Excursion Sets and Critical Points of Gaussian Random Fields

by

Weinan Qi

A thesis
presented to the University of Waterloo
in fulfillment of the
thesis requirement for the degree of
Doctor of Philosophy
in
Statistics

Waterloo, Ontario, Canada, 2022

Examining Committee Membership

The following served on the Examining Committee for this thesis. The decision of the Examining Committee is by majority vote.

External Examiner: Jian Kang
Professor, Dept. of Biostatistics
University of Michigan

Internal-External Member: Matthew Scott
Associate Professor, Dept. of Applied Mathematics
University of Waterloo

Internal Member: Alexander Schied
Professor, Dept. of Statistics and Actuarial Science
University of Waterloo

Tony Wirjanto
Professor, Dept. of Statistics and Actuarial Science
and School of Accounting and Finance
University of Waterloo

Supervisor(s): Paul Marriott
Professor, Dept. of Statistics and Actuarial Science
University of Waterloo

Yi Shen
Associate Professor, Dept. of Statistics and Actuarial Science
University of Waterloo

Author's Declaration

Abstract

Modeling the critical points of a Gaussian random field is an important challenge in stochastic geometry. In this thesis, we focus on stationary Gaussian random fields and study the locations and types of the critical points over high thresholds. Under certain conditions, we show that when the threshold tends to infinity and the searching area expands with a matching speed, both the locations of the local maxima and the locations of all the critical points above the threshold converge weakly to a Poisson point process. We then discuss the local behavior of the critical points by looking at the type of a critical point given there is another critical point close to it. We show if two critical points above u are very close one to each other, then they are most likely to be one local maximum and one saddle point with index $N-1$. We will further discuss the modeling of the critical points when the threshold is high but not very high. The proposed model has a hierarchical structure that can capture the positions of the global maxima and other critical points simultaneously. The performance of the proposed model is evaluated by the comparisons between the L functions.

Acknowledgements

First and foremost, I would like to express my deepest gratitude to my Ph.D. supervisors, Professor Paul Marriott and Professor Yi Shen, for their enduring support and guidance that helped me overcome countless difficulties in my Ph.D. career.

My sincere gratitude also goes to the examining committee members, Professor Jian Kang, Professor Matthew Scott, Professor Alexander Schied, and Professor Tony Wirjanto, for their time and effort to review my thesis.

I am very grateful to my master's supervisor, Professor Mahmoud Zarepour, for introducing me to the world of applied probability.

I would also like to thank the faculty and staff, Ms. Mary Lou Dufton, Ms. Lisa Baxter, and Mr. Greg Preston, in the Department of Statistics and Actuarial Science, for their help during my Ph.D. study.

I am highly indebted to my parents, for encouraging me to pursue my education. This thesis would not have been possible without their unconditional support and love for me.

Finally, special thanks to my loving girlfriend, Yan Su, and cute dog, Dadan, for accompanying me through the ups and downs of life.

Dedication

This thesis is dedicated to my parents: Zhongqiang Qi and Lili Wei, and my girlfriend: Yan Su, for their love and encouragement in countless ways.

Table of Contents

List of Figures

List of Tables

Chapter 1

Introduction

This thesis deals with the critical points and the excursion sets of stationary Gaussian random fields. Gaussian random fields, defined by the multivariate normality on any finite subset of some region $S \subset \mathbb{R}^N$, play an important role in the modeling in various areas such as astronomy ([6], [40]), biomedical imaging ([42], [19]), geography ([38], [5], [7]), etc. The topological information hidden in the high-dimensional parameter space of the Gaussian random field is considered from both Bayesian and frequentist perspectives. Especially, researchers are interested in the topological structure of the domain where the underlying Gaussian random field is above a predetermined threshold. This domain is called an excursion set (see Figure 1). More specifically, let $\{X(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^N\}$ $(N \geq 1)$ be a Gaussian random field indexed by $\mathbb{R}^N$. For any $u \in \mathbb{R}$ and $S \subset \mathbb{R}^N$, the excursion set of X with threshold u and search region S is defined by

$$A_u(X, S) := \{\boldsymbol{t} \in S : X(\boldsymbol{t}) > u\}.$$

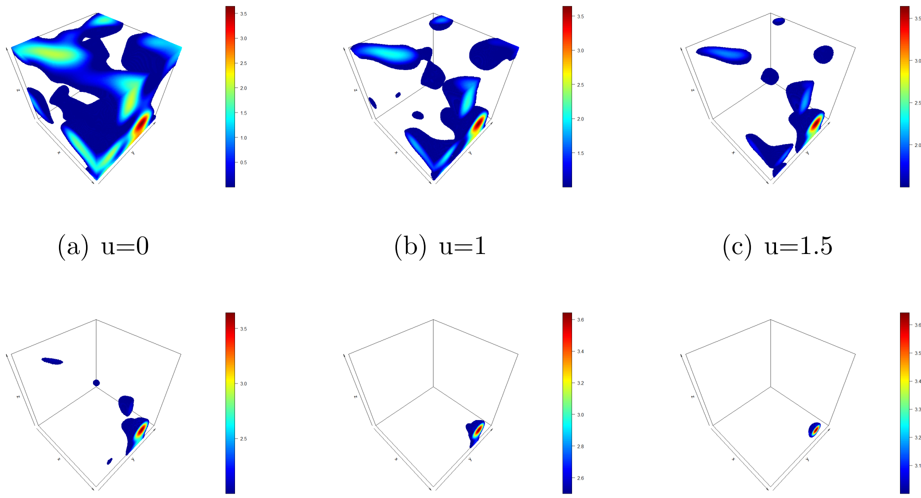
For a sample function f of X possessing up to second order partial derivatives, its gradient and Hessian matrix are denoted by ∇f and $\nabla^2 f$, respectively. A point $\boldsymbol{t} \in \mathbb{R}^N$ is said to be (the position of) a critical point of f if

$$\nabla f(\boldsymbol{t}) = 0.$$

If it further satisfies

$$\det\left(\nabla^2 f(\boldsymbol{t})\right) \neq 0 \quad \text{and} \quad \text{index}\left(\nabla^2 f(\boldsymbol{t})\right) = k,$$

where $0 \leq k \leq N$ and $\text{index}(\cdot)$ stands for the number of negative eigenvalues of a square matrix, then $\boldsymbol{t}$ is said to be a non-degenerate critical point of f with index (or type) k.



(a) u=0 (b) u=1 (c) u=1.5

(d) u=2 (e) u=2.5 (f) u=3

Figure 1.1: Excursion sets of a Gaussian random field indexed by $\mathbb{R}^3$. The Gaussian random field is centered with covariance function $\mathbb{E}[X(\boldsymbol{t})X(\boldsymbol{s})] = e^{-\|\boldsymbol{t}-\boldsymbol{s}\|_3^2}$, where $\|\cdot\|_3$ denotes the 3-dimensional Euclidean norm. The colored part is the excursion set in the search region $[0, 1]^3$ above the threshold u, where a warmer color stands for a higher value.

For each connected component of $A_u(f, \mathbb{R}^N)$, by Morse theory ([26]), its topology can be characterized by the number and types of the non-degenerate critical points it contains. Therefore, studying the distribution of non-degenerate critical points above u can be very useful for understanding the excursion set.

In general, we study the distribution of the critical points of a stationary Gaussian random field above a threshold from two aspects. The first aspect is to study the Poisson limit of the critical points as the threshold tends to infinity. When the threshold is very high, this result can provide a model for the critical points in the excursion set. However, this model cannot capture the local structure of the Gaussian random field near a high critical point. When the threshold is high but not very high, capturing such local structure plays an important role in modeling the critical points. Therefore, a correction is needed for this model. Thus, the second aspect is to study the interactions between high critical points to provide a theoretical basis for such a correction. In particular, for two high critical points, we will see the effect of their indices on the interaction between them. From the above two aspects, we propose a model for the critical points when the threshold is high but not very high.

The remainder of the thesis is organized as follows.

In Chapter 2, we introduce three important mathematical tools used in this thesis: Gaussian random fields, Point processes and Morse theory. In particular, we introduce the "metatheorem" ([4]), which is a generalization of Rice's formula ([33]), allowing us to derive an integral expression for the expected number of the critical points in an excursion set with a given index when $N \geq 1$. This result has profound implications for many subsequent studies and also plays an important role in this thesis.

In Chapter 3, we study the Poisson limits of the critical points of a stationary Gaussian random field in the excursion set as the threshold tends to infinity. If a critical point has full index, we call it a local maximum of X. If it is also the position of the global maximum of X in the connected component containing it, we call it a global maximum. Then we can construct three point processes using only the global maxima above the threshold, only the local maxima above the threshold, and all the critical points above the threshold, respectively. As u increases, the search region is expanded to keep the mean number of the local maxima at a constant level. We will then show that as $u \to \infty$, these point processes will all converge weakly to the same stationary Poisson point process. This generalization is mentioned in [1], but never formally proved. We further prove that the number of connected components of the excursion set will converge weakly to a Poisson random variable as $u \to \infty$.

In Chapter 4, we explore the interactions between the critical points above a high threshold for an isotropic Gaussian random field. This is achieved by studying the local behavior of the underlying random field in the vicinity of the origin, provided that a critical point above the threshold is located at the origin. We get the densities of the mean measures of the critical points with different indices. We then use these densities to construct two ratios and study their limiting behaviors. Combining our results for the two ratios, we find an important relationship between different types of critical points: a pair of very close critical points above a high threshold must consist of one local maximum and one critical point with index $N - 1$.

In Chapter 5, with the help of our previous results, we propose a model for the critical points in the excursion set when the threshold is high but not very high. In order to capture the global maxima and other critical points simultaneously, the model has a hierarchical structure which represents three components of the interactions under study. We evaluate its performance empirically on two isotropic Gaussian random fields with different covariance structures by checking their L functions, where the stationary Poisson point process model serves as a benchmark.

Chapter 2

Preliminaries

This chapter consists of three parts: Gaussian random fields, point processes and a brief introduction to Morse theory. For Gaussian random fields, we will mainly focus on two types of continuity and differentiability, and the metatheorem. For point processes, we will formally define Poisson point processes and the weak convergence of point processes. A sufficient condition for the existence of the weak limit of a series of point processes will also be given. Finally, we will define Morse functions and introduce Morse inequalities on an excursion set.

Since we are working on Gaussian random fields indexed by $\mathbb{R}^N$ $(N \geq 1)$, the following notations will be used throughout this chapter. Let $(\Omega, \mathcal{A}, P)$ be the probability space on which the random objects are defined. For any $n \geq 1$, endow $\mathbb{R}^n$ with the usual Euclidean norm $\|\cdot\|_n$. We often drop the subscript n when the dimension can be easily seen from the context. Denote by λ_n the n-dimensional Lebesgue measure, and by $\mathcal{B}(\mathbb{R}^n)$ the Borel σ-field on $\mathbb{R}^n$. For any open or closed set $T \subset \mathbb{R}^N$, denote by $C^k(T)$ $(k \geq 1)$ the set of all real-valued functions on T with continuous up to k-th order derivatives.

2.1 Gaussian Random Fields

2.1.1 The Definition of Gaussian Random Fields

A random field $\{X(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^N\}$ is said to be a Gaussian random field indexed by $\mathbb{R}^N$ if for any non-empty finite subset $\{\boldsymbol{t}_i, 1 \leq i \leq n\} \subset \mathbb{R}^N$, the distribution of $(X(\boldsymbol{t}_1), \ldots, X(\boldsymbol{t}_n))$ is multivariate Gaussian. X is said to be a Gaussian random field on $\mathbb{R}^N$ with mean function $\mu(\cdot)$ and covariance function $r(\cdot, \cdot)$ if for any $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^N$,

$$\mu(\boldsymbol{t}) := \mathbb{E}\left[X(\boldsymbol{t})\right] \quad \text{and} \quad r(\boldsymbol{s}, \boldsymbol{t}) = \text{Cov}\left[X(\boldsymbol{s}), X(\boldsymbol{t})\right].$$

From the above definition and Appendix A.1, we see that

1. X can be fully characterized by its behavior on the non-empty finite subsets of $\mathbb{R}^N$;
2. X on any non-empty finite subset follows a multivariate Gaussian distribution;
3. Any multivariate Gaussian distribution can be fully characterized by its mean vector and covariance matrix.

Therefore, the distribution of X can be fully characterized by its mean function $\mu(\cdot)$ and covariance function $r(\cdot, \cdot)$.

A Gaussian Random Field X is said to be homogeneous or stationary if its mean function $\mu(\cdot)$ is a constant and there exists a real-valued function $\rho(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^N$, such that

$$r(\boldsymbol{s}, \boldsymbol{t}) = \rho(\boldsymbol{s} - \boldsymbol{t})$$

for any $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^N$. With some abuse of notations, we still use $r(\cdot)$ instead of $\rho(\cdot)$ to denote the covariance function, i.e., in stationary cases,

$$r(\boldsymbol{s} - \boldsymbol{t}) = \text{Cov}\left[X(\boldsymbol{s}), X(\boldsymbol{t})\right].$$

If it is also true that $r(\boldsymbol{s}, \boldsymbol{t})$ is a function of the Euclidean distance $\|\boldsymbol{s} - \boldsymbol{t}\|$ only, then X is also called isotropic. As we mentioned before, without loss of generality, we will often work on the centered, stationary Gaussian random fields with unit variance, i.e.,

$$r(\boldsymbol{0}) = 1 \quad \text{and} \quad \mu(\boldsymbol{t}) = 0$$

for any $\boldsymbol{t} \in \mathbb{R}^N$. In this case, it is easy to see that $|r(\boldsymbol{t})| \leq 1$ for any $\boldsymbol{t} \in \mathbb{R}^N$. While $\text{Cov}\left[X(\boldsymbol{s}), X(\boldsymbol{t})\right]$ is defined on $\mathbb{R}^N \times \mathbb{R}^N$, $r(\boldsymbol{s} - \boldsymbol{t})$ is defined on $\mathbb{R}^N$ only. In the isotropic case, we shall go even further, and write $\text{Cov}\left[X(\boldsymbol{s}), X(\boldsymbol{t})\right] = r(\|\boldsymbol{s} - \boldsymbol{t}\|)$, where $r(\|\boldsymbol{s} - \boldsymbol{t}\|)$ is defined on $[0, \infty)$.

2.1.2 Almost Sure Continuity and Differentiability

Given a compact subset $T \subset \mathbb{R}^N$, a random field X is said to be almost surely continuous on T if it has a continuous sample function on T with probability one, i.e.,

$$P\left[\lim_{\boldsymbol{s}\in T, \|\boldsymbol{t}-\boldsymbol{s}\|\to 0} |X(\boldsymbol{t}) - X(\boldsymbol{s})| = 0, \forall \boldsymbol{t} \in T\right] = 1.$$

Suppose that X is a centered, stationary Gaussian random field with unit variance and covariance function $r(\cdot)$. For any given compact subset $T \subset \mathbb{R}^N$, define

$$p_T(u) := \sqrt{\max_{\boldsymbol{s},\boldsymbol{t}\in T, \|\boldsymbol{t}-\boldsymbol{s}\|\le u} (1 - r(\boldsymbol{t}-\boldsymbol{s}))}.$$

[4] introduced a sufficient condition for the almost sure continuity of $X(\boldsymbol{t})$ as follows.

Lemma 2.1.1. *(Theorem 1.4.1, [4]) If for some $\delta > 0$, either*

$$\int_0^\delta (-\log(u))^{\frac{1}{2}} dp_T(u) < \infty \quad or \quad \int_\delta^\infty p_T\left(e^{-u^2}\right) du < \infty, \tag{2.1}$$

then the Gaussian random field X has an almost surely continuous modification on T. A sufficient condition which makes (2.1) hold is that for some $0 < \gamma < \infty$, $\alpha, \beta > 0$ and any $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^N$ with $0 < \|\boldsymbol{t}-\boldsymbol{s}\| < \beta$,

$$\mathbb{E}\left[(X(\boldsymbol{t}) - X(\boldsymbol{s}))^2\right] \le \gamma \left|\log(\|\boldsymbol{t}-\boldsymbol{s}\|)\right|^{-(1+\alpha)}. \tag{2.2}$$

Note that the sufficient condition (2.2) only depends on the covariance function of X. This is an advantage for studying centered Gaussian random fields: all of their properties only depend on the covariance structure. Also note that this condition is actually independent from the choice of T. We will say X is almost surely continuous on $\mathbb{R}^N$ if it is almost surely continuous on any compact subset $T \subset \mathbb{R}^N$. In Example A.2.1, we show that a centered Gaussian random field with covariance function $r(\boldsymbol{t}) = e^{-\|\boldsymbol{t}\|^2}$, $\boldsymbol{t} \in \mathbb{R}^N$ satisfies all the conditions in Lemma 2.1.1.

A random field X is said to be almost surely k times continuously differentiable on an open subset $T \subset \mathbb{R}^N$ if its sample function has up to k-th order continuous derivatives in T with probability one, i.e.,

$$P\left[X \in C^k(T)\right] = 1.$$

At each point $\boldsymbol{t} \in T$, these derivatives are called almost sure derivatives of X at $\boldsymbol{t}$ since they exist as almost sure limits. Similarly, X is said to be almost surely k times continuously differentiable on a subset $A \subset \mathbb{R}^N$ if it is almost surely k times continuously differentiable on an open set containing A. In the next subsection, we will also introduce a sufficient condition for almost sure differentiability.

2.1.3 Mean Square Continuity and Differentiability

Another mode of continuity, which is more popular in the theory of random fields is the continuity in mean square. Given a point $\boldsymbol{t} \in \mathbb{R}^N$, a random field X is said to be continuous at $\boldsymbol{t}$ in mean square if for any sequence $\{\boldsymbol{t}_n, n \geq 1\}$ in $\mathbb{R}^N$ which satisfies $\lim_{n\to\infty} \|\boldsymbol{t} - \boldsymbol{t}_n\| = 0$, we have

$$\lim_{n\to\infty} \mathbb{E}\left[(X(\boldsymbol{t}) - X(\boldsymbol{t}_n))^2\right] = 0. \tag{2.3}$$

For any $A \subset \mathbb{R}^N$, if (2.3) holds for all $\boldsymbol{t} \in A$, then the random field X is said to be continuous on A in mean square.

Compared with the almost sure continuity, an important advantage of the mean square continuity is that it relates the continuity of a random field with the continuity of its covariance function. This can be explained by the following lemma.

Lemma 2.1.2. *(Theorem 2.2.1, [1]) A random field X is continuous in mean square at a given point $\boldsymbol{t}_0 \in \mathbb{R}^N$ if and only if its covariance function $r(\boldsymbol{s}, \boldsymbol{t})$ is continuous at the point $\boldsymbol{s} = \boldsymbol{t} = \boldsymbol{t}_0$. Therefore, if $r(\boldsymbol{s}, \boldsymbol{t})$ is continuous on every diagonal element $\boldsymbol{s} = \boldsymbol{t}$ in $\mathbb{R}^N \times \mathbb{R}^N$, then X is continuous in mean square on $\mathbb{R}^N$.*

Correspondingly, we can define mean square differentiability. Here we follow the definition in [4]. Choose a point $\boldsymbol{t} \in \mathbb{R}^N$ and k non-zero vectors $\boldsymbol{t}'_1, \ldots, \boldsymbol{t}'_k$ in $\mathbb{R}^N$. Let $\boldsymbol{t}' = (\boldsymbol{t}'_1, \ldots, \boldsymbol{t}'_k) \in \bigotimes^k \mathbb{R}^N$, the k-fold tensor product of $\mathbb{R}^N$. For any vector $H = (h_1, \ldots, h_k) \in \mathbb{R}^k$, denote by $\boldsymbol{t}'_H := (h_1 \boldsymbol{t}'_1, \ldots, h_k \boldsymbol{t}'_k)$. Then a random field X is said to have the kth-order mean square (partial) derivative in the direction $\boldsymbol{t}' = (\boldsymbol{t}'_1, \ldots, \boldsymbol{t}'_k)$ at $\boldsymbol{t}$ if the limit

$$D_{\boldsymbol{t}'_1,\ldots,\boldsymbol{t}'_k} X(\boldsymbol{t}) := \lim_{h_1,\ldots,h_k \to 0} \frac{1}{\prod_{i=1}^k h_i} F_{X,k}(\boldsymbol{t}, \boldsymbol{t}'_H) \tag{2.4}$$

exists in the mean square sense, where

$$F_{X,k}(\boldsymbol{t}, \boldsymbol{t}') = \sum_{a_1,\ldots,a_k \in \{0,1\}} (-1)^{k - \sum_{i=1}^k a_i} X\left(\boldsymbol{t} + \sum_{i=1}^k a_i \boldsymbol{t}'_i\right) \tag{2.5}$$

for any $\boldsymbol{t} \in \mathbb{R}^N$ and $\boldsymbol{t}' \in \bigotimes^k \mathbb{R}^N$, and the mean square limit in (2.4) is interpreted sequentially, i.e., first send h_1 to 0, then h_2, etc. Furthermore, X is said to be k times differentiable in the mean square sense on $A \subset \mathbb{R}^N$ if its kth-order mean square derivatives exist in all directions at any point in A.

Example 2.1.1. *When $k = 1$, it is easy to get the mean square limit*

$$D_{\boldsymbol{t}_1'} X(\boldsymbol{t}) = \lim_{h_1 \to 0} \frac{X(\boldsymbol{t} + h_1 \boldsymbol{t}_1') - X(\boldsymbol{t})}{h_1}.$$

When $k = 2$, this is

$$\begin{aligned} &D_{\boldsymbol{t}_1', \boldsymbol{t}_2'} X(\boldsymbol{t}) \\ &= \lim_{h_1, h_2 \to 0} \frac{X(\boldsymbol{t} + h_1 \boldsymbol{t}_1' + h_2 \boldsymbol{t}_2') - X(\boldsymbol{t} + h_1 \boldsymbol{t}_1') - X(\boldsymbol{t} + h_2 \boldsymbol{t}_2') + X(\boldsymbol{t})}{h_1 h_2} \\ &= \lim_{h_2 \to 0} \frac{1}{h_2} \left(\lim_{h_1 \to 0} \frac{X(\boldsymbol{t} + h_1 \boldsymbol{t}_1' + h_2 \boldsymbol{t}_2') - X(\boldsymbol{t} + h_2 \boldsymbol{t}_2')}{h_1} - \lim_{h_1 \to 0} \frac{X(\boldsymbol{t} + h_1 \boldsymbol{t}_1') - X(\boldsymbol{t})}{h_1} \right). \end{aligned}$$

Note that the only difference from the almost sure first-order directional derivative in the direction $\boldsymbol{t}'$ is that the limit here is in the mean square sense. The cases for $k > 2$ are similar.

From Example 2.1.1, we see that if the almost sure derivative and the mean square derivative at the same order in the same direction both exist, then they will coincide with each other almost surely. This can be easily derived from the fact that almost sure convergence and mean square convergence of a sequence of random variables will both result in convergence in probability, and therefore their limits are the same almost surely. For a Gaussian random field, by Lemma A.1.1, if an almost sure derivative exists, then the mean square derivative at the same order in the same direction also exists.

In general, for some directions $\boldsymbol{t}_1', \boldsymbol{t}_2' \in \mathbb{R}^N$, point $\boldsymbol{t} \in \mathbb{R}^N$ and random field $X(\boldsymbol{t})$, the equation

$$D_{\boldsymbol{t}_1', \boldsymbol{t}_2'} X(\boldsymbol{t}) = D_{\boldsymbol{t}_2', \boldsymbol{t}_1'} X(\boldsymbol{t}). \tag{2.6}$$

does not hold. However, we can derive the commutativity as in (2.6) from the coincidence of mean square derivatives and almost sure derivatives if almost sure differential operators are commutable for X.

By (A.5) and (2.4), the following result is not surprising.

Lemma 2.1.3. *([4]) Let X be a centered Gaussian random field on $\mathbb{R}^N$. If some mean square derivative of X exists in $\mathbb{R}^N$, then this mean square derivative is also a centered Gaussian random field on $\mathbb{R}^N$. Moreover, let $Z_1(\boldsymbol{t}) = X(\boldsymbol{t})$, and let $Z_2(\boldsymbol{t}), \ldots, Z_n(\boldsymbol{t})$ ($n \geq 2$) be any $n - 1$ mean square derivatives of X at $\boldsymbol{t} \in \mathbb{R}$. Then for any $\boldsymbol{t}_1, \boldsymbol{t}_2, \ldots, \boldsymbol{t}_n \in \mathbb{R}^N$, $(Z_1(\boldsymbol{t}_1), Z_2(\boldsymbol{t}_2), \ldots, Z_n(\boldsymbol{t}_n))$ is a centered Gaussian n-vector.*

The following lemma explains why the mean square derivatives are so popular in the theory of random fields. Let $\boldsymbol{e}_i$, $1 \leq i \leq N$ be the N-vector with all zeros except for a one at the i-th coordinate. For $1 \leq k \leq N$ and any $1 \leq i_1, \ldots i_k \leq N$, denote by $X_{i_1 \ldots i_k}$ the k-th mean square derivative of X in the direction $(\boldsymbol{e}_{i_1}, \ldots, \boldsymbol{e}_{i_k})$. Let $X_{i_1 \ldots i_k} = X$ if $k = 0$.

Lemma 2.1.4. *(Section 5.5, [4]) Let $\{X(\boldsymbol{t}),\ \boldsymbol{t} \in \mathbb{R}^N\}$ be a Gaussian random field with covariance function*

$$r(\boldsymbol{s}, \boldsymbol{t}) := \operatorname{Cov}\left[X(\boldsymbol{s}), X(\boldsymbol{t})\right],$$

where $\boldsymbol{s} = (s_1, \ldots, s_N)^T$ and $\boldsymbol{t} = (t_1, \ldots, t_N)^T$.

(i) For any positive integers k and $i_1, \ldots, i_k \in \{1, 2, \ldots, N\}$, the k times mean square derivatives $X_{i_1 \ldots i_k}(\boldsymbol{t})$ exists if and only if the derivative $\frac{\partial^{2k} r(\boldsymbol{s}, \boldsymbol{t})}{\partial s_{i_1} \cdots \partial s_{i_k} \partial t_{i_1} \cdots \partial t_{i_k}}$ exists and is finite at the point $(\boldsymbol{t}, \boldsymbol{t}) \in \mathbb{R}^{2N}$.

(ii) For some positive integer k, suppose that the derivative $\frac{\partial^{2k} r(\boldsymbol{s}, \boldsymbol{t})}{\partial s_{i_1} \cdots \partial s_{i_k} \partial t_{j_1} \cdots \partial t_{j_k}}$ exists and is finite for any $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^N$ and $i_1, \ldots, i_k, j_1, \ldots, j_k \in \{1, 2, \ldots, N\}$ (by (i), this implies all the mean square derivatives of X up to k-th order at any $\boldsymbol{t} \in \mathbb{R}^N$ exist). Then we have for any $0 \leq k_1, k_2 \leq k$ and $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^N$,

$$\operatorname{Cov}\left[X_{i_1 \ldots i_{k_1}}(\boldsymbol{s}), X_{j_1 \ldots j_{k_2}}(\boldsymbol{t})\right] = \frac{\partial^{k_1 + k_2} r(\boldsymbol{s}, \boldsymbol{t})}{\partial s_{i_1} \cdots \partial s_{i_{k_1}} \partial t_{j_1} \cdots \partial t_{j_{k_2}}}. \tag{2.7}$$

Consider a centered, stationary random field $\{X(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^N\}$. The following theorem shows that its covariance function $r(\boldsymbol{t})$ has an integral representation. For completeness, this theorem is stated in the framework of complex-valued random fields. Let $\mathcal{C}$ be the set of complex number. For any $z \in \mathcal{C}$, denote by $\overline{z}$ the complex conjugate of z. Then the covariance function r of X is defined by

$$r(\boldsymbol{t} - \boldsymbol{s}) := \mathbb{E}\left[X(\boldsymbol{s})\overline{X(\boldsymbol{t})}\right].$$

A well-known property of the covariance function r is its positive semi-definiteness, i.e., for any $n \geq 1$, $z_1, \ldots, z_n \in \mathcal{C}$, and $\boldsymbol{t}_1, \ldots, \boldsymbol{t}_n \in \mathbb{R}^N$, we have

$$\sum_{1 \leq i,j \leq n} z_i r(\boldsymbol{t}_i - \boldsymbol{t}_j) \overline{z}_j \geq 0.$$

Theorem 2.1.5. *(Theorem 5.4.1, [4]) For a continuous function $r(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^N$, the following are equivalent:*

1. r is positive semi-definite;

2. there exists a finite measure ν on the Borel σ-field $\mathcal{B}(\mathbb{R}^N)$ such that

$$r(\boldsymbol{t}) = \int_{\mathbb{R}^N} e^{i<\boldsymbol{t},\boldsymbol{x}>}\nu(d\boldsymbol{x}),$$

where $< \cdot, \cdot >$ denotes the inner product of two vectors.

By Theorem 2.1.5 and the bounded convergence theorem, we can further calculate (2.7) in Lemma 2.1.4 by

$$\begin{aligned}
\operatorname{Cov}\left[X_{i_1\ldots i_{k_1}}(\boldsymbol{s}), X_{j_1\ldots j_{k_2}}(\boldsymbol{t})\right] &= \frac{\partial^{k_1+k_2} r(\boldsymbol{t}-\boldsymbol{s})}{\partial s_{i_1}\cdots\partial s_{i_{k_1}}\partial t_{j_1}\cdots\partial t_{j_{k_2}}} \\
&= \int_{\mathbb{R}^N} \frac{\partial^{k_1+k_2} e^{i<\boldsymbol{t}-\boldsymbol{s},\boldsymbol{x}>}}{\partial s_{i_1}\cdots\partial s_{i_{k_1}}\partial t_{j_1}\cdots\partial t_{j_{k_2}}}\nu(d\boldsymbol{x}) \\
&= (-1)^{k_1} i^{k_1+k_2} \int_{\mathbb{R}^N} x_{i_1}\cdots x_{i_{k_1}} x_{j_1}\cdots x_{j_{k_2}} e^{i<\boldsymbol{t}-\boldsymbol{s},\boldsymbol{x}>}\nu(d\boldsymbol{x}).
\end{aligned}$$

Then by the symmetry of the covariance function, any odd-ordered partial derivative of the continuous covariance function $r(\boldsymbol{t})$ at $\boldsymbol{t} = \boldsymbol{0}$ is zero. Immediately, in Lemma 2.1.4, we have

$$\operatorname{Cov}[X(\boldsymbol{t}), X_i(\boldsymbol{t})] = r_i(\boldsymbol{0}) = 0 \quad \text{and} \quad \operatorname{Cov}[X_i(\boldsymbol{t}), X_{k\ell}(\boldsymbol{t})] = r_{ik\ell}(\boldsymbol{0}) = 0 \tag{2.8}$$

for any $1 \le i, k, \ell \le N$, which will be very useful in the following chapters.

Recall that in (2.5), we actually defined a real-valued function $F_{X,k}(\boldsymbol{t}, \boldsymbol{t}')$ on the space $\mathbb{R}^N \times \bigotimes^k \mathbb{R}^N$ for each sample function of X. Endow the space $\mathbb{R}^N \times \bigotimes^k \mathbb{R}^N$ with the norm

$$\|(\boldsymbol{t}, \boldsymbol{t}')\|_{N,k} := \|\boldsymbol{t}\|_N + \|\boldsymbol{t}'\|_{kN}.$$

Denote

$$B_{N,k}((\boldsymbol{t}, \boldsymbol{t}'), h) := \left\{\boldsymbol{v} \in \mathbb{R}^N \times \bigotimes^k \mathbb{R}^N : \|\boldsymbol{v} - (\boldsymbol{t}, \boldsymbol{t}')\|_{N,k} < h\right\}$$

and

$$A_{k,\rho} := A \times \{\boldsymbol{t}' : \|\boldsymbol{t}'\|_{kN} \in (1-\rho, 1+\rho)\}$$

for any subset $A \subset \mathbb{R}^N$ and $\rho > 0$. Now we are well prepared to introduce a sufficient condition for the almost sure continuous differentiability by the existence of mean square derivatives.

Theorem 2.1.6. *(Theorem 1.4.2, [4]) Let X be a centered Gaussian random field on any open subset $A \subset \mathbb{R}^n$ with kth-order mean square partial derivatives in all directions everywhere inside A. If there exist constants $0 < K < \infty$ and $\rho, \delta, h_0 > 0$ such that for $0 < \eta_1, \eta_2, h < h_0$,*

$$\mathbb{E}\left[(F_{X,k}(\boldsymbol{t}, \eta_1 \boldsymbol{t}') - F_{X,k}(\boldsymbol{s}, \eta_2 \boldsymbol{s}'))^2\right] < K\left(-\log\left(\|(\boldsymbol{t}, \boldsymbol{t}') - (\boldsymbol{s}, \boldsymbol{s}')\|_{N,k} + |\eta_1 - \eta_2|\right)\right)^{-(1+\delta)} \tag{2.9}$$

for all

$$\left\{((\boldsymbol{t}, \boldsymbol{t}'), (\boldsymbol{s}, \boldsymbol{s}')) \in A_{k,\rho} \times A_{k,\rho} : (\boldsymbol{s}, \boldsymbol{s}') \in B_{N,k}((\boldsymbol{t}, \boldsymbol{t}'), h)\right\},$$

then X has an almost surely k times continuously differentiable modification, i.e.,

$$P\left[X \in C^k(A)\right] = 1.$$

In Example A.2.2, we show that a centered Gaussian random field with covariance function $r(\boldsymbol{t}) = e^{-\|\boldsymbol{t}\|^2}$, $\boldsymbol{t} \in \mathbb{R}^N$ satisfies all the conditions in Theorem 2.1.6.

2.1.4 The Metatheorem

The construction of the point process we study in this thesis requires some knowledge of the expected number of the critical points with different indices in a Gaussian excursion set. The book [4] derived an important theorem which provides an integral expression in a more general setting. To introduce the theorem, we need the following concepts. The modulus of continuity ω_F of a real-valued function F on $(\mathbb{R}^N, \|\cdot\|)$ is defined as

$$\omega_F(\delta) = \sup_{\|\boldsymbol{t}-\boldsymbol{s}\|<\delta} |F(\boldsymbol{t}) - F(\boldsymbol{s})|, \quad \delta > 0.$$

The diameter of a set $B \subset \mathbb{R}^N$ is

$$\operatorname{diam}(B) := \sup\left\{\|\boldsymbol{t} - \boldsymbol{s}\| : \boldsymbol{s}, \boldsymbol{t} \in B\right\}.$$

The Hausdorff dimension of a set $A \subset \mathbb{R}^N$ is given by

$$H_{dim}(A) := \inf\left\{\alpha : \liminf_{\varepsilon \downarrow 0} \sum_i (\operatorname{diam} B_i)^\alpha = 0\right\},$$

where the infimum is taken over all collections $\{B_i\}$ of open balls in $\mathbb{R}^N$ such that

1. $A \subset \bigcup_i B_i$
2. $\operatorname{diam}(B_i) < \varepsilon$ for any i.

For some $N, K \geq 1$, let $\boldsymbol{f} := (f^1, \ldots, f^N)$ and $\boldsymbol{g} := (g^1, \ldots, g^K)$ be $\mathbb{R}^N$- and $\mathbb{R}^K$-valued random fields indexed by $\mathbb{R}^N$, respectively. Since we have used subscripts to denote the derivatives of random fields, here we use superscripts to denote components of vector-valued random fields. In this section, all the derivatives are supposed to exist in the almost sure sense. The gradient $\triangledown \boldsymbol{f}$ is defined as

$$\triangledown \boldsymbol{f} := \left(f_j^i\right)_{1 \leq i,j \leq N},$$

where

$$f_j^i := \frac{\partial f^i}{\partial t_j}$$

for $1 \leq i, j \leq N$. For any $\boldsymbol{u} \in \mathbb{R}^N$, we are interested in the number of points $\boldsymbol{t} \in T$, written as $N_{\boldsymbol{u}}(\boldsymbol{f}, \boldsymbol{g}; T, B)$, for which

$$\boldsymbol{f}(\boldsymbol{t}) = \boldsymbol{u} \quad \text{and} \quad \boldsymbol{g}(\boldsymbol{t}) \in B \subset \mathbb{R}^K, \tag{2.10}$$

where $T \subset \mathbb{R}^N$ is a compact subset with $\lambda_{N-1}(\partial T) < \infty$ and $B \subset \mathbb{R}^K$ is an open set whose boundary ∂B has Hausdorff dimension $K - 1$.

Theorem 2.1.7. *(Theorem 11.2.1, [4]) Let $\boldsymbol{f}, \boldsymbol{g}, T$ and B be as in (2.10) and $p_{\boldsymbol{t}}(\boldsymbol{x}, \boldsymbol{x}', \boldsymbol{y})$ be the density of $(\boldsymbol{f}(\boldsymbol{t}), \triangledown \boldsymbol{f}(\boldsymbol{t}), \boldsymbol{g}(\boldsymbol{t}))$. Suppose that the following conditions hold for some $\boldsymbol{u} \in \mathbb{R}^N$:*

(1) Continuities of sample paths:

(a) All components of $\boldsymbol{f}$, $\triangledown \boldsymbol{f}$, and $\boldsymbol{g}$ are almost surely continuous and have finite variances over T.

(b) The moduli of continuity with respect to the usual Euclidean norm for each components of $\boldsymbol{f}$, $\triangledown \boldsymbol{f}$, and $\boldsymbol{g}$ satisfy

$$P\left[\omega(\delta) > \varepsilon\right] = o(\delta^N) \quad \text{as } \delta \downarrow 0$$

for any $\varepsilon > 0$.

(2) Continuities of densities:

(a) *For all* $\boldsymbol{t} \in T$*, the marginal densities* $p_{\boldsymbol{t}}(\boldsymbol{x})$ *of* $\boldsymbol{f}(\boldsymbol{t})$ *are continuous at* $\boldsymbol{x} = \boldsymbol{u}$*.*

(b) *The conditional densities* $p_{\boldsymbol{t}}(\boldsymbol{x}|\nabla \boldsymbol{f}(\boldsymbol{t}), \boldsymbol{g}(\boldsymbol{t}))$ *of* $\boldsymbol{f}(\boldsymbol{t})$ *given* $\boldsymbol{g}(\boldsymbol{t})$ *and* $\nabla \boldsymbol{f}(\boldsymbol{t})$ *are uniformly bounded from above and uniformly continuous at* $\boldsymbol{x} = \boldsymbol{u}$ *for all* $\boldsymbol{t} \in T$*.*

(c) *The conditional densities* $p_{\boldsymbol{t}}(\boldsymbol{z}|\boldsymbol{f}(\boldsymbol{t}) = \boldsymbol{x})$ *of* $\det(\nabla \boldsymbol{f}(\boldsymbol{t}))$ *given* $\boldsymbol{f}(\boldsymbol{t}) = \boldsymbol{x}$ *are continuous for* $\boldsymbol{z}$ *and* $\boldsymbol{x}$ *in some neighborhoods of* $\boldsymbol{0}$ *and* $\boldsymbol{u}$*, respectively, uniformly in* $\boldsymbol{t} \in T$*.*

(d) *The conditional densities* $p_{\boldsymbol{t}}(\boldsymbol{y}|\boldsymbol{f}(\boldsymbol{t}) = \boldsymbol{x})$ *of* $\boldsymbol{g}(\boldsymbol{t})$ *given* $\boldsymbol{f}(\boldsymbol{t}) = \boldsymbol{x}$ *are continuous for all* $\boldsymbol{y}$ *and for* $\boldsymbol{x}$ *in a neighborhood of* $\boldsymbol{u}$*, uniformly in* $\boldsymbol{t} \in T$*.*

(3) *The moment condition:*

$$\sup_{\boldsymbol{t}\in T} \max_{1\le i,j\le N} \mathbb{E}\left[\left|f_j^i(\boldsymbol{t})\right|^N\right] < \infty$$

holds.

Then we have

$$\mathbb{E}\left[N_{\boldsymbol{u}}(\boldsymbol{f}, \boldsymbol{g}; T, B)\right] = \int_T \int_{\mathbb{R}^K} \int_{\mathbb{R}^{N(N+1)/2}} |\det(\boldsymbol{x}')|\, \mathbb{1}_B(\boldsymbol{y}) p_{\boldsymbol{t}}(\boldsymbol{u}, \boldsymbol{x}', \boldsymbol{y}) d\boldsymbol{x}' d\boldsymbol{y} d\boldsymbol{t},$$

where

$$\mathbb{1}_B(\boldsymbol{y}) := \begin{cases} 1 & \text{if } \boldsymbol{y} \in B, \\ 0 & \text{if } \boldsymbol{y} \notin B. \end{cases}$$

Note that in Theorem 2.1.7, the set B is forced to be open. This condition can be weakened by the following lemma.

Lemma 2.1.8. *(Lemma 11.2.12, [4]) Let* $\boldsymbol{f}$*,* $\boldsymbol{g}$*,* T *and* B *be as in (2.10). Suppose that conditions* (1)*,* (2*a*)*, and* (2*d*) *of Theorem 2.1.7 holds. Then, with probability one, there is no point* $\boldsymbol{t} \in T$ *satisfying* $\boldsymbol{f}(\boldsymbol{t}) - \boldsymbol{u} = \boldsymbol{0}$ *and* $\boldsymbol{g}(\boldsymbol{t}) \in \partial B$*.*

Also note that the continuities of densities and the moment condition will be automatically satisfied if $\boldsymbol{f}$ and $\boldsymbol{g}$ are both vector-valued Gaussian random fields such that the distribution of $(\boldsymbol{f}(\boldsymbol{t}), \nabla \boldsymbol{f}(\boldsymbol{t}), \boldsymbol{g}(\boldsymbol{t}))$ is non-degenerate. In fact, by the condition of Theorem 2.1.7, if we let $\boldsymbol{h}(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^N$ be the remaining components of $(\nabla \boldsymbol{f}(\boldsymbol{t}), \boldsymbol{g}(\boldsymbol{t}))$ after removing all the components of $\boldsymbol{g}(\boldsymbol{t})$ which are also components of $\nabla \boldsymbol{f}(\boldsymbol{t})$, then we only need the distribution of $(\boldsymbol{f}(\boldsymbol{t}), \boldsymbol{h}(\boldsymbol{t}))$ to be non-degenerate. Moreover, in Lemma 2.1.1, we have seen that the continuity of sample paths of a Gaussian random field can be fully controlled by its covariance structure. Therefore, [4] also derived the following corollary which looks more friendly for the setting of this thesis.

Corollary 2.1.9. *(Corollary 11.2.2, [4]) Let $\boldsymbol{f}$ and $\boldsymbol{g}$ be centered Gaussian random fields over T, as in (2.10), and let $\boldsymbol{h}$ be as described above such that for each $\boldsymbol{t} \in T$, the distributions of $(\boldsymbol{f}(\boldsymbol{t}), \boldsymbol{h}(\boldsymbol{t}))$ are non-degenerate. Denote by $C_{f_j^i}(\boldsymbol{s}, \boldsymbol{t})$ the covariance function of f_j^i for $1 \le i, j \le N$. Similarly, denote by $C_{g_j^i}(\boldsymbol{s}, \boldsymbol{t})$ the covariance function of g_j^i for $1 \le i \le K$ and $1 \le j \le N$. Suppose*

$$\begin{aligned}
\max_{1 \le i,j \le N} \left| C_{f_j^i}(\boldsymbol{t}, \boldsymbol{t}) + C_{f_j^i}(\boldsymbol{s}, \boldsymbol{s}) - 2C_{f_j^i}(\boldsymbol{s}, \boldsymbol{t}) \right| &\le K |\log(\|\boldsymbol{t} - \boldsymbol{s}\|)|^{-(1+\alpha)}, \\
\max_{1 \le i,j \le N} \left| C_{g_j^i}(\boldsymbol{t}, \boldsymbol{t}) + C_{g_j^i}(\boldsymbol{s}, \boldsymbol{s}) - 2C_{g_j^i}(\boldsymbol{s}, \boldsymbol{t}) \right| &\le K |\log(\|\boldsymbol{t} - \boldsymbol{s}\|)|^{-(1+\alpha)},
\end{aligned} \tag{2.11}$$

for some finite $K > 0$, some $\alpha > 0$, and all $\boldsymbol{s}, \boldsymbol{t} \in T$ such that $\|\boldsymbol{t} - \boldsymbol{s}\|$ is small enough. Then for any $\boldsymbol{u} \in \mathbb{R}^N$,

$$\mathbb{E}\left[N_{\boldsymbol{u}}(\boldsymbol{f}, \boldsymbol{g}; T, B)\right] = \int_T \int_{\mathbb{R}^K} \int_{\mathbb{R}^{N(N+1)/2}} |\det(\boldsymbol{x}')| \, \mathbb{1}_B(\boldsymbol{y}) p_{\boldsymbol{t}}(\boldsymbol{u}, \boldsymbol{x}', \boldsymbol{y}) d\boldsymbol{x}' d\boldsymbol{y} d\boldsymbol{t}.$$

Let $T \subset \mathbb{R}^N$ be a compact subset whose boundary has finite $(N-1)$-dimensional Lebesgue measure. Let $\{X(\boldsymbol{t}), \boldsymbol{t} = (t_1, \dots, t_N)^T \in T\}$ be a centered Gaussian random field on T possessing up-to second-order almost sure derivatives. Its gradient $\triangledown X$ and Hessian matrix $\triangledown^2 X$ are defined as

$$\triangledown X := (X_1, \dots, X_N) \text{ and } \triangledown^2 X := (X_{ij})_{1 \le i,j \le N},$$

respectively, where for $1 \le i, j \le N$,

$$X_i = \frac{\partial X}{\partial t_i} \text{ and } X_{ij} = \frac{\partial^2 X}{\partial t_i \partial t_j}.$$

For any $\boldsymbol{t} \in T$, a real-valued random vector

$$\boldsymbol{a}(\boldsymbol{t}) = (a_1(\boldsymbol{t}), \dots, a_{N(N+1)/2}(\boldsymbol{t}))^T$$

is said to be the usual vectorization of $\triangledown^2 X(\boldsymbol{t})$ if

$$a_{i+j(j-1)/2}(\boldsymbol{t}) = X_{ij}(\boldsymbol{t}) \text{ for any integers } 1 \le i \le j \le N.$$

(Note that if X_{ij}, $1 \le i, j \le N$ are all continuous on T, which is always the case in the main results of this thesis, then $\triangledown^2 X$ is symmetric and can be fully characterized by its usual vectorization) For convenience, we will not distinguish the $N \times N$ matrix $\triangledown^2 X$ from its usual vectorization in notations, but one can easily distinguish them from a given context.

By Lemma 2.1.3, $(\nabla^2 X(\boldsymbol{t}), \nabla X(\boldsymbol{t}), X(\boldsymbol{t}))$ is a centered Gaussian vector. Denote by $N_{\boldsymbol{v},B}(T)$ the number of points $\boldsymbol{t} \in T$, such that

$$\nabla X(\boldsymbol{t}) = \boldsymbol{v} \quad \text{and} \quad (\nabla^2 X(\boldsymbol{t}), X(\boldsymbol{t})) \in B, \tag{2.12}$$

where $\boldsymbol{v} \in \mathbb{R}^N$ and $B \subset \mathbb{R}^{N(N+1)/2+1}$ is an open set whose boundary has the Hausdorff dimension $N(N+1)/2$. Then by letting $\boldsymbol{f} = \nabla X$ and $\boldsymbol{g} = (\nabla^2 X, X)$ in Corollary 2.1.9, we have the following lemma.

Lemma 2.1.10. *(Corollary 11.2.2, [4]) Let X, T, $\boldsymbol{v}$ and B be as above. Suppose that the distribution of $(\nabla^2 X(\boldsymbol{t}), \nabla X(\boldsymbol{t}), X(\boldsymbol{t}))$ is non-degenerate for any $\boldsymbol{t} \in T$. For any $i, j \in \{1, \dots, N\}$ and $\boldsymbol{s}, \boldsymbol{t} \in T$, denote*

$$C_{ij}(\boldsymbol{s}, \boldsymbol{t}) := \text{Cov}\left[X_{ij}(\boldsymbol{s}), X_{ij}(\boldsymbol{t})\right].$$

Suppose

$$\max_{1 \le i,j \le N} |C_{ij}(\boldsymbol{t}, \boldsymbol{t}) + C_{ij}(\boldsymbol{s}, \boldsymbol{s}) - 2C_{ij}(\boldsymbol{s}, \boldsymbol{t})| \le K |\log(\|\boldsymbol{t} - \boldsymbol{s}\|)|^{-(1+\alpha)}, \tag{2.13}$$

for some finite $K > 0$, some $\alpha > 0$, and all $\boldsymbol{s}, \boldsymbol{t} \in T$ such that $\|\boldsymbol{t} - \boldsymbol{s}\|$ is small enough. Then we have

$$\mathbb{E}\left[N_{\boldsymbol{v},B}(T)\right] = \int_{\boldsymbol{t} \in T} \int_{(\boldsymbol{x}'', x) \in B} |\det(\boldsymbol{x}'')|\, p_{\boldsymbol{t}}(\boldsymbol{x}'', \boldsymbol{v}, x) d\boldsymbol{x}'' dx d\boldsymbol{t},$$

where $p_{\boldsymbol{t}}(\boldsymbol{x}'', \boldsymbol{x}', x)$ is the density of the Gaussian vector $(\nabla^2 X(\boldsymbol{t}), \nabla X(\boldsymbol{t}), X(\boldsymbol{t}))$.

Remark 2.1.11. As discussed in the paragraphs just below Theorem 11.2.1 in Page 268 of [4], Condition (2.13) is only used to ensure that Condition (1) in Theorem 2.1.7 holds. If X satisfies all the conditions in Lemma 2.1.10 except that the mean function of X is non-centered and continuously twice differentiable on T, then one can easily follow these paragraphs to show that Condition (1) in Theorem 2.1.7 still holds. As a result, Lemma 2.1.10 still holds for such a non-centered Gaussian random field.

Remark 2.1.12. By taking $\boldsymbol{f}(\boldsymbol{t}) = \nabla X(\boldsymbol{t})$ and $\boldsymbol{g}(\boldsymbol{t}) = (\nabla^2 X(\boldsymbol{t}), X(\boldsymbol{t}))$ for $\boldsymbol{t} \in T$ in Lemma 2.1.8, we have

$$P\left[\#\left\{\boldsymbol{t} \in T : X(\boldsymbol{t}) > u, \nabla X(\boldsymbol{t}) = \boldsymbol{0}, \det(\nabla^2 X(\boldsymbol{t})) = 0\right\} = 0\right] = 1.$$

Therefore, with probability one, all the critical points in the excursion set $A_u(X, T)$ are non-degenerate.

When X is also stationary, the following Lemma gives a sufficient condition for (2.13) to hold.

Lemma 2.1.13. *Let $\{X(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^N\}$ be a stationary Gaussian random field possessing up-to second-order almost sure derivatives. Let $r(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^N$ be the covariance function of X. Suppose that all of the sixth-order partial derivatives of $r(\boldsymbol{t})$ exist at $\boldsymbol{t} = \mathbf{0}_N$. Then Condition (2.13) holds for all $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^N$ such that $\|\boldsymbol{t} - \boldsymbol{s}\|$ is small enough.*

Proof. See Appendix A.2.3. □

In Example A.2.3, we take $\boldsymbol{f} = \nabla X$ and $\boldsymbol{g} = (\nabla^2 X, X)$, where X is a centered, stationary Gaussian random field with covariance function $r(\boldsymbol{t}) = e^{-\|\boldsymbol{t}\|^2}$, $\boldsymbol{t} \in \mathbb{R}^N$. Then we show that $\boldsymbol{f}$ and $\boldsymbol{g}$ satisfy all the conditions in Lemma 2.1.10.

2.2 Point Processes

2.2.1 Point Processes and Random Measures

In this section, we will review the basic concepts for point processes on $\mathbb{R}^N$ ($N \geq 1$). Endow $\mathbb{R}^N$ with the usual Euclidean norm. Denote by $\mathcal{K}(\mathbb{R}^N)$ the set of all compact subsets of $\mathbb{R}^N$. A measure μ on $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$ is said to be locally finite if $\mu(K) < \infty$ for all $K \in \mathcal{K}(\mathbb{R}^N)$. On $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$, denote by $M_+(\mathbb{R}^N)$ the set of all locally finite measures and by $M_p(\mathbb{R}^N)$ the subset of $M_+(\mathbb{R}^N)$ whose elements take values in $\mathbb{N}_+ = \{0, 1, 2, \ldots, \infty\}$, i.e.,

$$M_p\left(\mathbb{R}^N\right) := \left\{\mu \in M_+\left(\mathbb{R}^N\right) : \mu(B) \in \mathbb{N}_+, B \in \mathcal{B}\left(\mathbb{R}^N\right)\right\}.$$

The next step is to endow the space $M_+(\mathbb{R}^N)$ and $M_p(\mathbb{R}^N)$ with σ-fields. For $M_+(\mathbb{R}^N)$, it is reasonable to consider that the mappings $\mu \to \mu(B)$ for any $\mu \in M_+(\mathbb{R}^N)$ and $B \in \mathcal{B}(\mathbb{R}^N)$, should be all measurable with respect to this σ-field. Therefore, the σ-field, $\mathcal{M}_+(\mathbb{R}^N)$, can be simply taken as the smallest σ-field containing the sets of the form

$$\left\{\mu \in M_+\left(\mathbb{R}^N\right) : \mu(B) \in G\right\} \quad \text{for } B \in \mathcal{B}\left(\mathbb{R}^N\right) \text{ and } G \in \mathcal{B}\left([0, \infty]\right).$$

Similarly, we can also endow the space $M_p(\mathbb{R}^N)$ with the σ-field $\mathcal{M}_p(\mathbb{R}^N)$ which is the smallest σ-field containing the sets of the form

$$\left\{\mu \in M_p\left(\mathbb{R}^N\right) : \mu(B) \in G\right\} \quad \text{for } B \in \mathcal{B}\left(\mathbb{R}^N\right) \text{ and } G \in \mathcal{B}\left([0, \infty]\right).$$

A random measure or a point process on $\mathbb{R}^N$ is defined to be any measurable mapping from the probability space $(\Omega, \mathcal{F}, P)$ to $(M_+(\mathbb{R}^N), \mathcal{M}_+(\mathbb{R}^N))$ or $(M_p(\mathbb{R}^N), \mathcal{M}_p(\mathbb{R}^N))$, respectively. We shall allow a point process to take values outside $\mathbb{N}_+$ with probability zero. Therefore a point process can be considered as an almost surely $\mathbb{N}_+$-valued random measure. The following lemma shows a decomposition of a point process.

Lemma 2.2.1. *(Lemmas 2.1 and 2.3, [18]) A point process ξ on $\mathbb{R}^N$ can be decomposed as*

$$\xi = \sum_{j=1}^{\nu} \beta_j \delta_{\boldsymbol{\tau}_j},$$

where ν is a random variable taking values in $\mathbb{N}_+$, $\beta_j, 1 \le j \le \nu$, are random variables taking values in $\{1, 2, \dots\}$, $\boldsymbol{\tau}_j, 1 \le j \le \nu$, are random elements in $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$, and

$$\delta_{\boldsymbol{t}}(A) := \begin{cases} 1 & \boldsymbol{t} \in A, \\ 0 & \boldsymbol{t} \notin A \end{cases}$$

for any $\boldsymbol{t} \in \mathbb{R}^N$ and $A \in \mathcal{B}(\mathbb{R}^N)$.

A point process ξ on $\mathbb{R}^N$ is said to be simple if its distribution concentrates on the simple point measures of $M_p(\mathbb{R}^N)$:

$$P\left[\xi(\{\boldsymbol{t}\}) \le 1 \text{ for all } \boldsymbol{t} \in \mathbb{R}^N\right] = 1.$$

A Radon measure (the measure of compact sets is always finite) μ is said to be the mean measure (or the intensity measure) of ξ if it satisfies

$$\mu(B) = \mathbb{E}[\xi(B)] \text{ for any } B \in \mathcal{B}(\mathbb{R}^N).$$

Thus, if $\mu(B) = \infty$, then $\xi(B) = \infty$ almost surely.

2.2.2 Weak Convergence of Random Measures

To establish the weak convergence of random measures, the first step is to show that the space $(M_+(\mathbb{R}^N), \mathcal{M}_+(\mathbb{R}^N))$ is metrizable, i.e., we can find a metric on $M_+(\mathbb{R}^N)$ such that the σ-field $\mathcal{M}_+(\mathbb{R}^N)$ can be induced by this metric. To this end, we need to first topologize $M_+(\mathbb{R}^N)$ by introducing the so-called vague convergence. Let $\mathcal{F}^+(\mathbb{R}^N)$ be the set of measurable mappings from $(\mathbb{R}^N, \mathcal{B}(\mathbb{R}^N))$ to $([0,\infty], \mathcal{B}([0,\infty]))$. For any locally finite measure $\mu \in M_+(\mathbb{R}^N)$ and function $f \in \mathcal{F}^+(\mathbb{R}^N)$, we can define the following integral

$$\mu(f) := \int_{\mathbb{R}^N} f(\boldsymbol{t})\mu(d\boldsymbol{t}).$$

Denoted by $C_K^+(\mathbb{R}^N)$ the set of continuous, nonnegative functions with compact support. For $\mu_n, \mu \in M_+\left(\mathbb{R}^N\right)$, we say μ_n vaguely converges to μ if for all $f \in C_K^+(\mathbb{R}^N)$,

$$\lim_{n\to\infty} \mu_n(f) = \mu(f).$$

Then a sub-base of the vague topology on $M_+(\mathbb{R}^N)$ consists of the sets of the form

$$\{\mu \in M_+(\mathbb{R}^N) : \mu(f) \in (s,t)\}$$

for some $f \in C_K^+(\mathbb{R}^N)$ and $s < t$. Let $\mathcal{B}(M_+(\mathbb{R}^N))$ be the Borel σ-field generated by vaguely open subsets of $M_+(\mathbb{R}^N)$, and let $\mathcal{M}_+^v(\mathbb{R}^N)$ be the smallest σ-field containing the sets of the form

$$\left\{\mu \in M_+\left(\mathbb{R}^N\right) : \mu(f) \in G\right\} \text{ for } f \in C_K^+(\mathbb{R}^N) \text{ and } G \in \mathcal{B}\left([0,\infty]\right).$$

So far, we have established three σ-fields, $\mathcal{M}_+(\mathbb{R}^N)$, $\mathcal{B}(M_+(\mathbb{R}^N))$ and $\mathcal{M}_+^v(\mathbb{R}^N)$, on $M_+(\mathbb{R}^N)$. The following lemma shows the relationship between them.

Lemma 2.2.2. *([34]) We have the following relationships:*

1. $\mathcal{M}_+^v(\mathbb{R}^N) = \mathcal{M}_+(\mathbb{R}^N)$;
2. $\mathcal{M}_p(\mathbb{R}^N) = \{A \cap M_p(\mathbb{R}^N) : A \in \mathcal{M}_+^v(\mathbb{R}^N)\}$;
3. $M_p(\mathbb{R}^N)$ *is vaguely closed in* $M_+(\mathbb{R}^N)$;
4. $\mathcal{M}_+^v(\mathbb{R}^N) = \mathcal{B}(M_+(\mathbb{R}^N))$.

The second item in Lemma 2.2.2 implies that $\mathcal{M}_p(\mathbb{R}^N)$ is the same as the smallest σ-field containing the sets of the form

$$\left\{\mu \in M_p\left(\mathbb{R}^N\right) : \mu(f) \in G\right\} \quad \text{for } f \in C_K^+(\mathbb{R}^N) \text{ and } G \in \mathcal{B}([0,\infty]).$$

The third item in Lemma 2.2.2 implies that the topological information about $M_+(\mathbb{R}^N)$ can actually be inherited by $M_p(\mathbb{R}^N)$. So far, we have restated $\mathcal{M}_+(\mathbb{R}^N)$ in terms of vague convergence. The following lemma shows that $\mathcal{M}_+(\mathbb{R}^N)$ is metrizable.

Lemma 2.2.3. *(Proposition 3.17, [34]) The vague topology on $M_p(\mathbb{R}^N)$ or $M_+(\mathbb{R}^N)$ is metrizable as a complete, separable metric space.*

Remark 2.2.4. It is complicated to actually give the metric. One can refer to the proof of Proposition 3.17 in [34] for more details.

Now denote by $C(M_+(\mathbb{R}^N))$ the bounded, continuous real-valued functions on the complete, separable metrizable space $(M_+(\mathbb{R}^N), \mathcal{M}_+(\mathbb{R}^N))$. Then for random measures ξ_n and ξ on $\mathbb{R}^N$, we say that ξ_n converges weakly to ξ (written as $\xi_n \Rightarrow \xi$) if $\lim_{n\to\infty} \mathbb{E}[f(\xi_n)] = \mathbb{E}[f(\xi)]$ for any $f \in C(M_+(\mathbb{R}^N))$. Since we are more concerned about the weak convergence of point processes, the following lemma will be useful in the Chapter 3.

Lemma 2.2.5. *([18]) Let $\mathcal{J}$ be a basis of relatively compact sets such that $\mathcal{J}$ is closed under finite unions and intersections. Let ξ be a simple point process on $\mathbb{R}^N$ such that for any $I \in \mathcal{J}$, $P\left[\xi(\partial I) = 0\right] = 1$. Suppose that $\xi_n, n \geq 1$, are point processes on $\mathbb{R}^N$ and for all $I \in \mathcal{J}$*

- $\lim_{n\to\infty} P[\xi_n(I) = 0] = P[\xi(I) = 0]$*;*
- $\limsup_{n\to\infty} \mathbb{E}\left[\xi_n(I)\right] \leq \mathbb{E}\left[\xi(I)\right] < \infty$.

