin (2004b), Higher Criticism for detecting sparse heterogeneous mixtures, *The Annals of Statistics*, 32(3), 962–994.
- Efromovich, S. (1998), Data-driven efficient estimation of the spectral density, *Journal of the American Statistical Association*, 93(442), 762–769.

- Fan, J. (1996), Test of significance based on wavelet thresholding and Neyman's truncation, *Journal of the American Statistical Association*.
- Fazekas, I., and A. G. Kukush (2000), Infill asymptotics inside increasing domains for the least squares estimator in linear models, *Statistical inference for stochastic processes*, 3(3), 199–223.
- Feller, W. (2008), *An Introduction to Probability Theory and Its Applications*, Vol. 2, John Wiley & Sons.
- Feng, J., H. Xu, and S. Yan (2013), Online robust PCA via stochastic optimization, in *Advances in Neural Information Processing Systems*.
- Ferraty, F., and P. Vieu (2006), *Nonparametric functional data analysis: theory and practice*, vol. 76, Springer.
- Gao, Z., and S. Stoev (2018), Fundamental Limits of Exact Support Recovery in High Dimensions, *ArXiv e-prints*, arXiv:1811.05124.
- Gao, Z., and S. Stoev (2020), Fundamental Limits of Exact Support Recovery in High Dimensions, *Bernoulli*, 26(4), 2605 – 2638, doi:10.3150/20-BEJ1197.
- Gao, Z., and S. Stoev (2021), *Concentration of Maxima and Fundamental Limits in High-Dimensional Testing and Inference*, SpringerBriefs in Probability and Mathematical Statistics, Springer, Cham, doi:https://doi.org/10.1007/978-3-030-80964-5.
- Ginovyan, M. S. (2011), Efficient estimation of spectral functionals for continuous-time stationary models, *Acta Appl. Math.*, 115(2), 233–254, doi:10.1007/s10440-011-9617-7.
- Gnedenko, B. (1943), Sur la distribution limite du terme maximum d'une série aléatoire, *Annals of Mathematics*, pp. 423–453.
- Gruber, P. M. (2007), *Convex and Discrete Geometry*, vol. 336, Springer Science & Business Media.
- Guyon, X. (1982), Parameter estimation for a stationary process on a d -dimensional lattice, *Biometrika*, 69(1), 95–105.
- Hall, P., and P. Patil (1994), Properties of nonparametric estimators of autocovariance for stationary random fields, *Probability Theory and Related Fields*, 99(3), 399–424.
- Halliwel, L. J. (2015), The Lognormal Random Multivariate, in *Casualty Actuarial Society E-Forum*, Spring, p. 5.
- Hannan, E. J. (1970), *Multiple Time Series*, xi+536 pp., John Wiley and Sons, Inc., New York-London-Sydney.

- Hardy, G. H., J. E. Littlewood, G. Pólya, G. Pólya, D. Littlewood, et al. (1952), *Inequalities*, Cambridge university press.
- Hörmann, S., and P. Kokoszka (2012), Functional time series, in *Handbook of statistics*, vol. 30, pp. 157–186, Elsevier.
- Horváth, L., and P. Kokoszka (2012), *Inference for functional data with applications*, vol. 200, Springer Science & Business Media.
- Horváth, L., P. Kokoszka, and G. Rice (2014), Testing stationarity of functional time series, *Journal of Econometrics*, 179(1), 66–82.
- Hsing, T., and R. Eubank (2015), *Theoretical foundations of functional data analysis, with an introduction to linear operators*, vol. 997, John Wiley & Sons.
- Ingster, Y. I. (1998a), Minimax detection of a signal for l_n^p -balls, *Mathematical Methods of Statistics*, 7(4), 401–428.
- Ingster, Y. I. (1998b), Minimax detection of a signal for ℓ_n^p -balls, *Mathematical Methods of Statistics*, 7(4), 401–428.
- Isserlis, L. (1918), On a formula for the product-moment coefficient of any order of a normal frequency distribution in any number of variables, *Biometrika*, 12(1/2), 134–139.
- Ji, P., and J. Jin (2012), UPS delivers optimal phase diagram in high-dimensional variable selection, *The Annals of Statistics*, 40(1), 73–103.
- Jonker, M., A. King, J. Krupp, C. Rossow, A. Sperotto, and A. Dainotti (2017), Millions of Targets Under Attack: A Macroscopic Characterization of the DoS Ecosystem, in *Proceedings of the 2017 IMC*.
- Kartsioukas, R., Z. Gao, and S. Stoev (2021), On the rate of concentration of maxima in Gaussian arrays, *Extremes*, 24(1), 37–65, doi:10.1007/s10687-020-00399-8.
- Khintchine, A. (1934), Korrelationstheorie der stationären stochastischen Prozesse, *Mathematische Annalen*, 109(1), 604–615.
- Kinoshita, K., and S. I. Resnick (1991), Convergence of scaled random samples in \mathbb{R}^d , *The Annals of Probability*, pp. 1640–1663.
- Klopsch, A., C. Dietrich, and R. Springer (2020), A detailed look into the Mozi P2P IoT botnet, <https://www.botconf.eu/botconf-2020/schedule/>.
- Krebs, B. (2021a), KrebsOnSecurity Hit With Record DDos, <https://krebsonsecurity.com/2016/09/krebsonsecurity-hit-with-record-ddos/>.
- Krebs, B. (2021b), KrebsOnSecurity Hit By Huge New IoT Botnet Meris, <https://krebsonsecurity.com/2021/09/krebsonsecurity-hit-by-huge-new-iot-botnet-meris/>.

