Local Structure of Random Fields - Properties and Inference

by

Jinqi Shen

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Statistics)
in the University of Michigan
2019

Doctoral Committee:

Professor Tailen Hsing, Chair
Associate Professor Veronica Berrocal
Associate Professor XuanLong Nguyen
Professor Stilian A. Stoev

Jinqi Shen

jqshen@umich.edu

ORCID iD: 0000-0002-8961-5743

© Jinqi Shen 2019

For my mom, who gave me all her love since I was a child.

ACKNOWLEDGEMENTS

I want to give my biggest thanks to my advisor Prof. Tailen Hsing, for the warm and endless support, for the numerous time we spent together towards challenging problems, and most importantly, for showing me the right attitude towards career. Five years is a short period, but I have learnt so much from him and will benefit from it for the rest of my life. I also want to give special thanks to Prof. Stilian Stoev for the tremendously helpful guidance in my research. I am extremely grateful to my parents Yuhuan and Jiangang, and my girlfriend, Qinghua, who accompany me across all the academic and personal difficulties. Words can hardly express how precious those supports are and how much I love them. I also want to thank Prof. XuanLong Nguyen and Prof. Veronica Berrocal for their valuable time serving in the committee and their useful comments on my research. Finally, I really appreciate all the supports from my friends and the department. Those wonderful memories will never be forgotten in my life!

TABLE OF CONTENTS

Dedication . . . . . **ii**

Acknowledgments . . . . . **iii**

List of Figures . . . . . **vi**

List of Tables . . . . . **viii**

Abstract . . . . . **ix**

Chapter

1 Introduction **1**

1.1 Spatial statistics and kriging 1
1.2 Universal kriging and intrinsic stationarity 3
1.3 Fractal and self-similar process 4
1.4 Local properties of random fields and tangent fields 6
1.5 Outline of this thesis 8

2 Tangent Field and Its Properties **10**

2.1 Introduction 10
2.2 Definitions and notations 11
2.2.1 The space Λ_k 11
2.2.2 Spaces $S(\Lambda_k, E)$ and $\widetilde{S}(\Lambda_k, E)$ 12
2.2.3 Tangent field 17
2.3 Self-similarity 20
2.4 Intrinsic stationarity 26
2.4.1 A tool based on the Egorov and Luzin Theorems 26
2.4.2 Extension of Falconer's Theorem for IRF$_k$ 29
2.5 Spectral theory for tangent fields 33
2.5.1 Notations 33
2.5.2 Spectral measures of stationary and IRF$_k$ random fields 35
2.5.3 Integral representations of stationary and IRF$_k$ random fields . . . 36
2.5.4 Characterization of tangent fields when $T_r(X) = r \cdot X$ 38
2.5.5 Characterization of tangent fields when $T_r(e_i) = r^{\alpha_i} \cdot e_i$ 46
2.6 Examples 48
2.7 Potential inference methods 50

2.7.1 Estimator for $T_r(X) = r \cdot X$ . . . 50
2.7.2 Estimator when $T_r(e_i) = r^{\alpha_i} \cdot e_i$ . . . 52
2.8 Appendix . . . 55
2.8.1 Integration with respect to operator-valued operators . . . 55
2.8.2 PROOF for Theorem 2.5.2 . . . 63
2.8.3 PROOF for Theorem 2.5.3 . . . 73
2.8.4 PROOF for Theorem 2.5.4 . . . 82
2.8.5 PROOF for Theorem 2.5.5 . . . 85
2.8.6 Sufficient conditions for the existence of tangent field for scalar Gaussian random field . . . 86

3 Hurst Function Estimation . . . 107

3.1 Introduction . . . 107
3.2 Basic formulation of the estimator . . . 110
3.3 Asymptotic properties . . . 113
3.4 Backfitting estimation when σ^2 is unknown . . . 119
3.5 Selection of q and b . . . 123
3.5.1 Selection of q . . . 123
3.5.2 Selection of b . . . 123
3.6 A simulation study . . . 125
3.7 Discussions and extensions . . . 128
3.7.1 Non-gridded data . . . 128
3.7.2 Non-constant variance . . . 130
3.8 Appendix . . . 131
3.8.1 Proof of Theorem 3.3.1 . . . 131
3.8.2 Proofs of Theorem 3.3.2 and Theorem 3.3.3 . . . 141
3.8.3 Proofs of Theorem 3.3.4 and Corollary 3.3.5 . . . 147
3.8.4 Proofs for Theorem 3.4.1, Theorem 3.4.2 and Theorem 3.4.3 . . . 150
3.8.5 Proof for Theorem 3.8.4 . . . 158
3.8.6 Proofs for Lemma 3.8.5, Lemma 3.8.6 and Lemma 3.8.7 . . . 164
3.8.7 Proofs for Section 3.7.1 . . . 176
3.8.8 Proofs for Section 3.7.2 and additional discussions for non-constant variance . . . 179

Bibliography . . . 185

LIST OF FIGURES

Figure

1.1 Simulated path of mBm on $[0, 1]$. The first figure is the simulated path, while the latter is the corresponding Hurst function. 6

2.1 Simulation of $\widetilde{X}_H(t)$ on $\mathbb{R}^2$. $H(t) = 0.5 + 0.4\sin(2\pi t_2)$ and $g_t(\theta) = 2.5 + 2\sin(2\theta - t_1\pi/2)$. 49

2.2 Comparison of $\widehat{C}_1$ and C_1. The first two plots are estimated and true C_1 separably. The third plot is the ratio (True Value)/(Estimated Value) and one can see most of the values fall around 1. 53

2.3 Comparisons of estimated and true $e_i, i = 1, 2$ after appropriate normalization. The black points are true values while the red are estimated values. 55

3.1 Plots of σ_H^2 for $q = 1, 2, 3, 4$. 124

3.2 Empirical quantiles of $\widehat{H}_1(t), \widehat{H}_2(t), \mathrm{QV}(t), \mathrm{IR}(t)$ with $n = 10,000$ based on 1,000 simulation runs; green curves are 5% and 95% pointwise empirical quantiles, respectively; blue curves are 25% and 75% empirical quantiles, respectively; black curves are empirical means. 126

3.3 Plot of $\log(\mathrm{MISE})$ versus $\log n$ 127

3.4 Empirical histograms of $\log\left(\sqrt{ISE}\right)$ based on $\widehat{H}$ with the oracle bandwidth and that selected by LSCV 128

3.5 Empirical quantiles of $\widehat{H}, \widehat{H}_1, \widehat{H}_2$ on the circle $(x - .5)^2 + (y - .5)^2 = 0.25^2$ based on 300 simulation runs; $\theta = \arctan((y - .5)/(x - .5))$; red curves are the true function; green curves are 5% and 95% pointwise empirical quantiles, respectively; blue curves are 25% and 75% empirical quantiles, respectively; black curves are empirical means. 129

3.6 Plot of $\log(\mathrm{MISE})$ versus $\log n$ 179

3.7 Plot of $\log(\mathrm{MISE})$ versus $\log n$ for $\widehat{\log \sigma^2}$ 183

3.8 Empirical quantiles of $\widehat{\log \sigma^2}$ 100 simulation runs; red curve is the true function; green curves are 5% and 95% pointwise empirical quantiles, respectively; blue curves are 25% and 75% empirical quantiles, respectively; black curves are empirical means. 184

3.9 Empirical quantiles of $\widehat{H}_1$ and $\widehat{H}_2$ with plugged in $\widehat{\log \sigma^2}(t)$ based on 100 simulation runs; red curves are true functions; green curves are 5% and 95% pointwise empirical quantiles, respectively; blue curves are 25% and 75% empirical quantiles, respectively; black curves are empirical means. 184

LIST OF TABLES

Table

3.1 $\sqrt{MISE}$ for all estimators with a range of sample sizes 127
3.2 Frequencies with which various bandwidths are picked by LSCV and oracle for $\widehat{H}$ based on $1,000$ simulation runs 128

ABSTRACT

Advances in data collection and computation tools popularize localized modeling on temporal or spatial data. Similar to the connection between derivatives and smooth functions, one approach to studying the local structure of a random field is to look at the tangent field, which is a stochastic random field obtained as a limit of suitably normalized increment of the random field at any fixed location. This thesis develops theories for tangent fields of any order and new statistical tools for their inference.

Our first project focuses on various properties of tangent fields. In particular, we show that tangent fields are self-similar and intrinsically stationary. Those two properties, along with the assumption of mean-square continuity, allow us to fully characterize a tangent field via a spectral representation, which provides a systematic way to obtain useful models. Our extension of the spectral theory to abstract spaces, including function spaces, can be of interest on its own. We also connect our theories with common models in spatial statistics including the Matérn model and its variations. Preliminary inference methods are proposed along with simulation studies.

An important example of random fields with tangent fields is the multifractional Brownian motion which has been studied extensively. Our second project focuses on a wide range of issues concerning the estimation of the Hurst function of a multifractional Brownian motion when the process is observed on a regular grid. A theoretical lower bound for the minimax risk of this inference problem is established for a wide class of smooth Hurst functions. We also propose nonparametric estimators and show they are rate optimal. Implementation issues including how to overcome the presence of a nuisance parameter and

choose the tuning parameter from data have also been included. An extensive numerical study is conducted to compare our approach with other approaches. Some explorations about non-grid observations and non-constant variances are also included.

CHAPTER 1

Introduction

1.1 Spatial statistics and kriging

Spatial statistics involves a variety of techniques studying entities with their geographic properties. Those techniques, many still in their early stages of development, have been widely used in different areas of science including epidemiology (e.g., disease mapping), biology (e.g., plant distributions), geology (e.g., mineral distribution), meteorology (e.g., maps of temperature and precipitation) and so on.

A central problem in spatial statistics is the estimation or prediction of a variable of interest over a domain based on values observed at a limited number of points. This can be considered as an interpolation problem from a deterministic viewpoint. In statistics, there is a probabilistic approach known as kriging, named by Matheron (1963a) in honor of Danie Krige, who is a South African statistician and mining engineer giving nice discussions about how regression can improve predictions in the mining industry in Krige (1951).

Kriging is closely related to regression analysis and can be considered as best linear unbiased predictions (BLUPs) with various assumptions of mean and covariance structures. As illustrated in Cressie (1990), Krige did not derive the theory of the spatial BLUP and the technique of kriging was formally introduced independently by Matheron (1962, 1963b) and Gandin (1963).

Specifically, let $\{X(t), t \in D \subset \mathbb{R}^d\}$ be the random function of interest and $m(t)$, $C(s,t)$ be the mean and covariance functions of $X(t)$, respectively. Assume we observe $\{X(t_i), i = 1, \ldots, n\}$ and our target is to predict $X(t_0)$, where $t_0 \in D$ is any location. Assume $m(t) = \mu$ is a known constant. Then the *simple* kriging predictor is defined to be a linear predictor, i.e.,

$$\sum_{i=1}^{n} \lambda_i X(t_i) + \lambda_0, \tag{1.1}$$

that minimizes the mean-square prediction error

$$\mathbb{E}\left(X(t_0) - \sum_{i=1}^{n} \lambda_i X(t_i) - \lambda_0\right)^2.$$

It can be shown that the solution of $(\lambda_0, \ldots, \lambda_n)$ is

$$(\lambda_1, \ldots, \lambda_n) = \mathbf{c}^T\mathbf{C}^{-1}, \;\; \lambda_0 = (1 - \mathbf{c}^T\mathbf{C}^{-1}\mathbf{1})\mu,$$

where $\mathbf{c} = (C(t_0, t_1), \ldots, C(t_0, t_n))^T$, $\mathbf{C} = (C(t_i, t_j))$ and $\mathbf{1} = (1, \ldots, 1)^T$ (see e.g., Cressie (1990)).

If we let $m(t) = \mu$ to be an unknown constant, then (1.1) is no longer a predictor. One should restrict the class of linear predictor to be

$$\sum_{i=1}^{n} \lambda_i X(t_i), \text{ with } \sum_{i=1}^{n} \lambda_i = 1, \tag{1.2}$$

where the constraint $\sum_{i=1}^{n} \lambda_i = 1$ is to guarantee the estimator is unbiased. Under the criterion of BLUP, by minimizing

$$\mathbb{E}\left(X(t_0) - \sum_{i=1}^{n} \lambda_i X(t_i)\right)^2 \tag{1.3}$$

subject to (1.2), we could get the optimal values (e.g., Cressie (1990)) of λ_i's to be

$$(\lambda_1, \ldots, \lambda_n) = (\mathbf{c} + (1 - \mathbf{c}^T\mathbf{C}^{-1}\mathbf{1})(\mathbf{1}^T\mathbf{C}^{-1}\mathbf{1}^T)^{-1}\mathbf{1})^T\mathbf{C}^{-1}, \tag{1.4}$$

and the corresponding predictor is known as the *ordinary* kriging predictor. It can be shown the implementation of ordinary kriging predictor does not require all information of the covariance function. In fact, $\mathbf{c}$ and $\mathbf{C}$ in (1.4) can be replaced by $\boldsymbol{\gamma} = (\gamma(t_0, t_1), \ldots, \gamma(t_0, t_n))^T$ and $\boldsymbol{\Gamma} = (\gamma(t_i, t_j))$, where $\gamma(t_i, t_j) := \mathbb{E}\,(X(t_i) - X(t_j))^2$, named as the *variogram* function. We refer to the ordinary kriging when we talk about kriging most of the time.

In order to consistently estimate the covariance structure based on limited observations, additional assumptions are needed. One such assumption is to assume stationarity in covariance, i.e., $C(s, t) = C(s-t)$ is a function that only depends on $s-t$. In the realm of ordinary kriging, people usually assume stationarity on the variogram, i.e., $\gamma(s, t) = \gamma(s-t)$, as variogram is the only information needed to implement ordinary kriging.

1.2 Universal kriging and intrinsic stationarity

In practice, it can be very restrictive to assume assumption of $\mu(t) \equiv \mu$, where μ is a constant. The *universal* kriging predictor is introduced in Matheron (1969) to relax this assumption.

In *universal* kriging, the mean function is assumed to be

$$m(x) = \sum_{l=0}^{L} a_l f^l(x),$$

where $\{f^l\}$ is a set of known basis functions and a_l's are fixed but unknown coefficients. Usually the first basis function f^0 is set to be the constant function identically equal to 1, which guarantees that the constant-mean case is included in the model. Universal kriging then modifies the constraint (1.2) to the set of $L+1$ conditions

$$\sum_{i=1}^{n} \lambda_i f^l(t_i) = f^l(t_0), l = 0, 1, \ldots, L,$$

and minimize (1.3) with this set of conditions. We omit the derivations for simplicity and one can refer to Chapter 3 in Chilès and Delfiner (2012) for details. One common choice of $\{f^l\}$ is to take $\{f^l\}$ as the basis of polynomial functions up to some order k and this leads to the so-called *intrinsic* kriging predictor (e.g., Chapter 4.6 of Chilès and Delfiner (2012)). Interestingly, intrinsic kriging could be implemented with even less information about the covariance structure than ordinary kriging. In fact, its implementation solely relies on the so-called *generalized covariance function* and one only needs to assume stationarity on the generalized covariance function.

Specifically, generalized covariance function is defined as follows. Let Λ be the class of signed measures on $\mathbb{R}^d$ supported on finitely many points, i.e.,

$$\lambda(dx) = \sum_{i=1}^{n} c_i \delta_{x_i}(dx), \tag{1.5}$$

where $c_i \in \mathbb{R}$, $x_i \in \mathbb{R}^d$, $i = 1, \ldots, n$, $\delta_a(dx)$ is the Dirac measure at a and x_i's are distinct. For any $k \in \mathbb{N}$, define $\Lambda_k \subset \Lambda$ to be the class of signed measures with finite supports that annihilate all polynomials of degree k, i.e. $f(\lambda) = 0$, where f is any polynomial function with degree no larger than k and $f(\lambda) = \sum_i c_i f(x_i)$. Notice that $\lambda \in \Lambda_k$ can be considered as a generalization of $(k+1)$-th order differencing. For any $w \in \mathbb{R}^d$ and $\lambda = \sum_i c_i \delta_{x_i}$, let $w + \lambda := \sum_i c_i \delta_{w+x_i}$.

Matheron (1973) considered the random mapping X from Λ_k to $\mathbb{R}$ satisfying $\{X(\lambda), \lambda \in \Lambda_k\} \stackrel{d}{=} \{X(w+\lambda), \lambda \in \Lambda_k\}$, for any $w \in \mathbb{R}^d$. He defined it as intrinsic random function of order k (IRF$_k$) and proved that every mean-square continuous IRF$_k$ $X(t)$ has a representation $\widetilde{X}(t) : \mathbb{R}^d \mapsto \mathbb{R}$ such that $\widetilde{X}(\lambda) \equiv \int \widetilde{X}(t)\lambda(dt) = X(\lambda), \lambda \in \Lambda_k$. If X is mean-square continuous, we say a continuous and symmetric function K on $\mathbb{R}^d$ is a generalized covariance function of X if

$$\text{Cov}(X(\lambda), X(\mu)) = \int \lambda(dx)K(x-y)\mu(dy), \text{ for } \lambda, \mu \in \Lambda_k.$$

Notice that the variogram is a special case of the generalized covariance with $k = 0$. Matheron (1973) claimed that generalized covariance function exists for every mean-square continuous IRF$_k$ and has a specific spectral measure (with no clear proof, but possibly in his note in French). Sasvári (2009) and Berschneider (2012) generalized this spectral theory to abstract spaces. More details about this spectral theory will be provided in Chapter 2 with variations and extensions.

1.3 Fractal and self-similar process

Since Benoit Mandelbrot coined the word 'fractal' in 1975, it has been widely used in studying the roughness and self-similarity in nature. (Interestingly, Mandelbrot and Matheron were both students of Paul Lévy.) The concept of fractal is difficult to define formally, but the key feature, self-similarity, can be understood with little mathematical background. Consider zooming in on a picture, which usually leads to finer, previously invisible new structures. However, if this is done on fractals, due to self-similarity, no new detail will appear and the same pattern will repeat over and over.

This property of self-similarity arises naturally when we study features that are unchanged with appropriate scalings, for example the local structure of certain random fields in this thesis. One may also equip stochastic processes with the property of self-similarity (c.f. Lamperti (1962a)) and we can define self-similar processes as follows.

Definition 1.3.1. *A stochastic process $\{X(t), t \geq 0\}$ is said to be self-similar if for any $a > 0$, there exists $b > 0$, s.t.*

$$\{X(at)\} \stackrel{d}{=} \{bX(t)\}.$$

Lamperti (1962a) shows if $X(t)$ is "nontrivial" and "stochastically continuous" at $t = 0$, then there exists a unique exponent $H \geq 0$ such that $b = a^H$. Mandelbrot and Wallis

(1968) named this exponent as "Hurst exponent" in honor of Harold Edwin Hurst, who studied the volatile rain and drought conditions of the Nile and developed the empirical rescaled range methodology.

We say a stochastic process $X(t)$ has stationary increments if the distribution of $\{X(t+h) - X(t)\}$ is independent of t. In the realm of Gaussian processes, an important self-similar stochastic process is fractional Brownian motion (fBm). It was first introduced by Kolmogorov (1940) and then recognized by Mandelbrot and Van Ness (1968) for its statistical applications. Specifically, a fBm with Hurst parameter $H \in (0, 1)$ is a Gaussian process $B_H(t), t \geq 0$ satisfying $B_H(t) = 0$ and

$$\text{Cov}\left(B_H(t), B_H(s)\right) = \frac{1}{2}\left(|s|^{2H} + |t|^{2H} - |s-t|^{2H}\right), s, t \geq 0.$$

Notice that the famous Brownian motion is a special case of fBm with $H = 0.5$. It can be proved that B_H is the only self-similar Gaussian process with stationary increments (e.g., Taqqu (1981)). The Hurst index is closely related to both sample-path smoothness and the long-memory pattern of the process. Due to the flexibility provided by the Hurst index, fBm has been widely applied in hydrology, financial mathematics, network analysis, to name a few.

In many circumstances, however, a more flexible model is desirable that allows sample path smoothness to vary with time or location while retains some of the other features of fBm. The multifractional Brownian motion (mBm), independently introduced in Lévy-Véhel and Peltier (1995) and Benassi et al. (1997), is such an example. There are different variations of the definition of mBm (e.g., Stoev and Taqqu (2006), Herbin (2006)), but in the one-dimensional case, they are all Gaussian processes satisfying $\mathbb{E}[B_H(t)] = 0$ and

$$\text{Cov}(t, s) = \mathscr{D}(H(t), H(s))\left(|t|^{H(t)+H(s)} + |s|^{H(t)+H(s)} - |s-t|^{H(t)+H(s)}\right), s, t \geq 0,$$

where $H(\cdot)$ is a function ranging in $(0, 1)$ and $\mathscr{D}(\cdot, \cdot)$ is some smooth function. The connection between sample-path smoothness and H can be seen in Fig 1.1, where a larger $H(t)$ corresponds to a higher degree of smoothness.

Due to the importance of mBm, a number of papers try to address the estimation of H and most of these estimators are formulated by leveraging the relationship between H and moments of generalized differences of ded data. With no intention to provide a complete list, we mention Coeurjolly (2005), Bardet and Surgailis (2013) and Bertrand et al. (2013). To our knowledge, Bardet and Surgailis (2013) has the most comprehensive results with extensions to processes similar to mBm in some circumstances. However, none of them illustrates what the optimal convergence rate is and how to incorporate the smoothness

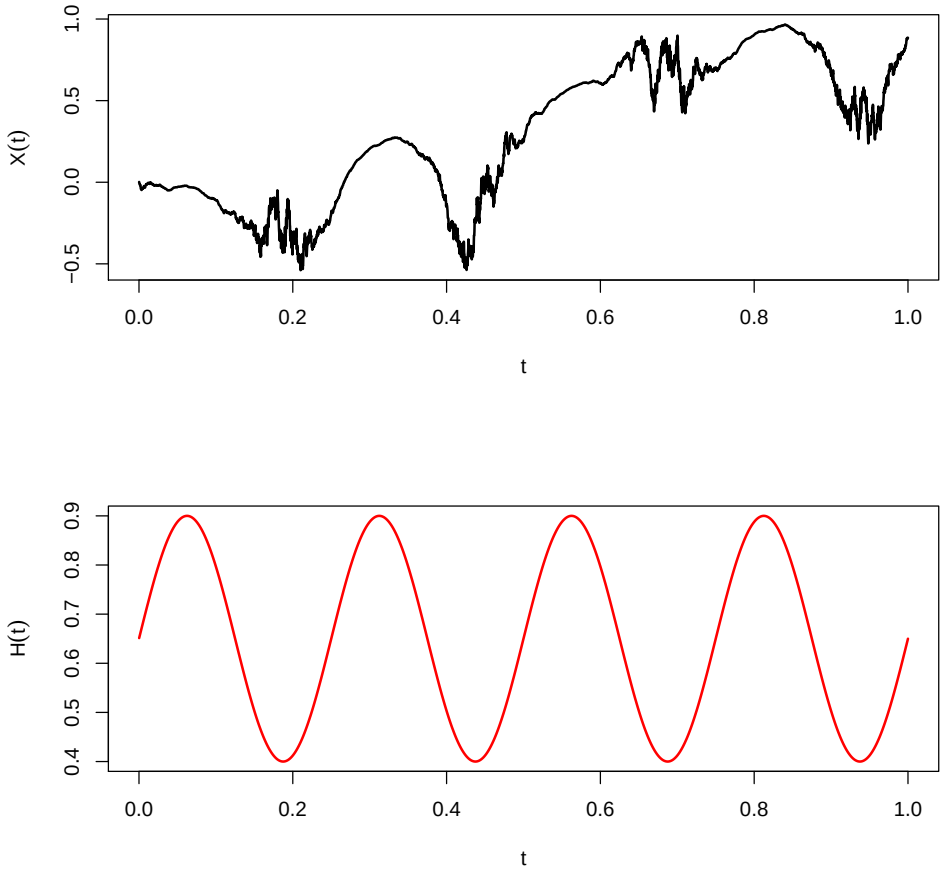


Figure 1.1: Simulated path of mBm on $[0, 1]$. The first figure is the simulated path, while the latter is the corresponding Hurst function.

property of H. We will try to address those problems with improved estimators in Chapter 3 of this thesis.

1.4 Local properties of random fields and tangent fields

The rapid development of modern technologies leads to the collection of spatial data in both larger scales and finer resolutions. For example, satellite-based climate and geospatial data could easily cover continents with high resolutions; see data available at NASAs Earth Science. Larger scales challenge the validity of the assumption of global stationarity and finer resolutions guarantee the amount of data for consistent modeling even in a relevantly small area. As a result, considerable progresses have been made on non-stationary model-

ing of spatial data. For instance, there is a large literature on methods based on the notion of deformation in Sampson and Guttorp (1992) and Anderes and Stein (2008), where non-stationarity is modeled by deforming the domain of a stationary process. Other common non-stationary models include convolutions of deterministic kernels with stationary processes (e.g., Higdon (1998), Higdon et al. (1999) and Fuentes (2002)) and uses of SPDEs with non-constant coefficients (e.g., Fuglstad et al. (2015) and Bolin (2014)).

For dense, gridded spatial data, a general framework for nonstationarity termed local intrinsic stationarity is developed in Hsing et al. (2016). Both mBm and deformation processes fall in this framework under certain conditions. Intuitively, this is more or less equivalent to assuming, mean function to be local polynomial function, which is reasonable due to Taylor's theorem, and generalized covariance function to be approximately stationary in each small neighborhood.

There are two high-level questions for this type of assumption:

(i) How general is this assumption? Is it possible to justify it intuitively?

(ii) Are there any other reasonable assumptions we could add to help inference?

In order to answers those questions, one needs to study the local structures of random fields and verify if local intrinsic stationarity is a common phenomenon. One approach in this direction is to consider the *tangent field* introduced in Falconer (2002, 2003). In Falconer (2002), $Y_z(t)$ is said to be a tangent field of $X(t)$ at location $z \in \mathbb{R}^d$ if there exist $c_n, r_n \to 0$ such that

$$\{c_n^{-1}(X(z + r_n t) - X(z)), t \in \mathbb{R}^d\} \xrightarrow{d} \{Y_z(t), t \in \mathbb{R}^d\},$$

where the convergence of functions is under the topology of uniform convergence on compact sets. Intuitively, tangent fields depict local features of stochastic processes at some location with diminishing neighborhoods. They proved that if $X(t)$ has tangent fields for all $z \in \mathbb{R}^d$, then

(i) Y_z is self-similar for all $z \in \mathbb{R}^d$, and

(ii) $Y_z(t + w) - Y_z(w) \stackrel{d}{=} Y_z(t), w \in \mathbb{R}^d$ for almost all $z \in \mathbb{R}^d$.

Notice that if there exists Y_z' such that $Y_z(t) = Y_z'(t) - Y_z'(0)$, then the second property is equivalent to Y_z' is IRF_0. Inspired by their work, we develop theories for tangent field of any order in Chapter 2, which characterize the local structure of random fields with mild conditions.

1.5 Outline of this thesis

Considering the length of this thesis, it may be helpful to provide an overview of main results and contributions.

The Chapter 2 of this thesis explores and characterizes local structures of random fields in a broad setting. Assuming on some Borel set $B \subset \mathbb{R}^d$, the stochastic random field $\{X(t), t \in \mathbb{R}^d\}$ has kth-order tangent fields at every location $t \in B$, i.e.,

$$\{c(z,r)X(t+r\lambda), \lambda \in \Lambda_k\} \xrightarrow{d} \{Y_z(\lambda), \lambda \in \Lambda_k\}, \text{ with } r \to 0,$$

where $c(z,r)$ is measurable as a function of z for any fixed r and $\{Y_z\}$ is non-trivial, we prove that for every $z \in B$, $\{Y_z\}$ is self-similar on Λ_k, i.e.

$$\{Y_z(r\lambda)\} \overset{d}{=} \{r^{\alpha(z)}Y_z(\lambda)\},$$

where $\alpha(z) > 0$; and for almost every $z \in B$, $\{Y_z\}$ is IRF$_k$, i.e.

$$\{Y_z(\lambda + w)\} \overset{d}{=} \{Y_z(\lambda)\}.$$

Therefore, the assumption of local intrinsic stationarity follows readily and we may even assume self-similarity locally, which aligns with the covariance structure studied in Hsing et al. (2016).

Based on those two properties and the characterization of IRF$_k$ due to Gel'fand and Vilenkin (1964) and Matheron (1973), we provide a concrete formula for spectral measures of tangent fields, involving only the local self-similarity exponent α and the local spectral measure σ. It is worth noting that those results are proved for $\{X(t)\}$ with image in a separable Hilbert space and readily to be applied in the domain of functional data analysis for functional-valued observations indexed by temporal or spatial locations. We also provide preliminary methods for estimating (α, σ) with dense, gridded data, although a close study of those estimators is still not available.

While the inference of tangent fields in general is still mysterious, in Chapter 3, we obtain a thorough study for estimating the self-similar exponent or the Hurst function H in mBm, leveraging the smoothness properties of H. Notice that mBm has its own tangent fields with mild conditions, and our study on mBm may be inspiring for inference in general settings.

Let $\lfloor x \rfloor$ be the largest integer no larger than x. For open set $B \in \mathbb{R}$ and constants $p \geq 0$ and $M \in (0, \infty)$, define $\mathscr{H}_p(B, M)$ as the space of $\lfloor p \rfloor$-times differentiable functions

$f : B \mapsto \mathbb{R}$ such that $f^{(\lfloor p \rfloor)}$ is Hölder continuous on B with $|f^{(\lfloor p \rfloor)}(x) - f^{(\lfloor p \rfloor)}(y)| \leq M|x-y|^{p-\lfloor p \rfloor}$ for all $x, y \in B$. Assume $H \in \mathscr{H}_p([0,1])$ for some $p > 1$, which depicts the smoothness of H. For simplicity, let $X(t)$ be a mBm on $[0,1]$ and we have gridded observations with interval $1/n, n \in \mathbb{N}$. We first prove that any estimator $\widehat{H}$ based on this gridded data has a lower bound $\left(n \log^2 n\right)^{-\frac{p}{2p+1}}$ on convergence rate (in L_q sense, where $q \in [1, \infty)$). To our knowledge, this is the first time such a lower bound is developed in the mBm context. We then formulate the nonparametric estimation of H based on properties of tangent fields and local polynomial methods (cf. Fan and Gijbels (1996)). Thorough asymptotic theories for our estimators under different scenarios are provided and the rate of our estimator matches the lower bound with appropriate parameters tuning, which makes it rate optimal. Data-driven bandwidth selection is also provided which is important for the implementation of those procedures. We will also discuss various extensions with non-grid observations and non-constant variances, and the minimax rate is still achievable under certain conditions. A numerical study is conducted to illustrate those results and compare with existing approaches.

CHAPTER 2

Tangent Field and Its Properties

In this chapter, we develop theory for tangent fields with any order. We prove tangent field is self-similar and IRF$_k$ and then characterize its spectral measure in general settings. Examples and preliminary inferences will also be given.

2.1 Introduction

The tangent process of a random field is the stochastic process obtained in the limit of suitably normalized increment of the random field at a fixed location. Falconer (2002, 2003) discovered a remarkable phenomenon about the structure of the tangent process. Broadly speaking, Falconer proved that the tangent process must be self-similar and with stationary increments. The self-similarity of the tangent field is not surprising and it is akin to many limit theorems such as Lamperti's seminal work Lamperti (1962b) (see also Davydov and Paulauskas (2017)) and a host of results on (univariate and multivariate) regular variation Gnedenko (1943); Meerschaert (1984); Hult and Lindskog (2006).

However, the property of stationary increments for the tangent process is unexpected. Naturally, the proof of this property in Falconer's works is quite challenging and involves delicate Lebesgue-density arguments. Motivated to understand this local asymptotic stationarity of increments phenomenon, here we provide a new proof based on the Lusin and Egorov theorems. Our proof generalizes the theory to k-th order tangent processes, arising naturally in the context of intrinsic random functions.

Broadly speaking, we establish that the k-th order tangent process must be self-similar and intrinsically stationary with Falconer's results corresponding to the case $k = 0$. These results show that continuous random fields with non-trivial k-th order tangent fields, behave locally like a self-similar k-th order intrinsic random function (IRF$_k$). The well-known characterization of IRF$_k$ due to Gel'fand and Vilenkin (1964) and popularized by Matheron (1973) (see also Chilès and Delfiner (2012)) entails a concrete formula for *all possible* "lo-

cal" generalized covariance functions, which involve the local self-similarity exponent α and the local *spectral measure* σ. These results identify the (infinite-dimensional) parameters (α, σ) as the *canonical* local parameters to estimate in the context of in-fill asymptotics for general mean-square continuous random fields. They also provide guidance for building flexible random field models with desired local properties. Some similar studies are provided in Dobrushin (1979) about the characterization of self-similar translation-invariant generalized random fields on $\mathscr{S}'(\mathbb{R}^d)$, where $\mathscr{S}'(\mathbb{R}^d)$ is the topological dual of Schwartz space $\mathscr{S}(\mathbb{R}^d)$ of all infinitely differentiable rapidly decreasing real functions on $\mathbb{R}^d$. Our study, on the other hand, considers random fields on measures with finite support, which is a subclass of their generalized random fields, but provides more detailed results.

The characterization of tangent fields can also be extended to separable Hilbert spaces and applied in functional data analysis. Our theory for functional-valued stationary or IRF$_k$ random fields is more or less overlapped with previous work. With no intention to provide a full complete list, we refer to Bochner (1948) and Khintchine (1934) for Bochner's theorem and Neeb (1998) for extensions to general spaces; Cramér (1942) for the spectral representation of stationary random field; Matheron (1973), Sasvári (2009) for the existence of general covariance of IRF$_k$ and its integral representation; Berschneider (2012) for the integral representation of IRF$_k$ in an abstract space. A comprehensive summary can be found in Berschneider and Sasvári (2018).

The rest of this chapter is organized as follows: In Section 2.2, we will introduce notations, topologies and a formal definition of tangent fields. Self-similarity and intrinsic stationarity of tangent field will be proved in Section 2.3 and Section 2.4 respectively. Section 2.5 will contain a refined spectral theory for tangent field with its unique representation in Gaussian. Two examples will be given in Section 2.6 to illustrate the connection of our theory with common spatial models. Finally, two preliminary inference methods are provided in Section 2.7 based on the our spectral theory and the periodogram kernel method in Panaretos and Tavakoli (2013).

2.2 Definitions and notations

2.2.1 The space Λ_k

Let Λ be the class of signed measures on $\mathbb{R}^d$ supported on finitely many points, i.e.,

$$\lambda(dx) = \sum_{i=1}^{n} c_i \delta_{x_i}(dx), \tag{2.1}$$

where $c_i \in \mathbb{R}$, $x_i \in \mathbb{R}^d$, $i = 1, \ldots, n$, $\delta_a(dx)$ is the Dirac measure at a and x_i's are distinct.

Introduce the following operations on the set of measures Λ:

(i) Scaling: For $r > 0$, define $r \cdot \lambda(dx) := \lambda(dx/r)$.

(ii) Translation: For $w \in \mathbb{R}^d$, define $(w + \lambda)(dx) := \lambda(dx - w)$.

Definition 2.2.1. *For any $k \in \mathbb{N}$, define $\Lambda^{(k)} \subset \Lambda$ to be the class of signed measures supported on finitely many points that annihilate all polynomials of degree k, i.e. $f(\lambda) = 0$, where f is any polynomial function with degree no larger than k.*

Since the vector space of polynomial functions is invariant to both scaling and translation, the set $\Lambda^{(k)}$ is closed with respect to these two operations. Observe also that $\Lambda_{k+1} \subset \Lambda_k$ for any $k \geq 0$.

2.2.2 Spaces $S(\Lambda_k, E)$ and $\widetilde{S}(\Lambda_k, E)$

Let (E, d) be a linear separable complete metric space (lsms) over $\mathbb{R}$ or $\mathbb{C}$, that is, a linear space where the scalar multiplication and addition are continuous with respect to the metric. Typical examples for (E, d) are the Euclidean space $\mathbb{R}^d$, or a Banach space, like the space $C([0, 1])$ of continuous functions equipped with the sup-norm, or $L^2(\mathbb{R})$. For any function f from $\mathbb{R}^d$ to E, let $f(\lambda) = \int f d\lambda$ whenever there is no ambiguity.

Remark. *By Birkhoff-Kakutani Theorem, without loss of generality, we can assume the metric d is translation invariant, that is, $d(a + x, a + y) = d(x, y)$ for any $a, x, y \in E$. By continuity w.r.t. scalar multiplication, we have $\lim_{n\to\infty} d(cx_n, cx) = 0$ for any $c \in \mathbb{R}$ or $\mathbb{C}$ and $x_n, x \in E$ with $\lim_{n\to\infty} d(x_n, x) = 0$. In this chapter, we additionally assume*

[M] For any $K > 0$,

$$\lim_{\delta\to 0} \sup_{|c|\leq K, d(x,0)<\delta} d(cx, 0) \to 0. \tag{2.2}$$

There exists example such that (2.2) does not hold. For example, consider a separable Hilbert space $\mathbb{H}$ with $\{e_i\}$ to be its complete orthogonal system (CONS). For any $x = \sum_i c_i e_i$, define

$$d(x, 0) = \sum_i |c_i|^{a_i},$$

where $a_i \in (0, 1)$ and $a_i \to 0$ and let $E = \{x \in \mathbb{H}, d(x, 0) < \infty\}$.

Consider the space S of linear functions $\xi(\lambda)$ that maps from Λ_k to E, where linearity means $\xi(\lambda_1 + \lambda_2) = \xi(\lambda_1) + \xi(\lambda_2), \lambda_1, \lambda_2 \in \Lambda_k$. The first question is finding a representation for these functions. Pick arbitrary points $t_1, \ldots, t_{M_k}$ such that the matrix $B_{M_k \times M_k} = (\boldsymbol{t}_1, \ldots, \boldsymbol{t}_{M_k})$ is of full rank where $\boldsymbol{t} = (t^{\boldsymbol{\alpha}}, \|\boldsymbol{\alpha}\|_1 \leq k)^T$ and $M_k = \binom{k+d}{k}$. For $t_i \in \mathbb{R}$, B has full rank so long as t_i's are distinct but more care is needed for $\mathbb{R}^d, d > 1$.

For each $t \in \mathbb{R}^d$, define

$$\lambda_t = \delta_t - (\delta_{t_1}, \ldots, \delta_{t_{M_k}}) B^{-1} \boldsymbol{t}. \tag{2.3}$$

It follows that

$$\left(\int u^{\boldsymbol{\alpha}} d\lambda_t(u), \|\boldsymbol{\alpha}\|_1 \leq k \right)^T = (I - BB^{-1})\boldsymbol{t} = 0,$$

and so λ_t annihilates all k-th degree polynomials and is therefore a member of Λ_k. For any function f, define

$$\widetilde{f}(t) := \int f d\lambda_t = f(t) - (f(t_1), \ldots, f(t_{M_k}))) B^{-1} \boldsymbol{t}. \tag{2.4}$$

Then

$$\left(\widetilde{f}(t_1), \ldots, \widetilde{f}(t_{M_k}) \right) = (f(t_1), \ldots, f(t_{M_k}))(I - B^{-1}B) = 0. \tag{2.5}$$

Thus, $\widetilde{f}$ is a "pinned-down" version of f. Conversely, if f is already pinned down, i.e., $f(t_i) = 0, 1 \leq i \leq M_k$, then $\widetilde{f} = f$. In particular,

$$f(\lambda_t) = \widetilde{f}(t) \tag{2.6}$$

and so we can think of λ_t as the evaluation functional of f. Thus, the precise class of the pinned-down function is

$$\widetilde{S} = \{\widetilde{f} : \widetilde{f}(t_i) = 0, i = 1 \ldots, t_{M_k}\}.$$

Next, we establish a one-to-one correspondence between S and $\widetilde{S}$. Consider mappings $\mathscr{J}$ from S to $\widetilde{S}$ and $\mathscr{K}$ from $\widetilde{S}$ to S defined by

$$(\mathscr{J} f)(t) = f(\lambda_t), \quad (\mathscr{K} \widetilde{f})(\lambda) = \int \widetilde{f} d\lambda. \tag{2.7}$$

Proposition 2.2.2. *$\mathscr{J}$ and $\mathscr{K}$ are inverses of each other.*

Proof. By (2.7) and (2.6),

$$\mathscr{J}(\mathscr{K}\widetilde{f})(t) = \widetilde{f}(t).$$

Conversely, supposing $\lambda = \sum_{i=1}^{m} c_i \delta_{x_i}$, by the linearity of f, we would have

$$\mathscr{K}(\mathscr{J}f)(\lambda) = \int f(\lambda_t)\lambda(dt) = f\left(\sum_{i=1}^{m} c_i \lambda_{x_i}\right).$$

By (2.3), write

$$\lambda_{x_i} = \delta_{x_i} - [\delta_{t_1}, \ldots, \delta_{t_{M_k}}]B^{-1}\boldsymbol{x}_i$$

where $\boldsymbol{x}_i = (x_i^{\boldsymbol{\alpha}}, \|\boldsymbol{\alpha}\|_1 \leq k)^T$. Since $\lambda \in \Lambda_k$,

$$\sum_{i=1}^{m} c_i \lambda_{x_i} = \lambda - [\delta_{t_1}, \ldots, \delta_{t_{M_k}}]B^{-1}\sum_{i=1}^{m} c_i \boldsymbol{x}_i = \lambda. \tag{2.8}$$

This shows that

$$\mathscr{K}(\mathscr{J}f)(\lambda) = f(\lambda)$$

and concludes the proof. □

The isomorphism in Proposition 2.2.2 provides a natural link between "abstract" and "ordinary" IRFs. The following lemma shows the relationship between representations among two different sets of $\{\lambda_1^{(i)}, \ldots, \lambda_{M_k}^{(i)}\}, i = 1, 2$.

Lemma 2.2.3. *For $i = 1, 2$, denote $B^{(i)} = (\boldsymbol{t}_1^{(i)}, \ldots, \boldsymbol{t}_{M_k}^{(i)})$, $\lambda_t^{(i)} = \delta_t - (\delta_{t_1^{(i)}}, \ldots, \delta_{t_{M_k}^{(i)}})(B^{(i)})^{-1}\boldsymbol{t}$, and $\widetilde{f}^{(i)}(t) = f(\lambda_t^{(i)})$. If $B^{(i)}, i = 1, 2$ both have full rank, we have*

$$\widetilde{f}^{(1)}(t) = \widetilde{f}^{(2)}(t) - (\widetilde{f}^{(2)}(t_1^{(1)}), \ldots, \widetilde{f}^{(2)}(t_{M_k}^{(1)}))(B^{(1)})^{-1}\boldsymbol{t}.$$

Proof. By (2.8), for any $\lambda = \sum_{j=1}^{m} c_j \delta_{x_j} \in \Lambda_k$, we have

$$\lambda = \sum_{j=1}^{m} c_j \lambda_{x_j}^{(i)}, i = 1, 2.$$

Specially,

$$\lambda_t^{(1)} = \delta_t - (\delta_{t_1^{(1)}}, \ldots, \delta_{t_{M_k}^{(1)}})(B^{(1)})^{-1}\boldsymbol{t} = \lambda_t^{(2)} - (\lambda_{t_1^{(1)}}^{(2)}, \ldots, \lambda_{t_{M_k}^{(1)}}^{(2)})(B^{(1)})^{-1}\boldsymbol{t}.$$

Therefore,

$$\begin{aligned} \widetilde{f}^{(1)}(t) &= f(\lambda_t^{(1)}) = f(\lambda_t^{(2)} - (\lambda_{t_1^{(1)}}^{(2)}, \ldots, \lambda_{t_{M_k}^{(1)}}^{(2)})(B^{(1)})^{-1}\boldsymbol{t} \\ &= \widetilde{f}^{(2)}(t) - (\widetilde{f}^{(2)}(t_1^{(1)}), \ldots, \widetilde{f}^{(2)}(t_{M_k}^{(1)}))(B^{(1)})^{-1}\boldsymbol{t}. \end{aligned}$$

□

For our purpose, we focus on continuous functions $f \in S$ in the sense that $f(\lambda_{t_n}) \to f(\lambda_t)$ whenever $t_n \to t$ for all t, or, equivalently, $\widetilde{f} \in \widetilde{S}$ such that $\widetilde{f}(t_n) \to \widetilde{f}(t)$ whenever $t_n \to t$ for all t. By Proposition 2.2.2, it suffices to consider $\widetilde{S}$ and denote this class of continuous functions $\widetilde{f}$ as

$$\widetilde{S}(\Lambda_k, E) = \{\widetilde{f} : \widetilde{f} \text{ is continuous with } \widetilde{f}(t_i) = 0, i = 1 \ldots, t_{M_k}\}.$$

The corresponding space $\mathscr{J}^{-1}(\widetilde{S}(\Lambda_k, E))$ will be denoted by $S(\Lambda_k, E)$.

A possible metric on $\widetilde{S}(\Lambda_k, E)$ is

$$\rho(\widetilde{f}, \widetilde{g}) = \sum_{i \geq 1} 2^{-i} \left(1 - \exp\left\{- \sup_{\|t\|_2 \leq i} d\left(\widetilde{f}(t), \widetilde{g}(t)\right)\right\}\right), \tag{2.9}$$

which induces local uniform convergence. It is easy to verify that the induced topology is complete and separable. Using $\mathscr{J}$, we define the corresponding metric on $S(\Lambda_k, E)$, also denoted as ρ for convenience. As such, $\mathscr{J}$ is an isometric isomorphism between the spaces. By Lemma 2.2.3 and [M], one can easily verify that those topologies induced by different sets of $\{t_i\}$ are equivalent.

One may also consider the topology on $S(\Lambda_k, E)$ defined by (2.9) as a factor topology or quotient topology induced from $(C(\mathbb{R}^d, E), \rho)$, where $C(\mathbb{R}^d, E)$ is the set of all continuous mapping from $\mathbb{R}^d$ to E. For $f, g \in C(\mathbb{R}^d, E)$, f, g is said to belong to the same equivalent class if and only if $f - g$ is a polynomial function with order no larger than k and we denote it as $f \sim g$. Notice that each equivalent class defined in this way will contain a representation in $\widetilde{S}(\Lambda_k, E)$ for any fixed $\{t_k\}$ satisfying the full-rank condition for $B_{M_k \times M_k}$. The map from $(C(\mathbb{R}^d, E), \rho)$ to $(\widetilde{S}(\Lambda_k, E), \rho)$ can be easily verified to be surjective, continuous and open and thus a quotient map.

The following proposition characterizes weak convergence on $(S(\Lambda_k, E), \rho)$.

Proposition 2.2.4. *Let $X_n, X : (\Omega, \mathscr{F}, P) \to S(\Lambda_k, E)$ and denote $\widetilde{X}_n = \mathscr{J}(X_n)$ and $\widetilde{X} = \mathscr{J}(X)$. Then $X_n \xrightarrow{d} X$ in $(S(\Lambda_k, E), \rho)$, or equivalently $\widetilde{X}_n \xrightarrow{d} \widetilde{X}$ in $(\widetilde{S}(\Lambda_k, E), \rho)$, if and only if*

(i) For any $m > 0$ and $u_1, \ldots, u_m$,

$$(\widetilde{X}_n(u_1), \ldots, \widetilde{X}_n(u_m)) \xrightarrow{d} (\widetilde{X}(u_1), \ldots, \widetilde{X}(u_m)).$$

(ii) For any compact set $K \subset \mathbb{R}^d$, $\widetilde{X}_n$ is strongly stochastically equicontinuous on K, that is for any $\eta, \epsilon > 0$, there exists $\delta > 0$, s.t.

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{\|t-u\|_2<\delta, t,u\in K} d(\widetilde{X}_n(t), \widetilde{X}_n(u)) > \eta \right) < \epsilon.$$

Remark: By Lemma 2.2.3 and [M], one can easily verify that both conditions (i) and (ii) are not related to the choice of $\{t_1, \ldots, t_{M_k}\}$.

Proof. Since convergence in $(\widetilde{S}(\Lambda_k, E), \rho)$ corresponds to the local uniform convergence, the key ingredients for showing necessity will be exactly the same for those used in showing weak convergence in the continuous function space, and should be a simple exercise of the almost sure convergence representation. The proof of sufficiency will also be similar. The following is a sketch with details omitted.

For any $u > 0$, denote

$$\widetilde{X}_{u,n}(t) = \widetilde{X}_n(t) I(|t| < u) + \widetilde{X}_n(u) I(|t| \geq u)$$

and

$$\widetilde{X}_u(t) = \widetilde{X}(t) I(|t| < u) + \widetilde{X}(u) I(|t| \geq u).$$

With $u \to \infty$, as $\rho(\widetilde{X}_u, \widetilde{X}) \xrightarrow{a.s.} 0$, $\sup_n \rho(\widetilde{X}_{u,n}, \widetilde{X}_n) \xrightarrow{a.s.} 0$ and Theorem 3.2 in Billingsley (1999), we only need to prove for every fixed $u > 0$, $\widetilde{X}_{u,n} \xrightarrow{d} \widetilde{X}_u$ as $n \to \infty$.

By an extension of Arzelà-Ascoli theorem (cf. Theorem 7.17 in Kelley (1955)), one can easily obtain the following lemma.

Lemma 2.2.5. *For any set $A \subset \widetilde{S}(\Lambda_k, E)$, A is compact if and only if*

(i) A is closed under the topology induced by (2.9)*.*

(ii) For any $t \in \mathbb{R}^d$, $\sup_{\widetilde{f}\in A} d(\widetilde{f}(t), 0) < \infty$ is bounded.

(iii) $\lim_{\delta\to 0}\sup_{\widetilde{f}\in A} w(\widetilde{f},\delta)\to 0$, *where* $w(\widetilde{f},\delta)=\sup_{\|s-t\|_2\le\delta} d(\widetilde{f}(s),\widetilde{f}(t))$.

Denote $P_{n,u}$ as the law of $X_{n,u}$. By this lemma and the same argument used in the proof of Theorem 7.3 of Billingsley (1999), one can prove $\{P_{n,u}\}$ is tight and thus relatively compact. Combined with condition (i) and Example 5.1 of Billingsley (1999), we can prove $\widetilde{X}_{n,u}\xrightarrow{d}\widetilde{X}_u$ and this completes our proof.

□

2.2.3 Tangent field

Suppose that $X(z):\mathbb{R}^d\mapsto E$ is a random field with continuous paths.

For $\lambda\in\Lambda$, we write

$$X(z,\lambda)=\int X(z+x)\lambda(dx), \tag{2.10}$$

and treat $X(z,\cdot)=\{X(z,\lambda),\ \lambda\in\Lambda\}$ as a random field in λ.

For simplicity, we focus on the class $\Lambda^{(k)}$, and then $X(z,\lambda)$ can be viewed as the generalized $(k+1)$-th order increment of X at z. We are interested in the local asymptotic behavior of these generalized increments, i.e.,

$$X(z,r\cdot\lambda)=\sum_{i=1}^{m}c_iX(z+rx_i)\text{ as }r\downarrow 0.$$

We define the scaling action on the space (E,d) as below.

Definition 2.2.6. *A family of operators $T_a:E\to E$, indexed by the multiplicative group $\mathbb{R}_+\equiv(0,\infty)$ is said to be a* scaling action *on E, if the following conditions hold:*

(i) *For all $a_1>0$ and $a_2>0$, we have $T_{a_1}\circ T_{a_2}=T_{a_1a_2}$.*

(ii) $T_1=\mathrm{id}$ *is the identity, and $T_a(0)=0$, for all $a>0$.*

(iii) $\{T_a\}$ *is continuous, i.e., $d(T_{a_n}(x_n),T_a(x))\to 0$, whenever $a_n\to a>0$ and $d(x_n,x)\to 0$.*

(iv) $\{T_a\}$ *is radially monotone, i.e., $d(T_{a_1}(x),0)<d(T_{a_2}(x),0)$, for all $0\neq x\in E$ and $0<a_1<a_2$.*

(v) *For all balls $B_r:=\{x\in E:\ d(x,0)<r\},\ r>0$ centered at the origin, we have*

$$\bigcup_{n=1}^{\infty}T_n(B_r)=E. \tag{2.11}$$

Remark. *Property (v) in Definition 2.2.6 can be replaced by the equivalent condition of continuity at* 0*. That is,* $d(T_a(x), 0) \to 0$ *as* $a \downarrow 0$*, for all* $x \in E$*. Indeed, if (v) holds, by the radial monotonicity property, to prove continuity at* 0*, it is enough to verify that* $d(T_{1/n}(x), 0) \to 0$*, for all* $x \in E$*. For every* $r := \epsilon > 0$*, however, Relation* (2.11) *implies that* $T_n(B_\epsilon) \ni x$*, for all sufficiently large* n*. This is equivalent to* $d(T_{1/n}(x), 0) < \epsilon$*, for all sufficiently large* n*. The converse argument showing that the continuity at* 0 *implies (v) is similar.*

Observe that most examples of rescaling operations of function spaces readily satisfy the above conditions. For example, if E is a normed space, a natural scaling action is the scalar multiplication:

$$T_a(x) := a \cdot x.$$

More generally, the scalar multiplication is a scaling action if the underlying metric is homogeneous, e.g., $d(a \cdot x, a \cdot y) = a^\gamma d(x, y),\ \gamma > 0$. Notice that the metric and the action need not be compatible, that is, $d(T_a x, 0)$ is in general not equal to $ad(x, 0)$ and therefore, $T_a B_r$ is in general not B_{ar}. Nevertheless, the radial monotonicity property implies that

$$T_{a_1}(B_r) \subset T_{a_2}(B_r), \quad \text{for all } 0 < a_1 < a_2. \tag{2.12}$$

Lastly, observe that since $T_a = T_{1/a}^{-1}$, we have that the mappings T_a are homeomorphisms and hence T_a maps open (closed) sets to open (closed) sets.

Given a scaling action $\{T_a,\ a \in \mathbb{R}_+\}$ on E, it is natural to consider its coordinate-wise extension on the space of E-valued functions S. Namely, the action $\widetilde{T}_a : S \to S$ is defined such that for all $f \in S$, and any $\lambda \in \Lambda_k$

$$\widetilde{T}_a(f)(\lambda) = T_a(f(\lambda)). \tag{2.13}$$

The following result shows that the coordinate-wise action is in fact a scaling action on $S(\Lambda_k, E)$.

Lemma 2.2.7. *For any scaling action* $\{T_a,\ a \in \mathbb{R}_+\}$ *on* (E, d)*, the coordinate-wise action* $\{\widetilde{T}_a,\ a \in \mathbb{R}_+\}$ *in* (2.13) *is a scaling action on the linear space* $S(\Lambda_k, E)$ *equipped with the metric* ρ *in* (2.9)*.*

Proof. Properties (i) and (ii) in Definition 2.2.6 are immediate. We now verify (iii). Consider the coordinate-wise action on $\widetilde{S}(\Lambda_k, E)$ also denoted as $\{\widetilde{T}_a,\ a \in \mathbb{R}_+\}$ for convenience. One can easily verify that $\mathscr{K}(\widetilde{T}_a(f)) = \widetilde{T}_a(\mathscr{K}(f))$. Let $f_n \to f$ in $S(\Lambda_k, E)$ and $a_n \to a > 0$. To show that $\rho(\widetilde{T}_{a_n}(f_n), \widetilde{T}_a(f)) \to 0$, it is enough to verify that for every

$K > 0$, we have

$$\sup_{\|t\|_2 \le K} d(\widetilde{T}_{a_n}(\widetilde{f}_n)(t), \widetilde{T}_a(\widetilde{f})(t)) \to 0, \quad \text{as } n \to \infty,$$

where $\widetilde{f}_n = \mathscr{K}(f_n)$ and $\widetilde{f} = \mathscr{K}(f)$.

In view of Lemma 2.4.2, it is enough to show that $\widetilde{T}_{a_n}(\widetilde{f}_n)(t_n) \to \widetilde{T}_a(\widetilde{f})(t)$, whenever $t_n \to t$ in $\overline{B}_K := \{t : \|t\|_2 \le K, t \in \mathbb{R}^d\}$. Notice, however, that $\widetilde{T}_{a_n}(\widetilde{f}_n)(t_n) = T_{a_n}(y_n)$ and $\widetilde{T}_a(\widetilde{f})(t) = T_a(y)$, where $y_n := \widetilde{f}_n(t_n)$ and $y := \widetilde{f}(t)$. By applying Lemma 2.4.2 again, but now to the locally converging functions $\widetilde{f}_n$ and $\widetilde{f}$, we have that $y_n \equiv \widetilde{f}_n(t_n) \to y \equiv \widetilde{f}(t)$ in E, whenever $t_n \to t$ in $\overline{B}_K$. Hence, the continuity of the scaling action $\{T_a\}$, yields $T_{a_n}(y_n) \to T_a(y)$ in E, which completes the proof of property (iii).

Let now $0 \neq f \in S(\Lambda_k, E)$. Proving property (iv) of Definition 2.2.6 amounts to showing that $\rho(\widetilde{T}_{a_1}(f), 0) < \rho(\widetilde{T}_{a_2}(f), 0)$, for all $0 < a_1 < a_2$. Observe that by property (iv) for $\{T_a\}$, for all $\lambda \in \widetilde{\Lambda}$, we have

$$d(\widetilde{T}_{a_1}(\widetilde{f})(t), 0) \equiv d(T_{a_1}(\widetilde{f}(t)), 0) \le d(T_{a_2}(\widetilde{f}(t)), 0) \equiv d(\widetilde{T}_{a_2}(\widetilde{f})(t), 0).$$

This implies that $\rho(\widetilde{T}_{a_1}(f), 0) \le \rho(\widetilde{T}_{a_2}(f), 0)$. We next argue that the inequality is strict. Since $f \neq 0$, we have $0 \neq \widetilde{f}(t) \in E$ for some $t \in \mathbb{R}^d$. Let $t \in \overline{B}_j$ for some large enough j. Since the suprema therein are attained, it is enough to show that

$$\max_{\mu \in \overline{B}_j} d(T_{a_1}(\widetilde{f}(t)), 0) =: d(T_{a_1}(\widetilde{f}(t_1)), 0) < \max_{t \in \overline{B}_j} d(T_{a_2}(\widetilde{f}(t)), 0) =: d(T_{a_2}(\widetilde{f}(t_2)), 0).$$

Observe that, $0 < d(T_{a_1}(\widetilde{f}(t), 0) \le d(T_{a_1}(\widetilde{f}(t_1)), 0)$ and hence $\widetilde{f}(t_1) \neq 0$. Thus, by the radial monotonicity of the action $\{T_a\}$, we have

$$d(T_{a_1}(\widetilde{f}(t_1)), 0) < d(T_{a_2}(\widetilde{f}(t_1)), 0) \le d(T_{a_2}(\widetilde{f}(t_2)), 0),$$

which yields (2.14) and completes the proof of (iv).

We now verify property (v). In view of Remark 2.2.3, it is equivalent to show that for all $f \in S(\Lambda_k, E)$, we have $\widetilde{T}_{1/n}(f) \to 0$ in $S(\Lambda_k, E)$. Suppose that this is not the case. Then, for some compact $K \subset \widetilde{\Lambda}$, some $\epsilon_0 > 0$ and a sequence $t_n \in K$, we have

$$d(T_{1/n}(\widetilde{f}(t_n)), 0) \ge \epsilon_0 > 0.$$

Since K is compact, for some $n' \to \infty$, we have $t_{n'} \to t_*$ and by the continuity of $\widetilde{f}$, we

have $\widetilde{f}(t_n) \to \widetilde{f}(t_*)$ in E. For all $\delta > 0$, fixed, the radial monotonicity implies that

$$0 < \epsilon_0 \leq \limsup_{n' \to \infty} d(T_{1/n'}(\widetilde{f}(t_{n'})), 0) \leq \lim_{n' \to \infty} d(T_\delta(\widetilde{f}(t_{n'})), 0) = d(T_\delta(\widetilde{f}(t_*)), 0).$$

Property (v), for the scaling action $\{T_a\}$, however, entails that $T_\delta(\widetilde{f}(t_*)) \to 0$ in E, as $\delta \downarrow 0$, which yields a contradiction with the above inequality and completes the proof. □

Given the natural result in Lemma 2.2.7, from now on we will denote the scaling action on E and its coordinate-wise extensions on $S(\Lambda_k, E)$, $\widetilde{S}(\Lambda_k, E)$ with the same letter $\{T_a,\ a \in \mathbb{R}_+\}$.

Definition 2.2.8. *We say that the random field $\{Y_z(x),\ x \in \mathbb{R}^d\}$ is a k-th order tangent field to X at z if, for some $c(z, r) > 0$,*

$$\left\{T_{c(z,r)}(X(z, r \cdot \lambda)), \lambda \in \Lambda^{(k)}\right\} \xrightarrow{d} \{Y_z(\lambda), \lambda \in \Lambda^{(k)}\}, \quad \text{as } r \downarrow 0, \tag{2.14}$$

where the convergence in distribution is in the space $S(\Lambda^{(k)}, E)$.

In the next two sections, we will prove the following two properties for tangent fields with some additional assumptions,

$$\{Y_z(r \cdot \lambda), \lambda \in \Lambda^{(k)}\} \stackrel{d}{=} \{T_{r^{\alpha(z)}}(Y_z(\lambda)), \lambda \in \Lambda^{(k)}\} \text{ (Self-Similarity)},$$

and

$$\{Y_z(w + \lambda), \lambda \in \Lambda^{(k)}\} \stackrel{d}{=} \{Y_z(\lambda), \lambda \in \Lambda^{(k)}\} \text{ (IRF}_k\text{)}.$$

2.3 Self-similarity

In this section, we prove the self-similar property for tangent fields. We shall need the following Slutsky-type lemma for scaling actions.

Lemma 2.3.1. *Let $\xi, \widetilde{\xi}$ and ξ_n be random elements taking values in a separable linear metric space (E, d). Let also $\{T_a,\ a \in (0, \infty)\}$ be a scaling action on E as in Definition 2.2.6.*

(i) *If ξ is non-zero (a.s.), then $T_{a'}(\xi) \stackrel{d}{=} T_{a''}(\xi)$ implies $a' = a''$.*

(ii) *Suppose that $\xi_n \xrightarrow{d} \xi$ and $T_{a_n}(\xi_n) \xrightarrow{d} \widetilde{\xi}$, as $n \to \infty$, for some sequence $a_n > 0$. If ξ and $\widetilde{\xi}$ are non-zero (a.s.), then $a_n \to a$, as $n \to \infty$, for some $a > 0$ and $\widetilde{\xi} \stackrel{d}{=} T_a(\xi)$.*

(iii) *If ξ is non-zero (a.s.), and $T_{a_n}(\xi) \xrightarrow{d} 0$, as $n \to \infty$, then $a_n \to 0$.*

Proof. Proof of part (i). Suppose that $a' < a''$. Then,

$$\xi \stackrel{d}{=} T_{a''}^{-1} \circ T_{a'}(\xi) \equiv T_{a'/a''}(\xi),$$

which implies $\xi \stackrel{d}{=} T_{c_n}(\xi)$, for all $n \in \mathbb{N}$, where $c_n = (a'/a'')^n \downarrow 0$. Thus, for every B_r, we obtain

$$\mathbb{P}(\xi \in B_r) = \mathbb{P}(T_{c_n}(\xi) \in B_r) = \mathbb{P}(\xi \in T_{1/c_n}(B_r)).$$

Since $1/c_n \to \infty$, applying Property (2.11) (recall (2.12)), we see that $\mathbb{P}(\xi \in B_r) = 1$, for all $r > 0$, which contradicts the assumption that ξ is non-zero.

(Proof of (ii)) We will first show that $\{a_n\}$ is bounded away from 0 and ∞. Indeed, suppose that $a_{n'} \to \infty$ for some $n' \to \infty$. Consider the balls

$$B_r := \{x \in E \,:\, d(x, 0) < r\}$$

and observe that all but countably many of them are continuity sets for the distribution of ξ. Indeed, the sets $\partial B_r := \overline{B}_r \setminus B_r,\ r > 0$ are pairwise disjoint in r and for each $\epsilon > 0$, there are at most $1/\epsilon$ distinct values for r, such that $\mathbb{P}(\xi \in \partial B_r) > \epsilon$.

For every $r > 0$ such that $\mathbb{P}(\xi \in \partial B_r) = 0$, since $\xi_{n'} \xrightarrow{d} \xi$, we have

$$\begin{aligned} \mathbb{P}(\xi \in B_r) &= \lim_{n' \to \infty} \mathbb{P}(\xi_{n'} \in B_r) = \lim_{n' \to \infty} \mathbb{P}(T_{a_{n'}}(\xi_{n'}) \in T_{a_{n'}}(B_r)) \\ &\geq \limsup_{n' \to \infty} \mathbb{P}(\widetilde{\xi}_{n'} \in T_m(B_r)), \end{aligned} \tag{2.15}$$

where $\widetilde{\xi}_{n'} := T_{a_{n'}}(\xi_{n'})$ and m is an arbitrary fixed integer. Here, we used the fact that $T_m(B_r) \subset T_{a_{n'}}(B_r)$, for all large enough n', by the radial monotonicity property of the scaling action.

Now, since T_m is a homeomorphism, we have $\partial T_m(B_r) = T_m(\partial B_r)$ are disjoint in $r > 0$, and by the above argument, for all but countably many r's, we have $\mathbb{P}(\widetilde{\xi} \in \partial T_m(B_r)) = 0$ and hence $\mathbb{P}(\widetilde{\xi}_{n'} \in T_m(B_r)) \to \mathbb{P}(\widetilde{\xi} \in T_m(B_r))$, as $n' \to \infty$. Therefore, in view of (2.15), we obtain

$$\mathbb{P}(\xi \in B_r) \geq \mathbb{P}(\widetilde{\xi} \in T_m(B_r)), \quad \text{for all } m \text{ and all but countably many } r > 0.$$

Relation (2.11), however, implies that $T_m(B_r) \uparrow E$ as $m \to \infty$, which implies

$$\mathbb{P}(\xi \in B_r) = 1 = \lim_{m\to\infty} \mathbb{P}(\widetilde{\xi} \in T_m(B_r)),$$

for all but countably many r. This implies that $\mathbb{P}(\xi = 0) = 1$, which is a contradiction.

We have thus shown that the sequence $\{a_n\}$ is bounded above. One can similarly show that $\{a_n\}$ is bounded away from 0. Indeed, by defining $\widetilde{a}_n := 1/a_n$, we see that $\xi_n = T_{\widetilde{a}_n}(\widetilde{\xi}_n) \xrightarrow{d} \xi$. Therefore, repeating the above argument with a_n, ξ_n and ξ replaced by $\widetilde{a}_n, \widetilde{\xi}_n$ and $\widetilde{\xi}$, respectively, we see that $\{\widetilde{a}_n \equiv 1/a_n\}$ is bounded.

We have thus shown that $\{a_n\}$ can only have positive cluster points. Suppose that $a_{n'} \to a' > 0$ and $a_{n''} \to a'' > 0$, for some sub-sequences $n', n'' \to \infty$. Since the space (E, d) is separable, by the Skorokhod-Dudley representation Theorem 3.30 on page 56 in Kallenberg (1997), on a suitable probability space we can define ξ^* and ξ^*_n such that

$$\xi_n \stackrel{d}{=} \xi_n, \quad \xi^* \stackrel{d}{=} \xi, \quad \text{and} \quad \xi^*_n \to \xi^*, \text{ almost surely.}$$

Thus, the continuity property (iii) in Definition 2.2.6, implies that

$$T_{a_{n'}}(\xi^*_{n'}) \to T_{a'}(\xi^*) \quad \text{and} \quad T_{a_{n''}}(\xi^*_{n''}) \to T_{a''}(\xi^*),$$

almost surely. Since also $T_{a_n}(\xi^*_n) \stackrel{d}{=} T_{a_n}(\xi_n) \xrightarrow{d} \widetilde{\xi}$, and $\xi \stackrel{d}{=} \xi^*$, we obtain

$$T_{a'}(\xi) \stackrel{d}{=} \widetilde{\xi} \stackrel{d}{=} T_{a''}(\xi). \tag{2.16}$$

By part (i), this is only possible if $a' = a''$. We have thus shown that the sequence $\{a_n\}$ has a unique cluster point $a = a' = a'' > 0$ and in view of (2.16), that $T_a(\xi) \stackrel{d}{=} \widetilde{\xi}$.

(Proof of (iii)) Suppose that $\limsup_{n\to\infty} a_n > 0$, i.e., for some subsequence $n' \to \infty$, we have $a_{n'} \geq \epsilon_0 > 0$, for all n'. Then, in view of (2.12), for all $r > 0$, we have

$$\mathbb{P}(T_{\epsilon_0}(\xi) \in B_r^c) \leq \mathbb{P}(T_{a_{n'}}(\xi) \in B_r^c).$$

Since $T_{a_{n'}}(\xi) \to^d 0$, the right-hand side vanishes, as $n' \to \infty$. On the other hand, since ξ is nonzero, so is $T_{\epsilon_0}(\xi)$ and the left-hand side is positive for sufficiently small $r > 0$. This contradiction yields $\limsup_{n\to\infty} a_n = 0$. □

We also prove the following result for continuity in both operators under the metric of ρ.

Lemma 2.3.2. *Let X, X_n be random functions in $S(\Lambda_k, E)$ such that $X_n \xrightarrow{d} X$. Assume [M] and then for any sequence $w_n \to 0$, $r_n \to 1$, $\{X_n(r_n \cdot (w_n + \lambda)), \lambda \in \Lambda_k\} \xrightarrow{d} \{X(\lambda), \lambda \in \Lambda_k\}$.*

Proof. Define

$$Y_n(\lambda) = X_n(r_n \cdot (w_n + \lambda)) \text{ and } \widetilde{Y}_n = \mathscr{J}(Y_n).$$

By Proposition 2.2.4, conditions (i) and (ii) hold for $\widetilde{X}_n$. We need to show that they also hold for $\widetilde{Y}_n$. The proof for (i) is easy and we focus on proving (ii).

First consider the relationship between $\widetilde{X}_n$ and $\widetilde{Y}_n$. Recall from (2.7) that

$$Y_n(\lambda_t) = X_n(r_n \cdot (w_n + \lambda_t)) = \int \widetilde{X}_n(r_n u + w_n)\lambda_t(du).$$

As $\widetilde{X}_n(t) = \widetilde{X}_n(\lambda_t)$, for any K and n large enough,

$$\begin{aligned} \sup_{\|t\|_2 \le K} d(\widetilde{Y}_n(t), \widetilde{X}_n(t)) &\le \sup_{\|t\|_2 \le K} \sum_i d(c_i(t)\widetilde{X}_n(r_n x_i(t) + w_n), c_i(t)\widetilde{X}_n(x_i(t))) \\ &\le (M_k + 1) \sup_{\|s\|_2, \|t\|_2 \le 2K, |c| \le C_K, \|s-t\|_2 \le \delta_n} d(c\widetilde{X}_n(s), c\widetilde{X}_n(t)), \end{aligned} \tag{2.17}$$

where we assume $\lambda_t = \sum_i c_i(t)x_i(t)$, C_K is the upper bound of $|c_i(t)|$ for $\|t\|_2 \le K$ and $\delta_n = |(r_n - 1)K + w_n| \to 0$. Therefore, for any $\eta, \widetilde{\epsilon} > 0$, by (ii) in Proposition 2.2.4 there exists n large enough such that,

$$\mathbb{P}\left(A_n := \left\{ \sup_{\|s\|_2, \|t\|_2 \le 2K, \|s-t\|_2 \le \delta_n} d(\widetilde{X}_n(s), \widetilde{X}_n(t)) < \widetilde{\eta} \right\}\right) > 1 - \epsilon.$$

Then according to [M] and (2.17), we have on A_n,

$$\sup_{\|t\|_2 \le K} d(\widetilde{Y}_n(t), \widetilde{X}_n(t)) \le (M_k + 1) f_{C_K}(\widetilde{\eta}),$$

where for any $C_K > 0$, $f_{C_K}(\eta) := \sup_{d(x,y)<\eta, |c|<C_K} d(cx, cy) \to 0$ as $\eta \to 0$. Thus, on A_n

$$\sup_{\|s\|_2, \|t\|_2 \le K, \|s-t\|_2 \le \delta_n} d(\widetilde{Y}_n(s), \widetilde{Y}_n(t)) \le 2 f_{C_K}(\widetilde{\eta}) + \widetilde{\eta}.$$

Thus, the second condition of Proposition 2.2.4 for $\widetilde{Y}_n$ follows. □

The following theorem shows the self-similarity property of tangent fields.