Then as $n \to \infty$,

$$\xi_n \Rightarrow \xi \text{ in } M_+(\mathbb{R}^N).$$

Here $P[\xi(\cdot) = 0]$ is called the avoidance functional of ξ on $\mathcal{B}(\mathbb{R}^N)$. The book [34] introduced the following result which shows the importance of avoidance functionals in characterizing point processes. Let $\mathcal{J}$ be a basis of relatively compact sets such that $\mathcal{J}$ is closed under finite unions and intersections. Suppose that ξ_1, ξ_2 are two simple point processes on $\mathbb{R}^N$ such that

$$P[\xi_1(I) = 0] = P[\xi_2(I) = 0]$$

for any $I \in \mathcal{J}$. Then ξ_1 and ξ_2 have the same distribution.

At the end of this section, we would like to introduce a useful application of the weak convergence of random measures. For any $f \in \mathcal{F}^+(\mathbb{R}^N)$, denote by D_f the set of all discontinuity points of f.

Lemma 2.2.6. *(Lemma 4.4, [18]) Let $\xi, \xi_1, \xi_2, \ldots$ be random measures on $\mathbb{R}^N$ such that $\xi_n \Rightarrow \xi$. Then for every bounded function $f \in \mathcal{F}^+(\mathbb{R}^N)$ with bounded support and satisfying $\xi(D_f)$=0 almost surely, we have $\xi_n(f)$ converges to $\xi(f)$ weakly as $n \to \infty$.*

2.2.3 Poisson Point Processes

A point process ξ on $\mathbb{R}^N$ is called a Poisson point process or a Poisson random measure if there exists a Radon measure μ on $\mathcal{B}(\mathbb{R}^N)$, such that

1. For any $B \in \mathcal{B}(\mathbb{R}^N)$ and $k \in \{0, 1, 2, \dots\}$,

$$P[\xi(B) = k] = \begin{cases} \dfrac{\mu(B)^k e^{-\mu(B)}}{k!} & \text{if } \mu(B) < \infty \\ 0 & \text{if } \mu(B) = \infty. \end{cases}$$

2. For any $n \geq 1$, if $B_1, \dots, B_n \in \mathcal{B}(\mathbb{R}^N)$ are mutually disjoint sets, then $\xi(B_i), 1 \leq i \leq n$ are mutually independent random variables.

It is easy to see μ is the mean measure of ξ. To show a Poisson limit for a series of point processes ξ_n on $\mathbb{R}^N$, we can further simplify Lemma 2.2.5 as follows.

Lemma 2.2.7. *([18]) Let $\mathcal{J}$ be the set of bounded open N-dimensional cubes and their finite unions and intersections. Suppose that ξ_n, $n \geq 1$, are point processes on $\mathbb{R}^N$ and there exists a locally finite measure $\mu \in M_+(\mathbb{R}^N)$ such that for all $I \in \mathcal{J}$*

- $\lim_{n\to\infty} P[\xi_n(I) = 0] = \exp\{-\mu(I)\}$;
- $\limsup_{n\to\infty} \mathbb{E}[\xi_n(I)] \leq \mu(I)$.

Then ξ_n converges weakly to a Poisson point process with mean measure μ, as $n \to \infty$.

2.3 Morse Functions and Inequalities

Let G be a non-empty open subset of the norm space $(\mathbb{R}^N, \|\cdot\|)$ $(N \geq 1)$. In this section, we temporarily drop randomness and focus on a real-valued function $f \in C^2(G)$. By convention, we denote the gradient and the Hessian matrix of f at $\boldsymbol{t} \in G$ by $\nabla f(\boldsymbol{t})$ and $\nabla^2 f(\boldsymbol{t})$, respectively. Then a point $\boldsymbol{t} \in G$ is said to be (the position of) a critical point of f with index (or type) k for some integer $0 \leq k \leq N$ if

$$\nabla f(\boldsymbol{t}) = 0 \ \text{ and } \ \text{index}\left(\nabla^2 f(\boldsymbol{t})\right) = k,$$

where for any square matrix $A \in \mathbb{R}^{n\times n}$, the index$(A)$ is defined to be the number of its negative eigenvalues counted with their multiplicities. Moreover, a point $\boldsymbol{t} \in G$ is said to be a (strict) local maximum of f if there exists a neighborhood $N_{\boldsymbol{t}} \subset G$ of $\boldsymbol{t}$, such that

$$f(\boldsymbol{t}) \geq f(\boldsymbol{s}) \quad (f(\boldsymbol{t}) > f(\boldsymbol{s})) \quad \text{for any } \boldsymbol{s} \in N_{\boldsymbol{t}}.$$

By treating f as a sample path, we can easily define the critical points and the local maxima of a two times almost surely differentiable random field indexed by G. Immediately, we can observe that a local maximum of f is also a critical point of f, and a critical point of f with index N must be a strict local maximum. A critical point $\boldsymbol{t}$ of f is said to be non-degenerate if

$$\det\left(\nabla^2 f(\boldsymbol{t})\right) \neq 0.$$

Then f is said to be a Morse function or non-degenerate on the open set G if every critical point of f in G is non-degenerate.

To formally state Morse inequalities, we need to first introduce some concepts in differential topology. One can refer to [26] and [1] for more details. M is said to be a topological N-manifold ($N \geq 1$) if it is Hausdorff, second countable, and for any $\boldsymbol{t} \in M$, there exists an open neighborhood of $\boldsymbol{t}$ which is homeomorphic to an open subset of $\mathbb{R}^N$. A chart (U, φ) on the topological space M is defined as an open subset $U \subset M$ together with a homeomorphism φ from G to an open subset of $\mathbb{R}^n$. For any two charts (U_1, φ_1) and (U_2, φ_2) on M such that $U_1 \cap U_2 \neq \emptyset$, the compositions $g_1 \circ g_2^{-1}$ and $g_2 \circ g_1^{-1}$ are called the transition maps of these two charts. If a collection of charts on M forms a cover of M, then this collection is called an atlas of M. Furthermore, if all of the transition maps in this atlas are k times differentiable, then this atlas is called a C^k-atlas ($k \geq 0$). Two C^k-atlas on M are said to be compatible if their union is still a C^k-atlas on M. An atlas on M is said to be maximal if it is only compatible with itself. Then the topological N-manifold M together with an maximal C^k-atlas on M forms an N-dimensional differentiable manifold of class C^k.

A compact set $K \subset \mathbb{R}^N$ is said to be a regular C^2-domain in $\mathbb{R}^N$ if its boundary ∂K is a regular $(N-1)$-dimensional differentiable manifold of class C^2 possessing a finite number of connected components (see [26], Chapters 1-5, for the formal definition of the regularity of a C^2-manifold). With the above definitions, we can further define the concept of being "admissible relative to a regular C^2-domain" as follows.

Definition 2.3.1. ([26]) A real-valued function $f(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^N$ ($N \geq 1$) is said to be admissible relative to a regular C^2-domain $K \subset \mathbb{R}^N$ if

1. $f \in C^2(G)$ for some open set $G \supset K$;

2. f has no critical points on ∂K;

3. the restrictions of f to K and ∂K, $f|_K$ and $f|_{\partial K}$ are both non-degenerate.

For a regular C^2-domain $K \subset \mathbb{R}^N$, denote by ∂K^+ the open subset of ∂K for which the directional derivative of f in the direction of the outward normal to ∂K at $\boldsymbol{t}$ is positive. For any integer $0 \leq i \leq N$, let $M(f, K, i)$ be the number of the critical points of f with index i in K, and let $CC(K)$ be the number of connected components of K. Similarly, we can let $M(f|_{\partial K^+}, \partial K^+, i)$, $0 \leq i \leq N-1$ be the number of the critical points of $f|_{\partial K^+}$ with index i in ∂K^+. The following theorem is a partial result of Morse inequalities which discusses the relationship between the number of critical points of a Morse function and the number of connected components of the area it lives on.

Theorem 2.3.2. *(Theorem 10.2, [26]) Let $f(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^N$ $(N \geq 1)$ be a Morse function on an open set containing the regular C^2-domain $K \subset \mathbb{R}^N$. Suppose that f is also admissible relative to K. Then we have*

$$\begin{aligned} &M(f, K, N) + M(f|_{\partial K^+}, \partial K^+, N-2) - M(f, K, N-1) - M(f|_{\partial K^+}, \partial K^+, N-1) \\ &\leq CC(K) \leq M(f, K, N) + M(f|_{\partial K^+}, \partial K^+, N-1) \end{aligned}$$

for $N \geq 2$, and

$$\begin{aligned} &M(f, K, N) - M(f, K, N-1) - M(f|_{\partial K^+}, \partial K^+, N-1) \\ &\leq CC(K) \leq M(f, K, N) + M(f|_{\partial K^+}, \partial K^+, N-1) \end{aligned}$$

for $N = 1$.

Remark 2.3.3. One can refer to Theorem 10.2 in [26] for a full version of Morse inequalities which contains N inequalities and one equation. We only provide a partial result here since its full version involves deep background in differential topology which is not essential to this thesis.

Theorem 2.3.2 is very general since it can be applied to any regular C^2-domain to which f is admissible relative. In this thesis, our interest is only focused on Morse-like inequalities of a connected component of an excursion set. More specifically, for the real-valued function $f(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^N$, the excursion set of f in the search region $S \subset \mathbb{R}^N$ above the threshold $u \in \mathbb{R}$ is defined by

$$A_u(f, S) := \{\boldsymbol{t} \in S : f(\boldsymbol{t}) > u\}.$$

Then the connected component D of the excursion set $A_u(f, S)$ is said to be interior if

$$D \cap \partial S = \emptyset.$$

However, from Definition 2.3.1, it is easy to see that any real-valued function f can never be admissible relative to D (if D is a regular C^2-domain) since $f|_{\partial D}$ is always constant and thus degenerate on ∂D. This means that we cannot directly apply Morse inequalities to such a connected component of the excursion set $A_u(f, S)$. Fortunately, [4] provides a modification of Morse inequalities for the excursion set $A_u(f, S)$ when the search region S is a regular stratified manifold (see Definition 9.2.2 in [4]). Since $D \cap \partial S = \emptyset$, the boundary condition of the search region S in this modification can also be omitted, and finally, we have a very concise result as follows.

Theorem 2.3.4. *(Corollary 9.3.5, [4]) Let f be a Morse function on an open set $G \subset \mathbb{R}^N$ $(N \geq 1)$, i.e. $f \in C^2(G)$ and every critical point of f in G is non-degenerate. Suppose that D is a bounded connected component of $A_u(f, G)$ for some $u \in \mathbb{R}$, such that $\partial D \cap \partial S = \emptyset$. Then*

$$M(f, D, N) - M(f, D, N - 1) \leq CC(D) \leq M(f, D, N).$$

Remark 2.3.5. Although Corollary 9.3.5 in [4] only corresponds to the last equation of Theorem 10.2 in [26], its proof also works for other N inequalities.

Chapter 3

Poisson Limits of the Connected Components and Critical Points

3.1 Introduction

When $N = 1$, there is rich literature for the functionals concerning the topological structures of excursion sets of a Gaussian process, such as the number of level crossing points, the number of local maxima, the height of maxima, etc. However, some of these functionals are hard to characterize for a general $N \geq 1$. This chapter is therefore dedicated to developing a new asymptotic theory for the extension of certain important functionals to high-dimensional parameter space under reasonable assumptions.

In this chapter, the functionals we are interested in are the numbers of the local maxima and connected components of the excursion set of a stationary Gaussian random field above a given threshold (see Figure 3.1). We define a family of point processes by the number of the local maxima in the excursion set when the threshold tends to infinity and the search region expands with a matching speed. Our first aim is to show that these point processes will converge weakly to a Poisson point process. While the techniques needed to rigorously prove this result are complicated, the basic intuition goes back to the well-known Poisson approximation to binomial distribution in which the relationship between the number and the probability of successes is selected to maintain a constant mean. In our context, considering that the increasing threshold will decrease the number of the local maxima above the threshold while the expanding region will work in the opposite direction, the matching speed is then selected to maintain a constant mean measure.

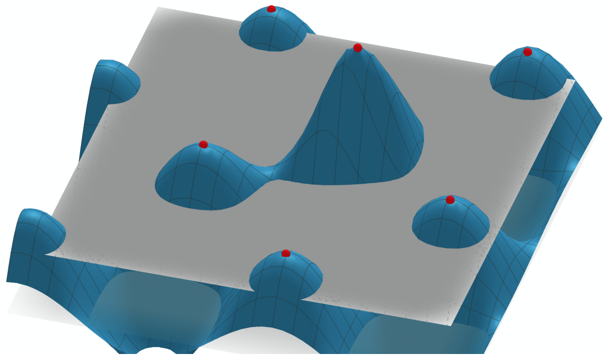


Figure 3.1: Connected components of an excursion set. The blue surface is a sample function of the random field, and the grey square is the search region. There are seven connected components and six local maxima in the excursion set.

To establish a Poisson limit, we define a "grid-block system" where the compatibility between the covariance structure of the underlying Gaussian random field, the matching speed of the expanding search region with the increasing threshold, and the existing results we use in our proof are fully considered. The system only needs quite mild compatibility conditions, which empower our theory to have a broad application. Although the underlying Gaussian random field is continuously parameterized, the grid-block system equips us with a powerful weapon to approximate continuity by discretization. This coincides with the spirit of the definition of the continuous-time Gaussian random field. Moreover, the asymptotic independence in the convergence to a Poisson limit can be fully captured by applying the normal comparison lemma ([23]) on the grids in blocks.

We also introduce a family of events with probability increasing to 1. By conditioning on these events and using a fundamental inequality, the avoidance functional of the local maxima in the excursion set can be related to the tail distribution of the maximum on an expanding region at an increasing value. This critical technique greatly simplifies our question and also plays an important role when we relate the number of local maxima with the number of connected components.

As mentioned above, our second aim is to show the Poisson limit for the number of connected components of the excursion set by relating it with the number of local maxima. Compared with local maxima, connected components cannot be characterized locally, and therefore resulting in the failure to construct a corresponding point process. Instead of constructing point processes directly by connected components and proving the weak convergence of point processes, we show the weak convergence of the number of connected components on expanding regions, i.e., the weak convergence of random variables. Then we only need to bound its distribution from both sides at the same threshold and prove that both sides will converge to the same value of a Poisson distribution. To achieve this objective, we construct another point process by the global maximum in each connected component and show that these point processes share the same Poisson limit with the point processes of the number of local maxima we have studied. In general, our theory actually provides a framework for the Poisson limit of functionals on the Gaussian excursion set which can be characterized either locally or nonlocally.

This work has a clear and immediate motivation from biostatistics. The choice of the functionals that we study is determined by a need to set the prior distribution in a Bayesian model ([19]). More precisely, in scalar-on-image regression models for brain imaging, researchers are interested in investigating the relationship between a scalar response, for example cognitive score, and brain images. In Bayesian literature, the conditional distribution of the scalar response $Y_i, 1 \leq i \leq n$, for n objects are modeled by

$$(Y_i | \boldsymbol{W}_i, \boldsymbol{X}_i, \boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma^2) \sim N\left(\sum_{k=1}^{q} \alpha_k \boldsymbol{W}_{i,k} + p_n^{-1/2} \sum_{j=1}^{p_n} \beta(\boldsymbol{s}_j) X_{ij}, \sigma^2\right),$$

where $N(\mu, \sigma)$ denotes a normal distribution with mean μ and covariance σ^2, $\boldsymbol{W}_i = (W_{i,1}, \dots, W_{i,k})$ is the vector of related scalar covariates, $\boldsymbol{X}_i = (X_{i,1}, \dots, X_{i,p_n})$ is the intensity value of the image measured at the locations $\boldsymbol{s}_1, \dots, \boldsymbol{s}_{p_n}$, and p_n is the number of voxels or pixels in the image. The choice of the prior distribution of regression coefficients $(\beta(\boldsymbol{s}_1), \dots, \beta(\boldsymbol{s}_{p_n}))$ is limited by the following observations

1. The locations $\boldsymbol{s}_1, \dots, \boldsymbol{s}_{p_n}$ are actually from a 2 or 3-dimensional lattice determined by the resolution of the image. Therefore β should distribute spatially.

2. The total effect of an image should not increase to infinity as the resolution increase. This results in the sparsity in regression coefficients, i.e., $\beta(\boldsymbol{s}_i) = 0$ for the majority of $1 \leq i \leq p_n$.

To capture the sparsity and spatiality of regression coefficients simultaneously, [19] suggested using a soft-thresholded Gaussian random field as the prior of β, where the support of β are modeled as the excursion set of the Gaussian random field indexed by $\mathbb{R}^2$ or $\mathbb{R}^3$. However we still need some rule for the selection of the threshold to match the prior degree of sparsity. Here the sparsity may be characterized by certain functionals of the activated regions of the brain, such as the number of the connected components and the number of the local maxima above the threshold, which are known from biology. Essentially, this chapter contributes to predict these random quantities in the Gaussian random field model for any given high threshold, hence helps to choose the prior distribution for the Gaussian random field in a Bayesian procedure.

The remainder of this chapter is organized as follows.

Section 3.2 mainly serves as the literature review of some existing results about Gaussian random fields. These results can be classified into two categories which correspond to the first two subsections. Section 3.2.1 starts from Corollary 3.2.1 which gives an integral expression for the expected number of the local maxima in a Gaussian excursion set. One simple asymptotic expression of this integral is shown in Lemma 3.2.2 which provides a theoretical basis for the expanding rate of the search region as we have discussed earlier. In Section 3.2.2, we introduce two other powerful asymptotic results of the global maxima of Gaussian random field in certain expanding systems. These asymptotic results are used to estimate the avoidance functional of the point process we construct. Finally, we will also introduce some useful results in Section 3.2.3 which are not about random fields.

In Section 3.3, we will construct a family of point processes for the local maxima of a qualified Gaussian random field on an expanding search region above an increasing threshold. The main result in this section, Theorem 3.3.2, shows the existence of a Poisson limit for this series of point processes when the expanding rate g_u for the search region matches the increasing threshold u. Our strategy is to approximate a continuously parameterized Gaussian random field and control its correlation structure by its values on the grid-block system we defined. Since the proof for Theorem 3.3.2 is very complicated, we also provide a clear outline of this proof in Section 3.3.2 for reader as a road map.

In Section 3.4, instead of directly working on a family of point processes, we study the number of connected components of a Gaussian excursion set with the expanding region $g_u S$ for any qualified $S \subset \mathbb{R}^N$ when the threshold u is high. In Section 3.4.1, we study the asymptotic behavior of the expected number of the connected components touching the boundary of the search region. In Section 3.4.2, the Poisson limit of the number of connected components of a Gaussian excursion set is derived from the relationship between the number of connected components and the number of local maxima, combined with the main results in the previous section. To further analyze the behavior of the number of connected components when the threshold is high, we first derive an asymptotic result about the expected number of the critical points in Theorem 3.4.5. Then in the last part of this section, a stronger relationship between the number of connected components and the number of local maxima will be clarified based on Theorem 3.4.5 and the Morse inequality of the excursion set we discussed in Section 2.3.

3.2 Literature Review

There is rich literature of functionals of Gaussian random fields indexed by $\mathbb{R}^N$. In this section, we mainly review some existing results about characterizations and asymptotic behavior of extremes of Gaussian random fields. We will also introduce some very useful results which are not about random fields in the last part of this section.

Throughout this section, endow $\mathbb{R}^N (N \geq 1)$ with the Euclidean norm $\|\cdot\|$, and denote by λ_N the N-dimensional Lebesgue measure. For a symmetric matrix $\boldsymbol{V} = (v_{ij})_{N\times N} \in \mathbb{R}^{N\times N}$, we say $\boldsymbol{W} = (w_1, \ldots, w_{N(N+1)/2}) \in \mathbb{R}^{N(N+1)/2}$ is a **usual vectorization** of $\boldsymbol{V}$ if

$$w_{i+j(j-1)/2} = v_{ij} \text{ for any integers } 1 \leq i \leq j \leq N.$$

Any vectorization of a symmetric matrix appearing in this section will follow the same way. We will not distinguish a matrix from its usual vectorization in notations, but one can easily distinguish them from a given context.

3.2.1 The Expected Number of the Local Maxima with Full Index above a Threshold

Let the random field $\{X(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^N\}$ be almost surely two times continuously differentiable, and let S be a compact subset of $\mathbb{R}^N$ with $\lambda_{N-1}(\partial S) < \infty$. Denote by $M_u(X, S)$ the number of the local maxima of a Gaussian random field X above the threshold u in the set S of the form:

$$\left\{\boldsymbol{t} \in S : X(\boldsymbol{t}) > u, \nabla X(\boldsymbol{t}) = \boldsymbol{0}, \text{index}\left(\nabla^2 X(\boldsymbol{t})\right) = N\right\},$$

i.e.,

$$M_u(X, S) := \#\left\{\boldsymbol{t} \in S : X(\boldsymbol{t}) > u, \nabla X(\boldsymbol{t}) = \boldsymbol{0}, \text{index}\left(\nabla^2 X(\boldsymbol{t})\right) = N\right\},$$

where the index of a matrix is defined as the number of its negative eigenvalue, and $\#$ stands for "the cardinality of". Note that given a particular realization of X, $M_u(X, S)$ is not necessarily equal to the true number of the local maxima of the random field $X(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^N$ above the threshold u in the search region S. This is because some local maximum $\boldsymbol{t}^*$ of the random field X can be degenerate, i.e.,

$$\det\left(\nabla^2 X(\boldsymbol{t}^*)\right) = 0.$$

We further assume that $\{X(\boldsymbol{t}), \boldsymbol{t} \in S\}$ satisfies all the conditions in Lemma 2.1.10, where $T = S$. Then by Remark 2.1.12, $M_u(X, S)$ is almost surely the number of the local maxima in the excursion set $A_u(X,S)$. Note that

$$M_u(X,S) = \#\{\boldsymbol{t} \in S : X(\boldsymbol{t}) > u, \nabla X(\boldsymbol{t}) = 0, \text{index}\left(\nabla^2 X(\boldsymbol{t})\right) = N\}.$$

In (2.12), if we let $\boldsymbol{v} = \boldsymbol{0}_N$, $T = S$, and

$$B = \{(\boldsymbol{x}'', x) \in \mathbb{R}^{N(N+1)/2+1} : \text{index}\,(\boldsymbol{x}'') = N, x > u\},$$

then we have the following important corollary.

Corollary 3.2.1. *(Lemma 11.7.1, [4]) Let X and S be as above. Then*

$$\mathbb{E}\left[M_u(X,S)\right] = \int_S \int_u^\infty \int_{D_N} \left|\det\left(\boldsymbol{x}''\right)\right| p_{\boldsymbol{t}}(\boldsymbol{x}'', \boldsymbol{0}, x) d\boldsymbol{x}'' dx d\boldsymbol{t},$$

where $p_{\boldsymbol{t}}(\boldsymbol{x}'', \boldsymbol{x}', x)$, $(\boldsymbol{x}'', \boldsymbol{x}', x) \in \mathbb{R}^{N(N+1)/2} \times \mathbb{R}^N \times \mathbb{R}$ is the density function of the Gaussian vector $(\nabla^2 X(\boldsymbol{t}), \nabla X(\boldsymbol{t}), X(\boldsymbol{t}))$ for any $\boldsymbol{t} \in S$, and

$$D_N := \{\boldsymbol{x}'' \in \mathbb{R}^{N(N+1)/2} : \text{index}\,(\boldsymbol{x}'') = N\}.$$

Corollary 3.2.1 provides an explicit integral expression for $\mathbb{E}[M_u(X,S)]$. However, it is still too complicated to calculate this integral due to the awkwardness of the domain D_N. A simplification of $\mathbb{E}\left[M_u(X,S)\right]$ will be very useful. In Corollary 3.2.1, if we further assume that X is stationary, then we have

$$\mathbb{E}\left[M_u(X,S)\right] = \lambda_N(S) \int_u^\infty \int_{D_N} \left|\det\left(\boldsymbol{x}''\right)\right| p_{\boldsymbol{0}}(\boldsymbol{x}'', \boldsymbol{0}, x) d\boldsymbol{x}'' dx. \tag{3.1}$$

We see that the remaining integral in (3.1) is actually independent of the choice of S. This integral can be further estimated by the following lemma.

Lemma 3.2.2. *(Theorem 6.3.1, [1]) Let X and S be as in Corollary 3.2.1. Suppose that X is also stationary with variance σ^2. Then*

$$\mathbb{E}\left[M_u(X,S)\right] = \frac{\lambda_N(S)\det\left(\boldsymbol{\Lambda}_X\right)^{1/2} u^{N-1}}{(2\pi)^{(N+1)/2}\sigma^{2N-1}} \exp\left(-\frac{u^2}{2\sigma^2}\right)\left(1 + O(u^{-1})\right),$$

where $\boldsymbol{\Lambda}_X$ is the covariance matrix of ∇X, and $O(u^{-1})$ is independent of the choice of S. More specifically, if we let

$$D_u(X,S) = \frac{\lambda_N(S)\det\left(\boldsymbol{\Lambda}_X\right)^{1/2} u^{N-1}}{(2\pi)^{(N+1)/2}\sigma^{2N-1}} \exp\left(-\frac{u^2}{2\sigma^2}\right), \tag{3.2}$$

then there exists a finite constant $C > 0$, which is independent of the choice of S, such that

$$\left|\frac{\mathbb{E}\left[M_u(X,S)\right]}{D_u(X,S)} - 1\right| \leq C u^{-1}.$$

3.2.2 Some Asymptotic Results of Maxima

In this section, we will introduce two asymptotic results about the maxima of a Gaussian random field over a compact search region expanding with an increasing threshold. Recall that in Lemma 2.2.7, $\mathcal{J}$ is defined as the set of bounded open N-dimensional cubes and their finite unions and intersections. Here we define

$$\overline{\mathcal{J}} := \left\{\overline{T} : T \in \mathcal{J}\right\},$$

where $\overline{T}$ is the closure of T. Therefore, any element, say J, of $\overline{\mathcal{J}}$ is a non-empty compact set whose boundary ∂J is piecewisely smooth (or more specifically, piecewisely flat) and $\lambda_{N-1}(\partial J) < \infty$. The first asymptotic result involves the definition of blowing-up systems as follows.

Definition 3.2.3. ([31]) A system $\{T_u, u \in \mathbb{R}\}$, $T_u \in \overline{\mathcal{J}}$, is said to blow up, if

1. $\lambda_N(T_u) \to \infty$ as $u \to \infty$;
2. there exist finite constants $L_1 > 0$ and $\delta_1 \geq 0$ such that for any $u \in \mathbb{R}$ and N-dimensional ball $B(\mathbf{0}_N, R)$ with radius $R > 1$ and centered at the origin,
$$\lambda_N(T_u \oplus B(\mathbf{0}_N, R)) - \lambda_N(T_u) \leq L_1 R^N (\lambda_{N-1}(\partial T_u))^{1+\delta_1}, \tag{3.3}$$
where the symbol $\oplus$ denotes a Minkowski summation operator,
$$A \oplus B = \{\boldsymbol{x} + \boldsymbol{y} : \boldsymbol{x} \in A, \boldsymbol{y} \in B\}. \tag{3.4}$$
3. there exist finite constants $L_2 > 0$ and $\alpha \in \mathbb{R}$ such that
$$\alpha(1 + \delta_1) < 1 \quad \text{and} \quad \lambda_{N-1}(\partial T_u) \leq L_2 \lambda_N(T_u)^{\alpha}. \tag{3.5}$$

The following lemma also serves as an example for blowing-up systems.

Lemma 3.2.4. *([31]) For any $K \in \mathcal{J}$ and $\alpha_u \to \infty$ as $u \to \infty$, $\{\alpha_u K, u \in \mathbb{R}\}$ is a blowing-up system with $\delta_1 = 0$ and $\alpha = (N-1)/N$.*

Lemma 3.2.5. *(Theorems 14.1 & 14.2, [31]) Suppose that $\{T_u, u > 0\}$ is a blowing-up system. Let $\{X(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^N\}$ be a centered, stationary, three times differentiable in the mean square sense, Gaussian random field with covariance function $r(\boldsymbol{t}) := \mathbb{E}[X(\mathbf{0})X(\boldsymbol{t})]$ and constant variance σ^2. Suppose that there exists a finite constant $\alpha > 0$ such that*

$$\int_{\mathbb{R}^N} |r(\boldsymbol{t})|^{\alpha} d\boldsymbol{t} < \infty.$$

Then for any $x \in \mathbb{R}$,

$$P\left[\max_{\boldsymbol{t}\in T_u}\left(\sigma^{-1}X(\boldsymbol{t})-\ell_{X,T_u}\right)\ell_{X,T_u}<x\right]=\exp\left\{-\exp\{-x\}\right\},$$

where ℓ_{X,T_u} is the largest solution to the equation

$$\frac{\lambda_N(T_u)\det\left(\boldsymbol{\Lambda}_X\right)^{1/2}}{(2\pi)^{(N-1)/2}\sigma^{N/2}}\ell^{N-1}\exp\left\{-\ell^2/2\right\}=1, \tag{3.6}$$

and $\boldsymbol{\Lambda}_X$ is the covariance matrix of ∇X.

It is worth mentioning that $D_u(X,S)$ in (3.2) and the left part of (3.6) are very similar. We will make the most of this point in the following section.

The second asymptotic result involves the definition of slowly blowing-up systems as follows.

Definition 3.2.6. The system $\{T_u, u\in\mathbb{R}\}$, where $T_u\subset\mathbb{R}^N$ is closed with $\lambda_N(\partial T_u)=0$ for any $u\in\mathbb{R}$, is said to blow up slowly with rate $\kappa>0$ if each T_u contains an N-dimensional unit cube and

$$\lambda_N(T_u)=O\left(\exp\left\{\kappa u^2/2\right\}\right)\quad\text{as } u\to\infty.$$

Remark 3.2.7. Note that the rate in Definition 3.2.6 is not unique. If the system $\{T_u, u\in\mathbb{R}\}$ blows up slowly with rate $\kappa_0>0$, then it also blows up slowly with rate κ for any $\kappa\ge\kappa_0$.

Lemma 3.2.8. *(Theorem 7.2, [31]) Let $\{X(\boldsymbol{t}), \boldsymbol{t}\in\mathbb{R}^N\}$ be a centered, stationary Gaussian random field with covariance function $r(\boldsymbol{t})$. Suppose that the following conditions hold:*

1. *there exists a non-degenerate matrix $\boldsymbol{Q}_X$ and a finite constant $\beta>0$ such that*

$$r(\boldsymbol{Q}_X\boldsymbol{t})=1-\|\boldsymbol{t}\|^\beta+o(\|\boldsymbol{t}\|^\beta)\quad\text{as } \|\boldsymbol{t}\|\to 0, \tag{3.7}$$

2. *$r(\boldsymbol{t})\to 0$ as $\|\boldsymbol{t}\|\to\infty$.*

Then there exists a finite constant $\kappa_X>0$ such that for any slowly blowing-up system $\{T_u, u\in\mathbb{R}\}$ with rate κ_X,

$$P\left[\max_{\boldsymbol{t}\in T_u}X(\boldsymbol{t})>u\right]=H_\beta\lambda_N(T_u)\det(\boldsymbol{Q}_X^{-1})u^N\Psi(u)(1+o(1))\quad\text{as } u\to\infty,$$

where H_β is a Pickands' constant (see [30] and [31]). Especially, $H_2=\pi^{-N/2}$.

When $\beta = 2$, the relationship between $\boldsymbol{\Lambda}_X$ in Lemma 3.2.5 and $\boldsymbol{Q}_X$ in Lemma 3.2.8 can be derived as follows. Note that by a Taylor's expansion at $\mathbf{0}$ and the definition of the mean square derivatives,

$$r(\boldsymbol{t}) = 1 - \frac{1}{2}\boldsymbol{t}^T\boldsymbol{\Lambda}_X\boldsymbol{t} + o(\|\boldsymbol{t}\|^2),$$

while by (3.7),

$$r(\boldsymbol{Q}_X\boldsymbol{t}) = 1 - \|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2).$$

Thus, we can take $\boldsymbol{Q}_X$ such that

$$\frac{1}{2}\boldsymbol{Q}_X^T\boldsymbol{\Lambda}_X\boldsymbol{Q}_X = \boldsymbol{I}_N,$$

where $\boldsymbol{I}_N$ is the identity matrix of size N. Therefore,

$$\det(\boldsymbol{Q}_X)^{-1} = 2^{-N/2}\det(\boldsymbol{\Lambda}_X)^{1/2}. \tag{3.8}$$

Example 3.2.1. *Let $\{X(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^N\}$ be a centered, stationary Gaussian random field with covariance function $r(\boldsymbol{t}) = e^{-\|\boldsymbol{t}\|^2}$. Since $r(\boldsymbol{t})$ has any order partial derivatives, by Lemma 2.1.4, X is three times differentiable in the mean square sense on $\mathbb{R}^N$. We also have*

$$\int_{\mathbb{R}^N} |r(\boldsymbol{t})|\, d\boldsymbol{t} = \int_{\mathbb{R}^N} e^{-\|\boldsymbol{t}\|^2} d\boldsymbol{t} = \left(\int_{\mathbb{R}} e^{-t_1^2} dt_1\right)^N = \pi^{N/2} < \infty.$$

Therefore, X satisfies all the conditions in Lemma 3.2.5. By Lemmas 2.1.3 and 2.1.4, X_j is also a centered, stationary Gaussian random field with covariance function

$$r_{X_j}(\boldsymbol{t}) = -r_{jj}(\boldsymbol{t}) = 2(1 - 2t_j^2)e^{-\|\boldsymbol{t}\|^2}.$$

Denote by $\sigma_{X_j}^2$ the variance of X_j. Then

$$\sigma_{X_j}^2 := -r_{jj}(\mathbf{0}) = 2.$$

Let $r_{X_j/\sigma_{X_j}}(\boldsymbol{t})$ be the covariance function of X_j/σ_{X_j}. Then

$$\begin{aligned}
\int_{\mathbb{R}^N} \left|r_{X_j/\sigma_{X_j}}(\boldsymbol{t})\right| d\boldsymbol{t} &= \int_{\mathbb{R}^N} \left|1 - 2t_j^2\right| e^{-\|\boldsymbol{t}\|^2} d\boldsymbol{t} \\
&= \left(\int_{\mathbb{R}} e^{-t_1^2} dt_1\right)^{N-1} \int_{\mathbb{R}} \left|1 - 2t_j^2\right| e^{-t_j^2} dt_j \\
&\le \pi^{(N-1)/2} \int_{\mathbb{R}} (1 + 2t_j^2) e^{-t_j^2} dt_j \\
&= 2\pi^{N/2} \\
&< \infty.
\end{aligned}$$

Therefore, X_j/σ_{X_j} also satisfies all the conditions in Lemma 3.2.5. Moreover, by a Taylor's expansion of the function e^{-x^2}, $x \in \mathbb{R}$ at $x = 0$, we have for any $\boldsymbol{t} \in \mathbb{R}^{N\times 1}$,

$$r(\boldsymbol{t}) = e^{-\|\boldsymbol{t}\|^2} = 1 - \|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)$$

and for any $1 \le j \le N$,

$$r_{X_j/\sigma_{X_j}}(\boldsymbol{Q}_1\boldsymbol{t}) = -\sigma_{X_j}^{-2} r_{jj}(\boldsymbol{Q}_1\boldsymbol{t}) = e^{-\|\boldsymbol{Q}_1\boldsymbol{t}\|^2}\left(1 - 2t_j^2/3\right) = 1 - \|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2),$$

where $\boldsymbol{Q}_1 \in \mathbb{R}^{N\times N}$ is a diagonal matrix with all diagonal elements equal to one except for a $1/\sqrt{3}$ at the j-th column. Finally, it is easy to see that

$$r(\boldsymbol{t}) \to 0 \quad \text{and} \quad r_{X_j/\sigma_{X_j}}(\boldsymbol{t}) \to 0$$

as $\|\boldsymbol{t}\| \to \infty$. Therefore, both X and X_j/σ_{X_j} satisfy all the conditions in Lemma 3.2.8.

3.2.3 Some Useful Lemmas

In this section, we will introduce three existing results which will play important roles in the proofs of the main results in this chapter. The first lemma gives a bound for the difference between the joint distribution functions of two sets of Gaussian random variables by their covariances.

Lemma 3.2.9. *(Theorem 4.2.1, [23]) Let $Y_1, \ldots, Y_n$ and $Z_1, \ldots, Z_n$ be standard Gaussian random variables with covariance matrices $\boldsymbol{\Lambda} = (\lambda_{ij})$ and $\Gamma = (\gamma_{ij})$, respectively. Denote $\rho_{ij} := \max(|\lambda_{ij}|, |\gamma_{ij}|)$ for any $1 \le i, j \le n$. Then*

$$\begin{aligned} &P\left[Y_1 \le t_1, \ldots, Y_n \le t_n\right] - P\left[Z_1 \le t_1, \ldots, Z_n \le t_n\right] \\ &\le \frac{1}{2\pi} \sum_{1\le i<j\le n} (\lambda_{ij} - \gamma_{ij})^+ \left(1 - \rho_{ij}^2\right)^{-1/2} \exp\left(-\frac{\frac{1}{2}\left(t_i^2 + t_j^2\right)}{1 + \rho_{ij}}\right), \end{aligned}$$

where $(x)^+ := \max(x, 0)$ for any $x \in \mathbb{R}$.

Let $\boldsymbol{\Lambda} \in \mathbb{R}^{N\times N}$ be a diagonal matrix with positive diagonal elements. Define

$$D_{k,x,\boldsymbol{\Lambda}} := \left\{\boldsymbol{v} \in \mathbb{R}^{N\times N} : \text{index}\left(\boldsymbol{v} - x\boldsymbol{\Lambda}\right) = k\right\}.$$

The second lemma can deal with the asymptotic behavior of $D_{k,x,\boldsymbol{\Lambda}}$ as $x \to \infty$.

Lemma 3.2.10. *(Lemma 6.3.1, [1]) Suppose that*

- $\boldsymbol{V}$ *is an* $N \times N$ *symmetric matrix;*
- $\boldsymbol{W}$ *is the usual vectorization of* $\boldsymbol{V}$ *with length* $L := N(N+1)/2$*;*
- $\boldsymbol{\Lambda}$ *is a diagonal matrix with positive diagonal elements;*
- $Y_u \in \mathbb{R}^L$ *is defined by*

$$\boldsymbol{W} \in Y_u \text{ is equivalent to } \boldsymbol{V} - u\boldsymbol{\Lambda} \text{ is negative definite.}$$

Then there exists a finite constant C *only determined by* $\boldsymbol{\Lambda}$*, such that for each pair* (r, u) *satisfying* $u > Cr$*,*

$$B_L(\mathbf{0}, r) \subset Y_u,$$

where $B_L(\mathbf{0}, r) = \{\boldsymbol{t} \in \mathbb{R}^L : \|\boldsymbol{t}\| \leq r\}$*.*

3.3 The Number of Local Maxima above a Threshold

In Section 3.2, we have defined

$$M_u(X, S) = \# \left\{ \boldsymbol{t} \in S : X(\boldsymbol{t}) > u, \nabla X(\boldsymbol{t}) = \mathbf{0}, \text{index}\left(\nabla^2 X(\boldsymbol{t})\right) = N \right\},$$

where the index of a matrix is defined as the number of its negative eigenvalues. By Remark 2.1.12 and Corollary 3.2.1, we do not need to distinguish $M_u(X, S)$ from the corresponding true number of the local maxima as long as the conditions in Corollary 3.2.1 are all satisfied.

With the help of the above definition, we will construct a series of point processes by expanding the search region for an increasing threshold. Then we will show the existence of a Poisson limit for this series of point processes when the rate of the expansion for the search region matches the rate of the increase of the threshold in a specific way. Throughout this section, for $N \geq 1$, we endow $\mathbb{R}^N$ with the Euclidean norm $\|\cdot\|$. Denote by λ_N the N-dimensional Lebesgue measure and by $\mathcal{B}(\mathbb{R}^N)$ the Borel σ-field of $(\mathbb{R}^N, \|\cdot\|)$. Denote by $\oplus$ the Minkowski summation operator (see (3.4)). For any $a > 0$ and $A \subset \mathbb{R}^N$, we define

$$aA := \{a\boldsymbol{t} : \boldsymbol{t} \in A\},$$

and let $\overset{\circ}{A}$ be the interior of A. For any $A, B \subset \mathbb{R}^N$, define $A - B := A \cap B^c$, i.e.,

$$A - B := \{a \in A : a \notin B\}.$$

Moreover, as in Section 3.2.2, let $\mathcal{J}$ be the set of N-dimensional cubes and their finite unions and intersections and define

$$\overline{\mathcal{J}} := \{\overline{T} : T \in \mathcal{J}\},$$

where $\overline{T}$ denotes the closure of T. Note that $\lambda_{N-1}(\partial K) < \infty$ for any $K \in \mathcal{J}$.

3.3.1 Main Results

Let $\{X(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^N\}$ $(N \geq 1)$ be a centered, stationary Gaussian random field with covariance function $r(\boldsymbol{t})$. Suppose that the following conditions hold:

1. the distribution of $(\nabla^2 X, \nabla X, X)$, written as $p(\boldsymbol{x}'', \boldsymbol{x}', x)$, is non-degenerate;
2. X is almost surely three times continuously differentiable;
3. X is four times differentiable in the mean square sense;
4. there exist finite constants $\alpha_1, \alpha_2 > 0$ such that
$$\int_{\mathbb{R}^N} |r(\boldsymbol{t})|^{\alpha_1} d\boldsymbol{t} < \infty,$$
and for any $1 \leq i \leq N$,
$$\int_{\mathbb{R}^N} \left| \frac{\partial^2 r(\boldsymbol{t})}{\partial t_i^2} \right|^{\alpha_2} d\boldsymbol{t} < \infty;$$
5. X satisfies the conditions of Lemma 3.2.8, i.e., there exists a non-degenerate matrix $\boldsymbol{Q}_X$ such that
$$r(\boldsymbol{Q}_X \boldsymbol{t}) = 1 - \|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2) \ \text{ as } \|\boldsymbol{t}\| \to 0,$$
and
$$r(\boldsymbol{t}) \to 0 \ \text{ as } \|\boldsymbol{t}\| \to \infty;$$
6. denote by $\sigma^2_{X_j} (\sigma_{X_j} > 0)$ the variance of X_j, $1 \leq j \leq N$, and then we assume that X_j/σ_{X_j} also satisfies the conditions of Lemma 3.2.8, i.e., there exists a non-degenerate matrix $\boldsymbol{Q}_{X_j/\sigma_{X_j}}$ such that
$$r_{X_j/\sigma_{X_j}} \left(\boldsymbol{Q}_{X_j/\sigma_{X_j}} \boldsymbol{t} \right) = 1 - \|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2) \ \text{ as } \|\boldsymbol{t}\| \to 0,$$
and
$$r_{X_j/\sigma_{X_j}}(\boldsymbol{t}) \to 0 \ \text{ as } \|\boldsymbol{t}\| \to \infty,$$
where $r_{X_j/\sigma_{X_j}}(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^N$ is the covariance function of X_j/σ_{X_j};

7. since $r(\boldsymbol{t}) \to 0$ as $\|\boldsymbol{t}\| \to \infty$, we can properly define

$$\overline{r}(x) := \max_{\|\boldsymbol{t}\| \geq x} |r(\boldsymbol{t})| \tag{3.9}$$

for any $x \geq 0$. Assume that there exists some finite constant $\beta > 0$ such that

$$\gamma(\beta, u) := u^{2(N+1)}\overline{r}\left(e^{\beta u^2}\right) \to 0 \quad \text{as } u \to \infty. \tag{3.10}$$

Example 3.3.1. *Let $\{X(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^N\}$ be a centered, stationary Gaussian random field with covariance function $r(\boldsymbol{t}) = e^{-\|\boldsymbol{t}\|^2}$. Then it is easy to see that Condition (3.10) holds. Moreover, from Examples 3.2.1, A.2.2 and A.2.3, we see that X satisfies all the above conditions.*

Since $\{X(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^N\}$ is stationary and almost surely three times continuously differentiable, all of the sixth-order partial derivatives of $r(\boldsymbol{t})$ exist at $\boldsymbol{t} = \mathbf{0}$. Then by Lemma 2.1.13, we are allowed to apply Corollary 3.2.1 to the Gaussian random field X on any compact set $S \subset \mathbb{R}^N$ with $\lambda_{N-1}(\partial S) < \infty$. From Section 3.2, we see that the above conditions also allow us to apply Lemmas 3.2.5 and 3.2.8 to the Gaussian random field X, and Lemmas 3.2.5 and 3.2.8 to Gaussian random fields X_j/σ_{X_j}, $1 \leq j \leq N$. In this section, we first fix a mean constant $c > 0$. Then by Corollary 3.2.1 and (3.1), we can choose a function, written as g_u, of u such that

$$\mathbb{E}\left[M_u(X, g_u K)\right] = c\lambda_N(K) \tag{3.11}$$

for all $u \in \mathbb{R}$ and any bounded subset $K \in \mathcal{B}(\mathbb{R}^N)$ with $\lambda_{N-1}(\partial K) < \infty$. Here we actually define a family of point processes $\{M_u(X, g_u\cdot), u \in \mathbb{R}\}$ on $\mathbb{R}^N$. For convenience, we say the pair (X, g_u) (or the triple (X, g_u, c) if we need to emphasis the mean constant c in (3.11)) is qualified if X and g_u satisfy all of the above conditions. By Lemma 3.2.2, we do not need to distinguish $M_u(X, g_u K)$ and $M_u(X, g_u \overline{K})$ for $K \in \mathcal{J}$ since

$$\mathbb{E}[M_u(X, g_u \partial K)] = 0.$$

The following are the main results in this section.

Theorem 3.3.1. *Let (X, g_u, c) be a qualified triple. Then*

$$\lim_{u \to \infty} P\left[M_u\left(X, g_u K\right) = 0\right] = \exp\left\{-c\lambda_N(K)\right\}$$

for any $K \in \mathcal{J}$.

Theorem 3.3.2. *Let (X, g_u, c) be a qualified triple. Then $M_u(X, g_u\cdot)$ converges weakly to a Poisson point process with mean measure $c\lambda_N$ as $u \to \infty$.*

Proof. This is an immediate consequence of Theorem 3.3.1, Lemma 2.2.7 and (3.11). □

In the rest of this section, we will provide a poof for Theorem 3.3.1. To this end, we have to first define a grid-block system as follows.

Definition 3.3.3. Let the pair (X, g_u) be qualified. $\{(G_u, B_u) \in 2^{\mathbb{R}^N} \times 2^{\mathbb{R}^N}, u \in \mathbb{R}\}$ is said to be a grid-block system adapted to (X, g_u) if there exist functions b_u, d_u and f_u of $u \in \mathbb{R}$ such that

1. $G_u = d_u \mathbb{Z}^N$.
2. $B_{u,\boldsymbol{k}} := b_u^{-1}\left(f_u[-0.5, 0.5]^N \oplus \{\boldsymbol{k}\}\right)$ for any $\boldsymbol{k} \in \mathbb{Z}^N$, and $B_u := \bigcup_{\boldsymbol{k}} B_{u,\boldsymbol{k}}$.
3. As $u \to \infty$, we have
   (a) $u^{-2}d_u^{-1} \to \infty$;
   (b) $b_u \to \infty$;
   (c) $0 < f_u \uparrow 1$;
   (d) $g_u b_u^{-1}(1 - f_u) \to \infty$;
   (e) $g_u b_u^{-1} = O(e^{\kappa u^2/2})$, for
   $$\kappa = \frac{1}{2N} \min\left\{1, \kappa_X, \kappa_{X_j}, 1 \leq j \leq N\right\}, \tag{3.12}$$
   where κ_X and κ_{X_j} are only determined by X (see Lemma 3.2.8);
   (f)
   $$g_u^{2N} d_u^{-2N} \overline{r}(2g_u b_u^{-1}(1 - f_u)) \exp\left\{-\frac{u^2}{1 + \overline{r}\left(2g_u b_u^{-1}(1 - f_u)\right)}\right\} \to 0,$$
   where $\overline{r}$ is defined in (3.9).
4. For any sufficiently large u,
   $$g_u b_u^{-1} \mathbb{Z}^N \subset d_u \mathbb{Z}^N.$$

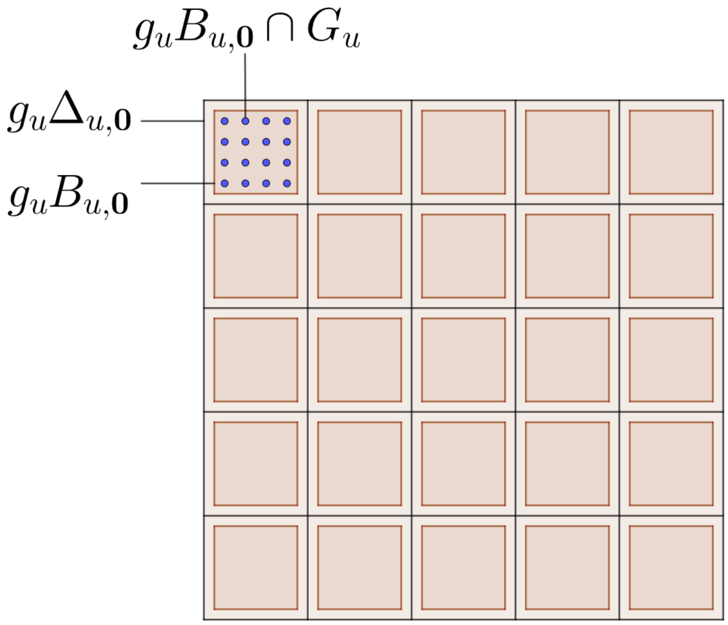


Figure 3.2: A part of the grid-block system $\{(G_u, B_u), u \in \mathbb{R}\}$ with the expanding rate g_u. For the blocks (squares) in the top-left corner of this part, the bigger one is $g_u\Delta_{u,\mathbf{0}}$, the smaller one is $g_u B_{u,\mathbf{0}}$, and $(g_u B_{u,\mathbf{0}}) \cap G_u$ consists of all the marked points in the figure.

Remark 3.3.4. Let $\varepsilon_u = d_u^{1/2}$. Then from Condition 3(a), we have

$$u^{-1} d_u^{-1} \varepsilon_u \to \infty \quad \text{and} \quad \varepsilon_u u \to 0$$

as $u \to \infty$. This property of ε_u will be very useful in the following sections. Define

$$\Delta_{u,\boldsymbol{k}} := b_u^{-1}([-0.5, 0.5]^N \oplus \{\boldsymbol{k}\}). \tag{3.13}$$

For example, Figure 3.3.1 shows the relationship between $g_u\Delta_{u,\mathbf{0}}$, $g_u B_{u,\mathbf{0}}$ and $(g_u B_{u,\mathbf{0}}) \cap G_u$. Conditions 3(b)-3(d) actually imply that as $u \to \infty$, $B_{u,\mathbf{0}}$ will be dominating in $\Delta_{u,\mathbf{0}}$, $g_u B_{u,\mathbf{0}}$ will be dominating in $\mathbb{R}^N$, and the shortest distance between expanded blocks $g_u B_{u,\boldsymbol{k}}$, $\boldsymbol{k} \in \mathbb{R}^N$ will increase to infinity. Condition 4 ensures that the center of each expanded block $g_u B_{u,\boldsymbol{k}}$, $\boldsymbol{k} \in \mathbb{Z}^N$ is a grid point in G_u, which implies that

$$(g_u B_{u,\boldsymbol{k}}) \cap G_u \neq \emptyset,$$

and the joint distribution of $\{X(\boldsymbol{t}), \boldsymbol{t} \in (g_u B_{u,\boldsymbol{k}}) \cap G_u\}$ is the same as the joint distribution of $\{X(\boldsymbol{t}), \boldsymbol{t} \in (g_u B_{u,\boldsymbol{0}}) \cap G_u\}$ for any $\boldsymbol{k} \in \mathbb{Z}^N$.

The following lemma shows the existence of the grid-block system for every qualified pair (X, g_u).

Lemma 3.3.5. *Every qualified pair (X, g_u) possesses an adapted grid-block system.*

Proof. See Appendix B.1.1. □

Immediately, we can construct a blowing-up system (see Definition 3.2.3) by a grid-block system. For $K \subset \mathbb{R}^N$, let

$$B_u(K) = \cup_{\Delta_{u,\boldsymbol{k}} \subset K, \boldsymbol{k} \in \mathbb{Z}^N} B_{u,\boldsymbol{k}}. \tag{3.14}$$

Then we have the following lemma.

Lemma 3.3.6. *Suppose that (X, g_u) is a qualified pair with the adapted grid-block system $\{(G_u, B_u), u \in \mathbb{R}\}$. Then for any non-empty $K \subset \mathcal{J}$, $g_u B_u(K)$ is a blowing-up system.*

Proof. From (3.14), we see that $B_u(K)$ is the union of finite number of disjoint identical sets $B_{u,\boldsymbol{k}}$, $\boldsymbol{k} \in \mathbb{Z}^N$. Therefore, it suffices to show that $g_u B_{u,\boldsymbol{0}}$ is a blowing-up systems. Note that

$$g_u B_{u,\boldsymbol{0}} = g_u b_u^{-1} f_u [-0.5, 0.5]^N.$$

By Definition 3.3.3, we have

$$g_u b_u^{-1} f_u \to \infty$$

as $u \to \infty$. It is easy to see that $[-0.5, 0.5]^N \in \mathcal{J}$. Then the proof is completed by directly applying Lemma 3.2.4 on $[-0.5, 0.5]^N$. □

In preparation for the proof of Theorem 3.3.1, there is the following lemma. This important lemma shows that a qualified search region can be approximated by a family of increasing sets in $\mathcal{J}$ from the inside of the search region and a family of decreasing sets in $\mathcal{J}$ from the outside of the search region.

Lemma 3.3.7. *For any bounded set $S \subset \mathbb{R}^N$ ($N \geq 1$) with $\lambda_{N-1}(\partial S) < \infty$, we can always find $J_{1,u}(S), J_{2,u}(S) \subset \mathcal{J}$ for any $u \in \mathbb{R}$ large enough, such that*

1. *$J_{1,u}(S) \subset \overset{\circ}{S} \subset S \subset J_{2,u}(S)$;*

2. $\lambda_N(J_{2,u}(S) - J_{1,u}(S)) \to 0$ *as* $u \to \infty$*;*

3. $\{g_u\overline{J_{2,u}(S) - J_{1,u}(S)}, u \in \mathbb{R}\}$ *is a blowing-up system.*

Proof. See Appendix B.1.2. ☐

3.3.2 The Outline for the Proof of Theorem 3.3.1

To make the proof for Theorem 3.3.1 easy to understand, we will first provide an outline of the proof. For space reasons, the proofs for the new lemmas appearing in the outline and the formal proof of Theorem 3.3.1 are all relegated to Appendix B.1.3. Given $K \in \mathcal{J}$, we use the following notations for convenience:

1. $H_u(K) := \{\max_{\boldsymbol{t}\in g_u\overline{K-J_{1,u}(K)}} X(\boldsymbol{t}) \leq u\}$, where $J_{1,u}(K)$ is as defined in Lemma 3.3.7.

2. Write $f_u \overset{u}{\approx} g_u$ if $|f_u - g_u| \to 0$ as $u \to \infty$.

3. Let $n_u(K)$ be the number of $\Delta_{u,\boldsymbol{k}}$ defined in (3.13) fully contained in K. Note that $\{\Delta_{u,\boldsymbol{k}}, \boldsymbol{k} \in \mathbb{R}^N\}$ forms a cover of $\mathbb{R}^N$ for each $u \in \mathbb{R}$, $\lambda_N(\Delta_{u,\boldsymbol{k}_1} \cap \Delta_{u,\boldsymbol{k}_2}) = 0$ for any $\boldsymbol{k}_1, \boldsymbol{k}_2 \in \mathbb{Z}^N$, and $\lambda_N(\Delta_{u,\boldsymbol{k}}) = b_u^{-N} \to 0$ as $u \to \infty$ for any $\boldsymbol{k} \in \mathbb{Z}^N$. Therefore, it is easy to see that
$$b_u^{-N} n_u(K) \overset{u}{\approx} \lambda_N(K). \tag{3.15}$$

Consider the following fundamental inequality. For any two events A and B with $P[A] > 0$, we have
$$\begin{aligned}
|P[B] - P[B|A]| &= \left|\frac{P[B]P[A] - P[BA]}{P[A]}\right| \\
&= \left|\frac{(P[BA] + P[BA^c])P[A] - P[BA]}{P[A]}\right| \\
&= \left|\frac{P[BA^c]P[A] - P[BA]P[A^c]}{P[A]}\right| \\
&= \left|P[BA^c] - \frac{P[BA]P[A^c]}{P[A]}\right| \\
&\leq P[A^c].
\end{aligned} \tag{3.16}$$

The outline for the proof of Theorem 3.3.1 is as follows.

1. Prove
$$P[H_u(K)] \to 1 \quad \text{as } u \to \infty. \tag{3.17}$$

2. Show
$$P\left[M_u(X, g_u K) = 0\right] \overset{u}{\approx} P\left[M_u(X, g_u B_u(K)) = 0\right], \tag{3.18}$$
where $B_u(K)$ is defined in (3.14).

3. By (3.17) and Inequality (3.16), we have
$$\begin{aligned} P\left[M_u(X, g_u B_u(K)) = 0\right] &\overset{u}{\approx} P\left[M_u(X, g_u B_u(K)) = 0 \;\middle|\; H_u(K)\right] \\ &= P\left[\max_{\boldsymbol{t}\in g_u B_u(K)} X(\boldsymbol{t}) \le u \;\middle|\; H_u(K)\right]. \end{aligned} \tag{3.19}$$

4. Again, by (3.17) and Inequality (3.16),
$$P\left[\max_{\boldsymbol{t}\in g_u B_u(K)} X(\boldsymbol{t}) \le u \;\middle|\; H_u(K)\right] \overset{u}{\approx} P\left[\max_{\boldsymbol{t}\in g_u B_u(K)} X(\boldsymbol{t}) \le u\right]. \tag{3.20}$$

5. Show
$$P\left[\max_{\boldsymbol{t}\in g_u B_u(K)} X(\boldsymbol{t}) \le u\right] \overset{u}{\approx} P\left[\max_{\boldsymbol{t}\in (g_u B_u(K))\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right]. \tag{3.21}$$

6. Since $B_u(K) = \cup_{\Delta_{u,\boldsymbol{k}}\subset K, \boldsymbol{k}\in\mathbb{Z}^N} B_{u,\boldsymbol{k}}$, we have
$$\begin{aligned} &P\left[\max_{\boldsymbol{t}\in (g_u B_u(K))\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right] \\ &= P\left[\bigcap_{\Delta_{u,\boldsymbol{k}}\subset K, \boldsymbol{k}\in\mathbb{Z}^N} \left\{\max_{\boldsymbol{t}\in (g_u B_{u,\boldsymbol{k}})\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right\}\right]. \end{aligned} \tag{3.22}$$

7. Show
$$\begin{aligned} &P\left[\bigcap_{\Delta_{u,\boldsymbol{k}}\subset K, \boldsymbol{k}\in\mathbb{Z}^N} \left\{\max_{\boldsymbol{t}\in (g_u B_{u,\boldsymbol{k}})\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right\}\right] \\ &\overset{u}{\approx} \prod_{\Delta_{u,\boldsymbol{k}}\subset K, \boldsymbol{k}\in\mathbb{Z}^N} P\left[\max_{\boldsymbol{t}\in (g_u B_{u,\boldsymbol{k}})\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right]. \end{aligned} \tag{3.23}$$

8. By the stationarity of X and Condition 4 in Definition 3.3.3, we have

$$
\begin{aligned}
&\prod_{\Delta_{u,\boldsymbol{k}}\subset K, \boldsymbol{k}\in\mathbb{Z}^N} P\left[\max_{\boldsymbol{t}\in(g_u B_{u,\boldsymbol{k}_i})\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right] \\
&= \left(P\left[\max_{\boldsymbol{t}\in(g_u B_{u,\mathbf{0}})\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right]\right)^{n_u(K)} \\
&= \left(1 - \frac{n_u(K) P\left[\max_{\boldsymbol{t}\in(g_u B_{u,\mathbf{0}})\cap G_u} X(\boldsymbol{t}) > u - d_u^{1/2}\right]}{n_u(K)}\right)^{n_u(K)}.
\end{aligned}
\tag{3.24}
$$

9. Show

$$
\begin{aligned}
b_u^N P\left[\max_{\boldsymbol{t}\in(g_u B_{u,\mathbf{0}})\cap G_u} X(\boldsymbol{t}) > u - d_u^{1/2}\right] &\overset{u}{\approx} b_u^N P\left[\max_{\boldsymbol{t}\in g_u B_{u,\mathbf{0}}} X(\boldsymbol{t}) > u\right] \\
&\to c \quad \text{as } u \to \infty.
\end{aligned}
\tag{3.25}
$$

10. By (3.15) and (3.25), we have

$$
\begin{aligned}
&n_u(K) P\left[\max_{\boldsymbol{t}\in(g_u B_{u,\mathbf{0}})\cap G_u} X(\boldsymbol{t}) > u - d_u^{1/2}\right] \\
&= b_u^{-N} n_u(K) b_u^N P\left[\max_{\boldsymbol{t}\in(g_u B_{u,\mathbf{0}})\cap G_u} X(\boldsymbol{t}) > u - d_u^{1/2}\right] \\
&\to c\lambda_N(K) \quad \text{as } u \to \infty.
\end{aligned}
\tag{3.26}
$$

11. Finally, combining (3.18)-(3.24) and (3.26) implies

$$
P\left[M_u(X, g_u K) = 0\right] \to \exp\left\{-c\lambda_N(K)\right\}.
$$

3.4 The Number of Connected Components of the Excursion Set

Let the pair (X, g_u) be qualified. For any given search region $S \subset \mathbb{R}^N$ ($N \ge 1$) and threshold $u \in \mathbb{R}$, we can define the excursion set $A_u(X, S) = \{\boldsymbol{t} \in S : X(\boldsymbol{t}) > u\}$. In the last section, we have established the Poisson limit of the series of point processes

$M_u(X, g_u\cdot)$. This actually implies that for any give $S \in \mathcal{B}(\mathbb{R}^N)$, $M_u(X, g_uS)$ follows a Poisson distribution as $u \to \infty$, i.e., the number of the local maxima of X with index N in the excursion set $A_u(X, g_uS)$ follows a Poisson distribution as $u \to \infty$.