- Kührer, M., T. Hupperich, C. Rossow, and T. Holz (2014), Exit from hell? reducing the impact of amplification ddos attacks, in *Proceedings of the 23rd USENIX Conference on Security Symposium*.
- Kuusela, M., and M. Stein (2018), Locally stationary spatio-temporal interpolation of Argo profiling float data, *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 474(2220), 20180,400.
- Lakhina, A., M. Crovella, and C. Diot (2004), Diagnosing Network-Wide Traffic Anomalies, *SIGCOMM Comput. Commun. Rev.*
- Li, D., P. M. Robinson, and H. L. Shang (2020), Long-range dependent curve time series, *Journal of the American Statistical Association*, 115(530), 957–971.
- Lin, X.-X., P. Lin, and E.-H. Yeh (2020), Anomaly detection/prediction for the Internet of Things: State of the art and the future, *IEEE Network*, 35(1), 212–218.
- Lotz, M., M. B. McCoy, I. Nourdin, G. Peccati, and J. A. Tropp (2018), Concentration of the intrinsic volumes of a convex body, *arXiv preprint arXiv:1810.12412*.
- Lucas, J. M., et al. (1990), Exponentially Weighted Moving Average Control Schemes: Properties and Enhancements, *Technometrics*.
- Maitra, T., and S. Bhattacharya (2020), Increasing Domain Infill Asymptotics for Stochastic Differential Equations Driven by Fractional Brownian Motion, *arXiv preprint arXiv:2005.09577*.
- Matsuda, Y., and Y. Yajima (2009), Fourier analysis of irregularly spaced data on \mathbb{R}^d , *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 71(1), 191–217.
- Merit Network, Inc. (2022), ORION: Observatory for Cyber-Risk Insights and Outages of Networks, <https://www.merit.edu/initiatives/orion-network-telescope/>.
- Meyer, C. D. (2000), *Matrix analysis and applied linear algebra*, vol. 71, Siam.
- Mieth, B., et al. (2016), Combining multiple hypothesis testing with machine learning increases the statistical power of genome-wide association studies, *Scientific reports*, 6(1), 36,671.
- Moore, D., C. Shannon, D. J. Brown, G. M. Voelker, and S. Savage (), Inferring Internet Denial-of-service Activity, *ACM Trans. Comput. Syst.*
- Moore, D., C. Shannon, G. Voelker, and S. Savage (2004), Network telescopes: Technical report, *Tech. rep.*, Cooperative Association for Internet Data Analysis (CAIDA).