Theorem 2.3.3. *Assume [M] and the following assumptions hold:*

[A1]: Fix $z \in \mathbb{R}^d$. There exist constants $r > 0, c(z,r) > 0$ and random fields $\{X(t)\}$ on $\mathbb{R}^d$ with continuous paths, such that

$$\left\{T_{c(z,r)}(X(z, r\cdot\lambda)), \lambda \in \Lambda^{(k)}\right\} \xrightarrow{d} \{Y_z(\lambda), \lambda \in \Lambda^{(k)}\}, \tag{2.18}$$

as $r \downarrow 0$, where $Y_z \in S(\Lambda_k, E)$.

[A2]: $Y_z \in S(\Lambda_k, E)$ is not equal to zero almost surely.

Then, for all $r > 0$, we have

$$\{Y_z(r\cdot\lambda), \lambda \in \Lambda^{(k)}\} \stackrel{d}{=} \{T_{r^{\alpha(z)}} Y_z(\lambda), \lambda \in \Lambda^{(k)}\}, \tag{2.19}$$

where $\alpha(z) > 0$ is some positive constant. Moreover, we have

$$c(z,r) = r^{-\alpha(z)}\ell_z(r),$$

where $\ell_z(r)$ is a slowly varying function at 0, i.e., for any fixed $s > 0$, $\lim_{r\to 0} \frac{\ell_z(r)}{\ell_z(sr)} = 1$.

Proof. For all fixed $s > 0$, by (2.18), as $r \downarrow 0$, we have

$$\widetilde{\xi}_r := \left\{T_{c(z,r)}(X(z, (sr)\cdot\lambda)), \lambda \in \Lambda^{(k)}\right\} \xrightarrow{d} \widetilde{\xi} := \left\{Y_z(s\cdot\lambda), \lambda \in \Lambda^{(k)}\right\}.$$

On the other hand, as $r \downarrow 0$,

$$\xi_r := \left\{T_{c(z,sr)}(X(z, (sr)\cdot\lambda)), \lambda \in \Lambda^{(k)}\right\} \xrightarrow{d} \xi := \left\{Y_z(\lambda), \lambda \in \Lambda^{(k)}\right\}.$$

Observe that

$$T_{c(z,r)}(X(z, (sr)\cdot\lambda)) = T_{\frac{c(z,r)}{c(z,sr)}} \circ T_{c(z,sr)}(X(z, (sr)\cdot\lambda))$$

and hence

$$T_{\frac{c(z,r)}{c(z,sr)}}(\xi_r) = \widetilde{\xi}_r \xrightarrow{d} \widetilde{\xi}.$$

By condition *[A2]*, both $\xi \equiv \{Y_z(\lambda)\}$ and $\widetilde{\xi} \equiv \{Y_z(s\cdot\lambda)\}$ are non-zero. Lemma 2.3.1, applied to the $S(\Lambda^{(k)}, E)$-valued random variables $\xi_r, \widetilde{\xi}_r, \xi$ and $\widetilde{\xi}$, implies

$$\frac{c(z,r)}{c(z,sr)} \to a(z,s), \quad \text{as } r \downarrow 0, \tag{2.20}$$

for some positive $a(z,s) > 0$. We have, moreover, $\widetilde{\xi} \stackrel{d}{=} T_{a(z,s)}(\xi)$, which reads

$$\left\{Y_z(s\cdot\lambda), \lambda \in \Lambda^{(k)}\right\} \stackrel{d}{=} \left\{T_{a(z,s)}(Y_z(\lambda)), \lambda \in \Lambda^{(k)}\right\}. \tag{2.21}$$

We will next show that $a(z,s) = s^{\alpha(z)}$, for some $\alpha(z) > 0$. Relation (2.21) readily implies that for all $s_1 > 0$ and $s_2 > 0$

$$a(z, s_1 s_2) = a(z, s_1)a(z, s_2). \tag{2.22}$$

Indeed, by (2.21),

$$\begin{aligned}\{T_{a(z,s_1s_2)}(Y_z(\lambda))\} &\stackrel{d}{=} \{Y_z((s_1s_2)\cdot\lambda)\} \stackrel{d}{=} \{T_{a(z,s_1)}(Y_z(s_2\cdot\lambda))\} \\ &\stackrel{d}{=} \{T_{a(z,s_1)} \circ T_{a(z,s_2)}(Y_z(\lambda))\} = \{T_{a(z,s_1)a(z,s_2)}(Y_z(\lambda))\}.\end{aligned}$$

Since Y_z is nonzero in view of Lemma 2.3.1.(i), the last relation yields (2.22).

The function $a(z,s)$ is also continuous in $s \in (0,\infty)$. Indeed, for any sequence $s_n \to s$, $s_n, s > 0$, by Lemma 2.3.2, we have $\{Y_z(s_n\cdot\lambda)\} \stackrel{d}{\to} \{Y_z(s\cdot\lambda)\}$. Therefore, by (2.21),

$$\{T_{a(z,s_n)}(Y_z(\lambda))\} \stackrel{d}{=} \{Y_z(s_n\cdot\lambda)\} \stackrel{d}{\to} \{Y_z(s\cdot\lambda)\} \stackrel{d}{=} \{T_{a(z,s)}(Y_z(\lambda))\}. \tag{2.23}$$

Since Y_z is nonzero, by applying Lemma 2.3.1.(ii) again, we obtain

$$a(z, s_n) \to a(z, s), \tag{2.24}$$

which shows the desired continuity.

By combining (2.22) and (2.24) with the continuity of $a(z,\cdot)$ and facts that $a(z,1) = 1$, it is easy to see that $a(z,s) = s^{\alpha(z)},\ s > 0$, for some $\alpha(z) \in (-\infty,\infty)$. This is in fact a special example of Cauchy's functional equation (c.f. Theorem 5.2.1 in Kuczma (2009)).

We will show next that $a(z,s_n) \to 0$ as $s_n \downarrow 0$, which necessarily implies $\alpha(z) > 0$. Indeed, with $s := 0$, (2.23) implies that $\{T_{a(z,s_n)}(Y(\lambda))\} \stackrel{d}{\to} 0 = \{Y(0\cdot\lambda)\}$, as $n \to \infty$. This, by Lemma 2.3.1.(iii) yields $a(z,s_n) \to 0$.

To conclude the proof, letting $c(z,r) =: r^{\alpha(z)}\ell_z(r)$, we see from equation (2.20), that for all $s > 0$

$$\frac{\ell_z(r)}{\ell_z(sr)} \to 1, \text{ as } r \downarrow 0.$$

This shows ℓ_z is a slowly varying function at 0. □

2.4 Intrinsic stationarity

In Falconer (2002) establishing the structure of tangent processes to general random fields with continuous paths, the most surprising finding (Theorem 3.6 in Falconer (2002)) is that almost everywhere the tangent fields have *stationary increments*. This result has a delicate measure-theoretic proof. Here, using independent arguments, we extend this fundamental result to the case of *intrinsic random functions of order* k (IRF$_k$), with Falconer's result corresponding to $k = 0$.

2.4.1 A tool based on the Egorov and Luzin Theorems

The following proposition has a key role in establishing the structure of tangent processes.

Proposition 2.4.1. *Let $A \subset \mathbb{R}^d$ be a Borel set with finite Lebesgue measure Leb$(A) < \infty$. Suppose that $F_n : A \to E$ is a sequence of Borel measurable functions into the separable metric space (E, ρ_E) such that*

$$F_n(z) \underset{n\to\infty}{\longrightarrow} G(z), \text{ for almost all } z \in A.$$

Then, for every $\epsilon > 0$, there exists a compact set $K_\epsilon \subset A$, such that Leb$(A \setminus K_\epsilon) < \epsilon$, the function G being continuous on K_ϵ, and

$$F_n(z_n) \to G(z), \text{ whenever } z_n \to z, \text{ for } z_n, z \in K_\epsilon. \tag{2.25}$$

The rest of this Section 2.4.1 is devoted to the proof of this proposition.

Lemma 2.4.2. *Let (K, ρ_K) be a compact metric space and (E, ρ_E) be a metric space. Suppose that f_n and $f : K \to E$ are measurable functions.*

If the function f is continuous, then

$$\sup_{x\in K} \rho_E(f_n(x), f(x)) \to 0, \quad \text{as } n \to \infty, \tag{2.26}$$

if and only if

$$f_n(x_n) \to f(x), \quad \text{whenever } x_n \to x. \tag{2.27}$$

Proof. ('if') Suppose that (2.27) holds and assume that (2.26) fails. Then, for some $\epsilon_0 > 0$, there exist an infinite sequence $n_k \in \mathbb{N}$ and $x_{n_k} \in K$, such that $\rho_E(f_{n_k}(x_{n_k}), f(x_{n_k})) \geq \epsilon_0 > 0$. It is easy to see that $\{x_{n_k}\}$ is also an infinite sequence, since for every $k_0 \in \mathbb{N}$, by (2.27), we have $f_{n_k}(x_{n_{k_0}}) \to f(x_{n_{k_0}})$, as $n_k \to \infty$.

The infinite sequence $\{x_{n_k}\}$ is included in the compact K, and hence it has a converging subsequence $x_{n_{k(m)}} \to x$. This, in view of (2.27), implies that $f_{n_k(m)}(x_{n_{k(m)}}) \to f(x)$ in E. Since f is continuous at x, however, $f(x_{n_{k(m)}}) \to f(x)$. This, by the triangle inequality, implies

$$\begin{aligned} 0 &\leq \rho_E(f_{n_k(m)}(x_{n_{k(m)}}), f(x_{n_{k(m)}})) \\ &\leq \rho_E(f_{n_k(m)}(x_{n_{k(m)}}), f(x)) + \rho_E(f(x), f(x_{n_{k(m)}})) \\ &\to 0, \text{ as } n_k(m) \to \infty. \end{aligned}$$

This contradicts the assumption that $\rho_E(f_{n_k(m)}(x_{n_{k(m)}}), f(x_{n_{k(m)}})) \geq \epsilon_0$.

('only if') Let $x_n \to x$. By the triangle inequality, we have that

$$\begin{aligned} \rho_E(f_n(x_n), f(x)) &\leq \rho_E(f_n(x_n), f(x_n)) + \rho_E(f(x_n), f(x)) \\ &\leq \sup_{x' \in K} \rho_E(f_n(x'), f(x')) + \rho_E(f(x_n), f(x)), \end{aligned}$$

which converges to zero by (2.26) and the continuity of f. □

The next result is a restatement of Theorem 7.5.1 in Dudley (2002).

Theorem 2.4.3 (Egorov). *Let $(A, \mathscr{A}, \mu)$ be a finite measure space and (Y, ρ_Y) be a separable metric space. Suppose that $f_n : A \to Y,\ n = 1, \cdots$ are measurable functions such that, for μ-almost all $x \in A$,*

$$f_n(x) \to f(x), \quad \textit{as } n \to \infty.$$

Then, for all $\epsilon > 0$, there exists a measurable set $A_\epsilon \subset A$, such that

$$\mu(A \setminus A_\epsilon) < \epsilon \quad \textit{and} \quad \sup_{x \in A_\epsilon} \rho_Y(f_n(x), f(x)) \to 0, \quad \textit{as } n \to \infty.$$

We present next a relatively general form of the classic Luzin's theorem stating that every Borel function is nearly continuous. The proof follows the elegant 3-line argument given in Theorem 1 on page 56 in Loeb and Talvila (2004). We provide a bit more detail and tailor the result to the case of metric spaces.

Theorem 2.4.4 (Luzin). *Let (X, ρ_X) be a metric space and (Y, ρ_Y) be a separable metric space. Let also $f : X \to Y$ be a Borel measurable function and μ be a finite Borel measure μ on X.*

For every $\epsilon > 0$, there exists a closed set $F \subset X$, such that $\mu(X \setminus F) < \epsilon$ and $f : F \to Y$ is continuous. If (X, ρ_X) is separable and complete, then the set F can be taken to be compact.

Proof. We will essentially unpack the argument on page 56 of Loeb and Talvila (2004) with small modifications.

By Theorem 7.1.3 on page 175 in Dudley (2002) every finite Borel measure μ on (X, ρ_X) is *closed regular*, that is, for every Borel set A in X, we have

$$\mu(A) = \sup\{\mu(F) \,:\, F \subset A,\ F - \text{closed}\}. \tag{2.28}$$

Recall that μ is called regular if the sets F above can be taken to be compact. Ulam's Theorem implies that if (X, ρ_X) is separable and complete, then μ is regular (c.f. Theorem 7.1.4 in Dudley (2002)).

We now fix an $\epsilon > 0$ and construct the closed set F. Since (Y, ρ_Y) is separable, it is second countable, i.e., its topology has a countable base. Namely, there exists a countable collection of open sets $\{V_n,\ n \in \mathbb{N}\}$ in Y such that every open set $V \subset Y$ can be represented as a union of V_n's, i.e., $V = \cup\{V_n \,:\, V_n \subset V,\ n \in \mathbb{N}\}$.

Following Loeb and Talvila Loeb and Talvila (2004), by (2.28) since μ is finite, we can find closed sets $F_n \subset f^{-1}(V_n)$ and $F_n' \subset X \setminus f^{-1}(V_n)$ in X (compact if μ is regular), such that

$$\mu(f^{-1}(V_n) \setminus F_n) < \frac{\epsilon}{2^{n+1}} \qquad \text{and} \qquad \mu([X \setminus f^{-1}(V_n)] \setminus F_n') < \frac{\epsilon}{2^{n+1}}.$$

Observe that

$$\mu(X \setminus (F_n \cup F_n')) = \mu(f^{-1}(V_n) \setminus F_n) + \mu([X \setminus f^{-1}(V_n)] \setminus F_n') < \frac{\epsilon}{2^n}.$$

Define $F := \cap_{n\in\mathbb{N}}(F_n \cup F_n')$ and notice that F is closed and in fact compact if (X, ρ_X) is separable and complete. The above relation implies moreover that

$$\mu(X \setminus F) \le \sum_{n\in\mathbb{N}} \mu(X \setminus (F_n \cup F_n')) < \sum_{n\in\mathbb{N}} \frac{\epsilon}{2^n} = \epsilon.$$

To complete the proof, it remains to show that $f : F \to Y$ is continuous. To this end, it is enough to show that for every $x \in F$ and every V_n such that $f(x) \in V_n$, there is an open set $U \ni x$ such that $f(U \cap F) \subset f(V_n)$. Suppose $f(x) \in V_n$ and consider the open set

$U := X \setminus F_n'$. Since $F \subset F_n \cap F_n'$ and $U \cap F_n' = \emptyset$, we have

$$U \cap F \subset U \cap F_n \subset f^{-1}(V_n),$$

which implies $f(U \cap F) \subset V_n$. We have thus established the desired continuity of f on F. □

Remark. *Loeb and Talvila's proof of Luzin's Theorem 2.4.4 is not constructive and it does not use approximation arguments based on the Tietze–Uryson Lemma and Egorov's theorem as many other proofs in the literature (see e.g., Theorem 7.5.2 in Dudley (2002)). This makes it possible to extend Luzin's theorem to functions taking values in an arbitrary separable metric space.*

We conclude this section with the proof of Proposition 2.4.1.

Proof of Proposition 2.4.1. By Egorov's Theorem (see Theorem 2.4.3), there is a Borel set $A_\epsilon \subset A$ such that $\mu(A \setminus A_\epsilon) < \epsilon/2$ and

$$\sup_{z \in A_\epsilon} \rho_K(F_n(z), G(z)) \to 0, \tag{2.29}$$

as $n \to \infty$. Observe that since the Lebesgue measure is closed regular (recall (2.28)), one can choose the set A_ϵ to be closed. Therefore, A_ϵ with the usual metric in $\mathbb{R}^d$ is a complete and separable metric space. Hence, we can apply Luzin's Theorem 2.4.4 to $X := A_\epsilon \subset \mathbb{R}^d$ and $Y := E$ to conclude that there is a further compact set $K_\epsilon \subset A_\epsilon$, such that $\mu(A_\epsilon \setminus K_\epsilon) < \epsilon/2$ and the function $G : K_\epsilon \to E$ is *continuous*.

Observe that

$$\mu(A \setminus K_\epsilon) = \mu(A \setminus A_\epsilon) + \mu(A_\epsilon \setminus K_\epsilon) < \epsilon.$$

By Lemma 2.4.2, the continuity of G on K_ϵ and the uniform convergence (2.29) imply (2.25). □

2.4.2 Extension of Falconer's Theorem for IRF$_k$

Let

$$F_n(z) := \text{Law}\left\{T_{c(z,1/n)} X\Big(z, \frac{1}{n} \cdot \lambda\Big),\ \lambda \in \Lambda^{(k)}\right\},$$

be the probability distribution of the rescaled and localized versions of process $X(z, \cdot)$ on the space $S(\Lambda^{(k)}, E)$ and similarly, let

$$G(z) := \text{Law}\left\{Y_z(\lambda),\ \lambda \in \Lambda^{(k)}\right\}.$$

Observe that $F_n(z)$ and $G(z)$ are functions of $z \in \mathbb{R}^d$ taking values in the set of all probability measures on $S(\Lambda^{(k)}, E)$. Thus, Assumption (2.14) (with $r := 1/n$) can be equivalently and succinctly expressed as

$$F_n(z) \xrightarrow{w} G(z), \quad n \to \infty. \tag{2.30}$$

We will utilize Proposition 2.4.1. To this end, we need to equip the space of probability measures $\mathscr{P}(S(\Lambda^{(k)}, E))$ on $S(\Lambda^{(k)}, E)$ with a separable and complete metric, which metrizes the weak convergence (2.30).

Lemma 2.4.5. *The Lévy-Prokhorov distance d_{LP} metrizes the weak convergence and the space $(\mathscr{P}(S(\Lambda^{(k)}, E)),\ d_{LP})$ is separable and complete.*

The result is a direct consequence of the fact that $(S(\Lambda^{(k)}, E), \rho)$ is separable and complete and Theorem 6.8 in Billingsley (1999).

Lemma 2.4.6. *Assume [M] and the following conditions:*

[A3]: For some $c(z, r) > 0$, $X(z)$ a continuous random fields on a Borel set B,

$$\left\{T_{c(z,r)}(X(z, r \cdot \lambda)), \lambda \in \Lambda^{(k)}\right\} \xrightarrow{d} \{Y_z(\lambda), \lambda \in \Lambda_k\}, \quad \text{as } r \downarrow 0, \tag{2.31}$$

where $Y_z \in S(\Lambda_k, E)$.

[A4] For each $n \in \mathbb{N}$, $c(z, 1/n)$ is Borel measurable with respect to z.

Then $F_n(z) : \mathbb{R}^d \mapsto \mathscr{P}(S(\Lambda^{(k)}, E), \rho)$ is Borel measurable.

Proof. For any $X \in \mathscr{P}(S(\Lambda^{(k)}, E))$, let $D = \text{Law}(X)$ and T'_a be an operator on $\mathscr{P}(S(\Lambda_k, E), \rho)$ such that

$$T'_a(D) = \text{Law}\{T_a(X)\},$$

or equivalently, for any Borel set A on $(S(\Lambda_k, E), \rho)$, let

$$T'_a(D)(A) = D(T_{-a}(A)).$$

For any $X_n \xrightarrow{d} X$ and $a_n \to a$, by Lemma 2.2.7 and Skorokhod-Dudley representation Theorem, one can easily prove that

$$T_{a_n}(X_n) \xrightarrow{d} T_a(X),$$

or equivalently,

$$T'_{a_n}(D_n) \to T'_a(D),$$

where $D_n = \text{Law}\{X_n\}$. Thus, if we consider T' as a function from $(a, D) \in \mathbb{R}_+ \otimes \mathscr{P}(S(\Lambda_k), E)$ to $\mathscr{P}(S(\Lambda_k), E)$ with box topology, then it is a continuous function.

Let

$$D_n(z) = \text{Law}\left\{X\Big(z, \frac{1}{n} \cdot \lambda\Big), \lambda \in \Lambda_k\right\}.$$

For any fixed n and any $z' \to z$, by Lemma 2.3.2, we have

$$\{X_n(z', \lambda), \lambda \in \Lambda_k\} = \{X_n(z, z' - z + \lambda), \lambda \in \Lambda_k\} \xrightarrow{d} \{X_n(z, \lambda), \lambda \in \Lambda_k\},$$

or equivalently, $D_n(z)$ is a continuous function from $\mathbb{R}^d$ to $\mathscr{P}(S(\Lambda_k, E), \rho)$.

Denote $H_n(z) := (c(z, 1/n), D_n(z))$, which is a mapping from $\mathbb{R}^d$ to $\mathbb{R}_+ \otimes \mathscr{P}(S(\Lambda_k), E)$. If we equip the product space with box topology, it is easy to verify that H_n is Borel measurable as $c(z, 1/n)$ is Borel measurable and D_n is continuous.

Finally, it is easy to verify $F_n(z) = T'_{c(z,1/n)}(D_n(z)) = T' \circ H_n(z)$, which implies $F_n(z)$ is a Borel measurable function as T' is continuous and H_n is Borel measurable. □

Theorem 2.4.7. *Assume [M], [A2], [A3] and [A4]. There exists a set U with 0 Lebesgue measure s.t. for all $z \in B \setminus U$ and $w \in \mathbb{R}^d$,*

$$\{Y_z(w + \lambda), \lambda \in \Lambda^{(k)}\} \stackrel{d}{=} \{Y_z(\lambda), \lambda \in \Lambda^{(k)}\}.$$

Proof. By assumption [A3], (2.30) holds for $z \in B$. By Proposition 2.4.1, for $\epsilon > 0$, we can have a compact set K_ϵ with $\text{Leb}(B \setminus K_\epsilon) < \epsilon$ such that, $F_n(z_n) \to G(z)$ so long as $z_n, z \in K_\epsilon$ and $z_n \to z$.

Denote K'_ϵ to be the subset in K_ϵ with the Lebesgue density to be 1 and by the Lebesgue's density theorem, $\text{Leb}(K_\epsilon \setminus K'_\epsilon) = 0$. By Lemma 3.5 in Falconer (2002), for any point $z \in K'_\epsilon$ and $w \in \mathbb{R}^d$, there exists a sequence s.t. $w_n \to w$ and $z_n := z + w_n/n \in K_\epsilon$.

We will have $F_n(z_n) \to G(z)$, as $n \to \infty$ for any $z \in K_\epsilon$ and z_n defined above, or equivalently,

$$\left\{T_{c_n(z_n)}(X(z_n, (1/n) \cdot \lambda)), \lambda \in \Lambda^{(k)}\right\} \xrightarrow{d} \{Y_z(\lambda), \lambda \in \Lambda^{(k)}\}, \tag{2.32}$$

where $c_n(z) := c(z, 1/n)$.

On the other hand, we have

$$\begin{aligned}
&\left\{T_{c_n(z_n)}(X(z_n,(1/n))\cdot\lambda),\lambda\in\Lambda^{(k)}\right\}\\
&=\left\{T_{\frac{c_n(z_n)}{c_n(z)}}\circ T_{c_n(z)}(X(z_n,(1/n)\cdot\lambda)),\lambda\in\Lambda^{(k)}\right\}\\
&=\left\{T_{\frac{c_n(z_n)}{c_n(z)}}\circ T_{c_n(z)}(X(z,(1/n)\cdot(w_n+\lambda))),\lambda\in\Lambda^{(k)}\right\}
\end{aligned}\tag{2.33}$$

and by Lemma 2.3.2,

$$\left\{T_{c_n(z)}(X(z,(1/n)\cdot(w_n+\lambda)),\lambda\in\Lambda^{(k)}\right\}\xrightarrow{d}\{Y_z(w+\lambda),\lambda\in\Lambda^{(k)}\}.\tag{2.34}$$

By the three equations above, assumption [A2], Lemma 2.3.1 and (iv) in Definition 2.2.6, there exists $a_z(\{w_n\})$ such that

$$\frac{c_n(z)}{c_n(z+w_n/n)}\to a_z(\{w_n\})\tag{2.35}$$

and

$$\{T_{a_z(\{w_n\})}(Y_z(\lambda)),\lambda\in\Lambda^{(k)}\}\overset{d}{=}\{Y_z(w+\lambda),\lambda\in\Lambda^{(k)}\},\tag{2.36}$$

where $z\in K_\epsilon$. One can verify that $a_z(\{w_n\})$ is not related to the choice of w_n. In fact, if there exists another $w_n'\to w$ and $z+w_n'/n\in K_\epsilon$, the arguments above will still hold with the right sides of equations (2.32) and (2.34) to be changed accordingly. Therefore, we will have

$$\begin{aligned}
\{Y_z(w+\lambda),\lambda\in\Lambda^{(k)}\}&\overset{d}{=}\{T_{a_z(\{w_n\})}(Y_z(\lambda)),\lambda\in\Lambda^{(k)}\}\\
&\overset{d}{=}\{T_{a_z(\{w_n'\})}(Y_z(\lambda)),\lambda\in\Lambda^{(k)}\},
\end{aligned}$$

which shows $a_z(\{w_n\})=a_z(\{w_n'\})$. Thus, we can just denote it as $a_z(w)$. We only need to verify $a_z(w)=1$ to finish our proof.

By (2.36), one can easily obtain that $a_z(0)=1$ and $a_z(w+u)=a_z(w)a_z(u)$. By Theorem 2.3.3, assumptions [A2] and [A3], we have

$$\{Y_z(r\cdot\lambda),\lambda\in\Lambda_k\}\overset{d}{=}\{T_{r^\alpha}(Y_z(\lambda)),\lambda\in\Lambda_k\},\ \alpha>0.$$

Let's now consider $Y_z(r\cdot(w+\lambda))$ with $r\in\mathbb{N}_+$.

On one hand

$$\{Y_z(r \cdot (w+\lambda))\} \stackrel{d}{=} \{T_{r^\alpha}(Y_z(w+\lambda))\} \stackrel{d}{=} \{T_{r^\alpha a_z(w)}(Y_z(\lambda))\}.$$

On the other hand

$$\{Y_z(r \cdot (w+\lambda))\} = \{Y_z((rw + r \cdot \lambda))\} \stackrel{d}{=} \{T_{a_z(rw)}(Y_z(r \cdot \lambda))\} \stackrel{d}{=} \{T_{a_z^r(w) r^\alpha}(Y_z(\lambda))\}.$$

Thus, we have $a_z(w) = a_z^r(w)$ for $\forall r > 0$ and this leads to $a_z(w) = 1$ as $a_z(w) > 0$.

One can easily verify that the derivations above hold for any $z \in \cup_n K'_{1/n}$, and thus we can take $U := \cap_{k=1}^{\infty}(B \setminus K'_{1/n})$, which is a set with measure 0. By now, we have proved our theorem. □

2.5 Spectral theory for tangent fields

Spectral theory plays an important role of studying random field with certain stationarity. As we have connected tangent field with certain self-similar IRF$_k$ classes, it becomes possible to characterize mean-square continuous tangent fields with a refined spectral theory.

In Section 2.5.1-2.5.3, we provide a self-contained spectral theory for stationary random field or IRF$_k$ on $\mathbb{R}^d$ with its image in a separable Hilbert space $\mathbb{H}$. After this, in Section 2.5.4 and 2.5.5, we will illustrate the spectral theory for tangent field with different scaling operators.

2.5.1 Notations

Let $\mathbb{H}$ be a Hilbert space. Denote by $\mathbb{T}$ the collection of trace class operators on $\mathbb{H}$. Recall that the trace norm of a trace class operator $\mathscr{T}$ is

$$\|\mathscr{T}\|_{tr} = \sum_{k=1}^{\infty} \langle (\mathscr{T}^* \mathscr{T})^{1/2} e_k, e_k \rangle$$

for any CONS $\{e_k\}$ of $\mathbb{H}$. The space $\mathbb{T}$ equipped with the trace norm is a Banach space. Also recall that an operator $\mathscr{T}$ is nonnegative definite, denoted by $\mathscr{T} \geq 0$, if $\langle f, \mathscr{T} f \rangle \geq 0$ for all $f \in \mathbb{H}$. In some instances we will focus on operators in $\mathbb{T}$ that are self-adjoint and nonnegative definite; an operator $\mathscr{T}$ is nonnegative definite, denoted by $\mathscr{T} \geq 0$, if $\mathscr{T}$ is self-adjoint and

$$\langle f, \mathscr{T} f \rangle \geq 0 \text{ for all } f \in \mathbb{H}.$$

Denote this class of operators by $\mathbb{T}_+ \subset \mathbb{T}$.

A collection of operators $\{\mathscr{K}(t), t \in \mathbb{R}^d\}$ is said to be positive definite, denoted by $\mathscr{K}(\cdot) \geq 0$, if

$$\sum_{i=1}^{n}\sum_{j=1}^{n} c_i\overline{c_j}\mathscr{K}(t_i - t_j) \geq 0 \text{ (operator positivity).} \tag{2.37}$$

Similarly, we say $\{\mathscr{K}(t), t \in \mathbb{R}^d\}$ is k-conditionally positive definite if (2.37) holds for all $\{(c_i, t_i)\}$ satisfying $\sum_i c_i\delta_{t_i} \in \Lambda_k$.

A $\mathbb{T}_+$-valued measure μ is a mapping from the class of Borel sets on $\mathbb{R}^d$ to $\mathbb{T}$ such that $\mu(B)$ is self-adjoint and nonnegative definite for any Borel set B, and satisfying the countable additivity property:

$$\mu\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty}\mu(A_i).$$

A sequence of finite $\mathbb{T}_+$-valued measures $\{\mu_n\}$ is said to be tight if the following two conditions hold:

(i) $\sup_n \|\mu_n(\mathbb{R}^d)\|_{tr} < \infty$;

(ii) for all $\epsilon > 0$, there is a compact set $K_\epsilon \subset \mathbb{R}^d$ such that

$$\sup_n \|\mu_n(\mathbb{R}^d \setminus K_\epsilon)\|_{tr} < \epsilon.$$

Let $\{X(t), t \in \mathbb{R}^d\}$ be an $\mathbb{H}$-valued random field, namely, each $X(t)$ is a random element in $\mathbb{H}$. Assume that $\mathbb{E}X(t) = 0$ and $\mathbb{E}\|X(t)\|^2 < \infty$ for each t. Define the (cross) covariance operator

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

where for any $f, g \in \mathbb{H}$, $(f \otimes g)h := \langle h, g\rangle f$. Here, for definiteness, let both the mean and covariance be defined in the Bochner sense (cf. Hsing and Eubank (2015)). We define $X(t)$ is mean-square continuous if and only if

$$\lim_{h\to 0} \|\mathbb{E}[(X(t+h) - X(t)) \otimes (X(t+h) - X(t))]\|_{tr} \to 0,$$

or equivalently, as

$$tr(A \otimes B) = \sum_i \langle A \otimes Be_i, e_i \rangle = \sum_i \langle A, e_i \rangle \langle e_i, B \rangle = \langle A, B \rangle, \tag{2.38}$$

$$\lim_{h \to 0} \mathbb{E}[\|X(t+h) - X(t)\|^2] \to 0.$$

Say that $X(t)$ is (weakly) stationary if there exists a stationary covariance function $C(h), h \in \mathbb{R}^d$, such that $C(h) = C(s, s+h)$.

Definition 2.5.1. *Let X be an IRF$_k$ with mean 0. A symmetric continuous function $K(h) : \mathbb{R}^d \mapsto$ is called a generalized covariance of X if*

$$\mathbb{E}[X(\lambda) \otimes X(\mu)] = \sum_{i,j} \lambda_i \overline{\mu_j} K(x_i - y_j)$$

for any pair of measures $\lambda = \sum_i \lambda_i \delta_{x_i}, \mu = \sum_j \mu_j \delta_{y_j} \in \Lambda_k$.

2.5.2 Spectral measures of stationary and IRF$_k$ random fields

In this part, we provide theory about spectral measures of positive-definite collection of operators and generalized covariance function of IRF$_k$. Proofs of results in Sections 2.5.2 and 2.5.3 can be found in the Appendix.

Theorem 2.5.2. *Let $\{\mathscr{K}(t), t \in \mathbb{R}^d\}$ be a collection of operators where*

(i) $\mathscr{K}(\cdot)$ is positive definite.

(ii) $\mathscr{K}(0)$ is trace class and $\lim_{t \to 0} \|\mathscr{K}(t) - \mathscr{K}(0)\|_{tr} = 0$.

Then there exists a unique $\mathbb{T}_+$-valued measure μ such that $\mathscr{K}(t) = \int e^{itx} \mu(dx)$.

Remark. *The integration with respect to $\mathbb{T}_+$-valued measure μ is defined in Section 2.8.1. We also include the definition and properties of Bochner integral that will be used in the proof and the connection between those two different integrals.*

With this extension of Bochner theorem, we can prove existence of the generalized covariance function for $\mathbb{H}$-valued IRF$_k$, based on ideas of Sasvári (2009).

Theorem 2.5.3. *Let $\mathbb{H}$ be a Hilbert space and $X(t) \in \mathbb{H}, t \in \mathbb{R}^d$ be mean-square continuous. Then $X(t)$ is IRF$_k$ if and only if it has a generalized covariance operator*

$$K(h) = \int \frac{e^{iu^T h} - I_B(u) P(u^T h)}{1 \wedge |u|^{2k+2}} \sigma(du) + Q(h)$$

where $P(x) = \sum_{j=0}^{2k+1}(ix)^j/j!$, B *is a bounded neighbour of* 0, $Q(h)$ *is a* k*-conditionally positive definite polynomial with degree no more than* $2k+2$ *and* σ *is a finite* $\mathbb{T}_+$*-valued measure with no point mass at* 0*. In addition, for a given* X*,* σ *is the only measure satisfying those conditions.*

2.5.3 Integral representations of stationary and IRF$_k$ random fields

As a parallel to the classical spectral representation (Cramér (1942)) for stationary random processes taking values in finite-dimensional spaces, in this section we will establish a representation

$$X(t) = \int e^{it^T x} d\xi(x)$$

for $X(t) \in \mathbb{H}$ being infinite dimensional. Before doing so, we need to first define the stochastic integral on the right-hand side. Define $\mathbb{L}^2(\Omega)$ as the L^2 space of all $\mathbb{H}$-valued random elements η on $(\Omega, \mathscr{F}, \mathbb{P})$ with $\mathbb{E}\|\eta\|^2 < \infty$, equipped with the inner product

$$\langle \eta_1, \eta_2 \rangle_\Omega = \mathbb{E}\langle \eta_1, \eta_2 \rangle.$$

Also, for any $\mathbb{H}$-valued random measure ξ on $\mathbb{R}^d$, define $\mathbb{H}(\xi)$ as the closure of $\mathrm{span}(\xi) := \{\sum_{i=1}^m c_i \xi(A_i), c_i \in \mathbb{C}, A_i \in \mathscr{B}(\mathbb{R}^d), m = 1, 2, \ldots\}$ in $\mathbb{L}^2(\Omega)$, where $\mathscr{B}(\mathbb{R}^d)$ is the set of all Borel sets on $\mathbb{R}^d$. Suppose ξ has the following properties:

(i) $\|\xi(A)\|_\Omega < \infty$ for all $A \in \mathscr{B}(\mathbb{R}^d)$.

(ii) $\|\xi(A_n)\|_\Omega \to 0$ if $A_n \to \emptyset$.

(iii) For $A, B \in \mathscr{B}(\mathbb{R}^d)$, $A \cap B = \emptyset$,

$$\langle \xi(A), \xi(B)) \rangle_\Omega = 0.$$

For a random orthogonal measure ξ and any $A \in \mathscr{B}(\mathbb{R}^d)$, define a set function

$$\nu_\xi(A) = \mathbb{E}[\xi(A) \otimes \xi(A)].$$

By the orthogonal increment property (iii), ν_ξ is a $\mathbb{T}_+$-valued measure on $\mathbb{R}^d$ and we refer to it as the structure measure of ξ. Also, define

$$\mu_\xi(A) = tr(\nu_\xi(A)).$$

It is easy to verify that

$$\|\xi(A)\|_\Omega^2 = \mu_\xi(A),\ A \in \mathscr{B}(\mathbb{R}^d).$$

For any step function $\psi(x) = \sum_{i=1}^l c_i I_{A_i}(x)$, where $A_i \in \mathscr{B}(\mathbb{R}^d), A_i \cap A_j = \emptyset$ when $i \neq j$, define the integral

$$\mathscr{I}_\xi(\psi) = \sum_{i=1}^l c_i \xi(A_i) \in \mathbb{H}(\xi).$$

It is clear that

$$\|\mathscr{I}_\xi(\psi)\|_\Omega^2 = \int |\psi(x)|^2 \mu_\xi(dt).$$

Therefore, $\mathscr{I}_\xi$ is an isometric linear mapping between step functions in the L^2 space $\mathbb{L}^2(\mathbb{R}^d, \mu_\xi)$ and $\mathbb{H}(\xi)$. As step functions ψ are dense in $\mathbb{L}^2(\mathbb{R}^d, \mu_\xi)$ and the integrals $\mathscr{I}_\xi(\psi)$ are dense in $\mathbb{H}(\xi)$, $\mathscr{I}$ can be uniquely extended to an isometric linear mapping between $\mathbb{L}^2(\mathbb{R}^d, \mu_\xi)$ and $\mathbb{H}(\xi)$. Write

$$\mathscr{I}_\xi(\phi) = \int \phi(x)\xi(dx),\ \phi \in \mathbb{L}^2(\mathbb{R}^d, \mu_\xi).$$

Theorem 2.5.4. *Assume $\mathbb{H}$ is a separable Hilbert space. Let $X(t) \in \mathbb{H}$ be a mean-square continuous stationary random field on $\mathbb{R}^d$ with spectral measure σ. Then there exists a uniquely determined $\mathbb{H}$-valued random orthogonal measure ξ on $(\mathbb{R}^d, \mathscr{B})$ with structure measure equal to σ such that, in $\mathbb{L}^2(\Omega)$,*

$$X(t) = \mathscr{I}_\xi(e^{it^T \cdot}) = \int e^{it^T x} \xi(dx). \tag{2.39}$$

Conversely, let ξ be a random orthogonal measure on $(\mathbb{R}^d, \mathscr{B})$. Then (2.39) defines a continuous stationary field X with spectral measure $\sigma = \mu_\xi$.

Similarly, we can obtain stochastic representation for IRF_k.

Theorem 2.5.5. *Let $k \in \mathbb{N}$. Assume $X : \mathbb{R}^d \to$ is mean-square continuous. Then X is IRF_k if and only if X has the following representation:*

$$X(t) = Y(t) + \int \frac{e^{it^T x} - I_B(x) P(t^T x)}{1 \wedge |x|^{k+1}} \xi(dx),$$

where

(i) $P(x) = \sum_{j=0}^{k}(ix)^j/j!$, *$B$ is a bounded neighbour of* 0.

(ii) *ξ is a uniquely determined random orthogonal measure ξ on $(\mathbb{R}^d, \mathscr{B})$ with structure measure σ, and σ is a finite $\mathbb{T}_+$-valued measure with no point mass at* 0.

(iii) *$Y(t)$ is random polynomial up to order $k+1$.*

(iv) *For any $\mu \in \Lambda_k$, we have*

$$Y(\mu) \perp \int \frac{e^{it^T x}\widehat{\mu}(x)}{1 \wedge |x|^{k+1}}\xi(dx).$$

2.5.4 Characterization of tangent fields when $T_r(X) = r \cdot X$

In this section, we consider the spectral measure of tangent field when $T_r(X) = r \cdot X$ and $E = \mathbb{H}$, a separable Hilbert space.

Proposition 2.5.6. *Let $E = \mathbb{H}$, d is the Euclidean distance, $T_r(X) = r \cdot X$ and conditions [A2]-[A4] hold. In addition, we assume $Y_z(x)$ is mean-square continuous. Then we have $\alpha(z) \in (0, k+1]$. Furthermore, when $\alpha(z) \in (0, k+1)$, the generalized covariance of $Y_z(x)$ is*

$$K(h) = \int \frac{\exp(iu^T h) - 1_B(u)P(u^T h)}{1 \wedge |u|^{2k+2}}\chi(du) + Q(h), \tag{2.40}$$

with χ a unique $\mathbb{T}_+$-valued measure with $\chi(\{0\}) = 0$ and such that for any Borel set A satisfying $\inf_{u\in A}\|u\|_2 > 0$, we have

$$\int_A \frac{\chi(du)}{1 \wedge |u|^{2k+2}} = r^{2\alpha(z)}\int_{rA} \frac{\chi(du)}{1 \wedge |u|^{2k+2}},\ r > 0, \tag{2.41}$$

$P(x) = \sum_{j=0}^{2k+1}(ix)^j/j$, $1_B(u)$ is the indicator function of an arbitrary neighborhood of $u = 0$ and $Q(h)$ is a k-conditionally positive definite polynomial of degree no more than $2k$.

If $\alpha(z) = k+1$, then $K(h) = Q(h)$ is a k-conditionally positive definite polynomial with degree to be $2k+2$.

Proof. By Theorem 2.4.7 and Theorem 2.5.3, one can have

$$K(h) = \int \frac{\exp(u^T h) - 1_B(u)P(u^T h)}{1 \wedge |u|^{2k+2}}\chi(du) + Q(h), \tag{2.42}$$

where

$$\int \chi(du) < \infty, \tag{2.43}$$

and $Q(h)$ is a k-conditionally positive definite polynomial function of degree up to $2k+2$. By Theorem 2.5.3, we know $K(h)$ is unique up to a polynomial of degree $\leq 2k$.

Consider the generalized covariance of $Y_z(r \cdot \lambda)$. One can easily verify that $K(rh)$ is its generalized covariance function. On the other hand, by Theorem 2.3.3, for any $\lambda_1, \lambda_2 \in \Lambda^{(k)}$ and $r > 0$, we have

$$\text{Cov}(Y_z(r \cdot \lambda_1), Y_z(r \cdot \lambda_2)) = r^{2\alpha(z)}\text{Cov}(Y_z(\lambda_1), Y_z(\lambda_2)),$$

which indicates $r^{2\alpha(z)}K(h)$ is another generalized covariance function of $Y_z(r \cdot \lambda)$. Therefore,

$$\begin{aligned}
K(rh) &= \int \frac{\exp(ru^Th) - 1_B(u)P(ru^Th)}{1 \wedge |u|^{2k+2}}\chi(du) + Q(rh) \\
&\approx r^{2\alpha(z)}K(h) \\
&= r^{2\alpha(z)} \int \frac{\exp(u^Th) - 1_B(u)P(u^Th)}{1 \wedge |u|^{2k+2}}\chi(du) + r^{2\alpha(z)}Q(h) \\
&= r^{2\alpha(z)} \int \frac{\exp(ru^Th) - 1_B(u)P(ru^Th)}{1 \wedge |ru|^{2k+2}}\chi(dru) + r^{2\alpha(z)}Q(h),
\end{aligned}$$

where $\approx$ means equal up to an polynomial with order no larger than $2k$. By uniqueness of the positive measure χ, we come to equation (2.41).

We then study the range of $\alpha(z)$ through the condition (2.43).

Denote $F_1(r) = \int_{B_r^c} \chi(du)$, for $r > 0$ and B_r is the ball centered at 0 with radius r. By (2.43), we have $F_1(r) < \infty$ and $F_1(r) \to 0$, as $r \to \infty$. It follows that $F_1(r) = 0$ for $r > 0$ if and only if $\int_A \chi(du) = 0$ for any Borel set A.

On the other hand, by (2.41), for $r \geq 1$,

$$F_1(r) = \int_{B_1^c} \chi(dru) = r^{-2\alpha(z)} \int_{B_1^c} \frac{(1 \wedge |ru|^{2k+2})\chi(du)}{1 \wedge |u|^{2k+2}} = r^{-2\alpha(z)} \int_{B_1^c} \chi(du) = r^{-2\alpha(z)}F_1(1).$$

Thus, we have $F_1(r) \to 0$ if and only if (i) $\alpha(z) > 0$ or (ii) $\int_A \chi(du) = 0$ for any Borel set A.

Similarly, consider $F_2(r) = \int_{B_r} \chi(du)$. We have

$$F_2(r) = r^{2k+2-2\alpha(z)}F_2(1), r \leq 1.$$

It follows readily $F_2(r) \to 0$ as $r \to 0$ if and only if (i) $\alpha(z) < k+1$ or (ii) $\int_A \chi(du) = 0$ for any Borel set A.

Therefore, if $\chi(du)$ is not 0 a.e., we must have $\alpha(z) \in (0, k+1)$.

Denote the coefficient of $|h|^{2k+2}$ is c_{2k+2} and $K'(h) = K(h) - c_{2k+2}|h|^{2k+2}$. By a trivial variation of Theorem 2.2 in Matheron (1973) and its proof, we will know $K'(h)/|h|^{2k+2} \to 0$ for $|h| \to \infty$. Therefore, $K(rh) \sim |h|^{2k+2}$ if and only if $c_{2k+2} \neq 0$. And thus, it is obvious that, when $K(h) \sim |h|^{2k+2}$, one must have $\alpha(z) = k+1$ and $\chi(du) = 0$ a.e.. Similarly, when $\alpha(z) < k+1$, one must have $c_{2k+2} = 0$.

By now we have proved our proposition. □

As the spectral measure has a self-similar structure, it becomes possible to rewrite the spectral representation in polar coordinate system as follows.

Proposition 2.5.7. *Assume the same conditions in Proposition 2.5.6. Additionally, assume X only takes value in $\mathbb{H}$ with real value scalar. When $\alpha(z) \notin \mathbb{Z}, \alpha \in (0, k+1)$, the generalized covariance of $Y_z(x)$ is*

$$K(h) = \int_{S^{d-1}} I(\alpha(z))|\langle \theta, h\rangle|^{2\alpha(z)}\sigma(d\theta) + Q(h); \tag{2.44}$$

when $\alpha(z) \in (0, k] \cap \mathbb{Z}$,

$$K(h) = \int_{S^{d-1}} P'_{\alpha(z)}(|\langle \theta, h\rangle|)\log|\langle \theta, h\rangle|\sigma(d\theta) + Q(h), \tag{2.45}$$

where

$$I(\alpha) = \int_0^\infty \frac{\cos(r) - P_{\lfloor\alpha\rfloor}(r)}{r^{2\alpha+1}}dr,$$

$P'_k(x)$ is the $k+1$th term of Taylor expansion of $\cos(x)$ at 0, $P_k(x) = \sum_{l=0}^k P'_k(x)$, $Q(h)$ is k-conditionally positive definite polynomial function up to order $2k$, $B = B_1(0)$, S_{d-1} is the sphere of a d-dimensional unit ball and $\sigma(\cdot)$ is a unique positive symmetric metric on S_{d-1}.

Remark. *When $\mathbb{H}$ is a separable Hilbert space with scalar in $\mathbb{C}$, it is not hard to write the spectral measure in polar coordinate system with similar derivations, but the result is more complicated and may not be very useful in practice.*

Proof. We only need to reconstruct the canonical form in (2.40) to the form we used in the proposition and discuss the special case for $\alpha(z) \in \mathbb{Z} \cap [1, k]$. As X only takes value in $\mathbb{H}$ with real scalar, we have the spectral measure of its generalized covariance to be symmetric.

For $u \in \mathbb{R}^d \backslash \{0\}$, denote $\theta = \frac{u}{\|u\|}$ and $r = \|u\|$, and by (2.41), one can verify that

$$\frac{\chi(du)}{1 \wedge |u|^{2k+2}} = r^{-(2\alpha(z)+1)} dr\sigma(d\theta),$$

where σ is some finite measure on S^{d-1}. Denote $\phi = \langle \theta, h \rangle$ and take $B = B_1(0)$. It follows

$$\begin{aligned}
K(h) &= \int_{S^{d-1}} \int_0^\infty \frac{\cos(r\phi) - 1_B(r) P_k(r\phi)}{r^{2\alpha(z)+1}} dr\sigma(d\theta) + Q(h) \\
&= \int_{S^{d-1}} \int_0^\infty \frac{\cos(\widetilde{r}) - 1_B(\widetilde{r}/|\phi|) P_k(\widetilde{r})}{\widetilde{r}^{2\alpha(z)+1}} |\phi|^{2\alpha(z)} d\widetilde{r}\sigma(d\theta) + Q(h) \\
&= \int_{S^{d-1}} \int_0^\infty \frac{\cos(r) - 1_B(r) P_k(r)}{r^{2\alpha(z)+1}} |\phi|^{2\alpha(z)} dr\sigma(d\theta) + Q(h) \\
&\quad + \int_{S^{d-1}} \int_0^\infty \frac{(1_B(r) - 1_B(r/|\phi|)) P_k(r)}{r^{2\alpha(z)+1}} |\phi|^{2\alpha(z)} dr\sigma(d\theta).
\end{aligned}$$

where the second line uses changing variables $\widetilde{r} = r|\phi|$ and the fact that $P_k(\cdot)$ is an even function.

If $\alpha(z) \notin \mathbb{Z}$ and $l = 0, \ldots, k$, we will have each term in the third part of the last line in the previous equation to be

$$\begin{aligned}
&\int_{S^{d-1}} \int_0^\infty \frac{(1_B(r) - 1_B(r/|\phi|)) r^{2l}}{r^{2\alpha(z)+1}} |\phi|^{2\alpha(z)} dr\sigma(d\theta) \\
= &- \int_{S^{d-1}} \int_1^{|\phi|} \frac{|\phi|^{2\alpha(z)}}{r^{2\alpha(z)-2l+1}} dr\sigma(d\theta) \\
= &\frac{1}{2\alpha(z) - 2l} \int_{S^{d-1}} (|\phi|^{-2\alpha(z)+2l} - 1) |\phi|^{2\alpha(z)} \sigma(d\theta) \\
= &\, C|h|^{2l} + \frac{1}{2l - 2\alpha(z)} \int_{S^{d-1}} |\langle \theta, h \rangle|^{2\alpha(z)} \sigma(d\theta).
\end{aligned}$$

If $\alpha(z) \in Z$ and $l = \alpha(z)$, by similar calculations as above, we will have

$$\begin{aligned}
&\int_{S^{d-1}} \int_0^\infty \frac{(1_B(r) - 1_B(r/|\phi|)) r^{2l}}{r^{2\alpha(z)+1}} |\phi|^{2\alpha(z)} dr\sigma(d\theta) \\
= &- \int_{S^{d-1}} \int_1^{|\phi|} |\phi|^{2\alpha(z)} / r dr\sigma(d\theta) \\
= &- \int_{S^{d-1}} |\phi|^{2\alpha(z)} \log |\phi| \sigma(d\theta).
\end{aligned}$$

Remark. *It may not be easy to see why the last term is still self-similar up to a polynomial*

of h. To verify it, we can have for any $c > 0$ and $\alpha(z) = l \in \mathbb{Z}$,

$$\begin{aligned} & \int_{S^{d-1}} |c\phi|^{2l} \log |c\phi| \sigma(d\theta) \\ = \; & c^{2l} \log(c) \phi^{2l} \int_{S^{d-1}} \sigma(d\theta) + c^{2l} \int_{S^{d-1}} \phi^{2l} \log(|\phi|) \sigma(d\theta), \end{aligned}$$

where the first term is a polynomial of h with order $2l$ and the second term represents the self-similarity with order $2l$. Notice that ϕ^{2l} can be absorbed into the polynomial with order no larger than $2k$ and we get our representation in the proposition.

Therefore, when $\alpha(z) \notin Z$,

$$K(h) = \int_{S^{d-1}} I_k(\alpha(z)) |\langle \theta, h \rangle|^{2\alpha(z)} \sigma(d\theta) + Q'(h),$$

where $Q'(h)$ is a k-conditionally positive definite polynomial function up to order $2k$ and

$$I_k(\alpha) = \int_0^\infty \frac{\cos(r) - 1_B(r) P_k(r)}{r^{2\alpha+1}} dr + \sum_{l=0}^{k} \frac{(-1)^l}{2l!(2l - 2\alpha)}. \tag{2.46}$$

If $l > \alpha$,

$$\int_0^1 r^{2l-2\alpha-1} dr = \frac{1}{2l - 2\alpha}; \tag{2.47}$$

if $l < \alpha$,

$$\int_1^\infty r^{2l-2\alpha-1} dr = -\frac{1}{2l - 2\alpha}. \tag{2.48}$$

Combine (2.46), (2.47) and (2.48), we obtain

$$I_k(\alpha) = \int_0^\infty \frac{\cos(r) - P_{\lfloor \alpha \rfloor}(r)}{r^{2\alpha+1}} dr.$$

One can see I_k is not related to k and thus we have the $I(\cdot)$ in our proposition. □

It is well known that fractional Brownian motion (fBm) is the only self-similar Gaussian process with stationary increment on $\mathbb{R}$. One may wonder whether the scalar valued tangent field, proved to be a self-similar IRF$_k$, will be similar to fBm. The remaining of this section will illustrate Gaussian tangent field when $E = \mathbb{R}$. Please refer to section 2.5.5 for $E = \mathbb{H}$, as it will be a special case of the discussion there.

It can be proved that the tangent field, if mean-sqaure continuous, is some generalization of fractional Brownian motion. We will illustrate it as follows.

In Perrin et al. (2001), a kth-order fraction Brownian motion (k-fBm) is defined in the following way:

$$B_H^{(k)}(t) = \frac{1}{\Gamma(H+\frac{1}{2})}\int_{\mathbb{R}}(t-s)_+^{H-1/2} + \sum_{l=0}^{k-1} C_l^{H-1/2}(-s)_+^{H-1/2-l}t^l dB(s),$$

where $C_\ell^r := -\frac{\prod_{i=0}^{\ell-1}(r-i)}{\ell!}$ for $\ell \geq 1$ and $C_0^r = -1$, or equivalently,

$$B_H^{(k)}(t) = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{1}{(i\omega)^{H+1/2}}\left[e^{it\omega} - \sum_{l=0}^{k-1}\frac{(it\omega)^l}{l!}\right] dW(\omega).$$

They proved that

- k-fBm is well defined for $H \in (k-1, k)$.
- The derivative of $B_H^{(k)}(t)$ is $B_{H-1}^{(k-1)}(t)$.

We prove that $B_H^{(k)}(t)$ is a self-similar IRF$_k$ as follows.

Fact 2.5.8. *For any $c \geq 0$, $t > 0$, $k \in \mathbb{N}_+$ and $H \in (k-1, k)$,*

$$B_H^{(k)}(c\lambda) \stackrel{d}{=} c^H B_H^{(k)}(\lambda), \text{ for any } \lambda \in \Lambda_{k'}, k' \in \mathbb{N}, \tag{2.49}$$

and

$$B_H^{(k)}(t+\lambda) \stackrel{d}{=} B_H^{(k)}(\lambda), \text{ for any } \lambda \in \Lambda_{k'}, k' \in \mathbb{N},\ k' \geq k-1. \tag{2.50}$$

Proof. To prove (2.49), we only need to prove $B_H^{(k)}(cx) \stackrel{d}{=} c^H B_H^{(k)}(x)$ and we can use induction. This property is obviously right for $B_H, H \in (0,1)$. Assume it holds for $B_H^{(k-1)}, H \in (k-2, k-1)$. Now, for any $H \in (k-1, k)$, we will have

$$\begin{aligned} B_H^{(k)}(cx) &= \int_0^{cx} B_{H-1}^{(k-1)}(s)ds = c\int_0^x B_{H-1}^{(k-1)}(cs)ds \\ &\stackrel{d}{=} c^{1-1+H}\int_0^x B_{H-1}^{(k-1)}(s)ds = c^H B_H^{(k)}(x). \end{aligned}$$

To prove (2.50), we will still use induction. Equation (2.50) holds for $k = 0$ because B_H has stationary increment. Assume it also holds for $B_H^{(k-1)}$. For any $\lambda = \sum_{i=1}^m c_i\delta_{x_i} \in$

$\Lambda_{k'}, k' \geq k-1 \geq 0$, we will have

$$\begin{aligned} B_H^{(k)}(t+\lambda) &= \sum_{i=1}^m c_i \int_0^{t+x_i} B_{H-1}^{(k-1)}(s)ds \\ &= \sum_{i=1}^m c_i \int_t^{t+x_i} B_{H-1}^{(k-1)}(s)ds = \int_0^1 B_{H-1}^{(k-1)}(\widetilde{\lambda}(s)+t)ds, \end{aligned}$$

where $\widetilde{\lambda}(s) = \sum_{i=1}^m c_i x_i \delta_{sx_i} \in \Lambda_{k'-1}$. Therefore,

$$\int_0^1 B_{H-1}^{(k-1)}(\widetilde{\lambda}(s)+t)ds \stackrel{d}{=} \int_0^1 B_{H-1}^{(k-1)}(\widetilde{\lambda}(s))ds = B_H^{(k)}(\lambda).$$

□

By Fact 2.5.8, we know that k-fBm is a H-order self-similar IRF$_{k-1}$. However, one problem around the k-fBm process is that H can only take value in $(k-1,k)$ to have the process well-defined, while our Propositions 2.5.6 and 2.5.7 suggest that a self-similar IRF$_{k-1}$ can have any self-similar order in $(0,k]$. Here, we present one way to define a self-similar IRF$_k$ with any self-similar index in $(0,k+1)$. (We will not include $k+1$ because it is trivially a random polynomial.)

By (2.3), given a set of points $t_1, \ldots, t_{M_k}$, for each t, we will have

$$\lambda_t = \delta_t - (\delta_{t_1}, \ldots, \delta_{t_{M_k}})B^{-1}\boldsymbol{t} \in \Lambda_k.$$

Now, the new process can be simply defined to be

$$\begin{aligned} \widetilde{B}_H^{(k+1)}(t) &:= B_H^{(k+1)}(\lambda_t) \\ &= \frac{1}{\Gamma(H+\frac{1}{2})}\int_{\mathbb{R}} (t-s)_+^{H-1/2} \\ &\quad -((t_1-s)_+^{H-1/2}, \cdots, (t_{M_k}-s)_+^{H-1/2})B^{-1}\boldsymbol{t}dB(s) \\ &= \frac{1}{2\pi}\int_{\mathbb{R}} \frac{1}{(i\omega)^{H+1/2}}\left[e^{it\omega} - (e^{it_1\omega}, \cdots, e^{it_{M_k}\omega})B^{-1}\boldsymbol{t}\right] dW(\omega). \end{aligned}$$

For any $\lambda \in \Lambda_k$, it follows that

$$\widetilde{B}_H^{(k+1)}(\lambda) = \frac{1}{\Gamma(H+\frac{1}{2})}\int_{\mathbb{R}} (\lambda-s)_+^{H-1/2} dB(s),$$

and thus, $\widetilde{B}_H^{(k+1)}(\lambda)$ can be trivially proved to be an H-order self-similar IRF$_k$. (It is only self-similar in Λ_k, but not in $\mathbb{R}$.)

To show $\widetilde{B}_H^{(k+1)}$ is well-defined for $H \in (0, k+1)$, we only need to prove the following lemma.

Lemma 2.5.9. *For any $t \in \mathbb{R}$, $H \in (0, k+1)$ and λ_t defined as in* (2.3)*, $f(s) := (\cdot - s)_+^{H-1/2}(\lambda_t) \in \mathbb{L}_2(\mathbb{R})$.*

Proof. Let $M = 2\max\{|t|, |t_1|, \cdots, |t_{M_k}|\}$. As $H > 0$, it is easy to see $2H - 1 > -1$ and thus $f(s) \in \mathbb{L}_2([-M, M])$. For $s > M$, we simply have $f(s) = 0$. For $s < -M$, we obtain

$$f(s) = (\cdot - s)^{H-1/2}(\lambda_t) = \varphi_{H-1/2,k}(s; \cdot)(\lambda_t),$$

where

$$\varphi_{r,k}(s; a) := (a - s)^r + \sum_{l=0}^{k} C_l^r (-s)^{r-l}(a)^l.$$

By Taylor theorem, for $s < -M$ and $M > a$, we have

$$|\varphi_{r,k}(s; a)| = \left| \int_0^a \frac{(a-u)^k}{k!} C_{k+1}^r (u - s)^{r-k-1} du \right| \leq C|s|^{r-k-1},$$

where C is related to $|a - M|, r, k$. Therefore, we can obtain a uniform constant C such that

$$\max_{a \in \{t, t_1, \cdots, t_{M_k}\}} |\varphi_{H-1/2,k}(s; a)| \leq C|s|^{H-k-3/2}.$$

As $H < k + 1$, we will have $2(H - k - 3/2) \leq -1$ and thus, $f(s) \in \mathbb{L}_2((-\infty, -M))$. Therefore, $f(s) \in \mathbb{L}_2(\mathbb{R})$. □

A further generalization is to define $\widetilde{B}_H^{(k+1)}$ on $\mathbb{R}^d$, and this can be simply done as

$$\widetilde{B}_H^{(k+1)}(t) \;\; := \;\; \int_{S^{d-1}} \int_{\mathbb{R}} (\langle \cdot, \theta \rangle - s)_+^{H-1/2}(\lambda_t) M(ds, d\theta),$$

or up to a constant scalar equivalently as,

$$\widetilde{B}_H^{(k+1)}(t) \;\; := \;\; \int_{\mathbb{R}^d} \frac{1}{(i\omega)^{H+d/2}} e^{i\langle \cdot, \omega \rangle}(\lambda_t) \widehat{W}(d\omega).$$

where $M(ds, d\theta)$ is an independently scattered Gaussian random measure on $\mathbb{R} \times S^{d-1}$ with control measure $ds\sigma(d\theta)$ for some finite and symmetric measure σ on the unit sphere, and

$\widehat{W}$ satisfies, for $\xi = r\theta, r > 0, \theta \in S_{d-1}$,

$$\mathbb{E}[|\widehat{W}(d\xi)|^2] = r^{d-1} dr\sigma(d\theta).$$

By Lemma 2.5.9, $\widetilde{B}_H^{(k+1)}$ above is well-defined for $H \in (0, k+1)$. Also, by its spectral representation, it is easy to see its general covariance function will be exactly the same as (2.40), which indicates it is the only self-similar IRF$_k$ in Gaussian up to scaling and random polynomial with degree no larger than k.

2.5.5 Characterization of tangent fields when $T_r(e_i) = r^{\alpha_i} \cdot e_i$

The scaling operator can be quite complex in general and a comprehensive result is out of the scope of this chapter. In this section, Assume T_r is linear operator on $\mathbb{H}$ and $T_r(e_i) = r^{\alpha_i} e_i$, with $\{e_i\}$ to be linear independent(may not be CONS) and $\mathbb{H} = \overline{\text{span}\{e_i\}}$. Notice that there always exists $\{y_i(s)\}$ such that

$$X(s,t) = \sum_i y_i(s) e_i(t).$$

Lemma 2.5.10. *Mean-square continuous $X(s,t)$ is T_r self-similar IRF$_k$ process, if and only if for any $r \geq 0$,*

$$(y_i(r \cdot \lambda + w), i = 1, \dots) \overset{d}{=} (r^{\alpha_i} y_i(\lambda), i = 1, \dots),$$

for any $\lambda \in \Lambda_k$ and $w \in \mathbb{R}^d$.

Proof. The 'if' part is trivial. For the other part, take any $e = \sum_i c_i e_i$, we will have

$$\sum_i c_i y_i(r \cdot \lambda + w) e_i = \langle X(r \cdot \lambda + w, \cdot), e \rangle \overset{d}{=} \langle T_r(X(\lambda, \cdot)), e \rangle = \sum_i c_i r^{\alpha_i} y_i(\lambda) e_i.$$

As e is arbitrary, we will obtain the conclusion in the lemma. □

Notice that y_i may not be independent. One can easily obtain that

$$\mathbb{E}[X(s_1, \cdot) \otimes X(s_2, \cdot)] = \sum_{i,j} \text{Cov}(y_i(s_1), y_j(s_2)) e_i \otimes e_j.$$

By Theorem 2.5.3, we will have

$$\mathbb{E}[X(s, \cdot) \otimes X(s+h, \cdot)] = \int \frac{\cos(u^T h) - 1_B(u) P_k(u^T h)}{1 \wedge |u|^{2k+2}} \sum_{j,k} \chi_{j,k}(du) e_j \otimes e_k + Q(h),$$

where $\chi_{j,k}(du) = \langle e_j, \chi(du)e_k\rangle$. By linear independence of $\{e_j \otimes e_k\}$, we will have $\chi_{j,j}$ coincides with the spectral measure for the generalized covariance of $y_j(s)$ in Theorem 2.5.6 and

$$\mathrm{Cov}(y_j(s), y_k(s+h)) = \int \frac{\cos(2\pi u^T h) - 1_B(u)P_k(2\pi u^T h)}{1 \wedge |u|^{2k+2}} \chi_{j,k}(du) + Q_{j,k}(h),$$

where $Q_{j,k}$ satisfies $Q(h) = \sum_{i,j} Q_{j,k} e_j \otimes e_k$ and each $\chi_{j,k}$ is unique by Theorem 2.5.3.

As $T_r(X(s)) \otimes T_r(X(t)) = T_r \circ (X(s) \otimes X(t)) \circ T_r^*$, by the same argument in the proof of Theorem 2.5.6, we will have for any Borel set A satisfying $\inf_{x\in A} d(x, 0) > 0$,

$$\int_A \frac{\chi(du)}{1 \wedge |u|^{2k+2}} = \int_{rA} \frac{T_r \circ \chi(du) \circ T_r^*}{1 \wedge |u|^{2k+2}},$$

or equivalently,

$$\int_A \frac{\chi_{j,k}(du)}{1 \wedge |u|^{2k+2}} = r^{\alpha_j+\alpha_k} \int_{rA} \frac{\chi_{j,k}(du)}{1 \wedge |u|^{2k+2}}.$$

It entails results similar to Proposition 2.5.7 but the form is complex and we omit it here.

Now, let's check the Gaussian case for those covariance structures. Let y_j to be the k-fBm we defined before, that is

$$y_j(t) = \int_{\mathbb{R}^d} \frac{1}{(iu)^{H+d/2}} e^{i\langle \cdot, u\rangle}(\lambda_t) \widehat{W}_j(du).$$

If we have

$$\mathbb{E}[\widehat{W}_j(du)\overline{\widehat{W}_k(du)}] = \chi_{j,k}(du), \tag{2.51}$$

then such kind of $X(s,t)$ will be the only Gaussian T_r-self-similar IRF$_k$ process up to scaling and random polynomial. Now, we show such kind of $\widehat{W}_j$ is achievable under mild conditions.

As $\chi(du)$ is T_+-valued Borel measure, there exists an eigen-decomposition of $\chi(du)$. Now, we further assume the eigen values are still Borel measures on $\mathbb{R}^d$, i.e.,

$$\chi(du) = \sum_i \lambda_i(du) e'_{i,u} \otimes e'_{i,u},$$

where $\lambda_i(du) \geq 0$ is a Borel measure, $\{e'_{i,u}\}$ is CONS for any fixed $u \in \mathbb{R}^d \setminus \{0\}$.

As $\{e_i\}$ is linear independent, for each e_i, it can be decomposed into $e_i^\perp$ and $e_i - e_i^\perp$ such that $e_i - e_i^\perp \in \overline{\mathrm{span}\{\{e_j, j = 1, \dots\} \setminus \{e_i\}\}}$ and $\langle e_i^\perp, e_j\rangle = 0, j \neq i$. WLOG, assume

$\|e_i^{\perp}\| = 1$, otherwise we can rescale e_i to achieve this.

Now, we can get $\widehat{W}_j$ in the following ways. We first obtain independent orthogonal random measure $\widehat{W}'_k$ with $\mathbb{E}[|\widehat{W}'_k(du)|^2] = \lambda_k(du)$. Then let

$$\widehat{W}_i = \sum_k C_{ik}(u)\widehat{W}'_k,$$

where $C_{ik}(u) = \langle e_i^{\perp}, e'_{k,u}\rangle$. Notice that

$$\chi_{i,j}(du) \quad = \quad \langle e_i^{\perp}, \chi(du)e_j^{\perp}\rangle = \sum_k \lambda_k(du)\langle e_i^{\perp}, e'_{k,u}\rangle\langle e'_{k,u}, e_j^{\perp}\rangle = \sum_k \lambda_k(du)C_{ik}(u)\overline{C_{jk}(u)}.$$

Therefore,

$$\mathbb{E}[\widehat{W}_i(du)\overline{\widehat{W}_j}(du)] = \sum_k \lambda_k(du)C_{ik}(u)\overline{C_{jk}(u)} = \chi_{i,j}(du).$$

2.6 Examples

In this section, we present a few random fields in Gaussian with tangent fields. The proof of the existence of tangent fields in the following examples are lengthy and provided in Section 2.8.6. Some other results about sufficient conditions of the existence of tangent fields are also supplied in that section.

As n-fBm is the representation of tangent field in Gaussian case, intuitively, if H and $\sigma(d\theta)$ are changing smoothly, the process may have tangent fields at almost every location. The following example shows this intuition is correct.

Example 2.6.1. *n-mBm.*

Consider

$$\widetilde{X}_H(t) = \int_{\mathbb{R}^d} \frac{e^{i\langle\omega,\lambda_t\rangle}}{|\omega|^{H(t)+d/2}} g_t\left(\frac{\omega}{|\omega|}\right) dW(\omega),$$

where $\mathbb{E}[|dW(\omega)|^2] = r^{d-1}drd\theta$. For some $l \in \mathbb{N}$, $H \in (0, l+1)$, $\lambda_t \in \Lambda_l$, we have $\widetilde{X}_H(t)$ is well-defined at location t. Assume $H \in C_{2[H(t)]+3}(V)$, where V is any bounded neighborhood around t; $g_s(\theta) \in C_{2[H(t)]+3}(V)$ with respect to s and for any given s, $g_s(\theta)$ is a bounded function. In Section 2.8.6, if $t \notin \{0, t_1, t_2, \ldots, t_{M_k}\}$, we can show that any order tangent field at location t exists. We can obtain the generalized covariance function

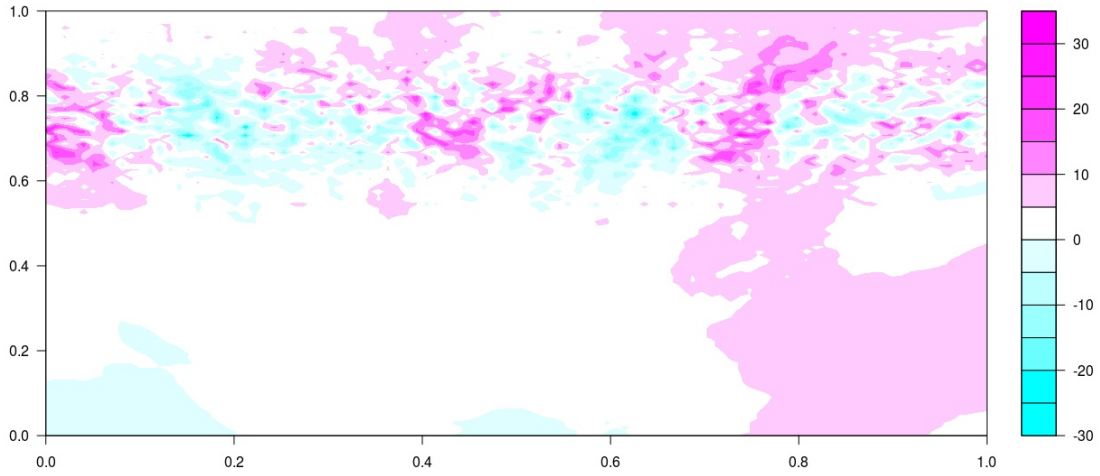


Figure 2.1: Simulation of $\widetilde{X}_H(t)$ on $\mathbb{R}^2$. $H(t) = 0.5 + 0.4\sin(2\pi t_2)$ and $g_t(\theta) = 2.5 + 2\sin(2\theta - t_1\pi/2)$.

to be

$$K(h) = \begin{cases} C_{z,k}|h|^{2l+2} & 0 \le k \le [H(t)] - 1, \\ C_{z,k}\int_{S^{d-1}} |\langle h, \theta\rangle|^{2H(t)} g_z(\theta) d\theta & k \ge [H(t)], H(t) \notin \mathbb{N}, \\ C_{z,k}\int_{S^{d-1}} |\langle h, \theta\rangle|^{2H(t)} \log(|\langle h, \theta\rangle|) g_z(\theta) d\theta & k \ge [H(t)], H(t) \in \mathbb{N}, \end{cases}$$

where k is the order of tangent field and $C_{z,k}$ is some constant. A quick and brutal simulation on $\mathbb{R}^2$ is given in Figure 2.1 for $H(t) = 0.5 + 0.4\sin(2\pi t_2)$ and $g_t(\theta) = 2.5 + 2\sin(2\theta - t_1\pi/2)$.