By Lemma 2.2.1, any point process ξ on $\mathbb{R}^N$ corresponds to a preset $\{\boldsymbol{\tau}_j, 1 \le j \le \nu\}$ which is independent of the choice of the set to be measured by ξ. Since the number of the connected components can not be described locally, it does not correspond to such a set in $\mathbb{R}^N$, and therefore, we cannot directly construct point processes from it. Instead of directly working on point processes, we study the number of the connected components of X in $A_u(X, g_uS)$ for any qualified $S \subset \mathbb{R}^N$. Its Poisson limit is then derived from the relationship between the number of the connected components and the number of the local maxima, combined with the main results in the last section. To this end, some notations will be inherited from the last section. Recall that we studied the avoidance functional $P[M_u(X, g_u\cdot) = 0]$ on $\mathcal{J}$, i.e., the set of bounded open cubes and their finite unions and intersections. For any bounded set $S \subset \mathbb{R}^N$ $(N \ge 1)$ with $\lambda_{N-1}(\partial S) < \infty$ as in Lemma 3.3.7, we can also define the event

$$H_u(S) := \left\{ \max_{\boldsymbol{t} \in g_u \overline{S - J_{1,u}(S)}} X(\boldsymbol{t}) \le u \right\}, \tag{3.27}$$

where $J_{1,u}(S)$ is defined in Lemma 3.3.7. The above notations will also play important roles in this section. Also recall that

$$M_u(X, S) := \# \left\{ \boldsymbol{t} \in S : X(\boldsymbol{t}) > u, \nabla X(\boldsymbol{t}) = \boldsymbol{0}, \text{index}\left(\nabla^2 X(\boldsymbol{t})\right) = N \right\},$$

where the index of a matrix is defined as the number of its negative eigenvalue. To indicate the index, we define the number of the critical points of X with index $0 \le k \le N$ above the threshold $u \in \mathbb{R}$ in the search region $S \subset \mathbb{R}^N$ (or equivalently, in the excursion set $A_u(X, S)$) as

$$M_u(X, S, k) := \# \left\{ \boldsymbol{t} \in S : X(\boldsymbol{t}) > u, \nabla X(\boldsymbol{t}) = \boldsymbol{0}, \text{index}\left(\nabla^2 X(\boldsymbol{t})\right) = k \right\}.$$

Obviously, we have

$$M_u(X, S, N) \equiv M_u(X, S).$$

Immediately, Corollary 3.2.1 can be generalized for any $M_u(X, S, k)$, $0 \le k \le N$.

Corollary 3.4.1. *(Lemma 11.7.1, [4]) Let $\{X(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^N\}$ be the same as in Corollary 3.2.1 and let S be a compact subset of $\mathbb{R}^N$ with $\lambda_{N-1}(S) < \infty$. Then for any $0 \le k \le N$,*

$$\mathbb{E}\left[M_{\boldsymbol{u}}(X, S, k)\right] = \int_S \int_u^\infty \int_{D_k} \left|\det\left(\boldsymbol{x}''\right)\right| p_{\boldsymbol{t}}(\boldsymbol{x}'', \boldsymbol{0}, x) d\boldsymbol{x}'' dx d\boldsymbol{t},$$

where

$$D_k := \{\boldsymbol{x}'' \in \mathbb{R}^{N(N+1)/2} : \text{index}\left(\boldsymbol{x}''\right) = k\}.$$

3.4.1 Expectation of the Number of Connected Components of the Excursion Set

Throughout this section, let $\{X(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^N\}$ be a Gaussian random field satisfying the conditions in Corollary 3.2.1, and let g_u satisfy (3.11). For any $K \in \mathcal{J}$, K is an open set and ∂K can be decomposed as

$$\partial K = \bigcup_{i=0}^{N-1} \bigcup_{J \in K_i} J,$$

where K_i is the set of i-dimensional open facets of ∂K. Here "open" means that K_i is an open subset of the i-dimensional Hyperplane H_J endowed with the usual Euclidean norm. Let $K_N := \{K\}$. From the definition of $\mathcal{J}$, we have

$$\# K_i < \infty \tag{3.28}$$

for $0 \le i \le N$, where "#" stands for "the cardinality of". For any $0 \le i \le N$ and $J \in K_i$, the facet boundary of J, written as F_J, is defined as

$$F_J := \begin{cases} \bigcup_{0 \le j \le i-1, J' \in K_j, J' \text{ is a facet of } J} J' & \text{if } 1 \le i \le N, \\ \emptyset & \text{if } i = 0. \end{cases}$$

Obviously, we have $\partial K = F_K$. As for $0 \le i \le N-1$ and $J \in K_i$, since $\partial J = \overline{J}$ in the space $(\mathbb{R}^N, \|\cdot\|)$, we do not have $\partial J = F_J$. Since each $J \in K_i$, $0 \le i \le N-1$ is an open subset of the i-dimensional Hyperplane H_J endowed with the usual Euclidean norm, it is easy to check that restriction $X|_{H_J}$ is also a Gaussian random field indexed by H_J, and satisfies all the conditions in Corollary 3.2.1, where we substitute $\mathbb{R}^N$ with H_J. Similarly, we denote by $M_u(X|_{H_J}, S_i, k)$ the number of the critical points of $X|_{H_J}$ with index $0 \le k \le i$ above the threshold $u \in \mathbb{R}$ in the search region $S_i \subset H_J$, i.e.,

$$M_u(X|_{H_J}, S_i, k) := \# \left\{ \boldsymbol{t} \in S_i : X|_{H_J}(\boldsymbol{t}) > u, \nabla X|_{H_J}(\boldsymbol{t}) = \boldsymbol{0}, \text{index}\left(\nabla^2 X|_{H_J}(\boldsymbol{t})\right) = k \right\}.$$

We also denote by $\widetilde{M}_u(X|_{H_J}, S_i)$ the true number of the local maxima of $X|_{H_J}$ in $S_i \subset H_J$ over $u \in \mathbb{R}$. For any $0 \le i \le N$, $J \in K_i$ and $u \in \mathbb{R}$, we define that

- $CC(A_u(X, g_u J))$ is the number of the connected components of $A_u(X, g_u J)$;
- $Ct(A_u(X, g_u J))$ is the number of the connected components of $A_u(X, g_u J)$ touching F_J (i.e., its intersection with F_J is not empty);

- $Cn(A_u(X, g_uJ))$ is the number of the connected components of $A_u(X, g_uJ)$ not touching F_J (i.e., its intersection with F_J is empty).

From the above definitions, it is easy to see that

1. if $J \in K_0$, then $F_J = \emptyset$ and $Ct(A_u(X, g_uJ)) = 0$;
2. $CC(A_u(X, g_uJ)) \equiv Ct(A_u(X, g_uJ)) + Cn(A_u(X, g_uJ))$.

Given $u \in \mathbb{R}$, the following are some useful observations about these quantities.

- For any $0 \le i \le N$ and $J \in K_i$, each connected component of $A_u(X, g_uJ)$ touching g_uF_J must contain at least one connected component of $A_u(X, g_uF_J)$, while obviously each connected component of $A_u(X, g_uF_J)$ cannot be the subset of two different connected components of $A_u(X, g_uJ)$ touching g_uF_J, simultaneously. Thus, we have

$$Ct(A_u(X, g_uJ)) \le CC(A_u(X, g_uF_J)). \tag{3.29}$$

- Each connected component of $A_u(X, g_u\partial K)$ must contain at least one connected component of $A_u(X, g_uJ)$ not touching g_uF_J for some $0 \le i \le N-1$ and $J \in K_i$, while obviously each connected component of $A_u(X, g_uJ)$ cannot be the subset of two different connected components of $A_u(X, g_u\partial K)$ simultaneously. Thus, we have

$$CC(A_u(X, g_u\partial K)) \le \sum_{i=0}^{N-1} \sum_{J \in K_i} Cn(A_u(X, g_uJ)). \tag{3.30}$$

- For any $1 \le i \le N$ and $J \in K_i$, let $X|_{H_J}$ be the restriction of X on J. Note that $X(\boldsymbol{t}) = u$ when $\boldsymbol{t}$ is in the boundary of a connected component of $A_u(X, g_uJ)$ not touching g_uF_J. Thus, the global maximum of each connected component of $A_u(X, g_uJ)$ not touching g_uF_J must exist and be a local maximum of $X|_{H_J}$, which implies

$$Cn(A_u(X, g_uJ)) \le \widetilde{M}_u(X|_{H_J}, g_uJ).$$

  Similarly, by Remark 2.1.12, we have

$$Cn(A_u(X, g_uJ)) \le M_u(X|_{H_J}, g_uJ) \tag{3.31}$$

  almost surely.

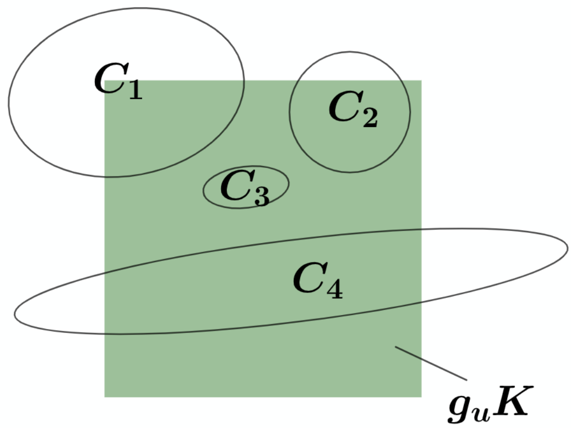


Figure 3.3: An example of $A_u(X, g_u\partial K)$ for some $u \in \mathbb{R}$. The interior of the shaded part is the search region $g_u K$, and the interior of each circle, written as C_i, $1 \le i \le 4$ is a connected component of $A_u(X, \mathbb{R}^N)$.

Example 3.4.1. *Figure 3.3 is an example of $A_u(X, g_u\partial K)$ for some $u \in \mathbb{R}$, where*

$$A_u(X, g_u\partial K) = \bigcup_{i=1}^{4} (C_i \cap g_u\partial K) .$$

From Figure 3.3, we have

1. $\sum_{J \in K_1} Ct(A_u(X, g_u K)) = 3$;
2. $CC(A_u(X, g_u\partial K)) = 4$;
3. $\sum_{J \in K_1} Cn(A_u(X, g_u J)) = 3$;
4. $\sum_{J \in K_0} Cn(A_u(X, g_u J)) = 1$.

Therefore, we can see that both (3.29) and (3.30) hold in this example.

Now we have made enough preparation for the following theorem.

Theorem 3.4.2. *Let X satisfy the conditions in Lemma 3.2.2, and let g_u satisfy (3.11). Then for any $K \in \mathcal{J}$,*

$$\lim_{u\to\infty} \mathbb{E}\left[Ct(A_u(X, g_u K))\right] = 0.$$

Proof. By (3.28), (3.29) and (3.30), we only need to show that for any $0 \leq i \leq N-1$ and $J \in K_i$,

$$\lim_{u\to\infty} \mathbb{E}\left[Cn(A_u(X, g_u J))\right] = 0.$$

If $i = 0$, then

$$\mathbb{E}\left[Cn(A_u(X, g_u J))\right] = P\left[X(\mathbf{0}) > u\right] \to 0, \qquad \text{as } u \to \infty.$$

If $1 \leq i \leq N-1$, then by (3.31),

$$Cn(A_u(X, g_u J)) \leq M_u(X|_{H_J}, g_u J)$$

almost surely, where $X|_{H_J}$ is the restriction of X on J. Thus, we only need to show that for any $1 \leq i \leq N-1$ and $J \in K_i$

$$\lim_{u\to\infty} \mathbb{E}\left[M_u(X|_{H_J}, g_u J)\right] = 0.$$

By Lemma 3.2.2 and (3.11), we have

$$\lim_{u\to\infty} g_u^N u^{N-1} \exp\left(-\frac{u^2}{2}\right) = C.$$

for some constant C. Then by Lemma 3.2.2, $g_u \to \infty$ as $u \to \infty$, and $1 \leq i \leq N-1$, we have

$$\begin{aligned}
\mathbb{E}\left[M_u(X|_{H_J}, g_u J)\right] &= \frac{\lambda_{i-1}(g_u J)\det\left(\mathbf{\Lambda}_{X|_{H_J}}\right)^{1/2} u^{i-1}}{(2\pi)^{(i+1)/2}} \exp\left(-\frac{u^2}{2}\right)\left(1 + O(u^{-1})\right) \\
&= g_u^N u^{N-1} \exp\left(-\frac{u^2}{2}\right) \frac{\lambda_{i-1}(J)\det\left(\mathbf{\Lambda}_{X|_{H_J}}\right)^{1/2} g_u^{i-N} u^{i-N}}{(2\pi)^{(i+1)/2}} \left(1 + O(u^{-1})\right) \\
&\to 0 \quad \text{ as } u \to \infty,
\end{aligned}$$

which completes the proof. □

3.4.2 Poisson Limit for the Number of Connected Components of the Excursion Set

Let the pair (X, g_u) be qualified. Endow $\mathbb{R}^N$ with the Euclidean norm $\|\cdot\|$. For any $\boldsymbol{t}_k = (t_{k1}, \ldots, t_{kN}) \in \mathbb{R}^N$, $k = 1, 2$ such that $\boldsymbol{t}_1 \neq \boldsymbol{t}_2$, $\boldsymbol{t}_1$ is said to be less than $\boldsymbol{t}_2$ in the lexicographical order on $\mathbb{R}^N$ if $t_{1n} < t_{2n}$ for some $1 \leq n \leq N$ and $t_{1i} = t_{2i}$ if $1 \leq i < n$ and $n > 1$. Since each bounded connected component of the excursion set $A_u(X, \mathbb{R}^N)$ is a non-empty open set and $X(\boldsymbol{t}) = u$ for all $\boldsymbol{t} \in \partial A_u(X, \mathbb{R}^N)$, the global maximum of X in this connected component must exist and be a local maximum. Note that there can be more than one global maximum taking the same value in this connected component. In this case, we only consider the global maximum with the smallest lexicographical order in this connected component. Denote by $G_u(X)$ the set of the global maxima in $A_u(X, \mathbb{R}^N)$ as above. Then by Remark 2.1.12 and Lemma 3.2.2, we have for any compact set T, with probability one, $G_u(X) \cap T$ is a finite set, and for any $\boldsymbol{t} \in G_u(X) \cap T$, $X(\boldsymbol{t})$ is a strict local maxima of X with index N. Thus, we can define a new family of point processes $\{N_u(X, g_u \cdot), u \in \mathbb{R}\}$ such that for any $S \in \mathcal{B}(\mathbb{R}^N)$, $N_u(X, S)$ counts the number of points in $G_u(X) \cap S$. From the above, it is easy to see that for any $S \in \mathcal{B}(\mathbb{R}^N)$,

$$N_u(X, g_u S) \leq \min\{CC(A_u(X, g_u S)), M_u(X, g_u S)\} \tag{3.32}$$

almost surely. The following theorem establishes the Poisson limit of $N_u(X, g_u \cdot)$ as $u \to \infty$.

Theorem 3.4.3. *Let (X, g_u, c) be a qualified triple, and let $N_u(X, g_u \cdot)$ be as above. Then as $u \to \infty$, $N_u(X, g_u \cdot)$ converge weakly to a Poisson point process with mean measure $c\lambda_N$.*

Proof. Note that for any $K \in \mathcal{J}$, the following two events are equivalent:

1. $\{N_u(X, g_u K) = 0\} \cap H_u(K)$;
2. $\{M_u(X, g_u K) = 0\} \cap H_u(K)$.

Therefore,

$$P\left[N_u(X, g_u K) = 0 \mid H_u(K)\right] = P\left[M_u(X, g_u K) = 0 \mid H_u(K)\right].$$

Then by Inequality (3.16) and Lemma B.1.5, we have

$$\begin{aligned}
\lim_{u\to\infty} P\left[N_u(X, g_u K) = 0\right] &= \lim_{u\to\infty} P\left[N_u(X, g_u K) = 0 \mid H_u(K)\right] \\
&= \lim_{u\to\infty} P\left[M_u(X, g_u K) = 0 \mid H_u(K)\right] \\
&= \lim_{u\to\infty} P\left[M_u(X, g_u K) = 0\right] \\
&= e^{-c\lambda_N(K)}.
\end{aligned} \tag{3.33}$$

Moreover, by (3.32), it is easy to see that for any $K \in \mathcal{J}$,

$$\limsup_{u\to\infty} \mathbb{E}\left[N_u(X, g_u K)\right] \le \lim_{u\to\infty} \mathbb{E}\left[M_u(X, g_u K)\right] = c\lambda_N(K) < \infty. \tag{3.34}$$

Finally, combining (3.33), (3.34) and Lemma 2.2.7 completes the proof. □

Combining Lemma 2.2.6, Corollary 3.2.1 and Theorem 3.4.3, we have for any bounded $S \in \mathcal{B}(\mathbb{R}^N)$ with $\lambda_{N-1}(\partial S) < \infty$ and non-negative integer k,

$$\lim_{u\to\infty} P\left[N_u(X, g_u S) = k\right] = \lim_{u\to\infty} P\left[M_u(X, g_u S) = k\right] = e^{-c\lambda_N(S)}\frac{c^k \lambda_N(S)^k}{k!}, \tag{3.35}$$

where, in Lemma 2.2.6, we let

$$f(\boldsymbol{t}) = \mathbb{1}_S(\boldsymbol{t}) := \begin{cases} 1 & \text{if } \boldsymbol{t} \in S, \\ 0 & \text{if } \boldsymbol{t} \notin S. \end{cases}$$

In consequence, we can study the Poisson limit of the number, denoted by $CC(A_u(X, g_u S))$, of the connected components of the excursion set $A_u(X, g_u S)$ by (3.35) and the relationship between $N_u(X, g_u S)$ and $CC(A_u(X, g_u S))$ for any bounded set $S \in \mathcal{B}(\mathbb{R}^N)$ with $\lambda_{N-1}(\partial S) < \infty$.

Theorem 3.4.4. *Let (X, g_u, c) be a qualified triple. For any bounded set $S \in \mathcal{B}(\mathbb{R}^N)$ with $\lambda_{N-1}(\partial S) < \infty$, $CC(A_u(X, g_u S))$ converge weakly to a Poisson random variable with mean $c\lambda_N(S)$.*

Proof. Note that for any bounded set $S \in \mathcal{B}(\mathbb{R}^N)$ with $\lambda_{N-1}(\partial S) < \infty$ and any non-negative integer k, the following events are equivalent:

1. $\{N_u(X, g_u S) = k\} \cap H_u(S)$;
2. $\{CC(A_u(X, g_u S)) = k\} \cap H_u(S)$.

Therefore,

$$P\left[N_u(X, g_u S) = k \mid H_u(S)\right] = P\left[CC(A_u(X, g_u S)) = k \mid H_u(S)\right].$$

By Lemmas 3.3.7 and B.1.5, we have

$$\lim_{u\to\infty} P\left[H_u(S)\right] = 1. \tag{3.36}$$

Then by Inequality (3.16), (3.35) and (3.36), we have

$$\begin{aligned}
\lim_{u\to\infty} P\left[CC(A_u(X,g_uS)) = k\right] &= \lim_{u\to\infty} P\left[CC(A_u(X,g_uS)) = k \mid H_u(S)\right] \\
&= \lim_{u\to\infty} P\left[N_u(X,g_uS) = k \mid H_u(S)\right] \\
&= \lim_{u\to\infty} P\left[N_u(X,g_uS) = k\right] \\
&= e^{-c\lambda_N(S)}\frac{c^k\lambda_N(S)^k}{k!}
\end{aligned}$$

for any non-negative integer k, and hence the proof is completed. □

3.4.3 A Further Relationship between the Number of Local Maxima and the Number of Connected Components

In this section, we will show some asymptotic results about the number of the critical points over some threshold. We will also relate these results to the number of the connected components of an excursion set we have discussed in the last section.

Lemma 3.4.5. *Let the pair (X, g_u) be qualified. Then for any bounded set $S \subset \mathbb{R}^N$ with $\lambda_{N-1}(\partial S) < \infty$, we have*

$$\lim_{u\to\infty} \mathbb{E}\left[M_u(X, g_uS, k)\right] = 0$$

for every $0 \le k \le N-1$.

Proof. See Appendix B.1.4. □

Theorem 3.4.6. *Let (X, g_u, c) be a qualified triple. Then as $u \to \infty$, $\sum_{k=0}^{N} M_u(X, g_u\cdot, k)$ converge weakly to a Poisson point process with mean measure $c\lambda_N$.*

Proof. Follow the proof of Theorem 3.4.3 and use Lemma 3.4.5. □

The following lemma is a simple application of Theorem 2.3.4. Recall that we have defined Morse functions in Section 2.3.

Lemma 3.4.7. *Let f be a Morse function on an open set containing the compact set $S \subset \mathbb{R}^N$ $(N \ge 1)$. Suppose that D is a connected component of $A_u(f, S)$ for some $u \in \mathbb{R}$ such that $\partial D \cap \partial S = \emptyset$. Suppose that $\boldsymbol{p}_1, \boldsymbol{p}_2 \in D$ are two local maxima of f with index N. Then there exists a critical point of f in D with index less than N.*

Combining Lemma 3.4.5 with Lemma 3.4.7, we can further get the following closer relationship between the number of the connected components and the number of the local maxima with index N when the threshold is very high.

Theorem 3.4.8. *Let the pair (X, g_u) be qualified, and let the set $S \subset \mathbb{R}^N$ $(N \geq 1)$ be bounded with $\lambda_{N-1}(\partial S) < \infty$. Then*

$$\lim_{u\to\infty} P[CC(A_u(X, g_uS)) = M_u(X, g_uS) = N_u(X, g_uS)] = 1.$$

Proof. By Lemma 3.4.5, we have

$$\lim_{u\to\infty} \mathbb{E}\left[\sum_{k=0}^{N-1} M_u(X, g_uS, k)\right] = 0,$$

which implies that

$$\lim_{u\to\infty} P\left[\sum_{k=0}^{N-1} M_u(X, g_uS, k) = 0\right] = 1.$$

By Lemmas 3.3.7 and B.1.5, we have

$$\lim_{u\to\infty} P\left[H_u(S)\right] = 1,$$

where $H_u(S)$ is defined in (3.27). Then by the Inequality (3.16), we have

$$\lim_{u\to\infty} P\left[\sum_{k=0}^{N-1} M_u(X, g_uS, k) = 0 \;\middle|\; H_u(S)\right] = 1.$$

By Lemma 3.4.7, it is easy to see that the event $\{\sum_{k=0}^{N-1} M_u(X, g_uS, k) = 0\} \cap H_u(S)$ implies the event $\{CC(A_u(X, g_uS)) = M_u(X, g_uS) = N_u(X, g_uS)\} \cap H_u(S)$. Thus,

$$\lim_{u\to\infty} P\left[CC(A_u(X, g_uS)) = M_u(X, g_uS) = N_u(X, g_uS) \mid H_u(S)\right] = 1.$$

Finally, again by Inequality (3.16), we have

$$\lim_{u\to\infty} P\left[CC(A_u(X, g_uS)) = M_u(X, g_uS) = N_u(X, g_uS)\right] = 1.$$

□

Chapter 4

Local Behavior of Critical Points

4.1 Introduction

In Section 3.4.3, we have explored that the critical points of a stationary Gaussian random field indexed by $\mathbb{R}^N$ above an extremely high threshold behave very similarly to a Poisson point process on $\mathbb{R}^N$, and each connected component of the excursion set tends to contain only one critical point which is the global maximum of the Gaussian random field in this connected component. However, when the threshold is high but not extremely high, the possibility for a connected component to contain more than one critical point increases. In this case, the main difference between the critical points above the threshold and a Poisson point process lies in their local behaviors which can be characterized by the interactions between different critical points. In fact, there are many excellent results on the local structure of a Gaussian random field near a high excursion point ([24], [10], [3]). In this chapter, this excursion point is chosen to be a critical point with unknown index, and we are interested in its impact on other critical points nearby.

When the Gaussian random field is isotropic, we can simply choose the origin as the given high critical point without loss of generality. Then conditional on the event that

$$\text{the origin is a critical point above the threshold } u,$$

we can define $f_{u,k}(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^N$ as the density of the mean measure of the critical points above u with index k ($0 \leq k \leq N$). The existence of these densities are guaranteed by Lemma 2.1.10 as shown later.

Note that all of these densities of mean measures will decrease to zero as $u \to \infty$ since a higher threshold can always screen out more critical points. Thus, their relative values carry more information about the interactions. In this chapter, we mainly consider two ratios. The first ratio is

$$\frac{\sum_{k \text{ even}} f_{u,k}(\boldsymbol{t})}{\sum_{k \text{ odd}} f_{u,k}(\boldsymbol{t})},$$

which reflects how a high critical point affects the signs of the determinants of the Hessian matrices of others. We will see its limiting behavior as $\|\boldsymbol{t}\| \to 0$ for any $u \in \mathbb{R}$. The second ratio is

$$\frac{\sum_{k=1}^{N-2} f_{u,k}(\boldsymbol{t})}{f_{u,N-1}(\boldsymbol{t}) + f_{u,N}(\boldsymbol{t})},$$

which reflects whether the critical points with index greater than or equal to $N-1$ are more likely to appear than others near a high critical point. We will establish its limiting behavior as $u \to \infty$ on a compact region. The combination of the results on these ratios reveals a profound relationship between the local maxima and critical points with index $N-1$ above a high threshold. This relationship also serves as a part of the intuition for the next chapter.

The remainder of this chapter is organized as follows. Section 4.2 reviews some existing results about Gaussian random fields and matrix perturbations. Section 4.3 introduces basic settings including notations and the conditions on the underlying Gaussian random field in this chapter. For convenience, we still call a Gaussian random field qualified if it satisfies these conditions. One should not confuse a "qualified" Gaussian random field with a "qualified triple" in Chapter 3. Next, there is a covariance matrix playing an important role in the construction of our theory, and we collect some useful properties of it in Section 4.4. Based on these properties, Sections 4.5 and 4.6 state and prove our main results about the two ratios above, respectively.

4.2 Literature Review

Let $\{X(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^N\}$ be a centered, isotropic Gaussian random field. Endow $\mathbb{R}^N$ with the usual Euclidean norm $\|\cdot\|$. Due to isotropy, there exists some function $\rho : [0, \infty) \to \mathbb{R}$ such that the covariance function of X can be written as

$$\mathbb{E}\left[X(\boldsymbol{s})X(\boldsymbol{t})\right] = \rho(\|\boldsymbol{t} - \boldsymbol{s}\|^2) \text{ for any } \boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^N.$$

Lemma 4.2.1. *(Proposition 3.3, [11]) Let $\{X(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^N\}$ be a centered, isotropic Gaussian random field possessing almost sure derivatives of up to second order and satisfying*

Condition (2.13) in Lemma 2.1.10. Then the Gaussian vector $(\nabla^2 X(\boldsymbol{t}), \nabla X(\boldsymbol{t}), X(\boldsymbol{t}))$ *is non-degenerate if and only if*

$$\rho^{(1)}(0) \neq 0 \text{ and } \frac{\rho^{(2)}(0)}{\rho^{(1)}(0)^2} > \frac{N}{N+2}.$$

For any positive integer n, endow $\mathbb{R}^{n\times n}$ with the spectral norm $\|\cdot\|_s$, i.e., for any $\boldsymbol{A} \in \mathbb{R}^{n\times n}$,

$$\|\boldsymbol{A}\|_s = \max\{\sqrt{\nu} : \nu \text{ is an eigenvalue of } \boldsymbol{A}^T\boldsymbol{A}\}$$

.

Lemma 4.2.2. *(Pages 405 and 411, [17]) Let* C_n *be the set of all positive semi-definite matrix in* $\mathbb{R}^{n\times n}$*. Then for any* $\boldsymbol{\Sigma} \in C_n$*, there exists a unique matrix* $\boldsymbol{A}$ *in* C_n *such that* $\boldsymbol{\Sigma} = \boldsymbol{A}\boldsymbol{A}^T$*. Here* $\boldsymbol{A}$ *is called the* ***non-negative square root*** *of* $\boldsymbol{\Sigma}$ *and can be written as* $\boldsymbol{\Sigma}^{1/2}$*. Moreover, let* $f_n : C_n \to C_n$*, where* $f(\boldsymbol{\Sigma}) = \boldsymbol{\Sigma}^{1/2}$*. Then* f_n *is continuous on the interior of* C_n*.*

Lemma 4.2.3. *(Theorem 5.1, [20]) Let* $\boldsymbol{T} : \mathbb{R} \to \mathbb{R}^{n\times n}$ *be continuous on* $A \subset \mathbb{R}$*. Then there exist continuous functions* $\mu_1, \dots, \mu_n$ *on* $x \in A$ *such that*

$$\boldsymbol{T}(x) = \boldsymbol{P}(x)\boldsymbol{\Lambda}(x)\boldsymbol{P}^T(x),$$

for any $x \in A$*, where* $\boldsymbol{\Lambda}(x) := \operatorname{diag}(\mu_1(x), \dots, \mu_n(x))$ *and* $\boldsymbol{P}(x)\boldsymbol{P}^T(x) = \boldsymbol{I}_n$*.*

4.3 Basic Settings

In the following sections, we will make heavy calculation of matrices and their sub-matrices. Let $\mathbb{N} := \{0, 1, \dots\}$ and $\mathbb{Z}^+ := \{1, 2, \dots\}$. For any matrix $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ ($m, n \in \mathbb{Z}^+$), denote by $\boldsymbol{M}_{(i)}$ its i-th row and by $\boldsymbol{M}^{(j)}$ its j-th column for $i = 1, \dots, m$ and $j = 1, \dots, n$. For any $a, b \in \mathbb{Z}^+$ such that $a \leq b$, denote by $a : b$ the set of all integers in $[a, b]$. For any $S_1 \subset 1 : m$ and $S_2 \subset 1 : n$, denote by $\boldsymbol{M}[S_1, S_2]$ the sub-matrix of $\boldsymbol{M}$ formed by the entries with row indices in S_1 and column indices in S_2. For a real vector $\boldsymbol{a} \in \mathbb{R}^n$ (row or column), we use $\boldsymbol{a}[i]$ to denote its i-th coordinate for $i = 1, \dots, n$. Denote by $\boldsymbol{I}_n$ the identity matrix of size n, by $\boldsymbol{0}_n$ the origin of $\mathbb{R}^n$, and by $\boldsymbol{0}_{m\times n}$ the origin of $\mathbb{R}^{m\times n}$. All vectors are column vectors by default.

The following conventional notations will frequently appear in this chapter. The Kronecker delta is defined by

$$\delta_{i,j} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}.$$

For any $n \in \mathbb{Z}^+$, let λ_n ($n \in \mathbb{Z}^+$) be the n-dimensional Lebesgue measure, and let $\|\cdot\|_n$ be the usual Euclidean norm of $\mathbb{R}^n$. For conciseness, We often drop the subscript n when the dimension can be easily seen from the context. For any matrix $\boldsymbol{M} \in \mathbb{R}^{m\times n}$ ($m, n \in \mathbb{Z}^+$), denote by $\text{Rank}\,(\boldsymbol{M})$ the rank of $\boldsymbol{M}$. Given a vector space V over $\mathbb{R}$, for any $S \subset V$, denote by $\text{span}\,(S)$ the linear span of S, i.e.,

$$\text{span}\,(S) := \left\{ \sum_{i=1}^{n} k_i \boldsymbol{v}_i \middle| n \in \mathbb{N}, k_i \in \mathbb{R}, \boldsymbol{v}_i \in S \right\}.$$

For any two non-negative real-valued functions $h(x)$ and $g(x)$, $x \in \mathbb{R}$, we write

$$h(x) = \Theta(g(x)) \text{ as } x \to \infty$$

if there exist positive constants k_1, k_2 and C such that for any $x > C$,

$$k_1 g(x) \le h(x) \le k_2 g(x).$$

For any two times differentiable function $f : \mathbb{R}^N \to \mathbb{R}$, let ∇f and $\nabla^2 f$ be the gradient and the Hessian matrix of f, respectively. Recall that in Section 2.3, the point $\boldsymbol{t} \in \mathbb{R}^N$ is said to be a critical point of f if $\nabla f(\boldsymbol{t}) = \boldsymbol{0}$. The index or type of a critical point $\boldsymbol{t}$ is defined to be the number of negative eigenvalues (counted with their multiplicities) of $\nabla^2 f(\boldsymbol{t})$. For example, a local maximum is typically a critical point with index N.

Recall that for any symmetric matrix $\boldsymbol{M} = (m_{ij}) \in \mathbb{R}^{N\times N}$, a vector $\boldsymbol{a} \in \mathbb{R}^{N(N+1)/2}$ is said to be the usual vectorization of $\boldsymbol{M}$ if

$$\boldsymbol{a}[i + j(j-1)/2] = m_{ij} \text{ for any } 1 \le i \le j \le N.$$

Definition 4.3.1. (Matriculation) A matrix $\boldsymbol{M} \in \mathbb{R}^{N\times N}$ is said to be the N-th order matriculation of a vector $\boldsymbol{a} \in \mathbb{R}^n$ ($n \ge N(N+1)/2$), written as $\boldsymbol{M} = \text{Matri}_N(\boldsymbol{a})$, if $\boldsymbol{a}[1 : (N(N+1)/2)]$ is the usual vectorization of $\boldsymbol{M}$.

It is noticeable that different vectors, with different coordinate values or even different lengths, can share the same N-th order matriculation since only the first $N(N+1)/2$ elements of them are considered.

Definition 4.3.2. An isotropic Gaussian random field $X(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^N$ ($N \ge 2$) with covariance structure

$$\rho(\|\boldsymbol{t} - \boldsymbol{s}\|^2) \equiv R(\boldsymbol{t} - \boldsymbol{s}) := \text{Cov}\,[X(\boldsymbol{s}), X(\boldsymbol{t})]$$

is said to be **qualified** if the following conditions are satisfied:

(1) X is centered with unit variance, i.e., $\mathbb{E}[X(\mathbf{0})] = 0$ and $\rho(0) = 1$.

(2) X has almost surely partial derivatives of up to second order.

(3) The sixth derivative of $\rho(x)$ at $x = 0$ exists, which implies there exists a constant $\delta_\rho > 0$ such that the fifth-order derivative of $\rho(x)$ exists and is bounded on $x \in [0, \delta_\rho^2]$.

(4) The distribution of
$$(\nabla^2 X(\boldsymbol{t}), \nabla X(\boldsymbol{t}), X(\boldsymbol{t}), \nabla X(\mathbf{0}), X(\mathbf{0}))$$
is non-degenerate for any $\boldsymbol{t} \in \mathbb{R}^N \setminus \{\mathbf{0}_N\}$, and the distribution of $(X_{111}(\mathbf{0}), \nabla X(\mathbf{0}))$ is also non-degenerate (note that by Lemma 2.1.4 and Condition (4), the mean square derivative $X_{111}(\mathbf{0})$ exists).

Remark 4.3.3. The above conditions allow us to apply Lemma 4.2.1 to X (see Lemma 2.1.13). If $\rho(x)$ is four times continuously differentiable on a compact set $T \subset \mathbb{R}^N \setminus \{\mathbf{0}_N\}$ with $\lambda_{N-1}(\partial T) < \infty$, then Conditions (1), (2) and (4) allow us to apply Lemma 2.1.10 to the Gaussian random field $X(\boldsymbol{t})$ conditional on $\{\nabla X(\mathbf{0}) = \mathbf{0}_N, X(\mathbf{0}) = z\}$ for any $z \in \mathbb{R}$ (see Appendix C.2 and Remark 2.1.11). Condition (3) is critical in the calculation of the asymptotic expansion of $\boldsymbol{\Sigma}(\boldsymbol{t})$ as $\|\boldsymbol{t}\| \to 0$ introduced later. Conditions (3) and (4) are also associated with the convergence speeds of those ordered eigenvalues of $\boldsymbol{\Sigma}(\boldsymbol{t})$ converging to 0 as $\|\boldsymbol{t}\| \to 0$.

Remark 4.3.4. Indeed, the assumption that X is qualified imposes some constraints on the derivatives of ρ. Let $\rho^{(i)}(x)$, $x \geq 0$ and $i \geq 1$ be the i-th derivative (if exists) of ρ at x. Firstly, by Condition (4) in Definition 4.3.2, X_1, X_{11} and X_{111} are all non-degenerate. Then by Lemma 2.1.4 and (C.1),
$$-2\rho^{(1)}(0) = \mathrm{Var}[X_1(\mathbf{0})] > 0,$$
by (C.4),
$$12\rho^{(2)}(0) = \mathrm{Var}[X_{11}(\mathbf{0})] > 0,$$
and by (C.5),
$$-120\rho^{(3)}(0) = \mathrm{Var}[X_{111}(\mathbf{0})].$$
This implies
$$\rho^{(1)}(0) < 0, \rho^{(2)}(0) > 0 \text{ and } \rho^{(3)}(0) < 0. \tag{4.1}$$

Secondly, let $\alpha = \rho^{(1)}(0)^{-1}\rho^{(2)}(0)^2$ and $\beta = \rho^{(3)}(0)$. Then by Condition (4) in Definition 4.3.2 and the Cauchy–Schwarz inequality,
$$(\mathrm{Cov}\,[X_{111}(\mathbf{0}), X_1(\mathbf{0})])^2 < \mathrm{Var}[X_{111}(\mathbf{0})]\mathrm{Var}[X_1(\mathbf{0})].$$

By Lemma 2.1.4, (C.2), (C.4) and (C.5), this is equivalent to

$$\left(12\rho^{(2)}(0)\right)^2 < \left(-120\rho^{(3)}(0)\right)\left(-2\rho^{(1)}(0)\right).$$

Then by $\rho^{(1)}(0) < 0$, we have

$$\alpha > \frac{5}{3}\beta. \tag{4.2}$$

Moreover, for any $r > 0$ and $\boldsymbol{t} \in \mathbb{R}^N$, let $B(\boldsymbol{t}, r)$ be the N-dimensional open ball centered at $\boldsymbol{t}$ with radius r. Then by the Cauchy–Schwarz inequality, we have for any $1 \le i \le N$ and $\boldsymbol{t} \in B(\mathbf{0}_N, \delta_\rho)$ (see δ_ρ in Definition 4.3.2),

$$|\mathrm{Cov}\left[X_i(\boldsymbol{t}), X_i(\mathbf{0})\right]| \le \sqrt{\mathrm{Var}[X_i(\boldsymbol{t})]\mathrm{Var}[X_i(\mathbf{0})]}.$$

By Lemma 2.1.4 and (C.2), this is equivalent to

$$\left|-\rho^{(1)}(\|\boldsymbol{t}\|^2) - 2t_i^2\rho^{(2)}(\|\boldsymbol{t}\|^2)\right| \le -\rho^{(1)}(0).$$

Note that the equal sign in the above inequality holds if and only if $X_i(\boldsymbol{t})$ and $X_i(\mathbf{0})$ are linearly dependent, which is impossible for any $\boldsymbol{t} \in B(\mathbf{0}_N, \delta_\rho) \setminus \{\mathbf{0}_N\}$ by Condition (4) in Definition 4.3.2. Thus, for any $1 \le i \le N$ and $\boldsymbol{t} \in B(\mathbf{0}_N, \delta_\rho) \setminus \{\mathbf{0}_N\}$,

$$\left|-\rho^{(1)}(\|\boldsymbol{t}\|^2) - 2t_i^2\rho^{(2)}(\|\boldsymbol{t}\|^2)\right| < -\rho^{(1)}(0). \tag{4.3}$$

Then for any $x \in (0, \delta_\rho^2]$, by taking $\boldsymbol{t} := (0, \ldots, 0, \sqrt{x}) \in \mathbb{R}^N$ and $i = 1$ in (4.3), we have

$$\left|\rho^{(1)}(x)\right| < -\rho^{(1)}(0). \tag{4.4}$$

Finally, by Lemma 4.2.1, we see

$$1 - \frac{\rho^{(1)}(0)^2}{3\rho^{(2)}(0)} > 1 - \frac{N+2}{3N} \ge 0.$$

Let X be qualified and $L := N(N+1)/2 + 2$. For any $\boldsymbol{t} \in \mathbb{R}^N \setminus \{\mathbf{0}_N\}$, by Lemma 2.1.4 and Appendix A.1,

$$(\nabla^2 X(\boldsymbol{t}), X(\boldsymbol{t}), X(\mathbf{0}) | \nabla X(\boldsymbol{t}) = \nabla X(\mathbf{0}) = \mathbf{0}_N)$$

(i.e., $(\nabla^2 X(\boldsymbol{t}), X(\boldsymbol{t}), X(\mathbf{0}))$ conditional on $\nabla X(\boldsymbol{t}) = \nabla X(\mathbf{0}) = \mathbf{0}_N$) is a Gaussian L-vector, and let $\boldsymbol{\Sigma}(\boldsymbol{t})$ be its covariance matrix. By (A.8) and Condition (4) in Definition 4.3.2, $\boldsymbol{\Sigma}(\boldsymbol{t})$ is positive-definite for any $\boldsymbol{t} \in \mathbb{R}^N \setminus \{\mathbf{0}_N\}$. Let

$$\lambda_1(\boldsymbol{t}) \ge \lambda_2(\boldsymbol{t}) \ge \cdots \ge \lambda_L(\boldsymbol{t}) > 0 \tag{4.5}$$

be the ordered eigenvalues of $\boldsymbol{\Sigma}(\boldsymbol{t})$. Then an eigen-decomposition of $\boldsymbol{\Sigma}(\boldsymbol{t})$ is given by

$$\boldsymbol{\Sigma}(\boldsymbol{t}) = \boldsymbol{P}(\boldsymbol{t})\boldsymbol{\Lambda}(\boldsymbol{t})\boldsymbol{P}^T(\boldsymbol{t}),$$

where $\boldsymbol{P}(\boldsymbol{t})$ is an $L \times L$ orthogonal real matrix and $\boldsymbol{\Lambda}(\boldsymbol{t}) := \mathrm{diag}\,(\lambda_1(\boldsymbol{t}), \ldots, \lambda_L(\boldsymbol{t}))$. One should note that $\boldsymbol{P}(\boldsymbol{t})$ in the above eigen-decomposition may be **not** unique since the ordering in (4.5) is not strict.

For any $\boldsymbol{t} \in \mathbb{R}^N \setminus \{\mathbf{0}_N\}$, denote

$$\boldsymbol{A}(\boldsymbol{t}) := \boldsymbol{P}(\boldsymbol{t})\boldsymbol{\Lambda}^{1/2}(\boldsymbol{t}) = \boldsymbol{P}(\boldsymbol{t})\mathrm{diag}\left(\lambda_1^{1/2}(\boldsymbol{t}), \ldots, \lambda_L^{1/2}(\boldsymbol{t})\right).$$

Immediately, we have for any $\boldsymbol{t} \in \mathbb{R}^N \setminus \{\mathbf{0}_N\}$,

$$\boldsymbol{\Sigma}(\boldsymbol{t}) = \boldsymbol{A}(\boldsymbol{t})\boldsymbol{A}^T(\boldsymbol{t}),$$

and $\boldsymbol{A}(\boldsymbol{t})$ can be uniquely determined by a version of $\boldsymbol{P}(\boldsymbol{t})$.

Denote by $\mathbb{S}^{N-1}$ the unit $(N-1)$-sphere. Since X is isotropic, we will focus on the behavior of $\boldsymbol{\Sigma}(\boldsymbol{u}r)$ along a given direction $\boldsymbol{u} = (u_1, \ldots, u_N)^T \in \mathbb{S}^{N-1}$ for $r > 0$. Thus, it would be more convenient and comprehensible to adopt the notations:

$$\boldsymbol{\Sigma}_{\boldsymbol{u}}(r) := \boldsymbol{\Sigma}_{\boldsymbol{u}}(r), \quad \boldsymbol{A}_{\boldsymbol{u}}(r) := \boldsymbol{A}_{\boldsymbol{u}}(r), \quad \boldsymbol{P}_{\boldsymbol{u}}(r) := \boldsymbol{P}_{\boldsymbol{u}}(r), \text{ and } \boldsymbol{\Lambda}_{\boldsymbol{u}}(r) := \boldsymbol{\Lambda}_{\boldsymbol{u}}(r),$$

which emphasize on them being the matrix-valued functions of r.

Let $M_L(\mathbb{R})$ be the space of all $L \times L$ real symmetric matrices endowed with the Frobenius norm $\|\cdot\|_F$, i.e., for any $\boldsymbol{A} = (a_{ij}) \in M_L(\mathbb{R})$, $\|\boldsymbol{A}\|_F = \sqrt{\sum_{i,j=1}^L a_{ij}^2}$. Endow $\mathbb{R}^N$ with the usual Euclidean norm. The following lemma describes the behavior of $\boldsymbol{\Sigma}_{\boldsymbol{u}}(r)$ as $r \to 0$.

Lemma 4.3.5. *Let X be qualified. Then for any direction $\boldsymbol{u} = (u_1, \ldots, u_N)^T \in \mathbb{S}^{N-1}$, $\boldsymbol{\Sigma}_{\boldsymbol{u}}(0) := \lim_{r \downarrow 0} \boldsymbol{\Sigma}_{\boldsymbol{u}}(r)$ exists, and the function $\boldsymbol{\Sigma}_{\boldsymbol{u}}(\cdot) : [0, \infty) \to M_L(\mathbb{R})$ is continuous on $[0, \delta_\rho]$, where δ_ρ is as defined in Definition 4.3.2. In addition, we have as $r \to 0$,*

$$\boldsymbol{\Sigma}_{\boldsymbol{u}}(r) = \boldsymbol{\Sigma}_{\boldsymbol{u},0} + \boldsymbol{\Sigma}_{\boldsymbol{u},2}r^2 + o(r^2),$$

where $\boldsymbol{\Sigma}_{\boldsymbol{u},0}, \boldsymbol{\Sigma}_{\boldsymbol{u},2} \in M_L(\mathbb{R})$ satisfy

1. *$\boldsymbol{\Sigma}_{\boldsymbol{u},0} = \boldsymbol{\Sigma}_{\boldsymbol{u}}(0)$ is positive semi-definite;*

2. *for any* $1 \le i_1 \le j_1 \le N$ *and* $1 \le i_2 \le j_2 \le N$*, we have*

$$\begin{aligned} &\boldsymbol{\Sigma}_{\boldsymbol{u},0}[i_1 + j_1(j_1-1)/2, i_2 + j_2(j_2-1)/2] \\ &= 4\rho^{(2)}(0)(\delta_{i_2,j_1}\delta_{i_1,j_2} + \delta_{i_1,i_2}\delta_{j_1,j_2} - \delta_{j_1,j_2}u_{i_1}u_{i_2} - \delta_{i_1,j_2}u_{j_1}u_{i_2} \\ &\quad - \delta_{i_2,j_1}u_{i_1}u_{j_2} - \delta_{i_1,i_2}u_{j_1}u_{j_2} + 2u_{i_1}u_{j_1}u_{i_2}u_{j_2}) \\ &\quad + \frac{8}{3}\rho^{(2)}(0)(\delta_{i_1,j_1} - u_{i_1}u_{j_1})(\delta_{i_2,j_2} - u_{i_2}u_{j_2}) \end{aligned} \tag{4.6}$$

and

$$\begin{aligned} &\boldsymbol{\Sigma}_{\boldsymbol{u},2}[i_1 + j_1(j_1-1)/2, i_2 + j_2(j_2-1)/2] \\ &= \left(2\alpha - \frac{14}{9}\beta\right)\delta_{i_1,j_1}\delta_{i_2,j_2} + \left(4\alpha - \frac{52}{9}\beta\right)\delta_{i_2,j_2}u_{i_1}u_{j_1} + \left(4\alpha - \frac{52}{9}\beta\right)\delta_{i_1,j_1}u_{i_2}u_{j_2} \\ &\quad + (2\alpha - 6\beta)\,\delta_{j_1,j_2}u_{i_1}u_{i_2} + (2\alpha - 6\beta)\,\delta_{i_1,j_2}u_{j_1}u_{i_2} + (2\alpha - 6\beta)\,\delta_{i_2,j_1}u_{i_1}u_{j_2} \\ &\quad + (2\alpha - 6\beta)\,\delta_{i_1,i_2}u_{j_1}u_{j_2} + \frac{64}{9}\beta u_{i_1}u_{j_1}u_{i_2}u_{j_2}. \end{aligned} \tag{4.7}$$

where $\alpha = \rho^{(1)}(0)^{-1}\rho^{(2)}(0)^2$ *and* $\beta = \rho^{(3)}(0)$*;*

3. *for any* $1 \le i_1 \le j_1 \le N$*, we have*

$$\boldsymbol{\Sigma}_{\boldsymbol{u},0}[i_1 + j_1(j_1-1)/2, L] = \boldsymbol{\Sigma}_{\boldsymbol{u},0}[i_1 + j_1(j_1-1)/2, L-1] = \frac{4}{3}\rho^{(1)}(0)(\delta_{i_1,j_1} - u_{i_1}u_{j_1}) \tag{4.8}$$

and

$$\begin{aligned} &\boldsymbol{\Sigma}_{\boldsymbol{u},2}[i_1 + j_1(j_1-1)/2, L] = \boldsymbol{\Sigma}_{\boldsymbol{u},2}[i_1 + j_1(j_1-1)/2, L-1] \\ &= \left(\frac{1}{3}\alpha' - \frac{1}{9}\beta'\right)\delta_{i_1,j_1} + \left(\frac{2}{3}\alpha' - \frac{14}{9}\beta'\right)u_{i_1}u_{j_1}, \end{aligned}$$

where $\alpha' = \rho^{(2)}(0)$ *and* $\beta' = \rho^{(1)}(0)\rho^{(2)}(0)^{-1}\rho^{(3)}(0)$*;*

4.

$$\begin{aligned} &\boldsymbol{\Sigma}_{\boldsymbol{u},0}[L-1, L-1] = \boldsymbol{\Sigma}_{\boldsymbol{u},0}[L, L-1] = \boldsymbol{\Sigma}_{\boldsymbol{u},0}[L-1, L] = \boldsymbol{\Sigma}_{\boldsymbol{u},0}[L, L] \\ &= 1 - \frac{1}{3}\rho^{(1)}(0)^2\rho^{(2)}(0)^{-1} \end{aligned} \tag{4.9}$$

and

$$\begin{aligned} &\boldsymbol{\Sigma}_{\boldsymbol{u},2}[L-1, L-1] = \boldsymbol{\Sigma}_{\boldsymbol{u},2}[L, L-1] = \boldsymbol{\Sigma}_{\boldsymbol{u},2}[L-1, L] = \boldsymbol{\Sigma}_{\boldsymbol{u},2}[L, L] \\ &= -\frac{1}{6}\rho^{(1)}(0) + \frac{5}{18}\rho^{(1)}(0)^2\rho^{(2)}(0)^{-2}\rho^{(3)}(0). \end{aligned}$$

Proof. Since the proof is long and computationally heavy, we include it in Appendix C.3. ☐

From Lemma 4.3.5, we see that the elements of $\mathbf{\Sigma}_{\boldsymbol{u}}(r)$ are at least two times continuously differentiable in r at 0. Intuitively, one may also expect that the diagonal matrix $\mathbf{\Lambda}_{\boldsymbol{u}}(r)$ of its ordered eigenvalues and the corresponding matrix $\boldsymbol{P}_{\boldsymbol{u}}(r)$ consisting of its eigenvectors to have the similar properties. In particular, when $\rho(x)$, $x \geq 0$ is real analytic on a neighborhood of 0, we can simply follow the proof of Lemma 4.3.5 to show that $\mathbf{\Sigma}_{\boldsymbol{u}}(r)$ is also real analytic on that neighborhood. Then by the first theorem in [22], both the eigenvalues and eigenvectors of $\mathbf{\Sigma}_{\boldsymbol{u}}(r)$ can be parameterized real analytically on a neighborhood of 0. As for the ordering of these eigenvalues, by noting the fact that zeros of a real analytic function indexed by $\mathbb{R}$ are isolated (see Corollary 1.2.5 in [21]), we can also show that $\mathbf{\Lambda}_{\boldsymbol{u}}(r)$ and $\boldsymbol{P}_{\boldsymbol{u}}(r)$ can be both real analytic on a neighborhood of 0.

However, the above result relies heavily on the assumption that ρ is real analytic, and is not guaranteed even when ρ is infinitely differentiable (see more explanation in [20]). Thus, the following condition can be regarded as a generalization of real analyticity of ρ on a neighborhood of 0, such that both the ordered eigenvalues and the corresponding eigenvectors of $\mathbf{\Sigma}_{\boldsymbol{u}}(r)$ can change smoothly enough under perturbation.

Definition 4.3.6. (Perturbation Condition) Let X be qualified, and let $\mathbf{\Sigma}(\boldsymbol{t})$, $\boldsymbol{P}(\boldsymbol{t})$ and $\mathbf{\Lambda}(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^N \setminus \{\mathbf{0}_N\}$ be the matrices as defined above. Then X is said to be **qualified under perturbation** if there exists a version of $\boldsymbol{P}(\boldsymbol{t})$ and $\mathbf{\Lambda}(\boldsymbol{t})$, such that for any direction $\boldsymbol{u} = (u_1, \ldots, u_N)^T \in \mathbb{S}^{N-1}$,

1. both $\boldsymbol{P}_{\boldsymbol{u}}(0) := \lim_{r\downarrow 0} \boldsymbol{P}_{\boldsymbol{u}}(r)$ and $\mathbf{\Lambda}_{\boldsymbol{u}}(0) := \lim_{r\downarrow 0} \mathbf{\Lambda}_{\boldsymbol{u}}(r)$ exist;
2. there exists a constant $\delta_{pc} \in (0, \delta_\rho]$ such that $\boldsymbol{P}_{\boldsymbol{u}}(r)$ and $\mathbf{\Lambda}_{\boldsymbol{u}}(r)$ are both continuous on $r \in [0, \delta_{pc}]$;
3. $\boldsymbol{P}_{\boldsymbol{u}}(r) = \boldsymbol{P}_{\boldsymbol{u},0} + \boldsymbol{P}_{\boldsymbol{u},1} r + o(r)$ as $r \to 0$ for some $\boldsymbol{P}_{\boldsymbol{u},0}$, $\boldsymbol{P}_{\boldsymbol{u},1} \in \mathbb{R}^{L\times L}$;
4. $\mathbf{\Lambda}_{\boldsymbol{u}}(r) = \mathbf{\Lambda}_{\boldsymbol{u},0} + \mathbf{\Lambda}_{\boldsymbol{u},1} r + \mathbf{\Lambda}_{\boldsymbol{u},2} r^2 + o(r^2)$ as $r \to 0$, where $\mathbf{\Lambda}_{\boldsymbol{u},j}$, $j = 0, 1, 2$ are all $L \times L$ real-valued diagonal matrices. More specifically, this is equivalent to
$$\lambda_{\boldsymbol{u},i}(r) = \lambda_{\boldsymbol{u},i,0} + \lambda_{\boldsymbol{u},i,1} r + \lambda_{\boldsymbol{u},i,2} r^2 + o(r^2),$$
where $\lambda_{\boldsymbol{u},i,j} \in \mathbb{R}$ for $i = 1, \ldots, L$ and $j = 0, 1, 2$, such that $\mathbf{\Lambda}_{\boldsymbol{u},0} := \operatorname{diag}(\lambda_{\boldsymbol{u},i,0}, i = 1, \ldots, L)$, $\mathbf{\Lambda}_{\boldsymbol{u},1} := \operatorname{diag}(\lambda_{\boldsymbol{u},i,1}, i = 1, \ldots, L)$, and $\mathbf{\Lambda}_{\boldsymbol{u},2} := \operatorname{diag}(\lambda_{\boldsymbol{u},i,2}, i = 1, \ldots, L)$.

In the following sections, if X is qualified under perturbation, then $\boldsymbol{P}(\boldsymbol{t})$ and $\boldsymbol{\Lambda}(\boldsymbol{t})$ are selected to be a version satisfying all the four conditions in Definition 4.3.6 by default.

Remark 4.3.7. Let X be qualified under perturbation and $\boldsymbol{u} \in \mathbb{S}^{N-1}$. Then by Definition 4.3.6, $\boldsymbol{A}_{\boldsymbol{u}}(r) := \boldsymbol{P}_{\boldsymbol{u}}(r)\boldsymbol{\Lambda}_{\boldsymbol{u}}^{1/2}(r)$, $r \geq 0$ is also continuous on $r \in [0, \delta_{pc}]$ and there exist $\boldsymbol{A}_{\boldsymbol{u},0}, \boldsymbol{A}_{\boldsymbol{u},1/2}, \boldsymbol{A}_{\boldsymbol{u},1} \in \mathbb{R}^{L\times L}$ such that ,

$$\boldsymbol{A}_{\boldsymbol{u}}(r) = \boldsymbol{A}_{\boldsymbol{u},0} + \boldsymbol{A}_{\boldsymbol{u},1/2}r^{1/2} + \boldsymbol{A}_{\boldsymbol{u},1}r + o(r)$$

as $r \to 0$. In particular, we have $\boldsymbol{A}_{\boldsymbol{u},0} = \boldsymbol{A}_{\boldsymbol{u}}(0)$.

In addition, by the continuities of $\boldsymbol{\Sigma}_{\boldsymbol{u}}(r)$, $\boldsymbol{A}_{\boldsymbol{u}}(r)$, $\boldsymbol{P}_{\boldsymbol{u}}(r)$ and $\boldsymbol{\Lambda}_{\boldsymbol{u}}(r)$ on $r \in [0, \delta_{pc}]$, it is easy to see

(i) $\boldsymbol{P}_{\boldsymbol{u},0}$ is orthogonal;

(ii) $\boldsymbol{\Sigma}_{\boldsymbol{u},0}$ has the eigen-decomposition

$$\boldsymbol{\Sigma}_{\boldsymbol{u},0} = \boldsymbol{P}_{\boldsymbol{u},0}\boldsymbol{\Lambda}_{\boldsymbol{u},0}\boldsymbol{P}_{\boldsymbol{u},0}^T;$$

(iii) $\boldsymbol{A}_{\boldsymbol{u},0} = \boldsymbol{P}_{\boldsymbol{u},0}\boldsymbol{\Lambda}_{\boldsymbol{u},0}^{1/2}$.

4.4 Covariance Structure

4.4.1 General Covariance Structure

The following lemma collects some useful results from Lemma 4.3.5 and Definition 4.3.6.

Lemma 4.4.1. *Let X be qualified under perturbation and $\boldsymbol{u} \in \mathbb{S}^{N-1}$. Then for any $1 \leq i \leq L$, we have*

(i) $(\boldsymbol{P}_{\boldsymbol{u},0}^{(i)})^T\boldsymbol{P}_{\boldsymbol{u},1}^{(i)} = 0$*;*

(ii) $\lambda_{\boldsymbol{u},i,1} = 0$*, i.e.,* $\boldsymbol{\Lambda}_{\boldsymbol{u},1} = \boldsymbol{0}_{L\times L}$*;*

(iii) if $\lambda_{\boldsymbol{u},i,0} = 0$*, then* $\lambda_{\boldsymbol{u},i,2} = (\boldsymbol{P}_{\boldsymbol{u},0}^{(i)})^T\boldsymbol{\Sigma}_{\boldsymbol{u},2}\boldsymbol{P}_{\boldsymbol{u},0}^{(i)}$*;*

(iv) $\boldsymbol{\Sigma}_{\boldsymbol{u},0}\boldsymbol{P}_{\boldsymbol{u},1}^{(i)} = \lambda_{\boldsymbol{u},i,0}\boldsymbol{P}_{\boldsymbol{u},1}^{(i)}$.