- Niranjana, R., V. A. Kumar, and S. Sheen (2020), Darknet traffic analysis and classification using numerical AGM and Mean Shift Clustering algorithm, *SN Computer Science*, 1(1), 1–10.
- Nishikaze, H., S. Ozawa, J. Kitazono, T. Ban, J. Nakazato, and J. Shimamura (2015), Large-scale monitoring for cyber attacks by using cluster information on darknet traffic features, *Procedia Computer Science*, 53, 175–182.
- Pace, R. K., R. Barry, and C. F. Sirmans (1998), Spatial statistics and real estate, *The Journal of Real Estate Finance and Economics*, 17, 5–13.
- Paganini, P. (2016), The hosting provider OVH continues to face massive DDoS attacks launched by a botnet composed at least of 150000 IoT devices., <http://securityaffairs.co/wordpress/51726/cyber-crime/ovh-hit-botnet-iot.html>.
- Panaretos, V. M., and S. Tavakoli (2013), Fourier analysis of stationary time series in function space, *The Annals of Statistics*, 41(2), 568–603.
- Pang, R., V. Yegneswaran, P. Barford, V. Paxson, and L. Peterson (2004), Characteristics of Internet Background Radiation, in *Proceedings of the 4th ACM SIGCOMM Conference on Internet Measurement*.
- Peccati, G., and M. S. Taqqu (2011), *Wiener Chaos: Moments, Cumulants and Diagrams: A survey with computer implementation*, vol. 1, Springer Science & Business Media.
- Percival, D. B., and A. T. Walden (2006), *Wavelet Methods for Time Series Analysis*, *Cambridge Series in Statistical and Probabilistic Mathematics*, vol. 4, xxvi+594 pp., Cambridge University Press, Cambridge, reprint of the 2000 original [MR1770693].
- Percival, D. B., and A. T. Walden (2020), *Spectral Analysis for Univariate Time Series*, *Cambridge Series in Statistical and Probabilistic Mathematics*, vol. 51, xxiv+691 pp., Cambridge University Press, Cambridge, doi:10.1017/9781139235723.
- Perlroth, N. (), Hackers Used New Weapons to Disrupt Major Websites Across U.S., <http://www.nytimes.com/2016/10/22/business/internet-problems-attack.html>.
- Pickands III, J. (1968), Moment convergence of sample extremes, *The Annals of Mathematical Statistics*, pp. 881–889.
- Politis, D. N. (2011), Higher-order accurate, positive semidefinite estimation of large-sample covariance and spectral density matrices, *Econometric Theory*, pp. 703–744.
- Ramírez, J., J. C. Segura, J. M. Górriz, and L. García (2007), Improved voice activity detection using contextual multiple hypothesis testing for robust speech recognition, *IEEE Transactions on Audio, Speech, and Language Processing*, 15(8), 2177–2189.

- Ramsay, J. O., and B. W. Silverman (2005), *Functional Data Analysis*, Springer.
- Rao, B. P. (2014), Characterization of gaussian distribution on a hilbert space from samples of random size, *Journal of Multivariate Analysis*, 132, 209–214.
- Resnick, S. I., and R. Tomkins (1973), Almost sure stability of maxima, *Journal of Applied Probability*, 10(2), 387–401.
- Robinson, P. (1983), Review of various approaches to power spectrum estimation, *Handbook of Statistics*, 3, 343–368.
- Roemmich, D., W. J. Gould, and J. Gilson (2012), 135 years of global ocean warming between the Challenger expedition and the Argo Programme, *Nature Climate Change*, 2(6), 425–428, doi:10.1038/nclimate1461.
- Rosenblatt, M. (1985), *Stationary Sequences and Random Fields*, Springer.
- Rosow, C. (2014), Amplification Hell: Revisiting Network Protocols for DDoS Abuse, in *Proceedings of the 2014 Network and Distributed System Security (NDSS) Symposium*.
- Rudin, W. (1991), *Functional analysis*, mcgrawhill, Inc, New York, 45, 46.
- Samarov, A. (1977), Lower bound on the risk for spectral density estimates, *Problemy Peredachi Informatsii*, 13(1), 67–72.
- Schuster, A. (1898), On the investigation of hidden periodicities with application to a supposed 26 day period of meteorological phenomena, *Terrestrial Magnetism*, 3(1), 13–41.
- Shaikh, F., E. Bou-Harb, J. Crichigno, and N. Ghani (2018), A Machine Learning Model for Classifying Unsolicited IoT Devices by Observing Network Telescopes, in *2018 14th International Wireless Communications and Mobile Computing Conference (IWCMC)*, pp. 938–943, doi:10.1109/IWCMC.2018.8450404.
- Shen, J., S. Stoev, and T. Hsing (2022), Tangent fields, intrinsic stationarity, and self similarity, *Electronic Journal of Probability*, 27, 1–56.
- Simon, B. (2015), *A Comprehensive Course in Analysis*, American Mathematical Society Providence, Rhode Island.
- Sun, L., R. V. Craiu, A. D. Paterson, and S. B. Bull (2006), Stratified false discovery control for large-scale hypothesis testing with application to genome-wide association studies, *Genetic Epidemiology: The Official Publication of the International Genetic Epidemiology Society*, 30(6), 519–530.
- Tanguy, K. (2015), Some superconcentration inequalities for extrema of stationary Gaussian processes, *Statistics & Probability Letters*, 106, 239–246.