In the example below, we connect our theory with the famous Matérn class.

Example 2.6.2. *As a generalization of the famous Matérn class, Stein (2005) proposed a model with the following covariance function:*

$$C(x, y) = \frac{c(x)c(y)}{|\Sigma(x, y)|^{1/2}} \mathscr{M}_{\nu(x,y)}(Q(x, y)^{1/2}),$$

where $c(x)$ is a smooth function for normalizing, $\Sigma(x, y) = (\Sigma(x) + \Sigma(y))/2$, $\nu(x, y) = (\nu(x) + \nu(y))/2$, $\Sigma(x)$ is a mapping from $\mathbb{R}^d$ to the space of $d \times d$ positive definite matrix, $\mathscr{M}_\nu(x) = x^\nu \mathscr{K}_\nu(x)$, with $\mathscr{K}$ being a modified Bessel function of order ν and $Q(x, y) = (x - y)^T \Sigma(x, y)^{-1}(x - y)$.

Assume in a bounded neighborhood V of z, $\Sigma(x)$ and $\nu(x)$ are smooth function, eigenvalues of $\Sigma(x)$ bounded from 0 and ∞ and $\nu(x) \in (l, l + 1)$ for some $l \in \mathbb{N}$. We show in

Section 2.8.6 that it has any order of tangent field at location z with generalized covariance function to be

$$K(h) = \begin{cases} C_{z,k}|h|^{2l+2} & k < l, \\ C_{z,k}|\Sigma^{-1/2}(z)h|^{2\nu(z)} & k \geq l, \end{cases}$$

where k is the order of tangent field and $C_{z,k}$ is some constant. With the same argument in page 82, 83 of Samorodnitsky and Taqqu (1994), one can show for any Σ strictly symmetric positive definite there exists a σ such that

$$|\Sigma^{-1/2}h|^{2\nu} = \int_{S^{d-1}} |\langle \theta, h \rangle|^{2\nu} \sigma(d\theta).$$

Therefore, the generalized covariance function of this model still satisfies our Proposition 2.5.7. We still don't know the conclusion for z such that $\nu(z) \in \mathbb{Z}$ as the computation is too complex.

2.7 Potential inference methods

We provide two possible estimation methods for tangent fields. However, proofs of their properties are out of the scope of this thesis. Both estimation methods are based on periodogram kernel method in Panaretos and Tavakoli (2013).

2.7.1 Estimator for $T_r(X) = r \cdot X$

Assume $X(s,t) : \mathbb{R} \times \mathbb{R} \mapsto \mathbb{R}$ is observed on grid points of $s \in \mathbb{N}$. Let $X(s,\cdot)$ be a mean-square continuous self-similar IRF$_k$ with $T_r(X) = r \cdot X$. Assume the spectral measure of $X(s,\cdot)$ has a density, or equivalently, the covariance operator function of $X(\mu,\cdot)$ is integrable for any $\mu \in \Lambda_k$. Then we know its generalized covariance operator function has the following decomposition.

$$K(h) = \int_{\mathbb{R}} \frac{C_1(e^{i\omega h} - P(\omega h))}{|\omega|^{2\alpha+1}} d\omega + Q(h),$$

where $P(h)$ is the first $2k+1$ terms of Taylor expansion of e^{ih}, $Q(h)$ are polynomial function up to order $2k+2$, C_1 is some trace class operator and $\alpha \in (0, k+1)$.

Then, for any $\mu \in \Lambda_l$, $l \geq k$, we will have

$$C_\mu(h) = C_1 \int_{\mathbb{R}} e^{i\omega h} |\widehat{\mu}(\omega)|^2 |\omega|^{-2\alpha-1} d\omega + C_2,$$

where C_2 is another trace class operator.

Notice that we only observe data on grid points and thus the spectral measure at locations $\{\omega + 2k\pi\}_{k\in\mathbb{Z}}$ are not distinguishable. In fact, if supporting points of μ are on $\mathbb{Z}$, we will have

$$\begin{aligned} C_\mu(n) &= C_1 \int_0^{2\pi} e^{i\omega n} \sum_{j\in\mathbb{Z}} \frac{|\widehat{\mu}(\omega + 2j\pi)|^2}{|\omega + 2j\pi|^{2\alpha+1}} d\omega + C_2 \\ &= C_1 \int_0^{2\pi} e^{i\omega n} |\widehat{\mu}(\omega)|^2 \sum_{j\in\mathbb{Z}} \frac{1}{|\omega + 2j\pi|^{2\alpha+1}} d\omega + C_2,\ n \in \mathbb{Z}, \end{aligned}$$

where the second equality is due to the fact that for any point measure μ with supporting points on $\mathbb{Z}$,

$$|\widehat{\mu}(\omega + 2j\pi)|^2 = |\widehat{\mu}(\omega)|^2, j \in \mathbb{Z}.$$

Suppose we are using T observations to construct the estimator.(In the case of tangent field, T can be considered as a window width.) For some $\omega \in [0, 2\pi)$, denote

$$\widetilde{X}^{(T)}_{\omega,\mu} = (2\pi T)^{-1/2} \sum_{t=0}^{T-1} X(\mu, \cdot) \exp(-\mathrm{i}\omega t),$$

and

$$p^{(T)}_{\omega,\mu} = \widetilde{X}^{(T)}_{\omega,\mu} \otimes \widetilde{X}^{(T)}_{\omega,\mu}.$$

By Proposition 2.6 in Panaretos and Tavakoli (2013), we will have uniformly for any $j \in \mathbb{N}$ satisfying $j \not\equiv 0 \bmod T$,

$$\mathbb{E}\left[p^{(T)}_{2\pi j/T,\mu}(t_1, t_2)\right] = f_{2\pi j/T,\mu}(t_1, t_2) + O(T^{-1}) \quad \text{in } L^2,$$

where

$$f_{\omega,\mu} = |\widehat{\mu}(\omega)|^2 \sum_{j\in\mathbb{Z}} \frac{C_1}{|\omega + 2j\pi|^{2\alpha+1}}.$$

Suppose we have calculated a bunch of $p^{(T)}_{2\pi j/T,\mu}$ for different $j \in S$, where $S \subset$

$\{1, 2, \ldots, T-1\}$ is a set of integers. To estimate α, we can calculate

$$\widetilde{p}^{(T)}_{2\pi j/T,\mu} = \log\left(\frac{p^{(T)}_{2\pi j/T,\mu}}{|\widehat{\mu}(2\pi j/T)|^2}\right) - \frac{1}{|S|}\sum_{l\in S}\log\left(\frac{p^{(T)}_{2\pi l/T,\mu}}{|\widehat{\mu}(2\pi l/T)|^2}\right).$$

Notice that for any $t_1, t_2 \in \mathbb{R}$,

$$\mathbb{E}[\widetilde{p}^{(T)}_{2\pi j/T,\mu}(t_1,t_2)] \approx g_\alpha(2\pi j/T) := \log\left(\sum_{l\in\mathbb{Z}}\frac{1}{|2\pi j/T + 2l\pi|^{2\alpha+1}}\right). \tag{2.52}$$

Therefore, to estimate α, one can minimize the following objective function by varying α,

$$\sum_{j\in S}\int_{\mathbb{R}^2} L\left(\widetilde{p}^{(T)}_{2\pi j/T,\mu}(t_1,t_2) - g_\alpha(2\pi j/T)\right)dt_1dt_2,$$

where L is any appropriate loss function.

After estimating α as $\widehat{\alpha}$, we can estimate C_1 as

$$\widehat{C}_1(t_1,t_2) = \frac{1}{|S|}\sum_{j\in S}\frac{p^{(T)}_{2\pi j/T,\mu}}{|\widehat{\mu}(2\pi j/T)|^2\exp\{g_{\widehat{\alpha}}(2\pi j/T)\}}.$$

Notice that for both $\widehat{\alpha}$ and $\widehat{C}_1$, one can use different μ's and take appropriate balances between them. We do a quick simulation for this estimator as follows. Consider the following random field to be estimated

$$X(s,t) = Y_1(s)e_1(t) + Y_2(s)e_2(t),$$

where Y_1, Y_2 are independent standard fractional Brownian motions with a common Hurst index $H = 0.4$, $e_1(t) = \cos(2\pi t)$ and $e_2(t) = \sin(2\pi t)$. We simulate with $s = 1, 2, \ldots, 3000$ and $t = 1/100, 2/100, \ldots, 1$. For simplicity, we take the loss function L to be square loss function and the optimization problem can be solved easily as a linear regression without intercept. We take $\mu = \delta_1 - \delta_0$ and thus $|\widehat{\mu}(\omega)|^2 = 2 - 2\cos(\omega)$.

For a single run, we get $\widehat{\alpha} = 0.401$ and the estimated $C_1 = e_1 \otimes e_1 + e_2 \otimes e_2$ are given in Figure 2.2.

2.7.2 Estimator when $T_r(e_i) = r^{\alpha_i}\cdot e_i$

In this section, we only consider the model $Y(s,t) = \sum_i X_i(s)e_i(t)$, where $X_i(s)$'s are independent and e_i's orthogonal. We still assume observations are on grid points and there-

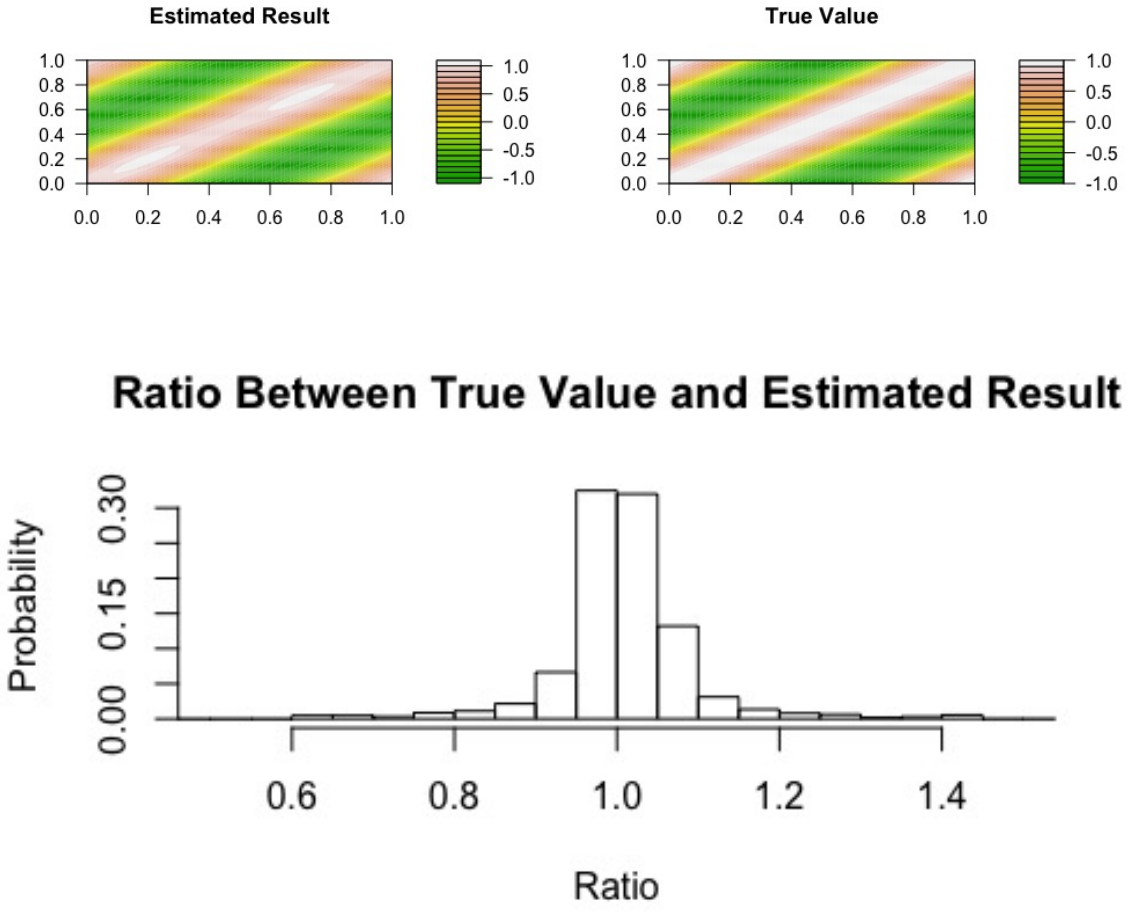


Figure 2.2: Comparison of $\widehat{C}_1$ and C_1. The first two plots are estimated and true C_1 separably. The third plot is the ratio (True Value)/(Estimated Value) and one can see most of the values fall around 1.

fore,

$$C_\mu(n) = \int_0^{2\pi} e^{i\omega n}|\widehat{\mu}(\omega)|^2 \sum_i \sum_{j\in\mathbb{Z}} \frac{c_i}{|\omega + 2j\pi|^{2\alpha_i+1}} e_i \otimes e_i d\omega + C_2,\ n \in \mathbb{Z},$$

where $c_i \in \mathbb{R}_+$. We still calculate a bunch of $p^{(T)}_{2\pi j/T,\mu}$ for different $j \in S$ as above. Notice that

$$\mathbb{E}\left[\frac{1}{|S|}\sum_{j\in S}\frac{p^{(T)}_{2\pi j/T,\mu}}{|\widehat{\mu}(2\pi j/T)|^2}\right] \approx \frac{1}{|S|}\sum_i \left(\sum_j \sum_{l\in\mathbb{Z}} \frac{1}{|2\pi j/T + 2l\pi|^{2\alpha_i+1}}\right) c_i e_i \otimes e_i.$$

Intuitively, we can do an eigen-decomposition of $\frac{1}{|S|}\sum_{j\in S}\frac{p^{(T)}_{2\pi j/T,\mu}}{|\widehat{\mu}(2\pi j/T)|^2}$ to estimate $e_i, i = 1, \ldots,$ denoted as $\widehat{e}_i$, then we can obtain

$$\widehat{\lambda}_{i,\mu}(w) = \langle \widehat{e}_i, p^{(T)}_{\omega,\mu}\widehat{e}_i\rangle.$$

Notice that

$$\widehat{\lambda}_{i,\mu}(w) \approx c_i \sum_{l\in\mathbb{Z}} \frac{1}{|2\pi j/T + 2l\pi|^{2\alpha_i+1}}.$$

Similar to the previous estimator, let

$$\widetilde{\lambda}_{i,\mu}(2j\pi/T) = \log(\widehat{\lambda}_{i,\mu}(2j\pi/T)) - \frac{1}{|S|}\sum_{j\in S}\log(\widehat{\lambda}_{i,\mu}(2j\pi/T)),$$

and minimize the following objective function for varying α_i

$$\sum_i \sum_{j\in S} L\left(\widetilde{\lambda}_{i,\mu}(2j\pi/T) - g_{\alpha_i}(2\pi j/T)\right),$$

where L is any appropriate loss function and g_α is defined in (2.52). This will give us the estimation of α_i as $\widehat{\alpha}_i$. Finally, we can obtain etimation of c_i as

$$\widehat{c}_i = \frac{1}{|S|}\sum_{j\in S}\frac{\widehat{\lambda}_{i,\mu}(2j\pi/T)}{\exp\{g_{\widehat{\alpha}_i}(2\pi j/T)\}}.$$

We do a quick simulation here with almost the same setting in the previous section except changing $\alpha = 0.4$ to $\alpha_1 = 0.3$ and $\alpha_2 = 0.6$ for Y_1, Y_2 respectively. We implement the method above with the loss function to be the square loss function and obtain $\widehat{\alpha}_1 =$

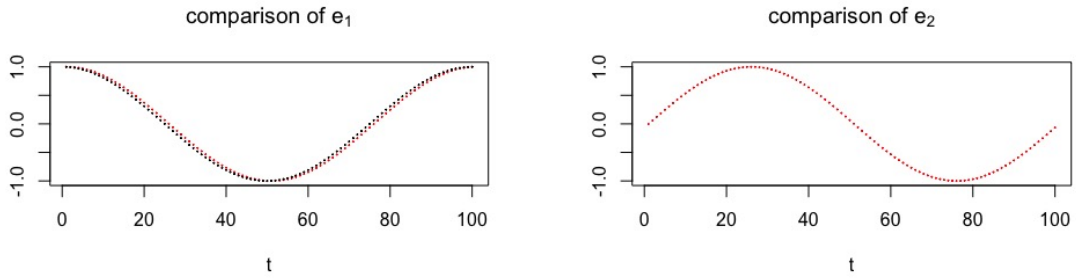


Figure 2.3: Comparisons of estimated and true $e_i, i = 1, 2$ after appropriate normalization. The black points are true values while the red are estimated values.

0.312, $\widehat{\alpha}_2 = 0.607$, $\widehat{c}_1 = 1.038$ and $\widehat{c}_2 = 1.108$. Other eigenvalues are much more smaller than the first two eigenvalues and are ignored as noise. Estimated e_1 and e_2 are given in figure 2.3.

2.8 Appendix

2.8.1 Integration with respect to operator-valued operators

For a $\mathbb{T}_+$-valued measure μ, define the positive and negative parts of μ, denoted by μ_+, μ_-, respectively, so that, for any Borel set A, $\mu(A) = \mu_+(A) - \mu_-(A)$ where $\mu \pm (A) \geq 0$.

Below we develop integration of a real or complex valued Borel measurable function with an operator valued measure μ. We follow the development of ordinary Lebesgue integration:

(i) For any real nonnegative simple function $f = \sum_{i=1}^k c_i I_{A_i}$, define $\int f d\mu = \sum_{i=1}^k c_i \mu(A_i)$.

(ii) For nonnegative measurable functions f, let

$$\int f d\mu = \lim_{n\to\infty} \int f_n d\mu \tag{2.53}$$

in $(\mathbb{T}, \|\cdot\|, \|_{tr})$ where $\{f_n\}$ is any sequence of simple functions such that

(a) $f_n \leq f_{n+1}$,

(b) $f_n(x) \uparrow f(x)$ for all x,

(c) $\int f_n d\mu \leq \mathscr{B}$ for all n for some $\mathscr{B} \in \mathbb{T}$.

We know how to construct a sequence $\{f_n\}$ satisfying (a) and (b) for any given f from standard measure theory. However, here we need the extra condition (c) to

ensure that the rhs of (2.53) is well defined: the limit exists and does not depend on the choice of $\{f_n\}$. This is done in Propositions 2.8.2 below.

(iii) For a general real measurable f, let

$$\int f d\mu = \int f_+ d\mu - \int f_- d\mu$$

provided one of the two terms on the right is finite. For a general complex f, let

$$\int f d\mu = \int f_{\rm re} d\mu + i \int f_{\rm im} d\mu$$

where $f_{\rm re}, f_{\rm im}$ be the real and imaginary parts, respectively. For a signed measure μ,

$$\int f d\mu = \int f d\mu_+ - \int f d\mu_-$$

where μ_+, μ_- be the positive and negative parts of μ, respectively.

Proposition 2.8.1. *Let $\mathscr{B}$ and $\mathscr{T}_n, n \geq 1$, be self-adjoint and nonnegative definite operators in $\mathbb{T}$ such that $\mathscr{T}_n \leq \mathscr{T}_{n+1} \leq \mathscr{B}$ for all n. Then $\mathscr{T}_n$ converges to a limit in $\mathbb{T}$.*

Proof. For $j \geq i$,

$$\|\mathscr{T}_i - \mathscr{T}_j\|_{tr} = \mathrm{tr}(\mathscr{T}_i - \mathscr{T}_j) = \|\mathscr{T}_i\|_{tr} - \|\mathscr{T}_j\|_{tr}.$$

Now, $\|\mathscr{T}_n\|_{tr}$ is nondecreasing and bounded by $\|\mathscr{B}\|_{tr}$ for all n. Hence, $\|\mathscr{T}_n\|_{tr}$ converges to some finite nonnegative limit. Consequently,

$$\lim_{n\to\infty} \sup_{j \geq i \geq n} \|\mathscr{T}_i - \mathscr{T}_j\|_{tr} = 0.$$

This shows that $\{\mathscr{T}_i\}$ is Cauchy and has a limit by the completeness of $\mathbb{T}$. □

Proposition 2.8.2. *Let f_n be a sequence of nonnegative simple functions with $f_n \leq f_{n+1}$ for each n and such that $\int f_n d\mu \leq \mathscr{B}$ for some $\mathscr{B} \in \mathbb{T}$.*

(a) $\int f_n d\mu$ converges in the space of trace class operators.

(b) If $\lim_{n\to\infty} f_n(x) \geq g(x)$ for all x for some simple function g, then $\lim_{n\to\infty} f_n d\mu \geq \int g d\mu$.

Proof. Note that nonnegative simple functions $g \leq f$ can be formulated as $f = \sum_{i=1}^{k} c_i I_{A_i}$, $g = \sum_{i=1}^{k} d_i I_{A_i}$ where $0 \leq c_i \leq d_i$ and the A_i are disjoint. Then it is easy to conclude that

$$\int g d\mu = \sum_{i=1}^{k} c_i \mu(A_i) \leq \sum_{i=1}^{k} d_i \mu(A_i) = \int f d\mu.$$

Thus, part (a) follows readily from Proposition 2.8.1.

To prove (b), let $g = \sum_{i=1}^{k} d_i I_{A_i}$, $A = \cup_{i=1}^{k} A_i$, and

$$E_n = \{x : f_n(x) + \epsilon > g(x)\}$$

for some fixed $\epsilon > 0$. Since

$$f_n \geq f_n I_{E_n \cap A} \geq (g - \epsilon) I_{E_n \cap A},$$

we have

$$\begin{aligned} \int f_n d\mu &\geq \int (g - \epsilon) I_{E_n \cap A} d\mu \\ &\geq \int g I_{E_n \cap A} d\mu - \epsilon \mu(A) \\ &= \sum_{i=1}^{k} c_i \mu(E_n \cap A_i) - \epsilon \mu(A). \end{aligned}$$

Letting $n \to \infty$, by the fact $E_n \cap A_i \uparrow A_i$,

$$\lim_{n \to \infty} \int f_n d\mu \geq \sum_{i=1}^{k} c_i \mu(A_i) - \epsilon \mu(A) = \int g d\mu - \epsilon \mu(A)$$

Since $\epsilon > 0$ is arbitrary, (b) follows. □

Proposition 2.8.3. *Let f be a nonnegative measurable function and $\{f_n\}$ be an increasing sequence of nonnegative simple functions satisfying (a)-(c) above. Then the limit $\lim_{n \to \infty} \int f_n d\mu$ does not depend on the particular sequence $\{f_n\}$.*

Proof. Suppose there are two sequences $\{f_n\}$ and $\{g_n\}$ of simple functions both satisfying (a)-(c). Then we have

$$\lim_{n \to \infty} f_n \geq g_m, \ \lim_{n \to \infty} g_n \geq g_m$$

for all m, and, by Proposition 2.8.2,

$$\lim_{n\to\infty}\int f_n d\mu \geq \int g_m d\mu,\ \lim_{n\to\infty}\int g_n d\mu \geq \int f_m d\mu.$$

The result follows by letting $m \to \infty$. □

Remark. Note that the condition (c) is readily fulfilled so long as f is bounded since $\mu(\mathbb{R})$ is assumed to be trace class. This is what we need in our application.

We now turn to Bochner's integrals. Below give a concise, but rigorous and self-contained, treatise concerning Bochner's integral. Additional details on Bochner's integrals can be found in texts such as Diestel and Uhl (1976), Dunford and Schwartz (1988) and Yosida (1980).

Suppose now that we have a function f on a measure space $(E, \mathscr{B}, \mu)$ that takes on values in a Banach space $\mathbb{X}$ with norm $\|\cdot\|$. In our application, $E = \mathbb{R}, \mathbb{X} = \mathbb{T}$ and μ is Lebesgue on $\mathbb{R}$. As usual, we start with simple functions.

Definition 2.8.4. *A function $f : E \to \mathbb{X}$ is called simple if it can be represented as*

$$f(\omega) = \sum_{i=1}^{k} I_{E_i}(\omega) g_i \tag{2.54}$$

for some finite k, $E_i \in \mathscr{B}$ and $g_i \in \mathbb{X}$.

Definition 2.8.5. *Any simple function $f(\omega) = \sum_{i=1}^{k} I_{E_i}(\omega) g_i$ with $\mu(E_i) < \infty$ for all i is said to be integrable and its Bochner integral is defined as*

$$\int_E f d\mu = \sum_{i=1}^{k} \mu(E_i) g_i. \tag{2.55}$$

It is not difficult to verify that this definition does not depend on the particular representation of f. In particular, the E_i can be chosen without loss to be disjoint. The previous definition is extended to a general measurable function from E to $\mathbb{X}$ as follows.

Definition 2.8.6. *A measurable function f is said to be Bochner integrable if there exists a sequence $\{f_n\}$ of simple and Bochner integrable functions such that*

$$\lim_{n\to\infty}\int_E \|f_n - f\| d\mu = 0. \tag{2.56}$$

In this case the Bochner integral of f is defined as

$$\int_E f d\mu = \lim_{n\to\infty} \int_E f_n d\mu. \tag{2.57}$$

To see that this definition has merit first observe from (2.55) and the triangle inequality that

$$\left\| \int_E f d\mu \right\| \leq \int_E \|f\| d\mu \tag{2.58}$$

for any simple function f. When applied to the simple function $f_n - f_m$ in (2.56) this inequality produces

$$\left\| \int_E f_n d\mu - \int_E f_m d\mu \right\| \leq \int_E \|f_n - f_m\| d\mu.$$

However, this upper bound is just a Lebesgue integral and the triangle inequality assures us that

$$\int_E \|f_n - f_m\| d\mu \leq \int_E \|f - f_n\| d\mu + \int_E \|f - f_m\| d\mu$$

which converges to zero as $m, n \to \infty$ by (2.56). This shows that $\{\int_E f_n d\mu\}$ is a Cauchy sequence and the completeness of $\mathbb{X}$ can now be invoked to conclude that the limit in (2.57) must exist. It is also independent of the approximating sequence since we can combine two approximating sequences into a third that must also be convergent.

The definition of the Bochner integral relies on the existence of an approximating sequence of simple and Bochner integrable functions $\{f_n\}$. This condition is easily verified if f is continuous, as we could construct step functions f_n. However, it may be difficult to verify for a more general setting. In the following we provide a sufficient condition that is applicable for many situations, including ours.

Theorem 2.8.7. *Let f be a measurable function from E to $\mathbb{X}$ with*

$$\int_E \|f\| d\mu < \infty. \tag{2.59}$$

Suppose that for each n there exists a finite-dimensional subspace $\mathbb{X}_n$ of $\mathbb{X}$ such that

$$\lim_{n\to\infty} \int_E \|f - g_n\| d\mu = 0 \tag{2.60}$$

for some measurable g_n taking value in $\mathbb{X}_n$. Then, there exist simple and Bochner integrable functions f_n such that (2.56) holds.

Proof. Define

$$\tilde{\mathbb{X}}_n = \mathbb{X}_n \cap \left\{ g \in \mathbb{X} : \|g\| \in [n^{-1}, n] \right\}$$

and

$$E_n = \left\{ t \in E : g_n(t) \in \tilde{\mathbb{X}}_n \right\}.$$

Markov's inequality produces

$$\mu(E_n) \leq n \int_{E_n} \|g_n\| d\mu \leq n \int_E \|g_n\| d\mu < \infty. \tag{2.61}$$

Since $\tilde{\mathbb{X}}_n$ is bounded and finite-dimensional, it is totally bounded. Thus, the Heine-Borel theorem can be invoked to see that there is a finite partition $B_i, 1 \leq i \leq k$, of $\tilde{\mathbb{X}}_n$ such that each B_i is in the Borel σ-field for $\mathbb{X}$ and has diameter less than $\epsilon = (n\mu(E_n))^{-1}$. For an arbitrary element $b_i \in B_i$ set

$$f_n(t) = \sum_{i=1}^{k} b_i I_{\{g_n(t) \in B_i\}}.$$

Note that $f_n = 0$ on E_n^c and hence (2.61) entails that f_n is simple and Bochner integrable. Also, by construction.

$$\|f_n(t) - g_n(t)\| \leq \max_{1 \leq i \leq k} \sup_{x \in B_i} \|b_i - x\| \leq (n\mu(E_n))^{-1} \tag{2.62}$$

for $\omega \in E_n$. Thus, by the triangle inequality,

$$\begin{aligned} & \int_E \|f_n - f\| d\mu \\ \leq & \int_E \|f_n - g_n\| d\mu + \int_E \|g_n - f\| d\mu \\ = & \int_{E_n} \|f_n - g_n\| d\mu + \int_{E_n^c} \|f_n - g_n\| d\mu + \int_E \|g_n - f\| d\mu. \end{aligned}$$

The first and third terms in the last expression tend to zero by (2.62) and (2.60), respectively, while the second term reduces to $\int_{E_n^c} \|g_n\| d\mu$ because $f_n = 0$ on E_n^c. Thus, to verify (2.56), we only need to establish that

$$\lim_{n \to \infty} \int_{E_n^c} \|g_n\| d\mu = 0.$$

Since

$$\|g_n\| \leq \|g_n - f\| + \|f\|,$$

using the assumption (2.60) it suffices to show

$$\lim_{n\to\infty} \int_E \|f\| I(\|g_n\| > n) d\mu = 0 \tag{2.63}$$

and that

$$\lim_{n\to\infty} \int_E \|f\| I(\|g_n\| < n^{-1}) d\mu = 0. \tag{2.64}$$

First, from (2.60) and Markov's inequality,

$$\mu(\|g_n\| > n) \leq n^{-1} \int_E \|g_n\| d\mu \to 0.$$

Since $\|f\|$ is integrable, (2.63) follows easily (cf. Exercise 5.6 of Resnick (1999)) from this relation. To show (2.64), for $\delta > 0$ write

$$\begin{aligned} \int_E \|f\| I(\|g_n\| < n^{-1}) d\mu &= \int_E \|f\| I(\|g_n\| < n^{-1}, \|f\| > \delta) d\mu \\ &\quad + \int_E \|f\| I(\|g_n\| < n^{-1}, \|f\| \leq \delta) d\mu \end{aligned}$$

so that

$$\begin{aligned} \limsup_{n\to\infty} \int_E \|f\| I(\|g_n\| < n^{-1}) d\mu &\leq \limsup_{n\to\infty} \int_E \|f\| I(\|g_n\| < n^{-1}, \|f\| > \delta) d\mu \\ &\quad + \int_E \|f\| I(\|f\| \leq \delta) d\mu. \end{aligned}$$

The first term on the right of the inequality is zero since

$$\begin{aligned} \mu(\|g_n\| < n^{-1}, \|f\| > \delta) &\leq \frac{\int_E \|f - g_n\| I(\|g_n\| < n^{-1}, \|f\| > \delta) d\mu}{\delta - n^{-1}} \\ &\leq \frac{\int_E \|f - g_n\| d\mu}{\delta - n^{-1}} \to 0 \end{aligned}$$

due to the triangle inequality $\|f - g_n\| \geq \|f\| - \|g_n\|$, Markov's inequality and (2.60). Hence, (2.64) follows now by letting $\delta \to 0$ and applying Lebesgue's dominated convergence theorem. □

Theorem 2.8.7 can be quite useful for proving existence of Bochner's integrals in general. For instance, suppose a measurable function f takes values in some separable Banach space $\mathbb{X}$ and $\int \|f(t)\| d\mu(t) < \infty$. By letting g_n be the projection of f onto the space spanned by the first n basis functions, the existence of the Bochner's integral $\int f d\mu$ in $\mathbb{X}$ follows at once from dominated convergence theorem and Theorem 2.8.7. In the same vein, the following corollary follows from Lemma 2.8.16 and Theorem 2.8.7.

Corollary 2.8.8. *Let $\mathscr{T}(t)$ be measurable from E to $\mathbb{T}$. If $\int_E \|\mathscr{T}(t)\|_{tr} d\mu(t) < \infty$, then the integral $\int_E \mathscr{T}(t) d\mu(t)$ is well defined as a Bochner's integral in $\mathbb{T}$.*

Another connection with Lebesgue integration is the following Bochner integral version of the dominated convergence theorem.

Theorem 2.8.9. *Let $\{f_n\}$ be a sequence of Bochner integrable functions in $\mathbb{X}$ that converges to some $f \in \mathbb{X}$. If there is a nonnegative Lebesgue integrable function g such that $\|f_n\| \leq g$ for all n a.e. μ, then f is Bochner integrable and $\int_E f d\mu = \lim_{n\to\infty} \int_E f_n d\mu$.*

Proof. In combination $\|f - f_n\| \leq 2g$ and $\|f - f_n\| \to 0$ allow us to apply the Lebesgue dominated convergence theorem to obtain $\int_E \|f - f_n\| d\mu \to 0$. Since the f_n are Bochner integrable we may find simple functions $\tilde{f}_n$ such that $\int_E \|f_n - \tilde{f}_n\| d\mu \to 0$. These new functions satisfy (2.56) because

$$\int_E \|f - \tilde{f}_n\| d\mu \leq \int_E \|f - f_n\| d\mu + \int_E \|f_n - \tilde{f}_n\| d\mu$$

and the theorem has been proved. □

The Bochner integral also has a feature similar to the monotonicity of the Lebesgue integral: namely,

Theorem 2.8.10. *If f is Bochner integrable, then $\| \int_E f d\mu\| \leq \int_E \|f\| d\mu$.*

Proof. Let $f_n = \sum_{i=1}^n g_i I_{E_i}(\omega)$ be a simple function with $E_i \cap E_j = \phi$ for $i \neq j$. Then,

$$\begin{aligned} \left\| \int_E f_n(\omega) d\mu \right\| &= \left\| \sum_{i=1}^n g_i \mu(E_i) \right\| \\ &\leq \sum_{i=1}^n \|g_i\| \mu(E_i) = \int_E \|f_n\| d\mu. \end{aligned}$$

So, if $\{f_n\}$ is a sequence of simple Bochner integrable functions that satisfies (2.56),

$$\begin{aligned}\left\|\int_E f d\mu\right\| &\leq \left\|\int_E f d\mu - \int_E f_n d\mu\right\| + \left\|\int_E f_n d\mu\right\| \\ &\leq \left\|\int_E f d\mu - \int_E f_n d\mu\right\| + \int_E \|f_n\| d\mu \\ &\leq \left\|\int_E f d\mu - \int_E f_n d\mu\right\| + \int_E \|f_n - f\| d\mu + \int_E \|f\| d\mu\end{aligned}$$

and the result follows upon taking limits with respect to n. □

Theorem 2.8.11. *Suppose now $\mathbb{X} = \mathbb{T}$ and that the assumptions of Theorem 2.8.7 hold for f. Then clearly $\int h(t)f(t)d\mu(t)$ is well defined for any bounded scalar function h. In particular, for indicators I_A,*

$$\lambda(A) := \int I_A(t)f(t)dt$$

is well defined and is a finite $\mathbb{T}$-valued signed measure, and

$$\int h(t)f(t)d\mu(t) = \int h\, d\lambda$$

holds for all bounded and measurable functions h.

Proof. It follows at once that

$$\int h(t)f(t)d\mu(t) = \int h\, d\lambda$$

for h equal to any indicator function and hence all simple functions. A standard argument going from simple to measurable functions then completes the proof. □

2.8.2 PROOF for Theorem 2.5.2

For simplicity, we will only illustrate proofs on $\mathbb{R}$ and the extension to $\mathbb{R}^d$ should be straightforward.

By Lemma 2.8.15, $\|\mathscr{K}(t)\|_{tr}$ is a continuous function and $\mathscr{K}(t)$ is therefore continuous in trace norm. We first assume that

$$\int_{\mathbb{R}} \|\mathscr{K}(t)\|_{tr}\, dt < \infty. \tag{2.65}$$

Let $\Pi_n = \sum_{i=1}^n e_i \otimes e_i$ be the projection operator onto $\mathrm{span}\{e_1, \ldots, e_n\}$ where $\{e_i\}$ is CONS of $\mathbb{H}$. Define $\mathscr{K}_n(t) = \Pi_n \mathscr{K}(t) \Pi_n$. By Lemma 2.8.13, $\|\mathscr{K}_n(t) - \mathscr{K}_n(s)\|_{tr} \leq \|\mathscr{K}(t) - \mathscr{K}(s)\|_{tr}$ and therefore $\mathscr{K}_n(t)$ is continuous in trace norm. Also by Lemma 2.8.13, it follows that

$$\int_{\mathbb{R}} \|\mathscr{K}_n(t)\|_{tr}\, dt \;\; \leq \;\; \int_{\mathbb{R}} \|\mathscr{K}(t)\|_{tr}\, dt < \infty.$$

By Corollary 2.8.8, the following Bochner's integrals in $\mathbb{T}$ is well defind:

$$\widehat{\mathscr{K}_n}(x) := (2\pi)^{-1} \int_t e^{-itx} \mathscr{K}_n(t) dt, \;\; \widehat{\mathscr{K}}(x) := (2\pi)^{-1} \int_t e^{-itx} \mathscr{K}(t) dt. \tag{2.66}$$

As $\mathscr{K}(t)$ is positive-definite, take $c_1 = 1$, $t = 0$, we will have $\mathscr{K}(0) \geq 0$ and self-adjoint. Take $c_1 = c_2 = 1$, $t_1 = 0$, $t_2 = t$, we will have $2\mathscr{K}(0) + \mathscr{K}(t) + \mathscr{K}(-t) \geq 0$, which implies $\mathscr{K}(t) + \mathscr{K}(-t)$ to be self-adjoint. Finally, take $c_1 = 1, c_2 = i, t_1 = t, t_2 = 0$, we will have $-i\mathscr{K}(t) + i\mathscr{K}(-t) \geq 0$ and self-adjoint. Therefore, we have the following two equations

$$\mathscr{K}^*(t) + \mathscr{K}^*(-t) = \mathscr{K}(t) + \mathscr{K}(-t)$$

and

$$i\mathscr{K}^*(t) - i\mathscr{K}^*(-t) = -i\mathscr{K}(t) + i\mathscr{K}(-t),$$

which immediately implies that

$$\mathscr{K}^*(t) = \mathscr{K}(-t). \tag{2.67}$$

By (2.67) and Lemma 2.8.17, we have

$$\begin{aligned} &\left\langle f, \widehat{\mathscr{K}}(t) g \right\rangle = \left\langle f, (2\pi)^{-1} \int_t e^{-itx} \mathscr{K}(t) dt g \right\rangle = \left\langle (2\pi)^{-1} \int_t e^{itx} \mathscr{K}(-t) dt f, g \right\rangle \\ = \;& \left\langle (2\pi)^{-1} \int_t e^{-itx} \mathscr{K}(t) dt f, g \right\rangle = \left\langle \widehat{\mathscr{K}}(t) f, g \right\rangle, \end{aligned}$$

and therefore, $\widehat{\mathscr{K}}(t)$ is self-adjoint. It follow from Lemma 2.8.16 and Theorem 2.8.10 that

$$\|\mathscr{K}_n(t) - \mathscr{K}(t)\|_{tr} \to 0 \text{ and } \|\widehat{\mathscr{K}_n}(x) - \widehat{\mathscr{K}}(x)\|_{tr} \to 0 \tag{2.68}$$

for all t, x.

For each $f, g \in \mathbb{H}$, defined the real-valued functions

$$K_{f,g}(t) = \langle f, \mathscr{K}(t)g \rangle \text{ and } K_f(t) = K_{f,f}(t).$$

Since $\mathscr{K}(\cdot)$ is positive definite and continuous at 0 in operator norm (entailed by continuity in trace norm), $K_f(t)$ is a positive-definite function and continuous at 0. It is also integrable by (2.65). Therefore, with the classical Bochner's Theorem, it follows that

$$K_f(t) = \int e^{itx} \widehat{K}_f(x) dx \quad \text{and} \quad \widehat{K}_f(x) := \frac{1}{2\pi} \int e^{-itx} K_f(t) dt, \tag{2.69}$$

where $\widehat{K}_f \geq 0$ and

$$\int \widehat{K}_f(x) dx = K_f(0), \tag{2.70}$$

which shows $\widehat{\mathscr{K}}$ is non-negative. Notice that, by (2.69), (2.67) and Lemma 2.8.12,

$$\begin{aligned}
\mathscr{K}_n(t) &= \sum_{i=1}^{n} \sum_{j=1}^{n} K_{e_i,e_j}(t)\, e_i \otimes e_j \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \left(\frac{i-1}{2}(K_{e_i} + K_{e_j}) - \frac{i}{2} K_{e_j+ie_i} + \frac{1}{2} K_{e_i+e_j} \right) e_i \otimes e_j \\
&= \sum_{i=1}^{n} \sum_{j=1}^{n} \int_x e^{itx} \left(\frac{i-1}{2}(\widehat{K}_{e_i} + \widehat{K}_{e_j}) - \frac{i}{2} \widehat{K}_{e_j+ie_i} + \frac{1}{2} \widehat{K}_{e_i+e_j} \right) dx\, e_i \otimes e_j \\
&= \int e^{itx} \widehat{\mathscr{K}_n}(x) dx,
\end{aligned}$$

and by (2.68), it implies

$$\mathscr{K}(t) = \int e^{itx} \widehat{\mathscr{K}}(x) dx. \tag{2.71}$$

To summarize, we have shown under (2.65) that

$$\mathscr{K}(t) = \int e^{itx} \widehat{\mathscr{K}}(x) dx \quad \text{and} \quad \widehat{\mathscr{K}}(x) = \frac{1}{2\pi} \int_t e^{-itx} \mathscr{K}(t) dt, \tag{2.72}$$

and $\widehat{\mathscr{K}}(x) \in \mathbb{T}_+$.

Next, relax (2.65) and define

$$\mathscr{K}_\sigma(t) = e^{-\sigma^2 t^2/2}\mathscr{K}(t) = \frac{1}{2\sigma^2}\int \mathscr{K}(t)e^{ity}e^{-y^2/(2\sigma^2)}dy. \tag{2.73}$$

Note that $\mathscr{K}(t)e^{ity}$ is positive definite in t. As a convex combination of positive definite functions,

$\mathscr{K}_\sigma(\cdot)$ is also positive definite. Since $\int_{\mathbb{R}} \|\mathscr{K}_\sigma(t)\|_{tr}dt < \infty$, it follows from (2.71) that

$$\mathscr{K}_\sigma(t) = \int_x e^{itx}\widehat{\mathscr{K}_\sigma}(x)dx \quad \text{and} \quad \widehat{\mathscr{K}_\sigma}(x) = \frac{1}{2\pi}\int_t e^{-itx}\mathscr{K}_\sigma(t)dt \tag{2.74}$$

for any $\sigma \geq 0$. Clearly, $\|\mathscr{K}_\sigma(t) - \mathscr{K}(t)\|_{tr} \to 0$. Thus, we only need to show that

$$\lim_{\sigma\to 0}\left\|\int_x e^{itx}\widehat{\mathscr{K}_\sigma}(x)dx - \int_x e^{itx}\mu(dx)\right\|_{tr} = 0$$

To achieve this, we first show that $\{\mu_\sigma\}$ is tight where $\mu_\sigma(\cdot) := \int_\cdot \widehat{\mathscr{K}_\sigma}(x)dx$. By (2.73), we have

$$\|\mu_\sigma(\mathbb{R})\|_{tr} = \|\mathscr{K}_\sigma(0)\|_{tr} = \|\mathscr{K}(0)\|_{tr} < \infty,$$

which implies (i) of the tightness definition. Next,

$$\begin{aligned}
\frac{1}{2T}\int_{-T}^{T}\mathscr{K}_\sigma(t)dt &= \frac{1}{2T}\int_{t=-T}^{T}\int_x e^{itx}\widehat{\mathscr{K}_\sigma}(x)dxdt \\
&= \int_x \frac{\sin Tx}{Tx}\widehat{\mathscr{K}_\sigma}(x)dx \\
&\leq \int_{|x|<b}\left|\frac{\sin Tx}{Tx}\right|\widehat{\mathscr{K}_\sigma}(x)dx + \int_{|x|\geq b}\left|\frac{\sin Tx}{Tx}\right|\widehat{\mathscr{K}_\sigma}(x)dx \\
&\leq \mu_\sigma(-b,b) + \frac{1}{Tb}\mu_\sigma(x:|x|>b) \\
&= \mu_\sigma(\mathbb{R}) - (1-1/(Tb))\mu_\sigma(x:|x|>b),
\end{aligned} \tag{2.75}$$

where "$\leq$" here is the operator inequality. By (2.75), with $b = 2/T$,

$$\begin{aligned}
\mu_\sigma(x:|x|>2/T) &\leq 2\left(\mu_\sigma(\mathbb{R}) - \frac{1}{2T}\int_{-T}^{T}\mathscr{K}_\sigma(t)dt\right) \\
&= \frac{1}{T}\int_{-T}^{T}(\mathscr{K}(0) - \mathscr{K}_\sigma(t))dt,
\end{aligned}$$

which implies that

$$\|\mu_\sigma(x : |x| > 2/T)\|_{tr} \leq \frac{1}{T}\int_{-T}^{T} \|\mathscr{K}(0) - \mathscr{K}_\sigma(t)\|_{tr}\, dt.$$

By the triangle inequality,

$$\|\mathscr{K}(0) - \mathscr{K}_\sigma(t)\|_{tr} \leq (1 - e^{-\sigma^2 t^2/2})\|\mathscr{K}(0)\|_{tr} + e^{-\sigma^2 t^2/2}\|\mathscr{K}(0) - \mathscr{K}(t)\|_{tr},$$

which tends to 0 by assumption (iii) of the theorem. This show that $\|\mu_\sigma(x : |x| > 2/T)\|_{tr} \to 0$ as $T \to \infty$ and establishes the tightness of $\{\mu_\sigma\}$.

For any measure μ and e_i, e_j in CONS, define $\mu^{(i,j)}(A) = \langle e_i, \mu(A)e_j\rangle$. By Lemma 2.8.19, for any sequence $\sigma_n \to 0$, there exists a finite $\mathbb{T}_+$-valued measure μ, and an integer sequence $n' \to \infty$, such that

$$\int_{\mathbb{R}} h(x)\mu^{(i,j)}_{\sigma_{n'}}(dx) \to \int_{\mathbb{R}} h(x)\mu^{(i,j)}(dx)$$

for all i, j and all bounded and continuous functions h. As each $\widehat{\mathscr{K}_{\sigma_n}}$ is symmetric, it follows that μ is a symmetric measure. By (ii) of Lemma 2.8.17, we have, for all i, j,

$$\begin{aligned}
\langle e_i, \mathscr{K}(t)e_j\rangle &= \lim_{n'\to\infty}\langle e_i, \mathscr{K}_{\sigma_{n'}}(t)e_j\rangle = \lim_{n'\to\infty}\left\langle e_i, \int_x e^{itx}\mu_{\sigma_{n'}}(dx)e_j\right\rangle \\
&= \lim_{n'\to\infty}\int_x e^{-itx}\left\langle e_i, \mu_{\sigma_{n'}}(dx)e_j\right\rangle = \lim_{n'\to\infty}\int_x e^{-itx}\mu^{(i,j)}_{\sigma_{n'}}(dx) \\
&= \int_x e^{-itx}\mu^{(i,j)}(dx) = \left\langle e_i, \int_x e^{itx}\mu(dx)e_j\right\rangle.
\end{aligned}$$

Thus, for any $f, g \in \mathbb{H}$,

$$\langle f, \mathscr{K}(t)g\rangle = \left\langle f, \int_x e^{itx}\mu(dx)g\right\rangle,$$

which entails that $\mathscr{K}(t) = \int_x e^{itx}\mu(dx)$. Finally, we have $K_f(t) = \int_x e^{itx}\langle f, \mu(dx)f\rangle$. Since this is a Fourier transform, it follows that $\langle f, \mu(A)f\rangle$ does not depend on the sequence σ_n or $\sigma_{n'}$ we take previously and thus μ is unique.

□

Lemma 2.8.12. *On a Hilbert space $\mathbb{H}$ allowing complex value coefficients, for any bounded*

linear operator A and $f, g \in \mathbb{H}$, we have

$$\langle f, Ag\rangle = \frac{i-1}{2}(\langle f, Af\rangle + \langle g, Ag\rangle) - \frac{i}{2}\langle g + if, A(g + if)\rangle + \frac{1}{2}\langle f + g, A(f + g)\rangle.$$

Proof. We have

$$\langle f + g, A(f + g)\rangle = \langle f, Af\rangle + \langle g, Ag\rangle + \langle f, Ag\rangle + \langle g, Af\rangle$$

and

$$\langle if + g, A(if + g)\rangle = \langle f, Af\rangle + \langle g, Ag\rangle + i\langle f, Ag\rangle - i\langle g, Af\rangle.$$

Cancel the term $\langle g, Af\rangle$ and we will obtain the representation in the lemma. □

Lemma 2.8.13. *For any trace class operator A, we have*

$$\|\Pi_1 A \Pi_2\|_{tr} \leq \|A\|_{tr},$$

where $\Pi_j, j = 1, 2$ is the projection operator from $\mathbb{H}$ to span$\{e_i, i \in I_j\}, j = 1, 2$, $|I_j| < \infty$ and $\{e_i\}$ is some fixed CONS.

Proof. As A is a compact operator, it has a singular value decomposition and $\|A\|_{tr} = \sum_i |\sigma_i|$, where σ_i's are singular values of A with $|\sigma_1| \geq |\sigma_2| \geq \cdots$.

$\Pi_1 A \Pi_2$ is a finite-rank operator and thus also has a singular value decomposition. Assume

$$\Pi_1 A \Pi_2 = \sum_i \sigma_i' f_i^{(1)} \otimes f_i^{(2)},$$

where $\{f_i^{(j)}\}, j = 1, 2$ are two CONS on span$\{e_i, i \in I_j\}, j = 1, 2$ and $|\sigma_1'| \geq |\sigma_2'| \geq \cdots$. We will have $\|\Pi_1 A \Pi_2\|_{tr} = \sum_i |\sigma_i'|$.

Suppose $\|\Pi_1 A \Pi_2\|_{tr} > \|A\|_{tr}$. Then there exists $m \in \mathbb{N}$ s.t. $|\sigma_m'| > |\sigma_m|$. Let $S = \text{span}\{f_1^{(2)}, \ldots, f_m^{(2)}\}$ and by Theorem 2.8.14 we will have

$$\begin{aligned}
\sigma_m^2 &\geq \min_{x \in S, \|x\|=1} \langle A^* A x, x\rangle = \min_{x \in S, \|x\|=1} \langle Ax, Ax\rangle = \min_{x \in S, \|x\|=1} \langle A\Pi_2 x, A\Pi_2 x\rangle \\
&\geq \min_{x \in S, \|x\|=1} \langle \Pi_1 A \Pi_2 x, \Pi_1 A \Pi_2 x\rangle = (\sigma_m')^2,
\end{aligned}$$

which is a contradiction and this completes our proof. □

Theorem 2.8.14. *(Courant min-max theorem, cf. Lieb and Loss (1997)). Let $\mathbb{H}$ be an infinite dimensional Hilbert and let A be a positive compact operator. If $k \in \mathbb{N}$ then*

$$\max_{dim\ S=k} \min_{x\in S, \|x\|=1} \langle Ax, x\rangle = \lambda_k(A),$$

and

$$\min_{dim\ S=k-1} \max_{x\in S^{\perp}, \|x\|=1} \langle Ax, x\rangle = \lambda_k(A).$$

Lemma 2.8.15. *Let $\{\mathscr{K}(t), t \in \mathbb{R}\}$ be a collection of operators. Assume (i)-(ii) as in Theorem 2.5.2. Then we have*

$$\lim_{\delta\to 0} \sup_{|t-s|<\delta} \|\mathscr{K}(t) - \mathscr{K}(s)\|_{tr} \to 0.$$

Proof. For any element $f \in$, denote $\mathscr{K}_f = \langle f, \mathscr{K} f\rangle$. It follows that $\mathscr{K}_f(t)$ is positive definite. By taking $c_1 = c_2 = 1$ and $t_1 = 0$, $t_2 = t$, one will have $Im(\mathscr{K}_f(t)) = -Im(\mathscr{K}_f(-t))$. Similarly, by taking $c_1 = 1$, $c_2 = i$, $t_1 = 0$, $t_2 = t$, we will have $Re(\mathscr{K}_f(t)) = Re(\mathscr{K}_f(-t))$. Therefore $\mathscr{K}_f(t) = \overline{\mathscr{K}_f(-t)}$. By taking $t_1 = t, t_2 = s, t_3 = 0$, we will have the following matrix to be nonnegative definite

$$\begin{pmatrix} \mathscr{K}_f(0) & \mathscr{K}_f(t-s) & \mathscr{K}_f(t) \\ \overline{\mathscr{K}_f(t-s)} & \mathscr{K}_f(0) & \mathscr{K}_f(s) \\ \overline{\mathscr{K}_f(t)} & \overline{\mathscr{K}_f(s)} & \mathscr{K}_f(0). \end{pmatrix}$$

As a result, its determinant will be nonnegative, that is

$$\begin{aligned} 0 \ \leq\ & \mathscr{K}_f^3(0) - \mathscr{K}_f(0)[|\mathscr{K}_f(t) - \mathscr{K}_f(s)|^2 + |\mathscr{K}_f(t-s)|^2] \\ & -2Re([\mathscr{K}_f(s)\mathscr{K}_f(t)(\mathscr{K}_f(0) - \mathscr{K}_f(t-s))]) \\ \leq\ & \mathscr{K}_f^3(0) - \mathscr{K}_f(0)[|\mathscr{K}_f(t) - \mathscr{K}_f(s)|^2 + |\mathscr{K}_f(t-s)|^2] - 2\mathscr{K}_f^2(0)|\mathscr{K}_f(0) - \mathscr{K}_f(t-s)|, \end{aligned}$$

where in the last line we use $|\mathscr{K}_f(s)|, |\mathscr{K}_f(t)| \leq \mathscr{K}_f(0)$. Rearranging terms then gives

$$\begin{aligned} |\mathscr{K}_f(t) - \mathscr{K}_f(s)|^2 \ &\leq\ \mathscr{K}_f^2(0) - |\mathscr{K}_f(t-s)|^2 + 2\mathscr{K}_f(0)|\mathscr{K}_f(0) - \mathscr{K}_f(t-s)| \\ &\leq\ 4\mathscr{K}_f(0)|\mathscr{K}_f(0) - \mathscr{K}_f(t-s)|, \end{aligned}$$

where we again use the fact that $|\mathscr{K}_f(t-s)| \leq \mathscr{K}_f(0)$.

Then, for CONS $\{e_i\}$ such that $\mathscr{K}(t) - \mathscr{K}(s)$ can be diagonalized, we have

$$\begin{aligned}
\|\mathscr{K}(t) - \mathscr{K}(s)\|_{tr} &= \sum_i |\mathscr{K}_{e_i}(t) - \mathscr{K}_{e_i}(s)| \\
&\leq \sum_i 2\sqrt{\mathscr{K}_{e_i}(0)|\mathscr{K}_{e_i}(0) - \mathscr{K}_{e_i}(t-s)|} \\
&\leq 2\sqrt{\sum_i \mathscr{K}_{e_i}(0) \sum_i |\mathscr{K}_{e_i}(0) - \mathscr{K}_{e_i}(t-s)|} \\
&\leq 2\sqrt{\|\mathscr{K}(0)\|_{tr}\|\mathscr{K}(0) - \mathscr{K}(t-s)\|_{tr}},
\end{aligned}$$

where the last line converges to 0 uniformly as $|t-s| \to 0$. □

Lemma 2.8.16. *Let $\mathscr{T}$ be an operator in $\mathbb{T}$ and Π_n be the projection operator $\Pi_n = \sum_{i=1}^n e_i \otimes e_i$ where $\{e_i\}$ is an arbitrary CONS of $\mathbb{H}$. Then*

$$\lim_{n\to\infty} \|\mathscr{T} - \Pi_n \mathscr{T} \Pi_n\|_{tr} = 0.$$

Proof. The result is a special case of Theorem 1 of Gelbaum and Gil de Lamadrid (1961). □

Lemma 2.8.17. *(a) For any bounded linear operator $\mathscr{T}$ and f for which $\int f(x)\mu(dx)$ is well defined, we have*

$$\int f(x)(\mathscr{T}^*\,\mu(dx)\mathscr{T}) = \mathscr{T}^* \left(\int f(x)\mu(dx)\right) \mathscr{T}$$

where $\mathscr{T}^*\,\mu(\cdot)\mathscr{T}$ denotes the measure $\nu(A) = \mathscr{T}^*\,\mu(A)\mathscr{T}$.

(b) For any $h \in \mathbb{H}$ and f for which $\int f(x)\mu(dx)$ is well defined, we have

$$\int \bar{f}(x)\langle h, \mu(dx)h\rangle = \left\langle h, \int f(x)\mu(dx)h \right\rangle,$$

where $\langle h, \mu(\cdot)h\rangle$ denotes the measure $\langle h, \mu(A)h\rangle$.

Proof. (a) is trivial for nonnegative simple functions. Then take limits as $f_n \to f$ and use the fact $\mathscr{T}$ is continuous. (b) is similarly proved. □

Lemma 2.8.18. *Let μ be a finite $\mathbb{T}_+$-valued measure, and $\{e_i\}$ a fixed CONS in $\mathbb{H}$. Then,*

$$\mu^{(i,j)}(A) := \langle e_i, \mu(A)e_j\rangle, \quad i,j \in \mathbb{N},$$

are signed measures such that $\mu^{(i,i)}$ are positive and for all $A \in \mathscr{B}(\mathbb{R})$, we have

$$|\mu^{(i,j)}(A)| \leq (\mu^{(i,i)}(A)\mu^{(j,j)}(A))^{1/2}. \tag{2.76}$$

Consequently

$$\sum_{i,j} \|\mu^{(i,j)}\|_{\mathrm{TV}}^2 \leq 4\|\mu(\mathbb{R})\|_{tr}^2 \equiv 4\left(\sum_i \mu^{(i,i)}(\mathbb{R})\right)^2. \tag{2.77}$$

Proof. The inequality in (2.76) is Cauchy-Schwartz. The σ-additivity of $\mu^{(i,j)}$ is a direct consequence of the (strong) σ-additivity of μ. The inequality (2.77) follows from the fact that by taking a supremum in (2.76) over A, we obtain

$$\|\mu^{(i,j)}\|_{TV} = 2\sup_A |\mu^{(i,j)}(A)|.$$

□

Lemma 2.8.19. *Let $\{\mu_n\}$ be be a class of finite $\mathbb{T}_+$-valued measures and tight. Then it is weakly relatively compact, in the sense that there exists a finite $\mathbb{T}_+$-valued measure μ, and an integer sequence $n' \to \infty$, such that*

$$\int_{\mathbb{R}} h(x)\mu_{n'}^{(i,j)}(dx) \to \int_{\mathbb{R}} h(x)\mu^{(i,j)}(dx), \tag{2.78}$$

for all bounded and continuous functions h.

Proof. For a finite signed measure ν from from the Jordan-Hahn theorem, we have

$$\nu(A) = \nu_+(A) - \nu_-(A),$$

where $\nu_\pm$ are finite positive measures. Let us define the finite positive measure $|\nu|(A) := \nu_+(A) + \nu_-(A)$. Observe that if $\{\nu_n\}$ is a sequence of finite signed measures such that $\{|\nu_n|\}$ is tight, then both $\{\nu_{n,+}\}$ and $\{\nu_{n,-}\}$ are and by Prokhorovs theorem there is a $n' \to \infty$, such that $\nu_{n',\pm} \to \nu_\pm$ weakly, for some finite positive measures $\nu_\pm$. This means that for all bounded and continuous h, we have

$$\int h(x)\nu_{n',\pm}(dx) \to \int h(x)\nu_\pm(dx).$$

Consequently, we obtain

$$\int_{\mathbb{R}} h(x)\nu_{n'}(dx) = \int h(x)\nu_{n',+}(dx) - \int h(x)\nu_{n',-}(dx) \to \int_{\mathbb{R}} h(x)\nu(dx),$$

where $\nu = \nu_+ - \nu_-$.

Thus, to prove (2.78), it remains to show that for all $i, j \in \mathbb{N}$, the sequence of finite signed measures $\{|\mu_n^{(i,j)}|,\ n \in \mathbb{N}\}$ is tight. This is a direct consequence of (2.77) and the tightness of the $\mathbb{T}_+$-valued sequence of measures $\{\mu_n\}$.

As $|\mu_n|^{(i,j)} \to |\mu|^{(i,j)}$ weakly, for any A with $|\mu^{(i,j)}|(\partial A) = 0$, we have

$$|\mu_{n'}^{(i,j)}|(A) \to |\mu^{(i,j)}|(A).$$

Denote

$$\mathscr{E} = \{A \subset \mathbb{R}, \mu_{n'}^{(i,j)}(A) \to \mu^{(i,j)}(A)\}.$$

Clearly, $\mathbb{R} \in \mathscr{E}$ and thus

$$|\mu^{(i,j)}|(\mathbb{R}) \leq \sup_n |\mu_n^{(i,j)}|(\mathbb{R}) < \infty,$$

which entails that the set $\mathscr{O} := \{x \in \mathbb{R}, \mu(x) > 0\}$ is countable. Then we know

$$\mathscr{S} := \{(a, b), a, b \notin \mathscr{O}\} \subset \mathscr{E}.$$

If $A \in \mathscr{E}$, we have $A^c \in \mathscr{E}$ as $\mathbb{R} \in \mathscr{E}$. Due to σ-additivity of $|\mu^{(i,j)}|$ and $|\mu_n^{(i,j)}|$, for disjoint $A_i \in \mathscr{E}, i = 1, 2, \ldots$, we have $\cup_i A_i \in \mathscr{E}$ and thus all open intervals belongs to $\mathscr{E}$. As the set for all open intervals is a π-system and $\mathscr{E}$ is a λ-system, $\mathscr{E}$ must contain all Borel sets on $\mathbb{R}$.

Finally, define $\mu(A)$ as $\mu(A)e_j = \sum_i \mu^{(i,j)}(A)e_i, j = 1, 2, \ldots$ and extend it to the whole space by linearity. As $\sup_j \sum_i |\mu^{(i,j)}(A)|^2 < \infty$ by (2.77), $\mu(A)$ is a bounded operator. Therefore, for any $f = c_i e_i \in \mathbb{H}$, $\mu(A)f = \sum_i c_i \mu(A)e_i = \sum_{i,j} c_i \mu^{(i,j)}(A)e_j$. We need to show

- $\mu(\cdot)$ satisfies σ-additivity.
- $\mu(A) \in \mathbb{T}_+$.

For disjoint Borel sets A_k, we have

$$\langle e_i, \mu(\cup_k A_k)e_j\rangle = \mu^{(i,j)}(\cup_k A_k) = \sum_k \mu^{(i,j)}(A_k) = \sum_k \langle e_i, \mu(A_k)e_j\rangle,$$

where the second and fourth equalities follow from the dominated convergence theorem and (2.77). Thus $\mu(\cdot)$ satisfies σ-additivity.

For any $f = \sum_i c_i e_i \in \mathbb{H}$, by dominated convergence,

$$\langle f, \mu(A) f\rangle = \sum_j c_j \overline{\sum_i c_i \mu^{(i,j)}(A)} = \lim_{n'\to\infty} \sum_j c_j \overline{\sum_i c_i \mu_{n'}^{(i,j)}(A)} = \lim_{n'\to\infty} \langle f, \mu_{n'}(A) f\rangle \geq 0,$$

and

$$\langle f, \mu(A) f\rangle = \lim_{n'\to\infty} \langle f, \mu_{n'}(A) f\rangle = \lim_{n'\to\infty} \langle \mu_{n'}(A) f, f\rangle = \langle \mu(A) f, f\rangle.$$

As $\mu(A)$ is nonnegative and self-adjoint, we have $\|\mu(A)\|_{tr} = \sum_k \langle e_k, \mu(A) e_k\rangle$, and thus by the dominated convergence theorem,

$$\|\mu(A)\|_{tr} = \sum_k \langle e_k, \mu(A) e_k\rangle = \sum_k \mu^{(k,k)}(A) = \lim_{n'\to\infty} \sum_k \mu_{n'}^{(k,k)}(A) \leq \sup_n \|\mu_n(\mathbb{R})\|_{tr} < \infty.$$

Therefore $\mu(A) \in \mathbb{T}_+$. □

2.8.3 PROOF for Theorem 2.5.3

For simplicity, we will only illustrate proofs on $\mathbb{R}$ and the extension to $\mathbb{R}^d$ should be straightforward.

For finitely supported measures $\mu = \sum_{i=1}^m c_i \delta_{x_i}$ and $\nu = \sum_{j=1}^n d_j \delta_{y_j}$, denote

$$\widetilde{\mu} = \sum_{i=1}^m \bar{c}_i \delta_{-x_i} \quad \text{and} \quad \mu * \nu = \sum_{i=1}^m \sum_{j=1}^n c_i d_j \delta_{x_i + y_j}.$$

For any $\mu, \nu \in \Lambda_k$, $X(\mu + t)$ and $X(\nu + t)$ are both stationary processes in t where their stationary covariance functions are denoted by C_μ, C_ν. Notice that

$$\mathrm{Cov}(X(\nu * \mu + t), X(\nu * \mu + t + h)) \;=\; \sum_{i=1}^n \sum_{j=1}^n d_i \bar{d}_j C_\mu(h + y_i - y_j) = C_\mu(\nu * \widetilde{\nu} + h).$$

Clearly, $X(\nu * \mu) = X(\mu * \nu)$, and hence we have

$$C_\nu(\mu * \widetilde{\mu} + h) = C_\mu(\nu * \widetilde{\nu} + h). \tag{2.79}$$

Denote by τ_μ the spectral measure of C_μ as indicated by Lemma 2.8.21 and Theorem 2.5.2.

Then

$$C_\nu(\mu * \widetilde{\mu} + h) = \int \sum_{k=1}^{m}\sum_{l=1}^{m} c_k\bar{c}_l e^{i(h+x_k-x_l)u}\tau_\nu(du) = \int e^{ihu}|\widehat{\mu}(u)|^2\tau_\nu(du)$$

where $\widehat{\mu}(u) = \int e^{iux}\mu(dx)$. It follows from Theorem 2.5.2, as measures,

$$|\widehat{\mu}|^2(u)\tau_\nu(du) = |\widehat{\nu}|^2(u)\tau_\mu(du). \tag{2.80}$$

By condition (i), Lemma 2.8.21 and a trivial extension of Lemma 2.8.13, we will have

$$C^{(i,j)}(s,t) := \langle e_i, C(s,t)e_j\rangle$$

is a continuous function. Take μ_0 to be the measure in Lemma 2.8.24, and by property 3 in that lemma, there exists $\{\mu_n\} \in \Lambda_k$ such that $\mu_n \to \mu_0$ weakly. Therefore,

$$\begin{aligned} & C^{(i,j)}(\mu_0 + a, \mu_0 + a + h) = \lim_{n\to\infty} C^{(i,j)}(\mu_n + a, \mu_n + a + h) \\ = & \lim_{n\to\infty} C^{(i,j)}(\mu_n, \mu_n + h) = C^{(i,j)}(\mu_0, \mu_0 + h), \end{aligned}$$

and thus, $X(\mu_0)$ is stationary. As a result, C_μ and τ_μ can be defined for that μ_0 and (2.79) and (2.80) hold for any $\mu, \nu \in \Lambda_k \cup \{\mu_0\}$.

We now define the measure

$$\sigma(du) = \begin{cases} \frac{\tau_{\mu_0}(du)}{|\widehat{\mu}_0(u)|^2}(1 \wedge |u|^{2k+2}) & u \neq 0, \\ 0 & u = 0. \end{cases} \tag{2.81}$$

We will show σ is a $\mathbb{T}_+$-valued measure in Proposition 2.8.20. By (2.80) and (2.81), for any $\mu \in \Lambda_k \cup \{\mu_0\}$, denote $S_\mu = \{u : |\widehat{\mu}(u)|^2 = 0\}$ and we will have

$$\int \frac{e^{ihu}|\widehat{\mu}(u)|^2}{1 \wedge |u|^{2k+2}}\sigma(du) + \sum_{u\in S_\mu} \tau_\mu(\{u\})e^{ihu} = \int e^{ihu}\tau_\mu(du). \tag{2.82}$$

Notice that by similar construction as we used for ν_s, for any fixed μ and location $u_0 \in S_\mu \setminus \{0\}$, we will have

$$\begin{aligned} \tau_\mu(\{u_0\}) &= \lim_{\epsilon\to 0}\int_{(u_0-\epsilon,u_0+\epsilon)} \tau_\mu(du) = \lim_{\epsilon\to 0}\int_{(u_0-\epsilon,u_0+\epsilon)} \frac{|\widehat{\mu}(u)|^2}{|\widehat{\mu}_0(u)|^2}\tau_{\mu_0}(du) \\ &= \frac{|\widehat{\mu}(u_0)|^2\tau_{\nu_{s_{u_0}}}(\{u_0\})}{|\widehat{\mu}_0(u_0)|^2} = 0. \end{aligned} \tag{2.83}$$

Therefore, (2.82) can be simplified as

$$\int \frac{e^{ihu}|\widehat{\mu}(u)|^2}{1 \wedge |u|^{2k+2}}\sigma(du) + \tau_\mu(\{0\}) = \int e^{ihu}\tau_\mu(du). \tag{2.84}$$

Denote

$$K_0(h) = \int \frac{e^{ihu} - I_B(u)P(uh)}{1 \wedge |u|^{2k+2}}\sigma(du),$$

where B is a bounded interval containing 0 and

$$P(x) = \sum_{j=0}^{2k+1} (ix)^j/j!.$$

In Proposition 2.8.20, we show that $K_0(h)$ is well defined for every h. Notice that for (2.84), $RHS = C_\mu(h)$ and

$$\begin{aligned} LHS &= \int \frac{\sum_{i=1}^m \sum_{j=1}^m c_i\bar{c}_j e^{i(h+x_i-x_j)u}}{1 \wedge |u|^{2k+2}}\sigma(du) + \tau_\mu(\{0\}) \\ &= \sum_{i=1}^m \sum_{j=1}^m c_i\bar{c}_j \int \frac{e^{i(h+x_i-x_j)u} - I_B(u)P_k(u(h+x_i-x_j))}{1 \wedge |u|^{2k+2}}\sigma(du) + \tau_\mu(\{0\}) \\ &= \sum_{i=1}^m \sum_{j=1}^m c_i\bar{c}_j K_0(h+t_i-t_k) + \tau_\mu(\{0\}), \end{aligned} \tag{2.85}$$

since $P(\mu * \widetilde{\mu}) = 0$.

One can notice that by now, if $\tau_\mu(\{0\}) = 0$ for any $\mu \in \Lambda_k$, then K_0 is a generalized covariance. The rest of the proof mainly focus on dealing with the issue of $\tau_\mu(\{0\})$ above.

For any $\mu, \nu \in \Lambda_k$, $s_0 \in \mathbb{R}$, denote

$$C_{\mu,\nu}(h) = \mathbb{E}[X(\mu + s_0) \otimes X(\nu + s_0 + h)],$$

and we will have for any $h_1, h_2 \in \mathbb{R}$,

$$C_{\mu_1,\mu_2}(h_1 - h_2) = \sum_{i,j} c_i^{(1)}\bar{c}_j^{(2)} C(s_0 + x_i^{(1)} + h_1, s_0 + x_j^{(2)} + h_2),$$

where $\mu_j = \sum_i c_i^{(j)}\delta_{x_i^{(j)}}, j = 1, 2$. For any $\mu, \nu \in \Lambda_k$, by the same calculation in Lemma 2.8.12,

we have

$$C_{\nu,\mu}(h) = \frac{i-1}{2}(C_\mu(h) + C_\nu(h)) - \frac{i}{2}C_{\mu+i\nu}(h) + \frac{1}{2}C_{\mu+\nu}(h).$$

Similarly, we can obtain $K_0(\nu * \widetilde{\mu} + h)$ as

$$\begin{aligned} K_0(\nu * \widetilde{\mu} + h) &= \frac{i-1}{2}(K_0(\mu * \widetilde{\mu} + h) + K_0(\nu * \widetilde{\nu} + h)) \\ &\quad -\frac{i}{2}K_0((\mu + i\nu) * \widetilde{(\mu + i\nu)} + h) + \frac{1}{2}K_0((\mu + \nu) * \widetilde{(\mu + \nu)} + h). \end{aligned}$$

Define

$$F(h_1, h_2) = C(s_0 + h_1, s_0 + h_2) - K_0(h_1 - h_2).$$

By (2.85), we have $C_\mu(h) - K_0(\mu * \widetilde{\mu} + h) = \tau_\mu(\{0\})$ and therefore,

$$\begin{aligned} & \sum_{i,j} c_i^{(1)} \bar{c}_j^{(2)} F(h_1 + x_i^{(1)}, h_2 + x_j^{(2)}) \\ = \;& C_{\mu,\nu}(h_1 - h_2) - K_0(\mu * \widetilde{\nu} + h_1 - h_2) \\ = \;& \frac{i-1}{2}(\tau_\mu(\{0\}) + \tau_\nu(\{0\})) - \frac{i}{2}\tau_{\mu+i\nu}(\{0\}) + \frac{1}{2}\tau_{\mu+\nu}(\{0\}), \end{aligned}$$

which is a constant unrelated to h_1, h_2.