Proof. By Definition 4.3.6 and the orthogonality of $\boldsymbol{P}_{\boldsymbol{u}}(r)$ for any $r \geq 0$, we have

$$\begin{aligned}\boldsymbol{I}_L &= \boldsymbol{P}_{\boldsymbol{u}}(r)^T \boldsymbol{P}_{\boldsymbol{u}}(r) \\ &= \left(\boldsymbol{P}_{\boldsymbol{u},0} + \boldsymbol{P}_{\boldsymbol{u},1} r + o(r)\right)^T \left(\boldsymbol{P}_{\boldsymbol{u},0} + \boldsymbol{P}_{\boldsymbol{u},1} r + o(r)\right) \\ &= \boldsymbol{I}_L + \left(\boldsymbol{P}_{\boldsymbol{u},0}^T \boldsymbol{P}_{\boldsymbol{u},1} + \boldsymbol{P}_{\boldsymbol{u},1}^T \boldsymbol{P}_{\boldsymbol{u},0}\right) r + o(r),\end{aligned}$$

which implies

$$\boldsymbol{P}_{\boldsymbol{u},0}^T \boldsymbol{P}_{\boldsymbol{u},1} + \boldsymbol{P}_{\boldsymbol{u},1}^T \boldsymbol{P}_{\boldsymbol{u},0} = \boldsymbol{0}_{L\times L}.$$

Thus, for any $1 \leq i \leq L$,

$$(\boldsymbol{P}_{\boldsymbol{u},0}^{(i)})^T \boldsymbol{P}_{\boldsymbol{u},1}^{(i)} = \frac{1}{2}\left(\boldsymbol{P}_{\boldsymbol{u},0}^T \boldsymbol{P}_{\boldsymbol{u},1} + \boldsymbol{P}_{\boldsymbol{u},1}^T \boldsymbol{P}_{\boldsymbol{u},0}\right)[i,i] = 0.$$

For (ii), note that

$$\begin{aligned}\boldsymbol{\Sigma}_{\boldsymbol{u}}(r)\boldsymbol{P}_{\boldsymbol{u}}(r)^{(i)} &= \left(\boldsymbol{\Sigma}_{\boldsymbol{u},0} + \boldsymbol{\Sigma}_{\boldsymbol{u},2} r^2 + o(r^2)\right)\left(\boldsymbol{P}_{\boldsymbol{u},0}^{(i)} + \boldsymbol{P}_{\boldsymbol{u},1}^{(i)} r + o(r)\right) \\ &= \boldsymbol{\Sigma}_{\boldsymbol{u},0}\boldsymbol{P}_{\boldsymbol{u},0}^{(i)} + \left(\boldsymbol{\Sigma}_{\boldsymbol{u},0}\boldsymbol{P}_{\boldsymbol{u},1}^{(i)}\right) r + O\left(r^2\right)\end{aligned}$$

and

$$\begin{aligned}\lambda_{\boldsymbol{u},i}(r)\boldsymbol{P}_{\boldsymbol{u}}(r)^{(i)} &= \left(\lambda_{\boldsymbol{u},i,0} + \lambda_{\boldsymbol{u},i,1} r + \lambda_{\boldsymbol{u},i,2} r^2 + o(r^2)\right)\left(\boldsymbol{P}_{\boldsymbol{u},0}^{(i)} + \boldsymbol{P}_{\boldsymbol{u},1}^{(i)} r + o(r)\right) \\ &= \lambda_{\boldsymbol{u},i,0}\boldsymbol{P}_{\boldsymbol{u},0}^{(i)} + \left(\lambda_{\boldsymbol{u},i,0}\boldsymbol{P}_{\boldsymbol{u},1}^{(i)} + \lambda_{\boldsymbol{u},i,1}\boldsymbol{P}_{\boldsymbol{u},0}^{(i)}\right) r + O(r^2).\end{aligned} \tag{4.10}$$

Then by $\boldsymbol{\Sigma}_{\boldsymbol{u}}(r)\boldsymbol{P}_{\boldsymbol{u}}(r)^{(i)} = \lambda_{\boldsymbol{u},i}(r)\boldsymbol{P}_{\boldsymbol{u}}(r)^{(i)}$, we get

$$\lambda_{\boldsymbol{u},i,0}\boldsymbol{P}_{\boldsymbol{u},1}^{(i)} + \lambda_{\boldsymbol{u},i,1}\boldsymbol{P}_{\boldsymbol{u},0}^{(i)} = \boldsymbol{\Sigma}_{\boldsymbol{u},0}\boldsymbol{P}_{\boldsymbol{u},1}^{(i)}. \tag{4.11}$$

By left-multiplying $(\boldsymbol{P}_{\boldsymbol{u},0}^{(i)})^T$ on the both sides of (4.11), (i) of this lemma, (i) and (ii) in Remark 4.3.7, and the symmetry of $\boldsymbol{\Sigma}_{\boldsymbol{u},0}$, we have for any $1 \leq i \leq L$,

$$\lambda_{\boldsymbol{u},i,1} = \left(\boldsymbol{P}_{\boldsymbol{u},0}^{(i)}\right)^T \boldsymbol{\Sigma}_{\boldsymbol{u},0}\boldsymbol{P}_{\boldsymbol{u},1}^{(i)} = \left(\boldsymbol{P}_{\boldsymbol{u},1}^{(i)}\right)^T \boldsymbol{\Sigma}_{\boldsymbol{u},0}\boldsymbol{P}_{\boldsymbol{u},0}^{(i)} = \lambda_{\boldsymbol{u},i,0}\left(\boldsymbol{P}_{\boldsymbol{u},1}^{(i)}\right)^T \boldsymbol{P}_{\boldsymbol{u},0}^{(i)} = 0.$$

For (iii), by (ii) and $\lambda_{\boldsymbol{u},i,0} = 0$, Equation (4.10) becomes

$$\begin{aligned}\lambda_{\boldsymbol{u},i}(r)\boldsymbol{P}_{\boldsymbol{u}}(r)^{(i)} &= \left(\lambda_{\boldsymbol{u},i,2} r^2 + o(r^2)\right)\left(\boldsymbol{P}_{\boldsymbol{u},0}^{(i)} + \boldsymbol{P}_{\boldsymbol{u},1}^{(i)} r + o(r)\right) \\ &= \lambda_{\boldsymbol{u},i,2}\boldsymbol{P}_{\boldsymbol{u},0}^{(i)} r^2 + o(r^2).\end{aligned}$$

Similarly, by $\boldsymbol{\Sigma}_{\boldsymbol{u}}(r)\boldsymbol{P}_{\boldsymbol{u}}(r)^{(i)} = \lambda_{\boldsymbol{u},i}(r)\boldsymbol{P}_{\boldsymbol{u}}(r)^{(i)}$, we can get

$$\boldsymbol{\Sigma}_{\boldsymbol{u},2}\boldsymbol{P}_{\boldsymbol{u},0}^{(i)} = \lambda_{\boldsymbol{u},i,2}\boldsymbol{P}_{\boldsymbol{u},0}^{(i)}. \tag{4.12}$$

Then by left-multiplying $(\boldsymbol{P}_{\boldsymbol{u},0}^{(i)})^T$ on the both sides of (4.12), we have for any $1 \le i \le L$,

$$\lambda_{\boldsymbol{u},i,2} = \left(\boldsymbol{P}_{\boldsymbol{u},0}^{(i)}\right)^T \boldsymbol{\Sigma}_{\boldsymbol{u},2}\boldsymbol{P}_{\boldsymbol{u},0}^{(i)}.$$

Finally, taking (ii) into Equation (4.11) yields (iv). □

Remark 4.4.2. Let X be qualified under perturbation and $\boldsymbol{u} \in \mathbb{S}^{N-1}$. By (ii) of Lemma 4.4.1, it is easy to see the matrix $\boldsymbol{A}_{\boldsymbol{u},1/2}$ in (iii) of Remark 4.3.7 is $\mathbf{0}_{L\times L}$, and then we have

$$\boldsymbol{A}_{\boldsymbol{u}}(r) = \boldsymbol{A}_{\boldsymbol{u},0} + \boldsymbol{A}_{\boldsymbol{u},1}r + o(r)$$

as $r \to 0$. In addition, by Definition 4.3.6 and (ii) of Lemma 4.4.1, we have

$$\lambda_{\boldsymbol{u},i}^{1/2}(r) = \begin{cases} \lambda_{\boldsymbol{u},i,0}^{1/2} + \dfrac{\lambda_{\boldsymbol{u},i,2}}{2(\lambda_{\boldsymbol{u},i,0})^{1/2}}r^2 + o\left(r^2\right), & \text{for any } 1 \le i \le \operatorname{Rank}(\boldsymbol{\Sigma}_{\boldsymbol{u},0}); \\ \lambda_{\boldsymbol{u},i,2}^{1/2}r + o(r), & \text{otherwise.} \end{cases}$$

Then by Definition 4.3.6, Remark 4.4.2, and $\boldsymbol{A}_{\boldsymbol{u}}(r) = \boldsymbol{P}_{\boldsymbol{u}}(r)\boldsymbol{\Lambda}_{\boldsymbol{u}}^{1/2}(r)$, we have

$$\boldsymbol{A}_{\boldsymbol{u},1}^{(i)} = \begin{cases} \lambda_{\boldsymbol{u},i,0}^{1/2}\boldsymbol{P}_{\boldsymbol{u},1}^{(i)}, & \text{for any } 1 \le i \le \operatorname{Rank}(\boldsymbol{\Sigma}_{\boldsymbol{u},0}); \\ \lambda_{\boldsymbol{u},i,2}^{1/2}\boldsymbol{P}_{\boldsymbol{u},0}^{(i)}, & \text{otherwise.} \end{cases} \tag{4.13}$$

For any $1 \le i \le j \le N$, $1 \le k \le N$ and $\boldsymbol{u} = (u_1, \dots, u_N)^T \in \mathbb{S}^{N-1}$, define $\boldsymbol{H}(\boldsymbol{u}) = (h_{k,\ell}(\boldsymbol{u})) \in \mathbb{R}^{N\times L}$ by

$$h_{k,i+j(j-1)/2}(\boldsymbol{u}) = \delta_{j,k}u_i + (1-\delta_{j,k})\delta_{i,k}u_j \text{ and } h_{k,L-1}(\boldsymbol{u}) = h_{k,L}(\boldsymbol{u}) = 0. \tag{4.14}$$

For example, when $N = 3$, we have

$$\boldsymbol{H}(\boldsymbol{u}) = \begin{pmatrix} u_1 & u_2 & 0 & u_3 & 0 & 0 & 0 & 0 \\ 0 & u_1 & u_2 & 0 & u_3 & 0 & 0 & 0 \\ 0 & 0 & 0 & u_1 & u_2 & u_3 & 0 & 0 \end{pmatrix}.$$

The following lemma collects some useful results about $\boldsymbol{H}(\boldsymbol{u})$.

Lemma 4.4.3. *Let X be qualified under perturbation and $\boldsymbol{u} \in \mathbb{S}^{N-1}$. Then*

(i) if $\boldsymbol{M} \in \mathbb{R}^{N\times N}$ is the N-th order matriculation of a vector $\boldsymbol{a} \in \mathbb{R}^n$ ($n \geq N(N+1)/2$), then

$$\boldsymbol{H}(\boldsymbol{u})\boldsymbol{a} = \boldsymbol{M}\boldsymbol{u};$$

(ii) $\boldsymbol{\Sigma}_{\boldsymbol{u},0}\boldsymbol{H}^T(\boldsymbol{u}) = \mathbf{0}_{L\times N}$;

(iii) $\boldsymbol{H}(\boldsymbol{u})\boldsymbol{A}_{\boldsymbol{u}}(r)^{(i)} = \mathbf{0}_N + o(r)$ as $r \to 0$ for any $1 \leq i \leq \mathrm{Rank}\,(\boldsymbol{\Sigma}_{\boldsymbol{u},0})$.

Proof. One can directly check (i) by (4.14). As for (ii), one can note that the expressions in (4.6) and (4.8) are both symmetric in i_1 and j_1. As a result, for any $1 \leq i \leq L$, the elements in $\mathrm{Matri}\,\left((\boldsymbol{\Sigma}_{\boldsymbol{u},0})^{(i)}\right)$ for which the row number i_1 is greater than the column number j_1 will also be described by these expressions. Thus, for any $1 \leq i \leq L$,

$$\mathrm{Matri}\,\left((\boldsymbol{\Sigma}_{\boldsymbol{u},0})^{(i)}\right)\boldsymbol{u} = \mathbf{0}_N.$$

Then by (i) of this lemma and $\boldsymbol{\Sigma}_{\boldsymbol{u},0} = \boldsymbol{\Sigma}_{\boldsymbol{u},0}^T$,

$$(\boldsymbol{\Sigma}_{\boldsymbol{u},0})_{(i)}\boldsymbol{H}^T(\boldsymbol{u}) = \left(\boldsymbol{H}(\boldsymbol{u})(\boldsymbol{\Sigma}_{\boldsymbol{u},0})^{(i)}\right)^T = \left(\mathrm{Matri}\,\left((\boldsymbol{\Sigma}_{\boldsymbol{u},0})^{(i)}\right)\boldsymbol{u}\right) = \mathbf{0}_N^T.$$

As for (iii), by (ii) of this lemma, we have

$$\boldsymbol{H}(\boldsymbol{u})\boldsymbol{A}_{\boldsymbol{u},0} = \mathbf{0}_{N\times L}. \tag{4.15}$$

Then by Remark 4.4.2, it suffices to show that for any $1 \leq i \leq \mathrm{Rank}\,(\boldsymbol{\Sigma}_{\boldsymbol{u},0})$,

$$\boldsymbol{H}(\boldsymbol{u})\boldsymbol{A}_{\boldsymbol{u},1}^{(i)} = \mathbf{0}_N. \tag{4.16}$$

Note that $\lambda_{\boldsymbol{u},i,0} \neq 0$ for any $1 \leq i \leq \mathrm{Rank}\,(\boldsymbol{\Sigma}_{\boldsymbol{u},0})$. By (iii) in Remark 4.3.7 and (4.15), we have for any $1 \leq i \leq \mathrm{Rank}\,(\boldsymbol{\Sigma}_{\boldsymbol{u},0})$,

$$\boldsymbol{H}(\boldsymbol{u})\boldsymbol{P}_{\boldsymbol{u},0}^{(i)} = \lambda_{\boldsymbol{u},i,0}^{-1/2}\boldsymbol{H}(\boldsymbol{u})\boldsymbol{A}_{\boldsymbol{u},0}^{(i)} = \mathbf{0}_N. \tag{4.17}$$

By (4.13), we have

$$\boldsymbol{A}_{\boldsymbol{u},1}^{(i)} = \lambda_{\boldsymbol{u},i,0}^{1/2}\boldsymbol{P}_{\boldsymbol{u},1}^{(i)},$$

which, together with (iv) of Lemma 4.4.1, implies for any $1 \leq i \leq \mathrm{Rank}\,(\boldsymbol{\Sigma}_{\boldsymbol{u},0})$,

$$\boldsymbol{A}_{\boldsymbol{u},1}^{(i)} \in \mathrm{span}\,\left\{\boldsymbol{P}_{\boldsymbol{u},0}^{(1)}, \ldots, \boldsymbol{P}_{\boldsymbol{u},0}^{(\mathrm{Rank}\,(\boldsymbol{\Sigma}_{\boldsymbol{u},0}))}\right\}.$$

Then by (4.17), (4.16) is immediate, and hence completes the proof. □

4.4.2 Properties of the Covariance Matrix along a Coordinate Axis

In the last section, we have explored some properties of $\mathbf{\Sigma}_{\boldsymbol{u}}(r)$ for any $\boldsymbol{u} \in \mathbb{S}^{N-1}$. Let $\boldsymbol{u}_0 := (0,\dots,0,1)^T \in \mathbb{R}^L$ which is the direction of the last coordinate axis. By Lemma 4.3.5, $\mathbf{\Sigma}_{\boldsymbol{u}_0}(0)$ has a simple form. This would be helpful in solving problems that depend on $\mathbf{\Sigma}_{\boldsymbol{u}}(0)$ but are independent of the choice of $\boldsymbol{u} \in \mathbb{S}^{N-1}$. In this section, we focus on the properties of the covariance matrix $\mathbf{\Sigma}_{\boldsymbol{u}_0}(r)$, $r \geq 0$. For conciseness, we will drop $\boldsymbol{u}_0$ from subscripts.

By Lemma 4.3.5, it is easy to see the form of $\mathbf{\Sigma}_0$ can be very simple after swapping some of its rows and the corresponding columns. For example, for $N = 4$ and $\rho(x) = \exp(-x)$, $x \geq 0$, i.e., the covariance function $R(\boldsymbol{t}) = \rho(\|\boldsymbol{t}\|^2) = \exp(-\|\boldsymbol{t}\|^2)$, $\boldsymbol{t} \in \mathbb{R}^4$, we have

$$\mathbf{\Sigma}_0 = \begin{pmatrix} \frac{32}{3} & 0 & \frac{8}{3} & 0 & 0 & \frac{8}{3} & 0 & 0 & 0 & 0 & -\frac{4}{3} & -\frac{4}{3} \\ 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{8}{3} & 0 & \frac{32}{3} & 0 & 0 & \frac{8}{3} & 0 & 0 & 0 & 0 & -\frac{4}{3} & -\frac{4}{3} \\ 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{8}{3} & 0 & \frac{8}{3} & 0 & 0 & \frac{32}{3} & 0 & 0 & 0 & 0 & -\frac{4}{3} & -\frac{4}{3} \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{4}{3} & 0 & -\frac{4}{3} & 0 & 0 & -\frac{4}{3} & 0 & 0 & 0 & 0 & \frac{2}{3} & \frac{2}{3} \\ -\frac{4}{3} & 0 & -\frac{4}{3} & 0 & 0 & -\frac{4}{3} & 0 & 0 & 0 & 0 & \frac{2}{3} & \frac{2}{3} \end{pmatrix}.$$

After swapping some rows and the corresponding columns, it is turned into

$$\boldsymbol{\Sigma}_0' = \begin{pmatrix} \frac{32}{3} & \frac{8}{3} & \frac{8}{3} & -\frac{4}{3} & -\frac{4}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{8}{3} & \frac{32}{3} & \frac{8}{3} & -\frac{4}{3} & -\frac{4}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ \frac{8}{3} & \frac{8}{3} & \frac{32}{3} & -\frac{4}{3} & -\frac{4}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{4}{3} & -\frac{4}{3} & -\frac{4}{3} & \frac{2}{3} & \frac{2}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ -\frac{4}{3} & -\frac{4}{3} & -\frac{4}{3} & \frac{2}{3} & \frac{2}{3} & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{pmatrix}.$$

Indeed, one can easily check that $\boldsymbol{\Sigma}_0'$ is the limit of the covariance matrix of

$$(X_{11}(\boldsymbol{u}_0 r), X_{22}(\boldsymbol{u}_0 r), X_{33}(\boldsymbol{u}_0 r), X(\boldsymbol{u}_0 r), X(\mathbf{0}), X_{12}(\boldsymbol{u}_0 r), X_{13}(\boldsymbol{u}_0 r), X_{23}(\boldsymbol{u}_0 r),$$
$$X_{14}(\boldsymbol{u}_0 r), X_{24}(\boldsymbol{u}_0 r), X_{34}(\boldsymbol{u}_0 r), X_{44}(\boldsymbol{u}_0 r) | \triangledown X(\boldsymbol{u}_0 r) = \triangledown X(\mathbf{0}) = \mathbf{0}_N)$$

as $r \to 0$. In general, we can rearrange the elements of the random vector

$$(\triangledown^2 X(\boldsymbol{t}), X(\boldsymbol{t}), X(\mathbf{0}) | \triangledown X(\boldsymbol{t}) = \triangledown X(\mathbf{0}) = \mathbf{0}_N)$$

such that the limiting covariance matrix $\boldsymbol{\Sigma}_0'$ of the random vector after the rearrangement has the form

$$\boldsymbol{\Sigma}_0' = \begin{pmatrix} \boldsymbol{B}_0 & \boldsymbol{B}_2 & \mathbf{0} & \mathbf{0} \\ \boldsymbol{B}_2^T & \boldsymbol{B}_3 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \boldsymbol{B}_1 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0}_{N\times N} \end{pmatrix}, \tag{4.18}$$

where $\boldsymbol{B}_0 \in \mathbb{R}^{(N-1)\times(N-1)}$, $\boldsymbol{B}_1 \in \mathbb{R}^{(N-1)(N-2)/2\times(N-1)(N-2)/2}$, $\boldsymbol{B}_2 \in \mathbb{R}^{(N-1)\times 2}$, and $\boldsymbol{B}_3 \in \mathbb{R}^{2\times 2}$ satisfy that

$$\boldsymbol{B}_0[i_0, j_0] = 4\left(\frac{2}{3} + 2\delta_{i_0,j_0}\right)\rho^{(2)}(0) \text{ for any } 1 \le i_0, j_0 \le N-1,$$

$$\boldsymbol{B}_1[i_1, j_1] = 4\rho^{(2)}(0)\delta_{i_1,j_1} \text{ for any } 1 \le i_1, j_1 \le (N-1)(N-2)/2,$$

$$\boldsymbol{B}_2[i_2, j_2] = \frac{4}{3}\rho^{(1)}(0) \text{ for any } 1 \le i_2 \le N-1 \text{ and } 1 \le j_2 \le 2,$$

and

$$\boldsymbol{B}_3[i_3, j_3] = 1 - \frac{\rho^{(1)}(0)^2}{3\rho^{(2)}(0)} \text{ for any } 1 \le i_3, j_3 \le 2.$$

Indeed, $\boldsymbol{B}_0$ corresponds to the elements in (4.6) with $1 \le i_1, j_1, i_2, j_2 \le N-1$, $i_1 = j_1$ and $i_2 = j_2$; $\boldsymbol{B}_1$ corresponds to the elements in (4.6) with $1 \le i_1, j_1, i_2, j_2 \le N-1$, $i_1 < j_1$ and $i_2 < j_2$; $\boldsymbol{B}_2$ corresponds to the elements in (4.8) with $1 \le i_1 = j_1 \le N-1$; $\boldsymbol{B}_3$ corresponds to the elements in (4.9). It is noticeable that $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_0'$ share the same eigenvalues (but different eigenspaces). The following lemma introduces some properties of the eigenvalues and eigenvectors of $\boldsymbol{\Sigma}_0$.

Lemma 4.4.4. *Let*

$$\boldsymbol{W} := \begin{pmatrix} a & b \\ c & d \end{pmatrix} \in \mathbb{R}^{2\times 2},$$

where

$$a = \frac{1}{3}(32 + 8(N-2))\rho^{(2)}(0), \quad b = \frac{8}{3}\rho^{(1)}(0),$$
$$c = \frac{4}{3}(N-1)\rho^{(1)}(0) \text{ and } d = 2\left(1 - \frac{\rho^{(1)}(0)^2}{3\rho^{(2)}(0)}\right).$$

Then we have

(i) The eigenvalues

$$\lambda_+ = \frac{a+d+\sqrt{(a-d)^2+4bc}}{2} \text{ and } \lambda_- = \frac{a+d-\sqrt{(a-d)^2+4bc}}{2}$$

of $\boldsymbol{W}$ are also two different eigenvalues of $\boldsymbol{\Sigma}_0$, i.e., there exist integers $1 \le l < s \le L$ such that

$$\lambda_{l,0} = \lambda_+ \text{ and } \lambda_{s,0} = \lambda_-.$$

Moreover, we have

$$\lambda_{l,0} > 8\rho^{(2)}(0) \text{ and } \lambda_{s,0} \ne 0.$$

(ii) 0 is an eigenvalue of $\boldsymbol{\Sigma}_0$ with multiplicity $N+1$ and its eigenvector, $\boldsymbol{p}_0$, must satisfy

$$\boldsymbol{p}_0[i + j(j-1)/2] = 0 \text{ for any } 1 \le i \le j \le N-1$$

and

$$\boldsymbol{p}_0[L-1] + \boldsymbol{p}_0[L] = 0.$$

(iii) $4\rho^{(2)}(0)$ *is an eigenvalue of* $\boldsymbol{\Sigma}_0$*, and if* $\lambda_{s,0} \neq 4\rho^{(2)}(0)$*, then the multiplicity of* $4\rho^{(2)}(0)$ *as an eigenvalue of* $\boldsymbol{\Sigma}_0$ *is* $(N-1)(N-2)/2$*.*

(iv) $8\rho^{(2)}(0)$ *is an eigenvalue of* $\boldsymbol{\Sigma}_0$*, and if* $\lambda_{s,0} \neq 8\rho^{(2)}(0)$*, then the multiplicity of* $8\rho^{(2)}(0)$ *as an eigenvalue of* $\boldsymbol{\Sigma}_0$ *is* $N-2$*.*

(v) For any nonzero eigenvalue of $\boldsymbol{\Sigma}_0$*, its eigenvector,* $\boldsymbol{p}_{nz}$*, must satisfy*

$$\boldsymbol{p}_{nz}[L-1] = \boldsymbol{p}_{nz}[L] \tag{4.19}$$

and

$$\boldsymbol{p}_{nz}[i + N(N-1)/2] = 0 \text{ for any } 1 \leq i \leq N. \tag{4.20}$$

(vi) If $\lambda_{s,0} \notin \{4\rho^{(2)}(0), 8\rho^{(2)}(0)\}$*, then any eigenvector,* $\boldsymbol{p}_*$*, of* $4\rho^{(2)}(0)$ *or* $8\rho^{(2)}(0)$ *must satisfy*

$$\boldsymbol{p}_*[L-1] = \boldsymbol{p}_*[L] = 0.$$

(vii) If $\lambda_{s,0} \notin \{4\rho^{(2)}(0), 8\rho^{(2)}(0)\}$*, then for* $\lambda_{l,0}$ *and* $\lambda_{s,0}$ *as eigenvalues of* $\boldsymbol{\Sigma}_0$*, any eigenvector,* $\boldsymbol{p}$*, of them must have the form:*

$$\boldsymbol{p}[i + j(j-1)/2] = \delta_{i,j} x, \text{ for any } 1 \leq i \leq j \leq N-1,$$

$$\boldsymbol{p}[i + N(N-1)/2] = 0 \text{ for any } 1 \leq i \leq N, \text{ and } \boldsymbol{p}[L-1] = \boldsymbol{p}[L] = y,$$

where x *and* y *are both non-zero.*

(viii) There exists a constant $C > 0$ *such that*

$$\tilde{\lambda}_{s,0} < 4\tilde{\rho}^{(2)}(0),$$

where $\tilde{\lambda}_{s,0}$ *is the analog of* $\lambda_{s,0}$ *defined using the* ***rescaled*** *covariance function*

$$\tilde{\rho}(\|\boldsymbol{t}\|^2) := \rho(C\|\boldsymbol{t}\|^2) \text{ for any } \boldsymbol{t} \in \mathbb{R}^N \setminus \{\boldsymbol{0}_N\}.$$

Proof. In (4.18), denote

$$\widetilde{\boldsymbol{B}} := \begin{pmatrix} \boldsymbol{B}_0 & \boldsymbol{B}_2 \\ \boldsymbol{B}_2^T & \boldsymbol{B}_3 \end{pmatrix} \in \mathbb{R}^{(N+1)\times(N+1)}.$$

Since $\det(\widetilde{\boldsymbol{B}} - \lambda \boldsymbol{I}_{N+1}) = 0$ for some $\lambda \in \mathbb{R}$ implies $\det(\boldsymbol{\Sigma}_0' - \lambda \boldsymbol{I}_L) = 0$, for (i), it suffices to show that an eigenvalue of $\boldsymbol{W}$ must also be an eigenvalue of $\widetilde{\boldsymbol{B}}$. Let $\tilde{\boldsymbol{q}} := (\tilde{q}_1, \ldots, \tilde{q}_{N+1})^T \in \mathbb{R}^{N+1} \setminus \{\boldsymbol{0}_{N+1}\}$ and $\lambda \in \mathbb{R}$ satisfy

$$(\widetilde{\boldsymbol{B}} - \lambda \boldsymbol{I}_{N+1})\tilde{\boldsymbol{q}} = 0. \tag{4.21}$$

It is easy to see that (4.21) implies

$$\tilde{q}_1 = \cdots = \tilde{q}_{N-1} \text{ and } \tilde{q}_N = \tilde{q}_{N+1}.$$

Then (4.21) is equivalent to

$$(\boldsymbol{W} - \lambda \boldsymbol{I}_2)(\tilde{q}_1, \tilde{q}_N)^T = \mathbf{0}_2.$$

Thus, λ_+ and λ_- are both eigenvalues of $\boldsymbol{\Sigma}_0$. Since $bc > 0$ and $\rho^{(2)}(0) > 0$ (see Remark 4.3.4), we have

$$\lambda_{l,0} - \lambda_{s,0} \geq \sqrt{4bc} > 0$$

and

$$\lambda_{l,0} = \frac{a + d + \sqrt{(a-d)^2 + 4bc}}{2} \geq \frac{a + d + |a - d|}{2} \geq \max(a, d) > 8\rho^{(2)}(0).$$

Moreover, if $\lambda_{s,0} = 0$, then $ad = bc$. By some calculation, this implies

$$\frac{\rho^{(2)}(0)}{\rho^{(1)}(0)^2} \leq \frac{N}{N+2}.$$

However, by Lemma 4.2.1 and Condition (4) in Definition 4.3.2, we can get

$$\frac{\rho^{(2)}(0)}{\rho^{(1)}(0)^2} > \frac{N}{N+2},$$

which leads to a contradiction. Thus, $\lambda_{s,0} \neq 0$.

As for (ii), we can observe from (4.18) that

$$L - N - 2 \leq \text{Rank}\,(\boldsymbol{\Sigma}_0) = \text{Rank}\,(\boldsymbol{\Sigma}_0') \leq L - N - 1,$$

and the only uncertainty of $\text{Rank}\,(\boldsymbol{\Sigma}_0)$ comes from $\widetilde{\boldsymbol{B}}$, where the last two rows (and columns) are the same, and hence, one of them can be dropped in the subsequent discussion. Thus,

$$\text{Rank}\,(\boldsymbol{\Sigma}_0) = L - N - 1 \text{ if and only if } \widetilde{\boldsymbol{B}}[1:N, 1:N] \text{ is non-degenerate.}$$

By solving the equation $\widetilde{\boldsymbol{B}}[1:N, 1:N]\boldsymbol{q} = \mathbf{0}_N$ for $\boldsymbol{q} = (q_1, \ldots, q_N)^T \in \mathbb{R}^N$, we see that the non-degeneracy of $\widetilde{\boldsymbol{B}}[1:N, 1:N]$ is equivalent to the equation $\boldsymbol{W}(q_1, q_N)^T = \mathbf{0}_2$ not having any non-trivial solutions, i.e.,

$$\det\,(\boldsymbol{W}) = \lambda_{l,0}\lambda_{s,0} \neq 0,$$

which is obvious by (i). Thus, $\mathrm{Rank}\,(\boldsymbol{\Sigma}_0) = L - N - 1$, and then the multiplicity of 0 is $N + 1$. By (4.18) and solving the equation $\boldsymbol{\Sigma}_0 \boldsymbol{p}_0 = \boldsymbol{0}_L$ for $\boldsymbol{p}_0 \in \mathbb{R}^L$, we have

$$\boldsymbol{p}_0[i + j(j-1)/2] = 0 \text{ for any } 1 \le i \le j \le N - 1 \tag{4.22}$$

and

$$\boldsymbol{p}_0[L-1] + \boldsymbol{p}_0[L] = 0. \tag{4.23}$$

As for (iii), similarly, we can observe from (4.18) that

$$\frac{1}{2}(N-1)(N-2) - 1 \le \mathrm{Rank}\,(\boldsymbol{\Sigma}_0 - 4\rho^{(2)}(0)\boldsymbol{I}_L) \le \frac{1}{2}(N-1)(N-2),$$

and the only uncertainty of $\mathrm{Rank}\,(\boldsymbol{\Sigma}_0 - 4\rho^{(2)}(0)\boldsymbol{I}_L)$ comes from $\mathrm{Rank}\,(\widetilde{\boldsymbol{B}}[1:N, 1:N] - 4\rho^{(2)}(0)\boldsymbol{I}_N)$, i.e.,

$$\mathrm{Rank}\,(\boldsymbol{\Sigma}_0 - 4\rho^{(2)}(0)\boldsymbol{I}_L) = \tfrac{1}{2}(N-1)(N-2) \text{ if and only if } \widetilde{\boldsymbol{B}}[1:N, 1:N] - 4\rho^{(2)}(0)\boldsymbol{I}_N \text{ is non-degenerate.}$$

By solving the equation $(\widetilde{\boldsymbol{B}}[1:N, 1:N] - 4\rho^{(2)}(0)\boldsymbol{I}_N)\boldsymbol{q}' = \boldsymbol{0}_N$ for $\boldsymbol{q}' = (q_1', \ldots, q_N')^T \in \mathbb{R}^N$, we see that the non-degeneracy of $\widetilde{\boldsymbol{B}}[1:N, 1:N] - 4\rho^{(2)}(0)\boldsymbol{I}_N$ is equivalent to

$$\det\,(\boldsymbol{W} - 4\rho^{(2)}(0)\boldsymbol{I}_2) \neq 0.$$

Then by (i), this is equivalent to $\lambda_{s,0} \neq 4\rho^{(2)}(0)$ as we desired. Note that the proof for (iv) is only an analog of (iii) by replacing $4\rho^{(2)}(0)$ with $8\rho^{(2)}(0)$.

As for (v), note that $\boldsymbol{j} = (0, \ldots, 0, 1, -1)^T \in \mathbb{R}^L$ is in the eigenspace of 0. Thus, by the orthogonality of eigenspaces, any eigenvector, $\boldsymbol{p}_{nz}$, of a nonzero eigenvalue satisfies $\boldsymbol{p}_{nz}[L-1] = \boldsymbol{p}_{nz}[L]$ as stated in (4.19). Then by (4.19), (4.22), (4.23) and the orthogonality of eigenspaces, (4.20) is immediate.

As for (vi), let $\boldsymbol{p}_{k,\ell}$, $1 \le k < \ell \le N-1$ be $(N-1)(N-2)/2$ vectors in $\mathbb{R}^L$ such that

$$\boldsymbol{p}_{k,\ell}[i + j(j-1)/2] = \delta_{i,k}\delta_{j,\ell} \text{ for any } 1 \le i < j \le N-1,$$

$$\boldsymbol{p}_{k,\ell}[i + N(N-1)/2] = 0 \text{ for any } 1 \le i \le N,$$

and

$$\boldsymbol{p}_{k,\ell}[L-1] = \boldsymbol{p}_{k,\ell}[L] = 0. \tag{4.24}$$

Then by (iii) and (4.18), it is easy to check that $\boldsymbol{p}_{k,\ell}$, $1 \le k < \ell \le N-1$ are linearly independent and form a basis of the eigenspace of $4\rho^{(2)}(0)$. By the orthogonality of eigenspaces and (4.20), this implies for any eigenvalue of $\boldsymbol{\Sigma}_0$ not equal to $4\rho^{(2)}(0)$, its eigenvector, $\boldsymbol{p}'$, must satisfy

$$\boldsymbol{p}'[i + j(j-1)/2] = 0 \text{ for } 1 \leq i < j \leq N. \tag{4.25}$$

Let $\boldsymbol{p}''_k$, $1 \leq k \leq N-2$ be $N-2$ linearly independent vectors in $\mathbb{R}^L$ such that for any $1 \leq k \leq N-2$,

$$\sum_{i=1}^{N-1} \boldsymbol{p}''_k[i + i(i-1)/2] = 0, \quad \boldsymbol{p}''_k[N + N(N-1)/2] = 0, \tag{4.26}$$

$$\boldsymbol{p}''_k[i + j(j-1)/2] = 0 \text{ for any } 1 \leq i < j \leq N,$$

and

$$\boldsymbol{p}''_k[L-1] = \boldsymbol{p}''_k[L] = 0. \tag{4.27}$$

Then by (iv) and (4.18), it is easy to check that $\boldsymbol{p}''_k$, $1 \leq k \leq N-2$ form a basis of the eigenspace of $8\rho^{(2)}(0)$. Then combining (4.24), (4.27) and the orthogonality of eigenspaces yields (vi).

As for (vii), by the orthogonality of eigenspaces, (vi), (4.20), (4.25) and (4.26), we have for any nonzero eigenvalue of $\boldsymbol{\Sigma}_0$ not equal to $4\rho^{(2)}(0)$ or $8\rho^{(2)}(0)$, its eigenvector, $\boldsymbol{p}$ must have the form

$$\boldsymbol{p}[i + j(j-1)/2] = \delta_{i,j} x, \tag{4.28}$$

for a constant $x \in \mathbb{R}$ and any $1 \leq i \leq j \leq N-1$. By (4.19), (4.20) and (4.28), the only thing left is to show $xy \neq 0$. Let $\boldsymbol{p}_l, \boldsymbol{p}_s \in \mathbb{R}^L$ be eigenvectors of $\lambda_{l,0}$ and $\lambda_{s,0}$ respectively, such that

$$\boldsymbol{p}_l[i + j(j-1)/2] = \delta_{i,j} x_l, \text{ for any } 1 \leq i \leq j \leq N-1,$$

$$\boldsymbol{p}_s[i + j(j-1)/2] = \delta_{i,j} x_s, \text{ for any } 1 \leq i \leq j \leq N-1,$$

$$\boldsymbol{p}_l[L-1] = \boldsymbol{p}_l[L] = y_l \text{ and } \boldsymbol{p}_s[L-1] = \boldsymbol{p}_s[L] = y_s.$$

Then it suffices to show $x_l y_l x_s y_s \neq 0$. Suppose $x_l y_l x_s y_s = 0$. By the orthogonality of eigenspaces,

$$(N-1) x_l x_s + 2 y_l y_s = 0.$$

Then we must have

$$x_l x_s = y_l y_s = 0.$$

Without loss of generality, suppose $x_l = 0$. Then $y_l \neq 0$ since $\boldsymbol{p}_l$ is nonzero. By checking the first rows of the both sides of the equation $(\boldsymbol{\Sigma}_0 - \lambda_{l,0} \boldsymbol{I}_L)\boldsymbol{p}_l = \boldsymbol{0}_L$, (4.8) and (i), we have

$$\frac{8}{3}\rho^{(1)}(0) y_l = (\boldsymbol{\Sigma}_0[1, L-1] + \boldsymbol{\Sigma}_0[1, L]) y_l = 0.$$

This implies $\rho^{(1)}(0) = 0$, which leads to a contradiction with Remark 4.3.4. Therefore, $x_l y_l x_s y_s \neq 0$ as we desired.

As for (viii), note that for any $C > 0$

$$\tilde{\lambda}_{s,0} = \frac{\tilde{a} + \tilde{d} - \sqrt{(\tilde{a} - \tilde{d})^2 + 4\tilde{b}\tilde{c}}}{2},$$

where

$$\tilde{a} = \frac{1}{3}(32 + 8(N-2))C^2\rho^{(2)}(0), \quad \tilde{b} = \frac{8}{3}C\rho^{(1)}(0),$$
$$\tilde{c} = \frac{4}{3}(N-1)C\rho^{(1)}(0) \text{ and } \tilde{d} = 2\left(1 - \frac{\rho^{(1)}(0)^2}{3\rho^{(2)}(0)}\right).$$

Then we can define

$$k_1 := \tilde{a}C^{-2}, \quad k_2 := \tilde{b}C^{-1}, \quad k_3 := \tilde{c}C^{-1}, \text{ and } k_4 := \tilde{d},$$

and by Remark 4.3.4, we have $k_i \neq 0$ for $i = 1, 2, 3, 4$. In particular, we have

$$k_1 = \frac{1}{3}(32 + 8(N-2))\rho^{(2)}(0) > 8\rho^{(2)}(0).$$

Then by (i), the inequality $\tilde{\lambda}_{s,0} < 4\rho^{(2)}(0)C^2$ holds if and only if

$$k_1C^2 + k_4 - \sqrt{(k_1C^2 - k_4)^2 + 4k_2k_3C^2} < 8\rho^{(2)}(0)C^2.$$

Thus, it suffices to have

$$f(C) := ((k_1 - 8\rho^{(2)}(0))C^2 + k_4)^2 - (k_1C^2 - k_4)^2 - 4k_2k_3C^2 < 0.$$

Note that f is a polynomial of C with degree four and its coefficient of C^4 is

$$(k_1 - 8\rho^{(2)}(0))^2 - k_1^2 = -16\rho^{(2)}(0)k_1 + 64\rho^{(2)}(0)^2 = 16\rho^{(2)}(0)(4\rho^{(2)}(0) - k_1) < 0.$$

Thus, there exists a constant $C_0 > 0$ such that $f(C) < 0$ for any $C > C_0$. This implies $\tilde{\lambda}_{s,0} < 4\rho^{(2)}(0)C^2 = 4\tilde{\rho}^{(2)}(0)$ for any $C > C_0$, and hence proved. □

Remark 4.4.5. Recall that $L = N(N+1)/2+2$. By Lemma 4.4.4, the sum of multiplicities of 0, $4\rho^{(2)}(0)$, $8\rho^{(2)}(0)$, $\lambda_{s,0}$, and $\lambda_{t,0}$ is equal to L, which implies they are the only eigenvalues of $\mathbf{\Sigma}_0$. The condition in (vii) of Lemma 4.4.4, i.e.,

$$\lambda_{s,0} \notin \{4\rho^{(2)}(0), 8\rho^{(2)}(0)\},$$

ensures that these eigenvalues are distinct.

However, if the problem of interest is independent of the choice of C in the covariance function $R(\boldsymbol{t}) = \rho(C\|\boldsymbol{t}\|^2)$, $\boldsymbol{t} \in \mathbb{R}^N$, then by (viii) of Lemma 4.4.4, we can assume that this condition always holds, since it can be achieved by a suitable rescaling.

Lemma 4.4.6. *Let X be qualified under perturbation. Then*

(i) there exists an integer $L - N \le i \le L$ such that $\lambda_{i,2} > 0$;

(ii) there exists an integer $L - N \le i \le L$ such that $\lambda_{i,2} = 0$.

Proof. By (ii) of Lemma 4.4.1 and (ii) of Lemma 4.4.4, $\lambda_{i,0} = \lambda_{i,1} = 0$ for any $L - N \le i \le L$. Then by Definition 4.3.6,

$$\lambda_{i,2} = \lim_{r\to 0} \frac{\lambda(r) - \lambda_{i,0} - \lambda_{i,1} r}{r^2} = \lim_{r\to 0} \frac{\lambda(r)}{r^2} \ge 0.$$

Thus, for (i), it suffices to show $\lambda_{i,2} \neq 0$ for some $L - N \le i \le L$. Let $\boldsymbol{p} \in \mathbb{R}^L$ satisfy

$$\boldsymbol{p}[k] = \delta_{k,N+N(N-1)/2} \text{ for } k = 1, \dots, L.$$

Then $\boldsymbol{p}^T = \boldsymbol{H}(\boldsymbol{u}_0)_{(N)}$, and by (ii) of Lemma 4.4.3,

$$\boldsymbol{p}^T \boldsymbol{\Sigma}_0 \boldsymbol{p} = 0.$$

Recall that in Lemma 4.3.5 and Remark 4.3.4, $\alpha = \rho^{(1)}(0)^{-1}\rho^{(2)}(0)^2$ and $\beta = \rho^{(3)}(0)$. Then

$$\boldsymbol{p}^T \boldsymbol{\Sigma}_2 \boldsymbol{p} = \boldsymbol{\Sigma}_2[N + N(N-1)/2, N + N(N-1)/2] = 18\alpha - 30\beta \neq 0,$$

which implies

$$\boldsymbol{p}^T \boldsymbol{\Sigma}(r) \boldsymbol{p} = \Theta(r^2) \text{ as } r \to 0. \tag{4.29}$$

By $\boldsymbol{p}^T = \boldsymbol{H}(\boldsymbol{u}_0)_{(N)}$, (iii) of Lemma 4.4.3, and (ii) of Lemma 4.4.4, we can also get

$$\begin{aligned}
&\boldsymbol{p}^T \boldsymbol{\Sigma}(r) \boldsymbol{p} \\
&= \left(\boldsymbol{p}^T \left(\boldsymbol{A}(r)[1:L, 1:(L-N-1)]; \boldsymbol{A}(r)[1:L, (L-N):L]\right)\right) \\
&\quad \left(\boldsymbol{p}^T \left(\boldsymbol{A}(r)[1:L, 1:(L-N-1)]; \boldsymbol{A}(r)[1:L, (L-N):L]\right)\right)^T \\
&= \left(\boldsymbol{p}^T (\boldsymbol{A}(r)[1:L, (L-N):L])\right) \left(\boldsymbol{p}^T (\boldsymbol{A}(r)[1:L, (L-N):L])\right)^T + o(r^2) \\
&= \left(\boldsymbol{p}^T \boldsymbol{P}(r) \boldsymbol{\Lambda}^{1/2}(r)[1:L, (L-N):L]\right) \left(\boldsymbol{p}^T \boldsymbol{P}(r) \boldsymbol{\Lambda}^{1/2}(r)[1:L, (L-N):L]\right)^T + o(r^2) \\
&= \left(\boldsymbol{p}^T \boldsymbol{P}(r)\right) \operatorname{diag}\left(0, \dots, 0, \lambda_{L-N}(r), \dots, \lambda_L(r)\right) \left(\boldsymbol{p}^T \boldsymbol{P}(r)\right)^T + o(r^2) \\
&= \left(\boldsymbol{p}^T \boldsymbol{P}(r)\right) \operatorname{diag}\left(0, \dots, 0, \lambda_{L-N,2}, \dots, \lambda_{L,2}\right) \left(\boldsymbol{p}^T \boldsymbol{P}(r)\right)^T r^2 + o(r^2),
\end{aligned}$$

where the semicolons represent juxtaposition operations on matrices. If $\lambda_{i,2} = 0$ for all integer $L - N \leq i \leq L$, then from the above equation, we see $\boldsymbol{p}^T \boldsymbol{\Sigma}(r) \boldsymbol{p} = o(r^2)$, which contradicts with (4.29).

As for (ii), by Lemma 4.3.5, we see

$$\boldsymbol{\Sigma}_0^{(i+N(N-1)/2)} = \boldsymbol{0}_L \text{ for any } 1 \leq i \leq N,$$

which contributes $O(r^{2N})$ in $\det(\boldsymbol{\Sigma}(r))$, and

$$\boldsymbol{\Sigma}_0^{(L-1)} = \boldsymbol{\Sigma}_0^{(L)} \text{ and } \boldsymbol{\Sigma}_2^{(L-1)} = \boldsymbol{\Sigma}_2^{(L)},$$

which contribute $o(r^2)$ in $\det(\boldsymbol{\Sigma}(r))$. Since $\boldsymbol{\Sigma}(r) = \boldsymbol{\Sigma}_0 + \boldsymbol{\Sigma}_2 r^2 + o(r^2)$, we have

$$\det(\boldsymbol{\Sigma}(r)) = o\left(r^{2N+2}\right) \text{ as } r \to 0.$$

Thus, there must be an integer $L - N \leq i \leq L$ such that $\lambda_i(r) = o(r^2)$, which implies $\lambda_{i,2} = 0$ as desired. □

Lemma 4.4.7. *Let X be qualified under perturbation. Denote $\boldsymbol{j} := (0, \ldots, 0, 1, -1)^T \in \mathbb{R}^L$. Then*

$$\lambda_{i,2} > 0 \textit{ for any } L - N \leq i \leq L - 1, \textit{ and } \lambda_{L,2} = 0.$$

Moreover, $\boldsymbol{P}_0^{(L)}$ and $\boldsymbol{j}$ are linearly dependent.

Proof. Note that by (ii) of Lemma 4.4.6, there exists an integer $L - N \leq i_* \leq L$ such that $\lambda_{i_*,2} = 0$, and by (4.7),

$$\begin{aligned} &\boldsymbol{\Sigma}_2[(N(N-1)/2+1) : N(N+1)/2, (N(N-1)/2+1) : N(N+1)/2] \\ &= \operatorname{diag}(2\alpha - 6\beta, \ldots, 2\alpha - 6\beta, 18\alpha - 30\beta)_{N \times N}. \end{aligned}$$

Then by (iii) of Lemma 4.4.1 and (ii) of Lemma 4.4.4, we have

$$\begin{aligned} 0 &= \left(\boldsymbol{P}_0^{(i_*)}\right)^T \boldsymbol{\Sigma}_2 \boldsymbol{P}_0^{(i_*)} \\ &= (2\alpha - 6\beta) \sum_{i=1}^{N-1} \boldsymbol{P}_0[i + N(N-1)/2, i_*]^2 + (18\alpha - 30\beta) \boldsymbol{P}_0[N + N(N-1)/2, i_*]^2. \end{aligned} \tag{4.30}$$

By Remark 4.3.4, we have $18\alpha - 30\beta > 0$ and $\beta < 0$, which implies $2\alpha - 6\beta > 0$. Then by (4.30), we have

$$\boldsymbol{P}_0[i, i_*] = 0 \text{ for any } 1 \leq i \leq N(N+1)/2,$$

and by (ii) of Lemma 4.4.4,

$$\boldsymbol{P}_0[L-1, i_*] + \boldsymbol{P}_0[L, i_*] = 0.$$

Since $\boldsymbol{P}_0$ is non-degenerate, we have

$$\boldsymbol{P}_0^{(i_*)} \text{ and } \boldsymbol{j} \text{ are linearly dependent.}$$

Note that this property holds for any column $L-N \leq i \leq L$ satisfying $\lambda_{i,2} = 0$. However, by the non-degeneracy of $\boldsymbol{P}_0$, there can only be one column of $\boldsymbol{P}_0$ satisfying this property. This means that there exists a unique integer $L-N \leq i_* \leq L$ such that $\lambda_{i_*,2} = 0$. By the continuity of $\boldsymbol{P}(r)$ at $r = 0$ and (4.5), we have $i^* = L$. □

4.5 Asymptotic Behavior as $r \to 0$

4.5.1 Main Result 1

Let X be qualified under perturbation. Then for any $z \in \mathbb{R}$ and compact set $T \subset \mathbb{R}^N \setminus \{\mathbf{0}_N\}$ with $\lambda_{N-1}(\partial T) < \infty$,

$$X(\boldsymbol{t}) \text{ conditional on } \nabla X(\mathbf{0}) = \mathbf{0}_N \text{ and } X(\mathbf{0}) = z$$

is still a Gaussian random field on $\boldsymbol{t} \in T$ and satisfies all the conditions in Lemma 2.1.10 except that it may not be centered. By (A.6), Condition (2) in Definition 4.3.2 and Lemma 2.1.4, one can easily check that for any $z \in \mathbb{R}$, the mean function of this new random field is continuously twice differentiable. Then by Remark 2.1.11, the result in Lemma 2.1.10 still holds for this new random field.

Consider a set $T \subset \mathbb{R}^N \setminus \{\mathbf{0}_N\}$ such that $T \cup \{\mathbf{0}_N\}$ is compact. Since T is not compact, we cannot directly apply Lemma 2.1.10 to get the expected number, denoted by $n(T)$, of the critical points in T for this new Gaussian random field. However, if we take $T_m := T \setminus B(\mathbf{0}_N, 2^{-m})$ for $m \geq 1$, where $B(\boldsymbol{t}, r)$ is the N-dimensional open ball centered at $\boldsymbol{t}$ with radius r for any $\boldsymbol{t} \in \mathbb{R}^N$ and $r > 0$, then these T_m (if non-empty) are all compact and form an increasing sequence of sets. This means that we can get the expected number, denoted by $n(T_m)$, of the critical points in each T_m by Lemma 2.1.10, and then $n(T) = \lim_{m\to\infty} n(T_m)$. To explore the integral expression of $n(T)$ given by Lemma 2.1.10, we need to define some notations.

For any $0 \le k \le N$, define

$$D_k := \Big\{ \boldsymbol{x} \in \mathbb{R}^{N(N+1)/2} : \text{Matri}_N(\boldsymbol{x}) \text{ is non-degenerate} \\ \text{and has exactly } k \text{ negative eigenvalues} \Big\}.$$

Let $p(z)$, $z \in \mathbb{R}$ be the density of $X(\mathbf{0})$, let $p(\boldsymbol{z})$, $\boldsymbol{z} \in \mathbb{R}^N$ be the density of $\nabla X(\mathbf{0})$, and let $p(\boldsymbol{z}, z)$, $(\boldsymbol{z}, z) \in \mathbb{R}^N \times \mathbb{R}$ be the density of $(\nabla X(\mathbf{0}), X(\mathbf{0}))$. Then by Lemma 2.1.4, $(\nabla X(\mathbf{0}), X(\mathbf{0}))$ are Gaussian with

$$\text{Cov}\left[X_i(\mathbf{0}), X(\mathbf{0})\right] = 0$$

for any $1 \le i \le N$, and thus, for any $(\boldsymbol{z}, z) \in \mathbb{R}^N \times \mathbb{R}$,

$$p(\boldsymbol{z}, z) = p(\boldsymbol{z})p(z).$$

For any $\boldsymbol{t} \in \mathbb{R}^N \setminus \{\mathbf{0}_N\}$, let $p_{\boldsymbol{t}}(\boldsymbol{x}, \boldsymbol{z})$, $(\boldsymbol{x}, \boldsymbol{z}) \in \mathbb{R}^N \times \mathbb{R}^N$ be the density of $(\nabla X(\boldsymbol{t}), \nabla X(\mathbf{0}))$, and let $p_{\boldsymbol{t}}(\boldsymbol{x}'', \boldsymbol{x}', x, \boldsymbol{z}, z)$, $(\boldsymbol{x}'', \boldsymbol{x}', x, \boldsymbol{z}, z) \in \mathbb{R}^{N(N+1)/2} \times \mathbb{R}^N \times \mathbb{R} \times \mathbb{R}^N \times \mathbb{R}$ be the density of $(\nabla^2 X(\boldsymbol{t}), \nabla X(\boldsymbol{t}), X(\boldsymbol{t}), \nabla X(\mathbf{0}), X(\mathbf{0}))$. For any $\boldsymbol{z}' \in \mathbb{R}^N$ and $z \in \mathbb{R}$, let $p_{\boldsymbol{t}}(\boldsymbol{x}'', \boldsymbol{x}', x|\boldsymbol{z}, z)$, $(\boldsymbol{x}'', \boldsymbol{x}', x) \in \mathbb{R}^{N(N+1)/2} \times \mathbb{R}^N \times \mathbb{R}$ be the density of $(\nabla^2 X(\boldsymbol{t}), \nabla X(\boldsymbol{t}), X(\boldsymbol{t}))$ conditional on $\nabla X(\mathbf{0}) = \boldsymbol{z}'$ and $X(\mathbf{0}) = z$.

Similarly, for any $\boldsymbol{t} \in \mathbb{R}^N \setminus \{\mathbf{0}_N\}$, $\boldsymbol{x}', \boldsymbol{z}' \in \mathbb{R}^N$, let $p_{\boldsymbol{t}}(\boldsymbol{x}'', x, z|\boldsymbol{x}', \boldsymbol{z}')$, $(\boldsymbol{x}'', x, z) \in \mathbb{R}^{N(N+1)/2} \times \mathbb{R} \times \mathbb{R}$ be the density of $(\nabla^2 X(\boldsymbol{t}), X(\boldsymbol{t}), X(\mathbf{0}))$ conditional on $\nabla X(\boldsymbol{t}) = \boldsymbol{x}'$ and $\nabla X(\mathbf{0}) = \boldsymbol{z}'$. In particular, the covariance matrix of the Gaussian density $p_{\boldsymbol{t}}(\boldsymbol{x}'', x, z|\mathbf{0}, \mathbf{0})$ is $\boldsymbol{\Sigma}(\boldsymbol{t})$ which we carefully studied in the previous sections. In addition, by Lemma 2.1.3 and (A.6), the mean vector of $p_{\boldsymbol{t}}(\boldsymbol{x}'', x, z|\mathbf{0}, \mathbf{0})$ is $\mathbf{0}_L$. Thus, we have

$$p_{\boldsymbol{t}}(\boldsymbol{x}'', x, z|\mathbf{0}, \mathbf{0}) = \frac{1}{\sqrt{(2\pi)^L \det\left(\boldsymbol{\Sigma}(\boldsymbol{t})\right)}} \exp\left(-\frac{1}{2}(\boldsymbol{x}'', x, z)\boldsymbol{\Sigma}(\boldsymbol{t})^{-1}(\boldsymbol{x}'', x, z)^T\right). \tag{4.31}$$

By Lemma 2.1.10, we have for any $0 \le k \le N$ and $u \in \mathbb{R}$, the density, $f_{u,k}(\boldsymbol{t})$ of the mean measure of the (non-degenerate, but we have mentioned in Remark 2.1.12 that a sample function on a compact set $T \subset \mathbb{R}^N$ with $\lambda_{N-1}(\partial T) < \infty$ does not have degenerate critical points with probability one) critical points of X above u with index k, conditional on $\nabla X(\mathbf{0}) = \mathbf{0}_N$ and $X(\mathbf{0}) > u$, is given by

$$\begin{aligned}
&f_{u,k}(\boldsymbol{t}) \\
&= P(X(\mathbf{0}) > u)^{-1} \int_{x,z>u} \int_{D_k} |\det\,(\mathrm{Matri}_N(\boldsymbol{x}''))|\, p_{\boldsymbol{t}}(\boldsymbol{x}'', \mathbf{0}, x|\mathbf{0}, z) p(z) d\boldsymbol{x}'' dx dz \\
&= P(X(\mathbf{0}) > u)^{-1} \int_{x,z>u} \int_{D_k} |\det\,(\mathrm{Matri}_N(\boldsymbol{x}''))|\, p_{\boldsymbol{t}}(\boldsymbol{x}'', \mathbf{0}, x, \mathbf{0}, z) p(\mathbf{0}, z)^{-1} p(z) d\boldsymbol{x}'' dx dz \\
&= P(X(\mathbf{0}) > u)^{-1} p(\mathbf{0})^{-1} p_{\boldsymbol{t}}(\mathbf{0}, \mathbf{0}) \int_{x,z>u} \int_{D_k} |\det\,(\mathrm{Matri}_N(\boldsymbol{x}''))|\, p_{\boldsymbol{t}}(\boldsymbol{x}'', x, z|\mathbf{0}, \mathbf{0}) d\boldsymbol{x}'' dx dz.
\end{aligned} \tag{4.32}$$

We can further define

$$f_{u,+}(\boldsymbol{t}) := \sum_{k \text{ even}} f_{u,k}(\boldsymbol{t})$$

and

$$f_{u,-}(\boldsymbol{t}) := \sum_{k \text{ odd}} f_{u,k}(\boldsymbol{t}),$$

i.e., to replace D_k in (4.32) with $\{\boldsymbol{x}'' \in \mathbb{R}^{N(N+1)/2} : \det(\mathrm{Matri}_N(\boldsymbol{x}'')) > 0\}$ and $\{\boldsymbol{x}'' \in \mathbb{R}^{N(N+1)/2} : \det(\mathrm{Matri}_N(\boldsymbol{x}'')) < 0\}$, respectively. It is trivial to show that $f_{u,\pm}(\boldsymbol{t})$ are both continuous functions of $\boldsymbol{t} \in \mathbb{R}^N \setminus \{\mathbf{0}_N\}$. The following is our first main result.

Theorem 4.5.1. *Let X be qualified under perturbation. Then for any $u \in \mathbb{R}$,*

$$\lim_{\|\boldsymbol{t}\| \to 0} \frac{f_{u,+}(\boldsymbol{t})}{f_{u,-}(\boldsymbol{t})} = 1.$$

4.5.2 Preparation for Main Result 1

Let X be qualified under perturbation. In this section, we only consider $\boldsymbol{t} = \boldsymbol{u}_0 r$, $r > 0$. Based on the knowledge of the covariance $\boldsymbol{\Sigma}(r)$ in the previous sections, we can establish two lemmas suggesting the asymptotic symmetry between the domains

$$D_{u,+}(r) := \left\{ \boldsymbol{y} \in \mathbb{R}^L : \boldsymbol{A}(r)_{(L-1)} \boldsymbol{y} > u, \boldsymbol{A}(r)_{(L)} \boldsymbol{y} > u, \frac{1}{r} \det\,(\mathrm{Matri}_N(\boldsymbol{A}(r)\boldsymbol{y})) > 0 \right\}$$

and

$$D_{u,-}(r) := \left\{ \boldsymbol{y} \in \mathbb{R}^L : \boldsymbol{A}(r)_{(L-1)} \boldsymbol{y} > u, \boldsymbol{A}(r)_{(L)} \boldsymbol{y} > u, \frac{1}{r} \det\,(\mathrm{Matri}_N(\boldsymbol{A}(r)\boldsymbol{y})) < 0 \right\}$$

as $r \to 0$. This asymptotic symmetry plays an essential role in our first main result.

We first introduce an important matrix. For any $r > 0$, let $a_{ij}(r)$ be the element of $\boldsymbol{A}(r)$ on the i-th row and j-th column for any $1 \leq i, j \leq L$. Define the map $\tau : \{1, \ldots, N\}^2 \to \{1, \ldots, N(N+1)/2\}$, where for any $1 \leq i \leq j \leq N$,

$$\tau(j, i) \equiv \tau(i, j) := i + j(j-1)/2.$$

Then for any $\boldsymbol{v} = (v_1, \ldots, v_N)^T \in \{1, \ldots, L\}^N$ and $r > 0$, we can define $\boldsymbol{B}^{\boldsymbol{v}}(r) := (b^{\boldsymbol{v}}_{i,j}(r))_{1 \leq i,j \leq N}$, where for any $1 \leq i, j \leq N$,

$$b^{\boldsymbol{v}}_{i,j}(r) = a_{\tau(i,j),v_i}(r).$$

For example, when $N = 3$,

$$\boldsymbol{B}^{\boldsymbol{v}}(r) := \begin{pmatrix} a_{1,v_1}(r) & a_{2,v_1}(r) & a_{4,v_1}(r) \\ a_{2,v_2}(r) & a_{3,v_2}(r) & a_{5,v_2}(r) \\ a_{4,v_3}(r) & a_{5,v_3}(r) & a_{6,v_3}(r) \end{pmatrix}.$$

In particular, if taking $v_1 = \cdots = v_N = i$ for some $1 \leq i \leq L$, then we have

$$\boldsymbol{B}^{\boldsymbol{v}}(r) = \mathrm{Matri}_N \left(\boldsymbol{A}(r)^{(i)} \right).$$

In general, for any $r > 0$, $\boldsymbol{B}^{\boldsymbol{v}}(r)$ can be written in the form:

$$\boldsymbol{B}^{\boldsymbol{v}}(r) = \begin{pmatrix} \mathrm{Matri}_N(\boldsymbol{A}(r)^{(v_1)})_{(1)} \\ \vdots \\ \mathrm{Matri}_N(\boldsymbol{A}(r)^{(v_N)})_{(N)} \end{pmatrix}. \tag{4.33}$$

Let Π_N be the set of permutations on $\{1, \ldots, N\}$. For any $\sigma \in \Pi_N$, define a map $\hat{\sigma} : \{1, \ldots, L\}^N \to \{1, \ldots, L\}^N$, where for any $\boldsymbol{v} \in \{1, \ldots, L\}^N$,

$$\hat{\sigma}(\boldsymbol{v}) = (v_{\sigma(1)}, \ldots, v_{\sigma(N)}).$$

With slight abuse of notation, we still write $\hat{\sigma}$ as σ, but one can easily distinguish them by the object it works on.