- Tian, Y., W. B. Ayers, H. Sang, W. D. McCain Jr, and C. Ehlig-Economides (2018), Quantitative evaluation of key geological controls on regional Eagle Ford shale production using spatial statistics, *SPE Reservoir Evaluation & Engineering*, 21(02), 238–256.
- Tsay, R. S. (2013), *Multivariate Time Series Analysis: With R and Financial Applications*, John Wiley & Sons.
- Tsybakov, A. B. (2008), *Introduction to Nonparametric Estimation*, Springer Science & Business Media.
- Voronoi, G. (1908), Nouvelles applications des paramètres continus à la théorie des formes quadratiques. Premier mémoire. Sur quelques propriétés des formes quadratiques positives parfaites., *Journal für die reine und angewandte Mathematik*, 1908(133), 97 – 102.
- Wahba, G. (1990), *Spline Models for Observational Data*, *CBMS-NSF Regional Conference Series in Applied Mathematics*, vol. 59, xii+169 pp., Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, doi:10.1137/1.9781611970128.
- Willinger, W., V. Paxson, and R. Reidi (2001), Long-Range Dependence and Data Network Traffic. Long-range Dependence: Theory and Applications, P. Doukhan, G. Oppenheim and MS Taqqu, eds.
- Wustrow, E., M. Karir, M. Bailey, F. Jahanian, and G. Huston (2010), Internet Background Radiation Revisited, in *Proceedings of the 10th ACM SIGCOMM Conference on Internet Measurement*.
- Xie, K., X. Li, X. Wang, G. Xie, J. Wen, J. Cao, and D. Zhang (2017), Fast tensor factorization for accurate Internet anomaly detection, *IEEE/ACM transactions on networking*, 25(6), 3794–3807.
- Xu, H., C. Caramanis, and S. Sanghavi (2010), Robust PCA via outlier pursuit, in *Advances in Neural Information Processing Systems*.
- Yan, D.-M., W. Wang, B. Lévy, and Y. Liu (2013), Efficient computation of clipped Voronoi diagram for mesh generation, *Computer-Aided Design*, 45(4), 843–852.
- Yarger, D. (2020a), Shiny applications accompanying “A functional-data approach to the Argo data”, Available at <https://sites.google.com/a/umich.edu/argostatistics/home/fdapaper>.
- Yarger, D., S. Stoev, and T. Hsing (2022), A functional-data approach to the Argo data, *The Annals of Applied Statistics*, 16(1), 216–246, doi:10.1214/21-AOAS1477.
- Yu, Y., T. Wang, and R. J. Samworth (2015), A useful variant of the Davis-Kahan theorem for statisticians, *Biometrika*.

- Zakroum, M., J. François, I. Chrisment, and M. Ghogho (2022), Monitoring Network Telescopes and Inferring Anomalous Traffic Through the Prediction of Probing Rates, *IEEE Transactions on Network and Service Management*.
- Zhang, L., G. Ding, Q. Wu, and Z. Han (2017), Spectrum sensing under spectrum misuse behaviors: A multi-hypothesis test perspective, *IEEE Transactions on Information Forensics and Security*, 13(4), 993–1007.
- Zhou, Z., X. Li, J. Wright, E. Candes, and Y. Ma (2010), Stable principal component pursuit, in *Information Theory Proceedings (ISIT), 2010 IEEE International Symposium on*, pp. 1518–1522, IEEE.
- Zhu, T., and D. N. Politis (2020), Higher-order accurate spectral density estimation of functional time series, *Journal of Time Series Analysis*, 41(1), 3–20.