By Lemma 2.8.27,

$$F(h_1, h_2) = \sum_{l=1}^{k} G_l^{(1)}(h_1)(h_2)^l + \sum_{l=1}^{k} G_l^{(2)}(h_2)(h_1)^l + C_0(h_1 - h_2)^{2k+2},$$

where $G_l^{(i)}$'s are arbitrary functions. This implies that for any $\mu, \nu \in \Lambda_k$,

$$h^{2k+2}(\mu * \widetilde{\mu})\tau_\nu(\{0\}) = h^{2k+2}(\nu * \widetilde{\nu})\tau_\mu(\{0\}).$$

Now, we can take arbitrary $\mu \in \Lambda_k$ such that $h^{2k+2}(\mu * \widetilde{\mu}) \neq 0$ and let $C_0 = \frac{\tau_\mu(0)}{h^{2k+2}(\mu * \widetilde{\mu})}$. Define

$$K(h) = K_0(h) + C_0 h^{2k+2}.$$

By similar calculations in (2.85), we can show K to be a generalized covariance function for X.

To prove uniqueness of σ, assume there is another σ' that can construct the generalized

covariance operator. Then by the same arguments in the proof of Proposition 2.8.20 and (2.84), we will have

$$\int \frac{e^{ihu}|\widehat{\mu}(u)|^2}{1 \wedge |u|^{2k+2}}\sigma(du) = \int \frac{e^{ihu}|\widehat{\mu}(u)|^2}{1 \wedge |u|^{2k+2}}\sigma'(du),$$

or equivalently

$$\frac{|\widehat{\mu}(u)|^2}{1 \wedge |u|^{2k+2}}\sigma(du) = \frac{|\widehat{\mu}(u)|^2}{1 \wedge |u|^{2k+2}}\sigma'(du).$$

By Lemma 2.8.24 and (ii) in the condition, take $\mu = \mu_0$ and we will have $\sigma = \sigma'$.

The lower order polynomial term in the representation of $K(h)$ is trivial.

The "if" part of this theorem is also trivial.

□

Proposition 2.8.20. *$\sigma(du)$ defined in (2.81) is a $\mathbb{T}_+$-valued measure and K_0 is well defined and finite whenever X is mean-square continuous.*

Proof. It suffices to show that there exists a neighborhood B containing 0 such that

$$\|\sigma(B)\|_{tr} = \left\| \int_B \frac{(1 \wedge |u|^{2k+2})\tau_{\mu_0}(du)}{|\widehat{\mu}_0(u)|^2} \right\|_{tr} < \infty, \tag{2.86}$$

$$\|\sigma(B^c)\|_{tr} = \left\| \int_{B^c} \frac{(1 \wedge |u|^{2k+2})\tau_{\mu_0}(du)}{|\widehat{\mu}_0(u)|^2} \right\|_{tr} < \infty, \tag{2.87}$$

$$\left\| \int_B \frac{\left|e^{ihu} - P(uh)\right|}{1 \wedge |u|^{2k+2}}\sigma(du) \right\|_{tr} < \infty, \tag{2.88}$$

and

$$\left\| \int_{B^c} \frac{1}{1 \wedge |u|^{2k+2}}\sigma(du) \right\|_{tr} < \infty. \tag{2.89}$$

By Taylor's expansion,

$$\left\| \int_B |e^{ihu} - P(uh)|\sigma(du) \right\|_{tr} \leq \left\| \frac{|h|^{2k+2}}{(2k+2)!} \int_B |u|^{2k+2}\sigma(du) \right\|_{tr}$$

and

$$2(1-\cos u) \geq u^2 - u^4/12 > u^2/2, \ |u| < \sqrt{6}.$$

Thus, taking $B = (-\sqrt{6}, \sqrt{6})$, $\mu = (\delta_1 - \delta_0)^{k+1}$, we will have

$$\begin{aligned} & \left\| \int_B \frac{(1 \wedge |u|^{2k+2}) \tau_{\mu_0}(du)}{|\widehat{\mu}_0(u)|^2} \right\|_{tr} \leq \left\| \int_B \frac{|u|^{2k+2} \tau_{\mu_0}(du)}{|\widehat{\mu}_0(u)|^2} \right\|_{tr} \\ \leq & \left\| \int_B \frac{|u|^{2k+2} \tau_{\mu}(du)}{|\widehat{\mu}(u)|^2} \right\|_{tr} \leq \left\| \int_B \tau_{\mu}(du) \right\|_{tr} < \infty, \end{aligned}$$

and

$$\begin{aligned} & \left\| \int_B \frac{|e^{ihu} - P(uh)|}{1 \wedge |u|^{2k+2}} \sigma(du) \right\|_{tr} \leq \left\| \frac{|h|^{2k+2}}{(2k+2)!} \int_B \frac{|u|^{2k+2}}{1 \wedge |u|^{2k+2}} \sigma(du) \right\|_{tr} \\ \leq & \frac{|h|^{2k+2}}{2(2k+2)!} \left\| \int_B \frac{|\widehat{\mu}(u)|^2}{1 \wedge |u|^{2k+2}} \sigma(du) \right\|_{tr} < \infty, \end{aligned}$$

where the last inequality is because both sides of (2.82) are finite for any h (and in particular for $h = 0$) and $\mu \in \Lambda_k$. This establishes (2.86) and (2.88).

To prove (2.89), take μ_0 as the measure in Lemma 2.8.24. By Lemma 2.8.24, we can pick δ such that $\inf_{u \in B^c} |\widehat{\mu}_0(u)| > \delta > 0$. Therefore,

$$\left\| \int_{B^c} \frac{(1 \wedge |u|^{2k+2}) \tau_{\mu_0}(du)}{|\widehat{\mu}_0(u)|^2} \right\|_{tr} \leq \left\| 1/\delta \int_{B^c} \tau_{\mu_0}(du) \right\|_{tr} < \infty,$$

$$\left\| \int_{B^c} \frac{1}{1 \wedge |u|^{2k+2}} \sigma(du) \right\|_{tr} \leq \left\| 1/\delta \int \frac{|\widehat{\mu}_0(u)|^2}{1 \wedge |u|^{2k+2}} \sigma(du) \right\|_{tr} < \infty$$

which proves (2.87) and (2.89).

□

Lemma 2.8.21. *If $X(t) \in \mathbb{H}$ is mean-square continuous, then $C(s,t)$ is continuous in trace norm.*

Proof. We first prove some Cauchy-Schwarz type inequalities for outer product operator. For any $A, B \in \mathbb{H}$ and $t \in \mathbb{R}$, where $\mathbb{H}$ is a real Hilbert space, we have

$$0 \leq tr((A - tB) \otimes (A - tB)) = tr(A \otimes A) + t^2 tr(B \otimes B) - (2t) Re(tr(A \otimes B)).$$

Minimized right side with respect to t and we will have

$$|Re(tr(A \otimes B))| \leq \sqrt{tr(A \otimes A)tr(B \otimes B)}. \tag{2.90}$$

For random elements $A, B \in \mathbb{H}$ with their covariance operator well-defined, assume the following singular value decomposition

$$\mathbb{E}[A \otimes B] = WDV,$$

with W, V unitaries and D a diagonal matrix of singular values. Let $U = V^*W^*$, which is also a unitary operator, and we will have

$$\begin{aligned} \|\mathbb{E}[A \otimes B]\|_{tr} &= tr(\mathbb{E}[A \otimes B]U) = \mathbb{E}[tr((A \otimes B)U)] = \mathbb{E}[tr(A \otimes U^*B)] \\ &\leq \mathbb{E}[\sqrt{tr(A \otimes A)tr((U^*B) \otimes (U^*B))}] \\ &= \mathbb{E}[\sqrt{tr(A \otimes A)tr(B \otimes B)}] \\ &\leq \sqrt{\|\mathbb{E}[A \otimes A]\|_{tr}\|\mathbb{E}[B \otimes B]\|_{tr}}. \end{aligned} \tag{2.91}$$

Now we are ready to prove the lemma.

$$\begin{aligned} & \|C(t+h, s+h') - C(t,s)\|_{tr} \\ = \; & \|\mathbb{E}[X(t+h) \otimes X(s+h')] - \mathbb{E}[X(t) \otimes X(s)]\|_{tr} \\ \leq \; & \|\mathbb{E}[(X(t+h) - X(t)) \otimes X(s+h')]\|_{tr} + \|\mathbb{E}[X(t) \otimes (X(s+h') - X(s))]\|_{tr} \\ \leq \; & \sqrt{\|\mathbb{E}[(X(t+h) - X(t)) \otimes (X(t+h) - X(t))]\|_{tr}\|\mathbb{E}[X(s+h') \otimes X(s+h')]\|_{tr}} \\ & + \sqrt{\|\mathbb{E}[(X(s+h') - X(s)) \otimes (X(s+h') - X(s))]\|_{tr}\|\mathbb{E}[X(t) \otimes X(t)]\|_{tr}}, \end{aligned}$$

where the final term converges to 0 if $h, h' \to 0$. □

Let

$$M_f(k) = \{\mu = \mu_1 * \cdots * \mu_{k+1} : \mu_i \in \Lambda_0\}.$$

Also, define the larger class

$$M_c(k) = \{\mu_1 * \cdots * \mu_{k+1} : 1(\mu_i) = 0,\ \mu_i\text{'s are finite signed measures with compact supports}\}.$$

It is easy to see that any $\mu \in M_c(k)$ annihilates polynomials of degree k.

Lemma 2.8.22. *(Cf. Lemma 2.5 of Sasvári (2009)) For any $k \in \mathbb{N}$, $M_f(k)$ is a dense set in $M_c(k)$ with respect to the weak topology.*

Lemma 2.8.23. *(Cf. Lemma 3.2 of Sasvári (2009)) For any open set O containing 0, there exists $\mu \in M_c(k)$ and $\delta > 0$ such that μ is symmetric, $\|\mu\|_{TV} < \infty$ and for any $u \in O^c$, $\widehat{\mu}(u) > \delta$.*

Proof. The proof is almost the same as the original proof. One only needs to notice that μ in the proof of their original lemma can be made to be symmetric under our scenario. □

Lemma 2.8.24. *There exists a measure μ on $\mathbb{R}$ satisfying*

(i) $\|\mu\|_{TV} < \infty$.

(ii) For any open set O containing 0, there exists $\delta > 0$ such that for any $u \in O^c$, $\widehat{\mu}(u) > \delta$.

(iii) There exists a sequence of $\{\mu_n\} \subset M_f(k)$ converging to μ weakly.

Proof. For $V_n = [-1/n, 1/n]$, by Lemma 2.8.23, there exists $\{\nu_n\} \subset M_c(k)$ be a set of symmetric measures satisfying $\|\mu_n(u)\|_{TV} \leq 1/2^n$ and $\mu_n(u) > 0$ for $u \notin V_n$. Then define $\mu = \sum_n \nu_n$. Easy to know properties (i) and (ii) will satisfy. Property (iii) is guaranteed by Lemma 2.8.22 and $\sum_{n=1}^{M} \nu_n \to \mu$ weakly. □

By a trivial extension of Lemma C.9.4 of Sasvári (2013), we can obtain the following lemma.

Lemma 2.8.25. *Let $f(h) : \mathbb{R} \mapsto \mathbb{T}$ be a continuous function in trace norm. If for some $k \in \mathbb{N}$, $f(\mu) = 0$ for any $\mu \in M_f(k)$, then $f(\lambda) = 0$ for any $\lambda \in \Lambda_k$.*

The following can be easily obtained from Proposition 3 in Berschneider (2012) and its proof.

Lemma 2.8.26. *Let $f(h) : \mathbb{R} \mapsto \mathbb{T}$. If for some $k \in \mathbb{N}$, $f(\lambda) = 0$ for any $\lambda \in \Lambda_k$, then we have the following decomposition*

$$f(h) = \sum_{i=0}^{k} a_i h^i,$$

furthermore, there exist $\{b_{ij}, i = 0, \ldots, k+1, j = 0, \ldots, M_k\}$ and $\{x_j, j = 0, \ldots, M_k\}$ such that

$$a_i = \sum_{j=0}^{M_k} b_{ij} f(x_j).$$

Lemma 2.8.27. *Assume $F(h_1, h_2) : \mathbb{R} \otimes \mathbb{R} \mapsto \mathbb{T}$ is continuous in trace norm. If $F(h_1 + \mu, h_2 + \nu) = C(\mu, \nu)$ for any two measures $\mu, \nu \in \Lambda_k$, then $F(h_1, h_2)$ has the following decomposition*

$$F(h_1, h_2) = \sum_{l=1}^{k} G_l^{(1)}(h_1)(h_2)^l + \sum_{l=1}^{k} G_l^{(2)}(h_2)(h_1)^l + C_0(h_1 - h_2)^{2k+2},$$

where $G_l^{(i)}$'s are arbitrary functions.

Proof. Fix $h_1, \mu \in \Lambda_k$, by Lemma 2.8.25, we will have for any $\nu \in \Lambda_{k+1}$, $F(h_1 + \mu, h_2 + \nu) = 0$, therefore by Lemma 2.8.26,

$$F(h_1 + \mu, h_2) = \sum_{i=0}^{k+1} a_i(h_1, \mu) h_2^i,$$

and

$$a_i(h_1, \mu) = \sum_{j=0}^{M_{k+1}} b_{ij} F(h_1 + \mu, x_j).$$

Denote

$$G_{1,i}(h_1) = \sum_{j=0}^{M_{k+1}} b_{ij} F(h_1, x_j),$$

and we will have for any $\mu \in \Lambda_k$,

$$F(h_1 + \mu, h_2) - \sum_{i=0}^{k+1} G_{1,i}(h_1 + \mu) h_2^i = 0.$$

Therefore, again by Lemma 2.8.26,

$$F(h_1, h_2) = \sum_{i=0}^{k+1} G_{1,i}(h_1) h_2^i + \sum_{i=0}^{k} G_{2,i}(h_2) h_1^i.$$

Notice that for any $\mu, \nu \in \Lambda_k$, $F(h_1 + \mu, h_2 + \nu) = G_{1,k+1}(h_1 + \mu) C(\nu)$ needs to be unrelated to h_1, and thus $G_{1,k+1}(\mu) = 0$ for all $\mu \in M_f(k+1)$. Using Lemma 2.8.25 and Lemma 2.8.26 again and we will get $G_{1,k+1}(h_1) = \sum_{i=0}^{k+1} a_i' h_1^i$. Reorganize the formula for

$F(h_1, h_2)$ and one will have

$$F(h_1, h_2) = \sum_{l=1}^{k} G_l^{(1)}(h_1)(h_2)^l + \sum_{l=1}^{k} G_l^{(2)}(h_2)(h_1)^l + C_0(h_1 - h_2)^{2k+2},$$

where $G_l^{(i)}$'s are arbitrary functions. □

2.8.4 PROOF for Theorem 2.5.4

For simplicity, we will only illustrate proofs on $\mathbb{R}$ and the extension to $\mathbb{R}^d$ should be straightforward.

For any $e \in \mathbb{H}$, denote

$$C_e(h) = \langle e, C(h)e \rangle,$$

$$C_0(h) = tr(C(h)),$$

$$\sigma_e(dx) = \langle e, \sigma(dx)e \rangle,$$

and

$$\sigma_0(dx) = tr(\sigma(dx)).$$

Then we have

$$\langle X(t), X(s) \rangle_\Omega = \mathbb{E}\langle X(t), X(s) \rangle = C_0(t-s) = \int e^{i(t-s)x} \sigma_0(dx),$$

where the second equality above is entailed by (2.38). Let L_f to be the linear span of the functions $\{e^{it\cdot}, t \in \mathbb{R}\}$. We claim that L_f is dense in $L^2(\sigma_0)$. To see this, assume $g \in L^2(\sigma_0)$ and

$$\int e^{itx} \overline{g(x)} \sigma_0(dx) = 0.$$

It is easy to know $\mu := \bar{g}\sigma_0$ is a finite measure and thus the equation above indicates $\widehat{\mu} = 0$,

or equivalently, $g = 0$ in $L^2(\sigma_0)$. Consider functions of the form

$$f(x) = \sum_{k=1}^{n} c_k e^{it_k x} \in L_f$$

and define the linear mapping $\mathscr{J}_X : L_f \to \mathbb{H}(X)$ by

$$\mathscr{J}_X(f) = \sum_{k=1}^{n} c_k X(t_k).$$

Thus,

$$\|\mathscr{J}_X(f)\|_\Omega = \sum_{j,k=1}^{n} c_j \overline{c_k} C_0(t_j, t_k) = \int f(x)\overline{f(x)}\sigma_0(dx) = \|f\|_{L^2(\sigma_0)},$$

and $\mathscr{J}_X$ can be readily extended to an isometric isomorphism, still denoted as $\mathscr{J}_X$, from $L^2(\sigma_0)$ into $\mathbb{H}(X)$.

Notice that if a sequence $\{f_j\}$ converges in $L^2(\sigma_0)$, it will also converge in $L^2(\sigma_e), e \in \mathbb{H}$. Similarly, for $A_n, A \in \mathbb{H}(X)$ and $e \in \mathbb{H}$, if $\lim_{n\to\infty} \|A_n - A\|_\Omega = 0$, we will also have

$$\lim_{n\to\infty} \langle e, \mathbb{E}[(A_n - A) \otimes (A_n - A)]e\rangle = 0.$$

Therefore, for $f, g \in L^2(\sigma_0)$ and $e \in \mathbb{H}$, we claim that

$$\langle e, \mathbb{E}[\mathscr{J}_X(f) \otimes \mathscr{J}_X(g)]e\rangle = \int f(x)g(x)\sigma_e(dx).$$

In fact, the equation above is obviously true for $f, g \in L_f$. Then the conclusion can be extended to whole $L^2(\sigma_0)$ by the fact that L_f is dense in $L^2(\sigma_0)$ and Cauchy-Schwartz inequality.

Now, combining with Lemma 2.8.12, we will have for any $e_1, e_2 \in \mathbb{H}$,

$$\langle e_1, \mathbb{E}[\mathscr{J}_X(f) \otimes \mathscr{J}_X(g)]e_2\rangle = \int f(x)g(x)\langle e_1, \sigma(dx)e_2\rangle.$$

Therefore,

$$\mathbb{E}[\mathscr{J}_X(f) \otimes \mathscr{J}_X(g)] = \int f(x)g(x)\sigma(dx) \text{ for all } f, g \in L_f. \tag{2.92}$$

Define ξ by

$$\xi(A) = \mathscr{J}_X(\mathbf{1}_A), A \in \mathscr{B}(\mathbb{R}).$$

By isometry, for any sequence of Borel sets A_n satisfying $A_n \to \emptyset$,

$$\|\xi(A_n)\|_\Omega = \|\mathbf{1}_{A_n}\|_{L^2(\sigma_0)} \to 0.$$

For any two Borel sets A and B such that their intersection is null, by (2.92), we have

$$\mathbb{E}[\xi(A)) \otimes (\xi(B))] = 0.$$

Finally, by (2.92), for any Borel set A,

$$\mathbb{E}[\xi(A)) \otimes (\xi(A))] = \int_A \sigma(dx).$$

Thus, we have proved $\xi(t)$ is a random orthogonal measure with structure measure σ and $\int e^{itx} d\xi(x)$ is therefore well-defined. It remains to show (2.39). Write

$$Z(t) := \mathscr{I}_\xi(e^{it\cdot}) = \int e^{itx} \xi(dx),$$

which is an element in $\mathbb{H}(\xi) \subset \mathbb{H}(X)$. We need to show that $Z = X$ as elements of $\mathbb{H}(X) \subset \mathbb{L}^2(\Omega)$. By (2.92), for any Borel set A,

$$\mathbb{E}[X(t) \otimes (\xi(A))] = \int e^{itx} \mathbf{1}_A(x) \sigma(dx).$$

For fixed s, let $g_n(x) = \sum_k c_k^n \mathbf{1}_{A_k^n}(x), n \in \mathbb{N}, A_k \cap A_l = \emptyset$ when $k \neq l$, be step functions converging to e^{isx} in $L^2(\sigma_0)$ as $n \to \infty$ satisfying conditions in Proposition 2.8.3. Then we have

$$\mathbb{E}[X(t) \otimes Z(s)] = \int e^{i(t-s)x} \sigma(dx) = \mathbb{E}[X(t) \otimes X(s)].$$

Thus, we have

$$\mathbb{E}\left[\sum_{i=1}^m c_i X(t_i) \otimes Y\right] = 0$$

for $Y := Z(s) - X(s)$, which is in $\mathbb{H}(X)$, and any linear combination $\sum_{i=1}^m c_i X(t_i)$. Since

linear combinations $\sum_{i=1}^{m} c_i X(t_i)$ are dense in $\mathbb{H}(X)$, we conclude that $\mathbb{E}(Y \otimes Y) = 0$ which implies that $Y = 0$.

To prove the uniqueness of ξ, let η be a random orthogonal measure on $\mathbb{R}$ with structure measure σ. Then we will have

$$\int e^{itx}\eta(dx) = \int e^{itx}\xi(dx).$$

As $\{e^{itx}\}$ is dense in $L^2(\sigma_0)$, by a similar arguments as (2.92), we have

$$\int h(x)\eta(dx) = \int h(x)\xi(dx), h \in L^2(\sigma_0).$$

Specially take $h = \mathbf{1}_B, B \in \mathscr{B}$, we obtain $\eta = \xi$.

The converse part should be trivial to prove.

□

2.8.5 PROOF for Theorem 2.5.5

For simplicity, we will only illustrate proofs on $\mathbb{R}$ and the extension to $\mathbb{R}^d$ should be straightforward.

Obviously conditions of Theorem 2.5.3 are satisfied and we can take σ to be the corresponding measure in that theorem, which is a finite T_+-valued measure. By (2.84), for any $\mu \in M_c(k) \cup \Lambda_k$, we have

$$\int \frac{e^{itx}|\widehat{\mu}(x)|^2}{1 \wedge |u|^{2k+2}}\sigma(dx) = \int e^{itx}\sigma_\mu(dx) - \sigma_\mu(\{0\}).$$

Take η to be the one in Lemma 2.8.24, by Theorem 2.5.4, we have a unique orthogonal random measure ξ_η with structure measure to be σ_η satisfying

$$X(\eta) = \int e^{itx}\xi_\eta(dx).$$

Now, define $\xi(\{0\}) = 0$ and for any $B \in \mathscr{B}$

$$\xi(B) = \int_{\mathbb{R}-\{0\}} \mathbf{1}_B(x)\frac{1 \wedge |x|^{k+1}}{\widehat{\eta}(x)}\xi_\eta(dx).$$

Then ξ is a well-defined orthogonal random measure with structure measure σ. Set

$$Y(t) := X(t) - \int \frac{e^{itx} - P(tx)}{1 \wedge |x|^{k+1}} d\xi(x).$$

The integral part exists by a very similar argument of Proposition 2.8.20. Therefore, for any $\mu \in \Lambda_k$, we have

$$Y(\mu) = \int e^{itx} d\xi_\mu(x) - \int_{\mathbb{R}-\{0\}} e^{itx} d\xi_\mu(x) = \xi_\mu(\{0\}). \tag{2.93}$$

By trivial extensions to the proof of Lemma C.9.4 in Sasvári (2013) toward mean-square continuity, Y is a random polynomial up to order $k+1$. The last property follows from (2.93) and ξ_μ is an orthogonal random measure.

Uniqueness can be proved in the same way as in the proof of Theorem 2.5.3. The converse part is trivial.

□

2.8.6 Sufficient conditions for the existence of tangent field for scalar Gaussian random field

In this section, we give some results about sufficient conditions for the existence of tangent field when the random field is a scalar Gaussian field. In the meanwhile, we will prove the existence of tangent field for those examples in Section 2.6.

According to Proposition 2.2.4, to prove the existence of tangent field, one needs to verify both finite dimensional convergence and some tightness condition. In many cases, the former is easier to be verified while the latter is much more difficult. The following lemma can be helpful in this aspect.

Lemma 2.8.28. *Let X_n, X be real-valued, zero-mean Gaussian random fields. If for any compact set $K \subset \mathbb{R}^d$, there exist $0 < C < \infty$ and $\gamma > 0$, such that for n large enough,*

$$\mathbb{E}[(X_n(s) - X_n(t))^2] \leq \frac{C}{|\log|s-t||^{1+\gamma}}, \text{ where } s, t \in K,$$

then we have for any $\eta, \epsilon > 0$, there exists $\delta > 0$, s.t.

$$\limsup_{n\to\infty} \mathbb{P}\left(\sup_{|s-t|<\delta, s,t\in K} |X_n(s) - X_n(t)| > \eta \right) < \epsilon.$$

Proof. This lemma is a direct result of the following fact.

Fact 2.8.29. *Let K be a compact subset of $\mathbb{R}^d$ and let $X(t), t \in K$ be a real-valued, zero-mean, Gaussian random field with continuous covariance function. If for some $0 < C < \infty$ and $\gamma > 0$,*

$$\mathbb{E}[(X(s) - X(t))^2] \leq \frac{C}{|\log |s-t||^{1+\gamma}}, \tag{2.94}$$

for all $s, t \in K$, then with probability one, X has continuous sample path over K. Furthermore, for any $\eta > 0$, we have

$$\omega(\delta, \eta, C, \gamma) := \delta^{-d}\mathbb{P}\left(\sup_{||t-s||<\delta} |X(t) - X(s)| > \eta\right) \to 0, \text{ as } \delta \to 0.$$

Its proof can be found in Theorem 3.4.1 in Adler (2010). □

In the next two propositions, we provide some sufficient conditions for showing the existence of tangent fields.

Proposition 2.8.30. *Assume $X(t)$ is a zero-mean Gaussian random fields on $\mathbb{R}^d$ with a covariance function satisfying the following decomposition in a bounded neighborhood V around $z \in \mathbb{R}^d$:*

$$C(z + x, z + y) = P_z(x, y) + H_z(x, y) + R_z(x, y),$$

where $P_z(x, y)$ is a polynomial function of x, y up to order $2k + 1, k \in \mathbb{N}$, $H_z(x - y)$ is defined to be

$$H_z(x, y) = \int_{S^{d-1}} |\langle x - y, \theta\rangle|^{2\alpha(z)} L_z(x, y, \theta)\sigma_z(d\theta),$$

$\alpha(z) \in (k, k + 1)$ and $L_z(x, y)$, $R_z(x, y)$ are symmetric, continuous and satisfy the following condition:

L1 : There exists function $\widetilde{L}_z(r)$ and a null set $\Delta \in S^{d-1}$, such that $\widetilde{L}_z(1/x)$ is slowly varying and for any $x, y \in \overline{V} - z$, $x \neq y$, $\theta \in S^{d-1} - \Delta$

$$\lim_{r\to 0} \frac{L_z(rx, ry, \theta)}{\widetilde{L}_z(r)} = 1.$$

L2 : For any $\delta > 0$, there exists a constant $C > 0$ and a null set $\Delta \in S^{d-1}$, such that

uniformly for any $x, y \in \overline{V} - z$ *and* $\theta \in S^{d-1} \setminus \Delta$,

$$\limsup_{r \to 0} |\langle x - y, \theta \rangle|^{\delta} \frac{|L_z(rx, ry, \theta)|}{|\widetilde{L}_z(r)|} < C.$$

L3 : For any $\delta > 0$, *there exist constants* $C, \gamma > 0$ *and a null set* $\Delta \in S^{d-1}$, *such that for all* $a \in \overline{V} - z$ *and* $r > 0$ *small enough,*

$$|\langle y - a, \theta \rangle|^{\delta} \left| \frac{L_z(ry, ra, \theta)}{\widetilde{L}_z(r)} - \frac{L_z(rx, ra, \theta)}{\widetilde{L}_z(r)} \right| \leq \frac{C}{|\log |x - y||^{1+\gamma}}$$

uniformly for $\theta \in S^{d-1} \setminus \Delta$, $x, y \in \overline{V} - z$ *and* $|\langle y - a, \theta \rangle| < |\langle x - a, \theta \rangle|$.

R1 : For any $x, y \in \overline{V} - z$,

$$\lim_{r \to 0} \frac{R_z(rx, ry)}{r^{2\alpha(z)} \widetilde{L}_z(r)} = 0.$$

R2 : There exist constants $C, \gamma > 0$ *such that for* $r > 0$ *small enough,*

$$\left| \frac{R_z(rx, ry)}{r^{2\alpha(z)} \widetilde{L}_z(r)} - \frac{R_z(rx, rx)}{r^{2\alpha(z)} \widetilde{L}_z(r)} \right| \leq \frac{C}{|\log |x - y||^{1+\gamma}}$$

uniformly for $x, y \in \overline{V} - z$.

R3 : There exist constants $C, \gamma > 0$ *such that for any* $a \in \overline{V} - z$ *and* $r > 0$ *small enough,*

$$\left| \frac{R_z(ry, ra)}{r^{2\alpha(z)} \widetilde{L}_z(r)} - \frac{R_z(rx, ra)}{r^{2\alpha(z)} \widetilde{L}_z(r)} \right| \leq \frac{C}{|\log |x - y||^{1+\gamma}}$$

uniformly for $x, y \in \overline{V} - z$.

Then $X(t)$ *has any order of tangent field at location* t. *If the order is no larger than* $k - 1$, *the tangent field is a random polynomial. If the order is greater than* $k - 1$, *the tangent field has a generalized covariance function* $K_z(h) = \int_{S^{d-1}} |\langle h, \theta \rangle|^{2\alpha(z)} \sigma_z(d\theta)$.

Remark.

- L1-L3 are designed such that $L(x, y, \theta) = \log(|\langle x - y, \theta \rangle|)$ can satisfy.
- R1-R3 are designed such that $R(x, y) = |x - y|^{2\alpha(z) + O(r)} - |x - y|^{2\alpha(z)}$ can satisfy.

The proof of Proposition 2.8.30 requires the following lemma.

Lemma 2.8.31. *For any compact set $K \subset \mathbb{R}^d$ and homogeneous function*

$$H(x) = \int_{S^{d-1}} |\langle x, \theta \rangle|^\nu \sigma(d\theta), \nu > 0,$$

there exists a constant $C > 0$ such that

$$|H(s) - H(t)| \leq C|s - t|^{\nu \wedge 1} \tag{2.95}$$

uniformly for $s, t \in K$.

Proof. We first prove the case with $d = 1$, that is

$$||s|^\nu - |t|^\nu| \leq C|s - t|^{\nu \wedge 1}.$$

WLOG, assume $s, t \geq 0$. The conclusion is obvious for $\nu \leq 1$ as the function $f(x) = x^\nu$ has non-increasing derivative on $(0, \infty)$. When $\nu > 1$, as $s, t \in K$, $f'(x)$ is bounded on K and thus $|s|^\nu - |t|^\nu \leq C|s - t|$.

For a general homogeneous function $H(x)$, we have

$$\begin{aligned} |H(s) - H(t)| &\leq \int_{S^{d-1}} ||\langle s, \theta \rangle|^\nu - |\langle t, \theta \rangle|^\nu | \sigma(d\theta) \\ &\leq C' \int_{S^{d-1}} |\langle s - t, \theta \rangle|^{\nu \wedge 1} \sigma(d\theta) \\ &\leq C|s - t|^{\nu \wedge 1}. \end{aligned}$$

□

PROOF for Proposition 2.8.30:

Step 1: Verifying convergence in finite dimensional distribution.

As $X(t)$ is a random field in Gaussian, this is equivalent to verify the convergence in covariance function.

Let $\lambda^{(1)}, \lambda^{(2)} \in \Lambda_l$ and $\lambda^{(i)} = \sum_l c_l^{(i)} \delta_{x_l^{(i)}}, i = 1, 2$.

If $l \geq k$, by assumptions [L1], [L2], [R1] and dominant convergence theorem, as $r \to 0$,

we have

$$\begin{aligned}
&\frac{\mathrm{Cov}(X(z+r\lambda^{(1)}),X(z+r\lambda^{(2)}))}{r^{2\alpha(z)}\widetilde{L}_z(r)}\\
=\;&\sum_{i,j}\frac{c_i^{(1)}c_j^{(2)}}{r^{2\alpha(z)}\widetilde{L}_z(r)}\left[H_z(rx_i^{(1)},rx_j^{(2)})+R_z(rx_i^{(1)},rx_j^{(2)})\right]\\
\to\;&\sum_{i,j}c_i^{(1)}c_j^{(2)}\int_{S^{d-1}}|\langle x_i^{(1)}-x_j^{(2)},\theta\rangle|^{2\alpha(z)}\sigma_z(d\theta),
\end{aligned}$$

which corresponds to the generalized covariance function $K(h)=\int_{S^{d-1}}|\langle h,\theta\rangle|^{2\alpha(z)}\sigma_z(d\theta)$.

If $l\le k-1$, by the same step we can have

$$\frac{\mathrm{Cov}(X(z+r\lambda^{(1)}),X(z+r\lambda^{(2)}))}{r^{2l+2}}\;\to\;\lim_{r\to 0}P_z(r\lambda^{(1)},r\lambda^{(2)})/r^{2l+2},$$

which is a function invariant to the shift of $\lambda^{(1)}$ and $\lambda^{(2)}$ and corresponds to the generalized covariance function $K_z(h)=|h|^{2l+2}$.

Step 2: Verifying the tightness condition.

We will use the sufficient conditions in Lemma 2.8.28 for the proof below.

We only consider the case when $l\ge k$, because the other situation will be trivial based on the proof for this case. For $s,t\in\overline{V}-z$ and $r>0$, denote

$$\lambda_s=\sum_i c_i(s)\delta_{x_i(s)}$$

and

$$C_r(s,t)=\sum_{i,j}c_i(s)c_j(t)\frac{H_z(rx_i(s),rx_j(t))+R_z(rx_i(s),rx_j(t))}{r^{2\alpha(z)}\widetilde{L}_z(r)}.$$

By Lemma 2.8.28, we only need to prove for any compact set K, there exists $C,\gamma>0$, such that

$$|C_r(s,t)-C_r(t,t)|\le\frac{C}{|\log|s-t||^{1+\gamma}}$$

uniformly for $s,t\in K$ with $r>0$ small enough.

Denote

$$A_{ij}^{(1)}(s,t)=c_i(s)c_j(t)\frac{H_z(rx_i(s),rx_j(t))}{r^{2\alpha(z)}\widetilde{L}_z(r)}$$

and

$$A_{ij}^{(2)}(s,t) = c_i(s)c_j(t)\frac{R_z(rx_i(s), rx_j(t))}{r^{2\alpha(z)}\widetilde{L}_z(r)}.$$

We only need to prove for any compact set K, there exists $C, \gamma > 0$, such that

$$|A_{ij}^{(m)}(s,t) - A_{ij}^{(m)}(t,t)| \leq \frac{C}{|\log|s-t||^{1+\gamma}}, i,j = 0,\ldots,M_l, m = 1,2.$$

We can easily observe the following facts:

- $x_0(s) = s$ and $x_i(s) = a_i, i \neq 0$, where $\{a_i\}_{i=1,\ldots,M_l}$ are those reference points for defining λ_s.
- $c_0(s) = 1$ and $c_i(s)$ is a polynomial function of s if $i \neq 0$.

If $i = j \neq 0$, then we have

$$A_{ij}^{(1)}(s,t) - A_{ij}^{(1)}(t,t) = 0,$$

and by [R1],

$$|A_{ij}^{(2)}(s,t) - A_{ij}^{(2)}(t,t)| = \left|(c_i(s) - c_i(t))c_j(t)\frac{R_z(ra_i, ra_j)}{r^{2\alpha(z)}\widetilde{L}_z(r)}\right| \leq C|s-t|.$$

If $i = j = 0$, by [L2],

$$\begin{aligned}
|A_{ij}^{(1)}(s,t) - A_{ij}^{(1)}(t,t)| &\leq C \sup_{\theta\in S^{d-1}\setminus\Delta} |\langle s-t,\theta\rangle|^{2\alpha(z)}\frac{|L_z(rs,rt,\theta)|}{|\widetilde{L}_z(r)|} \\
&\leq C \sup_{\theta\in S^{d-1}\setminus\Delta} |s-t|^{2\alpha(z)-\delta}|\langle s-t,\theta t\rangle|^{\delta}\frac{|L_z(rs,rt,\theta)|}{|\widetilde{L}_z(r)|} \\
&\leq C|s-t|^{2\alpha(z)-\delta},
\end{aligned}$$

and by [R2],

$$|A_{ij}^{(2)}(s,t) - A_{ij}^{(2)}(t,t)| = \frac{|R_z(rs,rt) - R_z(rt,rt)|}{r^{2\alpha(z)}\widetilde{L}_z(r)} \leq \frac{C}{|\log|s-t||^{1+\gamma}}.$$

If $i \neq j$ and $i = 0$, for $|t - a_j| < |s - a_j|$, by [L1], [L3] and Lemma 2.8.31,

$$\begin{aligned}
& |A_{ij}^{(1)}(s,t) - A_{ij}^{(1)}(t,t)| \\
= \; & \frac{|c_j(t)|}{r^{2\alpha(z)}\widetilde{L}_z(r)} |H_z(s, a_j) - H_z(t, a_j)| \\
\leq \; & \frac{|c_j(t)|}{\widetilde{L}_z(r)} \Big(\Big| \int_{S_s^{d-1}} (L_z(rs, ra_j, \theta) - L_z(rt, ra_j, \theta)) |\langle s - a_j, \theta \rangle|^{2\alpha(z)} \sigma_z(d\theta) \Big| \\
& + \Big| \int_{S_t^{d-1}} (L_z(rs, ra_j, \theta) - L_z(rt, ra_j, \theta)) |\langle t - a_j, \theta \rangle|^{2\alpha(z)} \sigma_z(d\theta) \Big| \\
& + \Big| \int_{S_s^{d-1}} L_z(rt, ra_j, \theta) (|\langle s - a_j, \theta \rangle|^{2\alpha(z)} - |\langle t - a_j, \theta \rangle|^{2\alpha(z)}) \sigma_z(d\theta) \Big| \\
& + \Big| \int_{S_t^{d-1}} L_z(rs, ra_j, \theta) (|\langle s - a_j, \theta \rangle|^{2\alpha(z)} - |\langle t - a_j, \theta \rangle|^{2\alpha(z)}) \sigma_z(d\theta) \Big| \Big) \\
\leq \; & \frac{C}{|\log |s - t||^{1+\gamma}} + C|s - t|^{2\alpha(z) \wedge 1},
\end{aligned}$$

where $S_t^{d-1} = \{\theta : |\langle s - a, \theta \rangle| \geq |\langle t - a, \theta \rangle|, \theta \in S^{d-1}\}$ and a similar definition for S_s^{d-1}.

For $A_{ij}^{(2)}(s,t)$, it is a direct result of [R3].

If $i \neq j$ and $i \neq 0$, by [L2] and smoothness of $c_i(\cdot)$ function,

$$\begin{aligned}
|A_{ij}^{(1)}(s,t) - A_{ij}^{(1)}(t,t)| & \leq C|c_i(s) - c_i(t)||c_j(t)||H_z(ra_i, rx_j(t))/\widetilde{L}_z(r)/r^{2\alpha(z)} \\
& \leq C|t - s|^{1-\delta},
\end{aligned}$$

and

$$|A_{ij}^{(2)}(s,t) - A_{ij}^{(2)}(t,t)| = |c_i(s) - c_i(t)||c_j(t)| \frac{R_z(ra_i, rx_j(t))}{r^{2\alpha(z)}\widetilde{L}_z(r)} \leq C|t - s|.$$

By now, we have discussed all different situations and verify the conditions for tightness.

□

Proposition 2.8.32. *Assume $X(t)$ is a zero-mean Gaussian random fields on $\mathbb{R}^d$ with a covariance function satisfying the following decomposition in a bounded neighborhood V of $z \in \mathbb{R}^d$:*

$$C(z + x, z + y) = P_z(x, y) + H_z(x, y) + R_z(x, y),$$

where $P_z(x, y)$ is a polynomial function of x, y up to order $2k + 1, k \in \mathbb{N}_+$, $H_z(x - y)$ is

defined to be

$$H_z(x,y) = \int_{S^{d-1}} |\langle x-y,\theta\rangle|^{2\alpha(z)} L_z(x,y,\theta)\sigma_z(d\theta),$$

$\alpha(z) = k > 0$ *and* $L_z(x,y)$, $R_z(x,y)$ *are continuous, symmetric and satisfy the following condition:*

LL1 : There exist functions $\widetilde{L}'_z(r)$ *and constant* C_z*, such that for any* $x, y \in \overline{V} - z$*,*

$$\lim_{r\to 0} \int_{S^{d-1}} L_z(rx,ry,\theta) - \widetilde{L}'_z(r) - C_z \log(|\langle x-y,\theta\rangle|)\sigma_z(d\theta) = 0,$$

and for some $x, y \in \overline{V} - z$

$$\int_{S^{d-1}} |\langle x-y,\theta\rangle|^{2k} C_z \log(\langle x-y,\theta\rangle)\sigma_z(d\theta) \neq 0.$$

LL2 : For any $\delta > 0$*, there exists a constant* $C > 0$ *and a null set* $\Delta \subset S^{d-1}$*, such that uniformly for any* $x, y \in \overline{V} - z$ *and* $\theta \in S^{d-1} \setminus \Delta$*,*

$$\limsup_{r\to 0} |\langle x-y,\theta\rangle|^{\delta} |L_z(rx,ry,\theta) - L'_z(r)| < C.$$

LL3 : There exist constants $C, \gamma > 0$ *and a null set* $\Delta \subset S^{d-1}$*, such that for any* $\delta > 0$ *and* $r > 0$ *small enough,*

$$|\langle y-a,\theta\rangle|^{\delta} \, |L_z(ry,ra,\theta) - L_z(rx,ra,\theta)| \leq \frac{C}{|\log|x-y||^{1+\gamma}}$$

uniformly for $\theta \in S^{d-1} \setminus \Delta$*,* $a, x, y \in \overline{V} - z$ *and* $|\langle y-a,\theta\rangle| < |\langle x-a,\theta\rangle|$*.*

RR $R_z(rx,ry)$ *satisfies [R1]-[R3] in Proposition 2.8.30 with* $\widetilde{L}_z(r) = 1$*.*

Then $X(t)$ *has any order of tangent field at location* t*. If the order is no larger than* $k-1$*, the tangent field is a random polynomial. If the order is no smaller than* k*, the tangent field has a generalized covariance function* $K_z(h) = C_z \int_{S^{d-1}} |\langle h,\theta\rangle|^{2\alpha(z)} \log|\langle h,\theta\rangle| \sigma_z(d\theta)$*.*

Proof. We only show finite dimensional convergence when the order is no smaller than k, because other parts will be almost the same as the proof of Proposition 2.8.30.

Let $\lambda^{(1)}, \lambda^{(2)} \in \Lambda_l$ and $\lambda^{(i)} = \sum_l c_l^{(i)} \delta_{x_l^{(i)}}, i = 1, 2.$

By assumptions [L1], [R1], as $r \to 0$,

$$\begin{aligned}
& \frac{\mathrm{Cov}(X(z + r\lambda^{(1)}), X(z + r\lambda^{(2)}))}{r^{2k}} \\
= & \sum_{i,j} \frac{c_i^{(1)} c_j^{(2)}}{r^{2k}} \left[H_z(rx_i^{(1)}, rx_j^{(2)}) + R_z(rx_i^{(1)}, rx_j^{(2)}) \right] \\
= & \sum_{i,j} \frac{c_i^{(1)} c_j^{(2)}}{r^{2k}} \Bigg[\int_{S^{d-1}} |\langle rx_i^{(1)} - rx_j^{(2)}, \theta \rangle|^{2k} \widetilde{L}_z'(r) \sigma_z(d\theta) \\
& + \int_{S^{d-1}} |\langle rx_i^{(1)} - rx_j^{(2)}, \theta \rangle|^{2k} (L_z(rx_i, rx_j, \theta) - \widetilde{L}_z'(r)) \sigma_z(d\theta) \\
& + R_z(rx_i^{(1)}, rx_j^{(2)}) \Bigg] \\
\to & \sum_{i,j} c_i^{(1)} c_j^{(2)} \int_{S^{d-1}} |\langle x_i^{(1)} - x_j^{(2)}, \theta \rangle|^{2k} C_z \log(|\langle x_i - x_j, \theta \rangle|) \sigma_z(d\theta),
\end{aligned}$$

which corresponds to the generalized covariance function $K(h) = C_z \int_{S^{d-1}} |\langle h, \theta \rangle|^{2k} \log |\langle h, \theta \rangle| \sigma_z(d\theta)$. □

The following two lemmas can be useful when we try to verify conditions of the previous two Propositions.

Lemma 2.8.33. *Let $f(s,t) \in C_{2k+3}(\overline{V}^2)$, where V is a bounded neighborhood around z. Denote the reminder of $(2k+1)$-order Taylor expansion of $f(s,t)$ at location (z,z) to be $R_1(s,t)$. Then $R_1(s,t)$ satisfies [R1], [R2] and [R3].*

Proof. We only need to prove for $\widetilde{L} = 1$, as in general case where $\widetilde{L}$ is a slowly varying function, the proof will be similar.

By Taylor Theorem in multiple dimensions, we have

$$R_1(s,t) = \sum_{|\beta|=2k+2} R_\beta(s,t)(s,t)^\beta,$$

and

$$R_\beta(s,t) = \frac{|\beta|}{\beta!} \int_0^1 (1-u)^{|\beta|-1} D^\beta f((z,z) + u(s,t)) du.$$

It is easy to know $R_\beta(s,t) \in C_1((\overline{V} - z)^{\otimes 2})$ and thus is bounded for $s, t \in \overline{V} - z$.

$R_1(rs, rt) \sim r^{2k+2}$ and thus [R1] is obvious.

For any $s,t \in \overline{V}-z$,

$$\frac{R_1(rs,rt)-R_1(rt,rt)}{r^{2H(z)}} = r^{2k+2-2\alpha(z)} \sum_{|\beta|=2k+2} (R_\beta(rs,rt)(s,t)^\beta - R_\beta(rt,rt)(t,t)^\beta).$$

As $R_\beta(s,t) \in C_1((\overline{V}-z)^{\otimes 2})$, [R2] follows directly. [R3] can be proved in the same way. □

Lemma 2.8.34. *Denote*

$$R_2(s,t) = D(s,t)\int_{S^{d-1}} |\langle s-t,\theta\rangle|^{\alpha(s,t)}\sigma_{s,t}(d\theta) - D(0,0)\int_{S^{d-1}} |\langle s-t,\theta\rangle|^{\alpha(0,0)}\sigma_{0,0}(d\theta),$$

where $D(s,t), \alpha(s,t) \in C_1(\overline{V}^{\otimes 2})$, and V is a bounded neighborhood around 0 and $\alpha(0,0) > 0$.For $\sigma_{s,t}$, assume $\sigma_{s,t} = \sigma_{t,s}$, $\int_{S^{d-1}} \sigma_{s,t}(d\theta) < \infty$, $\int_{S^{d-1}} \sigma_{s,t}(d\theta)$ continuous and for any $a \in V$ and $\delta > 0$, there exist $C, \gamma > 0$, such that

$$\left|\int_{S^{d-1}} |\langle t-a,\theta\rangle|^\delta(\sigma_{s,a}(d\theta)-\sigma_{t,a}(d\theta))\right| \le \frac{C}{|\log|s-t||^{1+\gamma}},$$

for all $s,t \in \overline{V}$. Then we have R_2 satisfies [R1], [R2] and [R3] with $\widetilde{L}_z(r) = 1$.

Proof. For [R2], we have

$$\begin{aligned}\frac{|R_2(rs,rt)-R_2(rt,rt)|}{r^{\alpha(0,0)}} &= \Big|D(rs,rt)\int_{S^{d-1}} |\langle s-t,\theta\rangle|^{\alpha(rs,rt)} r^{\alpha(rs,rt)-\alpha(0,0)}\sigma_{rs,rt}(d\theta) \\ &\quad -D(0,0)\int_{S^{d-1}} |\langle s-t,\theta\rangle|^{\alpha(0,0)}\sigma_{0,0}(d\theta)\Big| \\ &\le C|s-t|^{\alpha(0,0)-\delta},\end{aligned}$$

for some $\delta \in (0,\alpha(0,0))$.

For [R3], we have

$$\begin{aligned}\frac{|R_2(rs,ra)-R_2(rt,ra)|}{r^{\alpha(0,0)}} =& \Big|D(rs,ra)\int_{S^{d-1}} |\langle s-a,\theta\rangle|^{\alpha(rs,ra)} r^{\alpha(rs,ra)-\alpha(0,0)}\sigma_{rs,ra}(d\theta) \\ &- D(rt,ra)\int_{S^{d-1}} |\langle t-a,\theta\rangle|^{\alpha(rt,ra)} r^{\alpha(rt,ra)-\alpha(0,0)}\sigma_{rt,ra}(d\theta) \\ &+ D(0,0)\int_{S^{d-1}} |\langle t-a,\theta\rangle|^{\alpha(0,0)}\sigma_{0,0}(d\theta) \\ &- D(0,0)\int_{S^{d-1}} |\langle s-a,\theta\rangle|^{\alpha(0,0)}\sigma_{0,0}(d\theta)\Big|.\end{aligned} \tag{2.96}$$

One can show the last two lines satisfy [R3] by using Lemma 2.8.31. For the first two lines,

by $D(\cdot,\cdot) \in C_1(\overline{V}^{\otimes 2})$, we know

$$\begin{aligned}&\left|(D(rs,ra)-D(rt,ra))\int_{S^{d-1}}|\langle s-a,\theta\rangle|^{\alpha(rs,ra)}r^{\alpha(rs,ra)-\alpha(0,0)}\sigma_{rs,ra}(d\theta)\right| \\ &\leq C|s-t|.\end{aligned} \tag{2.97}$$

By mean value theorem,

$$|r^{\alpha(rs,ra)-\alpha(0,0)}-r^{\alpha(rt,ra)-\alpha(0,0)}| = |\log(r)r^{\alpha'}(\alpha(rs,ra)-\alpha(rt,ra))| \leq C|s-t|, \tag{2.98}$$

where α' is between $\alpha(rs,ra)-\alpha(0,0)$ and $\alpha(rt,ra)-\alpha(0,0)$.

Still by mean value theorem, uniformly for all $\theta \in S^{d-1}$,

$$\begin{aligned}&||\langle s-a,\theta\rangle|^{\alpha(rt,ra)}-|\langle s-a,\theta\rangle|^{\alpha(rs,ra)}| \\ =&|\log(|\langle s-a,\theta\rangle|)|\langle s-a,\theta\rangle|^{\alpha'}(\alpha(rs,ra)-\alpha(rt,ra))| \leq C|s-t|,\end{aligned} \tag{2.99}$$

where α' is between $\alpha(rs,ra)$ and $\alpha(rt,ra)$.

By Lemma 2.8.31, uniformly for all $\theta \in S^{d-1}$,

$$||\langle s-a,\theta\rangle|^{\alpha(rt,ra)}-|\langle t-a,\theta\rangle|^{\alpha(rt,ra)}| \leq C|s-t|^{\alpha(rt,ra)\wedge 1}. \tag{2.100}$$

Finally, by assumptions about $\sigma_{s,t}$,

$$\left|\int_{S^{d-1}}|\langle t-a,\theta\rangle|^{\alpha(rt,ra)}(\sigma_{s,a}(d\theta)-\sigma_{t,a}(d\theta))\right| \leq \frac{C}{|\log|s-t||^{1+\gamma}}. \tag{2.101}$$

With equations (2.97)-(2.101), we can prove the first two lines of (2.96) also satisfy condition [R3].

Due to the continuity of $D(s,t)$, $\alpha(s,t)$, $\int_{S^{d-1}}\sigma_{s,t}(d\theta)$ and very similar derivations as in (2.98) and (2.99), the condition [R1] is easy to be verified. □

Example 2.8.1. *(mBm) Multifractional Brownian motion $X(t)$ is defined to be a zero-mean Gaussian Random field on $(0,\infty)$ with covariance function to be*

$$C(s,t) = D(s,t)(|s|^{H(s)+H(t)}+|t|^{H(s)+H(t)}-|s-t|^{H(s)+H(t)}).$$

See details in Herbin (2006). Assume $H(s) \in C_3(V)$ and $H(s) \in (0,1)$, where V is a bounded neighborhood around $z > 0$. Then the lth order tangent field of X at location z is fractional Brownian motion with hurst index $H(z)$, where $l \geq 0$.

Proof. Take V to be a bounded neighborhood around z which doesn't contain 0. It is easy

to see $D(z+s,z+t)(|z+s|^{H(z+s)+H(z+t)}+|t|^{H(z+s)+H(z+t)}) \in C_3((\overline{V}-z)^{\otimes 2})$. Thus, by Lemma 2.8.33, we only need to prove

$$R_2(s,t) := D(z+s,z+t)|s-t|^{H(z+s)+H(z+t)} - D(z,z)|s-t|^{2H(z)}$$

satisfies [R1], [R2] and [R3], which is a direct result of Lemma 2.8.34.

□

Example 2.8.2. *n-mBm. Let*

$$X_H(t) = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{e^{it\omega} - \sum_{k=0}^{n-1}\frac{(it\omega)^k}{k!}}{|\omega|^{H(t)+1/2}} dW(\omega),$$

where $H(t) \in (n-1,n)$.

With similar calculations in Equation (12) in Perrin et al. (2001), one can show that the covariance function is

$$\begin{aligned} C(t,s) &= (-1)^n \frac{C_H^n(s,t)}{2}\Bigg\{|t-s|^{H(t)+H(s)} \\ &\quad -\sum_{j=0}^{n-1}(-1)^j \binom{H(t)+H(s)}{j}\left[\left(\frac{t}{s}\right)^j |s|^{H(t)+H(s)} + \left(\frac{s}{t}\right)^j |t|^{H(t)+H(s)}\right]\Bigg\}, \end{aligned}$$

where

$$C_H^n = \frac{1}{\Gamma(H(s)+H(t)+1)|\sin(\pi H)|}.$$

Therefore, if $H(t) \in C_{2n+1}$, by Lemma 2.8.33 and Lemma 2.8.34, we know the generalized covariance of $n-1$th order tangent field at location $t>0$ is $|h|^{2H(t)}$.

Similarly, one can prove the tangent field of the following process at location t when t is bounded away from those reference points of $\lambda_t \in \Lambda_{n-1}$ and 0, and $H(t)$ is not an integer,

$$\widetilde{X}_H(t) = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{e^{i\omega\lambda_t}}{(i\omega)^{H(t)+1/2}} dW(\omega) = \frac{1}{2\pi}\int_{\mathbb{R}} \frac{\left(e^{i\omega u} - \sum_{k=0}^{n-1}\frac{(i\omega u)^k}{k!}\right)\Big|_{u=\lambda_t}}{(i\omega)^{H(t)+1/2}} dW(\omega).$$

To prove the existence of tangent fields at location t_0 where $H(t_0) \in \mathbb{Z}$, we need the following two lemmas.

Lemma 2.8.35. *For any compact set $K \subset \mathbb{R}$, $\nu > 0$, there exists constant $C > 0$ and $\gamma > 0$*

such that

$$|s|^{\nu}|\log|s|-\log|t||\leq\frac{C}{|\log|s-t||^{1+\gamma}} \tag{2.102}$$

uniformly for $s,t\in K$ *and* $|s|\leq|t|$.

Proof. WLOG, assume $s,t\geq 0$. If $s\leq t/2$, as $t\log t$ is bounded, there exist $\delta\in(0,\nu/2)$, $C',C''>0$ such that

$$\begin{aligned}|s|^{\nu}|\log|s|-\log|t||&\leq|s|^{\nu/2}|t|^{\nu/2}|\log|s|-\log|t||\leq C's^{\nu/2-\delta}\leq C''|t-s|^{\nu/2-\delta}\\&\leq\frac{C}{|\log|s-t||^{1+\gamma}}.\end{aligned}$$

If $s>t/2$, for any $\gamma>0$, we have

$$\begin{aligned}&|s|^{\nu}|\log|s|-\log|t|||\log|s-t||^{1+\gamma}\\=\;&s^{\nu}\log(t/s)|\log(t/s-1)+\log(s)|^{1+\gamma}\\\leq\;&C's^{\nu}\log(t/s)(|\log(t/s-1)|^{1+\gamma}+|\log(s)|^{1+\gamma}).\end{aligned}$$

As $s\in(t/2,t)$, we have $\log(t/s)\in(0,\log 2)$. Therefore, $s^{\nu}\log(t/s)|\log(s)|^{1+\gamma}<\infty$.

For $\log(t/s)|\log(t/s-1)|^{1+\gamma}$, by changing variables with $t/s-1=u\in(0,1)$, we have it is equal to $|\log(1+u)\log^{1+\gamma}(u)|$, which is bounded as $\log(1+u)\sim u$ when $u\to 0$.

Thus $|s|^{\nu}|\log|s|-\log|t|||\log|s-t||^{1+\gamma}$ is bounded for any $\gamma>0$, which completes the proof of this lemma. □

Lemma 2.8.36. *For* $x\in(0,1]$, $\gamma\in\mathbb{R}$, *define*

$$K(x,\gamma)=\begin{cases}\frac{x^{\gamma}-1}{\gamma} & \text{if } \gamma\neq 0,\\ \log x & \text{if } \gamma=0.\end{cases}$$

Then $K(x,\gamma)$ *has the following properties:*

a. For any $\delta>\delta_0>0$, *there exists* $C>0$ *such that*

$$\sup_{x\in(0,1],|\gamma|<\delta_0}|K(x,\gamma)|x^{\delta}\leq C.$$

b. For any $\delta>\delta_0>0$, *there exists* $C>0$ *such that*

$$\sup_{x\in(0,1],|\gamma|<\delta_0}|K(x,\gamma)-\log x|x^{\delta}/|\gamma|\leq C.$$

c. *$K(x,\gamma)$ is a continuous function.*

d. *For $\forall x \neq 0$, $\frac{\partial K}{\partial \gamma}$ is continuous. Furthermore, for any $\delta > \delta_0 > 0$, there exists some constant $C > 0$ such that*

$$\sup_{x\in(0,1],|\gamma|<\delta_0} \left|\frac{\partial K}{\partial \gamma}(x,\gamma)\right| x^{\delta} < C.$$

Proof.

Part a

The conclusion is obvious when $\gamma = 0$, and we only prove the case when $\gamma \neq 0$.

Let $f_{\gamma,1}(x) = K(x,\gamma)x^{\delta}$. For $\gamma \neq 0$, we will have

$$f'_{\gamma,1}(x) = x^{\gamma+\delta-1} + \frac{\delta}{\gamma}x^{\delta-1}(x^{\gamma}-1),$$

$$\lim_{x\to 0} f_{\gamma,1}(x) = 0,$$

and

$$\lim_{x\to 1} f_{\gamma,1}(x) = 0.$$

Thus, for any given γ, $|f_{\gamma,1}(x)|$ is maximized when

$$f'_{\gamma,1}(x) = 0,$$

or equivalently,

$$x^{\gamma} = \frac{\delta}{\delta+\gamma}.$$

Therefore,

$$\sup_{x\in(0,1],|\gamma|<\delta_0,\gamma\neq 0} |K(x,\gamma)|x^{\delta} \leq \sup_{x\in(0,1],|\gamma|<\delta_0,\gamma\neq 0} \frac{x^{\delta}}{\delta+\gamma} \leq \frac{1}{2\delta}.$$

Part b

This conclusion is obvious for $\gamma = 0$, and we only prove the case when $\gamma \neq 0$.

For any fixed $\gamma \neq 0$, denote $f_{\gamma,2}(x) = (K(x,\gamma) - \log x)x^\delta/\gamma$. It is easy to see

$$f'_{\gamma,2}(x) = \frac{x^{\delta-1}\delta}{\gamma}\left(\frac{\delta+\gamma}{\delta\gamma}(x^\gamma - 1) - \log(x)\right),$$

$$f_{\gamma,2}(1) = 0, \tag{2.103}$$

and

$$\lim_{x\to 0} f_{\gamma,2}(x) = 0. \tag{2.104}$$

Therefore, for any given $\gamma > 0$, $|f_{\gamma,2}|$ is maximized when $f'_{\gamma,2} = 0$, or equivalently,

$$x^\gamma - 1 = \frac{\delta\gamma}{\delta+\gamma}\log(x). \tag{2.105}$$

Therefore,

$$\sup_{x\in(0,1],|\gamma|<\delta_0,\gamma\neq 0} \frac{|K(x,\gamma) - \log x|x^\delta}{|\gamma|} \leq \sup_{x\in(0,1],|\gamma|<\delta_0,\gamma\neq 0} \frac{|\frac{\delta}{\delta+\gamma}\log x - \log x|x^\delta}{|\gamma|}$$
$$= \sup_{x\in(0,1],|\gamma|<\delta_0,\gamma\neq 0} \frac{x^\delta|\log x|}{\delta+\gamma} < \infty.$$

part c

By **part b**, it is easy to see for any $\delta > 0$, $K(x,\gamma)x^\delta$ is a continuous function and thus $K(x,\gamma)$ is a continuous function.

part d

If $\gamma \neq 0$,

$$\frac{\partial K(x,\gamma)}{\partial\gamma}(x,\gamma) = \frac{x^\gamma\gamma\log x - x^\gamma + 1}{\gamma^2} = \frac{x^\gamma\log x^\gamma - x^\gamma + 1}{\gamma^2}. \tag{2.106}$$

Notice that when $\gamma > 0$, $\frac{\partial K(x,\gamma)}{\partial\gamma}(x,\gamma) \leq 0$ and when $\gamma < 0$, $\frac{\partial K(x,\gamma)}{\partial\gamma}(x,\gamma) \geq 0$.

If $\gamma = 0$, by using L'Hospital's rule twice,

$$\frac{\partial K(x,\gamma)}{\partial\gamma}(x,\gamma) = \lim_{r\to 0}\frac{\frac{x^r-1}{r} - \log x}{r} = \frac{\log^2 x}{2}.$$

Still, by L'Hospital's rule, one can verify that

$$\lim_{\gamma\to 0}\frac{x^\gamma\gamma\log x-x^\gamma+1}{\gamma^2}=\frac{\log^2 x}{2}.$$

Thus, $\frac{\partial K(x,\gamma)}{\partial\gamma}(x,\gamma)$ is continuous for γ.

Notice that when $\gamma\neq 0$,

$$\frac{\partial K(x,\gamma)}{\partial\gamma}(x,\gamma)=-\frac{K(x,\gamma)-\log x}{\gamma}+\frac{(x^\gamma-1)\log x}{\gamma}.$$

Thus, by **Part a** and **Part b**, it is obvious to see

$$\sup_{x\in(0,1],|\gamma|<\delta_0}\left|\frac{\partial K}{\partial\gamma}(x,\gamma)\right|x^\delta<C.$$

□

Example: n-mBm when $H(t)\in\mathbb{N}$.

Still, let

$$\widetilde{X}_H(t)=\frac{1}{2\pi}\int_{\mathbb{R}}\frac{e^{i\omega\lambda_t}}{|\omega|^{H(t)+1/2}}dW(\omega),$$

where $\lambda_t\in\Lambda_k$. We want to study the tangent field at location z where $H(z)\in\mathbb{N}\cap(0,k+1)$. Assume z is bounded away from those reference points of λ_t.

Take a bounded neighborhood V around z such that for $s\in V$, $H(s)\in(H(z)-1/2,H(z)+1/2)$ and it doesn't contain any reference points of λ_t. Assume $H\in C_{2H(z)+3}(V)$. For any two points $s,t\in\overline{V}-z$, we have

$$\begin{aligned}
&\mathrm{Cov}(\widetilde{X}_H(z+s),\widetilde{X}_H(z+t))\\
=\;&\frac{1}{4\pi^2}\int_{\mathbb{R}}\frac{e^{i\omega(\lambda_{z+s}-\lambda_{z+t})}}{|w|^{H(z+s)+H(z+t)+1}}d\omega\\
=\;&\frac{1}{4\pi^2}\sum_{i,j}c_i(z+s)c_j(z+t)\\
&*\int_{\mathbb{R}}\frac{e^{i\omega(x_i(z+s)-x_j(z+t))}-I_B(\omega)P_k(\omega(x_i(z+s)-x_j(z+t)))}{|w|^{H(z+s)+H(z+t)+1}}d\omega\\
:=\;&\frac{1}{4\pi^2}\sum_{i,j}C_{ij}(s,t).
\end{aligned}$$

where P_k is the first $k+1$ terms of the Taylor expansion of $\cos(x)$; $I_B(w)$ is the indicator

function with $B = \{x : |x| \leq 1\}$; $c_0(s) = 1$; $c_i(s)$ is a polynomial function of s when $i \geq 1$; $x_0(s) = s$; $x_i(s) = a_i$ when $i \geq 1$.

When $i \neq 0$ or $j \neq 0$, it is easy to see $C_{ij}(s,t) \in C_{2H(z)+3}((\overline{V}-z)^{\otimes 2})$, and thus by Lemma 2.8.33, we don't need to consider those terms in studying the tangent field.

For $C_{0,0}(s,t)$, if $s - t = 0$, we simply have $C_{0,0}$ to be a constant not related to s, t. When $s \neq t$, we have

$$\begin{aligned}
C_{0,0}(s,t) &= 2\int_{\mathbb{R}_+} \frac{e^{i\omega(s-t)} - I_B(\omega)P_k(w(s-t))}{\omega^{H(z+s)+H(z+t)+1}} d\omega \\
&= 2|s-t|^{H(z+s)+H(z+t)} \int_{\mathbb{R}_+} \frac{e^{i\omega} - I_B(\omega/|s-t|)P_k(w)}{\omega^{H(z+s)+H(z+t)+1}} d\omega \\
&= 2|s-t|^{H(z+s)+H(z+t)} \int_{\mathbb{R}_+} \frac{e^{i\omega} - I_B(\omega)P_k(w)}{\omega^{H(z+s)+H(z+t)+1}} d\omega \\
&\quad +2|s-t|^{H(z+s)+H(z+t)} \int_{|s-t|}^1 \frac{\sum_{n=0,\ldots,k,n\neq H(z)} (-1)^n \frac{w^{2n}}{(2n)!}}{\omega^{H(z+s)+H(z+t)+1}} d\omega \\
&\quad +2|s-t|^{H(z+s)+H(z+t)} \int_{|s-t|}^1 \frac{(-1)^{H(z)} \frac{w^{2H(z)}}{(2H(z))!}}{\omega^{H(z+s)+H(z+t)+1}} d\omega \\
&:= C_{0,0}^{(1)}(s,t) + C_{0,0}^{(2)}(s,t) + C_{0,0}^{(3)}(s,t).
\end{aligned}$$

Notice that $C_{0,0}^{(1)}(s,t) = D(s,t)|s-t|^{H(z+s)+H(z+t)}$, which satisfies conditions of Lemma 2.8.34. Also, for any terms in $C_{0,0}^{(2)}(s,t)$, we have

$$\int_{|s-t|}^1 \frac{w^{2n}}{\omega^{H(z+s)+H(z+t)+1}} d\omega = \int_{|s-t|}^1 w^{2n-(H(z+s)+H(z+t)+1)} d\omega = \frac{1 - |s-t|^{2n-(H(z+s)+H(z+t))}}{2n - (H(z+s)+H(z+t))}.$$

Thus, each term of $C_{0,0}^{(2)}(s,t)$ satisfies conditions of Lemma 2.8.34 and Lemma 2.8.33.

Finally for $C_{0,0}^{(3)}(s,t)$, we have

$$\begin{aligned}
C_{0,0}^{(3)}(s,t) &= \frac{2(-1)^{H(z)}}{(2H(z))!} |s-t|^{H(z+s)+H(z+t)} \int_{|s-t|}^1 \frac{w^{2H(z)}}{\omega^{H(z+s)+H(z+t)+1}} d\omega \\
&= \frac{2(-1)^{H(z)+1}}{(2H(z))!} K(|s-t|, H(s,t)) |s-t|^{H(z+s)+H(z+t)},
\end{aligned}$$

where $K(x,\gamma)$ is defined in Lemma 2.8.36 as an extension to the $\log$ function and $H(s,t) = 2H(z) - H(z+s) - H(z+t)$. By similar arguments in (2.98)-(2.100), we can verify $|s-t|^{H(z+s)+H(z+t)-2H(z)}$ satisfies [LL3] in Proposition 2.8.32 and $|rs-rt|^{H(z+rs)+H(z+rt)-2H(z)} \to 1$ as $r \to 0$, and therefore, we only need to verify that $K(|s-t|, H(s,t))$ satisfies [LL1]-[LL3] in Proposition 2.8.32.

For [LL1], set $\widetilde{L}'(r) = \log r$, and we will have for any $s, t \in \overline{V} - z, s \neq t$,

$$\begin{aligned} & \lim_{r\to 0} |K(r|s-t|, H(rs, rt)) - \widetilde{L}'(r) - \log|s-t| \\ = & \lim_{r\to 0} |K(r|s-t|, H(rs, rt)) - \log(r|s-t|)| \\ \leq & \lim_{r\to 0} |K(r|s-t|, Cr) - \log(r|s-t|)| \\ \leq & \lim_{r\to 0} C' r^{1-\delta} |s-t|^{-\delta} = 0, \end{aligned}$$

where $C', C > 0$ are some constants and the last line use **part b** of Lemma 2.8.36.

The condition [LL2] is a direct result of **part b** of Lemma 2.8.36.

For [LL3], if $|t-a| < |s-a|/2$, we have

$$\begin{aligned} & |t-a|^{\delta} |K(r|s-a|, H(rs, ra)) - K(r|t-a|, H(rt, ra))| \\ \leq & |t-a|^{\delta/2} |s-a|^{\delta/2} |K(r|s-a|, H(rs, ra)) - K(r|t-a|, H(rt, ra))| \\ \leq & C|t-a|^{\delta/4} |s-a|^{\delta/4} \\ \leq & C|t-a|^{\delta/4} \leq C|s-t|^{\delta/4}, \end{aligned}$$

where the second inequality uses **part a** of Lemma 2.8.36.

If $|t-a| \geq |s-a|/2$, we have

$$\begin{aligned} & |t-a|^{\delta} |K(r|s-a|, H(rs, ra)) - K(r|t-a|, H(rt, ra))| \\ \leq & |t-a|^{\delta} (|K(r|s-a|, H(rs, ra)) - K(r|t-a|, H(rs, ra))| \\ & + |K(r|t-a|, H(rs, ra)) - K(r|t-a|, H(rt, ra))|), \end{aligned}$$

and by **part d** of Lemma 2.8.36,

$$\begin{aligned} & |t-a|^{\delta} |K(r|t-a|, H(rs, ra)) - K(r|t-a|, H(rt, ra))| \\ \leq & C|H(rs, ra) - H(rt, ra)| \leq C|s-t|, \end{aligned}$$

If $H(rs,ra)\neq 0$, we have

$$\begin{aligned}
& |t-a|^{\delta}(|K(r|s-a|,H(rs,ra))-K(r|t-a|,H(rs,ra))|\,|\log|s-t||^{1+\gamma} \\
= \;& |t-a|^{\delta}\left|\frac{(r|s-a|)^{H(rs,ra)}-(r|t-a|)^{H(rs,ra)}}{H(rs,ra)}\right|\,|\log(|(s-a)/(t-a)|-1) \\
& +\log|t-a||^{1+\gamma} \\
\leq \;& C|t-a|^{\delta-H(rs,ra)}\left(\left|\frac{|(s-a)/(t-a)|^{H(rs,ra)}-1}{H(rs,ra)}\right|\right) \\
& *(|\log(|(s-a)/(t-a)|-1)|^{1+\gamma}+|\log|t-a||^{1+\gamma}) \\
\leq \;& C'\sup_{u\in(0,1],\alpha\in(-\delta_0,\delta_0),\alpha\neq 0}\left|\frac{(1+u)^{\alpha}-1}{\alpha}\log^{1+\gamma}(u)\right|+C''<\infty,
\end{aligned}$$

where $\delta_0<1$ and the last inequality is because for $f_\alpha(u):=\frac{(1+u)^\alpha-1}{\alpha}$, we have $f_\alpha(0)=0$ and $f'_\alpha(u)=(1+u)^{\alpha-1}\geq 2^{\alpha-1}$ for $u\in(1,2]$.