Now we can explain why $\boldsymbol{B}^{\boldsymbol{v}}(r)$ is important in the asymptotic symmetry between $D_{u,+}(r)$ and $D_{u,-}(r)$. Note that for any $\boldsymbol{y} \in \mathbb{R}^L$ and integers $1 \leq i, j \leq N$,

$$\mathrm{Matri}_N(\boldsymbol{A}(r)\boldsymbol{y})[i, j] = \sum_{k=1}^{L} a_{\tau(i,j),k}(r) y_k.$$

Then by the definition of determinant, we have

$$\begin{aligned}
&\det\left(\mathrm{Matri}_N(\boldsymbol{A}(r)\boldsymbol{y})\right)\\
&= \sum_{\sigma\in\Pi_N}(-1)^{\mathrm{sign}\,(\sigma)}\prod_{1\le n\le N}\left(\sum_{k=1}^{L} a_{\tau(n,\sigma(n)),k}(r)y_k\right)\\
&= \sum_{\boldsymbol{v}\in\{1,\ldots,L\}^N}\sum_{\sigma\in\Pi_N}(-1)^{\mathrm{sign}\,(\sigma)}a_{\tau(1,\sigma(1)),v_1}(r)\cdots a_{\tau(N,\sigma(N)),v_N}(r)y_{v_1}\cdots y_{v_N}\\
&= \sum_{\boldsymbol{v}\in\{1,\ldots,L\}^N}\sum_{\sigma\in\Pi_N}(-1)^{\mathrm{sign}\,(\sigma)}b^{\boldsymbol{v}}_{1,\sigma(1)}(r)\cdots b^{\boldsymbol{v}}_{N,\sigma(N)}(r)y_{v_1}\cdots y_{v_N}\\
&= \sum_{\boldsymbol{v}\in\{1,\ldots,L\}^N}\det\left(\boldsymbol{B}^{\boldsymbol{v}}(r)\right)y_{v_1}\cdots y_{v_N}\\
&= \frac{1}{N!}\sum_{\boldsymbol{v}\in\{1,\ldots,L\}^N}\left(\sum_{\sigma\in\Pi_N}\det\left(\boldsymbol{B}^{\sigma(\boldsymbol{v})}(r)\right)\right)y_{v_1}\cdots y_{v_N}.
\end{aligned}\tag{4.34}$$

This implies that the asymptotic symmetry between $D_{u,+}(r)$ and $D_{u,-}(r)$ is determined by the asymptotic behavior of the coefficients, $\sum_{\sigma\in\Pi_N}\det\left(\boldsymbol{B}^{\sigma(\boldsymbol{v})}(r)\right)$, of $y_{v_1}\cdots y_{v_N}$, $\boldsymbol{v}\in\{1,\ldots,L\}^N$ as $r\to 0$. To study asymptotic properties of $\sum_{\sigma\in\Pi_N}\det\left(\boldsymbol{B}^{\sigma(\boldsymbol{v})}(r)\right)$, the perturbation condition in Definition 4.3.6 will be intensively used, and we also need the following notations:

$$\begin{aligned}V_N := \{(v_1,\ldots,v_N)\in\{1,\ldots,L\}^N : L-N\le v_1\le L,\\ 1\le v_i\le L-N-1 \text{ for any } 2\le i\le N\}\end{aligned}$$

and

$$\widetilde{V}_N := \{\boldsymbol{v}\in\{1,\ldots,L\}^N : \sigma(\boldsymbol{v})\in V_N \text{ for some } \sigma\in\Pi_N\}.$$

Lemma 4.5.2. *Let X be qualified under perturbation. Then*

$$\max_{\boldsymbol{v}\in\{1,\ldots,L\}^N\setminus\widetilde{V}_N}\left|\sum_{\sigma\in\Pi_N}\det\left(\boldsymbol{B}^{\sigma(\boldsymbol{v})}(r)\right)\right| = o(r) \text{ as } r\to 0.$$

Proof. For any $\boldsymbol{v}=(v_1,\ldots,v_N)^T\in\{1,\ldots,L\}^N\setminus\widetilde{V}_N$, there are only two possible situations:

(1) $1\le v_i\le L-N-1$ for $i=1,\ldots,N$;

(2) there exist integers $1\le i,j\le N$ such that $i\ne j$ and $L-N\le v_i,v_j\le L$.

Since these two situations (as two separated subsets of $\{1, \ldots, L\}^N \setminus \widetilde{V}_N$) are closed under permutations in Π_N, it suffices to show that as $r \to 0$,

$$\det\left(\boldsymbol{B}^{\boldsymbol{v}}(r)\right) = o(r)$$

for any $\boldsymbol{v}$ in these two situations.

For Situation (1), note that by (ii) of Lemma 4.4.4,

$$\mathrm{Rank}\left(\boldsymbol{\Sigma}_0\right) = L - N - 1.$$

By (i) and (iii) of Lemma 4.4.3, for any $1 \le i \le N$,

$$\mathrm{Matri}\,\left(\boldsymbol{A}(r)^{(v_i)}\right)_{(i)} \boldsymbol{u}_0 = \left(H(\boldsymbol{u}_0)\boldsymbol{A}(r)^{(v_i)}\right)_{(i)} = \boldsymbol{0}_N^T + o(r) \text{ as } r \to 0.$$

Then by (4.33),

$$\boldsymbol{B}^{\boldsymbol{v}}(r)^{(N)} = \boldsymbol{B}^{\boldsymbol{v}}(r)\boldsymbol{u}_0 = \boldsymbol{0}_N + o(r) \text{ as } r \to 0.$$

This implies $\det\left(\boldsymbol{B}^{\boldsymbol{v}}(r)\right) = o(r)$ as $r \to 0$.

As for Situation (2), by (ii) of Lemma 4.4.1, (4.33) and that $\boldsymbol{A}(r) = \boldsymbol{P}(r)\boldsymbol{\Lambda}^{1/2}(r)$, there are two rows of $\boldsymbol{B}^{\boldsymbol{v}}(r)$ converge to $\boldsymbol{0}_N$ with the speed $O(r)$ as $r \to 0$. This also implies $\det\left(\boldsymbol{B}^{\boldsymbol{v}}(r)\right) = o(r)$ as $r \to 0$. □

Lemma 4.5.3. *Let X be qualified under perturbation. Then*

$$\max_{\boldsymbol{v} \in \widetilde{V}_N} \left| \sum_{\sigma \in \Pi_N} \det\,\left(\boldsymbol{B}^{\sigma(\boldsymbol{v})}(r)\right) \right| = \Theta(r) \;\; \textit{as } r \to 0. \tag{4.35}$$

Proof. The basic idea of this proof is to select a suitable $\boldsymbol{v} \in \widetilde{V}_N$ such that

$$\sum_{\sigma \in \Pi_N} \det\left(\boldsymbol{B}^{\sigma(\boldsymbol{v})}(r)\right) = \Theta(r) \text{ as } r \to 0.$$

First of all, we need to define some notations. For any $1 \le j \le L$, define

$$\boldsymbol{M}^j(0) := \mathrm{Matri}_N\left(\boldsymbol{A}_0^{(j)}\right), \quad \boldsymbol{M}^j(r) := \mathrm{Matri}_N\left(\boldsymbol{A}(r)^{(j)}\right),$$

$$\widetilde{\boldsymbol{M}}^j(0) := \mathrm{Matri}_N\left(\boldsymbol{P}_0^{(j)}\right), \text{ and } \widetilde{\boldsymbol{M}}^j(r) := \mathrm{Matri}_N\left(\boldsymbol{P}(r)^{(j)}\right).$$

Indeed, $\boldsymbol{M}^j(0)$ and $\widetilde{\boldsymbol{M}}^j(0)$ are the left limits of $\boldsymbol{M}^j(r)$ and $\widetilde{\boldsymbol{M}}^j(r)$ as $r \downarrow 0$, respectively. According to Remark 4.4.5, since the problem is independent of the rescaling of the covariance function, it is safe to use (vii) of Lemma 4.4.4 in this proof. By (i), (ii), and (vii) of Lemma 4.4.4, we can find an integer $1 \leq v^* \leq L - N - 1$ such that

$$\boldsymbol{P}_0[i + j(j-1)/2, v^*] = \delta_{i,j} x_{v^*} \text{ for any } 1 \leq i \leq j \leq N-1,$$

$$\boldsymbol{P}_0[i + N(N-1)/2, v^*] = 0 \text{ for any } 1 \leq i \leq N, \text{ and } \boldsymbol{P}_0[L-1, v^*] = \boldsymbol{P}_0[L, v^*] = y_{v^*},$$

where x_{v^*} and y_{v^*} are both non-zero. Then

$$\boldsymbol{M}^{v^*}(0) = \text{Matri}_N \left(\boldsymbol{A}_0^{(v^*)} \right) = \text{diag} \left(x_{v^*} \lambda_{v^*,0}^{1/2}, \ldots, x_{v^*} \lambda_{v^*,0}^{1/2}, 0 \right). \tag{4.36}$$

Our next step is to calculate $\sum_{\sigma \in \Pi_N} \det \left(\boldsymbol{B}^{\sigma(\boldsymbol{v})}(r) \right)$ for a general $\boldsymbol{v} \in \widetilde{V}_N$ using the above notations. By (4.33), for any $\boldsymbol{v} \in \{1, \ldots, L\}^N$,

$$\begin{aligned}
\sum_{\sigma \in \Pi_N} \det \left(\boldsymbol{B}^{\sigma(\boldsymbol{v})}(r) \right) &= \sum_{\sigma \in \Pi_N} \det \begin{pmatrix} \boldsymbol{M}^{v_{\sigma(1)}}(r)_{(1)} \\ \vdots \\ \boldsymbol{M}^{v_{\sigma(N)}}(r)_{(N)} \end{pmatrix} \\
&= \sum_{\sigma \in \Pi_N} (-1)^{\text{sign}(\sigma^{-1})} \det \begin{pmatrix} \boldsymbol{M}^{v_1}(r)_{(\sigma^{-1}(1))} \\ \vdots \\ \boldsymbol{M}^{v_N}(r)_{(\sigma^{-1}(N))} \end{pmatrix} \\
&= \sum_{\sigma \in \Pi_N} (-1)^{\text{sign}(\sigma)} \det \begin{pmatrix} \boldsymbol{M}^{v_1}(r)_{(\sigma(1))} \\ \vdots \\ \boldsymbol{M}^{v_N}(r)_{(\sigma(N))} \end{pmatrix}.
\end{aligned}$$

Note that for any matrix $\boldsymbol{A} = (a_{i,j}) \in \mathbb{R}^{N \times N}$,

$$\det(\boldsymbol{A}) := \sum_{\sigma \in \Pi_N} (-1)^{\text{sign}(\sigma)} a_{1,\sigma(1)} \cdots a_{N,\sigma(N)}.$$

Similarly, we can define an operation f on any symmetric matrices $\boldsymbol{A}_1, \ldots, \boldsymbol{A}_N \in \mathbb{R}^{N \times N}$ by

$$\begin{aligned}
f(\boldsymbol{A}_1, \cdots, \boldsymbol{A}_N) : &= \sum_{\sigma \in \Pi_N} (-1)^{\text{sign}(\sigma)} \det \begin{pmatrix} (\boldsymbol{A}_1)_{(\sigma(1))} \\ \vdots \\ (\boldsymbol{A}_N)_{(\sigma(N))} \end{pmatrix} \\
&\equiv \sum_{\sigma \in \Pi_N} (-1)^{\text{sign}(\sigma)} \det \left((\boldsymbol{A}_1)^{(\sigma(1))}, \cdots, (\boldsymbol{A}_N)^{(\sigma(N))} \right).
\end{aligned} \tag{4.37}$$

Then in our context, we have for any $r > 0$,

$$\sum_{\sigma \in \Pi_N} \det \left(\boldsymbol{B}^{\sigma(\boldsymbol{v})}(r) \right) = f\left(\boldsymbol{M}^{v_1}(r), \cdots, \boldsymbol{M}^{v_N}(r) \right).$$

Note that by Remark 4.4.2, we have

$$\boldsymbol{M}^i(r) = \begin{cases} \boldsymbol{M}^i(0) + O(r), & \text{for any } 1 \le i \le L - N - 1; \\ \lambda_{i,2}^{1/2} r \widetilde{\boldsymbol{M}}^i(r) + o(r), & \text{otherwise.} \end{cases}$$

For any $\boldsymbol{v} \in \widetilde{V}_N$, there exists an integer $1 \le k \le N$ such that $L - N \le v_k \le L$. Then

$$\begin{aligned} & \sum_{\sigma \in \Pi_N} \det \left(\boldsymbol{B}^{\sigma(\boldsymbol{v})}(r) \right) \\ &= \lambda_{v_k,2}^{1/2} r f \left(\boldsymbol{M}^{v_1}(r), \cdots, \boldsymbol{M}^{v_{k-1}}(r), \widetilde{\boldsymbol{M}}^{v_k}(r), \boldsymbol{M}^{v_{k+1}}(r), \cdots, \boldsymbol{M}^{v_N}(r) \right) + o(r) \\ &= (-1)^{N-k} \lambda_{v_k,2}^{1/2} r f \left(\boldsymbol{M}^{v_1}(r), \cdots, \widehat{\boldsymbol{M}^{v_k}(r)}, \cdots, \boldsymbol{M}^{v_N}(r), \widetilde{\boldsymbol{M}}^{v_k}(r) \right) + o(r) \\ &= (-1)^{N-k} \lambda_{v_k,2}^{1/2} r f \left(\boldsymbol{M}^{v_1}(0), \cdots, \widehat{\boldsymbol{M}^{v_k}(0)}, \cdots, \boldsymbol{M}^{v_N}(0), \widetilde{\boldsymbol{M}}^{v_k}(0) \right) + o(r), \end{aligned}$$

where the hat over a component of a vector means that component is discarded.

From the above calculation, we see that it suffices to show

$$\lambda_{v_k,2}^{1/2} f \left(\boldsymbol{M}^{v_1}(0), \cdots, \widehat{\boldsymbol{M}^{v_k}(0)}, \cdots, \boldsymbol{M}^{v_N}(0), \widetilde{\boldsymbol{M}}^{v_k}(0) \right) \neq 0$$

for some $\boldsymbol{v} \in \widetilde{V}_N$ (with $L - N \le v_k \le L$ for some $1 \le k \le N$). To take the advantage of (4.36) and by noting that $\lambda_{i,2} > 0$ for any $L - N \le i \le L - 1$ and $\lambda_{L,2} = 0$ (see Lemma 4.4.7), we select $\boldsymbol{v}$ having the form: $\boldsymbol{v} = (v^*, \ldots, v^*, v_0)$ for some $L - N \le v_0 \le L - 1$. Then it suffices to show that there exists an integer $L - N \le v_0 \le L - 1$ such that

$$f \left(\boldsymbol{M}^{v^*}(0), \cdots, \boldsymbol{M}^{v^*}(0), \widetilde{\boldsymbol{M}}^{v_0}(0) \right) \neq 0.$$

To this end, we first observe that

$$
\begin{aligned}
& f\left(\boldsymbol{M}^{v^*}(0), \cdots, \boldsymbol{M}^{v^*}(0), \widetilde{\boldsymbol{M}}^{v_0}(0)\right) \\
= & \sum_{\sigma \in \Pi_N} (-1)^{\operatorname{sign}(\sigma)} \det\left(\boldsymbol{M}^{v^*}(0)^{(\sigma(1))}, \cdots, \boldsymbol{M}^{v^*}(0)^{(\sigma(N-1))}, \widetilde{\boldsymbol{M}}^{v_0}(0)^{(\sigma(N))}\right) \\
= & \sum_{\sigma \in \Pi_N} (-1)^{\operatorname{sign}(\sigma)} \widetilde{\boldsymbol{M}}^{v_0}(0)[N, \sigma(N)] \\
& \det\left(\left(\boldsymbol{M}^{v^*}(0)^{(\sigma(1))}, \cdots, \boldsymbol{M}^{v^*}(0)^{(\sigma(N-1))}\right)[1:(N-1), 1:(N-1)]\right) \\
= & (N-1)! \sum_{j=1}^{N} (-1)^{N-j} \widetilde{\boldsymbol{M}}^{v_0}(0)[N, j] \\
& \det\left(\left(\boldsymbol{M}^{v^*}(0)^{(1)}, \cdots, \widehat{\boldsymbol{M}^{v^*}(0)}^{(j)}, \cdots, \boldsymbol{M}^{v^*}(0)^{(N)}\right)[1:(N-1), 1:(N-1)]\right) \\
= & (N-1)! \det\begin{pmatrix} \boldsymbol{M}^{v^*}(0)_{(1)} \\ \vdots \\ \boldsymbol{M}^{v^*}(0)_{(N-1)} \\ \widetilde{\boldsymbol{M}}^{v_0}(0)_{(N)} \end{pmatrix},
\end{aligned}
$$

where the second equation follows from the fact that $\boldsymbol{M}^{v^*}(0)_{(N)} = \mathbf{0}_N$ (see (4.36)), and the last equation is given by Laplace's expansion.

Now if $f(\boldsymbol{M}^{v^*}(0), \cdots, \boldsymbol{M}^{v^*}(0), \widetilde{\boldsymbol{M}}^{v_0}(0)) = 0$ for any integer $L - N \leq v_0 \leq L - 1$, then

$$
\det\begin{pmatrix} \boldsymbol{M}^{v^*}(0)_{(1)} \\ \vdots \\ \boldsymbol{M}^{v^*}(0)_{(N-1)} \\ \widetilde{\boldsymbol{M}}^{v_0}(0)_{(N)} \end{pmatrix} = 0,
$$

which, together with (4.36), implies

$$
\widetilde{\boldsymbol{M}}^{v_0}(0)[N, N] = \boldsymbol{u}_0^T \widetilde{\boldsymbol{M}}^{v_0}(0) \boldsymbol{u}_0 = 0.
$$

Then by (i) of Lemma 4.4.3, we have for any $L - N \leq v_0 \leq L - 1$,

$$
\boldsymbol{u}_0^T \boldsymbol{H}(\boldsymbol{u}_0) \boldsymbol{P}_0^{(v_0)} = \boldsymbol{u}_0^T \widetilde{\boldsymbol{M}}^{v_0}(0) \boldsymbol{u}_0 = 0.
$$

Note that $(\boldsymbol{u}_0^T \boldsymbol{H}(\boldsymbol{u}_0))^T = (\boldsymbol{H}(\boldsymbol{u}_0)_{(N)})^T$ is non-zero, and by (ii) of Lemma 4.4.3, it is in the zero space of $\boldsymbol{\Sigma}_0$ which is expanded by $\boldsymbol{P}_0^{(L-N)}, \ldots, \boldsymbol{P}_0^{(L)}$. Thus, the above equation implies $(\boldsymbol{H}(\boldsymbol{u}_0)_{(N)})^T$ and $\boldsymbol{P}_0^{(L)}$ must be linearly dependent. Then by Lemma 4.4.7, we get $\boldsymbol{H}(\boldsymbol{u}_0)_{(N)}$ and $(0, \dots, 0, 1, -1) \in \mathbb{R}^L$ are linearly dependent, which contradicts (4.14). This completes the proof. $\square$

Let X be qualified under perturbation. By (4.34), Lemmas 4.5.2, and 4.5.3, we have the following important observation: for any $\boldsymbol{y} = (y_1, \dots, y_L)^T \in \mathbb{R}^L$,

$$\lim_{r \to 0} r^{-1} \det\left(\text{Matri}_N(\boldsymbol{A}(r)\boldsymbol{y})\right) = \sum_{i=L-N}^{L} y_i K_i(y_1, \dots, y_{L-N-1}), \tag{4.38}$$

where each of $K_i(y_1, \dots, y_{L-N-1})$, $i = L-N, \dots, L$ is either a homogeneous polynomial of $y_1, \dots, y_{L-N-1}$ with degree $N-1$ or a zero function, and at least one of K_i, $L-N \leq i \leq L$ is not a zero function. Thus, $\sum_{i=L-N}^{L} y_i K_i(y_1, \dots, y_{L-N-1})$ is a homogeneous polynomial of $y_1, \dots, y_L$ with degree N. For any $r > 0$ and $\boldsymbol{y} \in \mathbb{R}^L$, define

$$h_r(\boldsymbol{y}) := r^{-1} \det\left(\text{Matri}_N(\boldsymbol{A}(r)\boldsymbol{y})\right). \tag{4.39}$$

We can also define

$$h_0(\boldsymbol{y}) := \sum_{i=L-N}^{L} y_i K_i(y_1, \dots, y_{L-N-1}). \tag{4.40}$$

Then $h_r(\boldsymbol{y})$, $r \geq 0$ are all polynomials of $y_1, \dots, y_L$, and (4.38) becomes for any $\boldsymbol{y} \in \mathbb{R}^L$,

$$h_0(\boldsymbol{y}) = \lim_{r \to 0} h_r(\boldsymbol{y}). \tag{4.41}$$

Since $\boldsymbol{A}(r)$ is continuous on $r \in [0, \delta_{pc}]$ (Definition 4.3.6), we have for any $\boldsymbol{y} \in \mathbb{R}^L$, $h_r(\boldsymbol{y})$ is a continuous function of $r \in [0, \delta_{pc}]$.

For any $u \in \mathbb{R}$, $r \geq 0$ and $i = L-1, L$, define

$$H_{i,u}(r) := \left\{\boldsymbol{y} \in \mathbb{R}^L : \boldsymbol{A}(r)_{(i)}\boldsymbol{y} > u\right\}. \tag{4.42}$$

For any $r \geq 0$, we also define

$$G_+(r) := \left\{\boldsymbol{y} \in \mathbb{R}^L : h_r(\boldsymbol{y}) > 0\right\} \text{ and } G_-(r) := \left\{\boldsymbol{y} \in \mathbb{R}^L : h_r(\boldsymbol{y}) < 0\right\}.$$

With the above notations, the domains $D_{u,\pm}(r)$ mentioned at the start of this section can also be defined as

$$D_{u,\pm}(r) := H_{L-1,u}(r) \cap H_{L,u}(r) \cap G_\pm(r), \tag{4.43}$$

for any $r \geq 0$. In particular, by $\boldsymbol{A}(0) = \boldsymbol{P}(0)\boldsymbol{\Lambda}^{1/2}(0)$, (ii) and (v) of Lemma 4.4.4,

$$\boldsymbol{A}_{(L-1)}(0) = \boldsymbol{A}_{(L)}(0).$$

Then we have

$$H_{L-1,u}(0) = H_{L,u}(0),$$

which implies

$$D_{u,\pm}(0) := H_{L,u}(0) \cap G_{\pm}(0). \tag{4.44}$$

The following lemma reveals the most important idea in this section: the asymptotic symmetry between $D_{u,\pm}(r)$ as $r \to 0$. Let $b > a$. For any $r_0 \in [a, b]$ and $A_r \subset \mathbb{R}^n$ ($n \geq 1$), $r \in [a, b]$, define

$$\liminf_{r \to r_0} A_r := \bigcup_{\delta > 0} \bigcap_{|r - r_0| < \delta} A_r \text{ and } \limsup_{r \to r_0} A_r := \bigcap_{\delta > 0} \bigcup_{|r - r_0| < \delta} A_r.$$

Lemma 4.5.4. *Let X be qualified under perturbation. Then for any $u \in \mathbb{R}$, we have*

(i) if $(y_1, \ldots, y_L) \in D_{u,\pm}(0)$, then $(y_1, \ldots, y_{L-N-1}, -y_{L-N}, \ldots, -y_L) \in D_{u,\mp}(0)$;

(ii) $D_{u,\pm}(0)$ are both non-empty open sets;

(iii) for any $r_0 \in [0, \delta_{pc}]$, we have

$$\lim_{r \to r_0} I_{D_{u,\pm}(r)}(\boldsymbol{y}) = I_{D_{u,\pm}(r_0)}(\boldsymbol{y}) \quad \textit{for almost all } \boldsymbol{y} \in \mathbb{R}^L,$$

where I stands for the indicator function of a set.

Proof. For (i), let $\boldsymbol{y} = (y_1, \ldots, y_L) \in D_{u,+}(0)$ and $\boldsymbol{y}' = (y_1, \ldots, y_{L-N-1}, -y_{L-N}, \ldots, -y_L)$. Then by (4.44),

$$\boldsymbol{A}_{(N)}(0)\boldsymbol{y} > u \text{ and } h_0(\boldsymbol{y}) > 0.$$

Note that by $\boldsymbol{A}(0) = \boldsymbol{P}(0)\boldsymbol{\Lambda}^{1/2}(0)$ and (ii) of Lemma 4.4.4, $\boldsymbol{A}(0)^{(i)} = 0$ for any $L - N \leq i \leq L$, which implies $\boldsymbol{A}_{(L)}(0)\boldsymbol{y}' > u$. In addition, by (4.40),

$$h_0(\boldsymbol{y}') = \sum_{L-N}^{L} (-y_i) K(y_1, \ldots, y_{L-N-1}) = -h_0(\boldsymbol{y}).$$

Thus, $\boldsymbol{y}' \in D_{u,-}(0)$. The proof of the other part is similar.

As for (ii), it is easy to see $D_{u,\pm}(0)$ are both open. By (i), $D_{u,\pm}(0)$ are either both empty or both non-empty. Since

$$D_{u,+}(0) \cup D_{u,-}(0) = H_{L,u}(0) \setminus \{\boldsymbol{y} \in \mathbb{R}^L : h_0(\boldsymbol{y}) = 0\},$$

their union is non-empty. Thus, $D_{u,\pm}(0)$ are both non-empty open sets, as required.

As for (iii), fix $r_0 \in [0, \delta_{pc}]$. By the continuity of $\boldsymbol{A}(r)$ on $r \in [0, \delta_{pc}]$, it is easy to see

$$H_{i,u}(r_0) \subset \liminf_{r\to r_0} H_{i,u}(r) \subset \limsup_{r\to r_0} H_{i,u}(r) \subset H_{i,u}(r_0) \cup \partial H_{i,u}(r_0)$$

for $i = L-1, L$, and

$$G_\pm(r_0) \subset \liminf_{r\to r_0} G_\pm(r) \subset \limsup_{r\to r_0} G_\pm(r) \subset G_\pm(r_0) \cup \partial G_\pm(r_0),$$

where ∂ stands for the boundary of a set. This implies

$$D_{u,\pm}(r_0) \subset \liminf_{r\to r_0} D_{u,\pm}(r) \subset \limsup_{r\to r_0} D_{u,\pm}(r) \subset D_{u,\pm}(r_0) \cup Q_0, \tag{4.45}$$

where

$$Q_0 := \partial H_{L-1,u}(r_0) \cup \partial H_{L,u}(r_0) \cup \partial G_\pm(r_0),$$

and it is easy to see $\lambda_L(Q_0) = 0$. Note that for any $\boldsymbol{y} \in (\liminf_{r\to r_0} D_{u,\pm}(r)) \cap D_{u,\pm}(r_0)$,

$$I_{D_{u,\pm}(r)}(\boldsymbol{y}) = 1 \text{ for any } n \geq 1 \text{ and } I_{D_{u,\pm}(r_0)}(\boldsymbol{y}) = 1,$$

while for any $\boldsymbol{y} \in (\liminf_{r\to r_0} D^c_{u,\pm}(r)) \cap D^c_{u,\pm}(r_0) = (\limsup_{r\to r_0} D_{u,\pm}(r))^c \cap D^c_{u,\pm}(r_0)$,

$$I_{D_{u,\pm}(r)}(\boldsymbol{y}) = 0 \text{ for any } n \geq 1 \text{ and } I_{D_{u,\pm}(r_0)}(\boldsymbol{y}) = 0,$$

where by (4.45),

$$\left(\liminf_{r\to r_0} D_{u,\pm}(r)\right) \cap D_{u,\pm}(r_0) = D_{u,\pm}(r_0)$$

and

$$\left(\limsup_{r\to r_0} D_{u,\pm}(r)\right)^c \cap D^c_{u,\pm}(r_0) = \left(\limsup_{r\to r_0} D_{u,\pm}(r)\right)^c \supset D^c_{u,\pm}(r_0) \cap Q^c_0.$$

Combining all of the above, we have for any $\boldsymbol{y} \in Q^c_0$, $\lim_{r\to r_0} I_{D_{u,\pm}(r)}(\boldsymbol{y}) = I_{D_{u,\pm}(r_0)}(\boldsymbol{y})$. This completes the proof of the lemma.

□

Definition 4.5.5. For any $n \geq 1$, a function $g : \mathbb{R}^n \to \mathbb{R}$ is said to be **regular** if it can be written in the form:

$$g(\boldsymbol{y}) = |\alpha(\boldsymbol{y})| \exp(-\beta(\boldsymbol{y})),$$

where $\alpha(\boldsymbol{y})$ is a non-zero polynomial of $y_1, \ldots, y_n$ and $\beta(\boldsymbol{y}) = \boldsymbol{y}^T \boldsymbol{\Sigma} \boldsymbol{y}$ for some positive-definite matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n \times n}$.

Remark 4.5.6. It is easy to check that a regular function on $\mathbb{R}^n$ must be non-negative, bounded and integrable. Moreover, the integral of a regular function on a non-empty open set must be positive.

Let X be qualified under perturbation. For any $r \geq 0$ and $\boldsymbol{y} \in \mathbb{R}^L$, define

$$g_r(\boldsymbol{y}) := |h_r(\boldsymbol{y})|\, p_L(\boldsymbol{y}), \tag{4.46}$$

where $h_r(\boldsymbol{y})$ is as defined in (4.38)-(4.41) and

$$p_L(\boldsymbol{y}) := \frac{1}{(2\pi)^{L/2}} \exp\left(-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{y}\right). \tag{4.47}$$

For any $r \geq 0$, since $h_r(\boldsymbol{y})$ is a nonzero polynomial of $y_1, \ldots, y_L$, $g_r(\boldsymbol{y})$ is a regular function of $\boldsymbol{y} \in \mathbb{R}^L$. In addition, since $\boldsymbol{A}(r)$ is continuous on $r \in [0, \delta_{pc}]$ (where δ_{pc} is as defined in Definition 4.3.6), all the coefficients of the polynomial h_r are continuous, and then uniformly bounded on $r \in [0, \delta_{pc}]$. Then by the triangle inequality, there exists a constant $C > 0$ such that for any $r \in [0, \delta_{pc}]$,

$$h_r(\boldsymbol{y}) \leq C \sum_{v_1,\ldots,v_N \in \{1,\ldots,L\}} |y_{v_1} \cdots y_{v_N}|\,.$$

Thus, for any $r \in [0, \delta_{pc}]$ and $\boldsymbol{y} \in \mathbb{R}^L$,

$$g_r(\boldsymbol{y}) \leq C \sum_{v_1,\ldots,v_N \in \{1,\ldots,L\}} |y_{v_1} \cdots y_{v_N}|\, p_L(\boldsymbol{y}), \tag{4.48}$$

where the right-hand side is a finite sum of regular functions.

4.5.3 Proof of Theorem 4.5.1

Proof. Fix $u \in \mathbb{R}$. Since X is isotropic, we only need to show

$$\lim_{r \to 0} \frac{f_{u,+}(\boldsymbol{u}_0 r)}{f_{u,-}(\boldsymbol{u}_0 r)} = 1.$$

For any $r > 0$, since $\boldsymbol{\Sigma(r)}$ is positive-definite and $\boldsymbol{\Sigma(r)} = \boldsymbol{A}(r)\boldsymbol{A}^T(r)$, $\boldsymbol{A}(r)$ is invertible. By the change of variable $\boldsymbol{y} = \boldsymbol{A}^{-1}(r)(\boldsymbol{x}'', x, z)^T$, (4.31) and (4.47), we have

$$\begin{aligned} p_{\boldsymbol{t}}(\boldsymbol{x}'', x, z|\boldsymbol{0}, \boldsymbol{0}) &= \frac{1}{\sqrt{(2\pi)^L \det(\boldsymbol{\Sigma}(\boldsymbol{t}))}} \exp\left(-\frac{1}{2}(\boldsymbol{x}'', x, z)\boldsymbol{\Sigma}(\boldsymbol{t})^{-1}(\boldsymbol{x}'', x, z)^T\right) \\ &= \frac{1}{\sqrt{(2\pi)^L \det(\boldsymbol{\Sigma}(\boldsymbol{t}))}} \exp\left(-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{y}\right) \\ &= \frac{1}{\sqrt{\det(\boldsymbol{\Sigma}(\boldsymbol{t}))}} p_L(\boldsymbol{y}). \end{aligned}$$

Then the ratio becomes

$$\begin{aligned} \frac{f_{u,+}(\boldsymbol{u}_0 r)}{f_{u,-}(\boldsymbol{u}_0 r)} &= \frac{\int_{x,z>u}\int_{\det(\text{Matri}_N(\boldsymbol{x}''))>0} |\det(\text{Matri}_N(\boldsymbol{x}''))|\, p_{\boldsymbol{t}}(\boldsymbol{x}'', x, z|\boldsymbol{0}, \boldsymbol{0}) d\boldsymbol{x}'' dx dz}{\int_{x,z>u}\int_{\det(\text{Matri}_N(\boldsymbol{x}''))<0} |\det(\text{Matri}_N(\boldsymbol{x}''))|\, p_{\boldsymbol{t}}(\boldsymbol{x}'', x, z|\boldsymbol{0}, \boldsymbol{0}) d\boldsymbol{x}'' dx dz} \\ &= \frac{\int_{D_{u,+}(r)} |\det(\text{Matri}_N(\boldsymbol{A}(r)\boldsymbol{y}))|\, p_L(\boldsymbol{y}) d\boldsymbol{y}}{\int_{D_{u,-}(r)} |\det(\text{Matri}_N(\boldsymbol{A}(r)\boldsymbol{y}))|\, p_L(\boldsymbol{y}) d\boldsymbol{y}} \\ &= \frac{\int_{D_{u,+}(r)} |r^{-1}\det(\text{Matri}_N(\boldsymbol{A}(r)\boldsymbol{y}))|\, p_L(\boldsymbol{y}) d\boldsymbol{y}}{\int_{D_{u,-}(r)} |r^{-1}\det(\text{Matri}_N(\boldsymbol{A}(r)\boldsymbol{y}))|\, p_L(\boldsymbol{y}) d\boldsymbol{y}} \\ &= \frac{\int_{D_{u,+}(r)} |h_r(\boldsymbol{y})| p_L(\boldsymbol{y}) d\boldsymbol{y}}{\int_{D_{u,-}(r)} |h_r(\boldsymbol{y})| p_L(\boldsymbol{y}) d\boldsymbol{y}} \\ &= \frac{\int_{D_{u,+}(r)} g_r(\boldsymbol{y}) d\boldsymbol{y}}{\int_{D_{u,-}(r)} g_r(\boldsymbol{y}) d\boldsymbol{y}} \\ &= \frac{\int_{\mathbb{R}^L} g_r(\boldsymbol{y}) I_{D_{u,+}(r)}(\boldsymbol{y}) d\boldsymbol{y}}{\int_{\mathbb{R}^L} g_r(\boldsymbol{y}) I_{D_{u,-}(r)}(\boldsymbol{y}) d\boldsymbol{y}}, \end{aligned} \tag{4.49}$$

where $D_{u,\pm}(r)$ are the same as in Lemma 4.5.4 (see also (4.43)), h_r, g_r and p_L are as defined in (4.39), (4.46) and (4.47).

Note that by (iii) of Lemma 4.5.4, for any $r_0 \in [0, \delta_{pc}]$, $g_r(\boldsymbol{y})I_{D_{u,\pm}(r)}(\boldsymbol{y})$ converges almost everywhere on $\mathbb{R}^L$ to $g_{r_0}(\boldsymbol{y})I_{D_{u,\pm}(r_0)}(\boldsymbol{y})$ as $r \to r_0$, and $g_r(\boldsymbol{y})I_{D_{u,\pm}(r)}(\boldsymbol{y})$ is dominated by the right-hand side of (4.48) which does not depend on r and is integrable over $\mathbb{R}^L$. Then by the dominated convergence theorem, we have

$$\lim_{r\to r_0} \int_{\mathbb{R}^L} g_r(\boldsymbol{y}) I_{D_{u,\pm}(r)}(\boldsymbol{y}) d\boldsymbol{y} = \int_{\mathbb{R}^L} g_{r_0}(\boldsymbol{y}) I_{D_{u,\pm}(r_0)}(\boldsymbol{y}) d\boldsymbol{y} = \int_{D_{u,\pm}(r_0)} g_{r_0}(\boldsymbol{y}) d\boldsymbol{y}.$$

Moreover, by (i) and (ii) of Lemma 4.5.4 and Remark 4.5.6, we have

$$\int_{D_{u,+}(0)} g_0(\boldsymbol{y})d\boldsymbol{y} = \int_{D_{u,-}(0)} g_0(\boldsymbol{y})d\boldsymbol{y} > 0.$$

Then

$$\lim_{r\to 0} \frac{f_{u,+}(\boldsymbol{u}_0 r)}{f_{u,-}(\boldsymbol{u}_0 r)} = \frac{\lim_{r\to 0}\int_{D_{u,+}(r)} g_r(\boldsymbol{y})d\boldsymbol{y}}{\lim_{r\to 0}\int_{D_{u,-}(r)} g_r(\boldsymbol{y})d\boldsymbol{y}} = \frac{\int_{D_{u,+}(0)} g_0(\boldsymbol{y})d\boldsymbol{y}}{\int_{D_{u,-}(0)} g_0(\boldsymbol{y})d\boldsymbol{y}} = 1.$$

This completes the proof of the theorem. □

Remark 4.5.7. In fact, the result in Theorem 4.5.1 also holds for $N = 1$. Following the proof of Lemma 4.3.5, one can easily check that the expression of $\boldsymbol{\Sigma}_0$ and $\boldsymbol{\Sigma}_2$ in Lemma 4.3.5 also holds for $N = 1$. Then for $N = 1$, all the results in Section 4.4.1 hold, and we have

$$\boldsymbol{\Sigma}_0 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & a & a \\ 0 & a & a \end{pmatrix},$$

where $a := 1 - \frac{1}{3}\rho^{(1)}(0)^2\rho^{(2)}(0)^{-1}$, and by Lemma 4.2.1, we have $a > 0$. By solving the equation $\det(\boldsymbol{\Sigma}_0 - \lambda \boldsymbol{I}_3) = 0$ for any $\lambda \in \mathbb{R}$, we have

$$\boldsymbol{\Lambda}_0 = \mathrm{diag}\,(2a, 0, 0),$$

which implies $\boldsymbol{A}_0^{(2)} = \boldsymbol{A}_0^{(3)} = \boldsymbol{0}_3$. Then by $\boldsymbol{\Sigma}_0 = \boldsymbol{A}_0\boldsymbol{A}_0^T$, we have

$$\boldsymbol{A}_0 = \begin{pmatrix} 0 & 0 & 0 \\ \sqrt{a} & 0 & 0 \\ \sqrt{a} & 0 & 0 \end{pmatrix}.$$

Thus,

$$\boldsymbol{A}(r)_{(1)} = \boldsymbol{A}_1 r + o(r).$$

As $N = 1$, $\mathrm{Matri}_N(\boldsymbol{A}(r)\boldsymbol{y}) = \boldsymbol{A}(r)_{(1)}\boldsymbol{y}$. Then by Remark 4.4.2,

$$\lim_{r\to 0} r^{-1}\det\,(\mathrm{Matri}_N(\boldsymbol{A}(r)\boldsymbol{y})) = \lim_{r\to 0}\frac{1}{r}\boldsymbol{A}(r)_{(1)}\boldsymbol{y} = (\boldsymbol{A}_1)_{(1)}\boldsymbol{y}.$$

Note that by Remark 4.4.2,

$$\boldsymbol{A}_1^{(i)} = \begin{cases} \lambda_{i,0}^{1/2}\boldsymbol{P}_1^{(i)}, & i = 1; \\ \lambda_{i,2}^{1/2}\boldsymbol{P}_0^{(i)}, & i = 2, 3. \end{cases}$$

Thus,

$$\lim_{r\to 0} r^{-1}\det\left(\text{Matri}_N(\boldsymbol{A}(r)\boldsymbol{y})\right) = \lambda_{1,0}^{1/2}\boldsymbol{P}_1[1,1]y_1 + \lambda_{2,2}^{1/2}\boldsymbol{P}_0[1,2]y_2 + \lambda_{3,2}^{1/2}\boldsymbol{P}_0[1,3]y_3. \quad (4.50)$$

Note that the dimension of the eigenspace of $2a$, as an eigenvalue of $\boldsymbol{\Sigma}_0$, is one. Thus, an eigenvector of $2a$ must have the form $k\boldsymbol{P}_0^{(1)}$ for some $k \neq 0$. By (iv) of Lemma 4.4.1, $\boldsymbol{P}_1^{(1)}$ is either $\mathbf{0}_3$ or an eigenvector of $2a$. Then by (i) of Lemma 4.4.1, we have

$$\boldsymbol{P}_1^{(1)} = \mathbf{0}_3.$$

Note that 0 is an eigenvalue of $\boldsymbol{\Lambda}_0$ with multiplicity $N+1$ for $N=1$. In addition, since $\boldsymbol{\Sigma}_0^{(1)} = \mathbf{0}_3$, $\boldsymbol{\Sigma}_0^{(2)} = \boldsymbol{\Sigma}_0^{(3)}$ and $\boldsymbol{\Sigma}_2^{(2)} = \boldsymbol{\Sigma}_2^{(3)}$, we can also get

$$\det\left(\boldsymbol{\Sigma}(r)\right) = o\left(r^{2N+2}\right)$$

for $N=1$. Then we can follow the proof of Lemma 4.4.6 to show that it also holds for $N=1$. This implies $\lambda_{2,2} > 0$ and $\lambda_{3,2} = 0$.

Therefore, (4.50) becomes

$$\lim_{r\to 0} r^{-1}\det\left(\text{Matri}_N(\boldsymbol{A}(r)\boldsymbol{y})\right) = \lambda_{2,2}^{1/2}\boldsymbol{P}_0[1,2]y_2.$$

Note that by Lemma 4.3.5,

$$\boldsymbol{\Sigma}_2 = \begin{pmatrix} 18\alpha - 30\beta & \alpha' - \frac{5}{3}\beta' & \alpha' - \frac{5}{3}\beta' \\ \alpha' - \frac{5}{3}\beta' & -b & -b \\ \alpha' - \frac{5}{3}\beta' & -b & -b \end{pmatrix},$$

where $b := \frac{1}{6}\rho^{(1)}(0) - \frac{5}{18}\rho^{(1)}(0)^2\rho^{(2)}(0)^{-2}\rho^{(3)}(0)$, and by (4.2), we have $b > 0$. Since $\boldsymbol{\Sigma}_0\boldsymbol{P}_0^{(3)} = 0$, we have

$$\boldsymbol{P}_0[2,3] + \boldsymbol{P}_0[3,3] = 0.$$

Then by (iii) of Lemma 4.4.1,

$$\begin{aligned} 0 = \lambda_{3,2} &= \left(\boldsymbol{P}_0^{(3)}\right)^T \boldsymbol{\Sigma}_2 \boldsymbol{P}_0^{(3)} \\ &= (18\alpha - 30\beta)\boldsymbol{P}_0[1,3]^2 + 2\left(\alpha' - \frac{5}{3}\beta'\right)\boldsymbol{P}_0[1,3](\boldsymbol{P}_0[2,3] + \boldsymbol{P}_0[3,3]) \\ &\quad - b(\boldsymbol{P}_0[2,3] + \boldsymbol{P}_0[3,3])^2 \\ &= (18\alpha - 30\beta)\boldsymbol{P}_0[1,3]^2. \end{aligned}$$

Since $18\alpha - 30\beta > 0$, we have $\boldsymbol{P}_0[1,3] = 0$ (and thus, Lemma 4.4.7 also holds for $N = 1$). By $(\boldsymbol{\Sigma}_0 - 2a\boldsymbol{I})\boldsymbol{P}_0^{(1)} = 0$ and $a \neq 0$, we have $\boldsymbol{P}_0[1,1] = 0$. Thus, if $\boldsymbol{P}_0[1,2] = 0$, then from the above, we can get $(\boldsymbol{P}_0)_{(1)} = (0,0,0)$. This implies $\boldsymbol{P}_0$ is degenerate, resulting in a contradiction. Thus, we have

$$\boldsymbol{P}_0[1,2] \neq 0.$$

Therefore, for $N = 1$, we can define

$$h_0(\boldsymbol{y}) := \lambda_{2,2}^{1/2}\boldsymbol{P}_0[1,2]y_2$$

which is a non-zero polynomial of $\boldsymbol{y}$ with degree one. Then we can also define $D_{u,\pm}(r)$ as in (4.43). Since $(y_1, y_2, y_3) \in D_{u,\pm}(0)$ is equivalent to $(y_1, -y_2, -y_3) \in D_{u,\mp}(0)$, one can follow the proof of Lemma 4.5.4 and show that it also holds for $N = 1$. The remaining proof is the same as that of Theorem 4.5.1.

4.6 Asymptotic Behavior as $u \to \infty$

4.6.1 Main Result 2

Let X be qualified under perturbation. Recall that Condition (3) in Definition 4.3.2 implies $\rho(x)$ is four times continuously differentiable on $[0, \delta_\rho^2]$, which makes $\boldsymbol{\Sigma}(r)$ continuous on $r \in [0, \delta_\rho]$ (see Lemma 4.3.5). Indeed, for any $\tilde{\delta}_\rho > 0$ such that

$$\rho(x) \text{ is four times continuously differentiable on } \left[0, \tilde{\delta}_\rho^2\right], \tag{4.51}$$

we can show that $\boldsymbol{\Sigma}(r)$ is continuous on $r \in [0, \tilde{\delta}_\rho]$ by a similar proof of Lemma 4.3.5.

Suppose that (4.51) holds for some $\tilde{\delta}_\rho > 0$. For any $r \geq 0$, let $\widetilde{\boldsymbol{A}}(r)$ be the non-negative square root of $\boldsymbol{\Sigma}(r)$. Recall that this means $\widetilde{\boldsymbol{A}}(r)$ is the unique positive semi-definite matrix such that

$$\boldsymbol{\Sigma}(r) = \widetilde{\boldsymbol{A}}(r)\widetilde{\boldsymbol{A}}^T(r).$$

Then by Lemma 4.2.2, $\widetilde{\boldsymbol{A}}(r)$ is continuous on $r \in (0, \tilde{\delta}_\rho]$. In addition, for any $r \geq 0$, we can get

$$\widetilde{\boldsymbol{A}}(r) = \boldsymbol{A}(r)\boldsymbol{P}^T(r). \tag{4.52}$$

Then by Definition 4.3.6,

$$\lim_{r\to 0}\widetilde{\boldsymbol{A}}(r) = \widetilde{\boldsymbol{A}}(0).$$

Thus, $\widetilde{\boldsymbol{A}}(r)$ is continuous on $r \in [0, \tilde{\delta}_\rho]$.

To establish our second main result, we also need

$$\boldsymbol{\Sigma}(r)[k + k(k-1)/2, i] < 0$$

for any $1 \le k \le N-1$, $r \ge 0$ and $i = L-1, L$. Note that by Lemma 4.3.5, $\boldsymbol{\Sigma}(0)[k+k(k-1)/2, i] = \frac{4}{3}\rho^{(1)}(0) < 0$ for any $1 \le k \le N-1$ and $i = L-1, L$. Thus, it is equivalent to requiring that the above inequality to hold for any $r > 0$. By the proof of Lemma 4.3.5 (see (C.10), (C.11), (C.46) and (C.47)), this is equivalent to

$$k_1(\boldsymbol{u}_0 r) > 0 \text{ and } 1 - k_1(\boldsymbol{u}_0 r)k_*(\boldsymbol{u}_0 r) - k_2^2(\boldsymbol{u}_0 r)r^4 > 0,$$

for any $r \in (0, \tilde{\delta}_\rho]$, where for any $\boldsymbol{t} \in \mathbb{R}^N$,

$$k_1(\boldsymbol{t}) = \frac{\rho^{(1)}(\|\boldsymbol{t}\|^2)}{\rho^{(1)}(0)}, \quad k_2(\boldsymbol{t}) = \frac{2\rho^{(2)}(\|\boldsymbol{t}\|^2)}{\rho^{(1)}(0)}, \text{ and } k_*(\boldsymbol{t}) = k_1(\boldsymbol{t}) + k_2(\boldsymbol{t})\|\boldsymbol{t}\|^2,$$

i.e., for any $x \in (0, \tilde{\delta}_\rho^2]$,

$$\rho^{(1)}(x) < 0 \text{ and } (\rho^{(1)}(x))^2 + 2\rho^{(1)}(x)\rho^{(2)}(x)x + 4(\rho^{(2)}(x))^2x^2 < (\rho^{(1)}(0))^2. \tag{4.53}$$

Example 4.6.1. *For any $a > 0$, we can show that condition (4.53) is satisfied when $\rho(x) = e^{-ax}$, $x \ge 0$. Since condition (4.53) is invariant under rescaling, it is equivalent to check this condition when $a = 1$. Note that $\rho^{(1)}(x) = -e^{-x} < 0$ and $\rho^{(2)}(x) = e^{-x}$, $x \ge 0$. Thus, it suffices to check for any $x > 0$,*

$$e^{2x} > 1 - 2x + 4x^2,$$

i.e., for any $t > 0$,

$$e^t > 1 - t + t^2.$$

Note that $e^t = \sum_{n=0}^{\infty} \frac{t^n}{n!}$. Thus, for any $t > 0$,

$$\begin{aligned} e^t &> 1 + t + \frac{1}{2}t^2 + \frac{1}{6}t^3 \\ &= (1 - t + t^2) + \left(2t + \frac{1}{6}t^3 - \frac{1}{2}t^2\right) \\ &> 1 - t + t^2, \end{aligned}$$

where the last inequality follows from the fact that $2t + \frac{1}{6}t^3 \ge \sqrt{\frac{1}{3}t^4} > \frac{1}{2}t^2$ for any $t > 0$.

Recall that by (4.32),

$$f_{u,k}(\boldsymbol{t}) := P(X(\mathbf{0}) > u)^{-1} p(\mathbf{0})^{-1} \int_{x,z>u} \int_{D_k} |\det(\boldsymbol{x}'')| p_{\boldsymbol{t}}(\boldsymbol{x}'', x, z | \mathbf{0}, \mathbf{0}) p_{\boldsymbol{t}}(\mathbf{0}, \mathbf{0}) d\boldsymbol{x}'' dx dz,$$

where

$$D_k := \Big\{ \boldsymbol{x} \in \mathbb{R}^{N(N+1)/2} : \text{Matri}_N(\boldsymbol{x}) \text{ is non-degenerate} \\ \text{and has exactly } k \text{ negative eigenvalues} \Big\}.$$

For any $0 \leq k \leq N$, $f_{u,k}(\boldsymbol{t})$ is positive and continuous (by the dominated convergence theorem) on $\boldsymbol{t} \in \mathbb{R}^N \setminus \{\mathbf{0}_N\}$.

For any $u > 0$ and $\boldsymbol{t} \in \mathbb{R}^N \setminus \{\mathbf{0}_N\}$, define

$$\Psi_u(\boldsymbol{t}) := \frac{\sum_{k=0}^{N-2} f_{u,k}(\boldsymbol{t})}{f_{u,N-1}(\boldsymbol{t}) + f_{u,N}(\boldsymbol{t})}.$$

Then $\Psi_u(\boldsymbol{t})$ is also continuous on $\boldsymbol{t} \in \mathbb{R}^N \setminus \{\mathbf{0}_N\}$. The following theorem describes the limiting behavior of $\Psi_u(\boldsymbol{t})$ as $\|\boldsymbol{t}\| \to 0$ or $u \to \infty$.

Theorem 4.6.1. *Let X be qualified under perturbation. Assume that X also satisfies (4.51) and (4.53) for some $\tilde{\delta}_\rho > 0$. Then we have*

(i) For any $u > 0$, the limit $\Psi_u(\mathbf{0}) := \lim_{\|\boldsymbol{t}\| \to 0} \Psi_u(\boldsymbol{t})$ exists (and thus, $\Psi_u(\boldsymbol{t})$ is well-defined and continuous on $\boldsymbol{t} \in \mathbb{R}^N$).

(ii) As $u \to \infty$,

$$\max_{\boldsymbol{t} \in B(\mathbf{0}_N, \tilde{\delta}_\rho)} \Psi_u(\boldsymbol{t}) \to 0.$$

4.6.2 Preparation for Main Result 2

By replacing $\boldsymbol{A}(r)$ in (4.42) with $\widetilde{\boldsymbol{A}}(r)$, we can similarly define

$$\widetilde{H}_{i,u}(r) := \left\{ \boldsymbol{y} \in \mathbb{R}^L : \widetilde{\boldsymbol{A}}(r)_{(i)} \boldsymbol{y} > u \right\}, \tag{4.54}$$

for any $r \geq 0$, $u > 0$ and $i = L-1, L$. It is noticeable that

- by (4.52) and (v) of Lemma 4.4.4, $\widetilde{H}_{L-1,u}(0)$ and $\widetilde{H}_{L,u}(0)$ coincide;

- by the continuity of $\boldsymbol{A}(r)$ on $r \in [0, \tilde{\delta}_\rho]$, for $i = L-1, L$,

$$\widetilde{H}_{i,u}(0) \subset \liminf_{r\to 0} \widetilde{H}_{i,u}(r) \subset \limsup_{r\to 0} \widetilde{H}_{i,u}(r) \subset \widetilde{H}_{i,u}(0) \cup \partial \widetilde{H}_{i,u}(0), \tag{4.55}$$

- for any $r > 0$, since $\boldsymbol{\Sigma}(r)$ is non-degenerate, the two $(L-1)$-dimensional hyper-planes $\partial \widetilde{H}_{L-1,u}(r)$ and $\partial \widetilde{H}_{L,u}(r)$ cannot be parallel, and thus, $\partial \widetilde{H}_{L-1,u}(r) \cap \partial \widetilde{H}_{L,u}(r) \neq \emptyset$.

By convention, the closure of a set $A \subset \mathbb{R}^L$ is denoted by $\overline{A}$. For any $r \geq 0$ and $u > 0$, let $V_u(r) := \widetilde{H}_{L-1,u}(r) \cap \widetilde{H}_{L,u}(r)$. Since $\overline{V}_u(r)$ is a convex set, the point in $\overline{V}_u(r)$ which minimizes the distance between the origin and a point in $\overline{V}_u(r)$ is unique. Thus, we can define

$$\hat{\boldsymbol{y}}_u(r) := \operatorname*{argmin}_{\boldsymbol{y} \in \overline{V}_u(r)} \|\boldsymbol{y}\|. \tag{4.56}$$

Let $\hat{\boldsymbol{y}}_{i,u}(r)$, $i = L-1, L$ be the projection of the origin on the $(L-1)$-dimensional hyper-plane $\partial \widetilde{H}_{i,u}(r)$, and let $\hat{\boldsymbol{y}}_{L-1,L,u}(r)$ be the projection of the origin on the hyper-plane $\partial \widetilde{H}_{L-1,u}(r) \cap \partial \widetilde{H}_{L,u}(r)$ (when $r > 0$, this hyper-plane is $(L-2)$-dimensional, where $L = N(N+1)/2 + 2 > 2$). Obviously, for any $r \geq 0$ and $u > 0$, we have

$$\hat{\boldsymbol{y}}_u(r) \in \left\{\hat{\boldsymbol{y}}_{L-1,u}(r), \hat{\boldsymbol{y}}_{L,u}(r), \hat{\boldsymbol{y}}_{L-1,L,u}(r)\right\}, \tag{4.57}$$

and in particular,

$$\hat{\boldsymbol{y}}_u(0) = \hat{\boldsymbol{y}}_{L-1,u}(0) = \hat{\boldsymbol{y}}_{L,u}(0). \tag{4.58}$$

Recall that the index of a critical point is defined to be the number of negative eigenvalues of its Hessian matrix.

Lemma 4.6.2. *Let X be qualified. Assume that X also satisfies (4.51) and (4.53) for some $\tilde{\delta}_\rho > 0$. Let $\hat{\boldsymbol{y}}_u(r)$, $\hat{\boldsymbol{y}}_{L-1,u}(r)$, $\hat{\boldsymbol{y}}_{L,u}(r)$ and $\hat{\boldsymbol{y}}_{L-1,L,u}(r)$ be as defined above. Then for any $u > 0$,*

(i) $\mathrm{Matri}_N(\widetilde{\boldsymbol{A}}(r)\hat{\boldsymbol{y}}_{i,u}(r))$*, $i = L-1, L$ has at least $N-1$ negative eigenvalues for any $r \in [0, \tilde{\delta}_\rho]$;*

(ii) $\mathrm{Matri}_N(\widetilde{\boldsymbol{A}}(r)\hat{\boldsymbol{y}}_{L-1,L,u}(r))$ *has at least $N-1$ negative eigenvalues for any $r \in (0, \tilde{\delta}_\rho]$.*

These, together with (4.57) and (4.58), imply that $\mathrm{Matri}_N(\widetilde{\boldsymbol{A}}(r)\hat{\boldsymbol{y}}_u(r))$ *has at least $N-1$ negative eigenvalues for any $r \in [0, \tilde{\delta}_\rho]$.*

Proof. For (i), fix $u > 0$ and $r \in [0, \tilde{\delta}_\rho]$. Since $\hat{\boldsymbol{y}}_{i,u}(r)$ is the projection of the origin on the hyper-plane $\partial \widetilde{H}_{i,u}(r)$ for $i = L-1, L$, it is easy to see

$$\hat{\boldsymbol{y}}_{i,u}(r) = \beta_{i,u}(r) \left(\widetilde{\boldsymbol{A}}(r)_{(i)} \right)^T$$

for some real number $\beta_{i,u}(r) \neq 0$. Then

$$0 < u = \widetilde{\boldsymbol{A}}(r)_{(i)} \hat{\boldsymbol{y}}_{i,u}(r) = \beta_{i,u}(r) \left\| \widetilde{\boldsymbol{A}}(r)_{(i)} \right\|^2 = \beta_{i,u}(r) \boldsymbol{\Sigma}(r)[i,i].$$

Since $\boldsymbol{\Sigma}(r)$ is positive semi-definite, we have $\beta_{i,u}(r) > 0$ for $i = L-1, L$, and then

$$\begin{aligned} \text{Matri}_N \left(\widetilde{\boldsymbol{A}}(r) \hat{\boldsymbol{y}}_{i,u}(r) \right) &= \beta_{i,u}(r) \text{Matri}_N \left(\widetilde{\boldsymbol{A}}(r) \left(\widetilde{\boldsymbol{A}}(r)_{(i)} \right)^T \right) \\ &= \beta_{i,u}(r) \text{Matri}_N \left(\boldsymbol{\Sigma}(r)^{(i)} \right). \end{aligned}$$

Note that by (i) and (ii) in Appendix C.4, (C.29) and (C.30), it is easy to check that $\text{Matri}_N(\boldsymbol{\Sigma}(r)^{(i)})$ is diagonal with the first $N-1$ diagonal elements equal to $\boldsymbol{\Sigma}(r)[1,i]$ for $i = L-1, L$. In addition, by Condition (4.53), we have $\boldsymbol{\Sigma}(r)[1,i] < 0$ for $i = L-1, L$. Thus, $\text{Matri}_N(\widetilde{\boldsymbol{A}}(r)\hat{\boldsymbol{y}}_{i,u}(r))$, $i = L-1, L$ has at least $N-1$ negative eigenvalues.

For (ii), fix $u > 0$ and $r \in (0, \tilde{\delta}_\rho]$. It is easy to see

$$\widetilde{\boldsymbol{A}}(r)_{(L-1)} \hat{\boldsymbol{y}}_{L-1,L,u}(r) = \widetilde{\boldsymbol{A}}(r)_{(L)} \hat{\boldsymbol{y}}_{L-1,L,u}(r) = u,$$

and for any $\boldsymbol{x} \in \mathbb{R}^L$ such that $\widetilde{\boldsymbol{A}}(r)_{(L-1)} \boldsymbol{x} = \widetilde{\boldsymbol{A}}(r)_{(L)} \boldsymbol{x} = u$,

$$\hat{\boldsymbol{y}}^T_{L-1,L,u}(r)(\boldsymbol{x} - \hat{\boldsymbol{y}}_{L-1,L,u}(r)) = 0.$$

This implies

$$\hat{\boldsymbol{y}}_{L-1,L,u}(r) = \beta'_{L-1,u}(r) \left(\widetilde{\boldsymbol{A}}(r)_{(L-1)} \right)^T + \beta'_{L,u}(r) \left(\widetilde{\boldsymbol{A}}(r)_{(L)} \right)^T,$$

for some constants $\beta'_{L-1,u}(r)$ and $\beta'_{L,u}(r)$. Then for $i = L-1, L$,

$$\begin{aligned} u &= \widetilde{\boldsymbol{A}}(r)_{(i)} \hat{\boldsymbol{y}}_{L-1,L,u}(r) \\ &= \widetilde{\boldsymbol{A}}(r)_{(i)} \left(\beta'_{L-1,u}(r) \left(\widetilde{\boldsymbol{A}}(r)_{(L-1)} \right)^T + \beta'_{L,u}(r) \left(\widetilde{\boldsymbol{A}}(r)_{(L)} \right)^T \right) \\ &= \beta'_{L-1,u}(r) \boldsymbol{\Sigma}(r)[i, L-1] + \beta'_{L,u}(r) \boldsymbol{\Sigma}(r)[i, L], \end{aligned}$$

i.e.,

$$\begin{pmatrix} \beta'_{L-1,u}(r) \\ \beta'_{L,u}(r) \end{pmatrix} = \left(\mathbf{\Sigma}(r)[(L-1):L,(L-1):L]\right)^{-1} \begin{pmatrix} u \\ u \end{pmatrix}. \tag{4.59}$$

Denote

$$a(r) := \mathbf{\Sigma}(r)[L-1, L-1] = \mathbf{\Sigma}(r)[L, L],$$

where the equality comes from the symmetry between $X(\mathbf{0})$ and $X(u_0 r)$ in the definition of $\mathbf{\Sigma}(r)$, and

$$b(r) := \mathbf{\Sigma}(r)[L-1, L] = \mathbf{\Sigma}(r)[L, L-1].$$

Since $\mathbf{\Sigma}(r)[(L-1):L,(L-1):L]$ is positive-definite, we have

$$a(r) > 0 \text{ and } a^2(r) - b^2(r) > 0.$$

Then by (4.59), we have

$$\beta'_{L-1,u}(r) = \beta'_{L,u}(r) = \frac{u}{a^2(r) - b^2(r)}(a(r) - b(r)) = \frac{u}{a(r) + b(r)} > 0.$$

Thus,

$$\begin{aligned}
&\text{Matri}_N\left(\widetilde{\boldsymbol{A}}(r)\hat{\boldsymbol{y}}_{L-1,L,u}(r)\right)\\
=&\frac{u}{a(r)+b(r)}\text{Matri}_N\left(\widetilde{\boldsymbol{A}}(r)\left(\widetilde{\boldsymbol{A}}(r)_{(L-1)} + \widetilde{\boldsymbol{A}}(r)_{(L)}\right)^T\right)\\
=&\frac{u}{a(r)+b(r)}\left(\text{Matri}_N\left(\mathbf{\Sigma}(r)^{(L-1)}\right) + \text{Matri}_N\left(\mathbf{\Sigma}(r)^{(L)}\right)\right)\\
=&\frac{u}{a(r)+b(r)}\left(\beta^{-1}_{L-1,u}(r)\text{Matri}_N(\widetilde{\boldsymbol{A}}(r)\hat{\boldsymbol{y}}_{L-1,u}(r)) + \beta^{-1}_{L,u}(r)\text{Matri}_N(\widetilde{\boldsymbol{A}}(r)\hat{\boldsymbol{y}}_{L,u}(r))\right),
\end{aligned}$$

which is diagonal and has at least $N-1$ negative eigenvalues by the proof of (i), and hence proved. □

Recall that in Section 4.5.1, we have defined for any $0 \le k \le N$,

$$\begin{aligned}D_k := \Big\{\boldsymbol{x} \in \mathbb{R}^{N(N+1)/2} : &\text{Matri}_N(\boldsymbol{x}) \text{ is non-degenerate}\\ &\text{and has exactly } k \text{ negative eigenvalues}\Big\}.\end{aligned}$$

For any $0 \le k \le N$ and $r > 0$, define

$$\widetilde{G}_k(r) := \left\{\boldsymbol{y} \in \mathbb{R}^L : \widetilde{\boldsymbol{A}}(r)\boldsymbol{y} \in \bigcup_{i=k}^{N} D_i\right\},$$

and let

$$\widetilde{G}_k(0) := \liminf_{r\to 0} \widetilde{G}_k(r).$$

Then for any $u > 0$, $0 \le k \le N$ and $r \ge 0$, we can further define

$$\widetilde{D}_{u,k}(r) := \widetilde{H}_{L-1,u}(r) \cap \widetilde{H}_{L,u}(r) \cap \widetilde{G}_k(r). \tag{4.60}$$

The following lemma describes the behavior of $\widetilde{D}_{u,k}(r)$ as $r \to r_0$ for any $r_0 \in [0, \tilde{\delta}_\rho]$, which is important for our second main result.

Lemma 4.6.3. *Let X be qualified under perturbation and satisfy (4.53) for some $\tilde{\delta}_\rho > 0$. Then for any $u > 0$, we have*

(i) $\widetilde{D}_{u,k}(0)$ has a non-empty interior for any $0 \le k \le N-1$;

(ii) for any $r_0 \in [0, \tilde{\delta}_\rho]$ and $0 \le k \le N$,

$$\lim_{r\to r_0} I_{\widetilde{D}_{u,k}(r)}(\boldsymbol{y}) = I_{\widetilde{D}_{u,k}(r_0)}(\boldsymbol{y}) \text{ for almost all } \boldsymbol{y} \in \mathbb{R}^L.$$

Proof. Fix $u > 0$. Without loss of generality, we can assume $\tilde{\delta}_\rho \le \delta_\rho$. For (i), since $\widetilde{D}_{u,j}(0) \subset \widetilde{D}_{u,i}(0)$ for any $0 \le i \le j \le N-1$, it suffices to show that $\widetilde{D}_{u,N-1}(0)$ contains a non-empty open set. By (i) of Lemma 4.6.2, $\mathrm{Matri}_N(\widetilde{\boldsymbol{A}}(0)\hat{\boldsymbol{y}}_{L,u}(0))$ has at least $N-1$ negative eigenvalues. Note that by Lemma 4.2.3 and the continuity of $\widetilde{\boldsymbol{A}}(r)$ on $r \in [0, \tilde{\delta}_\rho]$, the eigenvalues of $\mathrm{Matri}_N(\widetilde{\boldsymbol{A}}(r)\boldsymbol{y})$ are continuous on $(r, \boldsymbol{y}) \in [0, \tilde{\delta}_\rho] \times \mathbb{R}^L$. Thus, there exist constants $\delta > 0$ and $\gamma > 0$, such that $\mathrm{Matri}_N(\widetilde{\boldsymbol{A}}(r)\boldsymbol{y})$ has at least $N-1$ negative eigenvalues for any $r \in [0, \delta)$ and $\boldsymbol{y} \in B(\hat{\boldsymbol{y}}_{L,u}(0), \gamma)$, the L-dimensional open ball centered at $\hat{\boldsymbol{y}}_{L,u}(0)$ with radius γ. Note that for any $\boldsymbol{y}' \in \mathbb{R}^L$, if there exists $r'_n \downarrow 0$ such that

$$\det\left(\mathrm{Matri}_N\left(\widetilde{\boldsymbol{A}}\left(r'_n\right)\boldsymbol{y}'\right)\right) = 0 \text{ for any } n \ge 1,$$

then by (4.38)-(4.41), (4.52) and the continuity of $\boldsymbol{P}(r)$ on $r \in [0, \delta_{pc}]$,

$$h_0\left(\boldsymbol{P}^T(0)\boldsymbol{y}'\right) = \lim_{n\to\infty} h_{r'_n}\left(\boldsymbol{P}^T(r'_n)\boldsymbol{y}'\right) = \lim_{n\to\infty} \frac{1}{r'_n}\det\left(\mathrm{Matri}_N\left(\boldsymbol{A}\left(r'_n\right)\boldsymbol{P}^T(r'_n)\boldsymbol{y}'\right)\right) = 0.$$

Then by the orthogonality of $\boldsymbol{P}(0)$, we have

$$\boldsymbol{y}' \in Q_h := \left\{\boldsymbol{P}(0)\boldsymbol{y} : \boldsymbol{y} \in \mathbb{R}^L, h_0(\boldsymbol{y}) = 0\right\}.$$

Combining all of the above, we have

$$B(\hat{\boldsymbol{y}}_{L,u}(0), \gamma) \setminus Q_h \subset \widetilde{G}_{N-1}(0).$$

Then by (4.55),

$$\begin{aligned}\widetilde{D}_{u,N-1}(0) &= \liminf_{r\to 0} \widetilde{D}_{u,N-1}(r) \\ &= \liminf_{r\to 0} \widetilde{H}_{L-1,u}(r) \cap \liminf_{r\to 0} \widetilde{H}_{L,u}(r) \cap \widetilde{G}_{N-1}(0) \\ &\supset \left(\widetilde{H}_{L,u}(0) \cap B(\hat{\boldsymbol{y}}_{L,u}(0), \gamma)\right) \setminus Q_h.\end{aligned}$$

Since $\widetilde{H}_{L,u}(0)$ is open and $\hat{\boldsymbol{y}}_{L,u}(0) \in \partial\widetilde{H}_{L,u}(0)$, we have $\widetilde{H}_{L,u}(0) \cap B(\hat{\boldsymbol{y}}_{L,u}(0), \gamma)$ is a non-empty open set. Since $h_0(\boldsymbol{y})$ is a non-degenerate homogeneous polynomial of $y_1, \ldots, y_L$ with degree N, we have Q_h is closed and $\lambda_L(Q_h) = 0$. Thus, $(\widetilde{H}_{L,u}(0) \cap B(\hat{\boldsymbol{y}}_{L,u}(0), \gamma)) \setminus Q_h$ is also a non-empty open set, and hence (i) is proved.

As for (ii), we also fix $0 \le k \le N$. Since the eigenvalues of $\text{Matri}_N(\widetilde{\boldsymbol{A}}(r)\boldsymbol{y})$ are all continuous on $(r, \boldsymbol{y}) \in [0, \tilde{\delta}_\rho] \times \mathbb{R}^L$, we have

$$\widetilde{D}_{u,k}(0) \subset \liminf_{r\to 0} \widetilde{D}_{u,k}(r) \subset \limsup_{r\to 0} \widetilde{D}_{u,k}(r) \subset \widetilde{D}_{u,k}(0) \cup \partial\widetilde{H}_{L,u}(0) \cup Q_h. \tag{4.61}$$

The rest of the proof is only an analog of the proof of (iii) of Lemma 4.5.4. □

The following lemma is the last preparation for the second main result.

Lemma 4.6.4. *For any $a, b > 0$*

$$\sup_{k_1 > a, k_2 - k_1 > b} e^{\frac{1}{2}k_1^2 u^2} \int_{\|\boldsymbol{y}'\| \ge k_2 u} \exp\left(-\frac{1}{2}\boldsymbol{y}'^T \boldsymbol{y}'\right) d\boldsymbol{y}' \to 0 \text{ as } u \to \infty.$$

Proof. By the change of variable for spherical coordinates, we get

$$\int_{\|\boldsymbol{y}'\| \ge k_2 u} \exp\left(-\frac{1}{2}\boldsymbol{y}'^T \boldsymbol{y}'\right) d\boldsymbol{y}' = C\left(\int_{k_2 u}^{\infty} e^{-\frac{1}{2}r^2} r^{L-1} dr\right),$$

where C is positive and independent of k_2. Thus, we have

$$\begin{aligned} e^{\frac{1}{2}k_1^2u^2}\int_{\|\boldsymbol{y}'\|\geq k_2u}\exp\left(-\frac{1}{2}\boldsymbol{y}'^T\boldsymbol{y}'\right)d\boldsymbol{y}' &= C\left(\int_{k_2u}^{\infty}e^{-\frac{1}{2}(r^2-k_1^2u^2)}r^{L-1}dr\right)\\ &= C\left(\int_{k_2u}^{\infty}e^{-\frac{1}{2}(r+k_1u)(r-k_1u)}r^{L-1}dr\right)\\ &\leq C\left(\int_{k_2u}^{\infty}e^{-\frac{1}{2}(r+k_1u)bu}r^{L-1}dr\right)\\ &= Ce^{-\frac{1}{2}k_1bu^2}\left(\int_{k_2u}^{\infty}e^{-\frac{1}{2}bur}r^{L-1}dr\right)\\ &\leq Ce^{-\frac{1}{2}abu^2}\left(\int_{(a+b)u}^{\infty}e^{-\frac{1}{2}bur}r^{L-1}dr\right), \end{aligned}$$

which is independent of k_1, k_2, and converges to 0 as u tends to infinity.

□

4.6.3 Proof of Theorem 4.6.1

Proof. The proof of (i) is similar to that of Theorem 4.5.1. Fix $u > 0$. Since X is isotropic, it suffices to show that $\Psi_u(\boldsymbol{u}_0 r)$ converges as $r \to 0$. By the change of variable

$$\boldsymbol{y} = \widetilde{\boldsymbol{A}}^{-1}(r)(\boldsymbol{x}'', x, z)^T,$$

and (4.52), we have for any $r > 0$,

$$\begin{aligned}\Psi_u(\boldsymbol{u}_0 r) &= \frac{\sum_{k=0}^{N-2} f_{u,k}(\boldsymbol{u}_0 r)}{f_{u,N-1}(\boldsymbol{u}_0 r) + f_{u,N}(\boldsymbol{u}_0 r)} \\ &= \frac{\int_{x,z>u}\int_{\cup_{k=0}^{N-2} D_k} |\det(\boldsymbol{x}'')| p_{\boldsymbol{t}}(\boldsymbol{x}'', x, z|\boldsymbol{0},\boldsymbol{0}) p_{\boldsymbol{t}}(\boldsymbol{0},\boldsymbol{0}) d\boldsymbol{x}'' dx dz,}{\int_{x,z>u}\int_{D_{N-1}\cup D_N} |\det(\boldsymbol{x}'')| p_{\boldsymbol{t}}(\boldsymbol{x}'', x, z|\boldsymbol{0},\boldsymbol{0}) p_{\boldsymbol{t}}(\boldsymbol{0},\boldsymbol{0}) d\boldsymbol{x}'' dx dz} \\ &= \frac{\int_{\widetilde{D}_{u,0}(r)\setminus \widetilde{D}_{u,N-1}(r)} \left| r^{-1}\det\left(\text{Matri}_N\left(\widetilde{\boldsymbol{A}}(r)\boldsymbol{y}\right)\right)\right| p_L(\boldsymbol{y}) d\boldsymbol{y}}{\int_{\widetilde{D}_{u,N-1}(r)} \left| r^{-1}\det\left(\text{Matri}_N\left(\widetilde{\boldsymbol{A}}(r)\boldsymbol{y}\right)\right)\right| p_L(\boldsymbol{y}) d\boldsymbol{y}} \\ &= \frac{\int_{\widetilde{D}_{u,0}(r)\setminus \widetilde{D}_{u,N-1}(r)} \left| r^{-1}\det\left(\text{Matri}_N\left(\boldsymbol{A}(r)\boldsymbol{P}^T(r)\boldsymbol{y}\right)\right)\right| p_L\left(\boldsymbol{P}^T(r)\boldsymbol{y}\right) d\boldsymbol{y}}{\int_{\widetilde{D}_{u,N-1}(r)} \left| r^{-1}\det\left(\text{Matri}_N\left(\boldsymbol{A}(r)\boldsymbol{P}^T(r)\boldsymbol{y}\right)\right)\right| p_L\left(\boldsymbol{P}^T(r)\boldsymbol{y}\right) d\boldsymbol{y}} \\ &= \frac{\int_{\widetilde{D}_{u,0}(r)\setminus \widetilde{D}_{u,N-1}(r)} g_r\left(\boldsymbol{P}^T(r)\boldsymbol{y}\right) d\boldsymbol{y}}{\int_{\widetilde{D}_{u,N-1}(r)} g_r\left(\boldsymbol{P}^T(r)\boldsymbol{y}\right) d\boldsymbol{y}} \\ &= \frac{\int_{\widetilde{D}_{u,0}(r)\setminus \widetilde{D}_{u,N-1}(r)} \tilde{g}_r(\boldsymbol{y}) d\boldsymbol{y}}{\int_{\widetilde{D}_{u,N-1}(r)} \tilde{g}_r(\boldsymbol{y}) d\boldsymbol{y}},\end{aligned} \tag{4.62}$$

where $\widetilde{D}_{u,k}(r)$, $0 \le k \le N-1$ are the same as in Lemma 4.6.3 (see also (4.60)), g_r and p_L are as defined in (4.46) and (4.47), and for any $\boldsymbol{y} \in \mathbb{R}^L$ and $r \ge 0$,

$$\tilde{g}_r(\boldsymbol{y}) := g_r\left(\boldsymbol{P}^T(r)\boldsymbol{y}\right). \tag{4.63}$$

The fourth equality follows from the fact that $p_L(y) = p_L(\boldsymbol{P}^T(r)y)$. Then we can use (ii) of Lemma 4.6.3, which plays the role of (iii) of Lemma 4.5.4 in the proof of Theorem 4.5.1, to show that $\int_{\widetilde{D}_{u,k}(r)} \tilde{g}_r(\boldsymbol{y}) d\boldsymbol{y}$, $0 \le k \le N-1$ are continuous functions of $r \in [0, \tilde{\delta}_\rho]$. In addition, by Remark 4.5.6 and (i) of Lemma 4.6.3, we have for any $0 \le k \le N-1$,

$$\int_{\widetilde{D}_{u,k}(0)} \tilde{g}_0(\boldsymbol{y}) d\boldsymbol{y} > 0.$$

Then

$$\lim_{r\to 0} \Psi_u(\boldsymbol{u}_0 r) = \frac{\lim_{r\to 0}\int_{\widetilde{D}_{u,0}(r)\setminus \widetilde{D}_{u,N-1}(r)} \tilde{g}_r(\boldsymbol{y}) d\boldsymbol{y}}{\lim_{r\to 0}\int_{\widetilde{D}_{u,N-1}(r)} \tilde{g}_r(\boldsymbol{y}) d\boldsymbol{y}} = \frac{\int_{\widetilde{D}_{u,0}(0)\setminus \widetilde{D}_{u,N-1}(0)} \tilde{g}_0(\boldsymbol{y}) d\boldsymbol{y}}{\int_{\widetilde{D}_{u,N-1}(0)} \tilde{g}_0(\boldsymbol{y}) d\boldsymbol{y}}.$$

As for (ii), it suffices to show that for any $\varepsilon > 0$, there exists a constant $U > 0$ such that for any $r \in [0, \tilde{\delta}_\rho]$ and $u > U$,

$$\Psi_u(\boldsymbol{u}_0 r) < \varepsilon.$$

To this end, we need to introduce some notations and concepts.

We start from $\hat{\boldsymbol{y}}_u(r)$, $u > 0$ and $r \geq 0$ as defined in (4.56). It is easy to see $\hat{\boldsymbol{y}}_u(r) = u\hat{\boldsymbol{y}}_1(r)$ for any $u > 0$ and $r \geq 0$. By Lemma 4.6.2, there exists a positive function $\gamma(r)$ of $r \in [0, \tilde{\delta}_\rho]$ such that for any $\boldsymbol{y} \in B(\hat{\boldsymbol{y}}_1(r), \gamma(r))$, $\mathrm{Matri}_N(\widetilde{\boldsymbol{A}}(r)\boldsymbol{y})$ has at least $N-1$ negative eigenvalues. Since $\widetilde{\boldsymbol{A}}(r)$ is continuous on $r \in [0, \tilde{\delta}_\rho]$, $\{\hat{\boldsymbol{y}}_1(r), r \in [0, \tilde{\delta}_\rho]\}$ is compact and covered by $\{B(\hat{\boldsymbol{y}}_1(r), \gamma(r)), r \in [0, \tilde{\delta}_\rho]\}$. Then by the Heine–Borel theorem, there exists a finite open subcover $\{B(\hat{\boldsymbol{y}}_1(r), \gamma(r)), r \in \{r_1, \dots, r_n\}\}$ of $\{\hat{\boldsymbol{y}}_1(r) : r \in [0, \tilde{\delta}_\rho]\}$ for some positive integer n and $r_1, \dots, r_n \in [0, \tilde{\delta}_\rho]$. Let γ' be the distance between the two compact sets $\{\hat{\boldsymbol{y}}_1(r), r \in [0, \tilde{\delta}_\rho]\}$ and $\partial(\bigcup_{k=1}^n B(\hat{\boldsymbol{y}}_1(r_k), \gamma(r_k)))$, i.e.,

$$\gamma' := \min\left\{\|\hat{\boldsymbol{y}}_1(r) - \boldsymbol{y}\| : \boldsymbol{y} \in \partial\left(\bigcup_{k=1}^n B(\hat{\boldsymbol{y}}_1(r_k), \gamma(r_k))\right) \text{ and } r \in [0, \tilde{\delta}_\rho]\right\}. \tag{4.64}$$

It is easy to see $\gamma' > 0$. Then for any $r \in [0, \tilde{\delta}_\rho]$ and $\boldsymbol{y} \in B(\hat{\boldsymbol{y}}_1(r), \gamma')$, $\mathrm{Matri}_N(\widetilde{\boldsymbol{A}}(r)\boldsymbol{y})$ has at least $N-1$ negative eigenvalues. Therefore, for any $u > 0$, $\boldsymbol{y} \in B(\hat{\boldsymbol{y}}_u(r), \gamma' u)$ and $r \in [0, \tilde{\delta}_\rho]$, $\mathrm{Matri}_N(\widetilde{\boldsymbol{A}}(r)\boldsymbol{y})$ also has at least $N-1$ negative eigenvalues.

The next step is to define some useful distances. For any $u > 0$ and $r \in [0, \tilde{\delta}_\rho]$, let $\gamma_{2,u}(r)$ be the distance between the origin and the compact set $\partial B(\hat{\boldsymbol{y}}_u(r), \gamma' u) \cap \partial(\widetilde{H}_{L-1,u}(r) \cap \widetilde{H}_{L,u}(r))$, let

$$\gamma_{0,u}(r) := \|\hat{\boldsymbol{y}}_u(r)\|,$$

and let

$$\gamma_{1,u}(r) := \frac{1}{2}(\gamma_{0,u}(r) + \gamma_{2,u}(r)).$$

It is easy to see for any $u > 0$, $\gamma_{i,u}(r)$, $i = 0, 1, 2$ are all positive and continuous on $r \in [0, \tilde{\delta}_\rho]$ with $\gamma_{i,u}(r) = \gamma_{i,1}(r)u$. Thus, for any $u > 0$ and $i = 0, 1, 2$, we have

$$\min_{r \in [0, \tilde{\delta}_\rho]} \gamma_{i,u}(r) > 0. \tag{4.65}$$

Illustrations of these distances with $\widetilde{P}_{L-1,u}(r)$, $\widetilde{P}_{L,u}(r)$, $B(\hat{\boldsymbol{y}}_u(r), \gamma' u)$ and $B(\boldsymbol{0}_L, \gamma_{2,u}(r))$ are provided in Figure 4.6.3.

Fix $r \in [0, \tilde{\delta}_\rho]$ and $u > 0$. We have some results about these distances. Firstly, note that

$$\gamma_{2,u}(r) - \gamma_{1,u}(r) = \frac{1}{2}\left(\gamma_{2,u}(r) - \gamma_{0,u}(r)\right) = \frac{\gamma'^2 u^2}{2(\gamma_{2,u}(r) + \gamma_{0,u}(r))} > \frac{\gamma'^2 u^2}{4\max_{r \in [0, \tilde{\delta}_\rho]} \gamma_{2,1}(r)} > 0. \tag{4.66}$$

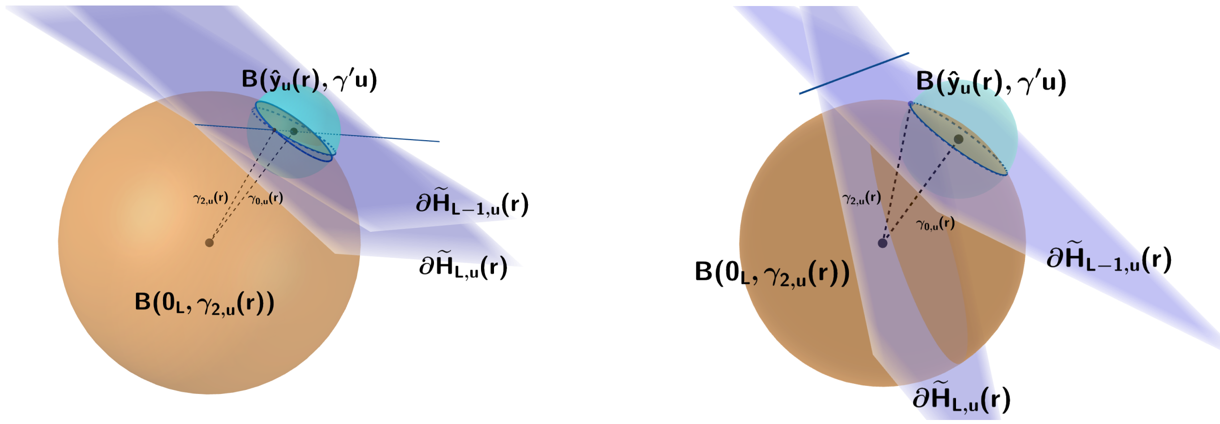


Figure 4.1: Illustrations of $\partial\widetilde{H}_{L-1,u}(r)$, $\partial\widetilde{H}_{L,u}(r)$, $B(\hat{\boldsymbol{y}}_u(r), \gamma' u)$ and $B(\mathbf{0}_L, \gamma_{2,u}(r))$ when $\hat{\boldsymbol{y}}_u(r) = \hat{\boldsymbol{y}}_{L-1,L,u}(r)$ (left) and $\hat{\boldsymbol{y}}_u(r) = \hat{\boldsymbol{y}}_{L-1,u}(r)$ (right).

From geometry, it easy to see

$$\gamma_{2,u}^2(r) = \gamma_{0,u}^2(r) + \gamma'^2 u^2.$$

In addition, for any $\boldsymbol{y} \in \mathbb{R}^L$ such that $\widetilde{\boldsymbol{A}}(\boldsymbol{u}_0 r)_{(i)}\boldsymbol{y} > u$ for $i = L-1, L$ and $\|\boldsymbol{y}\| < \gamma_{2,u}(r)$, suppose $\|\boldsymbol{y} - \hat{\boldsymbol{y}}_u(r)\| \geq \gamma' u$, then

$$\|\boldsymbol{y}\|^2 < \gamma_{2,u}^2(r) = \gamma_{0,u}^2(r) + \gamma'^2 u^2 \leq \|\hat{\boldsymbol{y}}_u(r)\|^2 + \|\boldsymbol{y} - \hat{\boldsymbol{y}}_u(r)\|^2.$$

By discussing each case in (4.57), this means that the origin and $\boldsymbol{y}$ are on the same side of the $(L-1)$-dimensional hyper-plane $\partial\widetilde{H}_{L-1,u}(r)$ or $\partial\widetilde{H}_{L,u}(r)$. Then we get $\widetilde{\boldsymbol{A}}(r)_{(L-1)}\boldsymbol{y} < u$ or $\widetilde{\boldsymbol{A}}(r)_{(L)}\boldsymbol{y} < u$, resulting in a contradiction. Thus, we must have $\|\boldsymbol{y} - \hat{\boldsymbol{y}}_u(r)\| < \gamma' u$, which implies

$$\widetilde{H}_{L-1,u}(r) \cap \widetilde{H}_{L,u}(r) \cap B(\hat{\boldsymbol{y}}_u(r), \gamma' u) \supset \widetilde{H}_{L-1,u}(r) \cap \widetilde{H}_{L,u}(r) \cap B(\mathbf{0}_L, \gamma_{2,u}(r)). \tag{4.67}$$

or equivalently,

$$\widetilde{H}_{L-1,u}(r) \cap \widetilde{H}_{L,u}(r) \cap B^c(\hat{\boldsymbol{y}}_u(r), \gamma' u) \subset \widetilde{H}_{L-1,u}(r) \cap \widetilde{H}_{L,u}(r) \cap B^c(\mathbf{0}_L, \gamma_{2,u}(r)).$$

Then by (4.64),

$$\widetilde{D}_{u,0}(r) \setminus \widetilde{D}_{u,N-1}(r) \subset \widetilde{H}_{L-1,u}(r) \cap \widetilde{H}_{L,u}(r) \cap B^c(\hat{\boldsymbol{y}}_u(r), \gamma' u) \subset B^c(\mathbf{0}_L, \gamma_{2,u}(r)), \tag{4.68}$$

where recall $\widetilde{D}_{u,k}(r)$, $0 \leq k \leq N-1$ was defined in (4.60).

Now we further define some useful subsets of $\mathbb{R}^L$. For any $r \in [0, \tilde{\delta}_\rho]$ and $u > 0$, define

$$C_u(r) := \widetilde{H}_{L-1,u}(r) \cap \widetilde{H}_{L,u}(r) \cap B(\mathbf{0}_L, \gamma_{1,u}(r)). \tag{4.69}$$

It is easy to see $C_u(r) = uC_1(r)$, $\lambda_L(C_1(r)) > 0$, and $\lambda_L(C_1(r))$ is a continuous function of $r \in [0, \tilde{\delta}_\rho]$. Then we have

$$C_{1,\tilde{\delta}_\rho} := \min_{r \in [0,\tilde{\delta}_\rho]} \lambda_L(C_1(r)) > 0. \tag{4.70}$$

For any $r \geq 0$ and $\boldsymbol{y} \in \mathbb{R}^L$, let

$$\tilde{h}_r(\boldsymbol{y}) := h_r\left(\boldsymbol{P}^T(r)\boldsymbol{y}\right),$$

where h_r is as defined in (4.39) and (4.40). Indeed, by (4.46) and (4.63), we have

$$\tilde{g}_r(\boldsymbol{y}) = \left|\tilde{h}_r(\boldsymbol{y})\right| p_L(\boldsymbol{y}) = r^{-1}\left|\det\left(\text{Matri}_N\left(\widetilde{\boldsymbol{A}}(r)\boldsymbol{y}\right)\right)\right| p_L(\boldsymbol{y}). \tag{4.71}$$

For any $r \geq 0$, define

$$Q_r := \left\{\boldsymbol{y} \in \mathbb{R}^L : \tilde{h}_r(\boldsymbol{y}) = 0\right\}$$

Since $h_r(\boldsymbol{y})$ is a polynomial of $\boldsymbol{y}$, we have

$$\lambda_L\left(Q_r\right) = 0. \tag{4.72}$$

By the continuity of $\widetilde{\boldsymbol{A}}(r)$ on $r \in [0, \tilde{\delta}_\rho]$, $\tilde{h}_r(\boldsymbol{y})$ is continuous on $r \in [0, \tilde{\delta}_\rho]$ for any $\boldsymbol{y} \in \mathbb{R}^L$ (not just continuous on $[0, \delta_{pc}]$ as $h_r(\boldsymbol{y})$). By (4.66) and (4.67), we also have

$$\begin{aligned} C_u(r) &\subset \widetilde{H}_{L-1,u}(r) \cap \widetilde{H}_{L,u}(r) \cap B(\mathbf{0}_L, \gamma_{2,u}(r)) \\ &\subset \widetilde{H}_{L-1,u}(r) \cap \widetilde{H}_{L,u}(r) \cap B(\hat{\boldsymbol{y}}_u(r), \gamma' u) \\ &\subset \widetilde{D}_{u,N-1}(r) \cup Q_r, \end{aligned} \tag{4.73}$$

where $\widetilde{D}_{u,N-1}(r)$ is as defined in (4.60). For any $\varepsilon > 0$, denote

$$I_\varepsilon := \left\{r \in [0, \tilde{\delta}_\rho] : \lambda_L\left(C_1(r) \cap \left\{\boldsymbol{y} \in \mathbb{R}^L : \left|\tilde{h}_r(\boldsymbol{y})\right| \geq \varepsilon\right\}\right)/\lambda_L(C_1(r)) \leq 0.5\right\}. \tag{4.74}$$

It is also easy to see $\lambda_L(C_1(r) \cap \{\boldsymbol{y} \in \mathbb{R}^L : |\tilde{h}_r(\boldsymbol{y})| \geq \varepsilon\})$ is a continuous function of $r \in [0, \tilde{\delta}_\rho]$. Then I_ε is compact for any $\varepsilon > 0$. Suppose $I_\varepsilon \neq \emptyset$ for any $\varepsilon > 0$, then $\{I_{1/n}, n \geq 1\}$ forms a decreasing sequence of non-empty compact set. By Cantor's intersection theorem, this implies

$$\bigcap_{n=1}^{\infty} I_{1/n} \neq \emptyset,$$

i.e., there exists a constant $r_0 \in [0, \tilde{\delta}_\rho]$ such that for any $n \geq 1$,

$$\lambda_L\left(C_1(r_0) \cap \{\boldsymbol{y} \in \mathbb{R}^L : \left|\tilde{h}_{r_0}(\boldsymbol{y})\right| \geq 1/n\}\right) \leq 0.5\lambda_L(C_1(r_0)).$$

By taking limits on the both sides of the above inequality as $n \to \infty$, we have

$$\lambda_L\left(C_1(r_0) \cap \{\boldsymbol{y} \in \mathbb{R}^L : \left|\tilde{h}_{r_0}(\boldsymbol{y})\right| > 0\}\right) \leq 0.5\lambda_L(C_1(r_0)). \tag{4.75}$$

However, by (4.70), (4.72) and (4.75), we have

$$\begin{aligned}
0 < C_{1,\tilde{\delta}_\rho} \leq \lambda_L(C_1(r_0)) &\leq \lambda_L\left(C_1(r_0) \cap \left\{\boldsymbol{y} \in \mathbb{R}^L : \left|\tilde{h}_{r_0}(\boldsymbol{y})\right| > 0\right\}\right) + \lambda_L(Q_{r_0}) \\
&= \lambda_L\left(C_1(r_0) \cap \left\{\boldsymbol{y} \in \mathbb{R}^L : \left|\tilde{h}_{r_0}(\boldsymbol{y})\right| > 0\right\}\right) \\
&\leq 0.5\lambda_L(C_1(r_0)),
\end{aligned}$$

resulting in a contradiction. Therefore, there exists some $\eta > 0$ independent of u such that $I_\eta = \emptyset$. For any $u > 0$ and $r \in [0, \tilde{\delta}_\rho]$, let

$$T_u(r) := C_u(r) \cap \left\{\boldsymbol{y} \in \mathbb{R}^L : \left|\tilde{h}_r(\boldsymbol{y})\right| \geq \eta u^N\right\}. \tag{4.76}$$

By (4.39), (4.40) and (4.69), it is easy to see

$$T_u(r) = T_1(r)u^L.$$

Then by (4.74),

$$\lambda_L(T_u(r)) = \lambda_L(T_1(r))u^L \geq \frac{1}{2}\lambda_L(C_1(r))u^L \geq \frac{1}{2}C_{1,\tilde{\delta}_\rho}u^L \tag{4.77}$$

By (4.73) and $\eta > 0$, we also have

$$T_u(r) \subset \widetilde{D}_{u,N-1}(r). \tag{4.78}$$

Now we return to the proof of (ii). Given $\varepsilon > 0$, we can choose U (as mentioned at the start of the proof of (ii)) as follows. Firstly, choose $\theta \in (0,1)$. Note that there exists a constant $\widetilde{C} > 0$ such that for any $r \in [0, \tilde{\delta}_\rho]$ and $\boldsymbol{y} \in \mathbb{R}^L$,

$$\tilde{h}_r(\boldsymbol{y}) \leq \widetilde{C} \sum_{v_1,\ldots,v_N \in \{1,\ldots,L\}} |y_{v_1} \cdots y_{v_N}|.$$

Thus, there exists $U_1 > 0$ such that

$$\tilde{h}_r(\boldsymbol{y}) \le \exp\left(\frac{1}{2}\theta \boldsymbol{y}^T \boldsymbol{y}\right), \tag{4.79}$$

for any $r \in [0, \tilde{\delta}_\rho]$ and $\boldsymbol{y} \in \mathbb{R}^L$ satisfying $\|\boldsymbol{y}\| \ge U_1 \min_{s\in[0,\tilde{\delta}_\rho]} \gamma_{2,1}(s)$ (see (4.65)). Then by (4.65), (4.66) and Lemma 4.6.4 with $k_2 = \gamma_{2,1}(r)$ and $k_1 = \gamma_{1,1}(r)$, there exists $U_2 > 0$ such that

$$\frac{(1-\theta)^{-\frac{L}{2}} \int_{uB^c(\boldsymbol{0}_L,\gamma_{2,1}(r))} \exp\left(-\frac{1}{2}\boldsymbol{y}'^T\boldsymbol{y}'\right) d\boldsymbol{y}'}{\frac{1}{2}\eta C_{1,\tilde{\delta}_\rho} u^{L+N} \exp\left(-\frac{1}{2}\gamma_{1,1}^2(r)u^2\right)} < \varepsilon, \tag{4.80}$$

for any $u \ge U_2$. Then we can simply take $U := \max(U_1, U_2)$. Indeed, for any $r \in [0, \tilde{\delta}_\rho]$ and $u > \max(U_1, U_2)$,

$$\begin{aligned}
&\Psi_u(\boldsymbol{u}_0 r) \\
&= \frac{\int_{\widetilde{D}_{u,0}(r)\setminus\widetilde{D}_{u,N-1}(r)} \tilde{g}_r(\boldsymbol{y})d\boldsymbol{y}}{\int_{\widetilde{D}_{u,N-1}(r)} \tilde{g}_r(\boldsymbol{y})d\boldsymbol{y}} \quad \text{(by (4.62))} \\
&= \frac{\int_{\widetilde{D}_{u,0}(r)\setminus\widetilde{D}_{u,N-1}(r)} \tilde{h}_r(\boldsymbol{y}) \exp\left(-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{y}\right) d\boldsymbol{y}}{\int_{\widetilde{D}_{u,N-1}(r)} \tilde{h}_r(\boldsymbol{y}) \exp\left(-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{y}\right) d\boldsymbol{y}} \quad \text{(by (4.71))} \\
&\le \frac{\int_{B^c(\boldsymbol{0}_L,\gamma_{2,u}(r))} \tilde{h}_r(\boldsymbol{y}) \exp\left(-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{y}\right) d\boldsymbol{y}}{\int_{T_u(r)} \tilde{h}_r(\boldsymbol{y}) \exp\left(-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{y}\right) d\boldsymbol{y}} \quad \text{(by (4.68) and (4.78))} \\
&\le \frac{\int_{B^c(\boldsymbol{0}_L,\gamma_{2,u}(r))} \exp\left(-\frac{1}{2}(1-\theta)\boldsymbol{y}^T\boldsymbol{y}\right) d\boldsymbol{y}}{\eta u^N \int_{T_u(r)} \exp\left(-\frac{1}{2}\boldsymbol{y}^T\boldsymbol{y}\right) d\boldsymbol{y}} \quad \text{(by (4.76) and (4.79))} \\
&\le \frac{(1-\theta)^{-\frac{L}{2}} \int_{uB^c(\boldsymbol{0}_L,\gamma_{2,1}(r))} \exp\left(-\frac{1}{2}\boldsymbol{y}'^T\boldsymbol{y}'\right) d\boldsymbol{y}'}{\frac{1}{2}\eta C_{1,\tilde{\delta}_\rho} u^{L+N} \exp\left(-\frac{1}{2}\gamma_{1,1}^2(r)u^2\right)} \quad \text{(by } \boldsymbol{y}' = \sqrt{1-\theta}\boldsymbol{y}\text{, (4.69), (4.76), (4.77))} \\
&< \varepsilon, \quad \text{(by (4.80))}
\end{aligned}$$

as desired.

□

4.6.4 A Corollary of the Main Results

Note that (ii) of Lemma 4.6.3 holds for $k = N$. Then by a similar proof for (i) of Theorem 4.6.1, the limit of $\frac{f_{u,N}(\boldsymbol{t})}{f_{u,N-1}(\boldsymbol{t})+f_{u,N}(\boldsymbol{t})}$ as $\|\boldsymbol{t}\| \to 0$ also exists for any $u > 0$. The following is an immediate corollary of Theorems 4.5.1 and 4.6.1.

Corollary 4.6.5. *Let X be qualified under perturbation and satisfy Condition (4.53). Then*

$$\lim_{u\to\infty}\lim_{\|\boldsymbol{t}\|\to 0}\frac{f_{u,N}(\boldsymbol{t})}{f_{u,N-1}(\boldsymbol{t})+f_{u,N}(\boldsymbol{t})}=\frac{1}{2}.$$

Remark 4.6.6. Since we are conditioning on having a critical point at the origin with unknown index, the ratio in Corollary 4.6.5 being very close to $\frac{1}{2}$ implies that a pair of very close critical points should consist of one local maximum and one critical point with index $N-1$.

Intuitively, a connected component of a high excursion set most likely contains exactly one critical point (one global maximum) or three critical points (two local maxima and one critical point with index $N-1$). Thus, for a connected component containing three critical points, Corollary 4.6.5 predicts that the critical point with index $N-1$ will be very close to one of the two local maxima.

Chapter 5

Modeling Critical Points

5.1 Introduction

Modeling the critical points of a Gaussian random field, X, is an important challenge in stochastic geometry and has various applications in astronomy ([6], [40]), biomedical imaging ([42], [19]), geography ([38], [5], [7]), etc. It is noticeable that many of these applications only consider critical points above a high threshold, u, i.e., the critical points in the excursion set

$$A_u(X, S) := \{\boldsymbol{t} \in S : X(\boldsymbol{t}) > u\},$$

for some search region S.

If the underlying Gaussian random field is known, then one can simply obtain a sample of the critical points in an excursion set by locating the critical points on the generated sample functions of the underlying Gaussian random field. However, in practice, one may only be able to observe critical points from a high excursion set, lacking information of the underlying Gaussian random field. In this case, it may not be practical to generate a sample function, as this requires estimating the mean and covariance functions of the underlying Gaussian random field, which can be very difficult. In this case, an above-threshold critical point model that does not require much information about the underlying Gaussian random field would be very useful. Moreover, generating a sample function of the underlying Gaussian random field, even if feasible, would be very time-consuming, while many real-world problems often require to produce a large number of samples of critical points in a short period of time but allowing concessions for the accuracy. Therefore, a relatively accurate but more efficient model can be better adapted to the needs of the practice.

From Theorem 3.4.3, we have learnt that as the threshold u tends to infinity and the search region expands with a matching speed, the point process consisting of (the position of) the global maximum of a Gaussian random field in each connected components of the excursion set will converge weakly to a stationary Poisson point process. An intuitive explanation for this convergence is that distances between the global maxima above u will increase as u increases, resulting in the approximate independence of these global maxima. By Theorem 3.4.6, the above convergence also holds for the point process consisting of the critical points of the Gaussian random field, and the Poisson limits in these two convergences have the same intensity. This is because when the threshold is extremely high, the critical points with index smaller than N should be much less compared with the local maxima as shown in Lemma 3.4.5, while Morse theory asserted that a connected component cannot have more than one local maximum without other types of critical points, and thus, the majority of the critical points above the threshold are the global maxima. This implies when u is extremely high, a Poisson point process can be a suitable model for these critical points above u.

It is natural to ask whether a Poisson point process can still be a suitable model for the critical points of a Gaussian random field above a not-so-high u which is more commonly met in practice. Indeed, under this setting, we can observe from simulations that not only the interactions between the global maxima can be nonnegligible, which causes the global maxima to deviate from a Poisson point process, but also a connected component of the excursion set is more likely to have a complicated structure in the sense that it contains more than one critical point, which further compromises the accuracy of modeling the critical points by a Poisson point process. The emergence of these phenomena poses a challenge to the modeling of the critical points.

Considering that some topological features of a sample function on an excursion set may not be detectable due to the limitation of measurement accuracy, and are often interpreted as noises in persistent homology (see [2]), only the critical points corresponding to detectable topological features are worth studying from both practical and theoretical perspectives. For convenience, we call such critical points persistent.

In this chapter, we will construct a model for the persistent critical points of an isotropic Gaussian random field indexed by $\mathbb{R}^N$ ($N \geq 2$) above a high but not very high threshold with the help of the results that we have obtained in the previous two chapters. Since u is still high, one should expect the Poisson heuristic still hold to certain degree. Moreover, the deviation from it comes from different mechanisms as discussed above. Therefore, the basic idea is to first study the underlying Poisson structure, the interactions between global maxima, the interactions within connected components separately, and then integrate them into one single model. This gives the model a hierarchical structure which allows to capture the global maxima and other critical points simultaneously. To the best of our knowledge, there is no similar study on the modeling of the critical points of a Gaussian random field above a threshold.

The remainder of this chapter is organized as follows. Section 5.2 introduces the hard-core process which serves as the starting point of the model, and the L function which is used to evaluate the performance of the model in the empirical study. Section 5.3 mainly defines the persistent critical points and some other useful concepts. The intuitions behind these concepts are also discussed. Section 5.4 elaborates in detail the construction of the model. To better understand the hierarchical structure of the model, we divided the construction into three stages which correspond to the underlying Poisson structure, the interactions between global maxima, and the interactions within connected components as mentioned above. The relationships between these stages and the intuitions behind them are also clarified. Section 5.5 explains the estimation of the parameters in the proposed model, and Section 5.6 checks the performance of the model using two isotropic Gaussian random fields indexed by $\mathbb{R}^2$.

5.2 Preliminaries

5.2.1 Hard-core Processes

A **hard-core process** is a point process whose points maintain a predetermined minimum distance $r_H \geq 0$ from one another. This distance is usually called the **hard-core distance** or the radius of the hard core. In practice, one can simulate a hard-core process by thinning an underlying process, i.e., deleting some points of the underlying process according to the hard-core distance and some **thinning rules**. For example, [25] introduced two hard-core processes, the first and the second Matérn processes, by thinning a Poisson point process. The constructions of the two Matérn processes both start from detecting all the pairs of points of a Poisson point process with distance less than r_H. However, their thinning rules are slightly different: the first Matérn process requires to delete the whole pair, while the second Matérn process only requires to delete the point with the lower height, where the heights of points are independently and identically sampled from the uniform distribution on $[0, 1]$ (written as $U(0, 1)$) and also independent of other source of randomness in this construction.

More specifically, equip $\mathbb{R}^N$ with the usual Euclidean norm $\| \cdot \|$. For any $\boldsymbol{t} \in \mathbb{R}^N$ and $r > 0$, let $B(\boldsymbol{t}, r)$ be the N-dimensional open ball centered at $\boldsymbol{t}$ with radius r, and let

$\mathring{B}(\boldsymbol{t}, r) := B(\boldsymbol{t}, r) \setminus \{\boldsymbol{t}\}$. A point $\boldsymbol{s} \in \mathbb{R}^N$ is said to be an r-**neighbor** of $\boldsymbol{t}$ if $\boldsymbol{s} \in \mathring{B}(\boldsymbol{t}, r)$. The intensity of a stationary point process on $\mathbb{R}^N$ is defined to be the average number of its points in a unit of area. Let $\{\boldsymbol{t}_i, i \geq 1\}$ be the points of a Poisson point process on $\mathbb{R}^N$ with intensity $\mu_P > 0$. Let $\{h_i, i \geq 1\}$ be a random sample from $U(0, 1)$, and h_i is called the height of $\boldsymbol{t}_i$. The thinning rule of a second Matérn process with hard-core distance $r_H > 0$ is

$$\begin{gathered}\text{a point } \boldsymbol{t}_i \text{ is only retained if it has no } r_H\text{-neighbors in } \{\boldsymbol{t}_i, i \geq 1\} \\ \text{that are higher than or equal to it.}\end{gathered} \tag{5.1}$$

Based on [29], the intensity μ_H of this second Matérn Process is given by

$$\mu_H = \frac{1 - e^{-\mu_P \pi r_H^2}}{\pi r_H^2}. \tag{5.2}$$

5.2.2 L Functions

Denote by $\mathcal{B}(\mathbb{R}^N)$ the Borel σ-field of $\mathbb{R}^N$. Let ξ be a point process on $\mathbb{R}^N$ with the intensity function $\mu : \mathbb{R}^N \to [0, \infty)$, i.e., for any $B \in \mathcal{B}(\mathbb{R}^N)$,

$$\mathbb{E}[\xi(B)] = \int_B \mu(\boldsymbol{s}) d\boldsymbol{s}.$$

Then the Ripley's K function (see [35] and ([36])) of ξ at a point $\boldsymbol{t} \in \mathbb{R}^N$ is defined by

$$K_{\boldsymbol{t}}(r) := \frac{1}{\mu(\boldsymbol{t})} \lim_{h \to 0} \mathbb{E}\left[\xi\left(\mathring{B}(\boldsymbol{t}, r)\right) \middle| \xi\left(B(\boldsymbol{t}, h)\right)\right], \quad r > 0. \tag{5.3}$$

If ξ is stationary, then the values of $\mu(\boldsymbol{t})$ and $K_{\boldsymbol{t}}(r)$ are independent of $\boldsymbol{t}$, and thus, the subscript $\boldsymbol{t}$ can be omitted. In particular, if ξ is a stationary Poisson point process, its Ripley's K function is

$$K(r) = \pi r^2, \quad r > 0, \tag{5.4}$$

which is independent of its intensity.

In practice, points of ξ are collected on a bounded search region, S. For any $n \geq 1$, let λ_n be the n-dimensional Lebesgue measure. Let $M := \xi(S)$, and let $\xi^S := \{\boldsymbol{t}_i, 1 \leq i \leq M\}$ be the points of ξ in S. By (5.3), it is straightforward to estimate Ripley's K function by the average number of r-neighbors over all the points in ξ^S per unit area. This is the intuition behind the classical empirical K function, $\widehat{K}(r)$, on S, suggested by [35] and [36]. More specifically,

$$\begin{aligned}\widehat{K}(r) :&= \frac{1}{\hat{\mu}}\frac{1}{M}\sum_{i=1}^{M}\sum_{j=1,j\neq i}^{M-1} I_{\{\|\boldsymbol{t}_i-\boldsymbol{t}_j\|<r\}}\\ &= \frac{\lambda_N(S)}{M(M-1)}\sum_{i=1}^{M}\sum_{j=1,j\neq i}^{M-1} I_{\{\|\boldsymbol{t}_i-\boldsymbol{t}_j\|<r\}},\end{aligned}$$

where I stands for the indicator function of an event, and the estimator $\hat{\mu} := \frac{M-1}{\lambda_N(S)}$ of the intensity μ (instead of using $\hat{\mu} := \frac{M}{\lambda_N(S)}$) is suggested by [12] for technical reasons.

However, the classical empirical K function on S at r fails to consider the r-neighbors outside S, which leads to the so-called edge effect bias. A correction method for this bias is called the border correction. This correction method requires shrinking the original search region to the extent that for any point in the search region, its r-neighbors are all included in the original one, and hence, the edge effect does not occur. However, when r is large, the border correction method may dramatically reduce the available data due to a significant shrinkage of the search region. In this case, a more efficient correction method needs to be considered.

To this end, we choose the isotropic correction method from several candidates suggested by [37]. This correction method can eliminate the edge effect bias without reducing available data. For any set $A \in \mathbb{R}^N$, let ∂A be the boundary of A. The empirical K function on S with the isotropic correction is defined by

$$\widehat{K}_{\text{iso}}(r) := \frac{\lambda_N(S)}{M(M-1)}\sum_{i=1}^{M}\sum_{j=1,j\neq i}^{M-1} w_{ij} I_{\{\|\boldsymbol{t}_i-\boldsymbol{t}_j\|<r\}}, \quad r > 0,$$

where w_{ij} represents the weight of the ordered pair $(\boldsymbol{t}_i, \boldsymbol{t}_j)$ and is defined by

$$w_{ij} := \frac{\lambda_{N-1}(\partial B(\boldsymbol{t}_i, \|\boldsymbol{t}_i - \boldsymbol{t}_j\|))}{\lambda_{N-1}(\partial B(\boldsymbol{t}_i, \|\boldsymbol{t}_i - \boldsymbol{t}_j\|) \cap S)}.$$

In fact, $\lambda_N(S)^{-1}\sum_{i=1}^{M}\sum_{j=1,j\neq i}^{M-1} w_{ij} I_{\{\|\boldsymbol{t}_i-\boldsymbol{t}_j\|<r\}}$ is an unbiased estimator of $\mu^2 K(r)$ when r is small relative to S (see [35] and [36]). Thus, for any $r > 0$, $\widehat{K}_{\text{iso}}(r)$ is an asymptotically unbiased estimator of $K(r)$ as S expands to $\mathbb{R}^N$. In particular, when S is a rectangle, $\lambda_N(S)^{-1}\sum_{i=1}^{M}\sum_{j=1,j\neq i}^{M-1} w_{ij} I_{\{\|\boldsymbol{t}_i-\boldsymbol{t}_j\|<r\}}$ is an unbiased estimator of $\mu^2 K(r)$ when r is less than or equal to half of the diagonal of S. In this case, an explicit formula for w_{ij} is provided in [15].

When ξ is a Poisson point process on $\mathbb{R}^2$ with intensity μ, [37] provided an approximate variance of its empirical K function on S with the isotropic correction as S expands to $\mathbb{R}^2$,

$$Var[\hat{K}_{\text{iso}}(r)] \approx \frac{2}{\mu}\left(\frac{\pi r^2}{\lambda_2(S)} + \frac{0.96\lambda_1(\partial S)r^3}{\lambda_2(S)^2} + 0.13\mu\frac{\lambda_1(\partial S)r^5}{\lambda_2(S)^3}\right).$$

This implies the variance of $\hat{K}_{\text{iso}}(r)$ on a fixed search region will explode as r increases, which makes it unsuitable for a graphical comparison. A workaround is to use the L function ([8])

$$L(r) := \sqrt{\frac{K(r)}{\pi}} - r, \quad r > 0.$$

This transformation, inspired by Ripley's K function of a stationary Poisson point process in (5.4), can stabilize the variance of the estimator. Since the L function of any stationary Poisson point process is the zero function, one can easily compare a point process with a stationary Poisson point process by comparing its L function with zero.

Let ξ_i, $1 \leq i \leq \ell$ be independent copies of ξ conditional on $\xi(S) \geq 2$, and let $\hat{K}_{\text{iso},i}(r)$ be the empirical K function of ξ_i on S with the isotropic correction. Then we can define the $\bar{L}$ function

$$\bar{L}(r) := \sqrt{\frac{\sum_{i=1}^{\ell} \hat{K}_{\text{iso},i}(r)}{\ell\pi}} - r, \quad r > 0.$$

From the above discussion, it is easy to see for any $r > 0$, $\bar{L}(r)$ is an asymptotic unbiased estimator of $L(r)$ as S expands to $\mathbb{R}^N$ and $\ell \to \infty$. To reduce the bias of $\bar{L}(r)$, one needs to select S as large as possible to reduce the bias of $\hat{K}_{\text{iso},i}(r)$.

Let $\bar{K} := \frac{1}{\ell}\sum_{i=1}^{\ell} \hat{K}_{\text{iso},i}(r)$. Since the variance of $\bar{K}$ can be estimated by $\widehat{\text{Var}}(\bar{K}) := \frac{\sum_{i=1}^{\ell}(\hat{K}_{\text{iso},i}(r)-\bar{K})^2}{\ell(\ell-1)}$, the variance of $\bar{L}(r)$ can be estimated by $\widehat{\text{Var}}(\bar{K})$ and the delta method, i.e.,

$$\widehat{\text{Var}}\left(\bar{L}(r)\right) := h'\left(\bar{K}\right)^2 \widehat{\text{Var}}\left(\bar{K}\right) = \frac{1}{4\pi\bar{K}} \frac{\sum_{i=1}^{\ell}(\hat{K}_{\text{iso},i}(r) - \bar{K})^2}{\ell(\ell-1)},$$

where $h(y) = \sqrt{\frac{y}{\pi}}$, $y \geq 0$.

5.3 Basic Settings

Let $\{X(\boldsymbol{t}), \boldsymbol{t} \in \mathbb{R}^N\}$ be an isotropic Gaussian random field and satisfy the conditions of Lemma 2.1.10. Assume that the covariance function $\text{Cov}\left[X(\boldsymbol{s}), X(\boldsymbol{t})\right] \geq 0$ for any

$\boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^N$. Define the **excursion set**, $A_u(X, S)$, of X at the threshold $u \in \mathbb{R}$ in the search region $S \subset \mathbb{R}^N$ by

$$A_u(X, S) := \{\boldsymbol{t} \in S : X(\boldsymbol{t}) > u\}.$$

Given $u \in \mathbb{R}$, the point $\boldsymbol{t} \in \mathbb{R}^N$ is said to be a (component) global maximum of X if it is the position of the global maximum of X in a bounded connected component of $A_u(X, \mathbb{R}^N)$. When u is reasonably high, all the connected components are almost surely bounded. Also, we call a connected component of $A_u(X, \mathbb{R}^N)$ "a connected component at u".

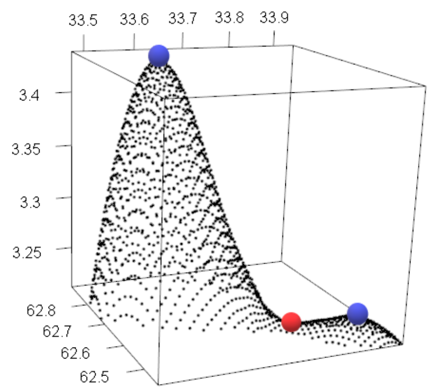


Figure 5.1: A sample function of a Gaussian random field above a high threshold. The connected component has three critical points: two local maxima (blue) and one saddle point (red).

Let f be a sample function of X. Recall that in Section 2.3, a point $\boldsymbol{t} \in \mathbb{R}^N$ is said to be a critical point of f with index k $(0 \leq k \leq N)$ if

$$\nabla f(\boldsymbol{t}) = 0 \quad \text{and} \quad \text{index}\left(\nabla^2 f(\boldsymbol{t})\right) = k,$$

where $\nabla f(\boldsymbol{t})$ and $\nabla^2 f(\boldsymbol{t})$ denote the gradient and the Hessian matrix of f respectively, and $\text{index}\,(\cdot)$ denotes the number of negative eigenvalues of a square matrix. A critical point $\boldsymbol{t}$ of f is said to be non-degenerate if

$$\det\left(\nabla^2 f(\boldsymbol{t})\right) \neq 0.$$

By Remark 2.1.12, with probability one, all the critical points of X in a bounded search region are non-degenerate. Thus, we can focus only on non-degenerate critical points of X in the remainder of this chapter.

Now let us review our previous results about critical points of an isotropic Gaussian random field above a high threshold u. In Lemma 3.4.5, we have shown that the majority of the critical points above u are local maxima. In Corollary 4.6.5, we have shown that if two critical points above u are very close one to each other, then they are most likely to be one local maximum and one saddle point with index $N-1$. These results imply that the number of the critical points with index less than $N-1$ should be negligible compared with the number of local maxima or the number of the saddle points with index $N-1$. Thus, when u is reasonably high, most of the connected components at u only contain local maxima and saddle points with index $N-1$. By Morse theory (see Corollary 9.3.5 in [4] and Theorem 9.1 in [26]), such a connected component typically contains either exactly one critical point which is its global maximum, or three critical points consisting of one global maximum, one local maximum and one saddle point with index $N-1$.

A connected component at u is said to be

- **simple** if it contains exactly one critical point which is its global maximum;
- **typical** if it is simple, or contains exactly three critical points: one global maximum, one local maximum and one saddle point with index $N-1$, which form a **family of three** at u.

One should note that, a family of three at u may not be a family of three at u' when $u' \neq u$. This is because if $u' > u$, then a family member may be lower than u'; if $u' < u$, the connected component at u' which contains the whole family may have more than three critical points (be non-typical).

Assume that f is a sample function of X, $u_1 \in \mathbb{R}$, and C_{u_1} is a typical connected component of $A_{u_1}(f, \mathbb{R}^N)$. Let $u_2 := \max_{\boldsymbol{t} \in C_{u_1}} f(\boldsymbol{t})$. For any $u \in [u_1, u_2)$, define

$$C_u := A_u\left(f, \mathbb{R}^N\right) \cap C_{u_1}.$$

Then as u increases, we can observe the change in the topology of C_u as follows. If C_{u_1} is a simple connected component, then C_u will gradually shrink to the position of the global maximum of f in C_{u_1} and there is no change of the topology of C_u during this process. However, if C_{u_1} contains a family of three at u_1, then the change is more complicated. When u just passes the value of the saddle point in the family of three, C_u will split into two simple connected components at u such that one of them contains the global maximum and the other one contains the local maximum. As u further increases until it just passes the value of the local maximum, the connected component containing the local maximum will disappear. Such changes of the topology of C_u can be captured by the so-called persistence barcode (see [14]) which is a widely used tool in persistence homology. Furthermore, if the difference between the values of the saddle point and local maximum are small and u is interpreted as time, then the simple connected component containing the local maximum can only "survive" for a short time, which corresponds to a short bar on the persistence barcode. Note that in persistence homology, a short bar, corresponding to a short-lived topological feature, is usually interpreted as noise (see [2]). We will follow this idea to eliminate these noises in our data.

More specifically, a family of three is said to be **short-lived** if the difference between the values of the local maximum and the saddle point is less than some predetermined small value δ_p. Let u_t be our target threshold. Considering that in practice, it is often hard to distinguish a typical connected component containing a short-lived family of three from a simple connected component, and in order to eliminate the interference of noises from short-lived topological features, it is more realistic and reasonable to modify every short-lived family of size three at any $u \in [u_t, \infty)$ (u is from low to high) into a single point, by deleting the local maximum and the saddle point in the family. After this modification, all the critical points above u_t are called **persistent**.

In summary, given an isotropic Gaussian random field X indexed by $\mathbb{R}^N$ and a target threshold $u_t > 0$ which is high but not very high, we are interested in the simulation of the persistent critical points of X above u_t.

5.4 Modeling Critical Points Using a Clustering Process

We call a persistent critical point **typical** if it is contained by a typical (after the modification) connected component at u_t. From the last section, we have learnt that the majority of the connected components at u_t are typical. As such, we will focus only on the distribution of persistent and typical critical points above u_t in our model design. In the remainder of this chapter, denote by $\xi^C_{u_t}$ the point process on $\mathbb{R}^N$ consisting of persistent and typical critical points above u_t, by $\xi^M_{u_t}$ the point process on $\mathbb{R}^N$ consisting of the persistent and typical global maxima above u_t, and for convenience, we will not distinguish a point process from the set of its points. Then it is easy to see $\xi^M_{u_t} \subset \xi^C_{u_t}$.

When the target threshold u_t is very high, Sections 3.4.2 and 3.4.3 suggest that the difference between $\xi^C_{u_t}$ and its subset $\xi^M_{u_t}$ is negligible, and both of them can be approximated by a Poisson point process in distribution. However, this is not the case when u_t is high but not very high. Figure 5.2 shows that the estimates of the L functions (see Section

5.2.2) of $\xi_{u_t}^C$ and $\xi_{u_t}^M$ both obviously deviate from the zero function, which is the theoretical L function of a stationary Poisson point process. Thus, more subtle models for $\xi_{u_t}^C$ and $\xi_{u_t}^M$ should be considered in this case.