If $H(rs,ra)=0$, then [LL3] is a direct result of Lemma 2.8.35.

Therefore, [LL3] is satisfied. And thus, the lth tangent field for $l\geq H(z)$ has generalized covariance function $K(h)=|h|^{2H(z)}\log|h|$.

Example 2.8.3. *n-mBm.*

Let's consider the extension of $\widetilde{X}_H(t)$ to multiple dimensions, that is

$$\widetilde{X}_H(t)=\int_{\mathbb{R}^d}\frac{e^{i\langle\omega,\lambda_t\rangle}}{|\omega|^{H(t)+d/2}}g_t\left(\frac{\omega}{|\omega|}\right)dW(\omega),$$

where $\mathbb{E}[|dW(\omega)|^2]=r^{d-1}drd\theta$; for some $k\in\mathbb{N}$, $H\in(0,k+1)$, $\lambda_t\in\Lambda_k$ such that $\widetilde{X}_H(t)$ is well-defined at location t. Assume $H\in C_{2[H(t)]+3}(V)$, where V is any bounded neighborhood around t; $g_s(\theta)\in C_{2[H(t)]+3}(V)$ with respect to s and for any given s, $g_s(\theta)$ is bounded. Now, by the same argument in the previous two examples, one can easily show that any order tangent field at location t exists. If the order $0\leq l\leq[H(t)]-1$, the generalized covariance function is $K(h)=|h|^{2l+2}$; if the order $l\geq[H(t)]$, the generalized covariance function is

$$K(h) \;=\; \begin{cases} \int_{S^{d-1}}|\langle h,\theta\rangle|^{2H(t)}g_z(\theta)d\theta & H(t)\notin\mathbb{N}, \\ \int_{S^{d-1}}|\langle h,\theta\rangle|^{2H(t)}\log(|\langle h,\theta\rangle|)g_z(\theta)d\theta & H(t)\in\mathbb{N}. \end{cases}$$

Example 2.8.4. *Stein's model.*

As a generalization of the famous Matérn class, Stein (2005) proposed a model with the

following covariance function:

$$C(x,y) = \frac{c(x)c(y)}{|\Sigma(x,y)|^{1/2}} \mathscr{M}_{\nu(x,u)}(Q(x,y)^{1/2}),$$

where $c(x)$ is a smooth function for normalizing, $\Sigma(x,y) = (\Sigma(x) + \Sigma(y))/2$, $\nu(x,y) = (\nu(x) + \nu(y))/2$, $\Sigma(x)$ is a mapping from $\mathbb{R}^d$ to the space of $d \times d$ positive definite matrix, $\mathscr{M}_\nu(x) = x^\nu \mathscr{K}_\nu(x)$, with $\mathscr{K}$ being a modified Bessel function of order ν and $Q(x,y) = (x-y)^T \Sigma(x,y)^{-1}(x-y)$.

By (9.6.2) and (9.6.10) in Abramowitz and Stegun (1964), one can have the following decomposition

$$\mathscr{M}_\nu(x) = \frac{\pi}{2^{\nu+1}\sin(\nu\pi)} \left(\sum_{l=0}^{\infty} \frac{x^{2l}}{4l!\Gamma(-\nu+l+1)} - \sum_{l=0}^{\infty} \frac{x^{2l+2\nu}}{4l!\Gamma(\nu+l+1)} \right).$$

Assume in a bounded neighborhood V of z, $\Sigma(x)$ and $\nu(x)$ are smooth function, eigenvalues of $\Sigma(x)$ bounded from 0 and ∞ and $\nu(x) \in (k, k+1)$ for some $k \in \mathbb{N}$.

Denote

$$\begin{aligned} C_1(x,y) \quad = \quad & \frac{c(x)c(y)\pi}{2^{\nu+1}\sin(\nu(x,y)\pi)|\Sigma(x,y)|^{1/2}} \left(\sum_{l=0}^{\infty} \frac{Q(x,y)^l}{4l!\Gamma(-\nu(x,y)+l+1)} \right. \\ & \left. - \sum_{l=1}^{\infty} \frac{Q(x,y)^{l+\nu(x,y)}}{4l!\Gamma(\nu(x,y)+l+1)} \right) \end{aligned}$$

and

$$C_2(x,y) = \frac{c(x)c(y)\pi}{2^{\nu+1}\sin(\nu(x,y)\pi)|\Sigma(x,y)|^{1/2}} \frac{Q(x,y)^{\nu(x,y)}}{4l!\Gamma(\nu(x,y)+l+1)}.$$

By studying the property of Γ function, one can show $C_1(x,y)$ is a smooth function on $(\overline{V}-z)^{\otimes 2}$. And by Lemma 2.8.33, the reminder after Taylor expansion satisfies conditions [R1], [R2] and [R3].

With the same argument in page 82, 83 of Samorodnitsky and Taqqu (1994), one can show there exists a $\sigma_{x,y}(d\theta)$ such that

$$|\Sigma(x,y)^{-1/2}x|^{2\nu(x,y)} = \int_{S^{d-1}} |\langle \theta, x \rangle|^{2\nu(x,y)} \sigma_{x,y}(d\theta)$$

for any Σ strictly symmetric positive definite.

Thus, to verify $C_2(x,y)$ satisfies the conditions of Lemma 2.8.34, we only need to show

those conditions for $\sigma_{x,y}$. Notice that by Lemma 2.8.31,

$$\begin{aligned}\left|\int_{S^{d-1}} |\langle x,\theta\rangle|^{\nu}(\sigma_{x,a}(d\theta)-\sigma_{y,a}(d\theta))\right| &\leq \left||\Sigma(x,a)^{-1/2}x|^{\nu}-|\Sigma(y,a)^{-1/2}x|^{\nu}\right| \\ &\leq \left|(\Sigma(x,a)^{-1/2}-\Sigma(y,a)^{-1/2})x\right|^{\nu\wedge 1} \\ &\leq C|x-y|,\end{aligned}$$

which completes the proof.

CHAPTER 3

Hurst Function Estimation

In this chapter, we give a thorough study for estimating the Hurst function in a multifractional Brownian motion. We first prove a lower bound for convergence rates and then propose estimators with convergence rates matching this lower bound under appropriate parameter tunings, implying they are rate optimal. Data-driven selection of parameters is also given for practical implementations. Simulations and some extensions are provided at the end.

3.1 Introduction

Since the introduction by Mandelbrot and Van Ness (1968), fractional Brownian motion (fBm) has found many applications, in hydrology, financial mathematics, network analysis, to name a few. Specifically, a fBm with Hurst parameter $H \in (0,1)$ is a Gaussian process $B_H(t), t \geq 0$, with stationary increments and satisfying $\mathbb{E}[B_H(t)] = 0$ and $\mathbb{E}[B_H^2(t)] = t^{2H}, t \geq 0$. In spatial statistics, the latter expression is referred to as a power variogram (cf. Stein (1999)). One of the appealing features of fBm is that the Hurst index characterizes the nature of dependence or, equivalently, sample path smoothness of the process globally. The book Nourdin (2012) contains a detailed introduction of the properties of fBm.

However, in many circumstances, a more flexible model is desirable that allows sample path smoothness to vary with time or location while retains some of the other key features of fBm. The multifractional Brownian motion (mBm) is such an example. The mBm was independently introduced in Lévy-Véhel and Peltier (1995) using a moving average type construction and in Benassi et al. (1997) based on a harmonizable integral representation. Cohen (1999) proved that these two definitions are equivalent up to a multiplicative deterministic function. Stoev and Taqqu (2006) proposed a generalizations of these two definitions. In this chapter, we use the definition of mBm in a general dimension d introduced in Herbin (2006).

For convenience, let $|\cdot|$ denote both absolute value and the Euclidean norm in $\mathbb{R}^d$. Let

$$D(H) = \left(\int_{\mathbb{R}^d} \frac{1 - \cos x_1}{|\boldsymbol{x}|^{2H+d}} d\boldsymbol{x} \right)^{\frac{1}{2}}, \ H \in (0,1),$$

where x_1 is the first component of the vector $\boldsymbol{x}$.

Definition 3.1.1. *The multifractional Brownian motion $\{X(\boldsymbol{t}), \boldsymbol{t} \in (0,1)^d\}$ is a zero-mean Gaussian process with covariance function*

$$C(\boldsymbol{t}, \boldsymbol{s}) = \sigma^2 \, \mathscr{D}(H(\boldsymbol{t}), H(\boldsymbol{s})) \left(|\boldsymbol{t}|^{H(\boldsymbol{t})+H(\boldsymbol{s})} + |\boldsymbol{s}|^{H(\boldsymbol{t})+H(\boldsymbol{s})} - |\boldsymbol{s} - \boldsymbol{t}|^{H(\boldsymbol{t})+H(\boldsymbol{s})} \right), \quad (3.1)$$

where $\sigma^2 \in (0, \infty)$, $H(\boldsymbol{t})$ is a Hölder continuous function with range in $(0,1)$, and

$$\mathscr{D}(H(\boldsymbol{t}), H(\boldsymbol{s})) = \{2D(H(\boldsymbol{t}))D(H(\boldsymbol{s}))\}^{-1} D^2 \left(\frac{H(\boldsymbol{t}) + H(\boldsymbol{s})}{2} \right).$$

The function H in this definition will be referred to as the Hurst function of the mBm. Note that the covariance function $C(\boldsymbol{t}, \boldsymbol{s})$ in (3.1) is normalized such that $\mathrm{Var}(X(\boldsymbol{t})) = \sigma^2 |\boldsymbol{t}|^{2H(\boldsymbol{t})}$. The sample paths of a mBm are still Hölder continuous but the degree of smoothness varies from point to point according to H. Also, properties of self similarity and stationary increments only hold in a local sense.

In this chapter, our primary concern is the estimation of H. We will address both cases where σ is known and unknown. We view σ as a nuisance parameter if it is unknown, in which case we also consider its estimation. While we focus on the case where σ is constant over the entire domain of $\{X(\boldsymbol{t})\}$, we will discuss some extensions where σ is allowed to vary with $\boldsymbol{t}$.

The problem of estimating $H(\boldsymbol{t})$ is challenging, at least non-standard for time series or spatial statistics due to the nonstationarity of the process mentioned above. Fortunately, self similarity and stationary increments still hold locally in some sense for the mBm, which ensures that H can be identified with probability one if the entire path of X is observed. Of course, one never observes the whole path in applications. Instead, we will follow the convention for this problem and assume for the most part that the data are observed on a regular grid. However, a brief discussion will be provided to address how this restriction might be relaxed.

A number of papers in the literature address the estimation of H. All of them focus on the case $d = 1$ and the estimators are formulated by considering the relationship between H and moments of functions of generalized differences of gridded data. With no intention to provide a complete list, we mention Coeurjolly (2005), Bardet and Surgailis (2013)

and Bertrand et al. (2013). To the best of our understanding, Bardet and Surgailis (2013) contains the most comprehensive results to date that unify the approaches of Coeurjolly (2005) and Bertrand et al. (2013); furthermore, it discusses the extension to a class of processes that behave like the mBm in a local sense. More details on these will be given in Sections 3.2 and 3.3. None of these papers approach the inference of H in a principled manner so as to thoroughy address issues such as how higher-order smoothness of H should be accounted for in the inference problem and how to formulate rate optimal estimators.

In the context of spatial data, a general framework for nonstationarity termed local intrinsic stationarity is developed in Hsing et al. (2016). The mBm falls in that framework. However, they focus on the scenario that H is twice continuously differentiable and the estimator introduced is not tailored to the mBm and consequently leads to sub-optimal rates. More importantly, their estimator does not satisfactorily address the subtle but important computational issues of the problem.

The main contributions and organization of the chapter are summarized as follows. Our goal is to explore a range of issues concerning the inference of the Hurst function. First, we formulate the nonparametric estimation of H based on gridded data in a general dimension d, taking into account the degree of smoothness of H, for both cases of known σ^2 (Section 3.2) and unknown σ^2 (Section 3.4). The existing results focus on $d = 1$ and essentially do not consider the smoothness of H beyond Hölder continuity with index 2. Second, we provide thorough asymptotic theories (e.g., Theorem 3.3.4 and Theorem 3.4.3) for our estimators under different scenarios. In that vein, we also establish for $d = 1$ a lower bound for the minimax risk of estimating H by all possible estimators assuming a broad class of H. This is the first time such a lower bound is developed in the mBm context. With proper tuning parameters, the rate of our estimator matches the lower bound, which makes it rate optimal. We also address the issue of data-driven bandwidth selection (Section 3.5), which is important for the implementation of the procedures. Some extensions are given in Section 3.7. Section 3.7.1 considers a non-gridded data scenario under which some of the key results established for gridded data continue to hold. Section 3.7.2 relaxes the assumption of constant σ by replacing σ^2 in (3.1) with $\sigma(\boldsymbol{t})\sigma(\boldsymbol{s})$ for some smooth, non-constant function $\sigma(\cdot)$. A numerical study is conducted (Section 3.6) to illustrate the results and compare with existing approaches. For clarity of presentation and to keep the paper under page limit, all proofs and technical details are given in Section 3.8.

3.2 Basic formulation of the estimator

In this section, we consider the estimation of the Hurst function $H(\boldsymbol{t}), \boldsymbol{t} \in (0,1)^d$ in the covariance of the mBm in (3.1). We first assume that $\sigma^2 \in (0,\infty)$ is known. In Section 3.4, we will address the case where σ^2 is unknown and, in Section 3.7.2, some extensions to non-constant σ^2.

Assume that we observe $X(\boldsymbol{t})$ for all $\boldsymbol{t}$ belonging to the grid

$$\Omega_n = \{(i_1, i_2, \ldots, i_d)/n, \text{ with } i_s = (j-.5)/n \text{ for } s = 1, \ldots, d, j = 1, \ldots, n\}.$$

For convenience, the generic notion $\boldsymbol{t}_i = (t_{i1}, \ldots, t_{id})$ will be used to denote the grid points. Define the differencing operator in the direction $\boldsymbol{h}$: for a function w,

$$\Delta_{\boldsymbol{h}} w(\boldsymbol{t}) := w(\boldsymbol{t}) - w(\boldsymbol{t}+\boldsymbol{h}) \quad \text{and} \quad \Delta_{\boldsymbol{h}}^j w(\boldsymbol{t}) := \Delta_{\boldsymbol{h}} \Delta_{\boldsymbol{h}}^{j-1} w(\boldsymbol{t}),\ j \geq 1.$$

It follows that, for any $q \in \mathbb{N}$,

$$\Delta_{\boldsymbol{h}}^q w(\boldsymbol{t}) = \sum_{i=0}^{q} (-1)^i \binom{q}{i} w(\boldsymbol{t} + i\boldsymbol{h}).$$

For the rest of this section let q be fixed and define

$$g(H, \boldsymbol{u}, \boldsymbol{h}) := -\frac{1}{2} \sum_{i=0}^{q} \sum_{j=0}^{q} (-1)^{i+j} \binom{q}{i} \binom{q}{j} |\boldsymbol{u} + (i-j)\boldsymbol{h}|^{2H}. \tag{3.2}$$

The choice of q will be discussed later in condition [A3] and in Section 3.5.

Let us consider the properties of X around a fixed $\boldsymbol{t}$. It is well known (cf. Falconer (2002)) that, as $n \to \infty$,

$$U_n(\boldsymbol{h}) := n^{H(\boldsymbol{t})} \Delta_{\boldsymbol{h}/n} X(\boldsymbol{t}) \xrightarrow{d} \sigma B_{H(\boldsymbol{t})}(\boldsymbol{h}), \tag{3.3}$$

where $B_{H(\boldsymbol{t})}$ is fBm with index $H(\boldsymbol{t})$, and $\xrightarrow{d}$ stands for convergence in distribution for the process $U_n(\boldsymbol{h})$ in the space of continuous functions endowed with the uniform metric

on any compact set. Consequently,

$$n^{H(\boldsymbol{t})}\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t}) = n^{H(\boldsymbol{t})}\sum_{i=0}^{q}(-1)^i\binom{q}{i}(X(\boldsymbol{t}+i\boldsymbol{h}/n)-X(\boldsymbol{t}))$$
$$\xrightarrow{d}\sigma\sum_{i=0}^{q}(-1)^i\binom{q}{i}B_{H(\boldsymbol{t})}(i\boldsymbol{h}).$$

Recall that

$$\mathrm{Cov}\left(B_{H(\boldsymbol{t})}(\boldsymbol{s}_1), B_{H(\boldsymbol{t})}(\boldsymbol{s}_2)\right) = \frac{1}{2}\left(|\boldsymbol{s}_1|^{2H(\boldsymbol{t})} + |\boldsymbol{s}_2|^{2H(\boldsymbol{t})} - |\boldsymbol{s}_1-\boldsymbol{s}_2|^{2H(\boldsymbol{t})}\right).$$

Thus, for any direction $\boldsymbol{h}$,

$$\mathbb{E}\left(\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t})\right)^2 \sim n^{-2H(\boldsymbol{t})}\sigma^2 g(H(\boldsymbol{t}),\boldsymbol{0},\boldsymbol{h}). \tag{3.4}$$

Note that the right hand side is a one-to-one function in $H(\boldsymbol{t})$. Thus, a plausible approach might be that, for large n, if we could estimate $\mathbb{E}\left(\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t})\right)^2$ well using a nonparametric approach based on differenced data $\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t}_i)$ for $\boldsymbol{t}_i$ in a small neighborhood of $\boldsymbol{t}$, then in principle we could also estimate $H(\boldsymbol{t})$ by inverting g. The choice of the direction $\boldsymbol{h}$ is relevant in two ways in multiple dimensions. Most importantly, since we consider gridded data, we must make sure that the $\boldsymbol{h}$ picked will lead to the full utilization of the data as differences are formed. Given that this is fulfilled, the choice of $\boldsymbol{h}$ may affect the asymptotics in a minor way but not the rate of convergence. An alternative approach is to consider a more general notion of differencing as in Coeurjolly (2005) and Bardet and Surgailis (2013) but that is unlikely to improve the rate either. In our setting, we conjecture that the optimal choice of $\boldsymbol{h}$ is a unit vector that parallels any of the d axes. We shall fix $\boldsymbol{h}$ to be such a vector in the remainder of this chapter.

Approaches similar in spirit to what was proposed above have been considered in the literature. For instance, in the case $d = 1$, Coeurjolly (2005) considers a kernel approach to estimate $\mathbb{E}(\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t}))^2$ by averaging those $(\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t}_i))^2$ for which $\boldsymbol{t}_i$ in a small neighborhood of $\boldsymbol{t}$. Unfortunately, the paper contains some technical issues which were later corrected by Bardet and Surgailis (2013). On the other hand, Hsing et al. (2016) considers a larger class of model for a general d and adopts local linear estimation. As explained in Section 1, none of the existing works satisfactorily address the wide range of statistical issues explored in this chapter.

A major problem with the approach motivated by (3.4) described earlier is that since the quantity in (3.4) is small for large n, a smoothing approach such as local polynomial

regression that takes into account of higher-order smoothness could yield negative values. When that happens there is no sensible way to define an estimate for $H(\boldsymbol{t})$. Note that the issue is non-existent if ones uses the Nadaraya-Watson estimator with a nonnegative kernel, but the approach would not account for higher order smoothness and the usual associated boundary issues would be exaggerated in higher dimensions. A natural remedy is to consider instead the estimation of $\mathbb{E}\log\left(\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t})\right)^2 = 2\mathbb{E}\log\left|\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t})\right|$. For $H \in (0,1)$ define

$$G(H;n,\boldsymbol{h}) := -2H\log n + \log\sigma^2 + \log g(H,\boldsymbol{0},\boldsymbol{h}) + \mathbb{E}\log\chi^2_1, \tag{3.5}$$

with χ^2_1 denoting a χ^2 random variable with one degree of freedom. Note that $g(H,\boldsymbol{0},\boldsymbol{h}) > 0$ for any Hurst index $H \in (0,1)$ and direction $\boldsymbol{h}$ by Lemma 3.8.10. We can also define $G(0;n,\boldsymbol{h})$ as $\lim_{H\downarrow 0} G(H;n,\boldsymbol{h})$. It is easy to see (cf. Lemma 3.8.5) that

$$2\mathbb{E}\log\left|\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t})\right| \approx G(H(\boldsymbol{t});n,\boldsymbol{h}). \tag{3.6}$$

Thus, our basic strategy is to estimate $\mathbb{E}\log\left|\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t})\right|$ and hence the quantity $G(H(\boldsymbol{t});n,\boldsymbol{h})$ nonparametrically based on $\log\left|\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t}_i)\right|$ for $\boldsymbol{t}_i$ in a small neighborhood of $\boldsymbol{t}$.

The nonparametric approach of our choice will be local polynomial regression (cf. Fan and Gijbels, 1996). The advantages of the approach have been extensively documented in the literature. For vectors $\boldsymbol{x} = (x_1,\ldots,x_d)^T \in \mathbb{R}^d$ and $\boldsymbol{i} = (i_1,\ldots,i_d)^T \in \{0,1,2,\ldots\}^d$, let

$$\boldsymbol{x}^{\boldsymbol{i}} = \prod_{l=1}^d x_l^{i_l} \quad \text{and} \quad \boldsymbol{i}! = \prod_{l=1}^d i_l!. \tag{3.7}$$

Fix $p \in [1,\infty)$ and let $\lceil p\rceil$ be the smallest integer no smaller than p. Sort the finite set $\{(j_1,\ldots,j_d)^T : j_l \in \{0,1,2,\ldots\}, \sum_{l=1}^d j_l \le \lceil p\rceil - 1\}$ in any manner and denote the sorted set as $\{\boldsymbol{j}_m, m = 1,\ldots,S\}$. However, we set $\boldsymbol{j}_1 = \boldsymbol{0}$ for convenience. Then, solve

$$\begin{aligned}&(\widehat{\beta}_{\boldsymbol{j}_1},\ldots,\widehat{\beta}_{\boldsymbol{j}_S})\\ &= \underset{\beta_{\boldsymbol{j}_1},\ldots,\beta_{\boldsymbol{j}_S}}{\operatorname{argmin}} \sum_i K\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)\left\{2\log\left|\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t}_i)\right| - \sum_{m=1}^S \beta_{\boldsymbol{j}_m}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)^{\boldsymbol{j}_m}\right\}^2,\end{aligned}$$

where K is a kernel function and $b > 0$ is bandwidth. The local polynomial regression

estimator of $G(H(\boldsymbol{t}); n, \boldsymbol{h})$ is $G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) := \widehat{\beta}_{\boldsymbol{j}_1}$. Define

$$\mathbf{A}(\boldsymbol{x}) = (\boldsymbol{x}^{\boldsymbol{j}_1}, \ldots, \boldsymbol{x}^{\boldsymbol{j}_S})^T \in \mathbb{R}^S,$$

and let $s_{\boldsymbol{t},p,b}(\boldsymbol{s})$ be the first element of the vector

$$K\left(\frac{\boldsymbol{s}-\boldsymbol{t}}{b}\right)\left(\sum_i K\left(\frac{\boldsymbol{t}_i-\boldsymbol{t}}{b}\right)\mathbf{A}\left(\frac{\boldsymbol{t}_i-\boldsymbol{t}}{b}\right)\mathbf{A}\left(\frac{\boldsymbol{t}_i-\boldsymbol{t}}{b}\right)^T\right)^{-1}\mathbf{A}\left(\frac{\boldsymbol{s}-\boldsymbol{t}}{b}\right).$$

It follows that

$$G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) = \sum_i 2 s_{\boldsymbol{t},p,b}(\boldsymbol{t}_i) \log\left|\Delta^q_{\boldsymbol{h}/n} X(\boldsymbol{t}_i)\right|. \tag{3.8}$$

Finally, the estimator of $H(\boldsymbol{t})$ that we will focus on is

$$\widehat{H}(\boldsymbol{t}) = G^{-1}\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})); n, \boldsymbol{h}\right). \tag{3.9}$$

When n is sufficiently large, the upper bound for $G(H; n, \boldsymbol{h})$ is $G(0; n, \boldsymbol{h}) < \infty$ and the lower bound is $-\infty$. As it is possible to have

$$G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) > G(0; n, \boldsymbol{h}) = \log\sigma^2 + \log g(0, \mathbf{0}, \boldsymbol{h}) + \mathbb{E}\log\chi^2_1,$$

we define $G^{-1}(x; n, \boldsymbol{h}) = 0$ for all the $x \geq G(0; n, \boldsymbol{h})$. The asymptotic properties of $\widehat{H}(\boldsymbol{t})$ will be considered in Section 3.3.

3.3 Asymptotic properties

This section contains two major asymptotic results as $n \to \infty$. Since we consider gridded data on a fixed bounded set, these results belong to the realm of the so-called fixed-domain or infill asymptotics (cf. Stein (1999)). Our first result is a uniform lower bound for the risk of estimating the Hurst function H that belongs to a class of smooth functions. The second result addresses the properties of the estimator $\widehat{H}$ defined in (3.9).

We begin by presenting a minimax bound for the risk of estimating H for the case $d = 1$. For functions f on $(0, 1)$ define

$$\|f\|_s = \left[\int_0^1 |f(t)|^s\, dt\right]^{1/s},\ s \in [1, \infty), \quad \text{and} \quad \|f\|_\infty = \sup_{t\in(0,1)} |f(t)|.$$

Let $\lfloor x \rfloor$ be the largest integer no larger than x. For open set $B \in \mathbb{R}^d$ and constants $p \geq 0$ and $M \in (0, \infty)$, define $\mathscr{H}_p(B, M)$ as the space of $\lfloor p \rfloor$-times differentiable functions $f : B \mapsto \mathbb{R}$ such that $f^{(\lfloor p \rfloor)}$ is Hölder continuous on B with $|f^{(\lfloor p \rfloor)}(\boldsymbol{x}) - f^{(\lfloor p \rfloor)}(\boldsymbol{y})| \leq M|\boldsymbol{x} - \boldsymbol{y}|^{p - \lfloor p \rfloor}$ for all $\boldsymbol{x}, \boldsymbol{y} \in B$.

Theorem 3.3.1. *Consider all estimator $\widetilde{H}_n$ of H based on data $\{X((i - 1/2)/n), i = 1, \ldots, n\}$. Then for any $\gamma \in (0, 1), s \in [1, \infty), p > 1$ and $M \in (0, \infty)$, there exists a $\delta \in (0, \infty)$ that only depends on s, p, M, γ such that*

$$\liminf_{n \to \infty} \inf_{\widetilde{H}_n} \sup_{H \in \mathscr{H}_p((0,1), M)} \mathbb{P}_H \left(\left\| \widetilde{H}_n - H \right\|_s > \delta \left(n \log^2 n \right)^{-\frac{p}{2p+1}} \right) > \gamma. \tag{3.10}$$

where $\mathbb{P}_H$ denotes the probability measure under H.

To the best of our knowledge, (3.10) is the first bound of the kind for the inference H. The proof borrows a familiar strategy from the development of minimax bounds in density and regression function estimation (cf. Chapter 2 of Tsybakov (2009)). The core of the proof is to compute a tight bound for the Kullback-Leibler (KL) divergence between two mBms with Hurst functions that are close. It is interesting to note that the minimax bound established in Theorem 3.3.1 is slightly faster, due to presence of the $\log n$ term, than the corresponding bounds in classical problems such as density and regression function estimation. While the lower bound is only established for $d = 1$, we conjecture the corresponding lower bound for $d = 2$ has the rate $\left(n^2 \log^2 n\right)^{-\frac{p}{2p+2}}$. Unfortunately, we have not been able to establish it so far. Below we will see that the lower bound for $d = 1$ can be attained by the estimator $\widehat{H}$ defined in (3.9) in Section 3.2.

We next proceed to consider the asymptotic properties of the estimator $\widehat{H}(\boldsymbol{t})$. For clarity, we list below the assumptions that will be frequently referred to in this and future sections. Let $\mathscr{H}_p(B) := \cup_{M=1}^{\infty} \mathscr{H}_p(B, M)$.

[K] K is a non-negative kernel function with support $B_1(\mathbf{0})$ and has continuous second-order partial derivatives, where $B_r(\boldsymbol{t}_0) = \{\boldsymbol{t} : |\boldsymbol{t} - \boldsymbol{t}_0| < r\}$.

[A1] $H \in \mathscr{H}_p((0, 1)^d)$ and is bounded away from 0 and 1.

[A2] $b = b_n$ is a bandwidth parameter varying with n, such that $nb \to \infty$ and $b \log^k n \to 0$ for all $k > 0$ as $n \to \infty$.

[A3] $p \geq q \geq 1$, where q is the order of differencing in defining $G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})$.

We also collect some common notations here for easy reference. Define

$$\psi(\boldsymbol{t}) := 2q - 2H(\boldsymbol{t}),\ \bar{\psi} = \inf_{\boldsymbol{t}} \psi(\boldsymbol{t}) \ \text{ and } \ \rho_n(\boldsymbol{t}) = \log(n)/n + n^{-\psi(\boldsymbol{t})}.$$

Also, let

$$\Omega_\delta = (0,1)^d - B_\delta(\mathbf{0})$$

where $B_\delta(\mathbf{0})$ is the d-dimensional ball centered at $\mathbf{0}$ with radius $\delta > 0$. We will focus on δ that are close to 0 (see next paragraph).

To investigate the asymptotic properties of $\widehat{H}(t)$, we consider the decomposition

$$\begin{aligned}
&G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - G(H(\boldsymbol{t}); n, \boldsymbol{h}) \\
&= \left\{ G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - \mathbb{E}\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) \right) \right\} \\
&\quad + \left\{ \mathbb{E}\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) \right) - G(H(\boldsymbol{t}); n, \boldsymbol{h}) \right\}.
\end{aligned}$$

The first term on the right-hand side is a centered random variable, which destermines the variance of the estimator. The second term corresponds to the bias. We will address the asymptotic behavior of these terms in the following two results. In doing so, our approach is to establish uniform bounds with respect to $\boldsymbol{t}$, which requires us to focus on $\boldsymbol{t} \in \Omega_\delta$ for an arbitrarily small but fixed δ. This is due to the fact that, since $\mathrm{Var}(X(\boldsymbol{t})) = \sigma^2 |\boldsymbol{t}|^{H(\boldsymbol{t})}$, the information contained in $X(\boldsymbol{t})$ becomes increasingly scarce as $\boldsymbol{t}$ approaches zero, and, consequently, the asymptotic theory for $G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})$ and $\widehat{H}(\boldsymbol{t})$ with $\boldsymbol{t}$ close to $\mathbf{0}$ has to be dealt with differently. This is completely unrelated to the usual boundary-effect issues in nonparametric estimation.

Our first result considers the rates of bias and variance of $G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})$. For convenience of presentation, define

$$T_1(n, b, \boldsymbol{t}) = \log(n) \left(b^p + (nb)^{-2 \wedge p} \right) + \rho_n(\boldsymbol{t}) \tag{3.11}$$

and

$$T_2(n, b, \boldsymbol{t}) = \begin{cases} (nb)^{-d}, & \text{if } 2\psi(\boldsymbol{t}) > d, \\ (nb)^{-d} \log(nb), & \text{if } 2\psi(\boldsymbol{t}) = d, \\ (nb)^{-2\psi(\boldsymbol{t})}, & \text{if } 2\psi(\boldsymbol{t}) < d. \end{cases} \tag{3.12}$$

In the remaining part of the chapter, the statement $f(n, b, \boldsymbol{t}) = O(g(n, b, \boldsymbol{t}))$ uniformly for

$\boldsymbol{t} \in \Omega_\delta$ means $\sup_{\boldsymbol{t}\in\Omega_\delta} \left|\frac{f(n,b,\boldsymbol{t})}{g(n,b,\boldsymbol{t})}\right| \leq C_\delta$ for some finite constant C_δ.

Theorem 3.3.2. *Suppose that the conditions [K], [A1]-[A3] hold. Then for any $\delta \in (0,1)$, we have uniformly for any $\boldsymbol{t} \in \Omega_\delta$,*

$$\mathbb{E}\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})\right) - G(H(\boldsymbol{t}); n, \boldsymbol{h}) = O\left(T_1(n, b, \boldsymbol{t})\right), \tag{3.13}$$

and

$$\mathrm{Var}\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})\right) = O\left(T_2(n, b, \boldsymbol{t})\right).$$

Remarks.

(i) In the result for bias, (3.13), the term $\rho_n(\boldsymbol{t})$ stands out as one that is absent in a classical nonparametric regression context. In some situations, e.g., when the dimension d exceeds 3, $\rho_n(\boldsymbol{t})$ could play a major role in deciding the estimation rate (cf. Remark (i) following Theorem 3.3.4 below).

(ii) Consider the case where $p \in \mathbb{N}$ and the term $\log(n)b^p$ dominates in $T_1(n, b, \boldsymbol{t})$, i.e.,

$$\frac{1}{n^{2\wedge p} b^{p+2\wedge p}} \to 0 \quad \text{and} \quad \frac{\rho_n(\boldsymbol{t})}{\log(n)b^p} \to 0. \tag{3.14}$$

The proof of Theorem 3.3.2 shows that if

$$\begin{aligned} R(\boldsymbol{t}) := \mathbf{e}_1^T \left(\int_{B_1(\mathbf{0})} K(\mathbf{z})\mathbf{A}(\mathbf{z})\mathbf{A}^T(\mathbf{z})d\mathbf{z}\right)^{-1} \times \\ \int_{B_1(\mathbf{0})} K(\mathbf{z})\mathbf{A}(\mathbf{z}) \left(\sum_{|\boldsymbol{\alpha}|=p} \frac{D^{\boldsymbol{\alpha}} H(\boldsymbol{t})}{\boldsymbol{\alpha}!}\mathbf{z}^{\boldsymbol{\alpha}}\right) d\mathbf{z} \end{aligned} \tag{3.15}$$

is well defined in $(0, \infty)$, where $\mathbf{e}_1$ stands for $(1, 0, \ldots, 0)^T$, then

$$\mathrm{Bias}\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})\right) \sim -2\log(n)b^p R(\boldsymbol{t}). \tag{3.16}$$

This bias expression is useful for deriving the asymptotic distribution of $G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})$ and $\widehat{H}(\boldsymbol{t})$ (cf. Corollary 3.3.5).

(iii) The variance in Theorem 3.3.2 was derived by analyzing the dependence of the process $n^{H(\boldsymbol{t})}\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t})$ in $\boldsymbol{t}$. The cases $2\psi(\boldsymbol{t}) > d$ and $2\psi(\boldsymbol{t}) < d$ can be referred to as the short and long memory cases while $2\psi(\boldsymbol{t}) = d$ is the borderline of the two.

Limit theorems in those cases are related to classical limit theorems in, for instance, Breuer and Major (1983) and Taqqu (1975). This remark also applies to Theorem 3.3.3 below.

Next we turn attention to the asymptotic distribution of $G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})$. Let the function $F(x) := \log(x^2) - \mathbb{E}\log\chi_1^2$ have the Hermite polynomial decomposition (cf. Taqqu (1975))

$$F(x) = \sum_{l=2}^{\infty} c_l H_l(x), \tag{3.17}$$

where H_l denotes the Hermite polynomial of order l. Define

$$\sigma_H^2 = \sum_{l=2}^{\infty} c_l^2 l! \sum_{\boldsymbol{j}\in\mathbb{Z}^d} \left(\frac{g(H, \boldsymbol{j}, \boldsymbol{h})}{g(H, \boldsymbol{0}, \boldsymbol{h})}\right)^l. \tag{3.18}$$

Theorem 3.3.3. *Assume that the conditions [K], [A1]-[A3] hold. Then, for all $\boldsymbol{t} \in (0,1)^d$,*

$$a_{n,b,\boldsymbol{t}} \left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - \mathbb{E}\, G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) \right) \xrightarrow{d} Z(\boldsymbol{t}),$$

where the normalizing constant $a_{n,b,\boldsymbol{t}}$ and the limit $Z(\boldsymbol{t})$ depend on $\psi(\boldsymbol{t})$, as follows.

(i) If $2\psi(\boldsymbol{t}) > d$, then $a_{n,b,\boldsymbol{t}} = (nb)^{d/2}$ and $Z(\boldsymbol{t}) \sim N\left(0, \xi^2(\boldsymbol{t})\right)$ where

$$\xi^2(\boldsymbol{t}) = \sigma_{H(\boldsymbol{t})}^2 \int_{B_1(\boldsymbol{0})} \omega^2(\mathbf{z}) d\mathbf{z}, \tag{3.19}$$

with

$$\omega(\mathbf{z}) = \mathbf{e}_1^T \left(\int_{B_1(\boldsymbol{0})} \mathbf{A}(\boldsymbol{\nu}) \mathbf{A}(\boldsymbol{\nu})^T K(\boldsymbol{\nu}) d\boldsymbol{\nu} \right)^{-1} \mathbf{A}(\mathbf{z}) K(\mathbf{z});$$

(ii) if $2\psi(\boldsymbol{t}) = d$, then $a_{n,b,\boldsymbol{t}} = (nb)^{d/2} \log^{-1/2}(nb)$ and $Z(\boldsymbol{t}) \sim N(\boldsymbol{0}, \xi^2(\boldsymbol{t}))$, where

$$\xi^2(\boldsymbol{t}) = \frac{V_d\, \omega^2(\boldsymbol{0})}{g^2(H(\boldsymbol{t}), \boldsymbol{0}, \boldsymbol{h}))} \int_{B_1(\boldsymbol{0})} \frac{c_{\boldsymbol{h}}^2(\boldsymbol{t}, \boldsymbol{x})}{|\boldsymbol{x}|^{d-1}} d\boldsymbol{x},$$

V_d is the volume of a d-dimensional unit ball and $c_{\boldsymbol{h}}(\boldsymbol{t}, \boldsymbol{u})$ is the coefficient of $|\boldsymbol{u}|^{-\psi(\boldsymbol{t})}$ in

$$-\frac{\partial^{2q}}{\partial\phi^q \partial\eta^q} \frac{1}{2} |\boldsymbol{u} + (\phi - \eta)\boldsymbol{h}|^{H(\boldsymbol{t}+\phi\boldsymbol{h}) + H(\boldsymbol{t}+\eta\boldsymbol{h})}|_{\phi=0, \eta=0};$$

(iii) if $2\psi(\boldsymbol{t}) < d$*, then* $a_{n,b,\boldsymbol{t}} = (nb)^{\psi(\boldsymbol{t})}$ *and the characteristic function of* $Z(\boldsymbol{t})$ *is*

$$\phi(u) = \exp\left\{\frac{1}{2}\sum_{k=2}^{\infty}\frac{(2iu)^k}{k}S_k\right\},$$

where

$$S_k = \int_{(B_1(\mathbf{0}))^{\otimes k}} \frac{\prod_{i=1}^{k} c_{\boldsymbol{h}}(\boldsymbol{t}, \mathbf{z}_i - \mathbf{z}_{i+1})|\mathbf{z}_i - \mathbf{z}_{i+1}|^{-\psi(\boldsymbol{t})}\omega(\mathbf{z}_i)}{g^k(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h})} d\mathbf{z}_1 \ldots d\mathbf{z}_k,$$

in which $\mathbf{z}_{k+1}$ *denotes* $\mathbf{z}_1$ *for notational convenience.*

Remarks. Recall that we let $\boldsymbol{h}$ be any unit vector parallel to one of the axes. It is worth pointing out that, by Lemma 3.8.11, $c_{\boldsymbol{h}}(\boldsymbol{t}, \boldsymbol{u})$ is actually a polynomial of $H(\boldsymbol{t})$ and $\langle \boldsymbol{u}, \boldsymbol{h}\rangle/|\boldsymbol{u}|$. Thus, $c_{\boldsymbol{h}}(\boldsymbol{t}, \boldsymbol{u})$ depends on $\boldsymbol{u}$ only through its direction $\boldsymbol{u}/|\boldsymbol{u}|$. For instance, for $d = 1$

$$c_h(t, u) = (-1)^{q-1}\prod_{l=0}^{2q-1}(2H(t) - l).$$

As a consequence of this and symmetry in integration, taking into account that $g(H, \mathbf{0}, \boldsymbol{h})$ does not depend on the direction of $\boldsymbol{h}$, the limits in (i)-(iii) do not depend on the direction of $\boldsymbol{h}$ either.

The combination of Theorems 3.3.2 and 3.3.3 leads to the following result for $\widehat{H}(\boldsymbol{t})$.

Theorem 3.3.4. *Suppose that the conditions [K], [A1]-[A3] hold. Then, uniformly for* $\boldsymbol{t} \in \Omega_\delta$*, where* δ *is an arbitrary constant in* $(0, 1)$*, we have*

$$\begin{aligned}\mathbb{E}\left(\widehat{H}(\boldsymbol{t}) - H(\boldsymbol{t})\right) = &-(2\log n)^{-1}\mathbb{E}\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - G(H(\boldsymbol{t}); n, \boldsymbol{h})\right) \\ &+ O\left((T_1(n, b, \boldsymbol{t}) + T_2(n, b, \boldsymbol{t}))/\log^2 n\right)\end{aligned}$$

and

$$\mathbb{E}\left(\widehat{H}(\boldsymbol{t}) - H(\boldsymbol{t})\right)^2 = O\left(\left(T_1^2(n, b, \boldsymbol{t}) + T_2(n, b, \boldsymbol{t})\right)/\log^2 n\right),$$

where T_1, T_2 *are as defined in (3.11) and (3.12). Moreover, for any* $\boldsymbol{t} \in (0, 1)^d$*, we have* $2a_{n,b,\boldsymbol{t}}\log(n)\left(\widehat{H}(\boldsymbol{t}) - G^{-1}\left(\mathbb{E}G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})\right)\right)$ *has the same asymptotic distribution as* $a_{n,b,\boldsymbol{t}}\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - \mathbb{E}G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})\right)$*, where* $a_{n,b,\boldsymbol{t}}$ *and the corresponding limits are the same as given in Theorem 3.3.3.*

Remarks.

(i) To compute the rate of $\widehat{H}(\boldsymbol{t})$ in various situations using Theorems 3.3.2 and 3.3.4, we focus on the case where $2\psi(\boldsymbol{t}) > d$ and (3.14) holds. Letting $b \sim (n^d \log^2 n)^{-\frac{1}{2p+d}}$ then leads to the rate $(n^d \log^2 n)^{-\frac{p}{2p+d}}$. The following are some key scenarios for which this rate holds:

- $d = 1, p \geq q = 1$ and $\sup_t H(t) < 0.75$
- $d = 2, p \geq q = 1$ and $\sup_{\boldsymbol{t}} H(\boldsymbol{t}) < 0.5$
- $d \leq 2, p \geq q \geq 2$ and any H
- $d = 3, p \in [1.5, 3), q = 1$ and $\sup_{\boldsymbol{t}} H(\boldsymbol{t}) < 0.25$
- $d = 3, 3 > p \geq q = 2$ and any H

If $d > 3$, the term $\rho_n(\boldsymbol{t})$ dominates the bias when picking the bandwidth b above and the rate calculations above does not apply (i.e., (3.14) fails). In future results, we will return to these 5 scenarios for comparisons.

(ii) For $d = 1, p \geq q \geq 2$ and any H or $d = 1, p \geq q = 1$ and $\|H\|_\infty < 0.75$, the rate $(n \log^2 n)^{-\frac{p}{2p+1}}$ matches the rate in the lower bound of Theorem 3.3.1 and therefore $\widehat{H}(t)$ is minimax optimal.

The following corollary describes the asymptotic distributions under the scenarios in Remark (i) above.

Corollary 3.3.5. *Suppose that the conditions [K], [A1], and [A3] hold. For $\boldsymbol{t} \in (0,1)^d$, assume additionally that $2\psi(\boldsymbol{t}) > d$, $p \in \mathbb{N}$, (3.14) holds for $b \sim (n^d \log^2 n)^{-\frac{1}{2p+d}}$ and that the quantity $R(\boldsymbol{t})$ in (3.15) is well defined in $[0, \infty)$. Then*

$$\left(n^p \log^{-1} n\right)^{\frac{d}{2p+d}} \left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - G(H(\boldsymbol{t}); n, \boldsymbol{h})\right) \xrightarrow{d} N\left(R(\boldsymbol{t}), \xi^2(\boldsymbol{t})\right)$$

and

$$2(n^d \log^2 n)^{\frac{p}{2p+d}} \left(\widehat{H}(\boldsymbol{t}) - H(\boldsymbol{t})\right) \xrightarrow{d} N\left(R(\boldsymbol{t}), \xi^2(\boldsymbol{t})\right),$$

where $\xi^2(\boldsymbol{t})$ is as given in (3.19).

3.4 Backfitting estimation when σ^2 is unknown

We have so far assumed that σ^2 in (3.1) is known. In this section we will address the case where σ^2 is unknown and construct a slightly different estimator than $\widehat{H}(\boldsymbol{t})$.

The construction consists of three steps.

(i) First, consider a less efficient estimator, $\widehat{H}_1(\boldsymbol{t})$, that does not require the knowledge of σ^2 to crudely estimate $H(\boldsymbol{t})$.

(ii) Based on $\widehat{H}_1(\boldsymbol{t})$, estimate $\log \sigma^2$ by an estimator denoted by $\widehat{\log \sigma^2}$.

(iii) Finally, estimate $H(\boldsymbol{t})$ again by $\widehat{H}(\boldsymbol{t})$ by plugging in $\widehat{\log \sigma^2}$ for $\log \sigma^2$. The final estimator is denoted as $\widehat{H}_2(\boldsymbol{t})$.

The details are given below.

The estimator of $H(\boldsymbol{t})$ in step (i) is defined as

$$\widehat{H}_1(\boldsymbol{t}) = \frac{G(\widehat{H(\boldsymbol{t})}; n, 2\boldsymbol{h}) - G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})}{2 \log 2}, \tag{3.20}$$

where $G(\widehat{H(\boldsymbol{t})}; n, 2\boldsymbol{h})$ and $G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})$ use the same bandwidth denoted as b_1. The proof of the following result is similar to that for Theorem 3.3.2 and 3.3.3.

Theorem 3.4.1. *Assume that [K], [A1]-[A3] hold with b in [A2] replaced by b_1. Then, for any $\delta \in (0, 1)$, we have uniformly for $\boldsymbol{t} \in \Omega_\delta$*

$$\mathbb{E}\left(\widehat{H}_1(\boldsymbol{t}) - H(\boldsymbol{t})\right) = O\left(b_1^p + \rho_n(\boldsymbol{t})\right)$$

and

$$\mathrm{Var}\left(\widehat{H}_1(\boldsymbol{t})\right) = O\left(T_2(n, b_1, \boldsymbol{t})\right),$$

where T_2 is defined in (3.12). Additionally, if $2\psi(\boldsymbol{t}) \geq d$, $\widehat{H}_1(\boldsymbol{t}) - \mathbb{E}[H(\boldsymbol{t})]$ is asymptotically normal.

Remarks. Under the 5 scenarios listed in remark (i) following Theorem 3.3.4, Theorem 3.4.1 implies that $\widehat{H}_1(\boldsymbol{t})$ has asymptotic bias $O(b_1^p)$ and variance $O((nb_1)^{-d})$. Thus, taking $b_1 \sim n^{-\frac{d}{2p+d}}$ leads to the optimal rate $n^{-\frac{dp}{2p+d}}$. Note that this rate is slower than that of $\widehat{H}(\boldsymbol{t})$ for the case where σ^2 is known.

Next, we proceed to estimate σ^2. For some $m = 1, 2, \ldots$, obtain all of $\widehat{H}_1(\boldsymbol{t})$ for $\boldsymbol{t} \in \Omega_m = \{(i_1, i_2, \ldots, i_d)/m, \text{ with } i_s = (j - 1/2)/m \text{ for } s = 1, \ldots, d, j = 1, \ldots, m\}$. Note that, in this step, we are still using the estimate $\widehat{H}_1$ obtained in the previous step based on data observed on Ω_n. Thus, m could be viewed as another tuning parameter. While it is possible to present our asymptotic results below for any choice of m, to streamline

presentation we will fix $m \sim 1/b_1$ (which is the optimal choice in some sense) from this point on.

Define

$$\begin{aligned}\widetilde{G}(H; n, \boldsymbol{h}) &:= G(H; n, \boldsymbol{h}) - \log \sigma^2 \\ &= -2H \log n + \log g(H, \mathbf{0}, \boldsymbol{h}) + \mathbb{E} \log \chi_1^2.\end{aligned} \tag{3.21}$$

Intuitively we could estimate $\log \sigma^2$ by $G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - \widetilde{G}(\widehat{H}_1(\boldsymbol{t}); n, \boldsymbol{h})$. However, while rare, numerically it is possible that $\widehat{H}_1(\boldsymbol{t}) \notin (0, 1)$, in which case the computation of $\log g\left(\widehat{H}_1(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h}\right)$ would be problematic (e.g., $g(1, \mathbf{0}, \boldsymbol{h}) = 0$). By [A1] there exists a constant $\gamma > 0$ such that

$$\inf_{\boldsymbol{t} \in [0,1]^d} \{H(\boldsymbol{t}) \wedge (1 - H(\boldsymbol{t}))\} > \gamma. \tag{3.22}$$

Define the thresholded estimator

$$\widehat{H}_1^{\gamma}(\boldsymbol{t}) = \widehat{H}_1(\boldsymbol{t}) \cdot I\left(\widehat{H}_1(\boldsymbol{t}) \in [0, 1 - \gamma/2]\right) + (1 - \gamma/2) \cdot I\left(\widehat{H}_1(\boldsymbol{t}) > 1 - \gamma/2\right),$$

and estimate $\log \sigma^2$ by

$$\begin{aligned}\widehat{\log \sigma^2} :=& \frac{1}{\#(\Omega_m \cap \Omega_\delta)} \sum_{\boldsymbol{t} \in \Omega_m \cap \Omega_\delta} \Big(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) + 2\widehat{H}_1(\boldsymbol{t}) \log n \\ &- \log g\left(\widehat{H}_1^{\gamma}(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h}\right) - \mathbb{E} \log \chi_1^2 \Big).\end{aligned}$$

In the following results, let

$$T_1'(n, b_1) = \log n \left(b_1^p + (nb_1)^{-2 \wedge p}\right) + n^{-1} \log^2 n + n^{-\bar{\psi}} \log n, \tag{3.23}$$

$$T_2'(n, b_1) = \begin{cases} (nb_1)^{-d}, & \text{if } 2\bar{\psi} > d, \\ (nb_1)^{-d} \log(nb_1), & \text{if } 2\bar{\psi} = d, \\ (nb_1)^{-2\bar{\psi}}, & \text{if } 2\bar{\psi} < d. \end{cases} \tag{3.24}$$

Theorem 3.4.2. *Assume that the conditions [K], [A1]-[A3] hold with b in [A2] replaced*

by b_1. We have

$$\mathbb{E}\left[\left(\widehat{\log\sigma^2}-\log\sigma^2\right)^2\right]=O\left((T_1'(n,b_1))^2+T_2'(n,b_1)\right).$$

Furthermore, if $T_2'(n,b_1)=O\left(T_1'(n,b_1)\right)$, then we have

$$\mathbb{E}\left(\widehat{\log\sigma^2}-\log\sigma^2\right)=O\left(T_1'(n,b_1)\right).$$

Remarks. For the cases $2\bar{\psi}>d$, $d\leq 2$ and $p\geq q\geq 2$ or $d=3$ and $3>p\geq q=2$, it can be seen that the convergence rate of $\widehat{\log\sigma^2}$ based on Theorem 3.4.2 is $n^{-\frac{dp}{2p+d}}(\log n)^{\frac{d}{2p+d}}$ if $b_1\sim(n^d\log^2 n)^{-\frac{1}{2p+d}}$.

Finally, we proceed to the final step and estimate $H(\boldsymbol{t})$ by

$$\widehat{H}_2(\boldsymbol{t})=\widetilde{G}^{-1}\left(G(\widehat{H(\boldsymbol{t})};n,\boldsymbol{h})-\widehat{\log\sigma^2};n,\boldsymbol{h}\right),$$

where the bandwidth used in $G(\widehat{H(\boldsymbol{t})};n,\boldsymbol{h})$ in this step is denoted as b_2.

Theorem 3.4.3. *Assume that the conditions [K], [A1]-[A3] hold with b in [A2] replaced by b_1 and b_2. Then uniformly for $\boldsymbol{t}\in\Omega_\delta,\delta\in(0,1)$,*

$$\begin{aligned}&\mathbb{E}\left(\widehat{H}_2(\boldsymbol{t})-H(\boldsymbol{t})\right)^2\\&=O\left(((T_1'(n,b_1))^2+T_2'(n,b_1)+T_1^2(n,b_2,\boldsymbol{t})+T_2(n,b_2,\boldsymbol{t}))/\log^2 n\right).\end{aligned}$$

Furthermore, if b_1 and b_2 are picked so that $T_2'(n,b_1)=O\left(T_1'(n,b_1)\right)$ and $T_2(n,b_2,\boldsymbol{t})=O\left(T_1(n,b_2,\boldsymbol{t})\right)$, then we have

$$\mathbb{E}\left(\widehat{H}_2(\boldsymbol{t})-H(\boldsymbol{t})\right)=O\left((T_1'(n,b_1)+T_1(n,b_2,\boldsymbol{t}))/\log n\right).$$

Remarks.

(i) Under the 5 scenarios listed in the first remark after Theorem 3.3.4, if $2\bar{\psi}>d$ and

$$b_1=b_2\sim(n^d\log^2 n)^{-\frac{1}{2p+d}},$$

the convergence rate of $\widehat{H}_2(\boldsymbol{t})$ will be $(n^d\log^2 n)^{-\frac{p}{2p+d}}$, which is the same as the convergence rate of $\widehat{H}(\boldsymbol{t})$ when σ^2 is known.

(ii) To the best of our knowledge, up to this point, the most complete asymptotic theory in the existing literature for the estimation of H when σ^2 is unknown were established

for the estimators QV and IR in Bardet and Surgailis (2013) for $d = 1$. If $p = 2$, the rate obtained in their equation (4.14) is $O_p(n^{-2/5+\varepsilon})$ for any $\varepsilon > 0$. This is slower than the optimal rate described in (i) for $\widehat{H}_2$, which is $O_p(n^{-2/5}(\log n)^{-4/5})$. As explained before, Bardet and Surgailis (2013) essentially does not take higher-order smoothness into account in defining the estimators.

3.5 Selection of q and b

In this section we consider the selection of q and b in $\widehat{H}$.

3.5.1 Selection of q

To consider the choice of q we examine the asymptotic mean squared error of $\widehat{H}(\boldsymbol{t})$, or equivalently, the asymptotic bias and variance of $G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})$ given by Theorem 3.3.4 under the assumption that the bandwidth b is chosen optimally. We will focus on the 5 scenarios in the first remark after Theorem 3.3.4 and $p \in \mathbb{N}$.

Let $b \sim (n^d \log^2 n)^{-\frac{1}{2p+1}}$, which is the bandwidth that leads to the optimal rate for $\widehat{H}$. When $p \in \mathbb{N}$, it can be seen from (3.46) that the dominant term of the bias is $R(b, \boldsymbol{t})$ which does not depend on q. Thus, we only need to check the effect of q on variance In all of the 5 scenarios, the variance is proportional to $\sigma^2_{H(\boldsymbol{t})}$ by Theorem 3.3.3. As $\sigma^2_{H(\boldsymbol{t})}$ depends on both q and $H(\boldsymbol{t})$, the effect of q may be different for different values of $H(\boldsymbol{t})$.

In Figure 3.1, we present plots of σ^2_H versus H for $d = 1$ and $q = 1, 2, 3, 4$, where the values of σ^2_H are numerically computed using 80 terms in the expansion. One observes from this plot that the values of σ^2_H for $q = 3, 4$ are uniformly larger than those for $q = 2$. For $q > 4$, σ^2_H increases progressively but the plots for those are omitted for clarity of presentation. Consequently, there is no need to let $q > 2$. We also observe that for $H \leq 0.66$, σ^2_H is the smallest when $q = 1$. However, for $H > 0.66$, σ^2_H increases rapidly if $q = 1$. To conclude, $q = 2$ is generally a safe choice, but if $H(\boldsymbol{t}) < 0.66$ for most of $\boldsymbol{t}$, $q = 1$ may lead to better estimation results. The discussions for $d \geq 2$ are similar.

3.5.2 Selection of b

In this subsection, we aim to construct a bandwidth selection method for $\widehat{H}$ when $d = 1$ and $q = 2$. For a fixed $\delta > 0$, define

$$\mathrm{MISE}(b) = \mathbb{E}\left(\int_{\Omega_\delta} \left[\widehat{H}_{n,b}(t) - H(t)\right]^2 dt\right),$$

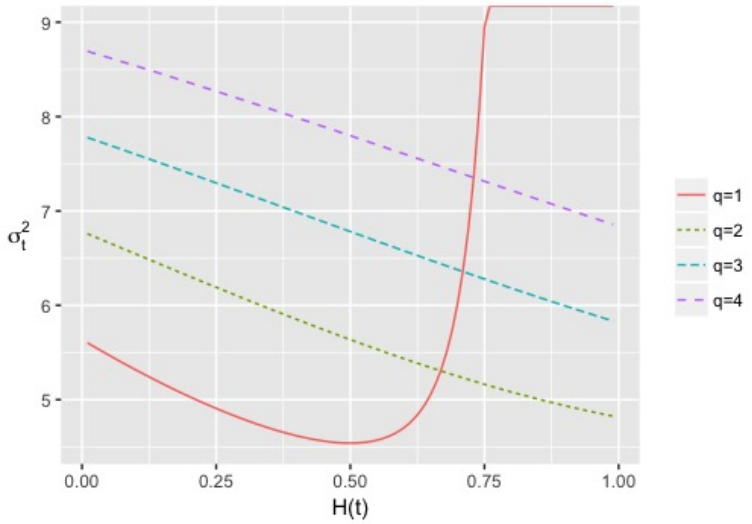


Figure 3.1: Plots of σ_H^2 for $q = 1, 2, 3, 4$.

which is a common criterion for measuring goodness of the fit. Denote by b^* the optimal bandwidth based on $\mathrm{MISE}(b)$. By Theorem 3.3.4 and the remarks that follow, we conclude that $b^* := O\left(n \log^2 n\right)^{-\frac{1}{2p+1}}$. Instead of minimizing $\mathrm{MISE}(b)$ directly, minimizing a discretized version of $\mathrm{MISE}(b)$ is usually preferable for computations. Clearly, we can drop the term $H^2(t)$ which does not depend on b. These lead to the objective function

$$R(b) = \frac{1}{\#(\Omega_m \cap \Omega_\delta)} \sum_{t \in \Omega_m \cap \Omega_\delta} \left(\widehat{H}_{n,b}^2(t) - 2\widehat{H}_{n,b}(t) H(t) \right)$$

for some large m where, as usual, $\Omega_m = \{(i - 1/2)/m, j = 1, \ldots, m\}$. Note that we highlight the dependence on data and bandwidth in the notion $\widehat{H}_{n,b}^2(t)$.

However, $R(b)$ is still not calculable as $H(t)$ is unknown. We follow an idea from Coeurjolly (2005) and replace $H(t)$ by an undersmoothed estimate. Specifically, let

$$\widehat{R}(b) = \frac{1}{\#(\Omega_m \cap \Omega_\delta)} \sum_{t \in \Omega_m \cap \Omega_\delta} \left(\widehat{H}_{n,b}^2(t) - 2\widehat{H}_{n,b}(t) \widetilde{H}_{n,b^2}(t) \right),$$

where

$$\widetilde{H}_{n,b^2}(t) = \frac{1}{\#(B_{b^{2p}}(t) \cap \Omega_m \cap \Omega_\delta)} \sum_{t' \in B_{b^{2p}}(t) \cap \Omega_m \cap \Omega_\delta} \widehat{H}_{n,b^2}(t'),$$

and $B_{b^{2p}}(t)$ is the interval centered at t with radius b^{2p}. Now, we can select our bandwidth

to be

$$\widehat{b}^* = \underset{b \in E}{\operatorname{argmin}}\, \widehat{R}(b),$$

where E is the interval defined by

$$E = \left[\kappa_1 n^{-\frac{p+1}{4p+2}} (\log n)^{\frac{p}{2p+1}}, \kappa_2 (n \log^2 n)^{-\frac{1}{2p+1}}\right], \tag{3.25}$$

with κ_1, κ_2 being any two positive constants. We will refer to this as the LSCV approach, a terminology borrowed from the density estimation literature. We expect $\widehat{R}(\widehat{b}^*)$ to closely approximate $R(b^*)$ if m is chosen large enough. For $\widehat{H}_2$, with $b_1 = b_2$, we can adopt the same strategy to select the bandwidth.

3.6 A simulation study

We primarily focus on the case $d = 1$ but will also briefly discuss the $d = 2$ case.

First let $d = 1$ and assume that $H(t) = 0.5 + 0.4\sin(4\pi t), t \in [0, 1]$, and $\sigma^2 = 1$. The estimators that we compare are $\widehat{H}, \widetilde{H}, \widehat{H}_1, \widehat{H}_2, \widetilde{H}_2$ plus the approaches QV and IR considered by Coeurjolly (2005), Bardet and Surgailis (2013) and Bertrand et al. (2013), where $\widetilde{H}, \widetilde{H}_2$ are $\widehat{H}$ and $\widehat{H}_2$ with bandwidth b selected by the LSCV approach in Section 5. Note that σ^2 is assumed known for $\widehat{H}$ and $\widetilde{H}$ but unknown for the other procedures. To implement our approaches, we set $p = 3$, $q = 2$ and use the Epanechnikov kernel. QV and IR do not produce estimation results in the boundary areas while all our approaches do. As such, all the ISEs are computed on the interval $[0.1, 0.9]$ to make the comparisons fair.

In the first set of simulations, we compare the four procedures that do not require the knowledge of σ^2, $\widehat{H}_1, \widehat{H}_2$ and QV and IR. The bandwidth parameter b for all estimators is taken from the set $0.3 \times 0.8^{0:19}$, and the results reported are based on the b's having the smallest MISEs. Obviously, this requires the knowledge of H and is not a data-driven bandwidth selector. The plots in Figure 3.2 are the empirical pointwise quartiles based on $n = 10,000$ and 1,000 simulation runs; the true H is also plotted in red. Among the four approaches, $\widehat{H}_2$ is the clearly winner and actually substantially improves upon its preliminary procedure $\widehat{H}_1$. One could also see that the performance of IR significantly lags the other procedures for small values of $H(t)$.

To visualize the convergence rates of all eight estimators, we let the number of observations vary from 1,000 to 10,000 and compute the $\sqrt{\text{MISE}}$s based on 1,000 simulation runs. The results are displayed in Table 3.1 and Figure 3.3, where Figure 3.3 contains the plots

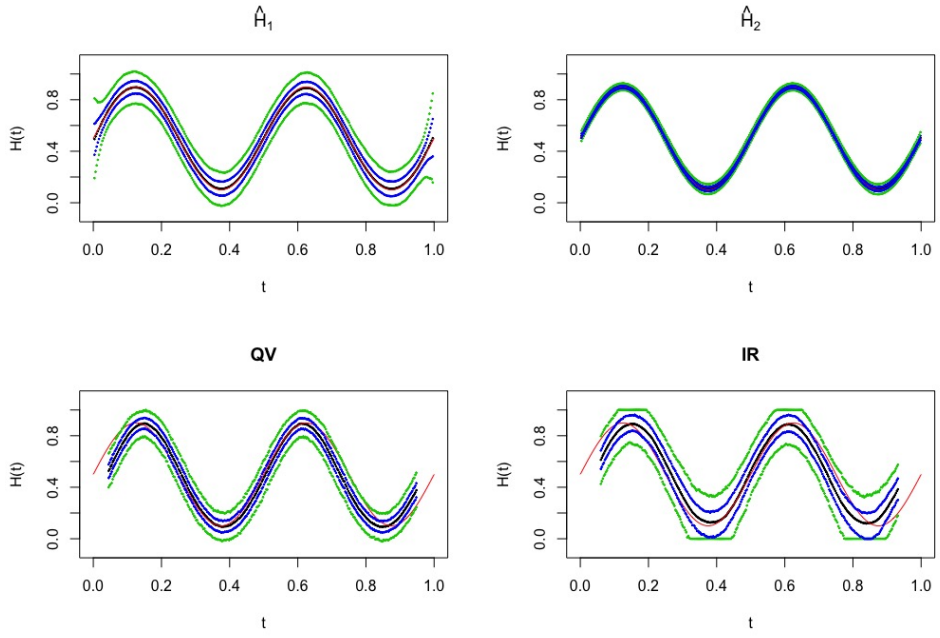


Figure 3.2: Empirical quantiles of $\widehat{H}_1(t), \widehat{H}_2(t), \mathrm{QV}(t), \mathrm{IR}(t)$ with $n = 10,000$ based on 1,000 simulation runs; green curves are 5% and 95% pointwise empirical quantiles, respectively; blue curves are 25% and 75% empirical quantiles, respectively; black curves are empirical means.

for the log MISEs. The bandwidths for $\widehat{H}, \widehat{H}_1, \widehat{H}_2$, QV and IR are selected optimally as described in the previous paragraph while $\widetilde{H}$ and $\widetilde{H}_2$ use data-driven bandwidth determined by LCSV. By conducting linear regression on the $\log\sqrt{\mathrm{MISE}}$s, we arrive at very crude estimates of the convergence rates, which are $n^{-0.54}$, $n^{-0.52}$, $n^{-0.44}$, $n^{-0.49}$, $n^{-0.47}$, $n^{-0.39}$ and $n^{-0.37}$ for $\widehat{H}$, $\widetilde{H}$, $\widehat{H}_1$, $\widehat{H}_2$, $\widetilde{H}_2$, QV and IR, respectively. It is important to note that although $\widehat{H}$ and $\widetilde{H}$ performs the best in this experiment assuming the true σ^2, they will have a large bias when σ^2 is misspecified. The performance of all other methods are not affected by the change of σ^2.

To evaluate the bandwidth selector LCSV in detail, we focus on $\widehat{H}$ with $n = 1,000$ and select b from the set $0.3 \times 0.8^{0:8}$. This is done $1,000$ times. We compare our selection approach with the "oracle" bandwidth, which is obtained by assuming we know the true $H(t)$ and select the bandwidth with smallest MISE for each run. Table 3.2 contains the number of times the individual bandwidths in $0.3 \times 0.8^{0:8}$ are selected by oracle and LSCV. Figure 3.4 contains the empirical histograms of the integrated squraed errors of $\widehat{H}$ and $\widetilde{H}$. It can be seen that the bandwidths selected by two approaches are fairly close. The result for $\widetilde{H}_2$ are very similar, and we omit them here. However, the bandwidth selection method does not work well for $\widetilde{H}_1$ and tends to select the smallest bandwidth in the range

of candidate bandwidths.

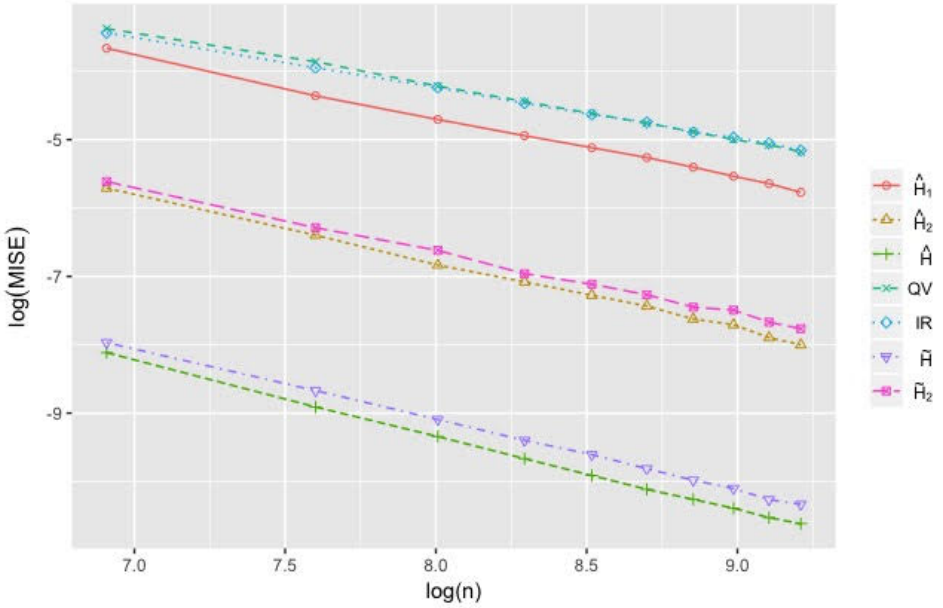


Figure 3.3: Plot of $\log(\text{MISE})$ versus $\log n$

n	QV	IR	$\widehat{H}_1$	$\widehat{H}_2$	$\widehat{H}$	$\widetilde{H}_2$	$\widetilde{H}$
1000	0.1849	0.1792	0.1602	0.0576	0.01574	0.0604	0.0186
2000	0.1451	0.1388	0.1131	0.0408	0.01060	0.0431	0.0131
3000	0.1214	0.1201	0.0951	0.0328	0.00845	0.0365	0.0106
4000	0.1085	0.1071	0.0845	0.0290	0.00736	0.0308	0.0091
5000	0.0993	0.0987	0.0774	0.0263	0.00656	0.0285	0.0082
6000	0.0926	0.0932	0.0720	0.0243	0.00595	0.0264	0.0074
7000	0.0869	0.0866	0.0671	0.0221	0.00545	0.0241	0.0068
8000	0.0821	0.0834	0.0628	0.0212	0.00504	0.0236	0.0064
9000	0.0790	0.0800	0.0595	0.0193	0.00473	0.0216	0.0059
10000	0.0751	0.0758	0.0558	0.0183	0.00449	0.0206	0.0057

Table 3.1: $\sqrt{MISE}$ for all estimators with a range of sample sizes

Finally, we present some simulation results for the case $d = 2$ to illustrate the performance of our three estimators $\widehat{H}$, $\widehat{H}_1$ and $\widehat{H}_2$. The model that we consider is the mBm on $[0,1] \times [0,1]$ with Hurst function $H(x,y) = 0.5 + 0.4\sin(4\pi x)\sin(2\pi y)$, where we assume that the process is observed on a regular grid of size 166×166. For ease of presentation, we only show the estimation results on the circle $(x-.5)^2 + (y-.5)^2 = 0.25^2$. As before we let $p = 3$ and $q = 2$ where the direction of differencing is taken as $\boldsymbol{h} = (0,1)$. With bandwidths equal to $0.15 \times 0.8^{0:5}$, 300 runs of simulations were carried out. Figure 3.5 presents

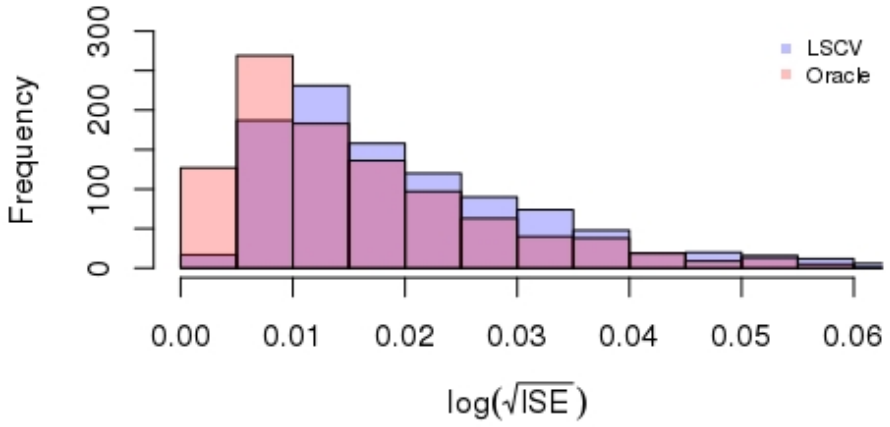


Figure 3.4: Empirical histograms of $\log\left(\sqrt{ISE}\right)$ based on $\widehat{H}$ with the oracle bandwidth and that selected by LSCV

	0.3×0.8^3	0.3×0.8^4	0.3×0.8^5	0.3×0.8^6	0.3×0.8^7	0.3×0.8^8
LSCV	434	413	115	29	4	5
Oracle	60	584	328	27	1	0

Table 3.2: Frequencies with which various bandwidths are picked by LSCV and oracle for $\widehat{H}$ based on $1,000$ simulation runs

the the empirical quantiles of the estimation results for this simulation experiment using the optimal bandwidths based on MISE. One can see that $\widehat{H}$ continues to perform quite well assuming that σ^2 is known, whereas for $\widehat{H}_1, \widehat{H}_2$ that do not require a priori knowledge of σ^2, $\widehat{H}_2$ outperforms $\widehat{H}_1$ by a wide margin.