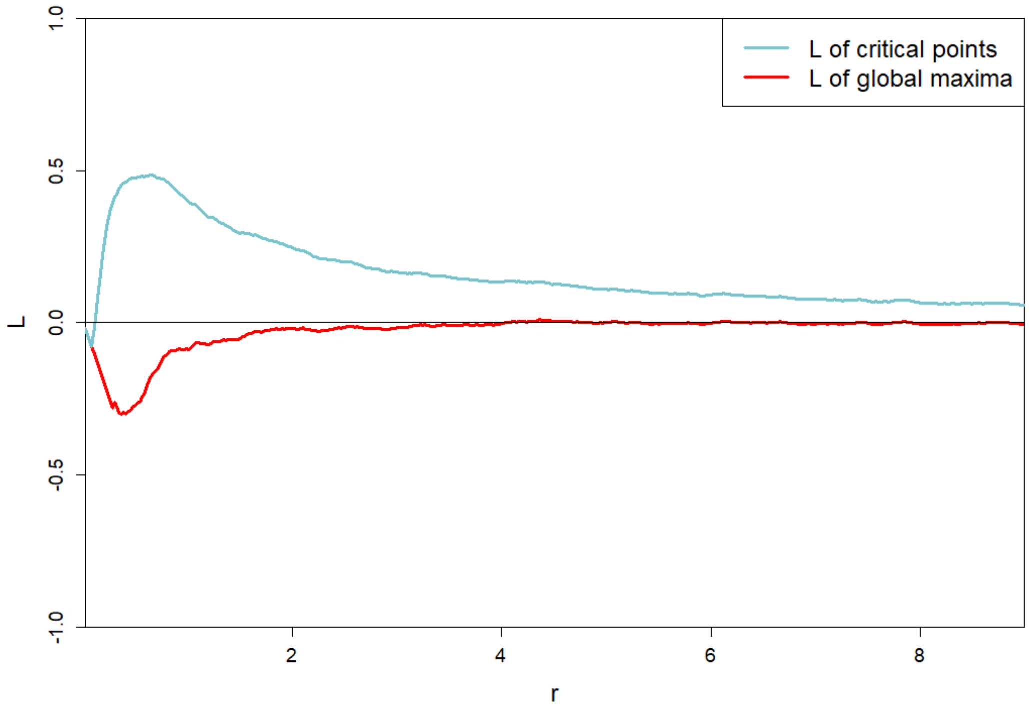


Figure 5.2: Estimates of L functions of the critical points $\xi_{u_t}^C$ and the global maxima $\xi_{u_t}^M$ for the covariance function $C(r) = \exp(-4r^2)$, $r \geq 0$ and $u_t = 3.5$.

A stochastic process is called a **clustering process** on $\mathbb{R}^N$ if it can be constructed by the following procedures:

- generate a parent process on $\mathbb{R}^N$ which may or may not be observed;
- each parent point produces a random number of observed daughter points, and these numbers are independently and identically distributed;
- the positions of daughter points relative to their respective parents are also independently and identically distributed.

For a clustering process, we call it a clustering process of order one if its parent process is not a clustering process. As an immediate generalization, we can define a **clustering process of order** n ($n \geq 2$) iteratively by letting its parent process be a clustering process of order $n-1$. One can also refer to [28] for a similar generalization of the Neyman–Scott process to higher orders.

In particular, a clustering process of order n with parents points being observed in every iteration can be divided into $n+1$ different groups. These groups are also called **generations** and denoted by G_i for $i = 1, \ldots, n+1$. For any $1 \leq k \leq n$, G_{k+1} is called the daughter of the accumulative union $\bigcup_{i=1}^{k} G_i$, and each point in G_{k+1} can only have one parent in $\bigcup_{i=1}^{k} G_i$.

In essence, the proposed model for the critical points $\xi_{u_t}^C$ is a clustering process of order two on $\mathbb{R}^N$ with parent points being observed in every iteration. This means that the model consists of three observed generations, G_i, $i = 1, 2, 3$ such that G_2 is the daughter of G_1, and G_3 is the daughter of $G_1 \cup G_2$.

The **intuition** behind the model is that one can find a similar structure in $\xi_{u_t}^C$. We first look at its subset $\xi_{u_t}^M$. When u_t is high but not very high, the Poisson characterization should be largely kept between points of $\xi_{u_t}^M$ which maintain a minimum distance, r'_H, from one another due to the decreasing dependence in distance. In fact, these points form a hard-core process with hard-core distance r'_H, and can be considered as the first generation, denoted by G'_1, in $\xi_{u_t}^C$. Thus, the behavior of G'_1 should be close to a second Matérn process with the same hard-core distance. Since points of $\xi_{u_t}^M$ are all born with natural heights (i.e., the values of X at these points), we can identify G'_1 from $\xi_{u_t}^M$ by applying the thinning rule (5.1) to $\xi_{u_t}^M$. An advantage of this thinning rule is that for each point in $\xi_{u_t}^M \setminus G'_1$, it has at least one r'_H-neighbor in G'_1, and then its closest r'_H-neighbor in G'_1 (by Lemma 3.2.2, with probability one, there are only finite local maxima in a bounded subset of $\mathbb{R}^N$) can be considered as its parent. Thus, $\xi_{u_t}^M \setminus G'_1$ is the second generation, denoted by G'_2, in $\xi_{u_t}^C$. Finally, if a global maximum in $G'_1 \cup G'_2 = \xi_{u_t}^M$ is the member of a family of three in $\xi_{u_t}^C$, then it can be considered as the parent of other family members, the local maximum and the saddle point. Thus, $\xi_{u_t}^C \setminus \xi_{u_t}^M$, consisting of the local maxima and saddle points of the families of three at u_t, can be considered as the third generation, denoted by G'_3, in $\xi_{u_t}^C$.

From the above intuition, we see that each generation G_i can be considered as a model for G'_i for $i = 1, 2, 3$. Moreover, note that the accumulative unions G'_1 (a hard-core process consisting of global maxima), $G'_1 \cup G'_2$ (i.e., $\xi_{u_t}^M$) and $G'_1 \cup G'_2 \cup G'_3$ (i.e., $\xi_{u_t}^C$) are all well-defined point processes with explicit geometric meanings. Thus, it is more helpful to divide the construction of our modeling into three stages, so that each stage can be considered as (and also named by) the modeling of one of these point processes.

5.4.1 Stage 1: Obtaining Global Maxima not Affected by the Hard-core Thinning

Let ξ be a stationary Poisson point process on $\mathbb{R}^N$ with intensity $\mu_P > 0$. Then we can apply the thinning rule (5.1) to ξ to generate a second Matérn process with hard-core distance $r_H > 0$. This second Matérn process is G_1.

5.4.2 Stage 2: Modeling Other Global Maxima

In this stage, we will construct a model for ξ^M_{ut} which is essentially a clustering process of order one. This model consists of two generations: G_1, the second Matérn process in Stage 1, and G_2, the daughter of G_1. Our algorithm to produce G_2 from G_1 is as follows.

For each parent in G_1, a Bernoulli trial with success rate $p_1 > 0$ is run to determine whether it produces daughters in its neighborhood: if successful, it produces one daughter; otherwise, it does not produce any daughters. These Bernoulli trials are mutually independent and also independent of all other randomness in the model. Once the Bernoulli trial succeeds, a distribution on $\mathbb{R}^N$ is needed to determine where to produce its daughter. Since the underlying random field is isotropic, this distribution can be characterized by the distance R between the daughter and the parent. More specifically, if the parent is at $\boldsymbol{P} \in \mathbb{R}^N$, then the position, $\boldsymbol{D}$, of its daughter can be given by

$$\boldsymbol{D} = \boldsymbol{P} + R\frac{\boldsymbol{V}}{\|\boldsymbol{V}\|}, \tag{5.5}$$

where $\boldsymbol{V}$ is a Gaussian N-vector with zero-mean and identity covariance matrix.

As far, the only thing left to determine G_2 is the distribution of R. Let $p(r)$, $r > 0$ be the probability density function of R. Intuitively, we see that $p(r)$ should be a model for the distribution of the distance, R', between a random point in G'_2 and its parent in G'_1, and $R' \leq r'_H$. Thus, it is reasonable to assume

$$p(r) = 0 \text{ for any } r \geq r_H.$$

Note that the distribution of R' is naturally weighted by $\lambda_{N-1}(\partial B(\mathbf{0}_N, r))$ at $R' = r$. Thus, equivalently but more naturally, we can consider the density function

$$q(r) \propto \frac{p(r)}{r^{N-1}}, \quad r > 0,$$

where "$\propto$" means "is proportional to". This should give us the "density" of finding a member of $G_2(r)$ at any given point with distance r from its parent.

Intuitively, the interaction between a point in G_1' and its daughter in G_2' (as defined in our intuition) is similar to the interatomic interaction since as the distance increases, they both go through the following three stages:

1. Strong repulsion at a very short distance: consider the situation where two global maxima above u_t are very close one to each other. The two global maxima should be both contained by a connected component at u which is still high (just slightly lower than u_t). By Morse theory, there must be a critical point in this connected component at u such that its index is smaller than N and its value is between u and u_t. However, from Lemma 3.4.5, we have learnt that the critical points above u with index smaller than N are very few compared with the local maxima above u. Therefore, such a situation should be rare, which corresponds to the strong repulsion at a very short distance.

2. Relatively strong attraction at a short distance: since the covariance function is positive, a high global maximum may have some "1ifting" effect on other points nearby. Thus, it is more likely to get another global maximum above u_t at a small distance from a high global maximum.

3. Decreasing attraction as the distance increases: as the distance increases, the "1ifting" effect will diminish due to the decreasing dependence, and the critical points behave more and more like in a Poisson point process.

Due to the similarity in the interactions discussed above, we turn to particle physics for a suitable model. An influential model to describe the interatomic interaction is the Morse potential ([27]) possessing the form

$$V_M(r) := D\left(1 - e^{-a(r-c)}\right)^2, \quad r > 0,$$

where D, a and c are all positive constants. This is in terms of the energy (potential). To turn energy into probability density, we use the Gibbs measure ([13]), which has the form

$$q_G(r) \propto \exp(-\beta V(r)), \quad r > 0,$$

where $\beta > 0$ and the function V is often interpreted as potential energy in physics applications. Then we can define the Gibbs-Morse density by

$$q_{GR}(r) \propto I_{\{0<r<r_H\}} \exp(-V_M(r)), \quad r > 0.$$

Plots of $q_{GR}(r)$ with $r_H = 2$ are provided in Figure 5.3. Finally, the density of R in the model is defined by

$$p(r) := Cr^{N-1} I_{\{0<r<r_H\}} \exp(-V_M(r)), \quad r > 0,$$

where

$$C := \left(\int_0^{r_H} r^{N-1} \exp(-V_M(r)) dr \right)^{-1}.$$

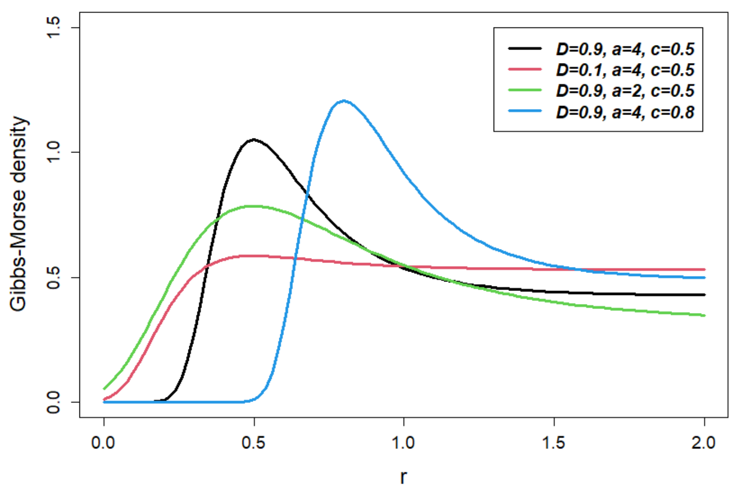


Figure 5.3: Plots of the Gibbs-Morse densities with $r_H = 2$.

5.4.3 Stage 3: Modeling Critical Points

In this stage, we will construct a model for $\xi^C_{u_t}$ consisting of the clustering process $G_1 \cup G_2$ of order one in Stage 2 and its daughter process G_3. Our algorithm to produce G_3 from $G_1 \cup G_2$ is as follows.

For each point in $G_1 \cup G_2$, a Bernoulli trial with success rate p_2 is run to determine whether it produces daughters in its neighborhood: if successful, it produces two daughters; otherwise, it does not produce any daughters (intuitively, if a global maximum is the member of a family of three, then it has two daughters: a local maximum and a saddle point in the same family; otherwise, it has no daughters). These Bernoulli trials are mutually independent and also independent of all other randomness in the model. Then G_3 is defined as the set of the daughters produced in this way.

In fact, a parent in $G_1 \cup G_2$ and its two daughters in G_3 also form a family of three in the model, which can be considered as the counterpart of a family of three in $\xi^C_{u_t}$. In each such family, denote by 0 the parent, by 1 and 2 the two daughters, by $\triangle 012$ the triangle formed by these three points, and by r_{ij} the distance between members i and j for $i, j = 1, 2, 3$. Denote by $\alpha_1 \in (\frac{\pi}{2}, \pi)$ the interior angle of $\triangle 123$ at 1. An illustration of a family of three of the model is provided in Figure 5.4.

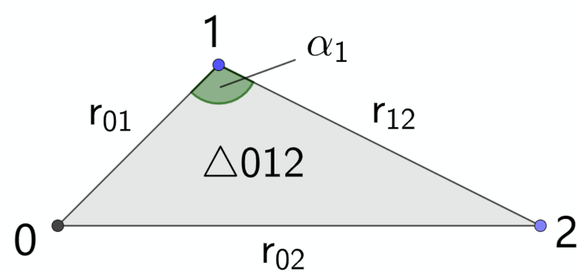


Figure 5.4: An illustration of a family of three of the proposed model.

Let $\theta_1 := W\alpha_1$, where W is a Rademacher random variable. These Rademacher random variables are mutually independent and also independent of all other randomness. Note that by the law of sines,

$$\frac{\sin(\theta_1)}{r_{02}} = \frac{\sin(W\alpha_1)}{r_{02}} = \frac{W}{2R_{123}},$$

where R_{123} is the radius of the circumscribed circle of $\triangle 123$. If $\frac{1}{R_{123}}$ is integrable, we must have

$$\mathbb{E}\left[\frac{\sin(\theta_1)}{r_{02}}\right] = \mathbb{E}[W]\mathbb{E}\left[\frac{1}{2R_{123}}\right] = 0.$$

Then we set the distribution of $(\log(r_{01}), \log(r_{02}), \frac{\sin(\theta_1)}{r_{02}})$ to be a multivariate Gaussian distribution, $\mathcal{N}(\boldsymbol{\nu}, \boldsymbol{\Sigma})$, with mean vector $\boldsymbol{\nu} := (\nu_1, \nu_2, 0)$ and covariance matrix

$$\boldsymbol{\Sigma} := \begin{pmatrix} \sigma_{11}^2 & \rho_{12}\sigma_{11}\sigma_{22} & 0 \\ \rho_{12}\sigma_{11}\sigma_{22} & \sigma_{22}^2 & 0 \\ 0 & 0 & \sigma_{33}^2 \end{pmatrix},$$

where $(\nu_1, \nu_2) \in \mathbb{R}^2$, $\rho_{12} \in (-1, 1)$, and $(\sigma_{11}, \sigma_{22}, \sigma_{33}) \in (0, \infty)^3$.

The motivation of the above setting comes from both theoretical analysis and simulation results. More precisely, the Rademacher random variable is a result of the symmetry implied by the isotropy. The joint normality and independence between $(\log(r_{01}), \log(r_{02}))$ and $\frac{\sin(\theta_1)}{r_{02}}$ are observed in simulation (for their counterparts in $\xi^C_{u_t}$) with several different covariance functions.

Note that r_{01}, r_{02} and θ_1, derived from the above Gaussian distribution, can only determine the shape of a family of three. We still need some directions to fully determine the positions of the two daughter points 1 and 2 relative to their parent 0. Since the underlying random field is isotropic, a uniform random direction, together with r_{02}, can determine the position of the daughter 2, but the daughter 1 can still be located at any point of a sphere of dimension $N-2$. The $(N-1)$-dimensional plane containing the sphere is perpendicular to $\vec{02}$, and the sphere is centered at the intersection of the plane with the line of $\vec{02}$. Therefore, we need another random vector uniformly distributed on the sphere to determine the position of the daughter 1. This completes the construction of our model.

5.5 Estimation of Parameters

In this section, we will illustrate the estimation of the parameters in the proposed model in two examples where the Gaussian random fields are defined on $\mathbb{R}^2$ with specific covariance structures. In total, there are 13 parameters in the proposed model, and we can divide them into the following groups to make the structure of this section clearer:

1. p_1 in Stage 2 and p_2 in Stage 3;

2. r_H and μ_P in Stage 1;

3. $\boldsymbol{\theta} := (D, a, c)$ in Stage 2;

4. $\boldsymbol{\gamma} := (\nu_1, \nu_2, \rho_{12}, \sigma^2_{11}, \sigma^2_{22}, \sigma^2_{33})$, i.e, the parameters in $\boldsymbol{\nu}$ and $\boldsymbol{\Sigma}$ in Stage 3.

First of all, we need to make some preparations.

5.5.1 Selections of the Hard-core Distance for an Isotropic Gaussian Random Field

The first preparation is to determine a range for the hard-core distance r'_H.

Recall that r'_H should be large enough to ensure the interaction between two points of G_1 are weak. Since the random field is Gaussian, this interaction is characterized by their covariance. Thus, to control this interaction, for a centered isotropic Gaussian random field X on $\mathbb{R}^2$ with covariance function $\rho(\|\boldsymbol{t}\|^2) := \mathrm{Cov}\,[X(\mathbf{0}), X(\boldsymbol{t})]$, $\boldsymbol{t} \in \mathbb{R}^2$, we require that r'_H satisfy

$$\rho(r'^2_H) \leq \alpha, \tag{5.6}$$

for some $\alpha > 0$.

5.5.2 Range of the Threshold

As stated earlier, our goal is to model the critical points when the threshold u is high but not very high. In our model, this is translated into requiring almost all the critical points above u to be still typical (see the beginning of Section 5.4). As a result, a value of u which would allow a satisfactory approximation using the proposed model should make the ratios

$$\frac{P(\xi^M_{u_t}(B(\mathbf{0}_N, r'_H)) \geq 2)}{P(\xi^M_{u_t}(B(\mathbf{0}_N, r'_H)) = 1)} \text{ and } \frac{P(\xi^M_{u_t}(B(\mathbf{0}_N, r'_H)) \geq 3)}{P(\xi^M_{u_t}(B(\mathbf{0}_N, r'_H)) = 2)}$$

both very small. Note that in the above two ratios, if replacing $\xi^M_{u_t}$ with a stationary Poisson point process, then both of them are almost proportional to the intensity of this Poisson point process. Considering that the behavior of $\xi^M_{u_t}$ should not be too far from that of a stationary Poisson point process when u_t is high, we can control these two ratios by controlling the expectation $\mathbb{E}[\xi^M_{u_t}(B(\mathbf{0}_N, r'_H))]$.

Since a global maximum must be a local maximum, it is sufficient to control the expected number $\mathbb{E}[M_{u_t}(X, B(\mathbf{0}_N, r'_H))]$ of the local maxima of X in $A_{u_t}(X, B(\mathbf{0}_N, r'_H))$. Recall that an integral expression of $\mathbb{E}[M_{u_t}(X, B(\mathbf{0}_N, r'_H))]$ is provided in Corollary 3.2.1. Since there is no simple expression of this integral, one can use a numerical method to get an approximation of this integral. Alternatively, one can use the asymptotic expression of this integral in Lemma 3.2.2, by which r'_H and u_t can be selected to satisfy

$$M(r'_H, u_t) := \frac{\lambda_N(B(\mathbf{0}_N, r'_H))\sqrt{\det(\boldsymbol{\Lambda})}\,u_t^{N-1}}{(2\pi)^{(N+1)/2}\sigma^{2N-1}} \exp\left(-\frac{u_t^2}{2\sigma^2}\right) < \beta \tag{5.7}$$

for some $\beta > 0$, where σ and $\boldsymbol{\Lambda}$ are as defined in Lemma 3.2.2.

In fact, (5.6) and (5.7) jointly determine a lower bound of the threshold to be considered as "high but not very high". In the following treatments, we take

$$\alpha = 0.0003 \text{ and } \beta = 0.15. \tag{5.8}$$

According to our empirical studies, the performance of the proposed model is not sensitive to the changes of α and β when they are both reasonably small.

5.5.3 Border Effect

To correctly identify G'_i, $i = 1, 2, 3$ in the search region S, we need to generate sample functions of X on a larger region $S' \supset S$ such that with a very large probability,

1. for every local maximum in S, the connected component it lies in is fully contained in S';

2. for any global maximum $\boldsymbol{t} \in S$, we have $B(\boldsymbol{t}, r'_H) \subset S'$.

5.5.4 Estimation of p_1 and p_2

For $i = 1, 2, 3$, denote by μ_i the intensity (i.e., the average number of points in a unit of area) of G_i, and by μ'_i the intensity of G'_i. Based on the definitions of p_1 and p_2 in Stages 1 and 2 respectively, we can derive that

$$p_1 = \frac{\mu_2}{\mu_1} \text{ and } p_2 = \frac{\mu_3}{\mu_1 + \mu_2}. \tag{5.9}$$

For $i = 1, 2, 3$, let $Y_i = \{Y_{i1}, \ldots, Y_{in}\}$ be independent copies of $G'_i(S)/\lambda_N(S)$, and let $\bar{Y}_i$ be the average of Y_i. For convenience, we also define $Y_4 := \{Y_{11} + Y_{21}, \ldots, Y_{1n} + Y_{2n}\}$, and similarly, let $\bar{Y}_4$ be the average of Y_4 (i.e., $\bar{Y}_4 = \bar{Y}_1 + \bar{Y}_2$). Then it is easy to see

$$\mathbb{E}[Y_{ij}] = \mu'_i$$

for any $1 \leq i \leq 4$ and $1 \leq j \leq n$. Thus, we can use the sample mean $\bar{Y}_i$ to estimate μ_i. Then, by (5.9), the consistent estimators of p_1 and p_2 can be

$$\hat{p}_1 := \frac{\bar{Y}_2}{\bar{Y}_1} \text{ and } \hat{p}_2 := \frac{1}{2}\frac{\bar{Y}_3}{\bar{Y}_4}, \tag{5.10}$$

respectively.

For $i = 1, 2, 3, 4$, let $S^2_{Y_i} := \frac{1}{n-1}\sum_{j=1}^{n}(Y_{ij} - \bar{Y}_i)^2$. For any $1 \leq i_1, i_2 \leq 4$ and $i_1 \neq i_2$, let $S_{Y_{i_1}Y_{i_2}} := \frac{1}{n-1}\sum_{j=1}^{n}(Y_{i_1 j} - \bar{Y}_{i_1})(Y_{i_2 j} - \bar{Y}_{i_2})$. Then the consistent estimators (see [43]) of the variances of $\hat{p}_1$ and $\hat{p}_2$ can be

$$\widehat{\mathrm{Var}}(\hat{p}_1) := \frac{1}{n}\left(\frac{S^2_{Y_2}}{\bar{Y}_1^2} + \frac{\bar{Y}_2^2 S^2_{Y_1}}{\bar{Y}_1^4} - \frac{2\bar{Y}_2 S_{Y_1 Y_2}}{\bar{Y}_1^3}\right)$$

and

$$\widehat{\text{Var}}(\hat{p}_2) := \frac{1}{4n}\left(\frac{S_{Y_3}^2}{\bar{Y}_4^2} + \frac{\bar{Y}_3^2 S_{Y_4}^2}{\bar{Y}_4^4} - \frac{2\bar{Y}_3 S_{Y_3Y_4}}{\bar{Y}_4^3}\right),$$

respectively.

5.5.5 Estimation of r_H and μ_P

Since the hard-core process G_1' is modeled by the second Matérn process G_1 and they share the same thinning rule, it is natural to set $r_H = r_H'$.

As for the estimation of μ_P, by (5.2), we have

$$\mu_1 = \mu_H = \frac{1 - e^{-\mu_P \pi r_H^2}}{\pi r_H^2}.$$

Since $\bar{Y}_1$ can be the estimator of μ_1, we set the estimator of μ_P

$$\hat{\mu}_P := -\frac{\log\left(1 - \pi r_H'^2 \bar{Y}_1\right)}{\pi r_H'^2}.$$

It is easy to see $\widehat{\text{Var}}(\bar{Y}_1) := S_{Y_1}^2/n$ is an unbiased estimator of the variance of the sample mean $\bar{Y}_1$. Then using the delta method, an estimator of the variance of $\hat{\mu}_P$ is

$$\widehat{\text{Var}}\left(\hat{\mu}_P\right) := g'\left(\bar{Y}_1\right)^2 \widehat{\text{Var}}\left(\bar{Y}_1\right) = \frac{S_{Y_1}^2}{n\left(1 - \pi r_H'^2 \bar{Y}_1\right)^2},$$

where $g(y) := -\frac{log(1-\pi r_H'^2 y)}{\pi r_H'^2}$, $y \geq 0$.

5.5.6 Estimation of D, a and c

We use the maximum likelihood estimation to estimate the parameters $\boldsymbol{\theta} = (D, a, c) \in (0, \infty)^3$ in the density

$$p(r) := C(\boldsymbol{\theta}) r^{N-1} I_{\{0<r<r_H\}} \exp\left(-D\left(1 - e^{-a(r-c)}\right)^2\right), \quad r > 0,$$

where

$$C(\boldsymbol{\theta}) := \left(\int_0^{r_H} r^{N-1} \exp\left(-D\left(1 - e^{-a(r-c)}\right)^2\right) dr\right)^{-1}.$$

Let $R_1, \ldots, R_k$ be independent copies of R', the distance between a random point in G'_2 and its parent in G'_1 (see Stage 2). Note that $p(r)$ is a model for the density of R'. Then the log-likelihood function of $p(r)$ based on $R_1, \ldots, R_k$ is

$$\ell(\boldsymbol{\theta}|R_1, \ldots, R_k) = k \log(C(\boldsymbol{\theta})) + \sum_{j=1}^{k} \left((N-1)\log(R_j) - D\left(1 - e^{-a(R_j - c)}\right)^2 \right).$$

Then the maximum likelihood estimator of $\boldsymbol{\theta}$ is defined by

$$\hat{\boldsymbol{\theta}} := \underset{\boldsymbol{\theta} \in (0,\infty)^3}{\operatorname{argmax}} \, \ell(\boldsymbol{\theta}|R_1, \ldots, R_k).$$

In practice, the maximum likelihood estimate of $\boldsymbol{\theta}$ can be numerically solved. By the asymptotic efficiency of the maximum likelihood estimator ([9]), the inverse of the Hessian matrix of the log-likelihood function at the maximum likelihood estimate can be used as the estimate of the covariance matrix of $\hat{\boldsymbol{\theta}}$.

5.5.7 Estimation of Parameters in $\boldsymbol{\nu}$ and $\boldsymbol{\Sigma}$

Let $\Gamma := \mathbb{R}^2 \times (-1, 1) \times (0, \infty)^3$. We also use the maximum likelihood estimation to estimate the parameters

$$\boldsymbol{\gamma} := (\nu_1, \nu_2, \rho_{12}, \sigma_{11}^2, \sigma_{22}^2, \sigma_{33}^2) \in \Gamma, \tag{5.11}$$

i.e, the parameters of the multivariate Gaussian distribution $\mathcal{N}(\boldsymbol{\nu}, \boldsymbol{\Sigma})$ with density

$$\rho_3(\boldsymbol{z}|\boldsymbol{\gamma}) = \frac{1}{\sqrt{(2\pi)^3 \det(\boldsymbol{\Sigma})}} \exp\left(-\frac{1}{2}(\boldsymbol{z} - \boldsymbol{\nu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{z} - \boldsymbol{\nu}) \right), \quad \boldsymbol{z} \in \mathbb{R}^3, \tag{5.12}$$

where $()^T$ represents the transpose of a matrix,

$$\boldsymbol{\nu} = (\nu_1, \nu_2, 0) \text{ and } \boldsymbol{\Sigma} = \begin{pmatrix} \sigma_{11}^2 & \rho_{12}\sigma_{11}\sigma_{22} & 0 \\ \rho_{12}\sigma_{11}\sigma_{22} & \sigma_{22}^2 & 0 \\ 0 & 0 & \sigma_{33}^2 \end{pmatrix}.$$

Let $\boldsymbol{Z}_i = (Z_{i1}, Z_{i2}, Z_{i3})$, $1 \leq i \leq m$ be a random sample from the distribution of the counterpart of $(\log(r_{01}), \log(r_{02}), \frac{\sin(\theta_1)}{r_{02}})$ in $\xi_{u_t}^C$ (see 5.4.3). Then by (5.12), the log-likelihood function based on this random sample is

$$\ell(\boldsymbol{\gamma}|\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m) = -\frac{3m}{2}\log(2\pi) - \frac{m}{2}\log(\det(\boldsymbol{\Sigma})) - \frac{1}{2}\sum_{i=1}^{m}(\boldsymbol{Z}_i - \boldsymbol{\nu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{Z}_i - \boldsymbol{\nu}).$$

Then the maximum likelihood estimator of $\boldsymbol{\gamma}$ based on $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m$ is defined by

$$\hat{\boldsymbol{\gamma}} := \operatorname*{argmax}_{\boldsymbol{\gamma} \in \Gamma} \ell(\boldsymbol{\gamma} | \boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_m).$$

Similarly as in Section 5.5.6, we can obtain the maximum likelihood estimate of $\boldsymbol{\gamma}$ by a numerical method, and the inverse of the Hessian matrix of the log-likelihood function at the maximum likelihood estimate can be used as the estimate of the covariance matrix of $\hat{\boldsymbol{\gamma}}$.

5.6 Empirical Analysis

5.6.1 Collecting Data from a Grid

In this section, we apply the proposed model to centered and isotropic Gaussian random fields indexed by $\mathbb{R}^N$. Two covariance functions are considered in this empirical study. The first one is a Gaussian covariance function (or a squared exponential covariance function) of the form

$$C_1(r) := \exp(-4r^2) \quad r > 0.$$

The second one is a Cauchy covariance function of the form

$$C_2(r) := \frac{1}{(1 + r^2)^3}.$$

Figure 5.5 shows how they decrease as r increases. Since $C_1(r)$ decreases much faster than $C_2(r)$, the interactions between two critical points can be very different in these two configurations, which allows us to verify the general applicability of the proposed model. In addition, for both covariance functions, the sample functions of the Gaussian random fields can be generated fast and easily using R ([32]) and the R package "RandomFields" ([39]). This makes it possible to later evaluate the performance of our model by comparing its outcome with the average of a large number of sample functions of the random fields.

According to Constraints (5.6) and (5.7), for each of the two covariance functions, say $C(r)$, we can select the hard-core distance r'_H and threshold u_t such that $C(r'_H) = \rho(r'^2_H) < \alpha = 0.0003$ and $M(r'_H, u_t) < \beta = 0.15$, as suggested in (5.8).

We also select $\delta_p = 0.01$ for the modification in Section 5.3 to get the persistent critical points above u_t. Note that the global maximum of a simple connected component at u_t could also be the local maximum of a short-lived family of three at u for some $u < u_t$. To

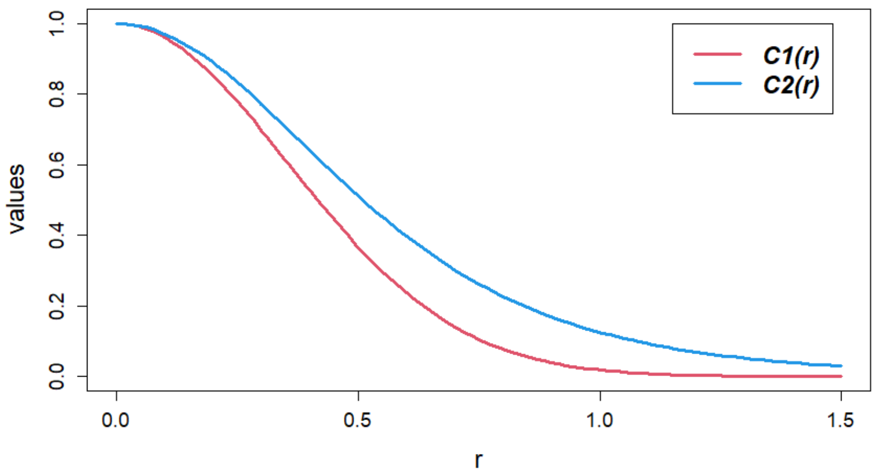


Figure 5.5: Covariance functions $C_i(r)$, $i = 1, 2$.

remove the effect of all the short-lived topological features from our data, we need to select a threshold $u_b < u_t$ and then modify every family of three at every $u \in [u_b, \infty)$ (from low to high). Here we select $u_b = 3.2$ for both covariance functions.

Let $S' = [0, 100]^2$. This region should be as large as possible, such that it allows a large search region S to diminish the bias of the estimator of a Ripley's L function (see Section 5.2.2). For $C_i(r)$, $i = 1, 2$, we generate 12000 and 6000 sample functions on S', respectively. Then based on these sample functions, S can be set as $[12, 88]^2$ such that the two properties in Section 5.5.3 hold. The selection results are all exhibited in Table 5.1.

Cov	r'_H	$\rho(r'^2_H)$	u_t	$M(r'_H, u_t)$	u_b	δ_p	S'	S	n
$C_1(r)$	2	0.0000	3.5	0.0069	3.2	0.01	$[0, 100]^2$	$[12, 88]^2$	12000
$C_2(r)$	4	0.0002	3.5	0.1466	3.2	0.01	$[0, 100]^2$	$[12, 88]^2$	6000

Table 5.1: Selections of r'_H, u_t, u_b, δ_p, S', S and the sample size n for $C_i(r)$, $i = 1, 2$.

Let $H := \left\{\left(\frac{i}{128}, \frac{j}{128}\right), i, j \in \mathbb{Z}\right\}$. Since we can only generate sample functions on a finite set, we take $S' \cap H$ as a discretization of S' and generate sample functions on it. Thus, for any $u \geq u_b$ and $K \subset S'$, a generated excursion set of a sample function f at u in K is $A_u(f, K) \cap H$.

For a connected component C_u of $A_u(f, \mathbb{R}^N)$, let m_k, $0 \leq k \leq N$ be the number of non-degenerate critical points of index k in C_u. By the definition of typical connected components, even if $m_N = 2$ and $m_{N-1} = 1$, C_u may not be typical since it may have $m_k > 0$ for some $0 \leq k \leq N-2$. However, when $N = 2$, by Corollary 9.3.5 in [4] and Theorem 9.1 in [26], we have

$$m_2 - m_1 + m_0 = 1 - b_1 \leq 1, \tag{5.13}$$

where $b_1 \geq 0$ is the 1st Betti number, a concept in homology and can be interpreted as the number of one-dimensional holes of C_{u_b}. Note that a typical connected component always satisfies $m_2 - m_1 = 1$ before or after the modification (see Section 5.3). Then by (5.13), we have $m_0 = 0$. This means that to locate all the persistent and typical critical points above u_t, it suffices to locate all the local maxima and saddle points above u_t.

Intuitively, for each sample function f, we can locate all the critical points in $A_{u_t}(f,S) \cap H$ by locating the positions where the gradients of f are very "close" to the zero vector. We call this the derivative method. However, it is hard to determine a unified threshold for the gradient to be close enough to 0 which works for every sample function. Inspired by the discussion on persistent critical points in Section 5.3, we can determine the positions, denoted by $PT_f(S, u_t)$, of all the persistent and typical critical points in $A_{u_t}(f,S) \cap H$ by the following **steps**:

1. Two points in $A_{u_b}(f,S') \cap H$ are considered to be in the same connected component if the distance between them is less than or equal to five grid distances. In this way, we can determine all the connected components of $A_{u_b}(f,S') \cap H$. Then we can determine the positions and values of the global maxima of these connected components.

2. Let C_{u_b} be a connected component of $A_{u_b}(f,S')$ not touching the boundary of S'. Define

   $$C_{u_b}^H = C_{u_b} \cap H \text{ and } u_m := \max_{\boldsymbol{t} \in C_{u_b}^H} f(\boldsymbol{t}).$$

   For any $u \in (u_b, u_m)$, define

   $$C_u^H := \left\{ \boldsymbol{t} \in C_{u_b}^H : f(\boldsymbol{t}) > u \right\}.$$

   As in Section 5.3, we look at the change of the number of connected components of C_u^H as u increases from u_b to u_m. The difference is that in practice, we can only increase u in a discrete manner: in the i-th iteration ($i \geq 1$),

   (a) count the number, denoted by n_u, of the connected components of C_u^H;

   (b) increase u by 0.001;

(c) if $u \geq u_m$, then stop.

Here we only increase u by a very small value 0.001 in each iteration. This is to ensure as much as possible that there is at most one critical point of C_{u_b} whose value is between u and $u+0.001$, and thus, n_u can increase or decrease by at most one in each iteration.

3. Assume that the above algorithm stops at the k-th iteration ($k \geq 1$). Let u_i be the value of u in the i-th iteration for $1 \leq i \leq k$ with $u_1 = u_b$. It is easy to see $\{C^H_{u_i}\}_{1\leq i\leq k}$ forms a decreasing sequence of finite sets. For any $A \subset \mathbb{R}^N$ and $\boldsymbol{t} \in \mathbb{R}^N$, the distance, $d_{\boldsymbol{t}}(A)$ between $\boldsymbol{t}$ and A is defined by

$$d_{\boldsymbol{t}}(A) := \sqrt{\inf_{\boldsymbol{s}\in A} \|\boldsymbol{t}-\boldsymbol{s}\|^2}.$$

Then for any $2 \leq i \leq k$,

- if $n_{u_i} = n_{u_{i-1}} + 1$, we can find a saddle point in $C^H_{u_i} \setminus C^H_{u_{i-1}}$: there must be only one connected component of $C^H_{u_{i-1}}$ splitting into two connected components of $C^H_{u_i}$, denoted by A_1 and B_1, and then the position of the saddle point is set to be a member of the set

$$\operatorname*{argmin}_{\boldsymbol{t}\in C^H_{u_i}\setminus C^H_{u_{i-1}}} \sqrt{d_{\boldsymbol{t}}(A_1)^2 + d_{\boldsymbol{t}}(B_1)^2}$$

since it should be very close to both connected components.

- if $n_{u_i} = n_{u_{i-1}} - 1$, we can find a local maximum in $C^H_{u_i} \setminus C^H_{u_{i-1}}$: there must be only one connected component of $C^H_{u_{i-1}}$ whose values are all less than u_i, and then the local maximum can be located at the global maximum of f in this connected component.

4. As far, we have obtained the positions of all the local maxima and saddle points in $A_{u_b}(f, S') \cap H$. Then we can

(a) modify all the short-lived families of three at u in S' from $u = u_b$ to $u = \infty$;

(b) remove the points whose values are less than u_t;

(c) remove the points in the non-typical connected components at u_t;

(d) remove the points outside S

to get the set $PT_f(S, u_t)$ of positions of all the persistent and typical critical points in $A_{u_t}(f, S) \cap H$.

The above algorithm cannot determine the number and positions of the local minima above u_t since n_u will not change when the threshold u passes the value of a local minimum. However, to measure the performance of our model, we still need to locate the positions of all the local minimum to obtain an estimate of the L function of the persistent critical points above u_t (which can be local minima). However, note that local minima can only appear in non-typical connected components at u_t whose number is very small. As such, we omit these connected components when estimating the L function.

In fact, $PT_f(S, u_t)$ represents a realization of $G'_1 \cup G'_2 \cup G'_3$ in S. One can also identify G'_3 in S from $PT_f(S, u_t)$ using the information of all the critical points in S' in Step 2. However, to identify G'_1 (and also G'_2) in S from $PT_f(S, u_t)$, one needs more information outside S. More specifically, this requires to first apply the thinning rule (5.1) to $PT_f(S', u_t)$ to identify the first generations in S' (not just S), and then only retain the first generations in S. One should note that the identification of the first generation in $S' \setminus S$ may not be accurate since it still needs information outside S' (see Section 5.5.3). However, the identification in S is accurate since for any point in $PT_f(S, u_t)$, all of its r'_H-neighbors in $PT_f(S', u_t)$ can be correctly identified.

Applying the above steps on every generated sample function gives a sample of (discretized) $G_i(S)$ for $i = 1, 2, 3$. Base on these samples, we can obtain the estimates of p_1, p_2 (see Section 5.5.4), μ_P (see Section 5.5.5), and their variances.

When applying the thinning rule (5.1) to $PT_f(S', u_t)$, for each deleted point, if its parent is in S, then the distance between them will be recorded. The collection of all these distances over all the sample functions forms a sample of R' which can be used to estimate D, a, c and their variances as in Section 5.5.6.

In the above steps, we can find all the families of three at u_t whose global maxima are in S. By decorating a "$'$" on each symbol (as shown in Figure 5.4), we can also define symbols for a family of three in $\xi^C_{u_t}$ where $0'$, $1'$ and $2'$ represent the global maximum, the saddle point and the local maximum, respectively. Define

$$\theta' := \operatorname{sign}\left((x_0 - x_1)(y_2 - y_1) - (x_2 - x_1)(y_0 - y_1)\right)\alpha'_1,$$

where (x_i, y_i) is the position of the point i' for $i = 1, 2, 3$. Then from these families, we can obtain a sample of $(\log(r'_{01}), \log(r'_{02}), \frac{\sin(\theta'_1)}{r'_{02}})$ which can be used to estimate parameters in $\boldsymbol{\nu}$, $\boldsymbol{\Sigma}$ and their variances as in Section 5.5.7.

5.6.2 Estimation Results

The estimation results of the parameters in our model are exhibited in Table 5.2.

	$C_1(r)$		$C_2(r)$	
	est.	s.d.	est.	s.d.
λ	0.003889	0.000008	0.002856	0.000010
p_1	0.024123	0.000308	0.088697	0.001029
D	0.134664	0.037141	1.380521	0.038231
a	3.991681	0.361824	2.549077	0.087810
c	0.754681	0.029225	0.665931	0.009528
p_2	0.002531	0.000098	0.009082	0.000304
ν_1	-1.189444	0.010756	-1.151694	0.010121
ν_2	-0.688618	0.009082	-0.672078	0.008687
ρ_{12}	0.875685	0.008909	0.867642	0.008178
σ_{11}^2	0.079131	0.004272	0.093530	0.004373
σ_{22}^2	0.056422	0.003044	0.068901	0.003221
σ_{33}^2	0.138004	0.007462	0.227393	0.010643

Table 5.2: The estimates (est.) of the parameters and their standard errors (s.d.).

Recall that in Stage 2, we use the Gibbs-Morse distribution to model the distribution of de-weighted R'. Based on the estimates of D, a and c, we provide a comparison between the fitted Gibbs-Morse distribution and the empirical distribution of de-weighted R' for each selected covariance function in Figure 5.6. One can see that for both covariance functions, the Gibbs-Morse distribution has an impressively good fit.

In Stage 3, we have the following hypotheses:

(i) $(\log(r_{01}), \log(r_{02}), \frac{\sin(\theta_1)}{r_{02}})$ is normally distributed;

(ii) $(\log(r_{01}), \log(r_{02}))$ and $\frac{\sin(\theta_1)}{r_{02}}$ are independent;

(iii) $\mathbb{E}[\frac{\sin(\theta_1')}{r_{02}'}] = 0$.

One should not that (i) and (ii) are the settings of our model, but (iii) only serves as the motivation for our model (and its version with all the "$'$" removed can be derived from the construction of the model). These hypotheses can be tested based on the sample of $(\log(r_{01}'), \log(r_{02}'), \frac{\sin(\theta_1')}{r_{02}'})$ obtained from the last section. We adopt the Henze-Zirkler Test ([16]) for (i), the distance correlation test ([41]) for (ii), and one sample t-test for (iii). The p-values and sample sizes are all exhibited in Table 5.3, and there are no rejections of the null hypotheses at the significance level 5%.

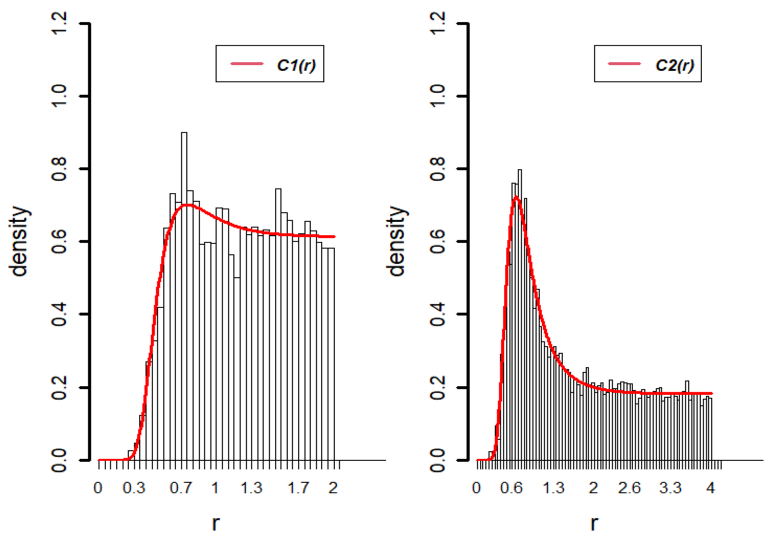


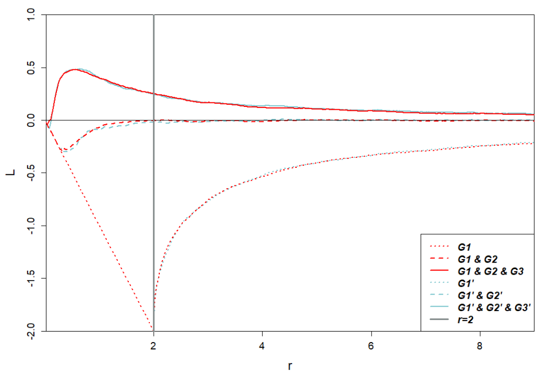
Figure 5.6: The fitted Gibbs-Morse densities and the histograms of the de-weighted R' for $C_i(r)$, $i = 1, 2$.

	$C_1(r)$		$C_2(r)$	
	p-value	n	p-value	n
HZ	0.17	684	0.24	913 (1)
dcor	0.19	684	0.96	913 (1)
t	0.28	684	0.95	913 (1)

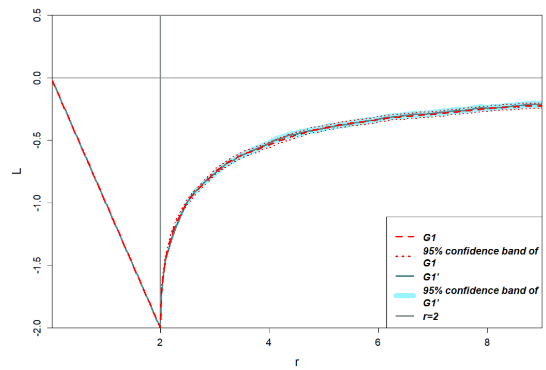
Table 5.3: The p-values and sample sizes of the hypothesis tests based on the sample of $(\log(r'_{01}), \log(r'_{02}), \frac{\sin(\theta'_1)}{r'_{02}})$. The number of outliers which have being removed in each hypothesis test is shown in the parentheses.

For each selected covariance function, we have generated 20000 samples of the critical points from the proposed model using the parameters estimated in Table 5.2. The quality of simulation are evaluated by a comparison between the L function of the proposed model and the L function of the critical points $\xi_{u_t}^C$. Since the proposed model has three stages, we can also compare the L functions of G_1, $G_1 \cup G_2$ with the L functions of G_1', $G_1' \cup G_2'$ $(= \xi_u^M)$, respectively. All the L functions are estimated by the $\bar{L}$ functions (see Section 5.2.2), and all these comparisons, together with the estimated 95% confidence band, are shown in Figures 5.7 and 5.8. We also use the stationary Poisson point process as the benchmark model, and recall that its theoretical L function is simply the zero function.

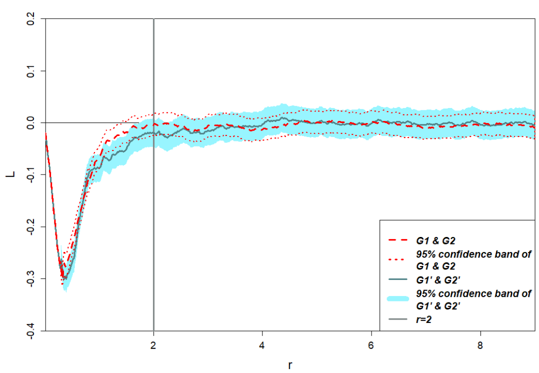
From Figures 5.7 and 5.8, we see that for every selected covariance functions and every stage, our model has an impressively good fit to the corresponding point process of the critical points of the isotropic Gaussian random field above a not-so-high threshold. In Figure 5.8, we see that there are still some small discrepancies between the estimated L functions of the proposed model and the estimated L functions of persistent critical points. These discrepancies start to get bigger from the hard-core distance. This may be because the Cauchy covariance function $C_2(r)$ decreases so slow with increasing distance that some interactions are not captured by the model, and also an L function, by its definition, will accumulate the discrepancy as distance increases.



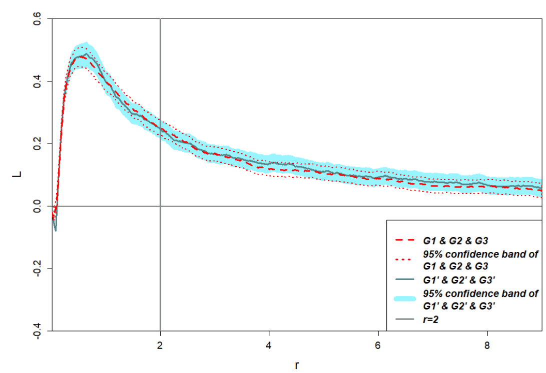
(a) Estimated L functions.



(b) Estimated L functions of G_1 and G_1' with their 95% confidence bands.

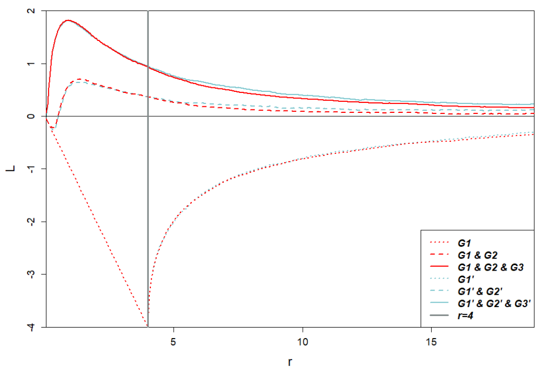


(c) Estimated L functions of $G_1 \cup G_2$ and $G_1' \cup G_2'$ with their 95% confidence bands.

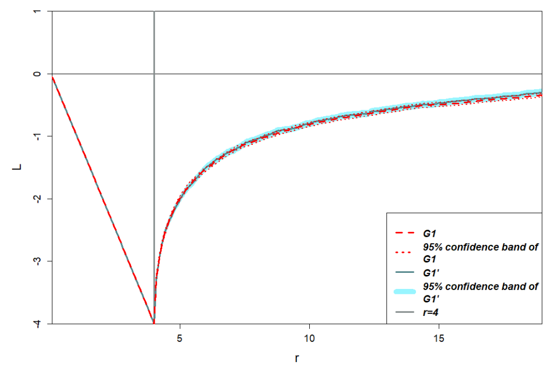


(d) Estimated L function of $G_1 \cup G_2 \cup G_3$ and $G_1' \cup G_2' \cup G_3'$ with their 95% confidence bands.

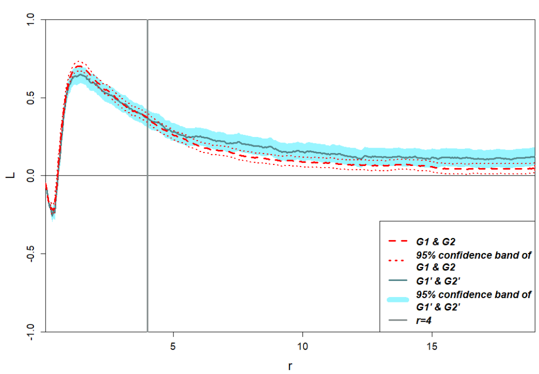
Figure 5.7: Comparisons between the estimated L functions of $\bigcup_{i=1}^{k} G_i$ and the estimated L functions of $\bigcup_{i=1}^{k} G_i'$ for $k = 1, 2, 3$ respectively, where the covariance function is $C_1(r) = \exp(-4r^2)$, $r \geq 0$.



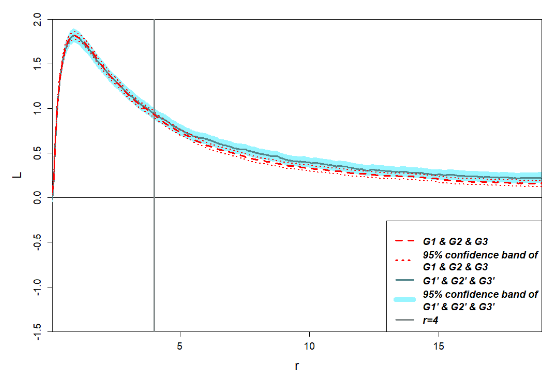
(a) Estimated L functions.



(b) Estimated L functions of G_1 and G_1' with their 95% confidence bands.



(c) Estimated L functions of $G_1 \cup G_2$ and $G_1' \cup G_2'$ with their 95% confidence bands.



(d) Estimated L function of $G_1 \cup G_2 \cup G_3$ and $G_1' \cup G_2' \cup G_3'$ with their 95% confidence bands.

Figure 5.8: Comparisons between the estimated L functions of $\bigcup_{i=1}^{k} G_i$ and the estimated L functions of $\bigcup_{i=1}^{k} G_i'$ for $k = 1, 2, 3$ respectively, where the covariance function is $C_2(r) = \frac{1}{(1+r^2)^3}$, $r \geq 0$. In estimating the L functions for $\bigcup_{i=1}^{k} G_i'$, $k = 1, 2, 3$, we dropped 85 connected components at $u_b = 3.2$ due to the numerical precision of the method used to identify the critical points.

References

[1] R. J. Adler. *The Geometry of Random Fields*. Wiley, New York, 1981.

[2] Robert J Adler, Omer Bobrowski, Matthew S Borman, Eliran Subag, and Shmuel Weinberger. Persistent homology for random fields and complexes. In *Borrowing strength: theory powering applications–a Festschrift for Lawrence D. Brown*, pages 124–143. Institute of Mathematical Statistics, 2010.

[3] Robert J Adler and Gennady Samorodnitsky. Climbing down Gaussian peaks. *The Annals of Probability*, 45(2):1160–1189, 2017.

[4] Robert J Adler, Jonathan E Taylor, et al. *Random fields and geometry*, volume 80. Springer, 2007.

[5] Jean-Marc Azaïs, José R León, and Joaquín Ortega. Geometrical characteristics of Gaussian sea waves. *Journal of applied probability*, 42(2):407–425, 2005.

[6] James M Bardeen, JR Bond, Nick Kaiser, and AS Szalay. The statistics of peaks of Gaussian random fields. *The Astrophysical Journal*, 304:15–61, 1986.

[7] Anastassia Baxevani and Igor Rychlik. Maxima for Gaussian seas. *Ocean Engineering*, 33(7):895–911, 2006.

[8] Julian Besag. Contribution to the discussion on Dr. Ripley's paper. *JR Stat. Soc.*, 39:193–195, 1977.

[9] George Casella and Roger L Berger. *Statistical inference*. Cengage Learning, 2021.

[10] Dan Cheng and Armin Schwartzman. Distribution of the height of local maxima of Gaussian random fields. *Extremes*, 18(2):213–240, 2015.

[11] Dan Cheng and Armin Schwartzman. Expected number and height distribution of critical points of smooth isotropic Gaussian random fields. *Bernoulli*, 24(4B):3422, 2018.

[12] Peter J Diggle. *Statistical analysis of spatial and spatio-temporal point patterns*. CRC press, 2013.

[13] Hans-Otto Georgii. *Gibbs measures and phase transitions*. de Gruyter, 2011.

[14] Robert Ghrist. Barcodes: the persistent topology of data. *Bulletin of the American Mathematical Society*, 45(1):61–75, 2008.

[15] François Goreaud and Raphaël Pélissier. On explicit formulas of edge effect correction for Ripley's k-function. *Journal of Vegetation Science*, 10(3):433–438, 1999.

[16] N Henze and B Zirkler. A class of invariant consistent tests for multivariate normality. *Communications in statistics-Theory and Methods*, 19(10):3595–3617, 1990.

[17] Roger A Horn and Charles R Johnson. *Matrix analysis*. Cambridge university press, 1985.

[18] Olav Kallenberg. *Random measures*. Academic Press, 1983.

[19] Jian Kang, Brian J. Reich, and Ana-Maria Staicu. Scalar-on-image regression via the soft-thresholded Gaussian process. *Biometrika*, 105(1):165–184, 2018.

[20] Tosio Kato. *Perturbation theory for linear operators*, volume 132. Springer Science & Business Media, 1980.

[21] Steven G Krantz and Harold R Parks. *A primer of real analytic functions*. Springer Science & Business Media, 2002.

[22] Andreas Kriegl, Peter W Michor, and Armin Rainer. Denjoy–Carleman differentiable perturbation of polynomials and unbounded operators. *Integral Equations and Operator Theory*, 71(3):407–416, 2011.

[23] Malcolm R Leadbetter, Georg Lindgren, and Holger Rootzén. *Extremes and related properties of random sequences and processes*. Springer Science & Business Media, 2012.

[24] Georg Lindgren. Local maxima of Gaussian fields. *Arkiv för matematik*, 10(1-2):195–218, 1972.

[25] B Matérn. Spatial variation: Meddelanden fran statens skogsforskningsinstitut. *Lecture Notes in Statistics*, 36:21, 1960.

[26] Marston Morse and Stewart S. Cairns. *Critical point theory in global analysis and differential topology: An introduction.* Academic Press, 1969.

[27] Philip M Morse. Diatomic molecules according to the wave mechanics. II. Vibrational levels. *Physical review*, 34(1):57, 1929.

[28] Jerzy Neyman and Elizabeth L Scott. Statistical approach to problems of cosmology. *Journal of the Royal Statistical Society: Series B (Methodological)*, 20(1):1–29, 1958.

[29] JE Paloheimo. On a theory of search. *Biometrika*, 58(1):61–75, 1971.

[30] James Pickands. Upcrossing probabilities for stationary Gaussian processes. *Transactions of the American Mathematical Society*, 145:51–73, 1969.

[31] Vladimir I. Piterbarg. *Asymptotic methods in the theory of Gaussian processes and fields*, volume 148. American Mathematical Soc., 1996.

[32] R Core Team. *R: A Language and Environment for Statistical Computing.* R Foundation for Statistical Computing, Vienna, Austria, 2022.

[33] Attilio J. Rainal. Origin of Rice's formula. *IEEE transactions on information theory*, 34(6):1383–1387, 1988.

[34] Sidney I. Resnick. *Extreme values, regular variation and point processes.* Springer, 2013.

[35] Brian D Ripley. The second-order analysis of stationary point processes. *Journal of applied probability*, 13(2):255–266, 1976.

[36] Brian D Ripley. Modelling spatial patterns. *Journal of the Royal Statistical Society: Series B (Methodological)*, 39(2):172–192, 1977.

[37] Brian D Ripley. *Statistical inference for spatial processes.* Cambridge university press, 1988.

[38] Peter A Rogerson. A statistical method for the detection of geographic clustering. *Geographical Analysis*, 33(3):215–227, 2001.

[39] Martin Schlather, Alexander Malinowski, Marco Oesting, Daphne Boecker, Kirstin Strokorb, Sebastian Engelke, Johannes Martini, Felix Ballani, Olga Moreva, Jonas Auel, Peter J Menck, Sebastian Gross, Ulrike Ober, Paulo Ribeiro, Brian D Ripley, Richard Singleton, Ben Pfaff, and R Core Team. *RandomFields: Simulation and Analysis of Random Fields*, 2022. R package version 3.3.14.

[40] Volker Springel, Simon DM White, Adrian Jenkins, Carlos S Frenk, Naoki Yoshida, Liang Gao, Julio Navarro, Robert Thacker, Darren Croton, John Helly, et al. Simulations of the formation, evolution and clustering of galaxies and quasars. *nature*, 435(7042):629–636, 2005.

[41] Gábor J Székely, Maria L Rizzo, and Nail K Bakirov. Measuring and testing dependence by correlation of distances. *The annals of statistics*, 35(6):2769–2794, 2007.

[42] Jonathan E Taylor and Keith J Worsley. Detecting sparse signals in random fields, with an application to brain mapping. *Journal of the American Statistical Association*, 102(479):913–928, 2007.

[43] GMP Van Kempen and LJ Van Vliet. Mean and variance of ratio estimators used in fluorescence ratio imaging. *Cytometry: The Journal of the International Society for Analytical Cytology*, 39(4):300–305, 2000.

APPENDICES

Appendix A

Appendix for Chapter 2

A.1 Properties of Gaussian Random Vectors

A real-valued random variable $X : (\Omega, \mathcal{A}, P) \to (\mathbb{R}, \mathcal{B}(\mathbb{R}^N))$ is said to be Gaussian (or normally distributed) if its probability density function, $\phi_X(x)$, can be written as

$$\phi_X(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \tag{A.1}$$

for some constants $\mu \in \mathbb{R}$ and $\sigma > 0$. Then by some calculations, it is easy to show that

$$\mathbb{E}[X] = \mu \quad \text{and} \quad \mathrm{Var}[X] = \sigma^2.$$

From above, we see that the distribution of the Gaussian random variable X can be fully characterized by its mean and variance, and thus, we can write $X \sim \mathcal{N}(\mu, \sigma^2)$, where "$\mathcal{N}''$ stands for "normally distributed". Especially, X is said to be a standard normal random variable, written as $X \sim \mathcal{N}(0, 1)$, when $\mathbb{E}[X] = 0$ and $\mathrm{Var}[X] = 1$. Then its probability density function will become

$$\phi(x) := \frac{1}{\sqrt{2\pi}} e^{-x^2}.$$

An important fact is: if $X \sim \mathcal{N}(\mu, \sigma^2)$ for some $\mu \in \mathbb{R}$ and $\sigma > 0$, then $\sigma^{-1}(X - \mu) \sim \mathcal{N}(0, 1)$. The simple transformation in this fact implies that many studies working on general normal distributions can actually be constrained on the standard normal distribution without loss of generality. Similar arguments will also apply when we introduce Gaussian random fields.