3.7 Discussions and extensions

In this section we briefly discuss the problems of relaxing the assumptions of gridded data and constant variance.

3.7.1 Non-gridded data

So far we focus on the estimation of the Hurst function of the mBm based on gridded data. In some specialized non-grid settings, our approach can be readily modified to produce estimators that have essentially the same rates of convergence as for the gridded case. To

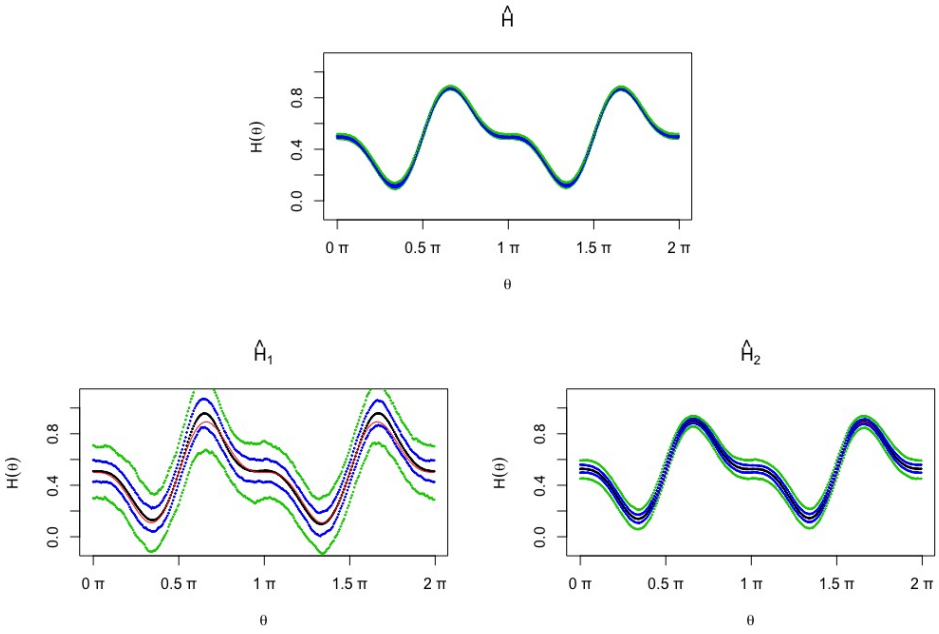


Figure 3.5: Empirical quantiles of $\widehat{H}, \widehat{H}_1, \widehat{H}_2$ on the circle $(x-.5)^2+(y-.5)^2 = 0.25^2$ based on 300 simulation runs; $\theta = \arctan((y-.5)/(x-.5))$; red curves are the true function; green curves are 5% and 95% pointwise empirical quantiles, respectively; blue curves are 25% and 75% empirical quantiles, respectively; black curves are empirical means.

demonstrate, we focus on $d = 1$ and follow the approach of Loh (2015). Assume that we have observations $X(t_i), 1 \le i \le n$, with $t_i = \varphi((i-1/2)/n)$, where φ satisfies the following condition:

[B] $\varphi \in \mathscr{H}_{p+1}([0,1]), p \ge 1$, is a strictly monotone and surjective mapping from $[0,1]$ to $[0,1]$, with its first order derivative $\varphi^{(1)}$ bounded away from 0.

Observe that [B] guarantees that the gaps between neighboring t_i's are of the order n^{-1}. Denote the modified estimators of $\widehat{H}, \widehat{H}_1$ and $\widehat{H}_2$ for this setting as $\widehat{H}', \widehat{H}_1'$ and $\widehat{H}_2'$, respectively, whose definitions involve some new notation to be introduced below. For any given $t \in (0,1)$ and $q = 1, 2, \ldots$, define $x_i(t) = \varphi(\varphi^{-1}(t) + i/n), i = 0, \ldots, q$, and $a_i = (-1)^i \binom{q}{i}$. Also, let the symbol $\Delta_{1/n}^q \mathfrak{X}(t)$ stand for the quantity $\sum_{i=0}^q a_i X(x_i(t))$. Observe that $x_i(t_j) = t_{i+j}$, which shows that $\Delta_{1/n}^q \mathfrak{X}(t_j)$ can be computed from data. It can be shown (cf. (3.6)) that

$$2\mathbb{E} \log \left| \Delta_{1/n}^q \mathfrak{X}(t) \right| \approx G(H(t); n, \varpi(t)),$$

where $\varpi(t) := \varphi^{(1)}(\varphi^{-1}(t))$. Thus, $G(H(t); n, \varpi(t))$ can be estimated nonparametrically by the local polynomial estimator based on the log-transformed differenced data

$2\log|\Delta^q_{1/n}\mathfrak{X}(t_j)|$, as was done for the gridded case. This motivates the new estimators $\widehat{H}'$, $\widehat{H}'_1$ and $\widehat{H}'_2$.

Specifically, the definition of $\widehat{H}'_1(t)$ is unchanged from $\widehat{H}_1(t)$ since the definition does not involve inversion of $G(H;n,h)$. For $\widehat{H}'$ and $\widehat{H}'_2$, however, since we need to solve $G(\widehat{H;n,\varpi}(t))$ for H, the function $\varpi(t)$ must be estimated. The condition [B] implies that $\varpi(t)$ is differentiable, from which it is easily concluded that $\varpi(t)$ can be estimated with precision $O(1/n)$ using the pairs $(i/n,t_i)$; for instance, a smoothed version of the naive estimator $\sum_i n(t_{i+1}-t_i)I_{(t_i,t_{i+1}]}(t)$ will suffice. Thus, to define $\widehat{H}'$ and $\widehat{H}'_2$, we first estimate $\varpi(t)$ by some $\widehat{\varpi}(t)$ and then proceed to estimate $H(t)$ in much the same way as in $\widehat{H}$ and $\widehat{H}_2$ using $\widehat{\varpi}(t)$ and the $\Delta^q_{1/n}\mathfrak{X}(t_j)$. Section 3.8.7 shows that the rates of convergence of $\widehat{H}'$, $\widehat{H}'_1$ and $\widehat{H}'_2$ are largely unchanged from their gridded-data counterparts $\widehat{H}$, $\widehat{H}_1$ and $\widehat{H}_2$. A simulation study is also included there to demonstrate this numerically.

We could go beyond the setting of [B] and consider data locations that are less regular; for instance, the t_i's are distributed as iid. uniform [0,1] or belong to a Poisson point process. In that case, one might be able to exclude those data for which the gaps are not of the order n^{-1} and also improvise in some ways in $\widehat{H}'$, $\widehat{H}'_1$ and $\widehat{H}'_2$ to still obtain meaningful estimators. However, much of what we know so far in that regard is still preliminary and will require more investigations.

3.7.2 Non-constant variance

Consider the model

$$Y(\boldsymbol{t}) = \sigma(\boldsymbol{t})X(\boldsymbol{t})$$

where X is mBm as defined in Definition 3.1.1 with $\sigma = 1$, and $\sigma(\boldsymbol{t})$ a function that determines the variance of $Y(\boldsymbol{t})$. Allowing $\sigma(\boldsymbol{t})$ to vary with $\boldsymbol{t}$ broadens the mBm process considerably. As before, assume that Y is observed on a grid.

If $\sigma(\boldsymbol{t})$ is known, one can simply divide $Y(\boldsymbol{t})$ by $\sigma(\boldsymbol{t})$ and all our estimators will apply. We only consider below the case when $\sigma(\boldsymbol{t})$ is unknown. Define

$$G(H;n,\boldsymbol{h},\sigma) := -2H\log n + \log\sigma^2 + \log g(H,\boldsymbol{0},\boldsymbol{h}) + \mathbb{E}\log\chi^2_1.$$

As the constant variance case, one can first obtain $G(H(\widehat{\boldsymbol{t}});n,\boldsymbol{h},\sigma(\boldsymbol{t}))$ by local polynomial regression based on data $\Delta^q_{\boldsymbol{h}/n}Y(\boldsymbol{t}_i)$. Intuitively, the properties of $G(H(\widehat{\boldsymbol{t}});n,\boldsymbol{h},\sigma(\boldsymbol{t}))$ should be very close to $G(\widehat{H(\boldsymbol{t});n,\boldsymbol{h}})$ based on $X(\boldsymbol{t})$ so long as $\sigma(\boldsymbol{t})$ is sufficiently smooth. To that end we define

[S] $\sigma(\cdot) \in \mathscr{H}_p((0,1)^d)$ and $\inf_{\boldsymbol{t}} \sigma(\boldsymbol{t}) > \delta$ for some $\delta > 0$.

And the following proposition follows readily.

Proposition 3.7.1. *Suppose [K], [A1]-[A3] and [S] hold. Additional, we assume $q \geq 2$ and $d \leq 3$. Then for any $\delta \in (0,1)$, we have uniformly for any $\boldsymbol{t} \in \Omega_\delta$,*

$$\mathbb{E}\left(G(H(\boldsymbol{t}); n, \boldsymbol{h}, \sigma(\boldsymbol{t}))\right) - G(H(\boldsymbol{t}); n, \boldsymbol{h}, \sigma(\boldsymbol{t})) = O\left(T_1(n, b, \boldsymbol{t})\right), \tag{3.26}$$

and

$$\mathrm{Var}\left(G(H(\boldsymbol{t}); n, \boldsymbol{h}, \sigma(\boldsymbol{t}))\right) = O\left((nb)^{-d}\right).$$

First we consider $\widehat{H}_1$ which does not require knowing $\sigma(\boldsymbol{t})$. Following the same steps as in the proof of Theorem 3.4.1 and Proposition 3.7.1, it is straightforward to conclude that the optimal convergence rate of $\widehat{H}_1$ based on Y is the same as that based on X when $d \leq 3$ and $q \geq 2$, which is $n^{-\frac{dp}{2p+d}}$. This is somewhat worse than the minimax rate $(n^d \log^2 n)^{-\frac{p}{2p+d}}$. One possible way to improve upon $\widehat{H}_1$ is to consider a modified version of $\widehat{H}_2$ obtained as follows. Step 1: estimate $H(\boldsymbol{t})$ by $\widehat{H}_1(\boldsymbol{t})$, step 2: based on $\widehat{H}_1$, estimate $\sigma(\boldsymbol{t})$ nonparametrically by some $\widehat{\sigma}(\boldsymbol{t})$, and step 3: based on $\widehat{\sigma}(\boldsymbol{t})$, estimate $H(\boldsymbol{t})$ by $\widehat{H}(\boldsymbol{t})$, i.e. $\widehat{H}(\boldsymbol{t}) = G^{-1}(G(H(\widehat{\boldsymbol{t}); n, \boldsymbol{h}}, \sigma(\boldsymbol{t})); n, \boldsymbol{h}, \widehat{\sigma}(\boldsymbol{t}))$. There are a few challenges in making this approach work. First, it involves the choices of three smoothing parameters, one for each of $\widehat{H}_1, \widehat{\sigma}$ and $\widehat{H}$. This makes the implementation difficult. Also, in order for $\widehat{H}_2(\boldsymbol{t})$ to achieve the rate $(n^d \log^2 n)^{-\frac{p}{2p+d}}$, it is necessary that $\widehat{\sigma}(\boldsymbol{t})$ first achieves the rate $n^{-\frac{dp}{2p+d}} (\log n)^{-\frac{d}{2p+d}}$. We conjecture that this is possible only if $H(\boldsymbol{t}) \leq .5$; see additional discussions and simulation results in Section 3.8.8. A complete solution of this problem is beyond the scope of this thesis.

3.8 Appendix

3.8.1 Proof of Theorem 3.3.1

The proof is based on the general approach introduced in Theorem 2.5 of Tsybakov (2009), which, in our context, can be described as follows. For candidate Hurst functions H_1 and H_2, let $K(H_1 \| H_2)$ be the the Kullback-Leibler divergence between mBms with Hurst functions H_1 and H_2 observed on $\Omega_n = \{(j - 1/2)/n, j = 1, \ldots, n\}$. Suppose that

$H_{i\delta}, i = 0, 1, \ldots, N_\delta$ are $N_\delta + 1$ candidate Hurst functions in $\mathscr{H}_p((0,1), M)$ satisfying

$$\|H_{j\delta} - H_{k\delta}\|_q \geq 2\delta > 0,\ 0 \leq j, k \leq N_\delta, \tag{3.27}$$

and

$$\frac{1}{N_\delta} \sum_{i=1}^{N_\delta} K(H_{i\delta} \| H_{0\delta}) \leq \alpha \log(N_\delta), \tag{3.28}$$

where $\alpha \in (0, 1/8)$. Then

$$\inf_{\widetilde{H}_n} \sup_{H \in \mathscr{H}_p((0,1),M)} \mathbb{P}_H \left(\|\widetilde{H}_n - H\|_q \geq \delta \right) \geq \frac{\sqrt{N_\delta}}{1 + \sqrt{N_\delta}} \left(1 - 2\alpha - \sqrt{\frac{2\alpha}{\log N_\delta}} \right) > 0.$$

We proceed to construct a set of functions that satisfy the conditions (3.27) and (3.28) above where δ and N_δ depend on n, and δ corresponds to the rate in the theorem.

First, let $H_{0,\delta} \equiv 1/2$, which is the Hurst function of the standard Brownian motion. Define

$$\kappa(x) = \begin{cases} \exp\left\{ -\frac{1}{x + \frac{1}{2}} - \frac{1}{x - \frac{1}{2}} \right\} & -\frac{1}{2} < x < \frac{1}{2} \\ 0 & \text{otherwise.} \end{cases}$$

It is clear that κ is a non-negative function with support $(-1/2, 1/2)$ and has derivatives of all orders. Let

$$M_1 = \sup_x \left| \kappa^{(\lfloor p \rfloor + 1)}(x) \right| / M \quad \text{and} \quad \phi(x) = \kappa(x)/M_1.$$

Also, for some sequence $m = m_n$, set

$$\phi_{mj}(x) = m^{-p} \phi(mx - j), j = 1, \ldots, m - 1.$$

We will specify m later and define δ and N_δ in terms of m.

It's clear that the support of ϕ_{mj} is $((j - 1/2)/m, (j + 1/2)/m)$, and therefore the ϕ_{mj}'s have non-overlapping supports. For any vector $\boldsymbol{a} = (a_1, \ldots, a_{m-1})$ with each $a_j = 0$ or 1, define

$$H_{\boldsymbol{a}}(t) = \frac{1}{2} + \sum_{j=1}^{m-1} a_j \phi_{mj}(x). \tag{3.29}$$

Clearly, for sufficiently large m and all $\boldsymbol{a}$,

$$\begin{aligned}
&\sup_{x,y\in(0,1)} \frac{|H_{\boldsymbol{a}}^{(\lfloor p\rfloor)}(x) - H_{\boldsymbol{a}}^{(\lfloor p\rfloor)}(y)|}{|x-y|^{p-\lfloor p\rfloor}} \\
&\leq \sup_{x,y\in[1/(2m),3/(2m)]} \frac{m^{-(p-\lfloor p\rfloor)}|\phi^{(\lfloor p\rfloor)}(mx-1) - \phi^{(\lfloor p\rfloor)}(my-1)|}{|x-y|^{p-\lfloor p\rfloor}} \\
&\leq M \sup_{x,y\in[1/(2m),3/(2m)]} \frac{m^{-(p-\lfloor p\rfloor)}|mx-my|}{|x-y|^{p-\lfloor p\rfloor}} \\
&\leq M \sup_{x,y\in[1/(2m),3/(2m)]} \frac{m^{-(p-\lfloor p\rfloor)}|mx-my|^{p-\lfloor p\rfloor}}{|x-y|^{p-\lfloor p\rfloor}} = M.
\end{aligned}$$

For different $\boldsymbol{a}$ and $\boldsymbol{a}'$, we have

$$\|H_{\boldsymbol{a}} - H_{\boldsymbol{a}'}\|_q = \left\|\sum_j (a_j - a_j')\phi_{mj}\right\|_q = m^{-p-1/q}\left(\sum_j |a_j - a_j'|\right)^{1/q}\|\phi\|_q.$$

Denote by D_1 the set of all vectors $\boldsymbol{a}$ such that if $\boldsymbol{a}, \boldsymbol{a}' \in D_1$,

$$\sum_j |a_j - a_j'| > \frac{m-1}{4}.$$

By the Varshamov-Gilbert bound, we have

$$\#(D_1) \geq \exp\{(m-1)/8\}.$$

Now take the $H_{i\delta}$'s to be the $H_{\boldsymbol{a}}, a \in D_1$, and $N_\delta = \#(D_1)$. It follows that

$$\|H_{i\delta} - H_{j\delta}\|_q \geq m^{-p}\left(\frac{m-1}{4m}\right)^{1/q}\|\phi\|_q.$$

Then let

$$\delta = m^{-p}\left(\frac{m-1}{4m}\right)^{1/q}\|\phi\|_q = O(m^{-p}).$$

It follows from Proposition 3.8.1 below with $a_n = O(m^{-p})$ and $b_n = O(m^{-(p-1)})$ that there exists some $C \in (0,\infty)$ such that, uniformly in i,

$$K(H_{i\delta}\|H_{0\delta}) \leq C(m^{-2(p-1)} + nm^{-2p})\log^2 n$$

for large n. Fix any $\alpha \in (0, 1/8)$. To ensure that (3.28) holds, it suffices to find m such that

$$\alpha \log(N_\delta) \geq \alpha \frac{m-1}{8} \geq C(m^{-2(p-1)} + nm^{-2p}) \log^2 n.$$

To achieve the slowest rate δ, we pick the smallest m for each n such that this inequality holds, in which case

$$m \sim (C/\alpha)(n \log^2 n)^{\frac{1}{2p+1}}.$$

The proof is complete.

□

The proof of Theorem 3.3.1 now hinges on the computation of $K(H_{i\delta} \| H_{0\delta})$, which is accomplished in the proposition below.

Proposition 3.8.1. *Let $H_0(t) \equiv 1/2$ and $H(t) = 1/2 + \phi(t)$, where $\sup_t |\phi(t)| \leq a_n$ and $\sup_t |\phi'(t)| \leq b_n$ with $(a_n + b_n) \log n \to 0$. Then there is a finite (universal) constant C such that*

$$\limsup_{n \to \infty} \left\{ \left(b_n^2 + n a_n^2 \right) \log^2 n \right\}^{-1} K(H \| H_0) \leq C. \tag{3.30}$$

Proof. Denote by X the mBm with Hurst function H. To simplify notation, we assume that the data is observed on $\{i/n, i = 1, \ldots, n\}$ instead of on Ω_n. This does not change the conclusion in the result. Define

$$Y_i = \sqrt{n} \left(X\left(\frac{i}{n}\right) - X\left(\frac{i-1}{n}\right) \right), \ i = 1, \ldots, n.$$

Let Σ be covariance matrix of $(Y_1, \ldots, Y_n)$. It is well known that

$$K(H \| H_0) = \frac{1}{2} \left[-\log |\Sigma| - n + \operatorname{tr}(\Sigma) \right]. \tag{3.31}$$

Let $D = (d_{ij})_{n \times n} = \Sigma - I$ and λ_i the eigenvalues of D. We will establish the following:

(i) $d_{ii} = O((a_n + n^{-1} b_n) \log n)$ uniformly for all i

(ii) $\sum_{i \neq j} d_{ij}^2 = O\left((b_n \log n)^2 + n a_n^2 \right)$

(iii) $\sum_{j=1}^n |d_{ij}| \to 0$, uniformly for all i.

It follows that

$$-\log|\Sigma| = -\sum_i \log(1+\lambda_i) = -\mathrm{tr}(D) + O(\mathrm{tr}(D^2)), \tag{3.32}$$

where the final equality uses the fact that $\lambda_i \to 0$ uniformly as $n \to \infty$, which follows from (i), (iii) and Gershgorin Circle Theorem. By (3.31),

$$K(H\|H_0) = O(\mathrm{tr}(D^2)) = O(\|D\|_F^2),$$

and (3.30) follows from (i) and (ii). The rest of the proof will focus on establishing (i)-(iii).

We first refer to the following identity from Ayache et al. (2000),

$$\begin{aligned}&\int_{\mathbb{R}} \frac{(e^{itx}-1)(e^{-isx}-1)}{|x|^{2H+1}}dx \\ &= \frac{1}{2}\left(\frac{\pi}{H\Gamma(2H)\sin(\pi H)}\right)(|t|^{2H}+|s|^{2H}-|t-s|^{2H}),\end{aligned} \tag{3.33}$$

which holds for all fixed constant $H \in (0,1)$, where $i = \sqrt{-1}$. Thus, the covariance function for the mBm with Hurst function H is

$$C(s,t) = \frac{1}{D(H(s))D(H(t))}\int_{\mathbb{R}} \frac{(e^{isx}-1)(e^{-itx}-1)}{|x|^{H(s)+H(t)+1}}dx,$$

where $D(H)$ is defined in Definition 3.1.1. By letting $t = s = 1$ in (3.33), we get $D^2(H) = \frac{\pi}{2H\Gamma(2H)\sin(\pi H)}$. Define

$$\Psi_x(t) = \frac{1-e^{itx}}{D(H(t))|x|^{H(t)}},$$

and write

$$\begin{aligned}d_{jk} &= n\left(C\left(\frac{j}{n},\frac{k}{n}\right)+C\left(\frac{j-1}{n},\frac{k-1}{n}\right)-C\left(\frac{j-1}{n},\frac{k}{n}\right)-C\left(\frac{j}{n},\frac{k-1}{n}\right)\right) \\ &\qquad - I(j=k) \\ &= n\int_{\mathbb{R}} \frac{1}{|x|}\Delta_{\frac{1}{n}}\Psi_x\left(\frac{j}{n}\right)\overline{\Delta_{\frac{1}{n}}\Psi_x\left(\frac{k}{n}\right)}dx - I(j=k).\end{aligned}$$

Now,

$$\begin{aligned}\Delta_{\frac{1}{n}}\Psi_x(t) &= \frac{1-e^{itx}}{D(H(t))|x|^{H(t)}} - \frac{1-e^{i(t-\frac{1}{n})x}}{D(H(t-\frac{1}{n}))|x|^{H(t-\frac{1}{n})}} \\ &= \frac{e^{i\left(t-\frac{1}{n}\right)x}-e^{itx}}{D(H(t))|x|^{H(t)}} - \left(\frac{1-e^{i\left(t-\frac{1}{n}\right)x}}{D(H(t))|x|^{H(t)}} - \frac{1-e^{i(t-\frac{1}{n})x}}{D(H\left(t-\frac{1}{n}\right))|x|^{H(t-\frac{1}{n})}}\right) \\ &:= f_1(x,t) + f_2(x,t).\end{aligned}$$

Thus, we can express d_{jk} in terms of f_1, f_2 as

$$\begin{aligned}d_{jk} = n\int_{\mathbb{R}} \frac{1}{|x|}\left(f_1\left(x,\frac{j}{n}\right)+f_2\left(x,\frac{j}{n}\right)\right)\overline{\left(f_1\left(x,\frac{k}{n}\right)+f_2\left(x,\frac{k}{n}\right)\right)}dx \\ - I(j=k)\end{aligned}$$

where overline stands for complex conjugate. By the mean-value theorem and dominated convergence theorem,

$$\begin{aligned}&\left\| n\int_{\mathbb{R}} \frac{1}{|x|} f_2\left(x,\frac{j}{n}\right)\overline{f_2\left(x,\frac{k}{n}\right)}dx\right\| \\ &= \left\| \frac{1}{n}\int_{\mathbb{R}} \frac{1}{|x|}\frac{(1-e^{i\frac{j-1}{n}x})(D'(H(\widetilde{t}_1))H'(\widetilde{t}_1)+D(H(\widetilde{t}_1))H'(\widetilde{t}_1)\log|x|)}{D(H(\widetilde{t}_1))|x|^{H(\widetilde{t}_1)}}\right. \\ &\quad \left. \times \overline{\frac{(1-e^{i\frac{k-1}{n}x})(D'(H(\widetilde{t}_2))H'(\widetilde{t}_2)+D(H(\widetilde{t}_2))H'(\widetilde{t}_2)\log|x|)}{D(H(\widetilde{t}_2))|x|^{H(\widetilde{t}_2)}}}dx\right\| \\ &\le \frac{M_1}{n}\left(\int_0^1 \frac{1}{x}\frac{x^2 b_n^2\log^2 x}{x^{1+2a_n}}dx + \int_1^\infty \frac{1}{x}\frac{b_n^2\log^2 x}{x^{1-2a_n}}dx\right)\end{aligned}$$

where $\widetilde{t}_1 \in [\frac{j-1}{n},\frac{j}{n}]$, $\widetilde{t}_2 \in [\frac{k-1}{n},\frac{k}{n}]$ and $M_1 > 0$ is a proper constant. The equations above require using mean-value theorem twice and consider the whole integral as the function to be differenced such that we can obtain $\widetilde{t}_1, \widetilde{t}_2$ unrelated to x. Since both integrals converge, we have uniformly for all $j, k \ge 1$,

$$\left\| n\int_{\mathbb{R}} \frac{1}{|x|} f_2\left(x,\frac{j}{n}\right)\overline{f_2\left(x,\frac{k}{n}\right)}dx\right\| = O(n^{-1}b_n^2). \tag{3.34}$$

Similarly, by Lemma 3.8.2, for some $\widetilde{t}_2 \in [\frac{k-1}{n}, \frac{k}{n}]$,

$$\begin{aligned}
&\left|\mathrm{Re}\left(n\int_{\mathbb{R}}\frac{1}{|x|}f_1\left(x,\frac{j}{n}\right)\overline{f_2\left(x,\frac{k}{n}\right)}dx\right)\right| \\
&= \Bigg|\mathrm{Re}\Bigg(\int_{\mathbb{R}}\frac{1}{|x|}\frac{e^{i(\frac{j}{n}-\frac{1}{n})x}-e^{i\frac{j}{n}x}}{D(H(\frac{j}{n}))|x|^{H(\frac{j}{n})}} \\
&\qquad\times\frac{\overline{(1-e^{i\frac{k-1}{n}x})}(D'(H(\widetilde{t}_2))H'(\widetilde{t}_2)+D(H(\widetilde{t}_2))H'(\widetilde{t}_2)\log|x|)}{D(H(\widetilde{t}_2))|x|^{H(\widetilde{t}_2)}}dx\Bigg)\Bigg| \qquad (3.35)\\
&\le M_2 n^{-1}b_n\left|\int_{\mathbb{R}}\frac{(e^{i(j-1)x}-e^{ijx})(1-e^{-i(k-1)x})(1+\log|x|+\log n)}{|x|^{1+H(\frac{j}{n})+H(\widetilde{t}_2)}}dx\right| \\
&= O(n^{-1}b_n\log n).
\end{aligned}$$

Note that the bound in (3.35) dominates that in (3.34).

To prove (i), we make use of (3.33) and perform a change of variables to get

$$\begin{aligned}
&n\int_{\mathbb{R}}\frac{1}{|x|}f_1\left(x,\frac{j}{n}\right)\overline{f_1\left(x,\frac{j}{n}\right)}dx \\
&= n\int_{\mathbb{R}}\frac{(1-e^{i\frac{1}{n}x})(1-e^{-i\frac{1}{n}x})}{|x|^{2H(\frac{j}{n})+1}D^2(H(\frac{j}{n}))}dx \qquad (3.36)\\
&= n^{1-2H(\frac{j}{n})}\int_{\mathbb{R}}\frac{(1-e^{ix})(1-e^{-ix})}{|x|^{2H(\frac{j}{n})+1}D^2(H(\frac{j}{n}))}dx \\
&= 1+O\left(a_n\log n\right).
\end{aligned}$$

Thus, (i) is established using (3.34)-(3.36).

To prove (ii), consider $j \neq k$ in which case we have

$$\begin{aligned}
&n\int_{\mathbb{R}}\frac{1}{|x|}f_1\left(x,\frac{j}{n}\right)\overline{f_1\left(x,\frac{k}{n}\right)}dx \\
&= \frac{n}{D(H(\frac{j}{n}))D(H(\frac{k}{n}))}\int_{\mathbb{R}}\frac{(e^{i\frac{j}{n}x}-e^{i\frac{j-1}{n}x})(e^{i\frac{k}{n}x}-e^{i\frac{k-1}{n}x})}{|x|^{1+H(\frac{j}{n})+H(\frac{k}{n})}}dx \\
&= \frac{n}{D(H(\frac{j}{n}))D(H(\frac{k}{n}))}\int_{\mathbb{R}}\frac{e^{i\frac{j-k}{n}x}\|1-e^{i\frac{1}{n}x}\|^2}{|x|^{1+H(\frac{j}{n})+H(\frac{k}{n})}}dx \\
&= \frac{n^{1-H(\frac{j}{n})-H(\frac{k}{n})}}{D(H(\frac{j}{n}))D(H(\frac{k}{n}))}\int_{\mathbb{R}}\frac{e^{i(j-k)x}\|1-e^{ix}\|^2}{|x|^{1+H(\frac{j}{n})+H(\frac{k}{n})}}dx.
\end{aligned}$$

It is easy to see that uniformly for all $j \neq k$, $\frac{n^{1-H(\frac{j}{n})-H(\frac{k}{n})}}{D(H(\frac{j}{n}))D(H(\frac{k}{n}))} = O(1)$, while, by Lemma 3.8.3,

the integral part is $O((j-k)^{-1}a_n)$. Thus, uniformly for all $j \neq k$,

$$n\int_{\mathbb{R}} \frac{1}{|x|} f_1\left(x, \frac{j}{n}\right) \overline{f_1\left(x, \frac{k}{n}\right)} dx = O\left((j-k)^{-1}a_n\right). \tag{3.37}$$

It follows that

$$\sum_{j\neq k}\left(n\int_{\mathbb{R}} \frac{1}{|x|} f_1\left(x, \frac{j}{n}\right) \overline{f_1\left(x, \frac{k}{n}\right)} dx\right)^2 \leq M_3 \sum_{l=1}^{n-1}(n-l)l^{-2}a_n^2 = O(na_n^2).$$

Combining this with (3.34) and (3.35) shows (ii).

To prove (iii), by (i), (3.34) and (3.35),

$$\begin{aligned}\sum_{j=1}^{n}|d_{ij}| &= O(a_n \log n) + \sum_{j\neq i} O\left((i-j)^{-1}a_n + n^{-1}b_n^2 + n^{-1}b_n \log n\right) \\ &= O((a_n+b_n)\log n) \to 0.\end{aligned}$$

This completes the proof. □

Lemma 3.8.2. *Suppose that $\delta_n > 0$ is a sequence of constants such that $n^{\delta_n} \to 1$. Then, uniformly for all $k, n \in \mathbb{N}$, $1 \leq k \leq n$,*

$$\begin{aligned}&\left|\int_{\mathbb{R}} \frac{1-e^{ikx}}{|x|^{2+\delta_n}}dx - \int_{\mathbb{R}} \frac{1-e^{i(k-1)x}}{|x|^{2+\delta_n}}dx\right| \\ &= \left|(k^{1+\delta_n} - (k-1)^{1+\delta_n})\int_{\mathbb{R}} \frac{1-e^{ix}}{|x|^{2+\delta_n}}dx\right| = O(1),\end{aligned}$$

and

$$\left|\int_{\mathbb{R}} \frac{(1-e^{ikx})\log|x|}{|x|^{2+\delta_n}}dx - \int_{\mathbb{R}} \frac{(1-e^{i(k-1)x})\log|x|}{|x|^{2+\delta_n}}dx\right| = O(\log n).$$

Proof. The result is trivial for $k = 1$. Assuming that $k > 1$ and n is large. By change of variables,

$$\begin{aligned}\left|\int_{\mathbb{R}} \frac{1-e^{ikx}}{|x|^{2+\delta_n}}dx - \int_{\mathbb{R}} \frac{1-e^{i(k-1)x}}{|x|^{2+\delta_n}}dx\right| &= \left|(k^{1+\delta_n} - (k-1)^{1+\delta_n})\int_{\mathbb{R}} \frac{1-e^{ix}}{|x|^{2+\delta_n}}dx\right| \\ &\leq \left|(1+\delta_n)k^{\delta_n}\int_{\mathbb{R}} \frac{1-e^{ix}}{|x|^{2+\delta_n}}dx\right| \\ &\leq 2(1+\delta_n)\int_{\mathbb{R}} \frac{1-e^{ix}}{|x|^{2+\delta_n}}dx = O(1),\end{aligned}$$

where the second line uses the Taylor expansion of $x^{1+\delta_n}$. Similarly,

$$\begin{aligned}
&\left|\int_{\mathbb{R}} \frac{(1-e^{ikx})\log|x|}{|x|^{2+\delta_n}}dx - \int_{\mathbb{R}} \frac{(1-e^{i(k-1)x})\log|x|}{|x|^{2+\delta_n}}dx\right| \\
&\leq \left|(k^{1+\delta_n}-(k-1)^{1+\delta_n})\int_{\mathbb{R}} \frac{(1-e^{ix})\log|x|}{|x|^{2+\delta_n}}dx\right| \\
&\quad + \left|(k^{1+\delta_n}\log k-(k-1)^{1+\delta_n}\log(k-1))\int_{\mathbb{R}} \frac{(1-e^{ix})}{|x|^{2+\delta_n}}dx\right| \\
&= O(\log n).
\end{aligned}$$

□

Lemma 3.8.3. *Let $\delta_n \in [0, 1/2)$ and $\delta_n \to 0$. Then, for all $k = 1, 2, \ldots$, we have*

$$\int_{\mathbb{R}} \frac{e^{ikx}\|1-e^{ix}\|^2}{|x|^{2+\delta_n}}dx = O\left(\frac{\delta_n}{k^{1-\delta_n}}\right).$$

Proof. By straightforward calculations,

$$\begin{aligned}
\int_{\mathbb{R}} \frac{e^{ikx}\|1-e^{ix}\|^2}{|x|^{2+\delta_n}}dx &= 2\int_{\mathbb{R}} \frac{\cos(kx)(1-\cos(x))}{|x|^{2+\delta_n}}dx \\
&= 4\int_0^\infty \frac{\cos(kx)-\frac{1}{2}[\cos((k+1)x)+\cos((k-1)x)]}{x^{2+\delta_n}}dx \\
&= 4\mathrm{Re}\left(\int_0^\infty f(x)dx\right)
\end{aligned}$$

where

$$f(z) = \frac{e^{ikz}-\frac{1}{2}[e^{i(k+1)z}+e^{i(k-1)z}]}{z^{2+\delta_n}}.$$

To evaluate this, we resort to contour integration. For $0 < r < R < \infty$, define the following four curves:

- $\gamma_1 = \{x : x \text{ from } r \text{ to } R\}$
- $\gamma_2 = \{Re^{i\theta} : \theta \text{ from } 0 \text{ to } \frac{\pi}{2}\}$
- $\gamma_3 = \{ix : x \text{ from } R \text{ to } r\}$
- $\gamma_4 = \{re^{i\theta} : \theta \text{ from } \frac{\pi}{2} \text{ to } 0\}$.

Denote by D the interior enclosed by the four curves on the complex plane. It is easy to see that $f(z)$ is a holomorphic function on $\bar{D}$, and thus the contour integral

$$\int_{\gamma_1} f(z)dz + \int_{\gamma_2} f(z)dz + \int_{\gamma_3} f(z)dz + \int_{\gamma_4} f(z)dz = 0. \tag{3.38}$$

Note that

$$\lim_{r\to 0, R\to\infty} \mathrm{Re}\left(\int_{\gamma_1} f(z)dz\right) = \mathrm{Re}\left(\int_0^\infty f(x)dx\right), \tag{3.39}$$

which is our target. On γ_2, $\mathrm{Im}(z) \geq 0$ and $\|e^{iz}\| \leq 1$. Thus,

$$\begin{aligned}
\left\|\int_{\gamma_2} f(z)dz\right\| &\leq \int_{\gamma_2} \|f(z)\|\|dz\| \\
&\leq \int_{\gamma_2} \frac{\|e^{ikz} - \frac{1}{2}[e^{i(k+1)z} + e^{i(k-1)z}]\|}{\|z^{2+\delta_n}\|}\|dz\| \\
&\leq \int_0^{\frac{\pi}{2}} \frac{2}{R^{2+\delta_n}} Rd\theta \\
&= \frac{\pi}{R^{1+\delta_n}} \to 0 \text{ as } R \to \infty.
\end{aligned} \tag{3.40}$$

Since $e^{iz} = 1 + iz - \frac{z^2}{2} + o(\|z\|^2)$, it follows that

$$\begin{aligned}
\left\|\int_{\gamma_4} f(z)dz\right\| &\leq \int_{\gamma_4} \|f(z)\|\|dz\| \\
&\leq \int_0^{\frac{\pi}{2}} \frac{r^2 + o(r^2)}{r^{2+\delta_n}} rd\theta \\
&= O(r^{1-\delta_n}) \to 0 \text{ as } r \to 0.
\end{aligned} \tag{3.41}$$

By (3.38) - (3.41), the integral in (3.39) is determined by $\int_{\gamma_3} f(z)dz$. It follows that

$$\begin{aligned}
\mathrm{Re}\left(\int_{\gamma_3} f(z)dz\right) &= \mathrm{Re}\left(i\int_R^r \frac{e^{-kx} - \frac{1}{2}[e^{-(k+1)x} + e^{-(k-1)x}]}{x^{2+\delta_n} e^{(2+\delta_n)\frac{i\pi}{2}}} dx\right) \\
&= \sin\left(\frac{\pi}{2}\delta_n\right) \int_r^R \frac{e^{-kx} - \frac{1}{2}[e^{-(k+1)x} + e^{-(k-1)x}]}{x^{2+\delta_n}} dx.
\end{aligned}$$

Thus,

$$\int_0^\infty f(x)dx = \sin\left(\frac{\pi}{2}\delta_n\right) \int_0^\infty \frac{\frac{1}{2}[e^{-(k+1)x} + e^{-(k-1)x}] - e^{-kx}}{x^{2+\delta_n}} dx.$$

As $\sin(\frac{\pi}{2}\delta_n) = O(\delta_n)$ when $\delta_n \to 0$, we only need to evaluate the integral on the right. For $x \geq 0$,

$$0 \leq e^{-kx}\left(\frac{1}{2}[e^x + e^{-x}] - 1\right) = e^{-kx}\left(\sum_{m=1}^{\infty} \frac{x^{2m}}{(2m)!}\right) \leq x^2 e^{-(k-1)x}.$$

With this we obtain a bound for the integral when $k > 1$ as

$$\begin{aligned}
&\int_0^\infty \frac{\frac{1}{2}[e^{-(k+1)x} + e^{-(k-1)x}] - e^{-kx}}{x^{2+\delta_n}} dx \\
&\leq \int_0^\infty \frac{e^{-(k-1)x}}{x^{\delta_n}} dx \\
&= \frac{1}{(k-1)^{1-\delta_n}} \int_0^\infty \frac{e^{-x}}{x^{\delta_n}} dx \\
&\leq \frac{1}{(k-1)^{1-\delta_n}} \left(\int_0^1 \frac{e^{-x}}{x^{1/2}} dx + \int_1^\infty e^{-x} dx\right) \\
&\leq \frac{C}{(k-1)^{1-\delta_n}} \\
&\leq \frac{C_2}{(k)^{1-\delta_n}},
\end{aligned}$$

where C, C_2 are some constants.

Finally, it is easy to prove that the integral is also bounded for $k = 1$ regardless of the choice of $\delta_n \in [0, 1/2]$. The proof of this lemma is complete. □

3.8.2 Proofs of Theorem 3.3.2 and Theorem 3.3.3

We begin by developing a few technical results needed for the proof. For convenience, let

$$\begin{aligned}
W_n(\boldsymbol{t}, \boldsymbol{h}) &:= \sigma^{-1} n^{H(\boldsymbol{t})} \Delta^q_{\boldsymbol{h}/n} X(\boldsymbol{t}), \\
C_n(\boldsymbol{t}, \boldsymbol{s}, \boldsymbol{h}) &:= \mathrm{Cov}\left(W_n(\boldsymbol{t}, \boldsymbol{h}), W_n(\boldsymbol{s}, \boldsymbol{h})\right).
\end{aligned}$$

Clearly,

$$C_n(\boldsymbol{t}, \boldsymbol{s}, \boldsymbol{h}) = \sigma^{-2} n^{H(\boldsymbol{t})+H(\boldsymbol{s})} \Delta^q_{\boldsymbol{h}/n,\boldsymbol{t}} \Delta^q_{\boldsymbol{h}/n,\boldsymbol{s}} C(\boldsymbol{t}, \boldsymbol{s}),$$

where $\Delta^q_{\boldsymbol{h}/n,\boldsymbol{t}}$ and $\Delta^q_{\boldsymbol{h}/n,\boldsymbol{s}}$ denote differencing with respect to $\boldsymbol{t}$ and $\boldsymbol{s}$, respectively.

The following theorem addresses the behavior of the covariance of $W_n(\boldsymbol{t}, \boldsymbol{h})$ for different ranges of the gap. There are some similarities between this theorem and Theorem

1 in Hsing et al. (2016) or Proposition 1 and 2 in Bardet and Surgailis (2013). However, the results here focus more on the mBm and provides the level of precision needed for our results. As noted before, the differencing direction $\boldsymbol{h}$ in this chapter is generally assumed to be a unit vector that parallels an axis. The following theorem is one instance where this is not assumed.

Theorem 3.8.4. *In the following let $\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{u}/n \in \Omega_\delta$. Under assumptions [A1]-[A3], for any $\delta \in (0,1)$ and $M>0$, there exist finite constants $C_\delta, C_{\delta,M}$ such that uniformly in $n, \boldsymbol{t}$,*

(i) $|C_n(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{u}/n, \boldsymbol{h}) - g(H(\boldsymbol{t}), \boldsymbol{u}, \boldsymbol{h})| \le C_{\delta,M}\rho_n(\boldsymbol{t})$ *for* $|\boldsymbol{u}| \le M$;

(ii) $|C_n(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{u}/n, \boldsymbol{h})| \le C_\delta |\boldsymbol{u}|^{-\psi(\boldsymbol{t})}$ *for $\boldsymbol{u}$ satisfying* $q|\boldsymbol{h}|+1 < |\boldsymbol{u}| < 2bn$;

(iii) $|C_n(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{u}/n, \boldsymbol{h})| \le C_\delta |\boldsymbol{u}|^{-\bar{\psi}}$ *for $\boldsymbol{u}$ satisfying* $|\boldsymbol{u}| > q|\boldsymbol{h}|+1$;

(iv) for $|\boldsymbol{u}| \to \infty$ *and* $|\boldsymbol{u}| < 2bn$,

$$C_n(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{u}/n, \boldsymbol{h}) = (c_{\boldsymbol{h}}(\boldsymbol{t}, \boldsymbol{u}) + O(b\log b))|\boldsymbol{u}|^{-\psi(\boldsymbol{t})},$$

where $c_{\boldsymbol{h}}(\boldsymbol{t}, \boldsymbol{u})$ is defined in Theorem 3.3.3.

Our approach for proving Theorems 3.3.2 and 3.3.3 starts from the following decomposition:

$$\begin{aligned} G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) &= \sum_i 2 s_{\boldsymbol{t},p,b}(\boldsymbol{t}_i) \log \left| \Delta^q_{\boldsymbol{h}/n} X(\boldsymbol{t}_i) \right| \\ &= I_1(\boldsymbol{t}, n, b) + \widetilde{\beta}_1(\boldsymbol{t}, \boldsymbol{h}; n, b) + \log \sigma^2, \end{aligned} \tag{3.42}$$

where

$$I_1(\boldsymbol{t}, n, b) = -2\log n \sum_i s_{\boldsymbol{t},p,b}(\boldsymbol{t}_i) H(\boldsymbol{t}_i)$$

and

$$\widetilde{\beta}_1(\boldsymbol{t}, \boldsymbol{h}; n, b) = 2\sum_i s_{\boldsymbol{t},p,b}(\boldsymbol{t}_i) \log |W_n(\boldsymbol{t}_i, \boldsymbol{h})|.$$

Denote

$$\widetilde{g}(H, \boldsymbol{h}) = \log(g(H, \boldsymbol{0}, \boldsymbol{h})) + \mathbb{E}\log \chi^2_1. \tag{3.43}$$

It is easy to see if $H(\boldsymbol{t})$ bounded away from 1 and $H(\boldsymbol{t}) \in \mathscr{H}_p((0,1)^d)$, then $\widetilde{g}(H(\boldsymbol{t}), \boldsymbol{h}) \in \mathscr{H}_p((0,1)^d)$. Clearly, $I_1(\boldsymbol{t}, n, b)$ approximates $-2(\log n)H(\boldsymbol{t})$. Similarly, the following lemma suggests that $\widetilde{\beta}_1(\boldsymbol{t}, \boldsymbol{h}; n, b)$ could be a reasonable estimator of $\widetilde{g}(H(\boldsymbol{t}), \boldsymbol{h})$.

Lemma 3.8.5. *Under assumption [A3], we have*

$$\mathbb{E}[2\log|W_n(\boldsymbol{t}, \boldsymbol{h})|] = \widetilde{g}(H(\boldsymbol{t}), \boldsymbol{h}) + O(\rho_n(\boldsymbol{t})).$$

Thus, the asymptotic properties of $G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})$ can be understood by conducting a careful analysis of $\widetilde{\beta}_1(\boldsymbol{t}, \boldsymbol{h}; n, b)$ and $I_1(\boldsymbol{t}, n, b)$, where both $I_1(\boldsymbol{t}, n, b)$ and $\widetilde{\beta}_1(\boldsymbol{t}, \boldsymbol{h}; n, b)$ contribute to the asymptotic bias whereas $\widetilde{\beta}_1(\boldsymbol{t}, \boldsymbol{h}; n, b)$ determines the asymptotic variance and asymptotic distribution. This will be done in the following two lemmas whose proofs along with the proof of Lemma 3.8.5 are given in Section 3.8.6.

Lemma 3.8.6. *Assume that [K], [A1], [A2] and [A3] hold. we have, uniformly for all* $\boldsymbol{t} \in [0,1]^d$*,*

$$I_1(\boldsymbol{t}, n, b) = -2\log(n)(H(\boldsymbol{t}) + R(\boldsymbol{t}, b)) + O\left(\frac{\log(n)}{(nb)^{2\wedge p}}\right)$$

and

$$R(\boldsymbol{t}, b) = O(b^p).$$

Furthermore, if $p \in \mathbb{N}$*, we have*

$$\begin{aligned} R(\boldsymbol{t}, b) =& \mathbf{e}_1^T \left(\int_{D_{\boldsymbol{t},b}} K(\mathbf{z})\mathbf{A}(\mathbf{z})\mathbf{A}^T(\mathbf{z})d\mathbf{z} \right)^{-1} \times \\ & \int_{D_{\boldsymbol{t},b}} K(\mathbf{z})\mathbf{A}(\mathbf{z}) \left(\sum_{|\boldsymbol{\alpha}|=p} R_{\boldsymbol{\alpha}}(\boldsymbol{t})(b\mathbf{z})^{\boldsymbol{\alpha}} \right) d\mathbf{z}, \end{aligned}$$

where $\mathbf{e}_1 = (1, 0, \ldots, 0)^T$*,* $\mathbf{z}^{\boldsymbol{\alpha}}$ *is as defined in (3.7),*

$$D_{\boldsymbol{t},b} = \{\mathbf{z} : \boldsymbol{t} + b\mathbf{z} \in \Omega_\delta\} \cap [0,1]^d,$$

and

$$R_{\boldsymbol{\alpha}}(\boldsymbol{t}) = \frac{|\boldsymbol{\alpha}|}{\boldsymbol{\alpha}!} \int_0^1 (1-s)^{|\boldsymbol{\alpha}|-1} D^{\boldsymbol{\alpha}} H(\boldsymbol{t} + sb\mathbf{z})ds.$$

Lemma 3.8.7. *For $\boldsymbol{t} \in (0,1)^d$, define*

$$\begin{aligned}&\mathbf{Z}(\boldsymbol{t}, n)\\ &:= (nb)^{-\frac{d}{2}} \left\{ \sum_i K\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) \left(2\log|W_n(\boldsymbol{t}_i, \boldsymbol{h})| - \widetilde{g}(H(\boldsymbol{t}_i), \boldsymbol{h})\right) \mathbf{A}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) \right\}.\end{aligned}$$

Assume that [A1], [A2] and [A3] hold. If $2\psi(\boldsymbol{t}) > d$, then

$$\mathbf{Z}(\boldsymbol{t}, n) \xrightarrow{d} N\left(\mathbf{0}, \sigma^2_{H(\boldsymbol{t})} \int_{\mathbb{R}^d} \mathbf{f}(\mathbf{z})\mathbf{f}^T(\mathbf{z}) d\mathbf{z}\right),$$

where $\mathbf{f}(\mathbf{z}) = K(\mathbf{z})\mathbf{A}(\mathbf{z})$ and $\sigma^2_{H(\boldsymbol{t})}$ is defined by (3.18); if $2\psi(\boldsymbol{t}) = d$, then

$$\log^{-\frac{1}{2}}(nb)\mathbf{Z}(\boldsymbol{t}, n) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\Sigma} = \frac{V_d \mathbf{f}(\mathbf{0})\mathbf{f}^T(\mathbf{0})}{g^2(H(\boldsymbol{t}, \mathbf{0}, \boldsymbol{h}))} \int_{B_1(\mathbf{0})} \frac{c^2_{\boldsymbol{h}}(\boldsymbol{t}, \boldsymbol{x})}{|\boldsymbol{x}|^{d-1}} d\boldsymbol{x};$$

if $2\psi(\boldsymbol{t}) < d$, then the characteristic function for $(nb)^{-\frac{d}{2}+\psi(\boldsymbol{t})}\mathbf{a}^T\mathbf{Z}(\boldsymbol{t}, n)$ converges to

$$\phi(u) = \exp\left\{\frac{1}{2}\sum_{k=2}^{\infty} \frac{(2iu)^k}{k} S_k\right\},$$

where

$$S_k = \int_{(B_1(\mathbf{0}))^{\otimes k}} \frac{\prod_{i=1}^k c_{\boldsymbol{h}}(\boldsymbol{t}, \mathbf{z}_i - \mathbf{z}_{i+1})|\mathbf{z}_i - \mathbf{z}_{i+1}|^{-\psi(\boldsymbol{t})} f_{\boldsymbol{a}}(\mathbf{z}_i)}{g^k(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h})} d\mathbf{z}_1 \ldots d\mathbf{z}_k,$$

in which $\mathbf{z}_{k+1} := \mathbf{z}_1$ and $f_{\boldsymbol{a}} = \boldsymbol{a}^T\mathbf{f}$. In addition, for each $\delta \in (0,1)$ and $\mathbf{a} \in \mathbb{R}^d$, there exist constants C_δ and $C_{\delta,\boldsymbol{a}}$ such that all $\boldsymbol{t} \in \Omega_\delta$,

$$|\mathbb{E}[\mathbf{Z}(\boldsymbol{t}, n)]| \leq C_\delta (nb)^{\frac{d}{2}} \rho_n(\boldsymbol{t}),$$

and

$$\mathrm{Var}(\mathbf{a}^T\mathbf{Z}(\boldsymbol{t}, n)) \leq \begin{cases} C_{\delta,\mathbf{a}} & \text{if } 2\psi(\boldsymbol{t}) > d, \\ C_{\delta,\mathbf{a}} \log(nb), & \text{if } 2\psi(\boldsymbol{t}) = d, \\ C_{\delta,\mathbf{a}} (nb)^{d-2\psi(\boldsymbol{t})}, & \text{if } 2\psi(\boldsymbol{t}) < d. \end{cases}$$

PROOFS FOR THEOREMS 3.3.2 AND 3.3.3:

To simplify notation, write $\widetilde{g}(\boldsymbol{t}) := \widetilde{g}(H(\boldsymbol{t}), \boldsymbol{h})$, where $\widetilde{g}(H(\boldsymbol{t}), \boldsymbol{h})$ is defined by (3.43), and

$$\widetilde{\boldsymbol{\beta}} := \underset{\beta_{\boldsymbol{j}_1},\ldots,\beta_{\boldsymbol{j}_S}}{\operatorname{argmin}} \sum_i K\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) \left\{ 2\log|W_n(\boldsymbol{t}_i, \boldsymbol{h})| - \sum_{m=1}^{S} \beta_{\boldsymbol{j}_m} \left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)^{\boldsymbol{j}_m} \right\}^2.$$

Thus we have,

$$\sum_i K\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) \left\{ 2\log|W_n(\boldsymbol{t}_i, \boldsymbol{h})| - \widetilde{\boldsymbol{\beta}}^T \mathbf{A}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) \right\} \mathbf{A}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) = 0.$$

Plug this equation into the definition of $\mathbf{Z}(\boldsymbol{t}, n)$, basically replacing $2\log|W_n(\boldsymbol{t}_i, \boldsymbol{h})|$ by $\widetilde{\boldsymbol{\beta}}^T \mathbf{A}$ there, and we obtain

$$\mathbf{Z}(\boldsymbol{t}, n) = (nb)^{-\frac{d}{2}} \left\{ \sum_i K\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) \left(\widetilde{\boldsymbol{\beta}}^T \mathbf{A}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) - \widetilde{g}(\boldsymbol{t}_i) \right) \mathbf{A}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) \right\}.$$

If $p \in \mathbb{N}$, denote the Taylor expansion of $\widetilde{g}(\boldsymbol{t} + b\boldsymbol{\nu})$ at $\boldsymbol{t}$ up to degree $p-1$ by $\breve{g}(\boldsymbol{t}, b\boldsymbol{\nu})$. Then we have

$$\breve{g}(\boldsymbol{t}, b\boldsymbol{\nu}) - \widetilde{g}(\boldsymbol{t} + b\boldsymbol{\nu}) = -\frac{b^p}{p!} \sum_{|\boldsymbol{i}|=p} [\boldsymbol{\nu}^{\boldsymbol{i}} D^{\boldsymbol{i}} \widetilde{g}(\boldsymbol{t})] + o((b\boldsymbol{\nu})^p).$$

Define

$$\mathbf{M}_n(b, \boldsymbol{t}) = (nb)^{-d} \sum_{\boldsymbol{\nu}=\frac{\boldsymbol{i}}{nb}, \boldsymbol{i}\in\Omega_{\boldsymbol{t}}} \mathbf{A}(\boldsymbol{\nu}) \mathbf{A}(\boldsymbol{\nu})^T K(\boldsymbol{\nu}),$$

where $\Omega_{\boldsymbol{t}} = \{\boldsymbol{i} : \boldsymbol{i} \in \mathbb{Z}^d, \boldsymbol{t} + \boldsymbol{i}/n \in (0,1)^d\}$. On one hand,

$$\begin{aligned}
&(nb)^{-d} \sum_{\boldsymbol{\nu}=\frac{\boldsymbol{i}}{nb}, \boldsymbol{i}\in\Omega_{\boldsymbol{t}}} \mathbf{A}(\boldsymbol{\nu}) K(\boldsymbol{\nu}) \left(\widetilde{\boldsymbol{\beta}}^T \mathbf{A}(\boldsymbol{\nu}) - \breve{g}(\boldsymbol{t}, b\boldsymbol{\nu}) \right) \\
&= (nb)^{-d} \sum_{\boldsymbol{\nu}=\frac{\boldsymbol{i}}{nb}, \boldsymbol{i}\in\Omega_{\boldsymbol{t}}} \mathbf{A}(\boldsymbol{\nu}) \mathbf{A}(\boldsymbol{\nu})^T K(\boldsymbol{\nu}) \begin{pmatrix} (\widetilde{\beta}_1 - \frac{b^{|\boldsymbol{j}_1|}}{\boldsymbol{j}_1!} \frac{\partial^{\boldsymbol{j}_1} \widetilde{g}(\boldsymbol{t})}{\partial \boldsymbol{t}^{\boldsymbol{j}_1}}) \\ \vdots \\ (\widetilde{\beta}_S - \frac{b^{|\boldsymbol{j}_S|}}{\boldsymbol{j}_S!} \frac{\partial^{\boldsymbol{j}_S} \widetilde{g}(\boldsymbol{t})}{\partial \boldsymbol{t}^{\boldsymbol{j}_S}}) \end{pmatrix} \\
&= \mathbf{M}_n(b, \boldsymbol{t}) \begin{pmatrix} (\widetilde{\beta}_1 - \frac{b^{|\boldsymbol{j}_1|}}{\boldsymbol{j}_1!} \frac{\partial^{\boldsymbol{j}_1} \widetilde{g}(\boldsymbol{t})}{\partial \boldsymbol{t}^{\boldsymbol{j}_1}}) \\ \vdots \\ (\widetilde{\beta}_S - \frac{b^{|\boldsymbol{j}_S|}}{\boldsymbol{j}_S!} \frac{\partial^{\boldsymbol{j}_S} \widetilde{g}(\boldsymbol{t})}{\partial \boldsymbol{t}^{\boldsymbol{j}_S}}) \end{pmatrix}.
\end{aligned}$$

On the other hand,

$$\begin{aligned}&(nb)^{-d}\sum_{\boldsymbol{\nu}=\frac{\boldsymbol{i}}{nb},\boldsymbol{i}\in\Omega_{\boldsymbol{t}}}\mathbf{A}(\boldsymbol{\nu})K(\boldsymbol{\nu})\left(\widetilde{\boldsymbol{\beta}}^T\mathbf{A}(\boldsymbol{\nu})-\breve{g}(\boldsymbol{t},b\boldsymbol{\nu})\right)\\&=(nb)^{-\frac{d}{2}}\mathbf{Z}(\boldsymbol{t},n)+B(\boldsymbol{t},b)+o\left(b^p\right)\end{aligned}$$

where

$$B(\boldsymbol{t},b)=-\frac{b^p}{p!}\sum_{|\boldsymbol{i}|=p}\left[(nb)^{-d}\sum_{\boldsymbol{\nu}\in\{\frac{\boldsymbol{j}}{nb},\boldsymbol{j}\in\Omega_{\boldsymbol{t}}\}}\mathbf{A}(\boldsymbol{\nu})K(\boldsymbol{\nu})\boldsymbol{\nu}^{\boldsymbol{i}}D^{\boldsymbol{i}}\widetilde{g}(\boldsymbol{t})\right]=O(b^p).$$

If $p\notin\mathbb{N}$, denote the Taylor expansion of $\widetilde{g}(\boldsymbol{t}+b\boldsymbol{\nu})$ at $\boldsymbol{t}$ up to degree $\lfloor p\rfloor$ by $\breve{g}(\boldsymbol{t},b\boldsymbol{\nu})$. Then we have

$$\breve{g}(\boldsymbol{t},b\boldsymbol{\nu})-\widetilde{g}(\boldsymbol{t}+b\boldsymbol{\nu})=O((b\boldsymbol{\nu})^p).$$

From this, we can still get $B(\boldsymbol{t},b)=O(b^p)$, although an explicit form is not available. Thus

$$\begin{pmatrix}(\widetilde{\beta}_1-\frac{b^{|\boldsymbol{j}_1|}}{\boldsymbol{j}_1!}\frac{\partial^{\boldsymbol{j}_1}\widetilde{g}(\boldsymbol{t})}{\partial\boldsymbol{t}^{\boldsymbol{j}_1}})\\\vdots\\(\widetilde{\beta}_S-\frac{b^{|\boldsymbol{j}_S|}}{\boldsymbol{j}_S!}\frac{\partial^{\boldsymbol{j}_S}\widetilde{g}(\boldsymbol{t})}{\partial\boldsymbol{t}^{\boldsymbol{j}_S}})\end{pmatrix}=\mathbf{M}_n(b,\boldsymbol{t})^{-1}\left((nb)^{-\frac{d}{2}}\mathbf{Z}(\boldsymbol{t},n)+B(\boldsymbol{t},b)+o\left(b^p\right)\right).$$

It is easily shown that

$$\mathbf{M}_n(b,\boldsymbol{t})\to\mathbf{M}:=\int_{B_1(\mathbf{0})}\mathbf{A}(\boldsymbol{\nu})\mathbf{A}(\boldsymbol{\nu})^TK(\boldsymbol{\nu})d\boldsymbol{\nu}$$

uniformly. Therefore,

$$\widetilde{\beta}_1=\widetilde{g}(\boldsymbol{t})+(1+o(1))\mathbf{e}_1^T\mathbf{M}^{-1}\left((nb)^{-\frac{d}{2}}\mathbf{Z}(\boldsymbol{t},n)+B(\boldsymbol{t},b)\right).\tag{3.44}$$

Combining this with equation (3.42) and Lemma 3.8.6, we obtain

$$\begin{aligned}
&G(\widehat{H(\boldsymbol{t});n},\boldsymbol{h}) \\
&= \widetilde{g}(H(\boldsymbol{t}),\boldsymbol{h}) - 2\log(n)H(\boldsymbol{t}) + \log\sigma^2 - 2\log(n)R(\boldsymbol{t},b) + O\left(\frac{\log(n)}{(nb)^{2\wedge p}}\right) \\
&\quad + \mathbf{e}_1^T\mathbf{M}^{-1}\left((nb)^{-\frac{d}{2}}\mathbf{Z}(\boldsymbol{t},n) + B(\boldsymbol{t},b)\right)(1+o(1)) \\
&= G(H(\boldsymbol{t});n,\boldsymbol{h}) - 2\log(n)R(\boldsymbol{t},b) + O\left(\frac{\log(n)}{(nb)^{2\wedge p}}\right) \\
&\quad + \mathbf{e}_1^T\mathbf{M}^{-1}\left((nb)^{-\frac{d}{2}}\mathbf{Z}(\boldsymbol{t},n) + B(\boldsymbol{t},b)\right)(1+o(1)),
\end{aligned} \tag{3.45}$$

where the last line uses the definitions of G and $\widetilde{g}$ in equations (3.5) and (3.43). As such, the bias of $G(\widehat{H(\boldsymbol{t});n},\boldsymbol{h})$ is

$$\begin{aligned}
&\text{Bias}\left(G(\widehat{H(\boldsymbol{t});n},\boldsymbol{h})\right) \\
&= O\left(\log(n)R(\boldsymbol{t},b) + \frac{\log(n)}{(nb)^{2\wedge p}} + (nb)^{-\frac{d}{2}}|\mathbb{E}\,\mathbf{Z}(\boldsymbol{t},n)| + \mathbf{e}_1^T\mathbf{M}^{-1}B(\boldsymbol{t},b)\right) \\
&= O\left(\log(n)\left(b^p + (nb)^{-2\wedge p}\right) + \rho_n(\boldsymbol{t})\right) = O(T_1(n,b,\boldsymbol{t})),
\end{aligned} \tag{3.46}$$

where the last line use facts that $R(\boldsymbol{t},b), B(\boldsymbol{t},b) = O(b^p)$ and $|\mathbb{E}\,\mathbf{Z}(\boldsymbol{t},n)| = O\left((nb)^{\frac{d}{2}}\rho_n(\boldsymbol{t})\right)$ in Lemma 3.8.7. Similarly,

$$\begin{aligned}
\text{Var}\left(G(\widehat{H(\boldsymbol{t});n},\boldsymbol{h})\right) &= O\left((nb)^{-d}\text{Var}\left(\mathbf{e}_1^T\mathbf{M}^{-1}\mathbf{Z}(\boldsymbol{t},n)\right)\right) \\
&= O(T_2(n,b,\boldsymbol{t})),
\end{aligned} \tag{3.47}$$

where the last step uses the result on the variance of $\mathbf{Z}$ in Lemma 3.8.7. The two bounds, (3.46) and (3.47), complete the proof for Theorem 3.3.2.

Finally, in view of (3.45), $G(\widehat{H(\boldsymbol{t});n},\boldsymbol{h}) - \mathbb{E}\,G(\widehat{H(\boldsymbol{t});n},\boldsymbol{h})$ has the same asymptotic distribution as that of $\mathbf{e}_1^T\mathbf{M}^{-1}\left(\mathbf{Z}(\boldsymbol{t},n) - \mathbb{E}\,\mathbf{Z}(\boldsymbol{t},n)\right)$. Applying Lemma 3.8.7, it is straightforward to obtain those asymptotic distributions listed in Theorem 3.3.3. This completes the proofs of Theorem 3.3.2 and Theorem 3.3.3.

□

3.8.3 Proofs of Theorem 3.3.4 and Corollary 3.3.5

PROOF for Theorem 3.3.4:

Let γ be the constant defined in (3.22). It follows that

$$G'(H; n, \boldsymbol{h}) = -2\log n + O(1) \quad \text{and} \quad G''(H; n, \boldsymbol{h}) = O(1)$$

for $H \in [\gamma/2, 1-\gamma/2]$. Below, when there is no ambiguity, we denote $G(\cdot; n, \boldsymbol{h})$ by G and the inverse by $G^{-1}(\cdot)$ to simplify notation. It is easy to verify that

$$\frac{\partial}{\partial \xi} G^{-1}(\xi) = \frac{1}{G'(G^{-1}(\xi))} = \frac{1}{G'(H)} = -\frac{1}{2\log n} + O\left(\frac{1}{\log^2 n}\right), \tag{3.48}$$

and

$$\frac{\partial^2}{\partial \xi^2} G^{-1}(\xi) = \left(\frac{1}{G'(G^{-1}(\xi))}\right)' = -\frac{G''(H)}{(G'(H))^3} = O\left(\frac{1}{\log^3 n}\right) \tag{3.49}$$

where $H = G^{-1}(\xi)$. Let A to be the event on which

$$G(1-\gamma/2; n, \boldsymbol{h}) \le G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) \le G(\gamma/2; n, \boldsymbol{h}).$$

On A, by (3.49) and Taylor's expansion, we have

$$\begin{aligned}
&\widehat{H}(\boldsymbol{t}) - H(\boldsymbol{t}) \\
&= G^{-1}\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})\right) - G^{-1}\left(G(H(\boldsymbol{t}); n, \boldsymbol{h})\right) \\
&= G^{-1}(G(H(\boldsymbol{t}); n, \boldsymbol{h}))' \left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - G(H(\boldsymbol{t}); n, \boldsymbol{h})\right) \\
&\quad + O\left(\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - G(H(\boldsymbol{t}); n, \boldsymbol{h})\right)^2 / \log^3 n\right).
\end{aligned} \tag{3.50}$$

For convenience of notation, define

$$V_n = \mathbb{E}\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - G(H(\boldsymbol{t}); n, \boldsymbol{h})\right)^2.$$

For n large enough, we have

$$\begin{aligned}
\mathbb{P}(A^c) &\le \mathbb{P}\left(\left|G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - G(H(\boldsymbol{t}); n, \boldsymbol{h})\right| \ge \frac{\gamma \log n}{2}\right) \\
&\le \left(\frac{\gamma \log n}{2}\right)^{-2} V_n.
\end{aligned} \tag{3.51}$$

By (3.48) and (3.50),

$$\begin{aligned}
&\mathbb{E}\left[\left(\widehat{H}(\boldsymbol{t})-H(\boldsymbol{t})\right)I_A\right]\\
&=G^{-1}(G(H(\boldsymbol{t});n,\boldsymbol{h}))'\mathbb{E}\left[\left(G(\widehat{H(\boldsymbol{t})};n,\boldsymbol{h})-G(H(\boldsymbol{t});n,\boldsymbol{h})\right)I_A\right]\\
&\quad+O(V_n/\log^3 n)\\
&=-(2\log n)^{-1}\mathbb{E}\left[G(\widehat{H(\boldsymbol{t})};n,\boldsymbol{h})-G(H(\boldsymbol{t});n,\boldsymbol{h})\right]+R_n\\
&\quad+O(T_1(n,b,\boldsymbol{t})/\log^2 n)+O(V_n/\log^3 n),
\end{aligned}\tag{3.52}$$

where, by the Cauchy-Schwarz inequality and (3.51),

$$\begin{aligned}
|R_n|&=O\left((\log n)^{-1}\right)\left|\mathbb{E}\left[\left(G(\widehat{H(\boldsymbol{t})};n,\boldsymbol{h})-G(H(\boldsymbol{t});n,\boldsymbol{h})\right)I_{A^c}\right]\right|\\
&\le O\left((\log n)^{-1}\right)\mathbb{E}^{1/2}\left[G(\widehat{H(\boldsymbol{t})};n,\boldsymbol{h})-G(H(\boldsymbol{t});n,\boldsymbol{h})\right]^2\mathbb{P}^{1/2}(A^c)\\
&=O(V_n/\log^2 n).
\end{aligned}\tag{3.53}$$

On the other hand, since both $\widehat{H}(\boldsymbol{t})$ and $H(\boldsymbol{t})\in[0,1]$,

$$\left|\mathbb{E}\left[\left(\widehat{H}(\boldsymbol{t})-H(\boldsymbol{t})\right)I_{A^c}\right]\right|\le 2\mathbb{P}(A^c)=O(V_n/\log^2 n).\tag{3.54}$$

By (3.52)-(3.54),

$$\begin{aligned}
\mathbb{E}\left[\widehat{H}(\boldsymbol{t})-H(\boldsymbol{t})\right]=&-(2\log n)^{-1}\mathbb{E}\left[G(\widehat{H(\boldsymbol{t})};n,\boldsymbol{h})-G(H(\boldsymbol{t});n,\boldsymbol{h})\right]\\
&+O((T_1(n,b,\boldsymbol{t})+V_n)/\log^2 n).
\end{aligned}$$

A similar argument gives

$$\mathbb{E}\left[\widehat{H}(\boldsymbol{t})-H(\boldsymbol{t})\right]^2=O(V_n/\log^2 n).$$

Finally, we turn to the asymptotic distributions. By Theorem 3.3.2,

$$\mathbb{E}G(\widehat{H(\boldsymbol{t})};n,\boldsymbol{h})-G(H(\boldsymbol{t});n,\boldsymbol{h})=o(\log n).$$

By this and the fact that $G(b;n,\boldsymbol{h})-G(a;n,\boldsymbol{h})=O(\log n)$ for any fixed $a<b$ in $[\gamma/2,1-\gamma/2]$, we conclude readily that

$$\mathbb{E}G(\widehat{H(\boldsymbol{t})};n,\boldsymbol{h})\in[G(1-\gamma/2;n,\boldsymbol{h}),G(\gamma/2;n,\boldsymbol{h})]$$

for large n. By the mean value theorem and (3.48), we have on the event A

$$\begin{aligned}
&\left(\widehat{H}(\boldsymbol{t}) - G^{-1}\left(\mathbb{E}G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})\right)\right)\log n \\
&= \left(G^{-1}\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})\right) - G^{-1}\left(\mathbb{E}G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})\right)\right)\log n \\
&= G^{-1}(\widetilde{\theta})'\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - \mathbb{E}G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})\right)\log n \\
&\sim -\frac{1}{2}\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - \mathbb{E}G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})\right).
\end{aligned} \tag{3.55}$$

The result on asymptotic distribution stated in the Theorem 3.3.4 is then a simple consequence of Slutsky's Theorem.

□

PROOF for Corollary 3.3.5:

The asymptotic distribution of $G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - G(H(\boldsymbol{t}))$ in Corollary 3.3.5 follows from Theorem 3.3.3 and the asymptotic bias in (3.16). Similarly to (3.55),

$$\begin{aligned}
&\left(G^{-1}\left(\mathbb{E}G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})\right) - H(\boldsymbol{t})\right)\log n \\
&= \left(G^{-1}\left(\mathbb{E}G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})\right) - G^{-1}\left(G(H(\boldsymbol{t}))\right)\right)\log n \\
&= G^{-1}(\widetilde{\theta})'\left(\mathbb{E}G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - G(H(\boldsymbol{t}))\right)\log n \\
&\sim -\frac{1}{2}\left(\mathbb{E}G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - G(H(\boldsymbol{t}))\right).
\end{aligned}$$

Thus, the asymptotic distribution of $(\widehat{H}(\boldsymbol{t}) - H(\boldsymbol{t}))\log n$ in Corollary 3.3.5 follows as for $G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) - G(H(\boldsymbol{t}))$ using (3.16) and (3.55).

□

3.8.4 Proofs for Theorem 3.4.1, Theorem 3.4.2 and Theorem 3.4.3

PROOF for Theorem 3.4.1:

The result for variance is a direct consequence of Theorem 3.3.2. For bias, by (3.42) and the fact that $I_1(\boldsymbol{t}, n, b)$ will be canceled in $\mathbb{E}[\widehat{H}_1]$, we have

$$\begin{aligned}
(2\log 2)\mathbb{E}[\widehat{H}_1] &= \mathbb{E}[\widetilde{\beta}_1(\boldsymbol{t}, 2\boldsymbol{h}; n, b_1) - \widetilde{\beta}_1(\boldsymbol{t}, \boldsymbol{h}; n, b_1)] \\
&= \log g(\boldsymbol{t}, \boldsymbol{0}, 2\boldsymbol{h}) - \log g(\boldsymbol{t}, \boldsymbol{0}, \boldsymbol{h}) + O(b^p + \rho_n(\boldsymbol{t})) \\
&= (2\log 2)H(\boldsymbol{t}) + O(b^p + \rho_n(\boldsymbol{t})),
\end{aligned}$$

where the second line uses (3.44).

The following is a sketch for the proof of the asymptotic distribution. We will omit the details here since they are straightforward. From the proof of Theorem 3.3.3, under proper normalization $a(n,b,\boldsymbol{t}) = (nb)^{-d/2}$, we have

$$a(n,b,\boldsymbol{t})\left(\widetilde{\beta}_1(\boldsymbol{t},2\boldsymbol{h};n,b) - \mathbb{E}[\widetilde{\beta}_1(\boldsymbol{t},2\boldsymbol{h};n,b)]\right) \xrightarrow{d} N\left(0,\xi^2(\boldsymbol{t})\right).$$

By similar derivations, we can actually conclude that the vector

$$a(n,b,\boldsymbol{t})\left(\widetilde{\beta}_1(\boldsymbol{t},\boldsymbol{h};n,b) - \mathbb{E}[\widetilde{\beta}_1(\boldsymbol{t},\boldsymbol{h};n,b)], \widetilde{\beta}_1(\boldsymbol{t},\boldsymbol{h};n,2b) - \mathbb{E}[\widetilde{\beta}_1(\boldsymbol{t},\boldsymbol{h};n,2b)]\right)$$

converges to a bivariate Gaussian distribution. Thus, the asymptotic variance of the difference of the components of this random vector must be in $[(2^{\frac{d}{2}}-1)^2\xi^2(\boldsymbol{t}), (2^{\frac{d}{2}}+1)^2\xi^2(\boldsymbol{t})]$.

□

PROOF for Theorem 3.4.2:

To simplify our notation in this proof, let C_δ denote a constant depending on δ which may vary from line to line.

By definitions, we have

$$\begin{aligned}\widehat{\log\sigma^2} =& \frac{1}{\#(\Omega_m\cap\Omega_\delta)}\sum_{\boldsymbol{t}\in\Omega_m\cap\Omega_\delta}\Bigg(G(\widehat{H(\boldsymbol{t})};n,\boldsymbol{h}) + 2\widehat{H}_1(\boldsymbol{t})\log n \\ & - \log g(\widehat{H}_1^\gamma(\boldsymbol{t}),\boldsymbol{0},\boldsymbol{h}) - \mathbb{E}[\log\chi_1^2]\Bigg),\end{aligned}$$

and

$$\log\sigma^2 = G(H(\boldsymbol{t});n,\boldsymbol{h}) + 2H(\boldsymbol{t})\log n - \log g(H(\boldsymbol{t}),\boldsymbol{0},\boldsymbol{h}) - \mathbb{E}[\log\chi_1^2].$$

Thus,

$$\begin{aligned}
&\widehat{\log\sigma^2} - \log\sigma^2 \\
=&\frac{1}{\#(\Omega_m \cap \Omega_\delta)} \sum_{\boldsymbol{t}\in\Omega_m\cap\Omega_\delta} \Bigg(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}) + 2\widehat{H}_1(\boldsymbol{t}) \log n \\
&- G(H(\boldsymbol{t}); n, \boldsymbol{h}) - 2H(\boldsymbol{t}) \log n \Bigg) \qquad (3.56) \\
&- \frac{1}{\#(\Omega_m \cap \Omega_\delta)} \sum_{\boldsymbol{t}\in\Omega_m\cap\Omega_\delta} \Bigg(\log g(\widehat{H}_1^{\gamma}(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h}) - \log g(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h}) \Bigg) \\
:=&I_1' + I_2'.
\end{aligned}$$

It is easy to see that

$$\mathbb{E}\left(\widehat{\log\sigma^2} - \log\sigma^2\right)^2 \le 2\mathbb{E}(I_1')^2 + 2\mathbb{E}(I_2')^2. \qquad (3.57)$$

We consider $\mathbb{E}(I_1')^2$ and $\mathbb{E}(I_2')^2$ separately.

Lemma 3.8.8. *Under the condition of Theorem 3.4.2, we have*

$$\mathbb{E}[I_1'] = O(T_1'(n, b_1)).$$

If $m \le 1/(3b_1)$,

$$\mathrm{Var}(I_1') \le C_\delta \log^2(n) \left(m^{-d} T_2'(n, b_1) + m^{2\bar{\psi}-2d} n^{-2\bar{\psi}} \sum_{l=1}^{m} l^{-2\bar{\psi}+d-1} \right);$$

if $m > 1/(3b_1)$,

$$\mathrm{Var}(I_1') \le C_\delta \log^2(n) \left(b_1^d T_2'(n, b_1) \sum_{l=1}^{2nb_1} l^{-2\bar{\psi}+d-1} + m^{2\bar{\psi}-2d} n^{-2\bar{\psi}} \sum_{l=1}^{m} l^{-2\bar{\psi}+d-1} \right).$$

Proof. The conclusion about bias is a direct result of Theorem 3.4.1 and Theorem 3.3.2.

From part (iii) in Theorem 3.8.4 and Lemma 3.8.10, uniformly for $\boldsymbol{t}, \boldsymbol{s} \in \Omega_\delta$ and $n|\boldsymbol{t} - \boldsymbol{s}| > q|\boldsymbol{h}| + 1$,

$$|r_n(\boldsymbol{s}, \boldsymbol{t})| \le n^{-2q+2\bar{H}} C_\delta |\boldsymbol{s} - \boldsymbol{t}|^{-2q}. \qquad (3.58)$$

Then we have

$$\begin{aligned}
&\mathrm{Var}\left(I_1'\right)\\
&=\mathrm{Var}\Bigg(\frac{1}{\#(\Omega_m\cap\Omega_\delta)}\sum_{\boldsymbol{t}\in\Omega_m\cap\Omega_\delta}\Bigg(G(\widehat{H(\boldsymbol{t});n},\boldsymbol{h})\\
&\quad+\frac{\log n}{2\log 2}\left(G(\widehat{H(\boldsymbol{t});n/2},\boldsymbol{h})-G(\widehat{H(\boldsymbol{t});n},\boldsymbol{h})\right)\Bigg)\Bigg)\\
&\le C_\delta\mathrm{Var}\left(\frac{\log n}{\#(\Omega_m\cap\Omega_\delta)}\sum_{\boldsymbol{t}\in\Omega_m\cap\Omega_\delta}G(\widehat{H(\boldsymbol{t});n},\boldsymbol{h})\right)\\
&\le C_\delta\log^2(n)m^{-2d}\mathrm{Var}\left(\sum_{\boldsymbol{t}\in\Omega_m\cap\Omega_\delta}G(\widehat{H(\boldsymbol{t});n},\boldsymbol{h})\right).
\end{aligned}\tag{3.59}$$

To calculate $\mathrm{Var}\left(\sum_{\boldsymbol{t}\in\Omega_m\cap\Omega_\delta}G(\widehat{H(\boldsymbol{t});n},\boldsymbol{h})\right)$, we need to do a further decomposition of $\mathrm{Cov}\left(G(\widehat{H(\boldsymbol{t});n},\boldsymbol{h}),G(\widehat{H(\boldsymbol{t}');n},\boldsymbol{h})\right)$ as

$$\begin{aligned}
&\left|\mathrm{Cov}\left(G(\widehat{H(\boldsymbol{t});n},\boldsymbol{h}),G(\widehat{H(\boldsymbol{t}');n},\boldsymbol{h})\right)\right|\\
&=\left|\mathrm{Cov}\left(\widetilde{\beta}_1(\boldsymbol{t},\boldsymbol{h};n,b_1),\widetilde{\beta}_1(\boldsymbol{t}',\boldsymbol{h};n,b_1)\right)\right|\\
&=\Bigg|\mathrm{Cov}\Bigg(\sum_{\boldsymbol{t_i}\in\Omega_\delta}2S_{\boldsymbol{t},p,b_1}(\boldsymbol{t_i})\log|W_n(\boldsymbol{t_i},\boldsymbol{h})|,\\
&\qquad\qquad\sum_{\boldsymbol{t_i}\in\Omega_\delta}2S_{\boldsymbol{t}',p,b_1}(\boldsymbol{t_i})\log|W_n(\boldsymbol{t_i},\boldsymbol{h})|\Bigg)\Bigg|\\
&\le C_\delta(nb_1)^{-2d}\Bigg|\mathrm{Cov}\Bigg(\sum_{|\boldsymbol{j}|\le nb_1,\boldsymbol{j}\in\mathbb{Z}^d}2\log|W_n(\boldsymbol{t}+\boldsymbol{j}/n,\boldsymbol{h})|,\\
&\qquad\qquad\sum_{|\boldsymbol{j}'|\le nb_1,\boldsymbol{j}'\in\mathbb{Z}^d}2\log W_n|\boldsymbol{t}'+\boldsymbol{j}'/n,\boldsymbol{h}|\Bigg)\Bigg|,
\end{aligned}\tag{3.60}$$

where the two equalities use the equation (3.42) and the ensuing derivations while the inequality uses the fact that $S_{\boldsymbol{t},p,b}(\boldsymbol{s})=O\left((nb)^{-d}\right)$ uniformly for any $b,\boldsymbol{t},\boldsymbol{s}$.

We need to discuss the impact of $|\boldsymbol{t}-\boldsymbol{t}'|$. For $\boldsymbol{t},\boldsymbol{t}'\in\Omega_m\cap\Omega_\delta$, $|\boldsymbol{t}-\boldsymbol{t}'|\ge 3b_1$, it is easy to verify that for n large enough,

$$|\boldsymbol{t}+\boldsymbol{j}/n-(\boldsymbol{t}'+\boldsymbol{j}'/n)|\ge\frac{1}{3}|\boldsymbol{t}-\boldsymbol{t}'|\ge b_1\ge\frac{q|\boldsymbol{h}|+1}{n}.$$

Then by arguments with Hermite polynomials similar to those in Lemma 3.8.15 and (3.58),

we have

$$\begin{aligned}
&\left|\mathrm{Cov}\left(\sum_{|\boldsymbol{j}|\le nb_1,\boldsymbol{j}\in\mathbb{Z}^d} 2\log|W_n(\boldsymbol{t}+\boldsymbol{j}/n,\boldsymbol{h})|,\sum_{|\boldsymbol{j}'|\le nb_1,\boldsymbol{j}'\in\mathbb{Z}^d} 2\log W_n|(\boldsymbol{t}'+\boldsymbol{j}'/n,\boldsymbol{h})|\right)\right|\\
&\le C_\delta \sum_{|\boldsymbol{j}|,|\boldsymbol{j}'|\le nb_1,\boldsymbol{j},\boldsymbol{j}'\in\mathbb{Z}^d} r_n^2(\boldsymbol{t}+\boldsymbol{j}/n,\boldsymbol{t}'+\boldsymbol{j}'/n)\\
&\le C_\delta(nb_1)^{2d}(n|\boldsymbol{t}-\boldsymbol{t}'|)^{-2\bar\psi}.
\end{aligned} \tag{3.61}$$

For fixed $\boldsymbol{t},\boldsymbol{t}'\in\Omega_m\cap\Omega_\delta$, $|\boldsymbol{t}-\boldsymbol{t}'|<3b_1$ and fixed $\boldsymbol{j}$, the number of different $\boldsymbol{j}'$ such that $|n(\boldsymbol{t}-\boldsymbol{t}')+\boldsymbol{j}-\boldsymbol{j}'|\le q|\boldsymbol{h}|+1$ is uniformly bounded. Thus, for $|\boldsymbol{t}-\boldsymbol{t}'|<3b_1$, we have

$$\begin{aligned}
&\left|\mathrm{Cov}\left(\sum_{|\boldsymbol{j}|\le nb_1,\boldsymbol{j}\in\mathbb{Z}^d} 2\log|W_n(\boldsymbol{t}+\boldsymbol{j}/n,\boldsymbol{h})|,\sum_{|\boldsymbol{j}'|\le nb_1,\boldsymbol{j}'\in\mathbb{Z}^d} 2\log|W_n(\boldsymbol{t}'+\boldsymbol{j}'/n,\boldsymbol{h})|\right)\right|\\
&\le C_\delta \sum_{|\boldsymbol{j}|,|\boldsymbol{j}'|\le nb_1,\boldsymbol{j},\boldsymbol{j}'\in\mathbb{Z}^d} r_n^2(\boldsymbol{t}+\boldsymbol{j}/n,\boldsymbol{t}'+\boldsymbol{j}'/n)\\
&\le C_\delta(nb_1)^d\left(1+\sum_{|\boldsymbol{j}|<2nb_1} n^{-2\bar\psi}|\boldsymbol{t}-\boldsymbol{t}'+\boldsymbol{j}/n|^{-2\bar\psi}I(n|\boldsymbol{t}-\boldsymbol{t}'+\boldsymbol{j}/n|>q|\boldsymbol{h}|+1)\right)\\
&\le C_\delta(nb_1)^d\sum_{l=1}^{\lceil 2nb_1\rceil} l^{-2\bar\psi+d-1}.
\end{aligned} \tag{3.62}$$

Combining (3.60), (3.61) and (3.62), we have

$$\begin{aligned}
&\left|\mathrm{Cov}\left(G(\widehat{H(\boldsymbol{t})};n,\boldsymbol{h}),G(\widehat{H(\boldsymbol{t}')};n,\boldsymbol{h})\right)\right|\\
&\le\begin{cases} C_\delta(n|\boldsymbol{t}-\boldsymbol{t}'|)^{-2\bar\psi} & \text{if } |\boldsymbol{t}-\boldsymbol{t}'|\ge 3b_1,\\ C_\delta(nb_1)^{-d}\sum_{l=1}^{\lceil 2nb_1\rceil} l^{-2\bar\psi+d-1} & \text{if } |\boldsymbol{t}-\boldsymbol{t}'|<3b_1.\end{cases}
\end{aligned} \tag{3.63}$$

We discuss the bound $3b_1$ instead of $2b_1+\frac{q|\boldsymbol{h}|+1}{n}$ since we need $|\boldsymbol{t}-\boldsymbol{t}'+(\boldsymbol{j}-\boldsymbol{j}')/n|\ge C|\boldsymbol{t}-\boldsymbol{t}'|$ for some constant $C>0$ in (3.61).

Now, if $m\le 1/(3b_1)$, $|\boldsymbol{t}-\boldsymbol{t}'|\ge 3b_1$ for all $\boldsymbol{t},\boldsymbol{t}_1\in\Omega_m,\boldsymbol{t}\ne\boldsymbol{t}_1$. Then by (3.59) and

(3.60) we have

$$
\begin{aligned}
&\mathrm{Var}(I_1') \\
&\le C_\delta \log^2(n) m^{-2d} \sum_{\boldsymbol{t}\in\Omega_m\cap\Omega_\delta} \mathrm{Var}\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h})\right) \\
&\quad + C_\delta \log^2(n) m^{-2d} \sum_{\boldsymbol{t},\boldsymbol{t}'\in\Omega_m\cap\Omega_\delta, \boldsymbol{t}\neq\boldsymbol{t}'} \mathrm{Cov}(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}), G(\widehat{H(\boldsymbol{t}')}; n, \boldsymbol{h})) \\
&\le C_\delta m^{-d} \log^2(n) T_2'(n, b_1) + C_\delta \log^2(n) m^{-2d} \left(\sum_{\boldsymbol{t},\boldsymbol{t}'\in\Omega_m\cap\Omega_\delta, \boldsymbol{t}\neq\boldsymbol{t}'} n^{-2\bar{\psi}} |\boldsymbol{t}-\boldsymbol{t}'|^{-2\bar{\psi}} \right) \\
&\le C_\delta m^{-d} \log^2(n) T_2'(n, b_1) + C_\delta \log^2(n) m^{-2d} \left(n^{-2\bar{\psi}} \sum_{l=1}^{m} l^{-2\bar{\psi}+d-1} m^{2\bar{\psi}} \right) \\
&= C_\delta m^{-d} \log^2(n) T_2'(n, b_1) + C_\delta \log^2(n) m^{2\bar{\psi}-2d} n^{-2\bar{\psi}} \sum_{l=1}^{m} l^{-2\bar{\psi}+d-1};
\end{aligned}
$$

if $m > 1/(3b_1)$, we can calculate the number of pairs of $\boldsymbol{t}, \boldsymbol{t}' \in \Omega_m$ satisfying $|\boldsymbol{t}-\boldsymbol{t}'| \le 3b_1$, which is no larger than $O\left(m^d \left(\frac{b_1}{1/m}\right)^d\right) = O(m^d (mb_1)^d)$. Thus

$$
\begin{aligned}
&\mathrm{Var}(I_1') \\
&\le C_\delta \log^2(n) m^{-2d} \sum_{\boldsymbol{t},\boldsymbol{t}'\in\Omega_m\cap\Omega_\delta, |\boldsymbol{t}-\boldsymbol{t}'|\le 3b_1} \mathrm{Cov}\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}), G(\widehat{H(\boldsymbol{t}')}; n, \boldsymbol{h})\right) \\
&\quad + C_\delta \log^2(n) m^{-2d} \sum_{\boldsymbol{t},\boldsymbol{t}'\in\Omega_m\cap\Omega_\delta, |\boldsymbol{t}-\boldsymbol{t}'|> 3b_1} \mathrm{Cov}\left(G(\widehat{H(\boldsymbol{t})}; n, \boldsymbol{h}), G(\widehat{H(\boldsymbol{t}')}; n, \boldsymbol{h})\right) \\
&\le C_\delta \log^2(n) m^{-2d} (nb_1)^{-d} \left(m^d (mb_1)^d \sum_{l=2q+1}^{\lceil 2nb_1 \rceil} l^{-2\bar{\psi}+d-1} \right) \\
&\quad + C_\delta \log^2(n) m^{-2d} \left(\sum_{\boldsymbol{t},\boldsymbol{t}'\in\Omega_m\cap\Omega_\delta, \boldsymbol{t}\neq\boldsymbol{t}'} n^{-2\bar{\psi}} |\boldsymbol{t}-\boldsymbol{t}'|^{-2\bar{\psi}} \right) \\
&\le C_\delta b_1^d \log^2(n) T_2'(n, b_1) \sum_{l=1}^{2nb_1} l^{-2\bar{\psi}+d-1} + C_\delta \log^2(n) m^{2\bar{\psi}-2d} n^{-2\bar{\psi}} \sum_{l=1}^{m} l^{-2\bar{\psi}+d-1}.
\end{aligned}
$$

The lemma now follows from the two bounds in Lemma 3.8.8. □

Lemma 3.8.9. *Under the conditions of Theorem 3.4.2,*

$$\mathbb{E}(I_2')^2 = O((T_1'(n,b_1))^2 + T_2'(n,b_1)).$$

Furthermore, if $T_2'(n,b_1) = O(T_1'(n,b_1))$*, then we have*

$$\mathbb{E}I_2' = O(T_1'(n,b_1)).$$

Remark. From this lemma and Lemma 3.8.8, we could see that, in terms of rate, the only term related to the choice of m is $\mathrm{Var}(I_1')$. We do a brief discussion about the choice of m here, although it is beyond the scope of Theorem 3.4.2. As the bound for $m > 1/(3b_1)$ is no better than the bound for $m \leq 1/(3b_1)$ and $m \sim 1/b_1$, there is no need to take $m \gg 1/b_1$. Interestingly, when $m \leq 1/b_1$, it is easy to verify that $m^{-d}T_2'(n,b_1) \geq m^{2\bar{\psi}-2d}n^{-2\bar{\psi}}\sum_{l=1}^{m} l^{-2\bar{\psi}+d-1}$ for all the choice of $\bar{\psi}$ and b_1 satisfying [A2]. Now, as $m^{-d}T_2'(n,b_1)$ is a monotone decreasing function of m, the optimal choice of m will be $m \sim 1/b_1$, which is what we take in Theorem 3.4.2.

Proof. By the Cauchy-Schwarz inequality,

$$\mathbb{E}(I_2')^2 \leq \frac{1}{\#(\Omega_m \cap \Omega_\delta)} \sum_{\boldsymbol{t}\in\Omega_m\cap\Omega_\delta} \mathbb{E}\left[\log g(\widehat{H}_1^{\gamma}(\boldsymbol{t}),\mathbf{0},\boldsymbol{h}) - \log g(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h})\right]^2.$$

So we only need a uniform bound of $\mathbb{E}\left[\log g(\widehat{H}_1^{\gamma}(\boldsymbol{t}),\mathbf{0},\boldsymbol{h}) - \log g(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h})\right]^2$ for $\boldsymbol{t} \in \Omega_\delta$.

Define the event A to be $\widehat{H}_1(\boldsymbol{t}) \in [H(\boldsymbol{t}) - \gamma/2, H(\boldsymbol{t}) + \gamma/2]$. As $\frac{\partial}{\partial H}\log g(H,\mathbf{0},\boldsymbol{h})$ is bounded on $[\gamma/2, 1-\gamma/2]$, we have

$$\begin{aligned}
&\mathbb{E}\left[\left(\log g(\widehat{H}_1^{\gamma}(\boldsymbol{t}),\mathbf{0},\boldsymbol{h}) - \log g(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h})\right)^2 I_A\right] \\
&\leq C_\delta \mathbb{E}\left[\left(\widehat{H}_1(\boldsymbol{t}) - H(\boldsymbol{t})\right)^2 I_A\right] \\
&\leq C_\delta \mathbb{E}\left[\left(\widehat{H}_1(\boldsymbol{t}) - H(\boldsymbol{t})\right)^2\right].
\end{aligned}$$

On the event A^c, we have

$$\mathbb{P}(A^c) \leq \mathbb{P}\left(|\widehat{H}_1(\boldsymbol{t}) - H(\boldsymbol{t})| \geq \frac{\gamma}{2}\right) \leq C_\delta \mathbb{E}\left[\left(\widehat{H}_1(\boldsymbol{t}) - H(\boldsymbol{t})\right)^2\right]. \tag{3.64}$$

As $\log g(\widehat{H}_1^{\gamma}(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h})$ is bounded and

$$\mathbb{E}\left[\left(\widehat{H}_1(\boldsymbol{t}) - H(\boldsymbol{t})\right)^2\right] = O((T_1'(n, b_1))^2 + T_2'(n, b_1)),$$

we have

$$\mathbb{E}\left[\left(\log g(\widehat{H}_1^{\gamma}(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h}) - \log g(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h})\right)^2 I_{A^c}\right] \leq O((T_1'(n, b_1))^2 + T_2'(n, b_1)).$$

This completes the proof for the first part of the lemma.

For the second part of the lemma, we can apply the fact that $\frac{\partial}{\partial H} \log g(H, \mathbf{0}, \boldsymbol{h})$ and $\frac{\partial^2}{\partial H^2} \log g(H, \mathbf{0}, \boldsymbol{h})$ are bounded on $[\gamma/2, 1 - \gamma/2]$. Using arguments similar to those in (3.52), (3.53), we obtain

$$\begin{aligned}
&\mathbb{E}\left[\left(\log g(\widehat{H}_1^{\gamma}(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h}) - \log g(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h})\right) I_A\right] \\
&= (\log g(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h}))' \mathbb{E}\left[\left(\widehat{H}_1(\boldsymbol{t}) - H(\boldsymbol{t})\right) I_A\right] + O((T_1'(n, b_1))^2 + T_2'(n, b_1)) \\
&= (\log g(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h}))' \mathbb{E}\left[\widehat{H}_1(\boldsymbol{t}) - H(\boldsymbol{t})\right] + O((T_1'(n, b_1))^2 + T_2'(n, b_1)) \\
&= O(T_1'(n, b_1)).
\end{aligned} \tag{3.65}$$

By (3.64) and the fact that $\log g(\widehat{H}_1^{\gamma}(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h})$ is bounded on A^c,

$$\mathbb{E}\left[\left(\log g(\widehat{H}_1^{\gamma}(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h}) - \log g(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h})\right) I_A\right] = O(T_1'(n, b_1)). \tag{3.66}$$

□

Now we return to the proof of our theorem. According to Lemma 3.8.8, if $m \sim 1/(3b_1)$, we have

$$\mathrm{Var}(I_1') = C_\delta \log^2(n) \left(b_1^d T_2'(n, b_1) + b_1^{2d - 2\bar{\psi}} n^{-2\bar{\psi}} \sum_{l=1}^{\lceil 1/b_1 \rceil} l^{-2\bar{\psi} + d - 1} \right).$$

We will establish $\mathrm{Var}(I_1') = o(T_2'(n, b_1))$ in the remaining proof. First, since $b_1^d \log^2(n) \to 0$, we have $b_1^d \log^2(n) T_2'(n, b_1) = o(T_2'(n, b_1))$. On the other hand, if $2\bar{\psi} > d$,

$$\begin{aligned}
&\log^2(n) b_1^{2d - 2\bar{\psi}} n^{-2\bar{\psi}} \sum_{l=1}^{\lceil 1/b_1 \rceil} l^{-2\bar{\psi} + d - 1} = O(\log^2(n) b_1^{2d - 2\bar{\psi}} n^{-2\bar{\psi}}) \\
&= O(\log^2(n) b_1^{2d} (n b_1)^{d - 2\bar{\psi}} (n b_1)^{-d}) = o((n b_1)^{-d});
\end{aligned}$$

if $2\bar{\psi} = d$, we have

$$\log^2(n) b_1^{2d-2\bar{\psi}} n^{-2\bar{\psi}} \sum_{l=1}^{\lceil 1/b_1 \rceil} l^{-2\bar{\psi}+d-1} = o((nb_1)^{-d} \log n);$$

if $2\bar{\psi} < d$,

$$\log^2(n) b_1^{2d-2\bar{\psi}} n^{-2\bar{\psi}} \sum_{l=1}^{\lceil 1/b_1 \rceil} l^{-2\bar{\psi}+d-1} = O(\log^2 n b_1^d n^{-2\bar{\psi}}) = o((nb_1)^{-2\bar{\psi}}).$$

Combining all the situations above, we have

$$\log^2(n) b_1^{2d-2\bar{\psi}} n^{-2\bar{\psi}} \sum_{l=1}^{\lceil 1/b_1 \rceil} l^{-2\bar{\psi}+d-1} = o(T_2'(n, b_1)).$$

Therefore, $\mathrm{Var}(I_1') = o(T_2'(n, b_1))$. The theorem is established by combining (3.56), (3.57), Lemma 3.8.8, 3.8.9 and remarks after 3.8.9.

PROOF for Theorem 3.4.3:

The proof is essentially identical to that of Theorem 3.3.4 and therefore will not be repeated.

□

3.8.5 Proof for Theorem 3.8.4

We generally assume that $\boldsymbol{h}$ is a unit vector that isparallel to one of the axis. However, the results in this section do not require that assumption.

Lemma 3.8.10. *Assume that [A3] holds. Then $g(H, \mathbf{0}, \boldsymbol{h})$ is bounded away from 0 for any fixed $\boldsymbol{h}$.*

Proof. A well-known result from Micchelli (1986) states the following:

Let $f(\boldsymbol{x}) = (-1)^k |\boldsymbol{x}|^\gamma$, $k \in \mathbb{N}_+, 2k - 2 < \gamma < 2k$. If for any n complex numbers $c_1, \ldots, c_n$, not all of them are zero, and any n distinct points $\boldsymbol{x}_1, \ldots, \boldsymbol{x}_n$ in $\mathbb{R}^d$ satisfying

$$\sum_{j=1}^{n} c_j \boldsymbol{x}_j^\alpha = 0, |\alpha| < k,$$

then we have

$$\sum_{j=1}^{n}\sum_{i=1}^{n} c_j \bar{c}_i f(\boldsymbol{x}_j - \boldsymbol{x}_i) > 0.$$

Applything this with $k = 1$, $\gamma = 2H$, $n = q + 1$, $\boldsymbol{x}_j = j\boldsymbol{h}$ and $c_i = \binom{q}{i-1}(-1)^{i-1}$, it is easy to conclude that $g(H, \mathbf{0}, \boldsymbol{h}) > 0$ for any fixed $H \in (0, 1)$. One can see that $g(H, \mathbf{0}, \boldsymbol{h})$ is continuous for $H \in (0, 1)$. By (3.22), $H(\boldsymbol{t}) \in [\gamma, 1 - \gamma]$. As $[\gamma, 1 - \gamma]$ is a compact set, $\min_{H\in[\gamma,1-\gamma]} g(H, \mathbf{0}, \boldsymbol{h}) > 0$.

This completes the proof of Lemma 3.8.10. □

Lemma 3.8.11. *Let $\boldsymbol{t}, \boldsymbol{h} \in \mathbb{R}^d$, $\boldsymbol{t}, \boldsymbol{h} \neq \mathbf{0}$ and assume $H(\boldsymbol{t}) \in \mathscr{H}_q((0, 1)^d)$, $q \in \mathbb{N}$. Then the coefficient of $|\boldsymbol{t}|^{H(\boldsymbol{t})-q}$ in $(|\boldsymbol{t} + \phi\boldsymbol{h}|^{H(\boldsymbol{t}+\phi\boldsymbol{h})})^{(q)}|_{\phi=0}$ is a polynomial in $H(\boldsymbol{t})$, $|\boldsymbol{h}|^2$ and $\langle \boldsymbol{t}, \boldsymbol{h}\rangle/|\boldsymbol{t}|$.*

Proof. We prove it by induction. The conclusion is trivial for $q = 0$. Now, assume the conclusion holds for $q = Q$ and we will prove for $q = Q + 1$. By straightforward calculations,

$$\begin{aligned}
&(|\boldsymbol{t} + \phi\boldsymbol{h}|^{H(\boldsymbol{t}+\phi\boldsymbol{h})})^{(Q+1)}|_{\phi=0} \\
&= (H(\boldsymbol{t} + \phi\boldsymbol{h})|\boldsymbol{t} + \phi\boldsymbol{h}|^{H(\boldsymbol{t}+\phi\boldsymbol{h})-2}\langle \boldsymbol{t} + \phi\boldsymbol{h}, \boldsymbol{h}\rangle \\
&\quad + \boldsymbol{h}^T \nabla H(\boldsymbol{t} + \phi\boldsymbol{h}) \log(|\boldsymbol{t} + \phi\boldsymbol{h}|)\, |\boldsymbol{t} + \phi\boldsymbol{h}|^{H(\boldsymbol{t}+\phi\boldsymbol{h})})^{(Q)}|_{\phi=0}.
\end{aligned}$$

Since this conclusion only involves the coefficient of $|\boldsymbol{t}|^{H(\boldsymbol{t})-Q-1}$ and any derivative taken on $H(\boldsymbol{t} + \phi\boldsymbol{h})$ will generate higher order polynomial of $|\boldsymbol{t}|$, it suffices to calculate

$$\begin{aligned}
&H(\boldsymbol{t})(|\boldsymbol{t} + \phi\boldsymbol{h}|^{H(\boldsymbol{t}+\phi\boldsymbol{h})-2}\langle \boldsymbol{t} + \phi\boldsymbol{h}, \boldsymbol{h}\rangle)^{(Q)}|_{\phi=0} \\
&= H(\boldsymbol{t}) \sum_{i=0}^{Q} [|\boldsymbol{t} + \phi\boldsymbol{h}|^{H(\boldsymbol{t}+\phi\boldsymbol{h})-2}]^{(i)} [\langle \boldsymbol{t} + \phi\boldsymbol{h}, \boldsymbol{h}\rangle]^{(Q-i)}|_{\phi=0}.
\end{aligned}$$

As every term in the summation on the right is a polynomial function of $H(\boldsymbol{t})$, $|\boldsymbol{h}|^2$ and $\langle \boldsymbol{t}, \boldsymbol{h}\rangle/|\boldsymbol{t}|$, the proof is accomplished by induction. □

Lemma 3.8.12. *Assume that [A1], [A2], [A3] hold. Define*

$$f(\phi, \eta; \boldsymbol{t}, \boldsymbol{s}, \boldsymbol{h}) = \frac{\partial^{2q}}{\partial\phi^q \partial\eta^q} C(\boldsymbol{t} + \phi\boldsymbol{h}, \boldsymbol{s} + \eta\boldsymbol{h}),$$

and

$$\widetilde{f}(\phi, \eta; \boldsymbol{t}, \boldsymbol{s}, \boldsymbol{h}) = \frac{\partial^{2q}}{\partial\phi^q \partial\eta^q} |\boldsymbol{t} + \phi\boldsymbol{h}|^{H(\boldsymbol{t}+\phi\boldsymbol{h})+H(\boldsymbol{s}+\eta\boldsymbol{h})} \mathscr{D}(H(\boldsymbol{t} + \phi\boldsymbol{h}), H(\boldsymbol{s} + \eta\boldsymbol{h})),$$

where C and $\mathscr{D}$ are as in Definition 3.1.1. Then there is a finite constant C_δ such that for all large enough n and uniformly for all $\boldsymbol{t}, \boldsymbol{s} \in \Omega_\delta$,

$$f(0,0;\boldsymbol{t},\boldsymbol{s},\boldsymbol{h}) \leq C_\delta |\boldsymbol{t}-\boldsymbol{s}|^{-2q+H(\boldsymbol{s})+H(\boldsymbol{t})};$$

furthermore, uniformly for $|\boldsymbol{t}-\boldsymbol{s}| \to 0$, we have

$$f(0,0;\boldsymbol{t},\boldsymbol{s},\boldsymbol{h}) = \sigma^2(c_{\boldsymbol{h}}(\boldsymbol{t},\boldsymbol{t}-\boldsymbol{s}) + O(|\boldsymbol{t}-\boldsymbol{s}|\log(|\boldsymbol{t}-\boldsymbol{s}|)))|\boldsymbol{t}-\boldsymbol{s}|^{-\psi(\boldsymbol{t})},$$

where $c_{\boldsymbol{h}}(\boldsymbol{t},\boldsymbol{u})$ is the coefficient of $|\boldsymbol{u}|^{-\psi(\boldsymbol{t})}$ in

$$-\frac{\partial^{2q}}{\partial\phi^q\partial\eta^q}\frac{1}{2}|\boldsymbol{u}+(\phi-\eta)\boldsymbol{h}|^{H(\boldsymbol{t}+\phi\boldsymbol{h})+H(\boldsymbol{t}+\eta\boldsymbol{h})}|_{\phi=0,\eta=0},$$

and, uniformly for all $\boldsymbol{s},\boldsymbol{t} \in \Omega_\delta$,

$$\widetilde{f}(0,0;\boldsymbol{t},\boldsymbol{s},\boldsymbol{h}) \leq C_\delta.$$

Proof. Let

$$I_1 := \mathscr{D}(H(\boldsymbol{t}+\phi\boldsymbol{h}),H(\boldsymbol{s}+\eta\boldsymbol{h}))|\boldsymbol{t}+\phi\boldsymbol{h}|^{H(\boldsymbol{t}+\phi\boldsymbol{h})+H(\boldsymbol{s}+\eta\boldsymbol{h})},$$

$$I_2 := \mathscr{D}(H(\boldsymbol{t}+\phi\boldsymbol{h}),H(\boldsymbol{s}+\eta\boldsymbol{h}))|\boldsymbol{s}+\eta\boldsymbol{h}|^{H(\boldsymbol{t}+\phi\boldsymbol{h})+H(\boldsymbol{s}+\eta\boldsymbol{h})},$$

and

$$I_3 := -\mathscr{D}(H(\boldsymbol{t}+\phi\boldsymbol{h}),H(\boldsymbol{s}+\eta\boldsymbol{h}))|\boldsymbol{t}-\boldsymbol{s}+(\phi-\eta)\boldsymbol{h}|^{H(\boldsymbol{t}+\phi\boldsymbol{h})+H(\boldsymbol{s}+\eta\boldsymbol{h})}.$$

Then we have

$$f(0,0;\boldsymbol{t},\boldsymbol{s},\boldsymbol{h}) = \sigma^2\frac{\partial^{2q}}{\partial\phi^q\partial\eta^q}(I_1+I_2+I_3)|_{\phi=0,\eta=0}$$

and

$$\tilde{f}(0,0;\boldsymbol{t},\boldsymbol{s},\boldsymbol{h}) = \frac{\partial^{2q}}{\partial\phi^q\partial\eta^q}I_1|_{\phi=0,\eta=0}.$$

As $\boldsymbol{t} \in \Omega_\delta$ and $|\boldsymbol{t}+\phi\boldsymbol{h}|$ is bounded away from 0 when $\phi \to 0$, we have $\tilde{f}(0,0;\boldsymbol{t},\boldsymbol{s},\boldsymbol{h}) =$

$O(1)$ uniformly. Similar arguments can be made for I_2 to deduce

$$f(0,0;\boldsymbol{t},\boldsymbol{s},\boldsymbol{h}) = \sigma^2 \frac{\partial^{2q}}{\partial\phi^q \partial\eta^q} I_3|_{\phi=0,\eta=0} + O(1).$$

If $|\boldsymbol{t}-\boldsymbol{s}| \to 0$, the dominant term in $f(0,0;\boldsymbol{t},\boldsymbol{s},\boldsymbol{h})$ is the part resulted from taking derivative with respect to ϕ and η in the base of $|\boldsymbol{t}-\boldsymbol{s}+(\phi-\eta)\boldsymbol{h}|^{H(\boldsymbol{t}+\phi\boldsymbol{h})+H(\boldsymbol{s}+\eta\boldsymbol{h})}$. By the definition of $c_{\boldsymbol{h}}(\boldsymbol{t},\boldsymbol{t}-\boldsymbol{s})$, we have

$$f(0,0;\boldsymbol{t},\boldsymbol{s},\boldsymbol{h}) = \sigma^2(c_{\boldsymbol{h}}(\boldsymbol{t},\boldsymbol{t}-\boldsymbol{s}) + O(|\boldsymbol{t}-\boldsymbol{s}|\log(|\boldsymbol{t}-\boldsymbol{s}|)))|\boldsymbol{t}-\boldsymbol{s}|^{-\psi(\boldsymbol{t})}.$$

Note that $f(0,0;\boldsymbol{t},\boldsymbol{s},\boldsymbol{h}) = O(|\boldsymbol{t}-\boldsymbol{s}|^{-2q+H(\boldsymbol{s})+H(\boldsymbol{t})})$ as $|\boldsymbol{t}-\boldsymbol{s}| \to 0$, and $f(0,0;\boldsymbol{t},\boldsymbol{s},\boldsymbol{h})$ is bounded for $|\boldsymbol{t}-\boldsymbol{s}|$ bounded away from 0 and $\boldsymbol{t},\boldsymbol{s} \in \Omega_\delta$. Thus, there is some constant C_δ such that for all $\boldsymbol{t},\boldsymbol{s} \in \Omega_\delta$,

$$f(0,0;\boldsymbol{t},\boldsymbol{s},\boldsymbol{h}) \le C_\delta|\boldsymbol{t}-\boldsymbol{s}|^{-2q+H(\boldsymbol{s})+H(\boldsymbol{t})}.$$

□

PROOF for Theorem 3.8.4:

Let $a_j = (-1)^j\binom{q}{j}, j = 0,\ldots,q$, then

$$C_n(\boldsymbol{t},\boldsymbol{t}+\boldsymbol{u}/n,\boldsymbol{h}) = n^{H(\boldsymbol{t})+H(\boldsymbol{t}+\frac{\boldsymbol{u}}{n})}(A_1+A_2+A_3),$$

where

$$A_1 = -\sum_{j_1=0}^{q}\sum_{j_2=0}^{q} a_{j_1}a_{j_2}\left|\frac{j_1\boldsymbol{h}}{n} - \frac{\boldsymbol{u}}{n} - \frac{j_2\boldsymbol{h}}{n}\right|^{H\left(\boldsymbol{t}+\frac{j_1\boldsymbol{h}}{n}\right)+H\left(\boldsymbol{t}+\frac{\boldsymbol{u}}{n}+\frac{j_2\boldsymbol{h}}{n}\right)} \mathscr{D}\left(H\left(\boldsymbol{t}+\frac{j_1\boldsymbol{h}}{n}\right), H\left(\boldsymbol{t}+\frac{\boldsymbol{u}}{n}+\frac{j_2\boldsymbol{h}}{n}\right)\right),$$

$$A_2 = \sum_{j_1=0}^{q}\sum_{j_2=0}^{q} a_{j_1}a_{j_2}\left|\boldsymbol{t}+\frac{j_1}{n}\boldsymbol{h}\right|^{H\left(\boldsymbol{t}+\frac{j_1}{n}\boldsymbol{h}\right)+H\left(\boldsymbol{t}+\frac{\boldsymbol{u}}{n}+\frac{j_2}{n}\boldsymbol{h}\right)} \mathscr{D}\left(H\left(\boldsymbol{t}+\frac{j_1}{n}\boldsymbol{h}\right), H\left(\boldsymbol{t}+\frac{\boldsymbol{u}}{n}+\frac{j_2}{n}\boldsymbol{h}\right)\right),$$

and

$$A_3 = \sum_{j_1=0}^{q}\sum_{j_2=0}^{q} a_{j_1}a_{j_2}\left|\boldsymbol{t}+\frac{\boldsymbol{u}}{n}+\frac{j_2}{n}\boldsymbol{h}\right|^{H\left(\boldsymbol{t}+\frac{j_1}{n}\boldsymbol{h}\right)+H\left(\boldsymbol{t}+\frac{\boldsymbol{u}}{n}+\frac{j_2}{n}\boldsymbol{h}\right)} \\ \mathscr{D}\left(H\left(\boldsymbol{t}+\frac{j_1}{n}\boldsymbol{h}\right), H\left(\boldsymbol{t}+\frac{\boldsymbol{u}}{n}+\frac{j_2}{n}\boldsymbol{h}\right)\right).$$

Note that $\mathscr{D}(.,.)$ is a smooth function with bounded first-order derivatives when $H(\boldsymbol{t})$ is bounded from 0 and 1. Using the fact that $n^{1/n}-1 = O\left(n^{-1}\log n\right)$, we conclude that, uniformly for all $\boldsymbol{t}$ and bounded $\boldsymbol{u}$,

$$\begin{aligned} &n^{H(\boldsymbol{t})+H\left(\boldsymbol{t}+\frac{\boldsymbol{u}}{n}\right)}A_1 - g(H(\boldsymbol{t}),\boldsymbol{u},\boldsymbol{h}) \\ &= O\left(|(j_1-j_2)\boldsymbol{h}-\boldsymbol{u}|^{2H(\boldsymbol{t})}n^{-1}\log n\right) = O\left(n^{-1}\log n\right). \end{aligned} \tag{3.67}$$

By the result on $\widetilde{f}$ in lemma 3.8.12, we have uniformly for all $\boldsymbol{t}\in\Omega_\delta$.

$$n^{H(\boldsymbol{t})+H\left(\boldsymbol{t}+\frac{\boldsymbol{u}}{n}\right)}A_2 = O\left(n^{H(\boldsymbol{t})+H\left(\boldsymbol{t}+\frac{\boldsymbol{u}}{n}\right)-2q}\right) = O\left(n^{-\psi(\boldsymbol{t})}+n^{-1}\log n\right), \tag{3.68}$$

and similarly

$$n^{H(\boldsymbol{t})+H(\boldsymbol{t}+\frac{\boldsymbol{u}}{n})}A_3 = O\left(n^{-\psi(\boldsymbol{t})}+n^{-1}\log n\right). \tag{3.69}$$

By (3.67)-(3.69), (i) follows.

Next, consider (ii). By the result about f in Lemma 3.8.12, we conclude that, uniformly for $\frac{q|\boldsymbol{h}|+1}{n} < |\boldsymbol{t}-\boldsymbol{s}| \le 2b$ and $\boldsymbol{t},\boldsymbol{s}\in\Omega_\delta$,

$$\begin{aligned} &\Delta^q_{\boldsymbol{h}/n,1}\Delta^q_{\boldsymbol{h}/n,2}\sigma^{-2}C(\boldsymbol{t},\boldsymbol{s}) \\ &= \sigma^{-2}\int_0^{1/n}\cdots\int_0^{1/n}\frac{\partial^{2q}}{\partial\phi_1\ldots\partial\phi_q\partial\eta_1\ldots\partial\eta_q}C\left(\boldsymbol{t}+\sum_{i=1}^q\phi_i\boldsymbol{h},\boldsymbol{s}+\sum_{i=1}^q\eta_i\boldsymbol{h}\right)d\boldsymbol{\phi}d\boldsymbol{\eta} \\ &= \sigma^{-2}\int_0^{1/n}\cdots\int_0^{1/n} f\left(0,0,\boldsymbol{t}+\sum_{i=1}^q\phi_i\boldsymbol{h},\boldsymbol{s}+\sum_{i=1}^q\eta_i\boldsymbol{h},\boldsymbol{h}\right)d\boldsymbol{\phi}d\boldsymbol{\eta} \\ &\sim \frac{1+o(1)}{n^{2q}}\int_{[0,1]^{2d}} c_{\boldsymbol{h}}\left(\boldsymbol{t},\boldsymbol{u}+\sum_{i=1}^q(\phi_i-\eta_i)\boldsymbol{h}\right)\left|\frac{\boldsymbol{u}}{n}+\frac{1}{n}\sum_{i=1}^q(\phi_i-\eta_i)\boldsymbol{h}\right|^{-\psi(\boldsymbol{t})}d\boldsymbol{\phi}d\boldsymbol{\eta}, \end{aligned} \tag{3.70}$$

where the last line uses Lemma 3.8.12, and $\boldsymbol{u}=n(\boldsymbol{t}-\boldsymbol{s})$ satisfying $q|\boldsymbol{h}|+1<|\boldsymbol{u}|<2nb$.

One can verify that for some constant C,

$$\int_0^1 \cdots \int_0^1 \left| \boldsymbol{u} + \sum_{i=1}^{q} (\phi_i - \eta_i)\boldsymbol{h} \right|^{-\psi(\boldsymbol{t})} d\boldsymbol{\phi} d\boldsymbol{\eta} \leq C(|\boldsymbol{u}| - q|\boldsymbol{h}|)^{-\psi(\boldsymbol{t})}.$$

Thus, when n is large enough, as $\sup_{\boldsymbol{s},\boldsymbol{t}} |c_{\mathbf{h}}(\boldsymbol{s},\boldsymbol{t})| < \infty$,

$$\begin{aligned}
&|C_n(\boldsymbol{t},\boldsymbol{s})| \\
&= \left| \sigma^{-2} n^{H(\boldsymbol{t})+H(\boldsymbol{s})} \Delta^q_{\boldsymbol{h}/n,1} \Delta^q_{\boldsymbol{h}/n,2} C(\boldsymbol{t},\boldsymbol{s}) \right| \\
&\sim (1+o(1)) \left| \int_{[0,1]^{2d}} c_{\boldsymbol{h}} \left(\boldsymbol{t}, \boldsymbol{u} + \sum_{i=1}^{q} (\phi_i - \eta_i)\boldsymbol{h} \right) \left| \boldsymbol{u} + \sum_{i=1}^{q} (\phi_i - \eta_i)\boldsymbol{h} \right|^{-\psi(\boldsymbol{t})} d\boldsymbol{\phi} d\boldsymbol{\eta} \right| \\
&\leq C_\delta (|\boldsymbol{u}| - q|\boldsymbol{h}|)^{-\psi(\boldsymbol{t})}.
\end{aligned}$$

This shows (ii).

For (iii), similar to the derivation above, for $n|\boldsymbol{t}-\boldsymbol{s}| > q|\boldsymbol{h}| + 1$ and $\boldsymbol{t}, \boldsymbol{s} \in \Omega_\delta$, we have

$$\begin{aligned}
&|C_n(\boldsymbol{t},\boldsymbol{s})| \\
&= \left| \sigma^{-2} n^{H(\boldsymbol{t})+H(\boldsymbol{s})} \int_0^{1/n} \cdots \int_0^{1/n} f\left(0, 0, \boldsymbol{t} + \sum_{i=1}^{q} \phi_i \boldsymbol{h}, \boldsymbol{s} + \sum_{i=1}^{q} \eta_i \boldsymbol{h}, \boldsymbol{h} \right) d\boldsymbol{\phi} d\boldsymbol{\eta} \right| \\
&\leq C'' n^{-2q+H(\boldsymbol{t})+H(\boldsymbol{s})} \int_0^1 \cdots \int_0^1 \left| \boldsymbol{t} - \boldsymbol{s} + \frac{1}{n} \sum_{i=1}^{q} (\phi_i - \eta_i)\boldsymbol{h} \right|^{-2q+H(\boldsymbol{t})+H(\boldsymbol{s})} d\boldsymbol{\phi} d\boldsymbol{\eta} \\
&\leq C'' n^{-2q+H(\boldsymbol{t})+H(\boldsymbol{s})} (|\boldsymbol{t}-\boldsymbol{s}| - q|\boldsymbol{h}|/n)^{-2q+H(\boldsymbol{t})+H(\boldsymbol{s})} \\
&\leq C''(n|\boldsymbol{t}-\boldsymbol{s}| - q|\boldsymbol{h}|)^{-2q+H(s)+H(t)} \\
&\leq C''(n|\boldsymbol{t}-\boldsymbol{s}| - q|\boldsymbol{h}|)^{-\bar{\psi}} \\
&\leq C_\delta (n|\boldsymbol{t}-\boldsymbol{s}|)^{-\bar{\psi}},
\end{aligned}$$

where the third line uses Lemma 3.8.12. This shows (iii).

Furthermore, when $|\boldsymbol{u}| \to \infty$, by Lemma 3.8.11,

$$c_{\boldsymbol{h}} \left(\boldsymbol{t}, \boldsymbol{u} + \sum_{i=1}^{q} (\phi_i - \eta_i)\boldsymbol{h} \right) \to c_{\boldsymbol{h}} (\boldsymbol{t}, \boldsymbol{u}),$$

and

$$\left| \boldsymbol{u} + \sum_{i=1}^{q} (\phi_i - \eta_i)\boldsymbol{h} \right|^{-\psi(\boldsymbol{t})} \to |\boldsymbol{u}|^{-\psi(\boldsymbol{t})}.$$

Combining the previous two equations with (3.70), we have

$$C_n(\boldsymbol{t}, \boldsymbol{s}) = (c_{\boldsymbol{h}}(\boldsymbol{t}, \boldsymbol{t} - \boldsymbol{s}) + O(b \log b))|\boldsymbol{u}|^{-\psi(\boldsymbol{t})}.$$

Thus, (iv) is established and the proof is complete.

□

3.8.6 Proofs for Lemma 3.8.5, Lemma 3.8.6 and Lemma 3.8.7

In the remaining part of the proof, as in the chapter, we assume that $\boldsymbol{h}$ is a unit vector that parallels an axis.

PROOF for Lemma 3.8.5:

By Theorem 3.8.4, we have

$$\mathbb{E}[W_n^2(\boldsymbol{t}, \boldsymbol{h})] = g(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h}) + O(\rho_n). \tag{3.71}$$

As $W_n(\boldsymbol{t}, \boldsymbol{h})$ follows a normal distribution with mean 0, we have

$$2\mathbb{E}|\log W_n(\boldsymbol{t}, \boldsymbol{h})| = 2 \log |\mathbb{E}W_n(\boldsymbol{t}, \boldsymbol{h})| + \mathbb{E} \log \chi_1^2. \tag{3.72}$$

The result follows from Lemma 3.8.10 and (3.71).

□

PROOF for Lemma 3.8.6:

Denote $D_{\boldsymbol{t},b} = \{\mathbf{z} : \boldsymbol{t} + b\mathbf{z} \in \Omega_\delta\} \cap [0,1]^d$. Since K, $\mathbf{A}$ and H have continuous second-order derivatives, it is not hard to find that

$$\begin{aligned}
&(nb)^{-d} \sum_i K\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) \mathbf{A}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) \mathbf{A}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right)^T \\
&= \int_{D_{\boldsymbol{t},b}} K(\mathbf{z})\, \mathbf{A}(\mathbf{z})\, \mathbf{A}^T(\mathbf{z})\, d\mathbf{z} + O\left(\frac{1}{(nb)^{2\wedge p}}\right)
\end{aligned}$$

and

$$\begin{aligned}
&(nb)^{-d} \sum_i K\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) \mathbf{A}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) H(\boldsymbol{t}_i) \\
&= \int_{D_{\boldsymbol{t},b}} K(\mathbf{z}) \mathbf{A}(\mathbf{z}) H(\boldsymbol{t} + b\mathbf{z}) d\mathbf{z} + O\left(\frac{1}{(nb)^{2\wedge p}}\right).
\end{aligned}$$

By Theorem 2 in Newman (1974), we have

$$I_1(\boldsymbol{t}, n, b) = 2\mathbf{e}_1^T \left(\int_{D_{\boldsymbol{t},b}} K(\mathbf{z})\mathbf{A}(\mathbf{z})\mathbf{A}^T(\mathbf{z})d\mathbf{z} \right)^{-1} \times \int_{D_{\boldsymbol{t},b}} K(\mathbf{z})\mathbf{A}(\mathbf{z})H(\boldsymbol{t}+b\mathbf{z})d\mathbf{z} \log\left(\frac{1}{n}\right) + O\left(\frac{\log(n)}{(nb)^{2\wedge p}}\right).$$

By Taylor's Theorem, when $p \in \mathbb{N}$, we have

$$H(\boldsymbol{t}+b\mathbf{z}) = \mathbf{A}(\mathbf{z})^T\mathbf{C}(\boldsymbol{t}) + \sum_{|\boldsymbol{\alpha}|=p} R_{\boldsymbol{\alpha}}(\boldsymbol{t})(b\mathbf{z})^{\boldsymbol{\alpha}},$$

where $\mathbf{C}(\boldsymbol{t}) \in \mathbb{R}^p$ of which the first element is $H(\boldsymbol{t})$, and

$$R_{\boldsymbol{\alpha}}(\boldsymbol{t}) = \frac{|\boldsymbol{\alpha}|}{\boldsymbol{\alpha}!} \int_0^1 (1-s)^{|\boldsymbol{\alpha}|-1} D^{\boldsymbol{\alpha}} H(\boldsymbol{t}+sb\mathbf{z})ds.$$

Thus,

$$\begin{aligned}
&2\mathbf{e}_1^T \left(\int_{D_{\boldsymbol{t},b}} K(\mathbf{z})\mathbf{A}(\mathbf{z})\mathbf{A}^T(\mathbf{z})d\mathbf{z} \right)^{-1} \int_{D_{\boldsymbol{t},b}} K(\mathbf{z})\mathbf{A}(\mathbf{z})H(\boldsymbol{t}+b\mathbf{z})d\mathbf{z} \log\left(\frac{1}{n}\right) \\
&= 2\log\left(\frac{1}{n}\right) H(\boldsymbol{t}) + 2\log\left(\frac{1}{n}\right) \mathbf{e}_1^T \left(\int_{D_{\boldsymbol{t},b}} K(\mathbf{z})\mathbf{A}(\mathbf{z})\mathbf{A}^T(\mathbf{z})d\mathbf{z} \right)^{-1} \times \\
&\qquad \int_{D_{\boldsymbol{t},b}} K(\mathbf{z})\mathbf{A}(\mathbf{z}) \left(\sum_{|\boldsymbol{\alpha}|=p} R_{\boldsymbol{\alpha}}(\boldsymbol{t})(b\mathbf{z})^{\boldsymbol{\alpha}} \right) d\mathbf{z} \\
&= -2\log(n)(H(\boldsymbol{t}) + \mathbf{R}(\boldsymbol{t}, b)).
\end{aligned}$$

It is obvious that $\mathbf{R}(\boldsymbol{t}, b) = O(b^p)$ uniformly for all $\boldsymbol{t} \in \Omega_\delta$.

When $p \notin \mathbb{N}$, we have

$$H(\boldsymbol{t}+b\mathbf{z}) = \mathbf{A}(\mathbf{z})^T\mathbf{C}(\boldsymbol{t}) + O((b\boldsymbol{z})^p),$$

uniformly, where $\mathbf{C}(\boldsymbol{t}) \in \mathbb{R}^{\lceil p \rceil}$ of which the first element is $H(\boldsymbol{t})$. By the same calculation above, we can still get $\mathbf{R}(\boldsymbol{t}, b) = O(b^p)$, although an explicit formula of $\mathbf{R}(\boldsymbol{t}, b)$ is not available.

□

Lemma 3.8.13. *Let $\boldsymbol{t}, \boldsymbol{t} + \boldsymbol{u}/n \in \Omega_\delta$ and assume that [A1]-[A3] hold. If $|\boldsymbol{u}| \to \infty$ and*

$|\boldsymbol{u}| < 2bn$, *then*

$$g(H(\boldsymbol{t}), \boldsymbol{u}, \boldsymbol{h}) = (c_{\boldsymbol{h}}(\boldsymbol{t}, \boldsymbol{u}) + o(1))|\boldsymbol{u}|^{-\psi(\boldsymbol{t})}.$$

Proof. With the same argument as (3.70) and proof for (iv) in Theorem 3.8.4 except for ignoring σ^2 and changing $C(\boldsymbol{t}, \boldsymbol{s})$ to $f(\boldsymbol{t}, \boldsymbol{s}) = -\frac{1}{2}|\boldsymbol{t} - \boldsymbol{s}|^{H(\boldsymbol{t})+H(\boldsymbol{s})}$, we can get our lemma. □

Lemma 3.8.14. *Define*

$$Y_n(\boldsymbol{t}, \boldsymbol{h}) = \frac{W_n(\boldsymbol{t}, \boldsymbol{h})}{\sqrt{\mathrm{Var}(W_n(\boldsymbol{t}, \boldsymbol{h}))}},$$

and

$$r_n(\boldsymbol{t}, \boldsymbol{s}; \boldsymbol{h}) = \mathrm{Cov}(Y_n(\boldsymbol{t}, \boldsymbol{h}), Y_n(\boldsymbol{s}, \boldsymbol{h})).$$

Assume [A1], [A2], [A3] and f to be a bounded function with supporting set on $B_1(\mathbf{0})$ and bounded first order derivative. Then, if $2\psi(\boldsymbol{t}) > d$, for $k \geq 2$ and uniformly for all $\boldsymbol{t} \in \Omega_\delta$, we have

$$\begin{aligned}
&(nb)^{-d} \sum_{\boldsymbol{t}_i, \boldsymbol{t}_j \in \Omega_n} f\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) f\left(\frac{\boldsymbol{t}_j - \boldsymbol{t}}{b}\right) r_n^k(\boldsymbol{t}_i, \boldsymbol{t}_j) \\
&\to \int_{D_{\boldsymbol{t},b}} f^2(\mathbf{z}) d\mathbf{z} \sum_{\boldsymbol{j} \in \mathbb{Z}^d} \left(\frac{g(H(\boldsymbol{t}), \boldsymbol{j}, \boldsymbol{h})}{g(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h})}\right)^k.
\end{aligned}$$

If $2\psi(\boldsymbol{t}) = d$, uniformly for all $\boldsymbol{t} \in \Omega_\delta$, we have

$$\begin{aligned}
&(nb)^{-d} \log^{-1}(nb) \sum_{\boldsymbol{t}_i, \boldsymbol{t}_j \in \Omega_n} f\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) f\left(\frac{\boldsymbol{t}_j - \boldsymbol{t}}{b}\right) r_n^2(\boldsymbol{t}_i, \boldsymbol{t}_j) \\
&\to \frac{V_d f^2(\mathbf{0})}{g(H(\boldsymbol{t}, \mathbf{0}, \boldsymbol{h}))} \int_{B_1(\mathbf{0})} \frac{c_{\boldsymbol{h}}^2(\boldsymbol{t}, \boldsymbol{x})}{|\boldsymbol{x}|^{d-1}} d\boldsymbol{x}.
\end{aligned}$$

If $2\psi(\boldsymbol{t}) < d$, for $k \geq 2$ and uniformly for all $\boldsymbol{t} \in \Omega_\delta$, we have

$$d_n^{-k} \sum_{\text{distinct } \boldsymbol{t}_{i_1},\ldots,\boldsymbol{t}_{i_k} \in \Omega_n} \prod_{l=1}^{k} r_n\left(\boldsymbol{t}_{i_l}, \boldsymbol{t}_{i_{l+1}}\right) f\left(\frac{\boldsymbol{t}_{i_l} - \boldsymbol{t}}{b}\right)$$
$$\to \int_{(B_1(\mathbf{0}))^{\otimes k}} \frac{\prod_{i=1}^{k} c_{\boldsymbol{h}}(\boldsymbol{t}, \mathbf{z}_i - \mathbf{z}_{i+1})|\mathbf{z}_i - \mathbf{z}_{i+1}|^{-\psi(\boldsymbol{t})} f(\mathbf{z}_i)}{g^k(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h})} d\mathbf{z}_1 \ldots d\mathbf{z}_k,$$

where $d_n = (nb)^{d-\psi(\boldsymbol{t})}$ and $\mathbf{z}_{k+1} := \mathbf{z}_1$.

Proof. By the first part of Theorem 3.8.4 and Lemma 3.8.10, we have for any fixed $\boldsymbol{j}$ and $\boldsymbol{t} \in \Omega_\delta$,

$$r_n(\boldsymbol{t}, \boldsymbol{t} + \boldsymbol{j}/n; \boldsymbol{h}) = \frac{g(H(\boldsymbol{t}), \boldsymbol{j}, \boldsymbol{h})}{g(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h})} + O(\rho_n(\boldsymbol{t})). \tag{3.73}$$

By the second part of Theorem 3.8.4 and Lemma 3.8.10, we have uniformly for $\boldsymbol{t} \in \Omega_\delta$ and $\boldsymbol{j} \leq nb$,

$$r_n(\boldsymbol{t}, \boldsymbol{t} + \boldsymbol{j}/n; \boldsymbol{h}) \leq C|\boldsymbol{j}|^{-\psi(\boldsymbol{t})}, \tag{3.74}$$

where C is some constant. By Fatou's Lemma,

$$\begin{aligned}
\sum_{\boldsymbol{j} \in \mathbb{Z}^d} \left(\frac{g(H(\boldsymbol{t}), \boldsymbol{j}, \boldsymbol{h})}{g(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h})}\right)^k &= \lim_{m\to\infty} \sum_{\boldsymbol{j} \in \mathbb{Z}^d \cap B_m(\mathbf{0})} \left(\frac{g(H(\boldsymbol{t}), \boldsymbol{j}, \boldsymbol{h})}{g(H(\boldsymbol{t}), \mathbf{0}, \boldsymbol{h})}\right)^k \\
&\leq \lim_{m\to\infty} \liminf_{n\to\infty} \sum_{\boldsymbol{j} \in \mathbb{Z}^d \cap B_m(\mathbf{0})} r_n^k(\boldsymbol{t}, \boldsymbol{t} + \frac{\boldsymbol{j}}{n}; \boldsymbol{h}) \\
&\leq C \lim_{m\to\infty} \sum_{\boldsymbol{j} \in \mathbb{Z}^d \cap B_m(\mathbf{0})} |\boldsymbol{j}|^{-k\psi(\boldsymbol{t})} \\
&\leq \sum_{\boldsymbol{j} \in \mathbb{Z}^d} |\boldsymbol{j}|^{-k\psi(\boldsymbol{t})} < \infty.
\end{aligned} \tag{3.75}$$

While $\boldsymbol{t}$ is arbitrary in Ω_δ, for convenience we will assume that $\boldsymbol{t}$ is a grid point in Ω_n; if $\boldsymbol{t}$ is not in Ω_n then we could replace $\boldsymbol{t}$ by the closest point in Ω_n and the bias will be a lower order term. With this simplification, we relabel the points $\boldsymbol{t}_i$ as $\boldsymbol{t} + \boldsymbol{i}/n$ with $\boldsymbol{i}$ in some

subset $\Omega_{\boldsymbol{t}}$ of $\mathbb{Z}^d$, such that $\boldsymbol{t}+\boldsymbol{i}/n \in \Omega_\delta \cap B_b(\boldsymbol{t})$. For a fixed positive m, write

$$\begin{aligned}
&(nb)^{-d} \sum_{\boldsymbol{i},\boldsymbol{j}\in\Omega_{\boldsymbol{t}}} f\left(\frac{\boldsymbol{i}}{nb}\right) f\left(\frac{\boldsymbol{j}}{nb}\right) r_n^k(\boldsymbol{t}+\boldsymbol{i}/n, \boldsymbol{t}+\boldsymbol{j}/n) \\
&= \int_{D_{\boldsymbol{t},b}} f^2(\mathbf{z}) d\mathbf{z} \sum_{\boldsymbol{j}\in\mathbb{Z}^d} \left(\frac{g(H(\boldsymbol{t}),\boldsymbol{j},\boldsymbol{h})}{g(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h})}\right)^k + \epsilon_{n,m,1} + \cdots + \epsilon_{n,m,5}
\end{aligned}$$

where

$$\epsilon_{n,m,1} = (nb)^{-d} \left(\sum_{\boldsymbol{i},\boldsymbol{j}} - \sum_{\boldsymbol{i}} \sum_{|\boldsymbol{j}-\boldsymbol{i}|\le m} \right) f\left(\frac{\boldsymbol{i}}{nb}\right) f\left(\frac{\boldsymbol{j}}{nb}\right) r_n^k(\boldsymbol{t}+\boldsymbol{i}/n, \boldsymbol{t}+\boldsymbol{j}/n),$$

$$\begin{aligned}
\epsilon_{n,m,2} = (nb)^{-d} \Bigg(&\sum_{\boldsymbol{i}} \sum_{|\boldsymbol{j}|\le m} f\left(\frac{\boldsymbol{i}}{nb}\right) f\left(\frac{\boldsymbol{i}+\boldsymbol{j}}{nb}\right) r_n^k(\boldsymbol{t}+\boldsymbol{i}/n, \boldsymbol{t}+(\boldsymbol{i}+\boldsymbol{j})/n) \\
&- \sum_{\boldsymbol{i}} \sum_{|\boldsymbol{j}|\le m} f^2\left(\frac{\boldsymbol{i}}{nb}\right) r_n^k(\boldsymbol{t}+\boldsymbol{i}/n, \boldsymbol{t}+(\boldsymbol{i}+\boldsymbol{j})/n \Bigg),
\end{aligned}$$

$$\begin{aligned}
\epsilon_{n,m,3} = &(nb)^{-d} \sum_{\boldsymbol{i}} \sum_{|\boldsymbol{j}|\le m} f^2\left(\frac{\boldsymbol{i}}{nb}\right) r_n^k(\boldsymbol{t}+\boldsymbol{i}/n, \boldsymbol{t}+(\boldsymbol{i}+\boldsymbol{j})/n) \\
&- \int_{D_{\boldsymbol{t},b}} f^2(\mathbf{z}) d\mathbf{z} \sum_{|\boldsymbol{j}|\le m} r_n^k(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{j}/n),
\end{aligned}$$

$$\begin{aligned}
\epsilon_{n,m,4} = &\int_{D_{\boldsymbol{t},b}} f^2(\mathbf{z}) d\mathbf{z} \sum_{|\boldsymbol{j}|\le m} r_n^k(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{j}/n) \\
&- \int_{D_{\boldsymbol{t},b}} f^2(\mathbf{z}) d\mathbf{z} \sum_{|\boldsymbol{j}|\le m} \left(\frac{g(H(\boldsymbol{t}),\boldsymbol{j},\boldsymbol{h})}{g(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h})}\right)^k,
\end{aligned}$$

$$\begin{aligned}
\epsilon_{n,m,5} = &\int_{D_{\boldsymbol{t},b}} f^2(\mathbf{z}) d\mathbf{z} \sum_{|\boldsymbol{j}|\le m} \left(\frac{g(H(\boldsymbol{t}),\boldsymbol{j},\boldsymbol{h})}{g(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h})}\right)^k \\
&- \int_{D_{\boldsymbol{t},b}} f^2(\mathbf{z}) d\mathbf{z} \sum_{|\boldsymbol{j}|\in\mathbb{Z}^d} \left(\frac{g(H(\boldsymbol{t}),\boldsymbol{j},\boldsymbol{h})}{g(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h})}\right)^k.
\end{aligned}$$

We will show that for $i = 1, \ldots, 5$,

$$\lim_{m\to\infty} \limsup_{n\to\infty} |\epsilon_{n,m,i}| = 0. \tag{3.76}$$

For $\epsilon_{n,m,1}$,

$$\begin{aligned}
|\epsilon_{n,m,1}| &= (nb)^{-d} \sum_{\boldsymbol{i}} \sum_{\boldsymbol{j}>m} \left| f\left(\frac{\boldsymbol{i}}{nb}\right) f\left(\frac{\boldsymbol{i}+\boldsymbol{j}}{nb}\right) r_n^k(\boldsymbol{t}+\boldsymbol{i}/n, \boldsymbol{t}+(\boldsymbol{i}+\boldsymbol{j})/n) \right| \\
&\leq C(nb)^{-d} \sum_{\boldsymbol{i}} \left| f\left(\frac{\boldsymbol{i}}{nb}\right) \right| \sum_{|\boldsymbol{j}|>m} |\boldsymbol{j}|^{-k\psi(\boldsymbol{t})} \\
&\leq Cm^{d-1-k\psi(\boldsymbol{t})}
\end{aligned}$$

for some $C < \infty$.

$$\begin{aligned}
|\epsilon_{n,m,2}| \leq (nb)^{-d} \sum_{\boldsymbol{i}} \sum_{|\boldsymbol{j}|\leq m} &\left| f\left(\frac{\boldsymbol{i}}{nb}\right) \left[f\left(\frac{\boldsymbol{i}+\boldsymbol{j}}{nb}\right) - f\left(\frac{\boldsymbol{i}}{nb}\right) \right] \right| \times \\
&r_n^k\left(\boldsymbol{t}+\frac{\boldsymbol{i}}{n}, \boldsymbol{t}+\frac{\boldsymbol{i}+\boldsymbol{j}}{n}\right) \\
= O((nb)^{-1}).&
\end{aligned}$$

Also, $\epsilon_{n,m,3} = O((nb)^{-1} + b)$ is a direct result of Riemann approximation and (3.73). That (3.76) holds for $i = 4, 5$ follows from (3.73) and (3.75), respectively. Thus, the first part of this lemma is established.

For the second part of this lemma, by Lemma 3.8.13 and Lemma 3.8.10, we know that as n and $|\boldsymbol{i}-\boldsymbol{j}| \to \infty$ and $\boldsymbol{t}_{\boldsymbol{i}}, \boldsymbol{t}_{\boldsymbol{j}} \in B_b(\boldsymbol{t})$,

$$r_n(\boldsymbol{t}_{\boldsymbol{i}}, \boldsymbol{t}_{\boldsymbol{j}}) = \left(\frac{c_{\boldsymbol{h}}(\boldsymbol{t}, \boldsymbol{i}-\boldsymbol{j})}{g(H(\boldsymbol{t}), \boldsymbol{0}, \boldsymbol{h}))} + o(1) \right) |\boldsymbol{i}-\boldsymbol{j}|^{-\psi(\boldsymbol{t})}. \tag{3.77}$$

One can verify that for all $\boldsymbol{t} \in (0,1)^d$, there exists a $N \geq 1$, such that for all $n \geq N$, $\Omega_{\boldsymbol{t}} = B_{nb}(\boldsymbol{0}) \cap \mathbb{Z}^d$ and

$$n^{-d} \log^{-1}(n) \sum_{\boldsymbol{i},\boldsymbol{j}\in[-n,n]^d, \boldsymbol{i}\neq\boldsymbol{j}} |\boldsymbol{i}-\boldsymbol{j}|^{-d} = O(1). \tag{3.78}$$

Denote

$$\epsilon'_{n,m,\delta,1} = \frac{1}{(nb)^d \log(nb)} \sum_{\boldsymbol{i},\boldsymbol{j} \in \Omega_{\boldsymbol{t}}} f\left(\frac{\boldsymbol{i}}{nb}\right) f\left(\frac{\boldsymbol{j}}{nb}\right) r_n^2(\boldsymbol{t} + \boldsymbol{i}/n, \boldsymbol{t} + \boldsymbol{j}/n) - \frac{1}{(nb)^d \log(nb)} \sum_{\substack{\boldsymbol{i},\boldsymbol{j} \in B_{nb}(\boldsymbol{0}) \\ |\boldsymbol{i}-\boldsymbol{j}| \geq m}} f\left(\frac{\boldsymbol{i}}{nb}\right) f\left(\frac{\boldsymbol{j}}{nb}\right) r_n^2(\boldsymbol{t} + \boldsymbol{i}/n, \boldsymbol{t} + \boldsymbol{j}/n),$$

$$\epsilon'_{n,m,\delta,2} = \frac{1}{(nb)^d \log(nb)} \sum_{\substack{\boldsymbol{i},\boldsymbol{j} \in B_{nb}(\boldsymbol{0}) \\ |\boldsymbol{i}-\boldsymbol{j}| \geq m}} f\left(\frac{\boldsymbol{i}}{nb}\right) f\left(\frac{\boldsymbol{j}}{nb}\right) r_n^2(\boldsymbol{t} + \boldsymbol{i}/n, \boldsymbol{t} + \boldsymbol{j}/n) - \frac{1}{(nb)^d \log(nb)} \sum_{\substack{\boldsymbol{i},\boldsymbol{j} \in B_{nb}(\boldsymbol{0}) \\ |\boldsymbol{i}-\boldsymbol{j}| \geq m}} f\left(\frac{\boldsymbol{i}}{nb}\right) f\left(\frac{\boldsymbol{j}}{nb}\right) \frac{c_{\boldsymbol{h}}^2(\boldsymbol{t}, \boldsymbol{i}-\boldsymbol{j})}{g^2(H(\boldsymbol{t}), \boldsymbol{0}, \boldsymbol{h}))} |\boldsymbol{i}-\boldsymbol{j}|^{-d},$$

$$\epsilon'_{n,m,\delta,3} = \frac{1}{(nb)^d \log(nb)} \sum_{\substack{\boldsymbol{i},\boldsymbol{j} \in B_{nb}(\boldsymbol{0}) \\ |\boldsymbol{i}-\boldsymbol{j}| \geq m}} f\left(\frac{\boldsymbol{i}}{nb}\right) f\left(\frac{\boldsymbol{j}}{nb}\right) \frac{c_{\boldsymbol{h}}^2(\boldsymbol{t}, \boldsymbol{i}-\boldsymbol{j})}{g^2(H(\boldsymbol{t}), \boldsymbol{0}, \boldsymbol{h}))} |\boldsymbol{i}-\boldsymbol{j}|^{-d} - \frac{1}{(nb)^d \log(nb)} \sum_{\substack{\boldsymbol{i},\boldsymbol{j} \in B_{nb}(\boldsymbol{0}) \\ m \leq |\boldsymbol{i}-\boldsymbol{j}| \leq \delta nb}} f\left(\frac{\boldsymbol{i}}{nb}\right) f\left(\frac{\boldsymbol{j}}{nb}\right) \frac{c_{\boldsymbol{h}}^2(\boldsymbol{t}, \boldsymbol{i}-\boldsymbol{j})}{g^2(H(\boldsymbol{t}), \boldsymbol{0}, \boldsymbol{h}))} |\boldsymbol{i}-\boldsymbol{j}|^{-d},$$

$$\epsilon'_{n,m,\delta,4} = \frac{1}{(nb)^d \log(nb)} \sum_{\substack{\boldsymbol{i},\boldsymbol{j} \in B_{nb}(\boldsymbol{0}) \\ m \leq |\boldsymbol{i}-\boldsymbol{j}| \leq \delta nb}} \left(f\left(\frac{\boldsymbol{i}}{nb}\right) f\left(\frac{\boldsymbol{j}}{nb}\right) - f^2(\boldsymbol{0}) \right) \times \frac{c_{\boldsymbol{h}}^2(\boldsymbol{t}, \boldsymbol{i}-\boldsymbol{j})}{g^2(H(\boldsymbol{t}), \boldsymbol{0}, \boldsymbol{h}))} |\boldsymbol{i}-\boldsymbol{j}|^{-d},$$

$$\epsilon'_{n,m,\delta,5} = \frac{f^2(\boldsymbol{0})}{(nb)^d \log(nb)} \sum_{\substack{\boldsymbol{i},\boldsymbol{j} \in B_{nb}(\boldsymbol{0}) \\ m \leq |\boldsymbol{i}-\boldsymbol{j}| \leq \delta nb}} \frac{c_{\boldsymbol{h}}^2(\boldsymbol{t}, \boldsymbol{i}-\boldsymbol{j})}{g^2(H(\boldsymbol{t}), \boldsymbol{0}, \boldsymbol{h}))} |\boldsymbol{i}-\boldsymbol{j}|^{-d} - \frac{f^2(\boldsymbol{0})}{(nb)^d \log(nb)} \int_{\substack{\boldsymbol{x},\mathbf{y} \in B_{nb}(\boldsymbol{0}) - B_1(\boldsymbol{0}) \\ m \leq |\boldsymbol{x}-\mathbf{y}| \leq \delta nb}} \frac{c_{\boldsymbol{h}}^2(\boldsymbol{t}, \boldsymbol{x}-\mathbf{y})}{g^2(H(\boldsymbol{t}), \boldsymbol{0}, \boldsymbol{h}))} |\boldsymbol{x}-\mathbf{y}|^{-d} d\boldsymbol{x} d\mathbf{y},$$

$$\epsilon'_{n,m,\delta,6} = \frac{f^2(\mathbf{0})}{(nb)^d \log(nb)} \int_{\substack{\boldsymbol{x},\mathbf{y}\in B_{nb}(\mathbf{0})-B_1(\mathbf{0}) \\ m\le|\boldsymbol{x}-\mathbf{y}|\le\delta nb}} \frac{c^2_{\boldsymbol{h}}(\boldsymbol{t},\boldsymbol{x}-\mathbf{y})}{g^2(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h}))}|\boldsymbol{x}-\mathbf{y}|^{-d}d\boldsymbol{x}d\mathbf{y}$$
$$- \frac{V_d f^2(\mathbf{0})}{g^2(H(\boldsymbol{t},\mathbf{0},\boldsymbol{h}))}\int_{|\boldsymbol{x}|=1} c^2_{\boldsymbol{h}}(\boldsymbol{t},\boldsymbol{x})d\boldsymbol{x}.$$

We will show

$$\lim_{\delta\to 0}\lim_{m\to\infty}\lim_{n\to\infty}|\epsilon'_{n,m,\delta,i}| = 0, i = 1,\ldots,6. \tag{3.79}$$

First,

$$|\epsilon'_{n,m,\delta,1}| \le (nb)^{-d}\log^{-1}(nb)\sum_{\boldsymbol{i}\in B_{nb}(\mathbf{0})} V_d m^d \sup_{\mathbf{z}}|f(\mathbf{z})|^2$$
$$= O\left(\frac{m^d}{\log(nb)}\right)\to 0 \text{ as } n\to\infty.$$

That (3.79) holds for $i = 2$ is a consequence of (3.77) and (3.78). Next,

$$|\epsilon'_{n,m,\delta,3}| \le \frac{\sup_{\boldsymbol{z}} f^2(\mathbf{z})\sup_{\boldsymbol{z}} c^2_{\boldsymbol{h}}(\boldsymbol{t},\boldsymbol{z})}{(nb)^d\log(nb)g^2(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h})}\sum_{\boldsymbol{i}\in B_{nb}(\mathbf{0}),|\boldsymbol{j}|\in[\delta nb,nb]}|\boldsymbol{j}|^{-d}$$
$$= O\left(\frac{\log(\delta)}{\log(nb)}\right)\to 0 \text{ as } n\to\infty.$$

That (3.79) holds for $i = 4$ follows from (3.78), smoothness of $f(.)$ and boundness of $c_{\boldsymbol{h}}(.,.)$. That (3.79) holds for $i = 5$ is due to (3.78), boundedness and smoothness of $c_{\boldsymbol{h}}(.,.)$ and

$$\lim_{m\to\infty}\sup_{|\boldsymbol{i}|>m}\left|\frac{c^2_{\boldsymbol{h}}(\boldsymbol{t},\boldsymbol{i})|\boldsymbol{i}|^{-d}}{\int_{[\boldsymbol{i},\boldsymbol{i}+\mathbf{1}]}c^2_{\boldsymbol{h}}(\boldsymbol{t},\boldsymbol{x})|\boldsymbol{x}|^{-d}d\boldsymbol{x}} - 1\right| = 0.$$

Specifically, there exists an $M_1 > 0$ such that

$$(nb)^{-d}\log^{-1}(nb)\int_{\substack{\boldsymbol{x},\mathbf{y}\in B_{nb}(\mathbf{0})-B_1(\mathbf{0}) \\ m\le|\boldsymbol{x}-\mathbf{y}|\le\delta nb}} f^2(\mathbf{0})\frac{c^2_{\boldsymbol{h}}(\boldsymbol{t},\boldsymbol{x}-\mathbf{y})}{g^2(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h}))}|\boldsymbol{x}-\mathbf{y}|^{-d}d\boldsymbol{x}d\mathbf{y} < M_1.$$

Thus,

$$\lim_{m\to\infty}|\epsilon'_{n,m,\delta,5}| \le (nb)^{-d}\log^{-1}(nb)f^2(\mathbf{0})\lim_{m\to\infty}|S_{n,m}/I_{n,m}-1|I_{n,m}$$
$$\le M_1\lim_{m\to\infty}\sup_{|\boldsymbol{i}|>m}\left|\frac{c^2_{\boldsymbol{h}}(\boldsymbol{t},\boldsymbol{i})|\boldsymbol{i}|^{-d}}{\int_{[\boldsymbol{i},\boldsymbol{i}+\mathbf{1}]}c^2_{\boldsymbol{h}}(\boldsymbol{t},\boldsymbol{x})|\boldsymbol{x}|^{-d}d\boldsymbol{x}} - 1\right| = 0,$$

where $S_{n,m}$ and $I_{n,m}$ represent the summation part and integration part, respectively, in $\epsilon'_{n,m,\delta,5}$.

Finally, for $\epsilon'_{n,m,\delta,6}$, letting

$$\boldsymbol{\Theta} := (\cos\theta_1, \sin\theta_1\cos\theta_2, \ldots, \prod_{i=1}^{d-2}\sin\theta_i\cos\theta_{d-1}, \prod_{i=1}^{d-1}\sin\theta_i)$$

we have

$$\begin{aligned}
&\lim_{\delta\to 0}\lim_{n\to\infty}\frac{f^2(\mathbf{0})}{(nb)^d\log(nb)}\int_{\substack{\boldsymbol{x},\mathbf{y}\in B_{nb}(\mathbf{0})-B_1(\mathbf{0})\\ m\le|\boldsymbol{x}-\mathbf{y}|\le\delta nb}}\frac{c_{\boldsymbol{h}}^2(\boldsymbol{t},\boldsymbol{x}-\mathbf{y})}{g^2(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h}))}|\boldsymbol{x}-\mathbf{y}|^{-d}d\boldsymbol{x}d\mathbf{y}\\
&=\lim_{\delta\to 0}\lim_{n\to\infty}\frac{f^2(\mathbf{0})}{(nb)^d\log(nb)}\int_{\boldsymbol{x}\in B_{nb}(\mathbf{0})-B_1(\mathbf{0})}\int_{m\le|\mathbf{y}|\le\delta nb}\frac{c_{\boldsymbol{h}}^2(\boldsymbol{t},\mathbf{y})}{g^2(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h}))}|\mathbf{y}|^{-d}d\mathbf{y}d\boldsymbol{x}\\
&=\lim_{\delta\to 0}\lim_{n\to\infty}\frac{V_d f^2(\mathbf{0})}{\log(nb)}\int_{[0,2\pi]^d}\int_m^{\delta nb}\frac{c_{\boldsymbol{h}}^2(\boldsymbol{t},\boldsymbol{\Theta})}{g^2(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h}))}|r|^{-1}\prod_{i=1}^{d-2}\sin^{d-i-1}\theta_i dr d\boldsymbol{\theta}\\
&=V_d f^2(\mathbf{0})\int_{[0,2\pi]^d}\frac{c_{\boldsymbol{h}}^2(\boldsymbol{t},\boldsymbol{\Theta})}{g^2(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h}))}\prod_{i=1}^{d-2}\sin^{d-i-1}\theta_i d\boldsymbol{\theta}\\
&=V_d f^2(\mathbf{0})\int_0^1\int_{[0,2\pi]^d}\frac{c_{\boldsymbol{h}}^2(\boldsymbol{t},\boldsymbol{\Theta})}{g^2(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h}))}\prod_{i=1}^{d-2}\sin^{d-i-1}\theta_i d\boldsymbol{\theta}dr\\
&=\frac{V_d f^2(\mathbf{0})}{g^2(H(\boldsymbol{t},\mathbf{0},\boldsymbol{h}))}\int_{B_1(\mathbf{0})}\frac{c_{\boldsymbol{h}}^2(\boldsymbol{t},\boldsymbol{x})}{|\boldsymbol{x}|^{d-1}}d\boldsymbol{x},
\end{aligned}$$

where the change of variables on the second line uses the properties that $c_{\boldsymbol{h}}^2(\boldsymbol{t},\boldsymbol{x})$ does not depend on $|\boldsymbol{x}|$, as indicated by Lemma 3.8.11. This concludes the proof of the second part of the lemma.

For the third part of the lemma, by (3.77), we have for all n distinct $\boldsymbol{t_i}, \boldsymbol{t_j} \in B_b(\boldsymbol{t})$,

$$|r_n(\boldsymbol{t_i},\boldsymbol{t_j})| \le C(\boldsymbol{t},\boldsymbol{h})|\boldsymbol{i}-\boldsymbol{j}|^{-\psi(\boldsymbol{t})}.$$

Let

$$\begin{aligned}
\epsilon''_{n,m,1} = d_n^{-k}&\sum_{\text{distinct } \boldsymbol{t}_{\boldsymbol{i}_1},\ldots,\boldsymbol{t}_{\boldsymbol{i}_k}\in\Omega_n}\prod_{l=1}^{k} r_n\left(\boldsymbol{t}_{\boldsymbol{i}_l},\boldsymbol{t}_{\boldsymbol{i}_{l+1}}\right) f\left(\frac{\boldsymbol{t}_{\boldsymbol{i}_l}-\boldsymbol{t}}{b}\right)\\
&- d_n^{-k}\sum_{\text{distinct } \boldsymbol{t}_{\boldsymbol{i}_1},\ldots,\boldsymbol{t}_{\boldsymbol{i}_k}\in\Omega_n,|\boldsymbol{i}_s-\boldsymbol{i}_t|>m}\prod_{l=1}^{k} r_n\left(\boldsymbol{t}_{\boldsymbol{i}_l},\boldsymbol{t}_{\boldsymbol{i}_{l+1}}\right) f\left(\frac{\boldsymbol{t}_{\boldsymbol{i}_l}-\boldsymbol{t}}{b}\right),
\end{aligned}$$

$$\epsilon''_{n,m,2} = d_n^{-k} \sum_{\text{distinct } \boldsymbol{t}_{\boldsymbol{i}_1},\ldots,\boldsymbol{t}_{\boldsymbol{i}_k}\in\Omega_n, |\boldsymbol{i}_s-\boldsymbol{i}_t|>m} \prod_{l=1}^{k} r_n\left(\boldsymbol{t}_{\boldsymbol{i}_l}, \boldsymbol{t}_{\boldsymbol{i}_{l+1}}\right) f\left(\frac{\boldsymbol{t}_{\boldsymbol{i}_l}-\boldsymbol{t}}{b}\right)$$
$$- d_n^{-k} \sum_{\text{distinct } \boldsymbol{i}_1,\ldots,\boldsymbol{i}_k\in\Omega_n, |\boldsymbol{i}_s-\boldsymbol{i}_t|>m} \frac{\prod_{l=1}^{k} c_{\boldsymbol{h}}(\boldsymbol{t}, \boldsymbol{i}_l-\boldsymbol{i}_{l+1})|\boldsymbol{i}_l-\boldsymbol{i}_{l+1}|^{-\psi(\boldsymbol{t})} f(\boldsymbol{i}_l)}{g^k(H(\boldsymbol{t}), \boldsymbol{0}, \boldsymbol{h})},$$

$$\epsilon''_{n,m,3} = d_n^{-k} \sum_{\text{distinct } \boldsymbol{i}_1,\ldots,\boldsymbol{i}_k\in\Omega_n, |\boldsymbol{i}_s-\boldsymbol{i}_t|>m} \frac{\prod_{l=1}^{k} c_{\boldsymbol{h}}(\boldsymbol{t}, \boldsymbol{i}_l-\boldsymbol{i}_{l+1})|\boldsymbol{i}_l-\boldsymbol{i}_{l+1}|^{-\psi(\boldsymbol{t})} f(\boldsymbol{i}_l)}{g^k(H(\boldsymbol{t}), \boldsymbol{0}, \boldsymbol{h})}$$
$$- \int_{(B_1(\boldsymbol{0}))^{\otimes k}} \frac{\prod_{i=1}^{k} c_{\boldsymbol{h}}(\boldsymbol{t}, \mathbf{z}_i-\mathbf{z}_{i+1})|\mathbf{z}_i-\mathbf{z}_{i+1}|^{-\psi(\boldsymbol{t})} f(\mathbf{z}_i)}{g^k(H(\boldsymbol{t}), \boldsymbol{0}, \boldsymbol{h})} d\mathbf{z}_1 \ldots d\mathbf{z}_k.$$

We will prove

$$\lim_{m\to\infty}\lim_{n\to\infty} |\epsilon''_{n,m,i}| = 0, i = 1, 2, 3. \tag{3.80}$$

It follows that

$$\begin{aligned}|\epsilon''_{n,m,1}| &\le km^d V_d C^k(\boldsymbol{t}, \boldsymbol{h}) \sup\left[\frac{f^2 c_{\boldsymbol{h}}}{g(H(\boldsymbol{t}, \boldsymbol{0}, \boldsymbol{h}))}\right]^k \times \\ &\qquad d_n^{-k} \sum_{\boldsymbol{i}_1,\ldots,\boldsymbol{i}_k\in\Omega_{\boldsymbol{t}}} 1^{-\psi(\boldsymbol{t})}|\boldsymbol{i}_2-\boldsymbol{i}_3|^{-\psi(\boldsymbol{t})}\ldots|\boldsymbol{i}_k-\boldsymbol{i}_1|^{-\psi(\boldsymbol{t})} \\ &= O(d_n^{-1}) \to 0.\end{aligned}$$

That (3.80) holds for $i = 2$ is a direct result of (3.77). That (3.80) holds for $i = 3$ follows from arguments similar to those in the proofs above regarding $\epsilon'_{n,m,5}$ and $\epsilon'_{n,m,6}$, and we omit the details here. □

Lemma 3.8.15. *Assume [A1], [A2], [A3] and f to be a bounded function with support on $B_1(\boldsymbol{0})$. Then for $k \ge 2$ and uniformly for all $\boldsymbol{t} \in \Omega_\delta$, we have*

$$\sum_{\boldsymbol{i},\boldsymbol{j}} f\left(\frac{\boldsymbol{t_i}-\boldsymbol{t}}{b}\right) f\left(\frac{\boldsymbol{t_j}-\boldsymbol{t}}{b}\right) r_n^k(\boldsymbol{t}_i, \boldsymbol{t}_j) \le \begin{cases}(nb)^{2d-2\psi(\boldsymbol{t})}C(f,\delta), & \text{if } 2\psi(\boldsymbol{t}) < d,\\ (nb)^d \log(nb) C(f,\delta), & \text{if } 2\psi(\boldsymbol{t}) = d,\\ (nb)^d C(f,\delta), & \text{if } 2\psi(\boldsymbol{t}) > d,\end{cases}$$

where $C(f, \delta)$ depends only on δ and f.

Proof. By the second part of Theorem 3.8.4 and Lemma 3.8.10, we have uniformly for

$\boldsymbol{t} \in \Omega_\delta$ and $|\boldsymbol{j}| \leq nb$,

$$r_n\left(\boldsymbol{t}, \boldsymbol{t} + \frac{\boldsymbol{j}}{n}; \boldsymbol{h}\right) \leq C|\boldsymbol{j}|^{-\psi(\boldsymbol{t})}. \tag{3.81}$$

For $|\boldsymbol{t}_i - \boldsymbol{t}| < b$, $|\boldsymbol{j}| \leq nb$, as $H(\boldsymbol{t})$ has bounded first order derivative and $n^b \to 1$, we know that when n is large enough, we have

$$r_n\left(\boldsymbol{t}_i, \boldsymbol{t}_i + \frac{\boldsymbol{j}}{n}; \boldsymbol{h}\right) \leq C|\boldsymbol{j}|^{-\psi(\boldsymbol{t}_i)} \leq C|\boldsymbol{j}|^{-\psi(\boldsymbol{t})}|nb|^{Mb} \leq C'|\boldsymbol{j}|^{-\psi(\boldsymbol{t})},$$

where $\sup_{\boldsymbol{t}} |H'(\boldsymbol{t})| \leq M$. Thus,

$$\begin{aligned}
&\sum_{\boldsymbol{i},\boldsymbol{j}} f\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) f\left(\frac{\boldsymbol{t}_j - \boldsymbol{t}}{b}\right) r_n^k(\boldsymbol{t}_i, \boldsymbol{t}_j) \\
&\leq \sup_{\mathbf{z}} |f(\mathbf{z})|^2 \sum_{\boldsymbol{i},\boldsymbol{j}} r_n^2(\boldsymbol{t}_i, \boldsymbol{t}_j) \\
&\leq \sup_{\mathbf{z}} |f(\mathbf{z})|^2 \sum_{\boldsymbol{i},\boldsymbol{j}} (C')^2 |\boldsymbol{j}|^{-2\psi(\boldsymbol{t})} \\
&\leq \begin{cases} (nb)^d C(f,\delta), & \text{if } 2\psi(\boldsymbol{t}) > d, \\ (nb)^d \log(nb) C(f,\delta), & \text{if } 2\psi(\boldsymbol{t}) = d, \\ (nb)^{2d-2\psi(\boldsymbol{t})} C(f,\delta), & \text{if } 2\psi(\boldsymbol{t}) < d. \end{cases}
\end{aligned}$$

The proof is complete. □

The following is a well-known result (cf. Breuer and Major (1983)) needed in some parts of the proofs below.

Lemma 3.8.16. *If ξ and η are jointly Gaussian random variables, $\mathbb{E}\xi = \mathbb{E}\eta = 0$, $\mathbb{E}\xi^2 = \mathbb{E}\eta^2 = 1$, $\mathbb{E}\xi\eta = r$, then*

$$\mathbb{E}H_k(\xi)H_l(\tau) = \delta(k,l)r^k k!,$$

where δ denotes the Kronecker delta.

PROOF for Lemma 3.8.7:

Define $Y_n(\boldsymbol{t}, \boldsymbol{h})$ and r_n as in Lemma 3.8.14. For all $\mathbf{a}$ with $|\mathbf{a}| = 1$,

$$\mathbf{a}^T(\mathbf{Z}(\boldsymbol{t}, n) - \mathbb{E}\mathbf{Z}(\boldsymbol{t}, n)) = (nb)^{-\frac{d}{2}} \sum_i K\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right) F(Y_n(\boldsymbol{t}_i, \boldsymbol{h}))\mathbf{a}^T\mathbf{A}\left(\frac{\boldsymbol{t}_i - \boldsymbol{t}}{b}\right).$$

where F is as defined in (3.17). Denote $f_{\mathbf{a}}(\boldsymbol{x}) = K(\boldsymbol{t})\mathbf{a}^T\mathbf{A}(\boldsymbol{t})$, which is a bounded function with a second derivative, and by (3.72) we can rewrite $\mathbf{a}^T(\mathbf{Z}(\boldsymbol{t},n) - \mathbb{E}[\mathbf{Z}(\boldsymbol{t},n)])$ as

$$\mathbf{a}^T(\mathbf{Z}(\boldsymbol{t},n) - \mathbb{E}[\mathbf{Z}(\boldsymbol{t},n)]) = (nb)^{-\frac{d}{2}} \left\{ \sum_{\boldsymbol{i}\in\Omega_{\boldsymbol{t}}} f_{\mathbf{a}}\left(\frac{\boldsymbol{i}}{nb}\right) F(Y_n(\boldsymbol{t}_i,\boldsymbol{h})) \right\}.$$

Denote

$$\sigma^2_{l,f}(\boldsymbol{t}) = \int_{D_{\boldsymbol{t},b}} f^2(\mathbf{z})d\mathbf{z} \sum_{\boldsymbol{j}\in\mathbb{Z}^d} \left(\frac{g(H(\boldsymbol{t}),\boldsymbol{j},\boldsymbol{h})}{g(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h})}\right)^l, l \geq 2,$$

and

$$\sigma^2_f(\boldsymbol{t}) = \sum_{l=2}^{\infty} c_l^2 l! \sigma^2_{l,f}(\boldsymbol{t}) = \sigma^2_{H(\boldsymbol{t})} \int_{D_{\boldsymbol{t},b}} f^2(\mathbf{z})d\mathbf{z}.$$

By derivations similar to the proof of Theorem 1 in Breuer and Major (1983) (changing all $r(.)$ in their proof to $r_n(.)$; using (3.74) when proving convergence to 0; using Lemma 3.8.14 when proving the limit of summation of $r_n(.)$; distributing $f_{\boldsymbol{a}}(.)$ properly in the proof), we get

$$\mathbf{a}^T\mathbf{Z}(\boldsymbol{t},n) \xrightarrow{d} N\left(0, \sigma^2_{f_{\mathbf{a}}}(\boldsymbol{t})\right).$$

As $\mathbf{a}$ is arbitrary, we have

$$\mathbf{Z}(\boldsymbol{t},n) \xrightarrow{d} N\left(\mathbf{0}, \sigma^2_{H(\boldsymbol{t})} \int_{D_{\boldsymbol{t},b}} \mathbf{f}(\mathbf{z})\mathbf{f}^T(\mathbf{z})d\mathbf{z}\right).$$

Similarly, following Theorem 1' in Breuer and Major (1983), we prove when $2\psi(\boldsymbol{t}) = d$,

$$\log^{-\frac{1}{2}}(nb)\mathbf{Z}(\boldsymbol{t},n) \xrightarrow{d} N(\mathbf{0},\boldsymbol{\Sigma}),$$

where

$$\boldsymbol{\Sigma} = \frac{V_d \mathbf{f}(\mathbf{0})\mathbf{f}^T(\mathbf{0})}{g^2(H(\boldsymbol{t}),\mathbf{0},\boldsymbol{h})} \int_{|\boldsymbol{x}|=1} c^2_{\boldsymbol{h}}(\boldsymbol{t},\boldsymbol{x})d\boldsymbol{x}.$$

When $2\psi(\boldsymbol{t}) < d$, following the proof of Proposition 6.1 in Taqqu (1975) and changing the matrix $D_N(1)$ in that proof to a matrix whose diagonal elements are $f(\frac{\boldsymbol{i}}{nb})$ and conducting some other corresponding modifications, we can establish the convergence of the

characteristic function of $(nb)^{-\frac{d}{2}+\psi(\boldsymbol{t})}\mathbf{a}^T\mathbf{Z}$ to

$$\phi(u) = \exp\left\{\frac{1}{2}\sum_{k=2}^{\infty}\frac{(2iu)^k}{k}S_k\right\}.$$

To establish the uniform bound of the bias, as $\mathbf{A}\left(\frac{\boldsymbol{t}_i-\boldsymbol{t}}{b}\right)$ and $K\left(\frac{\boldsymbol{t}_i-\boldsymbol{t}}{b}\right)$ are bounded, we have uniformly for $\boldsymbol{t}\in\Omega_\delta$

$$\mathbb{E}[\mathbf{Z}(\boldsymbol{t},n)] = O(2\mathbb{E}[\log|W_n(t_i)| - \widetilde{g}(t_i)]) = O((nb)^{\frac{d}{2}}\rho_n).$$

For the variance, it follows from Lemma 3.8.16 that

$$\begin{aligned}
&(nb)^d\mathbb{E}\left(\mathbf{a}^T\mathbf{Z}(\boldsymbol{t},n) - \mathbf{a}^T\mathbb{E}[\mathbf{Z}(\boldsymbol{t},n)]\right)^2\\
&= \mathbb{E}\left(\sum_{\boldsymbol{i}\in\Omega_{\boldsymbol{t}}} f_{\mathbf{a}}\left(\frac{\boldsymbol{i}}{nb}\right)F(Y_n(\boldsymbol{t}_i,\boldsymbol{h}))\right)^2\\
&= \sum_{\boldsymbol{i},\boldsymbol{j}\in\Omega_{\boldsymbol{t}}} f_{\mathbf{a}}\left(\frac{\boldsymbol{i}}{nb}\right)f_{\mathbf{a}}\left(\frac{\boldsymbol{j}}{nb}\right)\sum_{l=2}^{\infty}c_l^2 l! r_n^l(t_i,t_j),
\end{aligned}$$

which, by Lemma 3.8.15, is bounded by

$$\begin{aligned}
&\sup_{\mathbf{z}}|f_{\mathbf{a}}(\mathbf{z})|^2\sum_{l=2}^{\infty}c_l^2 l!\sum_{\boldsymbol{i},\boldsymbol{j}\in\Omega_{\boldsymbol{t}}} r_n^2(t_i,t_j)\\
&\le \begin{cases} C_{\delta,\mathbf{a}}(nb)^d, & \text{if } 2\psi(\boldsymbol{t}) > d,\\ C_{\delta,\mathbf{a}}(nb)^d\log(nb), & \text{if } 2\psi(\boldsymbol{t}) = d\\ C_{\delta,\mathbf{a}}(nb)^{2d-2\psi(\boldsymbol{t})}, & \text{if } 2\psi(\boldsymbol{t}) < d. \end{cases}
\end{aligned}$$

□

3.8.7 Proofs for Section 3.7.1

This section provides theoretical support for Section 3.7.1, which considers the non-grid data setting of Loh (2015). Let condition [B] and $\Delta_{1/n}^q\mathfrak{X}(t)$ be as defined in Section 3.7.1. The main result of this section is the following proposition, which serves the same role as that of Theorem 3.8.4 for the gridded case.

Proposition 3.8.17. *In the following let $t, t+u/n \in \Omega_\delta$. Under assumptions [A1]-[A3] and [B], for any $\delta\in(0,1)$ and $M>0$, there exist finite constants $C_\delta, C_{\delta,M}$ such that uniformly in n,t,*

(i) $|\mathrm{Cov}(\Delta_{1/n}^q \mathfrak{X}(t), \Delta_{1/n}^q \mathfrak{X}(t+u/n)) - g(H(t), u, \varpi(t))| \le C_{\delta,M}\rho_n(t)$ *for* $|u| \le M$*;*

(ii) $|\mathrm{Cov}(\Delta_{1/n}^q \mathfrak{X}(t), \Delta_{1/n}^q \mathfrak{X}(t+u/n))| \le C_\delta |u|^{-\psi(t)}$ *for* u *satisfying* $q+1 < |u| < 2bn$.

(iii) $|\mathrm{Cov}(\Delta_{1/n}^q \mathfrak{X}(t), \Delta_{1/n}^q \mathfrak{X}(t+u/n))| \le C_\delta |u|^{-\bar{\psi}}$ *for* u *satisfying* $|u| > q+1$.

Proof. We only provide a sketch here since the arguments are similar to those in the proof of Theorem 3.8.4. Define

$$f'(\eta_1, \eta_2; t, s, h) = \frac{\partial^{2q}}{\partial \eta_1^q \partial \eta_2^q} \left(C(\varphi(\tilde{t} + \eta_1 h), \varphi(\tilde{s} + \eta_2 h)) \right)$$

and

$$\begin{aligned} \tilde{f}'(\eta_1, \eta_2; t, s, h) &= \frac{\partial^{2q}}{\partial \eta_1^q \partial \eta_2^q} \Bigg(|\varphi(\tilde{t} + \eta_1 h)|^{H\circ\varphi(\tilde{t}+\eta_1 h) + H\circ\varphi(\tilde{s}+\eta_2 h)} \\ &\quad \times \mathscr{D}(H \circ \varphi(\tilde{t} + \eta_1 h), H \circ \varphi(\tilde{s} + \eta_2 h)) \Bigg), \end{aligned}$$

where $\tilde{t} = \varphi^{-1}(t)$ and $\tilde{s} = \varphi^{-1}(s)$. As $t \in \Omega_\delta$ and the derivative of φ is bounded from 0 and ∞, $\tilde{t}$ will also be bounded from 0.

By the same argument used in Lemma 3.8.12, we can establish the following: for large n, there is some finite constant C_δ such that

$$f'(0, 0; t, s, h) \le C_\delta |t - s|^{-2q + H(s) + H(t)} \tag{3.82}$$

and

$$\widetilde{f}'(0, 0; t, s, h) \le C_\delta \tag{3.83}$$

uniformly for all $s, t \in \Omega_\delta$. Denote $x_i = x_i(t)$ and $x_i' = x_i(t + u/n)$ for simplicity. As in the proof of Theorem 3.8.4, write

$$\begin{aligned} & \mathrm{Cov}(\Delta_{1/n}^q \mathfrak{X}(t), \Delta_{1/n}^q \mathfrak{X}(t+u/n)) \\ = \; & n^{H(t)+H(t+\frac{u}{n})} \sum_{i,j=0}^{q} a_i a_j \mathscr{D}(H(x_i), H(x_j')) \Big(|x_i|^{H(x_i)+H(x_j')} \\ & + |x_j'|^{H(x_i)+H(x_j')} - |x_i - x_j'|^{H(x_i)+H(x_j')} \Big) \\ =: \; & n^{H(t)+H(t+\frac{u}{n})} (A_2' + A_3' + A_1') \end{aligned}$$

where A_1', A_2', A_3' are defined in a similar way as A_1, A_2, A_3. It follows that both $n^{H(t)+H(t+\frac{u}{n})} A_2'$

and $n^{H(t)+H(t+\frac{u}{n})}A_3'$ are $O(n^{-\psi(t)})$, which are exactly the same as what was established for A_2 and A_3 in the proof of Theorem 3.8.4. With bounded u, we have for each term in A_1'

$$\begin{aligned}
& n^{H(t)+H(t+\frac{u}{n})}\mathscr{D}(H(x_i),H(x_j'))|x_i-x_j'|^{H(x_i)+H(x_j')} \\
= \;& n^{2H(t)+O(n^{-1})}\left(\frac{1}{2}+O(n^{-1})\right)\left|\frac{\varpi(t)(i-j)+u}{n}+O(n^{-2})\right|^{2H(t)+O(n^{-1})} \\
= \;& n^{O(n^{-1})}\left(\frac{1}{2}+O(n^{-1})\right)\left|\varpi(t)(i-j)+u+O(n^{-1})\right|^{2H(t)+O(n^{-1})} \\
= \;& \frac{1}{2}|\varpi(t)(i-j)+u|^{2H(t)}+O(n^{-1}\log n).
\end{aligned}$$

Therefore,

$$|A_1'-g(H(t),u,\varpi(t))|=O(n^{-1}\log n).$$

Combining the above, (i) of the proposition is proved.

Applying (3.82) and (3.83), the proofs for (ii) and (iii) are essentially the same as those for the corresponding parts in Theorem 3.8.4, and are omitted here. □

Applying Proposition 3.8.17, we obtain

$$2\mathbb{E}\log\left|\Delta_{1/n}^q\mathfrak{X}(t)\right|=G(H(t);n,\varpi(t))+O(\rho_n(t)). \tag{3.84}$$

Proposition 3.8.17 together with (3.84) and the fact that $\sup_t|\widehat{\varpi}(t)-\varpi(t)|=O(n^{-1})$ show that the rates of convergence of the estimators $\widehat{H}'$, $\widehat{H}_1'$ and $\widehat{H}_2'$ are the same as those for $\widehat{H}$, $\widehat{H}_1$ and $\widehat{H}_2$.

We conducted a simulation study for which the results are presented in Figure 3.6. We assume $\varphi(x)=(x^2+0.5x)/1.5$; note that $\varphi'(1)=5\varphi'(0)$ so that the sizes of the gaps in the data do vary quite a bit from location to location. For $\widehat{H}'$ and $\widehat{H}_2'$, we used a naive estimator $\widehat{\varpi}(t)$ of $\varpi(t)$ where $\widehat{\varpi}(t)=\sum_i n(t_{i+1}-t_i)I_{(t_i,t_{i+1}]}(t)$. Figure 3.6 below presents the simulation results of $\log(\text{MISE})$ versus sample size for $\widehat{H}'$, $\widehat{H}_1'$ and $\widehat{H}_2'$ based on non-grid data and those for $\widehat{H}$, $\widehat{H}_1$ and $\widehat{H}_2$ based on gridded data. As can be seen, the rates for $\widehat{H}'$, $\widehat{H}_1'$ and $\widehat{H}_2'$ are somewhat but not substantially worse than those for $\widehat{H}$, $\widehat{H}_1$ and $\widehat{H}_2$.

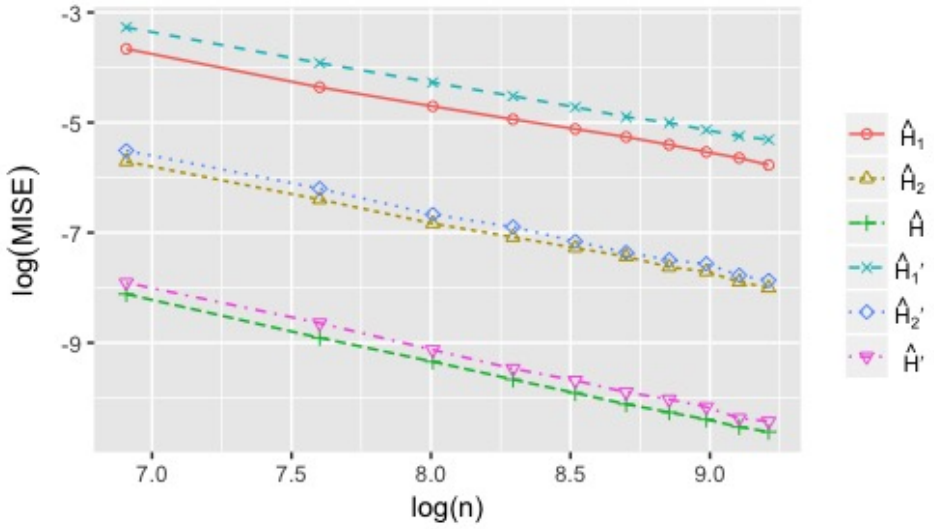


Figure 3.6: Plot of $\log(\mathrm{MISE})$ versus $\log n$

3.8.8 Proofs for Section 3.7.2 and additional discussions for non-constant variance

Define

$$\begin{aligned} W_{n,X}(\boldsymbol{t},\boldsymbol{h}) &:= n^{H(\boldsymbol{t})}\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t}),\\ W_{n,Y}(\boldsymbol{t},\boldsymbol{h}) &:= \sigma^{-1}(\boldsymbol{t})n^{H(\boldsymbol{t})}\Delta^q_{\boldsymbol{h}/n}Y(\boldsymbol{t}),\\ C_{n,X}(\boldsymbol{t},\boldsymbol{s},\boldsymbol{h}) &:= \mathrm{Cov}\left(W_{n,X}(\boldsymbol{t},\boldsymbol{h}), W_{n,X}(\boldsymbol{s},\boldsymbol{h})\right),\\ C_{n,Y}(\boldsymbol{t},\boldsymbol{s},\boldsymbol{h}) &:= \mathrm{Cov}\left(W_{n,Y}(\boldsymbol{t},\boldsymbol{h}), W_{n,Y}(\boldsymbol{s},\boldsymbol{h})\right). \end{aligned}$$

Lemma 3.8.18. *In the following let $\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{u}/n \in \Omega_\delta$. Under assumptions [A1]-[A3], [S] and $q \geq 2$, for any $\delta \in (0,1)$ and $M > 0$, there exist finite constants $C_\delta, C_{\delta,M}$ such that uniformly in $n, \boldsymbol{t}$,*

(i) $|C_{n,Y}(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{u}/n, \boldsymbol{h}) - g(H(\boldsymbol{t}), \boldsymbol{u}, \boldsymbol{h})| \leq C_{\delta,M}\rho_n(\boldsymbol{t})$ *for* $|\boldsymbol{u}| \leq M$*;*

(ii) $|C_{n,Y}(\boldsymbol{t}, \boldsymbol{t}+\boldsymbol{u}/n, \boldsymbol{h})| \leq C_\delta(|\boldsymbol{u}|^{-\psi(\boldsymbol{t})} + n^{-1}|\boldsymbol{u}|^{-(q+1)+2H(\boldsymbol{t})} + n^{-2})$ *for* $\boldsymbol{u}$ *satisfying* $q|\boldsymbol{h}| + 1 < |\boldsymbol{u}| < 2bn$.

Proof. By induction, we have

$$\begin{aligned} \Delta^q_{\boldsymbol{h}/n}Y(\boldsymbol{t}) &= \sum_{l=0}^{q}\binom{q}{l}\Delta^l_{\boldsymbol{h}/n}X(\boldsymbol{t})\Delta^{q-l}_{\boldsymbol{h}/n}\sigma(\boldsymbol{t}+l\boldsymbol{h}/n)\\ &:= \sigma(\boldsymbol{t})\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t}) + \epsilon(\boldsymbol{t}). \end{aligned}$$

Uniformly, we have $\text{Var}(\Delta^l_{\boldsymbol{h}/n}X(\boldsymbol{t})) = O(n^{-2H(\boldsymbol{t})}), l \geq 1$, and $\Delta^l_{\boldsymbol{h}/n}\sigma(\boldsymbol{t}) = O(n^{-2}), l \geq 2$. Thus, for any $q \geq 2$ and $l < q$,

$$\begin{aligned}
&\text{Var}(\Delta^l_{\boldsymbol{h}/n}X(\boldsymbol{t})\Delta^{q-l}_{\boldsymbol{h}/n}\sigma(\boldsymbol{t} + l\boldsymbol{h}/n)) \\
&= O(n^{-2H(\boldsymbol{t})I_{(l\neq 0)}-2\min(2,q-l)}) \\
&= O(n^{-2H(\boldsymbol{t})-2})
\end{aligned}$$

uniformly. Therefore,

$$|C_{n,Y}(\boldsymbol{t}, \boldsymbol{t} + \boldsymbol{u}/n, \boldsymbol{h}) - C_{n,X}(\boldsymbol{t}, \boldsymbol{t} + \boldsymbol{u}/n, \boldsymbol{h})| = O(n^{-1})$$

uniformly. This equation and (i) in Theorem 3.8.4 establish (i) in Lemma 3.8.18.

To prove (ii), we consider in detail the following expression

$$\text{Cov}(\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t}), \Delta^l_{\boldsymbol{h}/n}X(\boldsymbol{t} + \boldsymbol{u}/n)), 0 \leq l \leq q.$$

Define

$$f_{q,l}(\phi, \eta; \boldsymbol{t}, \boldsymbol{s}, \boldsymbol{h}) = \frac{\partial^{q+l}}{\partial\phi^q\partial\eta^l}C(\boldsymbol{t} + \phi\boldsymbol{h}, \boldsymbol{s} + \eta\boldsymbol{h}).$$

As $q \geq 2$, following the same steps as in Lemma 3.8.12, it can be shown that when n is large enough we have uniformly for $\boldsymbol{t}, \boldsymbol{s} \in \Omega_\delta$,

$$f_{q,l}(0, 0; \boldsymbol{t}, \boldsymbol{s}, \boldsymbol{h}) \leq C_\delta|\boldsymbol{t} - \boldsymbol{s}|^{-(q+l)+H(\boldsymbol{s})+H(\boldsymbol{t})}.$$

Then by the same steps in the proof of (ii) of Theorem 3.8.4, we have uniformly for $0 \leq l \leq q$, $q|\boldsymbol{h}| + 1 < |\boldsymbol{u}| < 2bn$,

$$|n^{2H(\boldsymbol{t})}\text{Cov}(\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t}), \Delta^l_{\boldsymbol{h}/n}X(\boldsymbol{t} + \boldsymbol{u}/n))| \leq C_\delta|\boldsymbol{u}|^{-(q+l)+2H(\boldsymbol{t})}.$$

Thus,

$$|n^{2H(\boldsymbol{t})}\text{Cov}(\Delta^q_{\boldsymbol{h}/n}X(\boldsymbol{t}), \epsilon(\boldsymbol{t} + \boldsymbol{u}/n)| \leq C_\delta|\boldsymbol{u}|^{-(q+1)+2H(\boldsymbol{t})}n^{-1}. \tag{3.85}$$

Combining (3.85), part (ii) of Theorem 3.8.4 and the fact that $\text{Var}(\epsilon(\boldsymbol{t})) = O(n^{-2-2H(\boldsymbol{t})})$, part (ii) in Lemma 3.8.18 is proved. □

PROOF for Proposition 3.7.1:

The proof for bias follows exactly as that for bias in Theorem 3.3.2 using (i) in Lemma 3.8.18. For the variance, using the notations in Lemma 3.8.15, we only need to demonstrate $\sum_{\boldsymbol{i},\boldsymbol{j}} f\left(\frac{\boldsymbol{t}_i-\boldsymbol{t}}{b}\right) f\left(\frac{\boldsymbol{t}_j-\boldsymbol{t}}{b}\right) r_n^2(\boldsymbol{t}_i,\boldsymbol{t}_j) = O((nb)^d)$, which follows from (ii) of Lemma 3.8.18 and $q \geq 2, d \leq 3$.

□

With Proposition 3.7.1 and arguments similar to those in Theorem 3.4.1, we can prove that the convergence rate of $\widehat{H}_1$ based on $Y(\boldsymbol{t})$ is the same as those based on $X(\boldsymbol{t})$ when $d \leq 3$ and $q \geq 2$, namely, $n^{-\frac{dp}{2p+d}}$. However, H_1 does not achieve the minimax rate $(n^d \log^2 n)^{-\frac{p}{2p+d}}$. One possible way to improve this is to adopt a modified version of $\widehat{H}_2$, as follows:

Step 1: estimate $H(\boldsymbol{t})$ by $\widehat{H}_1$,
step 2: based on $\widehat{H}_1$, estimate $\sigma(\boldsymbol{t})$ nonparametrically by some $\widehat{\sigma}(\boldsymbol{t})$, and
step 3: with $\widehat{\sigma}(\boldsymbol{t})$ replacing true $\sigma(\boldsymbol{t})$, estimate $H(\boldsymbol{t})$ by $\widehat{H}(\boldsymbol{t})$.

To verify the rate of this approach, we present the following intuitive arguments. We focus on the case $d = 1$. Suppose that we first apply $\widehat{H}_1$ with $p_1, q_1 \geq 2$ and b_1. We have

$$\begin{aligned}
& \log\left(n^{\widehat{H}_1(t_i)}\Delta_{1/n}Y(t_i)\right)^2 \\
= \; & \log\sigma^2(t_i) + \log\left(n^{\widehat{H}_1(t_i)}\Delta_{1/n}X(t_i)\right)^2 + O_p(n^{-2+2H(t)}) \\
= \; & \log\sigma^2(t_i) + 2(\widehat{H}_1(t_i) - H(t_i))\log n + \log\chi_{1,i}^2 \\
& + o_p(n^{-1-\delta_1}) + O(\rho_n(t_i)) \\
= \; & \log\sigma^2(t_i) + 2(\widehat{H}_1(t_i) - H(t_i))\log n + \epsilon_i + \mathbb{E}[\log\chi_1^2] + o_p(n^{-1-\delta_1})
\end{aligned}$$

where the third line uses (i) in Theorem 3.8.4 with $\delta_1 > 0$ being a small constant,

$$\chi_{1,i}^2 = \left(\Delta_{1/n}X(t_i)\right)^2 / \sigma^2(\Delta_{1/n}X(t_i))$$

and

$$\epsilon_i = \log\chi_{1,i}^2 - \mathbb{E}[\log\chi_1^2].$$

Intuitively, one can estimate $\log\sigma^2(t)$ by conducting another local polynomial regression esstimate based on the data $\log\left(n^{\widehat{H}_1(t_i)}\Delta_{1/n}Y(t_i)\right)^2 - \mathbb{E}[\log\chi_1^2]$, with parameters p_2, b_2. We first consider the variance of $\widehat{\log\sigma^2(t)}$. Note that by (iii) of Theorem 3.8.4 and the fact that ϵ_i is a function of a Gaussian variable with Hermite rank 2, if $H(t) < 0.5$ uniformly, we would have $\mathrm{Cov}(\epsilon_i, \epsilon_j) = O(|i-j|^{-2\bar{\psi}}) = o(|i-j|^{-2})$. According to Hall and

Hart (1990), the variance contributed by ϵ_i is $O(nb_2)^{-1}$. Also, by (3.63) and the ensuing arguments (replacing m by nb_2 or n appropriately), the variance contributed by $\widehat{H}_1(t_i)$ can be calculated to be $O((nb_2)^{-1}\log^2 n)$. The bias should be $O(b_1^{p_1}+b_2^{p_2})$ since both steps are local polynomial regression. Therefore, the convergence rate of $\widehat{\log\sigma^2(t)}$ will only be different from $O\left(n^{-\frac{p_2}{2p_2+1}}\right)$ by a log term and the estimator would favor a small b_1. Theoretically, if $\sigma(t)$ has higher order smoothness than $H(t)$ and H stays close to the interval $(0,1/2)$ so that we can estimate $\sigma(t)$ at a rate faster than the minimax rate, then we conjecture that the estimated $\sigma(t)$ could be used in place of the true $\sigma(t)$ for the inversion that would then lead to the minimax rate for estimating $H(t)$.

One may also consider higher-order differencing for $\widehat{\log\sigma^2(t)}$; that is

$$\begin{aligned}
& \log\left(n^{\widehat{H}_1(t_i)}\Delta_{1/n}^{q_2}Y(t_i)\right)^2 \\
= \;& \log\sigma^2(t_i)+\log\left(n^{\widehat{H}_1(t_i)}\Delta_{1/n}^{q_2}X(t_i)\right)^2+O_p(n^{-1-\delta}) \\
= \;& \log\sigma^2(t_i)+2(\widehat{H}_1(t_i)-H(t_i))\log n+\epsilon_i+\mathbb{E}[\log\chi_1^2] \\
& +g(H(t_i),0,1)+O_p(n^{-1-\delta}),
\end{aligned}$$

and the local polynomial regression based on data

$$\log\left(n^{\widehat{H}_1(t_i)}\Delta_{1/n}^{q}Y(t_i)\right)^2-\mathbb{E}[\log\chi_1^2]-g(\widehat{H}_1(t_i),0,1).$$

This is plausible but to analyze its properties, a much tighter version of Lemma 3.8.9 is needed for analyzing the variance contributed by $\widehat{H}_1(t_i)$ in $g(\widehat{H}_1(t_i),0,1)$. These ideas will be pursued in future work.

We have done some simulations for this situation with $d=1$ and $q_2=1$, $H(t)=0.25+0.2\sin(4\pi t)$ and $\sigma(t)=(\sin(3\pi t)+2)^2, t\in(0,1)$. Note that $H(t)\in(0,0.5)$. We chose $q_1=2$, $p_1=p_2=3$ and a varying of bandwidths b_1 and b_2.

We first present the simulation result for estimating $\log\sigma^2(t)$. The number of grid points ranges from 5,000 to 10,000 and we repeat the experiment for 200 times for each case. The calculated $\log\sqrt{MISE}$ for $\log(\sigma^2)$ is plotted in Figure 3.7. The plot shows a polynomial convergence rate, but a crude estimation of the convergence rate is $n^{-0.34}$, which is slower than the conjectured rate $n^{-\frac{3}{2\times3+1}}\approx n^{-0.43}$. We could also obtained the quantile plots in Figure 3.8 for 10,000 grid points with the optimal bandwidth $b_1=0.05$ and $b_2=0.4$. As a comparison, with the same setup, the optimal bandwidth for directly applying $\widehat{H}_1$ is 0.28, which is much larger than the optimal b_1 mentioned before, indicating an undersmoothing of $\widehat{H}_1$ is desired when we estimate $\log(\sigma^2(t))$.

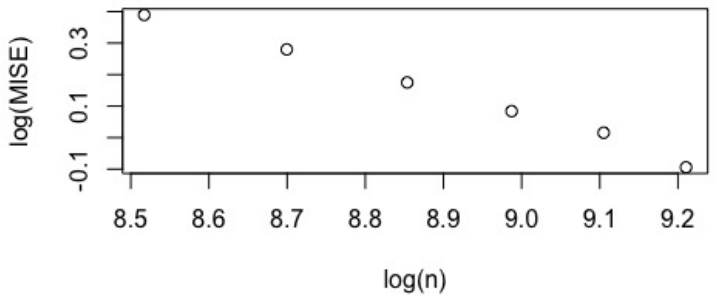


Figure 3.7: Plot of $\log(\mathrm{MISE})$ versus $\log n$ for $\widehat{\log \sigma^2}$

We also conducted simulations for calculating $\widehat{H}_2$ with $\widehat{\log \sigma^2(t)}$ plugged in for the true value (with $b_1 = 0.05$, $b_2 = 0.4$) under $p = 3$, $q = 2$ and varying b_3. The smallest $\sqrt{MISE}$ for $\widehat{H}_2$ is 0.0526. For comparison, the smallest $\sqrt{MISE}$ for directly applying $\widehat{H}_1$ is 0.0629. The corresponding empirical quantile plots are given in Figure 3.9. Although $\widehat{H}_2$ performed better than directly applying $\widehat{H}_1$, it is not clear whether there is a difference in convergence rates.

□

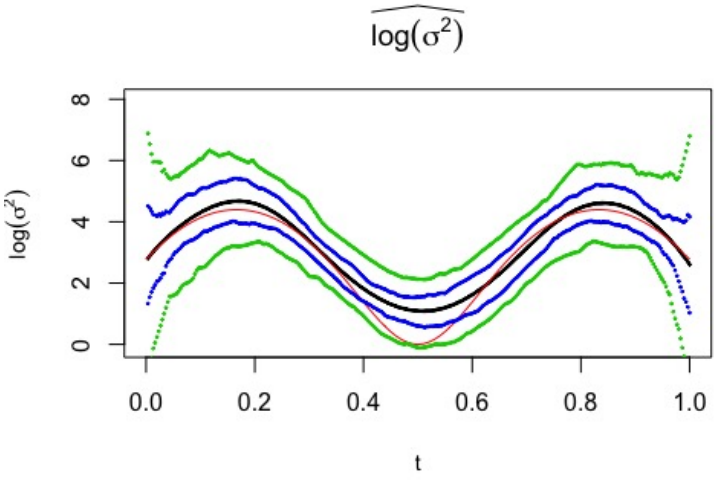


Figure 3.8: Empirical quantiles of $\widehat{\log \sigma^2}$ 100 simulation runs; red curve is the true function; green curves are 5% and 95% pointwise empirical quantiles, respectively; blue curves are 25% and 75% empirical quantiles, respectively; black curves are empirical means.

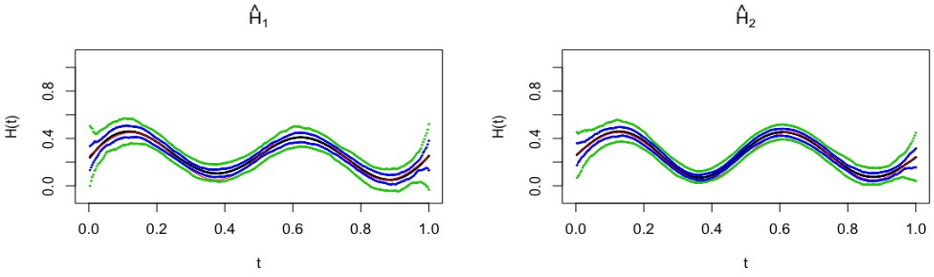


Figure 3.9: Empirical quantiles of $\widehat{H}_1$ and $\widehat{H}_2$ with plugged in $\widehat{\log \sigma^2(t)}$ based on 100 simulation runs; red curves are true functions; green curves are 5% and 95% pointwise empirical quantiles, respectively; blue curves are 25% and 75% empirical quantiles, respectively; black curves are empirical means.

BIBLIOGRAPHY

M. Abramowitz and I. A. Stegun. *Handbook of Mathematical Functions with Formulas, Graphs, and Mathematical Tables*, volume 55 of *National Bureau of Standards Applied Mathematics Series*. For sale by the Superintendent of Documents, U.S. Government Printing Office, Washington, D.C., 1964.

R. J. Adler. *The Geometry of Random Fields*, volume 62 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2010. ISBN 978-0-898716-93-1. doi: 10.1137/1.9780898718980.ch1. URL https://doi.org/10.1137/1.9780898718980.ch1.

E. B. Anderes and M. L. Stein. Estimating deformations of isotropic Gaussian random fields on the plane. *Ann. Statist.*, 36(2):719–741, 2008. ISSN 0090-5364. doi: 10.1214/009053607000000893. URL https://doi.org/10.1214/009053607000000893.

A. Ayache, S. Cohen, and J. L. Véhel. The covariance structure of multifractional brownian motion, with application to long range dependence. In *Acoustics, Speech, and Signal Processing, 2000. ICASSP'00. Proceedings. 2000 IEEE International Conference on*, volume 6, pages 3810–3813. IEEE, 2000.

J.-M. Bardet and D. Surgailis. Nonparametric estimation of the local Hurst function of multifractional Gaussian processes. *Stochastic Process. Appl.*, 123(3):1004–1045, 2013. ISSN 0304-4149. URL https://doi.org/10.1016/j.spa.2012.11.009.

A. Benassi, S. Jaffard, and D. Roux. Elliptic Gaussian random processes. *Rev. Mat. Iberoamericana*, 13(1):19–90, 1997. ISSN 0213-2230. URL https://doi.org/10.4171/RMI/217.

G. Berschneider. Spectral representation of intrinsically stationary fields. *Stochastic Process. Appl.*, 122(12):3837–3851, 2012. ISSN 0304-4149. doi: 10.1016/j.spa.2012.07.005. URL https://doi.org/10.1016/j.spa.2012.07.005.

G. Berschneider and Z. Sasvári. Spectral theory of stationary random fields and their generalizations. A short historical survey. In *Indefinite Inner Product Spaces, Schur Analysis, and Differential Equations*, volume 263 of *Oper. Theory Adv. Appl.*, pages 217–235. Birkhäuser/Springer, Cham, 2018.

P. R. Bertrand, M. Fhima, and A. Guillin. Local estimation of the Hurst index of multifractional Brownian motion by increment ratio statistic method. *ESAIM Probab. Stat.*, 17:307–327, 2013. ISSN 1292-8100. URL https://doi.org/10.1051/ps/2011154.

P. Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statistics: Probability and Statistics. John Wiley & Sons, Inc., New York, second edition, 1999. ISBN 0-471-19745-9. doi: 10.1002/9780470316962. URL https://doi.org/10.1002/9780470316962. A Wiley-Interscience Publication.

S. Bochner. *Vorlesungen über Fouriersche Integrale*. Chelsea Publishing Company, 1948.

D. Bolin. Spatial Matérn fields driven by non-Gaussian noise. *Scand. J. Stat.*, 41(3):557–579, 2014. ISSN 0303-6898. doi: 10.1111/sjos.12046. URL https://doi.org/10.1111/sjos.12046.

P. Breuer and P. Major. Central limit theorems for nonlinear functionals of Gaussian fields. *J. Multivariate Anal.*, 13(3):425–441, 1983. ISSN 0047-259X. URL https://doi.org/10.1016/0047-259X(83)90019-2.

J.-P. Chilès and P. Delfiner. *Geostatistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, second edition, 2012. ISBN 978-0-470-18315-1. URL https://doi.org/10.1002/9781118136188. Modeling spatial uncertainty.

J.-F. Coeurjolly. Identification of multifractional Brownian motion. *Bernoulli*, 11(6):987–1008, 2005. ISSN 1350-7265. URL https://doi.org/10.3150/bj/1137421637.

S. Cohen. From self-similarity to local self-similarity: the estimation problem. In *Fractals: Theory and Applications in Engineering*, pages 3–16. Springer, London, 1999.

H. Cramér. On harmonic analysis in certain functional spaces. *Ark. Mat. Astr. Fys.*, 28B (12):7, 1942. ISSN 0004-2080.

N. Cressie. The origins of kriging. *Math. Geol.*, 22(3):239–252, 1990. ISSN 0882-8121. doi: 10.1007/BF00889887. URL https://doi.org/10.1007/BF00889887.

Y. Davydov and V. Paulauskas. Lamperti type theorems for random fields, 2017. arXiv:1705.00182.

J. Diestel and J. J. Uhl, Jr. The Radon-Nikodym theorem for Banach space valued measures. *Rocky Mountain J. Math.*, 6(1):1–46, 1976. ISSN 0035-7596. doi: 10.1216/RMJ-1976-6-1-1. URL https://doi.org/10.1216/RMJ-1976-6-1-1.

R. L. Dobrushin. Gaussian and their subordinated self-similar random generalized fields. *Ann. Probab.*, 7(1):1–28, 1979. ISSN 0091-1798. URL http://links.jstor.org/sici?sici=0091-1798(197902)7:1<1:GATSSR>2.0.CO;2-Y&origin=MSN.

R. M. Dudley. *Real Analysis and Probability*, volume 74 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 2002. ISBN 0-521-00754-2. doi: 10.1017/CBO9780511755347. URL https://doi.org/10.1017/CBO9780511755347. Revised reprint of the 1989 original.

N. Dunford and J. T. Schwartz. *Linear Operators. Part I*. Wiley Classics Library. John Wiley & Sons, Inc., New York, 1988. ISBN 0-471-60848-3. General theory, With the assistance of William G. Bade and Robert G. Bartle, Reprint of the 1958 original, A Wiley-Interscience Publication.

K. J. Falconer. Tangent fields and the local structure of random fields. *J. Theoret. Probab.*, 15(3):731–750, 2002. ISSN 0894-9840. URL https://doi.org/10.1023/A:1016276016983.

K. J. Falconer. The local structure of random processes. *J. London Math. Soc. (2)*, 67(3): 657–672, 2003. ISSN 0024-6107. doi: 10.1112/S0024610703004186. URL https://doi.org/10.1112/S0024610703004186.

J. Fan and I. Gijbels. *Local Polynomial Modelling and its Applications*, volume 66 of *Monographs on Statistics and Applied Probability*. Chapman & Hall, London, 1996. ISBN 0-412-98321-4.

M. Fuentes. Spectral methods for nonstationary spatial processes. *Biometrika*, 89(1):197–210, 2002. ISSN 0006-3444. doi: 10.1093/biomet/89.1.197. URL https://doi.org/10.1093/biomet/89.1.197.

G.-A. Fuglstad, F. Lindgren, D. Simpson, and H. v. Rue. Exploring a new class of non-stationary spatial Gaussian random fields with varying local anisotropy. *Statist. Sinica*, 25(1):115–133, 2015. ISSN 1017-0405.

L. S. Gandin. *Objective Analysis of Meteorological Fields: Gidrometeorotogicheskoe Izdatel'stvo (GIMIZ), Leningrad (translated by Israel Program for Scientific Translations, Jerusalem, 1965)*. 1963.

B. R. Gelbaum and J. Gil de Lamadrid. Bases of tensor products of Banach spaces. *Pacific J. Math.*, 11:1281–1286, 1961. ISSN 0030-8730. URL http://projecteuclid.org/euclid.pjm/1103036914.

I. M. Gel'fand and N. Y. Vilenkin. *Generalized Functions. Vol. 4: Applications of harmonic analysis*. Translated by Amiel Feinstein. Academic Press, New York - London, 1964, 1964.

B. Gnedenko. Sur la distribution limite du terme maximum d'une série aléatoire. *Ann. of Math. (2)*, 44:423–453, 1943. ISSN 0003-486X.

P. Hall and J. D. Hart. Nonparametric regression with long-range dependence. *Stochastic Processes and Their Applications*, 36(2):339–351, 1990.

E. Herbin. From N parameter fractional Brownian motions to N parameter multifractional Brownian motions. *Rocky Mountain J. Math.*, 36(4):1249–1284, 2006. ISSN 0035-7596. doi: 10.1216/rmjm/1181069415. URL https://doi.org/10.1216/rmjm/1181069415.

D. Higdon. A process-convolution approach to modelling temperatures in the north atlantic ocean. *Environmental and Ecological Statistics*, 5(2):173–190, 1998.

D. Higdon, J. Swall, and J. Kern. Non-stationary spatial modeling. *Bayesian statistics*, 6 (1):761–768, 1999.

T. Hsing and R. Eubank. *Theoretical Foundations of Functional Data Analysis, with an Introduction to Linear Operators*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2015. ISBN 978-0-470-01691-6. doi: 10.1002/9781118762547. URL https://doi.org/10.1002/9781118762547.

T. Hsing, T. Brown, and B. Thelen. Local intrinsic stationarity and its inference. *Ann. Statist.*, 44(5):2058–2088, 2016. ISSN 0090-5364. URL https://doi.org/10.1214/15-AOS1402.

H. Hult and F. Lindskog. Regular variation for measures on metric spaces. *Publ. Inst. Math. (Beograd) (N.S.)*, 80(94):121–140, 2006. ISSN 0350-1302. doi: 10.2298/PIM0694121H. URL http://dx.doi.org/10.2298/PIM0694121H.

O. Kallenberg. *Foundations of modern probability*. Probability and Its Applications (New York). Springer-Verlag, 1997.

J. L. Kelley. *General Topology*. D. Van Nostrand Company, Inc., Toronto-New York-London, 1955.

A. Khintchine. Korrelationstheorie der stationären stochastischen prozesse. *Mathematische Annalen*, 109(1):604–615, 1934.

A. N. Kolmogorov. Wienersche Spiralen und einige andere interessante Kurven im Hilbertschen Raum. *C. R. (Doklady) Acad. Sci. URSS (N.S.)*, 26:115–118, 1940.

D. G. Krige. A statistical approach to some mine valuation and allied problems on the witwatersrand. 1951.

M. Kuczma. *An Introduction to the Theory of Functional Equations and Inequalities*. Birkhäuser Verlag, Basel, second edition, 2009. ISBN 978-3-7643-8748-8. doi: 10.1007/978-3-7643-8749-5. URL https://doi.org/10.1007/978-3-7643-8749-5. Cauchy's equation and Jensen's inequality, Edited and with a preface by Attila Gilányi.

J. Lamperti. Semi-stable stochastic processes. *Trans. Amer. Math. Soc.*, 104:62–78, 1962a. ISSN 0002-9947. doi: 10.2307/1993933. URL https://doi.org/10.2307/1993933.

J. Lamperti. Semi-stable stochastic processes. *Trans. Amer. Math. Soc.*, 104:62–78, 1962b. ISSN 0002-9947. doi: 10.2307/1993933. URL https://doi.org/10.2307/1993933.

J. Lévy-Véhel and R. Peltier. Multifractional brownian motion: definition and preliminary results. *Rapport de recherche de lINRIA n2645*, 1995.

E. H. Lieb and M. Loss. *Analysis*, volume 14 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 1997. ISBN 0-8218-0632-7.

P. A. Loeb and E. Talvila. Lusin's theorem and Bochner integration. *Sci. Math. Jpn.*, 60 (1):113–120, 2004. ISSN 1346-0862.

W.-L. Loh. Estimating the smoothness of a gaussian random field from irregularly spaced data via higher-order quadratic variations. *The Annals of Statistics*, 43(6):2766–2794, 2015.

B. B. Mandelbrot and J. W. Van Ness. Fractional Brownian motions, fractional noises and applications. *SIAM Rev.*, 10:422–437, 1968. ISSN 0036-1445. URL https://doi.org/10.1137/1010093.

B. B. Mandelbrot and J. R. Wallis. Noah, joseph, and operational hydrology. *Water resources research*, 4(5):909–918, 1968.

G. Matheron. *Traité de Géostatistique Appliquée. 1 (1962)*, volume 1. Editions Technip, 1962.

G. Matheron. Principles of geostatistics. *Economic geology*, 58(8):1246–1266, 1963a.

G. Matheron. *Traité de Géostatistique Appliquée. 2. Le krigeage*. Editions Technip, 1963b.

G. Matheron. *Le Krigeage Universel*, volume 1. École nationale supérieure des mines de Paris Paris, 1969.

G. Matheron. The intrinsic random functions and their applications. *Advances in Appl. Probability*, 5:439–468, 1973. ISSN 0001-8678. doi: 10.2307/1425829. URL https://doi.org/10.2307/1425829.

M. M. Meerschaert. *Multivariate Domains of Attraction and Regular Variation*. PhD thesis, University of Michigan, Ann Arbor, 1984.

C. A. Micchelli. Interpolation of scattered data: distance matrices and conditionally positive definite functions. *Constr. Approx.*, 2(1):11–22, 1986. ISSN 0176-4276. URL https://doi.org/10.1007/BF01893414.

K.-H. Neeb. Operator-valued positive definite kernels on tubes. *Monatsh. Math.*, 126(2): 125–160, 1998. ISSN 0026-9255. doi: 10.1007/BF01473583. URL https://doi.org/10.1007/BF01473583.

M. Newman. How to determine the accuracy of the output of a matrix inversion program. *J. Res. Nat. Bur. Standards Sect. B*, 78B:65–68, 1974. ISSN 0160-1741.

I. Nourdin. *Selected Aspects of Fractional Brownian Motion*. Springer, Milano, 2012.

V. M. Panaretos and S. Tavakoli. Fourier analysis of stationary time series in function space. *Ann. Statist.*, 41(2):568–603, 2013. ISSN 0090-5364. doi: 10.1214/13-AOS1086. URL https://doi.org/10.1214/13-AOS1086.

E. Perrin, R. Harba, C. Berzin-Joseph, I. Iribarren, and A. Bonami. nth-order fractional brownian motion and fractional gaussian noises. *IEEE Transactions on Signal Processing*, 49(5):1049–1059, 2001.

S. I. Resnick. *A Probability Path*. Birkhäuser Boston, Inc., Boston, MA, 1999. ISBN 0-8176-4055-X.

G. Samorodnitsky and M. S. Taqqu. *Stable Non-Gaussian Random Processes*. Stochastic Modeling. Chapman & Hall, New York, 1994. ISBN 0-412-05171-0. Stochastic models with infinite variance.

P. D. Sampson and P. Guttorp. Nonparametric estimation of nonstationary spatial covariance structure. *Journal of the American Statistical Association*, 87(417):108–119, 1992.

Z. Sasvári. Correlation functions of intrinsically stationary random fields. In *Modern Analysis and Applications. The Mark Krein Centenary Conference. Vol. 1: Operator theory and related topics*, volume 190 of *Oper. Theory Adv. Appl.*, pages 451–470. Birkhäuser Verlag, Basel, 2009. doi: 10.1007/978-3-7643-9919-1_28. URL https://doi.org/10.1007/978-3-7643-9919-1_28.

Z. Sasvári. *Multivariate Characteristic and Correlation Functions*, volume 50 of *De Gruyter Studies in Mathematics*. Walter de Gruyter & Co., Berlin, 2013. ISBN 978-3-11-022398-9; 978-3-11-022399-6. doi: 10.1515/9783110223996. URL https://doi.org/10.1515/9783110223996.

M. L. Stein. *Interpolation of Spatial Data*. Springer Series in Statistics. Springer-Verlag, New York, 1999. ISBN 0-387-98629-4. doi: 10.1007/978-1-4612-1494-6. URL https://doi.org/10.1007/978-1-4612-1494-6. Some theory for Kriging.

M. L. Stein. Nonstationary spatial covariance functions,. Technical Report 21, University of Chicago, Center for IntegratingStatistical and Environmental Science, 2005.

S. A. Stoev and M. S. Taqqu. How rich is the class of multifractional Brownian motions? *Stochastic Process. Appl.*, 116(2):200–221, 2006. ISSN 0304-4149. URL https://doi.org/10.1016/j.spa.2005.09.007.

M. S. Taqqu. Weak convergence to fractional Brownian motion and to the Rosenblatt process. *Z. Wahrscheinlichkeitstheorie und Verw. Gebiete*, 31:287–302, 1975. URL https://doi.org/10.1007/BF00532868.

S. Taqqu. Self-similar processes and related ultraviolet and infrared catastrophes. In *Random Fields, Vol. I, II (Esztergom, 1979)*, volume 27 of *Colloq. Math. Soc. János Bolyai*, pages 1057–1096. North-Holland, Amsterdam-New York, 1981.

A. B. Tsybakov. *Introduction to Nonparametric Estimation*. Springer Series in Statistics. Springer, New York, 2009. ISBN 978-0-387-79051-0. URL https://doi.org/10.1007/b13794. Revised and extended from the 2004 French original, Translated by Vladimir Zaiats.

K. Yosida. *Functional Analysis*, volume 123 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences]*. Springer-Verlag, Berlin-New York, sixth edition, 1980. ISBN 3-540-10210-8.