The characteristic function of $X \sim \mathcal{N}(\mu, \sigma^2)$ is defined by

$$\varphi_X(t) := \mathbb{E}\left[e^{itX}\right] = e^{it\mu - \frac{1}{2}\sigma^2 t^2}. \tag{A.2}$$

Due to the simple structure of the characteristic function of a Gaussian random variable, we have the following lemma.

Lemma A.1.1. *Let $X_k \sim \mathcal{N}(0, \sigma_k^2)$, $k \geq 1$. Then as $k \to \infty$,*

(i) if X_k converges to X_∞ in distribution, then

$$X_\infty \sim \mathcal{N}(0, \sigma_\infty^2),$$

where $\sigma_\infty^2 := \lim_{k\to\infty} \sigma_k^2 < \infty$.

(i) X_k converges to X_∞ in probability if and only if X_k converges to X_∞ in the mean square sense.

Proof. For (i), by the Lévy's continuity theorem, we have for any $t \in \mathbb{R}$,

$$\lim_{k\to\infty} \varphi_{X_k}(t) = \varphi_{X_\infty}(t).$$

By (A.2), $\varphi_{X_\infty}(t)$ must have the form

$$\varphi_{X_\infty}(t) = e^{-\frac{1}{2}\sigma_\infty^2 t^2},$$

where $\sigma_\infty^2 := \lim_{k\to\infty} \sigma_k^2 < \infty$.

As for (ii), the "if" direction is trivial. For the "only if" direction, by (i),

$$\lim_{k\to\infty} \mathbb{E}\left[X_k^4\right] = 3 \lim_{k\to\infty} \sigma_k^4 = 3\sigma_\infty^4 < \infty.$$

Then it is easy to show $\{(X_k - X_\infty)^2, k \geq 1\}$ is uniformly integrable. Since $(X_k - X_\infty)^2$ also converges to 0 in probability, we have $(X_k - X_\infty)^2$ converges to 0 in L^1, as desired. □

Denote by $\Psi(x)$ the tail distribution function of a standard Gaussian random variable $X \sim \mathcal{N}(0, 1)$, i.e.,

$$\Psi(x) := P\left[X > x\right] = \int_x^\infty \phi(u) du. \tag{A.3}$$

It is easy to see that

$$\lim_{x\to\infty} \Psi(x) = 0.$$

However, researchers are more concerned about the speed of this convergence in many asymptotic theories. Since there is no explicit expression for $\Psi(x)$, alternatively, we choose to bound $\Psi(x)$ by the following inequality (see also (1.2.2) in [4])

$$\left(\frac{1}{x}-\frac{1}{x^3}\right)\phi(x) < \Psi(x) < \frac{1}{x}\phi(x). \tag{A.4}$$

To show the lower bound in (A.4), we need to make the change of variable $u = x + v/x$. Then by the fact that $e^{-y} > 1 - y$ for all $y \geq 0$, we have

$$\begin{aligned}\int_x^\infty e^{-\frac{u^2}{2}}du &= \int_0^\infty x^{-1}e^{-(x^2+2v+v^2/x^2)/2}dv \\ &= x^{-1}e^{-x^2/2}\int_0^\infty e^{-(2v+v^2/x^2)/2}dv \\ &\geq x^{-1}e^{-x^2/2}\int_0^\infty e^{-v}(1-v^2/(2x^2))dv \\ &\geq x^{-1}e^{-x^2/2}(1-x^{-2}) \\ &= (x^{-1}-x^{-3})e^{-x^2/2}.\end{aligned}$$

Moreover, observe that

$$\int_x^\infty e^{-\frac{u^2}{2}}du \leq \int_x^\infty \frac{u}{x}e^{-\frac{u^2}{2}}du = x^{-1}e^{-x^2/2},$$

which implies the upper bound in (A.4). This inequality is rough but enough for the thesis.

A random n-vector $\boldsymbol{X} = (X_1, \ldots, X_n)^T, n \geq 1$, is said to be multivariate Gaussian distributed (or a Gaussian n-vector) with mean vector $\boldsymbol{\mu} \in \mathbb{R}^{n\times 1}$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$, written as $\boldsymbol{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, if every linear combination of its components $\{X_i, 1 \leq i \leq n\}$ is almost surely a constant or follows a univariate Gaussian distribution. More specifically, for every $\boldsymbol{a} \in \mathbb{R}^{n\times 1}$, we have

$$\boldsymbol{a}^T\boldsymbol{X} \sim \mathcal{N}(\boldsymbol{a}^T\boldsymbol{\mu}, \boldsymbol{a}^T\boldsymbol{\Sigma}\boldsymbol{a}),$$

where the univariate normal distribution $\mathcal{N}(\mu, 0)$ with mean $\mu \in \mathbb{R}$ and zero variance denotes a point mass on μ (almost surely equal to μ). The above definition implies that any multivariate Gaussian distribution can also be fully characterized by its mean vector and covariance matrix. By taking $\boldsymbol{a}_i \in \mathbb{R}^n$, $1 \leq i \leq n$, as the n-vector with all coordinates being zeros except for a one at the i-th coordinate, we see that each marginal $X_i = \boldsymbol{a}_i^T\boldsymbol{X}$ follows the univariate Gaussian distribution $\mathcal{N}(\mu_i, \sigma_{ii}^2)$. Especially, if $\boldsymbol{\Sigma}$ is positive-definite, then

$\boldsymbol{X} = (X_1, \ldots, X_n)^T, n \geq 1$, is said to be non-degenerate multivariate Gaussian distributed (or a non-degenerate Gaussian n-vector), and the joint density function of $X_1, \ldots, X_n$ at any point $\boldsymbol{t} \in \mathbb{R}^{n\times 1}$ can be written as

$$\phi_{\boldsymbol{X}}(\boldsymbol{t}) = \frac{1}{(2\pi)^{n/2}\det(\boldsymbol{\Sigma})^{1/2}} \exp\left\{-\frac{1}{2}(\boldsymbol{t}-\boldsymbol{\mu})^T\boldsymbol{\Sigma}^{-1}(\boldsymbol{t}-\boldsymbol{\mu})\right\}.$$

The following facts (see Section 1.2, [4]) about Gaussian n-vectors will be very useful in the thesis.

1. Let $\boldsymbol{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma})$, i.e., $\boldsymbol{X}$ is a column Gaussian n-vector with mean vector $\boldsymbol{\mu} \in \mathbb{R}^{n\times 1}$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$. Then for any $\boldsymbol{A} \in \mathbb{R}^{m\times n}$, we have

$$\boldsymbol{A}\boldsymbol{X} \sim \mathcal{N}_m(\boldsymbol{A}\boldsymbol{\mu}, \boldsymbol{A}\boldsymbol{\Sigma}\boldsymbol{A}^T).$$

2. Let $\boldsymbol{X}_k \sim \mathcal{N}_n(\boldsymbol{\mu}_k, \boldsymbol{\Sigma}_k)$, $k \geq 1$. Assume that the sequence $\{\boldsymbol{X}_k\}_{k=1}^{\infty}$ converges in mean square, i.e., there exists a random n-vector $\boldsymbol{X}$ such that

$$\lim_{k\to\infty} \mathbb{E}\left[\|\boldsymbol{X}_k - \boldsymbol{X}\|_n^2\right] = 0. \tag{A.5}$$

Then there exist a n-vector $\boldsymbol{\mu} \in \mathbb{R}^{n\times 1}$ and a positive semi-definite matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$, such that

$$\|\boldsymbol{\mu}_k - \boldsymbol{\mu}\|_n \to \infty \quad \text{and} \quad \|\boldsymbol{\Sigma}_k - \boldsymbol{\Sigma}\|_{n^2} \to \infty$$

as $k \to \infty$, and

$$\boldsymbol{X} \sim \mathcal{N}_n(\boldsymbol{\mu}, \boldsymbol{\Sigma}).$$

3. Let $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ be a Gaussian n-vector ($n \geq 2$) with mean vector $\boldsymbol{\mu} = (\mu_1, \ldots, \mu_n)^T \in \mathbb{R}^{n\times 1}$ and covariance matrix $\boldsymbol{\Sigma} \in \mathbb{R}^{n\times n}$. We separate the coordinates of $\boldsymbol{X}$ into two parts

$$\boldsymbol{X}_1 = (X_1, \ldots, X_{n_1})^T \quad \text{and} \quad \boldsymbol{X}_2 = (X_{n_1+1}, \ldots, X_n)^T$$

for some integer $1 \leq n_1 < n$. Let

$$\boldsymbol{\mu}_1 = (\mu_1, \ldots, \mu_{n_1}) \quad \text{and} \quad \boldsymbol{\mu}_2 = (\mu_{n_1+1}, \ldots, \mu_n).$$

Let $\boldsymbol{\Sigma}_{ij}$ be the cross-covariance matrix of $\boldsymbol{X}_i$ and $\boldsymbol{X}_j$, $i, j = 1, 2$, i.e.,

$$\boldsymbol{\Sigma}_{ij} = \mathbb{E}\left[(\boldsymbol{X}_i - \boldsymbol{\mu}_i)(\boldsymbol{X}_j - \boldsymbol{\mu}_j)^T\right].$$

Then we have

$$\boldsymbol{X}_1 \sim \mathcal{N}_{n_1}(\boldsymbol{\mu}_1, \boldsymbol{\Sigma}_{11}), \quad \boldsymbol{X}_2 \sim \mathcal{N}_{n-n_1}(\boldsymbol{\mu}_2, \boldsymbol{\Sigma}_{22}),$$

and

$$\boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix}.$$

Moreover, for $i, j = 1, 2$ and $i \neq j$, the conditional distribution of $\boldsymbol{X}_i$ given $\boldsymbol{X}_j = \boldsymbol{x}_j$ is $\mathcal{N}_{n_i}(\boldsymbol{\mu}_{i|j}, \boldsymbol{\Sigma}_{i|j})$, where

$$\boldsymbol{\mu}_{i|j} = \boldsymbol{\mu}_i + \boldsymbol{\Sigma}_{ij}\boldsymbol{\Sigma}_{jj}^{-1}(\boldsymbol{x}_j - \boldsymbol{\mu}_j), \tag{A.6}$$

$$\boldsymbol{\Sigma}_{i|j} = \boldsymbol{\Sigma}_{ii} - \boldsymbol{\Sigma}_{ij}\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{ji}, \tag{A.7}$$

and $\boldsymbol{\Sigma}_{jj}^{-1}$ is a generalized inverse of $\boldsymbol{\Sigma}_{jj}$, i.e., $\boldsymbol{\Sigma}_{jj}^{-1}$ satisfies

$$\boldsymbol{\Sigma}_{jj}\boldsymbol{\Sigma}_{jj}^{-1}\boldsymbol{\Sigma}_{jj} = \boldsymbol{\Sigma}_{jj}.$$

Note that if $\boldsymbol{\Sigma}_{22}$ is non-degenerate, then

$$\begin{aligned} \boldsymbol{\Sigma} &= \begin{pmatrix} \boldsymbol{\Sigma}_{11} & \boldsymbol{\Sigma}_{12} \\ \boldsymbol{\Sigma}_{21} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \\ &= \begin{pmatrix} \boldsymbol{I}_k & \boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1} \\ \boldsymbol{0}_{(n-k)\times k} & \boldsymbol{I}_{n-k} \end{pmatrix} \begin{pmatrix} \boldsymbol{\Sigma}_{1|2} & \boldsymbol{0}_{k\times(n-k)} \\ \boldsymbol{0}_{(n-k)\times k} & \boldsymbol{\Sigma}_{22} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_k & \boldsymbol{0}_{k\times(n-k)} \\ \boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{21} & \boldsymbol{I}_{n-k} \end{pmatrix}, \end{aligned} \tag{A.8}$$

where $\boldsymbol{I}_m$ is the identity matrix of size m for any positive integer m, and $\boldsymbol{0}_{i\times j}$ is the $i \times j$ matrix of zeros for any positive integers i and j. Thus, if $\boldsymbol{\Sigma}$ is non-degenerate, then $\boldsymbol{\Sigma}_{1|2}$ is also non-degenerate.

4. Let $(X_1, \ldots, X_n, Y_1, \ldots, Y_m)$ be a non-degenerate Gaussian $(n+m)$-vector. Let $\boldsymbol{X} := (X_1, \ldots, X_n)^T$ and $\boldsymbol{Y} := (Y_1, \ldots, Y_m)^T$. Then $\boldsymbol{X}$ is a non-degenerate Gaussian n-vector, and $\boldsymbol{Y}$ is a non-degenerate Gaussian m-vector. Let $\phi_{\boldsymbol{X}}$, $\phi_{\boldsymbol{Y}}$ and $\phi_{\boldsymbol{X},\boldsymbol{Y}}$ be the probability density functions of $\boldsymbol{X}$, $\boldsymbol{Y}$ and $(\boldsymbol{X}, \boldsymbol{Y})$, respectively. Assume $\mathrm{Cov}[X_i, Y_j] = 0$ for any $1 \leq i \leq n$ and $1 \leq j \leq m$. Then

$$\phi_{\boldsymbol{X},\boldsymbol{Y}}(\boldsymbol{x}, \boldsymbol{y}) = \phi_{\boldsymbol{X}}(\boldsymbol{x})\phi_{\boldsymbol{X}}(\boldsymbol{y}) \tag{A.9}$$

for any $\boldsymbol{x} \in \mathbb{R}^n$ and $\boldsymbol{y} \in \mathbb{R}^m$.

Here Fact 1 also implies a similar transformation as in univariate cases. Let $\boldsymbol{X}$ be a non-degenerate column Gaussian n-vector with mean vector $\boldsymbol{\mu}$ and positive-definite covariance matrix $\boldsymbol{\Sigma}$. Since $\boldsymbol{\Sigma}$ is a real positive-definite symmetric matrix, we can always find a matrix $\boldsymbol{Q} \in \mathbb{R}^{n\times n}$ such that

$$\boldsymbol{Q}\boldsymbol{\Sigma}\boldsymbol{Q}^T = \boldsymbol{I}_n.$$

Then

$$\boldsymbol{Q}(\boldsymbol{X} - \boldsymbol{\mu}) \sim \mathcal{N}_n(\boldsymbol{0}, \boldsymbol{I}_n).$$

A.2 Examples for Chapter 2

A.2.1 An Example for Lemma 2.1.1

Example A.2.1. *Suppose* $r(\boldsymbol{t}) = e^{-\|\boldsymbol{t}\|^2}$*,* $\boldsymbol{t} \in \mathbb{R}^N$*. It is easy to see that* $1 - e^{-x} \leq x$ *for any* $x \geq 0$*, and* $x \leq (-\log(x))^{-2}$ *for any* $0 < x < 1$*. If* $0 < \|\boldsymbol{t} - \boldsymbol{s}\| < 1$*, then*

$$\begin{aligned}
\mathbb{E}\left[(X(\boldsymbol{t}) - X(\boldsymbol{s}))^2\right] &= 2\left(1 - e^{-\|\boldsymbol{t}-\boldsymbol{s}\|^2}\right) \\
&\leq 2\|\boldsymbol{t} - \boldsymbol{s}\|^2 \\
&\leq 2\left(-\log\left(\|\boldsymbol{t} - \boldsymbol{s}\|^2\right)\right)^{-2} \\
&= \frac{1}{2}\left|\log\left(\|\boldsymbol{t} - \boldsymbol{s}\|\right)\right|^{-2}
\end{aligned}$$

Therefore, (2.2) holds for $\alpha = 1$ *and* $\gamma = \frac{1}{2}$ *and any* $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^N$ *with* $0 < \|\boldsymbol{t} - \boldsymbol{s}\| < 1$*. By Lemma 2.1.1, we conclude that a centered, stationary Gaussian random field with covariance function* $r(\boldsymbol{t}) = e^{-\|\boldsymbol{t}\|^2}$*,* $\boldsymbol{t} \in \mathbb{R}^N$*, is almost surely continuous on* $\mathbb{R}^N$*.*

A.2.2 An Example for Theorem 2.1.6

Example A.2.2. *Let* X *be a centered Gaussian random field with covariance function* $r(\boldsymbol{t}) = e^{-\|\boldsymbol{t}\|^2}$*,* $\boldsymbol{t} \in \mathbb{R}^N$*. For any positive integer* k*, we now show that* X *satisfies all the conditions in Theorem 2.1.6. Since* $r(\boldsymbol{t})$ *has any order partial derivatives, by Lemma 2.1.4,* X *is* k *times differentiable in the mean square sense on* $\mathbb{R}^N$*. Let* $\rho > 0$*,* $\delta = 3$*,* $h_0 = \frac{1}{2}$ *and* $K = k2^{2k+3}(1+\rho)^2$*. The only thing left is to check (2.9) for any* $0 < \eta_1, \eta_2, h < h_0$ *and*

$$\left\{((\boldsymbol{t}, \boldsymbol{t}'), (\boldsymbol{s}, \boldsymbol{s}')) \in \mathbb{R}^N_{k,\rho} \times \mathbb{R}^N_{k,\rho} : (\boldsymbol{s}, \boldsymbol{s}') \in B_{N,k}((\boldsymbol{t}, \boldsymbol{t}'), h)\right\},$$

such that $\|(\boldsymbol{t}, \boldsymbol{t}') - (\boldsymbol{s}, \boldsymbol{s}')\|^2_{N,k} \neq 0$*. Since* $h_0 = \frac{1}{2}$*, it is easy to see*

$$0 < \|(\boldsymbol{t}, \boldsymbol{t}') - (\boldsymbol{s}, \boldsymbol{s}')\|^2_{N,k} + |\eta_1 - \eta_2|^2 < h^2 + \frac{1}{4} < 1.$$

Since $x \leq (-\log(x))^{-2}$ *for any* $0 < x < 1$*,*

$$\begin{aligned}
&\left(\|(\boldsymbol{t}, \boldsymbol{t}') - (\boldsymbol{s}, \boldsymbol{s}')\|^2_{N,k} + |\eta_1 - \eta_2|^2\right) \\
&\leq \left(-\log\left(\sqrt{\|(\boldsymbol{t}, \boldsymbol{t}') - (\boldsymbol{s}, \boldsymbol{s}')\|^2_{N,k} + |\eta_1 - \eta_2|^2}\right)\right)^{-4} \\
&\leq \left(-\log\left(\|(\boldsymbol{t}, \boldsymbol{t}') - (\boldsymbol{s}, \boldsymbol{s}')\|_{N,k} + |\eta_1 - \eta_2|\right)\right)^{-4}.
\end{aligned} \tag{A.10}$$

Then by (2.5), $0 < \eta_1 < 1$, $\|\boldsymbol{s}'\|_{kN} < 1+\rho$, $k \ge 1$, and the fact that $1 - e^{-y} \le y$ for all $y \ge 0$, we have

$$
\begin{aligned}
&\mathbb{E}\left[(F_{X,k}(\boldsymbol{t},\eta_1\boldsymbol{t}') - F_{X,k}(\boldsymbol{s},\eta_2\boldsymbol{s}'))^2\right] \\
&\le 2^k \sum_{a_i\in\{0,1\},1\le i\le k} \mathbb{E}\left[\left(X\left(\boldsymbol{t}+\eta_1\sum_{i=1}^k a_i\boldsymbol{t}_i'\right) - X\left(\boldsymbol{s}+\eta_2\sum_{i=1}^k a_i\boldsymbol{s}_i'\right)\right)^2\right] \\
&= 2^{k+1} \sum_{a_i\in\{0,1\},1\le i\le k} \left(1-\exp\left\{-\left\|\boldsymbol{t}+\eta_1\sum_{i=1}^k a_i\boldsymbol{t}_i' - \boldsymbol{s} - \eta_2\sum_{i=1}^k a_i\boldsymbol{s}_i'\right\|^2\right\}\right) \\
&\le 2^{k+1} \sum_{a_i\in\{0,1\},1\le i\le k} \left(1-\exp\left\{-2\left(\|\boldsymbol{t}-\boldsymbol{s}\|^2 + k\sum_{i=1}^k a_i^2\left\|\eta_1\boldsymbol{t}_i' - \eta_2\boldsymbol{s}_i'\right\|^2\right)\right\}\right) \\
&\le 2^{2k+1}\left(1-\exp\left\{-2\left(\|\boldsymbol{t}-\boldsymbol{s}\|^2 + k\left\|\eta_1\boldsymbol{t}' - \eta_2\boldsymbol{s}'\right\|_{kN}^2\right)\right\}\right) \\
&\le 2^{2k+1}\left(1-\exp\left\{-4\left(\|\boldsymbol{t}-\boldsymbol{s}\|^2 + k\left\|\eta_1\boldsymbol{t}' - \eta_1\boldsymbol{s}'\right\|_{kN}^2 + k\left\|\eta_1\boldsymbol{s}' - \eta_2\boldsymbol{s}'\right\|_{kN}^2\right)\right\}\right) \\
&\le 2^{2k+1}\left(1-\exp\left\{-4k(1+\rho)^2\left(\|\boldsymbol{t}-\boldsymbol{s}\|^2 + \left\|\boldsymbol{t}' - \boldsymbol{s}'\right\|_{kN}^2 + |\eta_1-\eta_2|^2\right)\right\}\right) \\
&\le k2^{2k+3}(1+\rho)^2\left(\|(\boldsymbol{t},\boldsymbol{t}')-(\boldsymbol{s},\boldsymbol{s}')\|_{N,k}^2 + |\eta_1-\eta_2|^2\right) \\
&\le k2^{2k+3}(1+\rho)^2\left(-\log\left(\|(\boldsymbol{t},\boldsymbol{t}')-(\boldsymbol{s},\boldsymbol{s}')\|_{N,k} + |\eta_1-\eta_2|\right)\right)^{-4} \\
&= K\left(-\log\left(\|(\boldsymbol{t},\boldsymbol{t}')-(\boldsymbol{s},\boldsymbol{s}')\|_{N,k} + |\eta_1-\eta_2|\right)\right)^{-(1+\delta)}.
\end{aligned}
$$

A.2.3 Proof of Lemma 2.1.13

Proof. Since X is stationary, it suffices to show that there exist finite constants $K > 0$ and $\alpha > 0$ such that

$$\max_{1\le i_1,i_2,i_3,i_4\le N} |r_{i_1i_2i_3i_4}(\mathbf{0}) - r_{i_1i_2i_3i_4}(\boldsymbol{t})| \le K|\log(\|\boldsymbol{t}\|)|^{-(1+\alpha)}$$

for all $\|\boldsymbol{t}\| > 0$ small enough. Since $x \le (-\log(x))^{-2}$ for any $0 < x < 1$, we have $\|\boldsymbol{t}\| < (-\log(\|\boldsymbol{t}\|))^{-2}$ when $0 < \|\boldsymbol{t}\| < 1$. Thus, it suffices to show that there exists a constant $C > 0$ such that

$$\max_{1\le i_1,i_2,i_3,i_4\le N} |r_{i_1i_2i_3i_4}(\mathbf{0}) - r_{i_1i_2i_3i_4}(\boldsymbol{t})| \le C\|\boldsymbol{t}\| \tag{A.11}$$

for all $\|\boldsymbol{t}\| > 0$ small enough. Since all of the sixth-order partial derivatives of $r(\boldsymbol{t})$ exist at $\boldsymbol{t} = \mathbf{0}$, there exists a constant $\varepsilon > 0$ small enough such that for any $1 \le i_1, i_2, i_3, i_4, i_5 \le N$,

$r_{i_1i_2i_3i_4}(\boldsymbol{t})$ is continuous and $r_{i_1i_2i_3i_4i_5}(\boldsymbol{t})$ is bounded on $\boldsymbol{t} \in B(\boldsymbol{0}, \varepsilon)$, where $B(\boldsymbol{0}, \varepsilon)$ is the N-dimensional open ball centered at the origin with radius ε. Then by the mean value theorem, (A.11) is immediate, and hence proved.

□

A.2.4 An Example for Lemma 2.1.10

Example A.2.3. *Let X be a centered, stationary Gaussian random field with covariance function $r(\boldsymbol{t}) = e^{-\|\boldsymbol{t}\|^2}$, $\boldsymbol{t} \in \mathbb{R}^N$. By Lemma 2.1.4, X has any order mean square partial derivatives. From Example A.2.2, we see that X is almost surely k times differentiable for any positive integer k. Let $\boldsymbol{f} = \nabla X$ and $\boldsymbol{g} = (\nabla^2 X, X)$. Let T be a compact set as in Lemma 2.1.10. In this example, we will show that $\boldsymbol{f}$ and $\boldsymbol{g}$ on T satisfy all the conditions in Lemma 2.1.10.*

We first note that all of the sixth-order partial derivatives of $r(\boldsymbol{t})$ exist at $\boldsymbol{t} = \boldsymbol{0}$. Then by Lemma 2.1.13, Condition (2.13) holds for all $\boldsymbol{s}, \boldsymbol{t} \in T$ such that $\|\boldsymbol{t} - \boldsymbol{s}\|$ is small enough. The only thing left is to show that the joint distribution of $(\boldsymbol{f}, \boldsymbol{h}) = (\nabla X, \nabla^2 X, X)$ is non-degenerate. For any $p \geq 1$ and $1 \leq i_1, \ldots, i_p \leq N$, define $r_{i_1,\ldots,i_p}(\boldsymbol{t}) := \frac{\partial^p r(\boldsymbol{t})}{\partial t_{i_1} \cdots \partial t_{i_p}}$, $\boldsymbol{t} \in \mathbb{R}^N$. By Lemma 2.1.3 and (2.8), the joint distribution of $(\nabla X, \nabla^2 X, X)$ is Gaussian, and for any $1 \leq i, k, \ell \leq N$ and $\boldsymbol{t} \in \mathbb{R}^N$,

$$\mathrm{Cov}\left[X(\boldsymbol{t}), X_i(\boldsymbol{t})\right] = r_i(\boldsymbol{0}) = 0 \quad \textit{and} \quad \mathrm{Cov}\left[X_i(\boldsymbol{t}), X_{k\ell}(\boldsymbol{t})\right] = r_{ik\ell}(\boldsymbol{0}) = 0.$$

Therefore, we only need to show that the joint distribution of $(\nabla^2 X(\boldsymbol{t}), X(\boldsymbol{t}))$ and the distribution of $\nabla X(\boldsymbol{t})$ are non-degenerate for any $\boldsymbol{t} \in \mathbb{R}^N$. Suppose

$$aX(\boldsymbol{t}) = \sum_{1 \leq i,j \leq N} a_{ij} X_{ij}(\boldsymbol{t})$$

for some $\boldsymbol{t} \in \mathbb{R}^N$, $a, a_{ij} \in \mathbb{R}$, $1 \leq i, j \leq N$, where for any $1 \leq i, j \leq N$, $a_{ij} = a_{ji}$ and $|a| + \sum_{1 \leq i,j \leq N} |a_{ij}| \neq 0$. Then for any $1 \leq k, \ell \leq N$,

$$\mathrm{Cov}\left[aX(\boldsymbol{t}), X_{k\ell}(\boldsymbol{t})\right] = r_{k\ell}(\boldsymbol{0}) = -2a\delta_{k\ell}$$

and simultaneously,

$$\begin{aligned}\operatorname{Cov}\left[aX(\boldsymbol{t}), X_{k\ell}(\boldsymbol{t})\right] &= \sum_{1\le i,j\le N} a_{ij} r_{ijk\ell}(\mathbf{0}) \\ &= 4 \sum_{1\le i,j\le N} a_{ij}\left(\delta_{ij}\delta_{kl} + \delta_{jk}\delta_{il} + \delta_{ik}\delta_{jl}\right). \\ &= 4 \sum_{1\le i\le N} a_{ii}\delta_{kl} + 8a_{k\ell}.\end{aligned}$$

Thus for any $1 \le k, \ell \le N$,

$$-2a\delta_{k\ell} = 4 \sum_{1\le i\le N} a_{ii}\delta_{kl} + 8a_{k\ell},$$

and then,

$$a_{k\ell} = -\left(\frac{a}{4} + \frac{1}{2}\sum_{1\le i\le N} a_{ii}\right)\delta_{k\ell}. \tag{A.12}$$

From (A.12), we have $a_{k\ell} = 0$ *when* $k \neq \ell$. *Taking summation over all* k, ℓ *such that* $1 \le k = \ell \le N$ *on the both sides of (A.12), we have for any* $1 \le k \le N$,

$$a_{kk} = \frac{1}{N}\sum_{1\le i\le N} a_{ii} = -\frac{a}{2(N+2)}.$$

If $a \neq 0$, *then*

$$X(\boldsymbol{t}) = -\frac{1}{2(N+2)}\sum_{1\le i\le N} X_{ii}(\boldsymbol{t})$$

for any $\boldsymbol{t} \in \mathbb{R}^N$*. Therefore,*

$$\begin{aligned}
1 &= r(\mathbf{0}) \\
&= \left(\frac{1}{2(N+2)}\right)^2 \sum_{1\le i\le N} \sum_{1\le k\le N} \mathrm{Cov}\left[X_{ii}(\boldsymbol{t}), X_{kk}(\boldsymbol{t})\right] \\
&= \left(\frac{1}{2(N+2)}\right)^2 \sum_{1\le i\le N} \sum_{1\le k\le N} r_{iikk}(\mathbf{0}) \\
&= 4\left(\frac{1}{2(N+2)}\right)^2 \sum_{1\le i\le N} \sum_{1\le k\le N} (1+2\delta_{ik}) \\
&= 4\left(\frac{1}{2(N+2)}\right)^2 (N^2+2N) \\
&= \frac{N}{N+2} \\
&< 1,
\end{aligned}$$

which leads to a contradiction. If $a = 0$*, then by Lemma 2.1.4, we have*

$$\begin{aligned}
0 &= \sum_{1\le i\le N} \mathrm{Cov}\left[X_{ii}(\boldsymbol{t}) X(\boldsymbol{t})\right] \\
&= \sum_{1\le i\le N} r_{ii}(\mathbf{0}) \\
&= 2N,
\end{aligned}$$

which also leads to a contradiction. Therefore, the joint distribution of $(\nabla^2 X(\boldsymbol{t}), X(\boldsymbol{t}))$ *is non-degenerate for any* $\boldsymbol{t} \in \mathbb{R}^N$*. If* $\sum_{1\le i\le N} a_i X_i(\boldsymbol{t}) = 0$ *for some* $\boldsymbol{t} \in \mathbb{R}^N$*, then for any* $1 \le j \le N$*, we have*

$$\begin{aligned}
0 &= \mathrm{Cov}\left[\sum_{1\le i\le N} a_i X_i(\boldsymbol{t}), X_j(\boldsymbol{t})\right] \\
&= -\sum_{1\le i\le N} a_i r_{ij}(\boldsymbol{t}) \\
&= -2a_j,
\end{aligned}$$

which implies that the joint distribution of $\nabla X(\boldsymbol{t})$ *is non-degenerate for any* $\boldsymbol{t} \in \mathbb{R}^N$*.*

Appendix B

Appendix for Chapter 3

B.1 Proofs for Sections 3.3 and 3.4

B.1.1 Proof of Lemma 3.3.5

Lemma B.1.1. *Every qualified pair (X, g_u) possesses an adapted grid-block system.*

Proof. Let $b_u = g_u^{1-\kappa/2}$ and $1 - f_u = g_u^{-\kappa/4}$, where κ are defined in (3.12). Then Conditions 3(b)-3(d) in Definition 3.3.3 automatically hold. As for Condition 3(e), note that by Lemma 3.2.2, we have

$$\lim_{u\to\infty} g_u^N u^{N-1} e^{-u^2/2} = L_0 \tag{B.1}$$

for some finite constant $L_0 > 0$. Thus

$$g_u^N \leq L_1 u^{-N+1} e^{u^2/2}$$

for some finite constant $L_1 > 0$. Then

$$\begin{aligned} g_u b_u^{-1} &= g_u^{\kappa/2} \\ &= O\left(\left(u^{-N+1} e^{u^2/2}\right)^{\kappa/(2N)}\right) \\ &= O\left(u^{(-N+1)\kappa/(2N)} e^{\kappa u^2/(4N)}\right) \\ &= O\left(e^{\kappa u^2/2}\right). \end{aligned}$$

Let $d_u = u^{-2}\gamma(\beta, x_u)^{1/(4N)}$, where $\gamma(\beta, u)$, $u > 0$ is defined in Condition (3.10), and $x_u > 0$ satisfies $e^{\beta x_u^2} = 2g_u^{\kappa/4}$. It is easy to see that $x_u \to \infty$ as $u \to \infty$. Then by Condition (3.10), we have

$$u^{-2}d_u^{-1} = \gamma(\beta, x_u)^{-1/(4N)} \to \infty$$

as $u \to \infty$, which is Condition 3(a). For Condition 3(f), note that from (B.1), we also have

$$e^{\varepsilon u^2} \leq L_2 g_u$$

for some $0 < \varepsilon < 1/(2N)$ and some finite constant $L_2 > 0$. Then

$$L_2^{-1}e^{\varepsilon u^2} \leq g_u = \left(\frac{1}{2}e^{\beta x_u^2}\right)^{4/\kappa},$$

which implies

$$u^2 = O\left(x_u^2\right)$$

as $u \to \infty$. Again by Condition (3.10), we have

$$\begin{aligned}
u^2\overline{r}(2g_u b_u^{-1}(1-f_u)) &= u^2\overline{r}(2g_u^{\kappa/4}) \\
&= O\left(x_u^2\overline{r}\left(e^{\beta x_u^2}\right)\right) \\
&= O\left(x_u^{2N+2}\overline{r}\left(e^{\beta x_u^2}\right)\right) \\
&= o(1)
\end{aligned}$$

as $u \to \infty$, and then

$$\begin{aligned}
&g_u^{2N}d_u^{-2N}\overline{r}(2g_u b_u^{-1}(1-f_u))\exp\left\{-\frac{u^2}{1+\overline{r}\left(2g_u b_u^{-1}(1-f_u)\right)}\right\} \\
&= O\left(u^{-2N+2}u^{4N}\gamma(\beta, x_u)^{-1/2}\overline{r}\left(2g_u b_u^{-1}(1-f_u)\right)\exp\left\{\frac{u^2\overline{r}\left(2g_u b_u^{-1}(1-f_u)\right)}{1+\overline{r}\left(2g_u b_u^{-1}(1-f_u)\right)}\right\}\right) \\
&= O\left(u^{2(N+1)}\overline{r}(2g_u b_u^{-1}(1-f_u))\gamma(\beta, x_u)^{-1/2}\right) \\
&= O\left(x_u^{2(N+1)}\overline{r}(e^{\beta x_u^2})\gamma(\beta, x_u)^{-1/2}\right) \\
&= O\left(\gamma(\beta, x_u)\gamma(\beta, x_u)^{-1/2}\right) \\
&= o(1)
\end{aligned}$$

as $u \to \infty$. Therefore, Condition 3(f) holds. Finally, to meet Condition 4 in Definition 3.3.3, we only need to change d_u to d'_u by

$$d'_u = g_u b_u^{-1}/\lfloor g_u b_u^{-1} d_u^{-1}\rfloor,$$

where for any $x \in \mathbb{R}$, $\lfloor x \rfloor$ denotes the largest integer less than or equal to x. Let $N_u = \lfloor g_u b_u^{-1} d_u^{-1} \rfloor$. Then

$$d'_u = \left(1 + \frac{g_u b_u^{-1} d_u^{-1} - N_u}{N_u}\right) d_u,$$

where $0 \le g_u b_u^{-1} d_u^{-1} - N_u < 1$. This implies

$$0 < d'_u - d_u \le \frac{d_u}{N_u} \le \frac{d_u}{g_u b_u^{-1} d_u^{-1} - 1} \to 0$$

as $u \to \infty$. Then we have

$$\begin{aligned}
&g_u^{2N} (d'_u)^{-2N} \overline{r}(2 g_u b_u^{-1}(1 - f_u)) \exp\left\{-\frac{u^2}{1 + \overline{r}\,(2 g_u b_u^{-1}(1 - f_u))}\right\} \\
&\le g_u^{2N} d_u^{-2N} \overline{r}(2 g_u b_u^{-1}(1 - f_u)) \exp\left\{-\frac{u^2}{1 + \overline{r}\,(2 g_u b_u^{-1}(1 - f_u))}\right\} \\
&= o(1)
\end{aligned}$$

as $u \to \infty$, and

$$\begin{aligned}
u^{-2}(d'_u)^{-1} &= u^{-2} d_u^{-1} \frac{d_u}{d'_u} \\
&\ge u^{-2} d_u^{-1} \frac{d_u}{d_u + d_u/(g_u b_u^{-1} d_u^{-1} - 1)} \\
&\to \infty
\end{aligned}$$

as $u \to \infty$. □

B.1.2 Proof of Lemma 3.3.7

Lemma B.1.2. *For any bounded set $S \subset \mathbb{R}^N$ $(N \ge 1)$ with $\lambda_{N-1}(\partial S) < \infty$, we can always find $J_{1,u}(S), J_{2,u}(S) \subset \mathcal{J}$ for any $u \in \mathbb{R}$ large enough, such that*

1. $J_{1,u}(S) \subset \overset{\circ}{S} \subset S \subset J_{2,u}(S)$;
2. $\lambda_N(J_{2,u}(S) - J_{1,u}(S)) \to 0$ *as* $u \to \infty$;
3. $\{g_u \overline{J_{2,u}(S) - J_{1,u}(S)}, u \in \mathbb{R}\}$ *is a blowing-up system.*

Proof. For any positive integer i and $\boldsymbol{k} \in \mathbb{Z}^N$, let

$$\Delta'_{i,\boldsymbol{k}} := 2^{-i}\left([0,1]^N \oplus \{\boldsymbol{k}\}\right),$$

where the prime on the superscript is used to distinguish this notation from (3.13). Then each $\Delta'_{i,\boldsymbol{k}}$ is an N-dimensional cube with the side length 2^{-i}. For any positive integers n and i, let

$$C_{n,i} := \left\{\boldsymbol{t} = (t_1, \ldots, t_N) \in \mathbb{R}^N : t_j \in \left[0, 2^{-i}\right] \text{ for } 1 \le j \le n, \text{ and } t_j = 0 \text{ for } n < j \le N\right\}.$$

Then each $C_{n,i}$ is an n-dimensional cube with side length 2^{-i}. For the consistency of the notations in following, we let $C_{0,i} = \{\mathbf{0}\}$, where $\mathbf{0}$ is the origin in $\mathbb{R}^N$. Then we have the following facts:

1. for $n \ge 0$, $\lambda_n(C_{n,i}) = 2^{-in}$;
2. for $n \ge 1$, $\lambda_{n-1}(\partial C_{n,i}) = b_{n,n-1} 2^{-i(n-1)} = n 2^{-in+i+1}$;
3. for $n \ge 1$, we have

$$\lambda_n((aC_{n-1,i}) \oplus B_n(\mathbf{0}, R)) = \lambda_n(B_n(\mathbf{0}, R)) + 2\sum_{k=0}^{n-1} b_{n-1,k} R^{n-k} \lambda_k(C_{k,i}) a^k$$

   for any $a > 0$,

where for any integer $0 \le k \le n$, $b_{n,k} := 2^{n-k}\binom{n}{k}$ is the number of k-dimensional faces of $C_{n,i}$ if $n \ge 1$, and $b_{n,k} := 0$ if $n = 0$. Immediately, we have

$$\Delta'_{i,\mathbf{0}} = C_{N,i}$$

for any positive integer i.

For any positive integer i, we can define

$$K_{1,i}(S) := \bigcup \left\{\Delta' \in \left\{\Delta'_{i,\boldsymbol{k}}, \boldsymbol{k} \in \mathbb{Z}^N\right\} : \Delta' \subset \mathring{S}\right\}.$$

By the definitions above, it is easy to see that

$$\bigcup_{i=1}^{\infty} K_{1,i}(S) \subset \mathring{S}. \tag{B.2}$$

For any $\boldsymbol{t} \in \overset{\circ}{S}$, there exists some neighborhood $N_{\boldsymbol{t}}$ of $\boldsymbol{t}$ such that $N_{\boldsymbol{t}} \subset S$, and then we can find some $\Delta' \in \{\Delta'_{i,\boldsymbol{k}} : i \in \{1, 2, \dots\}, \boldsymbol{k} \in \mathbb{Z}^N\}$ such that $\Delta' \subset N_{\boldsymbol{t}}$. Therefore, $\boldsymbol{t} \in \Delta' \subset \cup_{i=1}^{\infty} K_{1,i}(S)$, which together with (B.2) implies

$$\bigcup_{i=1}^{\infty} K_{1,i}(S) = \overset{\circ}{S}.$$

Note that $\lambda_{N-1}(\partial S) < \infty$ implies that

$$\lambda_N(\partial S) = 0.$$

Thus,

$$\lim_{i \to \infty} \lambda_N(K_{1,i}(S)) = \lambda_N \left(\bigcup_{i=1}^{\infty} K_{1,i}(S) \right) = \lambda_N(S). \tag{B.3}$$

Moreover, let $K_{1,0}(S) = \emptyset$, and for any positive integer i, define

$$L_i(S) := K_{1,i}(S) - K_{1,i-1}(S)$$

and

$$a_i(S) := \# \left\{ \Delta' \in \left\{ \Delta'_{i,\boldsymbol{k}}, \boldsymbol{k} \in \mathbb{Z}^N \right\} : \Delta' \subset \overline{L}_i(S) \right\}.$$

where "#″ stands for "the cardinality of". Then by Fact 1, for any positive integer m, we have

$$K_{1,m}(S) = \bigcup_{i=1}^{m} L_i(S) \text{ and } \lambda_N(K_{1,m}(S)) = \sum_{i=1}^{m} \lambda_N(L_i(S)) = \sum_{i=1}^{m} a_i(S) 2^{-iN} < \lambda_N(S) < \infty.$$

By the subadditivity of λ_{N-1}, for any positive integer m, we have

$$\begin{aligned} \lambda_{N-1}\left(\partial K_{1,m}(S)\right) &\le \lambda_{N-1}\left(\partial K_{1,m-1}(S) \cup \partial L_m(S)\right). \\ &\le \lambda_{N-1}\left(\partial K_{1,m-1}(S)\right) + \lambda_{N-1}\left(\partial L_m(S)\right). \end{aligned}$$

Note that by Fact 2, for any positive integer i, we have

$$\lambda_{N-1}\left(\partial L_i(S)\right) \le a_i(S) N 2^{-iN+i+1},$$

and then

$$\begin{aligned} \lambda_{N-1}\left(\partial K_{1,m}(S)\right) &\le \sum_{i=1}^{m} \lambda_{N-1}\left(\partial L_i(S)\right) \\ &\le 2N \sum_{i=1}^{m} a_i(S) 2^{-i(N-1)}. \end{aligned}$$

Since S is bounded, we can find a positive integer $b > 0$ such that $S \subset [-2^b, 2^b]^N$. Then $S_b := [-2^{b+1}, 2^{b+1}]^N - S$ is also a bounded set with

$$\lambda_{N-1}(\partial S_b) \leq \lambda_{N-1}\left(\partial[-2^{b+1}, 2^{b+1}]^N\right) + \lambda_{N-1}(\partial S) < \infty.$$

The choice of S_b ensures that

$$\partial[-2^b, 2^b]^N \subset K_{1,1}(S_b).$$

For any positive integer m, define

$$K_{2,m}(S) := [-2^b, 2^b]^N - K_{1,m}(S_b).$$

Then it is easy to see that

$$K_{1,m}(S), K_{2,m}(S) \in \mathcal{J} \quad \text{and} \quad K_{1,m}(S) \subset \overset{\circ}{S} \subset S \subset K_{2,m}(S).$$

By (B.3), we also have

$$\begin{aligned}
\lambda_N(K_{2,m}(S)) - \lambda_N(S) &= \lambda_N(K_{2,m}(S) - S) \\
&= \lambda_N\left(\left([-2^b, 2^b]^N - S\right) - K_{1,m}(S_b)\right) \\
&\leq \lambda_N\left(\left([-2^{b+1}, 2^{b+1}]^N - S\right) - K_{1,m}(S_b)\right) \\
&= \lambda_N(S_b - K_{1,m}(S_b)) \\
&\to 0 \quad \text{as} \quad m \to \infty,
\end{aligned}$$

which together with (B.3) implies that

$$\lim_{m\to\infty} \lambda_N(K_{2,m}(S)) - \lambda_N(K_{1,m}(S)) = 0. \tag{B.4}$$

Given any $0 < r < N^{-1}$, we can find a positive-integer valued function $n_r(u)$, $u \in \mathbb{R}$ increasing slowly enough as $u \to \infty$, such that

$$g_u^{-r} 2^{n_r(u)} \leq \left(\lambda_N\left(K_{2,n_r(u)}(S)\right) - \lambda_N\left(K_{1,n_r(u)}(S)\right)\right)^{\alpha_0} \tag{B.5}$$

for sufficiently large u, where $\alpha_0 = (N-1+r)/N$. The definition of n_r also implies that

$$\begin{aligned}
g_u^N \lambda_N(K_{2,n_r(u)}(S) - K_{1,n_r(u)}(S)) &\geq g_u^N \left(g_u^{-r} 2^{n_r(u)}\right)^{1/\alpha_0} \\
&= g_u^{\frac{N(N-1)}{N-1+r}} 2^{\frac{N n_r(u)}{N-1+r}} \\
&\to \infty
\end{aligned}$$

as $u \to \infty$, and

$$g_u 2^{-n_r(u)} = g_u^{1-r}\left(g_u^{-r}2^{n_r(u)}\right)^{-1} \to \infty \tag{B.6}$$

as $u \to \infty$. Now let $J_{1,u}(S) = K_{1,n_r(u)}(S)$ and $J_{2,u}(S) = K_{2,n_r(u)}(S)$. So far we have shown that

1. $J_{1,u}(S) \subset \overset{\circ}{S} \subset S \subset J_{2,u}(S)$;
2. $\lambda_N(J_{2,u}(S) - J_{1,u}(S)) \to 0$ as $n \to \infty$;
3. $\lambda_N(g_u\overline{J_{2,u}(S) - J_{1,u}(S)}) \to \infty$ as $n \to \infty$.

The only thing left is to show that the system $\{g_u\overline{J_{2,u}(S) - J_{1,u}(S)}, u \in \mathbb{R}\}$ satisfies Conditions (3.3) and (3.5). For Condition (3.5), by (B.5) and Facts 1 and 2, we have

$$\begin{aligned}
&\lambda_{N-1}\left(g_u\partial J_{1,u}(S)\right)\\
&= \lambda_{N-1}\left(g_u\partial K_{1,n_r(u)}(S)\right)\\
&\leq 2Ng_u^{N-1}\sum_{i=1}^{n_r(u)} a_i(S)2^{-i(N-1)}\\
&= 2Ng_u^{N-1+r}\sum_{i=1}^{n_r(u)} (g_u^{-r}2^i)a_i(S)2^{-iN}\\
&\leq 2Ng_u^{N-1+r}\left(\lambda_N\left(K_{2,n_r(u)}(S)\right) - \lambda_N\left(K_{1,n_r(u)}(S)\right)\right)^{\alpha_0}\sum_{i=1}^{n_r(u)} a_i(S)2^{-iN}\\
&\leq 2N\lambda_N(S)\left(g_u^N\left(\lambda_N\left(K_{2,n_r(u)}(S)\right) - \lambda_N\left(K_{1,n_r(u)}(S)\right)\right)\right)^{\alpha_0}\\
&= 2N\lambda_N(S)\lambda_N\left(g_u\overline{J_{2,u}(S) - J_{1,u}(S)}\right)^{\alpha_0}.
\end{aligned} \tag{B.7}$$

Note that

$$K_{2,n_r(u)}(S) = [-2^b, 2^b]^N - K_{1,n_r(u)}(S_b) \quad \text{and} \quad \partial[-2^b, 2^b]^N \subset K_{1,n_r(u)}(S_b).$$

Then by the subadditivity of λ_{N-1}, we have

$$\lambda_{N-1}\left(\partial K_{2,n_r(u)}(S)\right) \leq \lambda_{N-1}\left(\partial K_{1,n_r(u)}(S_b)\right).$$

Similarly, we have

$$
\begin{aligned}
&\lambda_{N-1}\left(g_u \partial J_{2,u}(S)\right) \\
&= \lambda_{N-1}\left(g_u \partial K_{2,n_r(u)}(S)\right) \\
&\leq \lambda_{N-1}\left(g_u \partial K_{1,n_r(u)}(S_b)\right) \\
&\leq 2N g_u^{N-1+r}\left(\lambda_N\left(K_{2,n_r(u)}(S)\right) - \lambda_N\left(K_{1,n_r(u)}(S)\right)\right)^{\alpha_0} \sum_{i=1}^{n_r(u)} a_i(S_b) 2^{-iN} \\
&\leq 2N \lambda_N(S_b) \lambda_N\left(g_u \overline{J_{2,u}(S) - J_{1,u}(S)}\right)^{\alpha_0}.
\end{aligned} \tag{B.8}
$$

Again, by the subadditivity of λ_{N-1} and combining (B.7) with (B.8), we have

$$
\begin{aligned}
\lambda_{N-1}\left(g_u \partial \overline{J_{2,u}(S) - J_{1,u}(S)}\right) &\leq \lambda_{N-1}\left(g_u \partial J_{2,u}(S) \cup g_u \partial J_{1,u}(S)\right) \\
&\leq \lambda_{N-1}\left(g_u \partial J_{2,u}(S)\right) + \lambda_{N-1}\left(g_u \partial J_{1,u}(S)\right) \\
&= O\left(\lambda_N\left(g_u \overline{J_{2,u}(S) - J_{1,u}(S)}\right)^{\alpha_0}\right).
\end{aligned} \tag{B.9}
$$

As for Condition (3.3), from the construction of $J_{1,u}(S)$ and $J_{2,u}(S)$, we see that if

$$
\boldsymbol{t} \in \left(g_u \overline{J_{2,u}(S) - J_{1,u}(S)}\right) \oplus B_N(\mathbf{0}, R) - g_u \overline{J_{2,u}(S) - J_{1,u}(S)}
$$

for some $R > 0$, then

$$
B_N(\boldsymbol{t}, R) \cap \left(g_u \partial \overline{J_{2,u}(S) - J_{1,u}(S)}\right) \neq \emptyset.
$$

From the construction of $J_{1,u}(S)$ and $J_{2,u}(S)$, we see that there exists a finite set $G_u(S) \subset \mathbb{Z}^N$ such that

$$
\overline{J_{2,u}(S) - J_{1,u}(S)} = \bigcup_{\boldsymbol{k} \in G_u(S)} \Delta'_{n_r(u),\boldsymbol{k}}.
$$

Therefore,

$$
\begin{aligned}
&\lambda_N\left(\left(g_u \overline{J_{2,u}(S) - J_{1,u}(S)}\right) \oplus B_N(\mathbf{0}, R)\right) - \lambda_N\left(g_u \overline{J_{2,u}(S) - J_{1,u}(S)}\right) \\
&\leq \lambda_N\left(\left(g_u \partial \overline{J_{2,u}(S) - J_{1,u}(S)}\right) \oplus B_N(\mathbf{0}, R)\right) \\
&\leq \frac{\lambda_{N-1}\left(g_u \partial \overline{J_{2,u}(S) - J_{1,u}(S)}\right)}{\lambda_{N-1}\left(g_u C_{N-1,n_r(u)}\right)} \lambda_N\left(\left(g_u C_{N-1,n_r(u)}\right) \oplus B_N(\mathbf{0}, R)\right).
\end{aligned}
$$

Then by Facts 1 and 3 and (B.6), we have

$$
\begin{aligned}
&\lambda_N\left(\left(g_u\overline{J_{2,u}(S)-J_{1,u}(S)}\right)\oplus B_N(\mathbf{0},R)\right)-\lambda_N\left(g_u\overline{J_{2,u}(S)-J_{1,u}(S)}\right)\\
&\leq \frac{\lambda_{N-1}\left(g_u\partial\overline{J_{2,u}(S)-J_{1,u}(S)}\right)}{\lambda_{N-1}\left(g_uC_{N-1,n_r(u)}\right)}\lambda_N\left(\left(g_uC_{N-1,n_r(u)}\right)\oplus B_N(\mathbf{0},R)\right)\\
&=\lambda_{N-1}\left(g_u\partial\overline{J_{2,u}(S)-J_{1,u}(S)}\right)\frac{\lambda_N(B_N(\mathbf{0},R))+2\sum_{k=0}^{N-1}b_{N-1,k}R^{N-k}\lambda_k(C_{k,n_r(u)})g_u^k}{g_u^{N-1}\lambda_{N-1}\left(C_{N-1,n_r(u)}\right)}\\
&=R^NO\left(\lambda_{N-1}\left(g_u\partial\overline{J_{2,u}(S)-J_{1,u}(S)}\right)\sum_{k=0}^{N-1}\frac{\lambda_k(C_{k,n_r(u)})g_u^k}{g_u^{N-1}\lambda_{N-1}\left(C_{N-1,n_r(u)}\right)}\right)\\
&=R^NO\left(\lambda_{N-1}\left(g_u\partial\overline{J_{2,u}(S)-J_{1,u}(S)}\right)\sum_{k=0}^{N-1}\left(g_u2^{-n_r(u)}\right)^{-(N-1-k)}\right)\\
&=R^NO\left(\lambda_{N-1}\left(g_u\partial\overline{J_{2,u}(S)-J_{1,u}(S)}\right)\right).
\end{aligned}
\tag{B.10}
$$

Finally, combining (B.9) with (B.10) implies that $\{g_u\overline{J_{2,u}(S)-J_{1,u}(S)}, u\in\mathbb{R}\}$ is a blowing-up system with $\delta_1=0$ and $\alpha=\alpha_0$ in Definition 3.2.3.

□

B.1.3 Proof of Theorem 3.3.1

Throughout this section, let (X,g_u) be a qualified pair with the adapted grid-block system $\{(G_u,B_u)\in 2^{R^N}\times 2^{R^N}, u\in\mathbb{R}\}$ as defined in Definition 3.3.3.

Lemma B.1.3. *Let $H(x)=x^{N-1}\exp\{-x^2/2\}$, and $x_u,y_u\to\infty$ as $u\to\infty$. Then*

1. *if $H(x_u)/H(y_u)\to\infty$ as $u\to\infty$, we have*

$$y_u^2-x_u^2\to\infty \quad \text{as } u\to\infty;$$

2. *if $H(x_u)/H(y_u)\to C$ for some constant $C>0$ as $u\to\infty$, we have*

$$x_u=O(y_u)\quad\text{and}\quad y_u=O(x_u).$$

Proof. For the first part, we assume that there exist a finite constant $C > 0$ and a subsequence $u_n \uparrow \infty$ such that $y_{u_n}^2 - x_{u_n}^2 < C$. Since $H(x_u)/H(y_u) \to \infty$, we can also assume $x_{u_n} < y_{u_n}$ for sufficiently large n. Then

$$\frac{H(x_{u_n})}{H(y_{u_n})} = \left(\frac{x_u}{y_u}\right)^{N-1} \exp\left\{-\frac{x_u^2 - y_u^2}{2}\right\} \le \exp\left\{\frac{C}{2}\right\} < \infty,$$

which leads to a contradiction. For the second part, if there exists a subsequence $u_n \uparrow \infty$ such that $x_{u_n}/y_{u_n} \uparrow \infty$, then for sufficiently large n,

$$\left(\frac{x_{u_n}}{y_{u_n}}\right)^2 - 1 > \frac{x_{u_n}}{y_{u_n}} > \frac{x_{u_n}}{y_{u_n}^3},$$

and therefore

$$x_{u_n}^2 - y_{u_n}^2 > \frac{x_{u_n}}{y_{u_n}}.$$

Then we have

$$\frac{H(x_{u_n})}{H(y_{u_n})} = \left(\frac{x_{u_n}}{y_{u_n}}\right)^{N-1} \exp\left\{-\frac{x_{u_n}^2 - y_{u_n}^2}{2}\right\} < \left(\frac{x_{u_n}}{y_{u_n}}\right)^{N-1} \exp\left\{-\frac{x_{u_n}}{2y_{u_n}}\right\} \to 0 \quad \text{as } n \to \infty,$$

which leads to a contradiction. The other half, $y_u = O(x_u)$, can be proved symmetrically by observing $H(y_u)/H(x_u) \to 1/C$. □

Recall that by Lemma 3.2.5 and Lemma 3.3.6, $\ell_{X,g_u B_u(K)}$ denotes the solution of (3.6) for the blowing-up system $\{g_u B_u(K), u \in \mathbb{R}\}$ and the Gaussian random field X, where $K \in \mathcal{J}$. Similarly, for $1 \le j \le N$, $\ell_{X_j, g_u \overline{K}}$ denotes the solution of (3.6) for the blowing-up system $\{g_u \overline{K}, u \in \mathbb{R}\}$ and Gaussian random field X_j. Then we have the following lemma.

Lemma B.1.4. *There exist some finite positive constants c_1 and c_2 such that as $u \to \infty$,*

1. $\frac{H(\ell_{X_j, g_u \overline{K}})}{H(u)} \to c_1$;
2. $\frac{H(\ell_{X, g_u B_u(K)})}{H(u)} \to c_2$.

In general, for any blowing-up system $\{T_u, u \in \mathbb{R}\}$, there exists a finite constant $c_3 > 0$ such that

$$\frac{H(\ell_{X,T_u})}{H(u)} \frac{\lambda_N(T_u)}{g_u^N} \to c_3$$

as $u \to \infty$.

Proof. The proof of this lemma makes use of the similarity between Lemmas 3.2.5 and 3.2.2. Note that by Lemma 3.2.2 and the definition of g_u (see (3.11)), we have as $u \to \infty$,

$$H(u)\lambda_N(g_u K) \to k_1 \tag{B.11}$$

for some finite constant $k_1 > 0$. By Lemma 3.2.5, we have

$$H(\ell_{X_j, g_u \overline{K}})\lambda_N(g_u K) = k_2, \tag{B.12}$$

$$H(\ell_{X, g_u B_u(K)})\lambda_N(g_u B_u(K)) = k_3 \tag{B.13}$$

and

$$H(\ell_{X,T_u})\lambda_N(T_u) = k_4 \tag{B.14}$$

for any $1 \leq j \leq N$ and some finite constants $k_2, k_3, k_4 > 0$. By $0 < f_u \to 1$ and (3.15), we have

$$\frac{\lambda_N(B_u(K))}{\lambda_N(K)} \to 1. \tag{B.15}$$

Finally combining (B.11)-(B.15) completes the proof. □

Lemma B.1.5. *Let $\{T_u, u \in \mathbb{R}\}$ be a blowing-up system. If*

$$\lim_{u\to\infty} g_u^{-N} \lambda_N(T_u) = 0,$$

then

$$\lim_{u\to\infty} P\left[\max_{\boldsymbol{t}\in T_u} X(\boldsymbol{t}) > u\right] = 0.$$

Proof. Let $Z_{X,T_u} = \max_{\boldsymbol{t}\in T_u} \left(X(\boldsymbol{t}) - \ell_{X,T_u}\right)\ell_{X,T_u}$. Then it suffices to show as $u \to \infty$,

$$\left(Z_{X,T_u}/\ell_{X,T_u} + \ell_{X,T_u}\right)^2 - u^2 \xrightarrow{p} -\infty,$$

where "$\xrightarrow{p}$" denotes the convergence in probability. According to Lemma 3.2.5, Z_{X,T_u} converges in distribution. Note that $\ell_{X,T_u} \to \infty$ as $u \to \infty$. Then it suffices to show $u^2 - \ell^2_{X,T_u} \to \infty$ as $u \to \infty$. Since

$$\lim_{u\to\infty} g_u^{-N} \lambda_N(T_u) = 0,$$

by Lemma B.1.4, we have

$$\frac{H(\ell_{X,T_u})}{H(u)} \to \infty \quad \text{as } u \to \infty.$$

Then by Lemma B.1.3,

$$u^2 - \ell^2_{X,T_u} \to \infty \quad \text{as } u \to \infty,$$

and hence the proof is completed. □

Lemma B.1.6. *For any $1 \le j \le N$,*

- *if $h_u/u \to \infty$ as $u \to \infty$, then*

$$\lim_{u\to\infty} P\left[\max_{\boldsymbol{t}\in g_u\overline{K}} |X_j(\boldsymbol{t})| > h_u\right] = 0;$$

- *for any blowing-up system $\{T_u, u \in \mathbb{R}\}$, if $h_u/\ell_{X_j,T_u} \to \infty$ as $u \to \infty$, then*

$$\lim_{u\to\infty} P\left[\max_{\boldsymbol{t}\in T_u} |X_j(\boldsymbol{t})| > h_u\right] = 0.$$

Proof. Let $Z_{X_j,g_u\overline{K}} = \max_{\boldsymbol{t}\in g_u\overline{K}}\left(\sigma_j^{-1}X_j(\boldsymbol{t}) - \ell_{X_j,g_u\overline{K}}\right)\ell_{X_j,g_u\overline{K}}$, where $\sigma_j = \mathrm{Var}[X_j(\mathbf{0})]$. According to Lemma 3.2.5, $Z_{X_j,g_u\overline{K}}$ converges in distribution and $\ell_{X_j,g_u\overline{K}} \to \infty$ as $u \to \infty$. By Lemma B.1.4, we have

$$\frac{H(\ell_{X_j,g_u\overline{K}})}{H(u)} \to c_1 \quad \text{as } u \to \infty.$$

Then by Lemma B.1.3, we have

$$\ell_{X_j,g_u\overline{K}} = O(u).$$

Since $h_u/u \to \infty$ as $u \to \infty$,

$$h_u/\ell_{X_j,g_u\overline{K}} \to \infty$$

as $u \to \infty$, which implies that

$$Z_{X_j,g_u\overline{K}}/\ell_{X_j,g_u\overline{K}} + \ell_{X_j,g_u\overline{K}} - \sigma_j^{-1}h_u \xrightarrow{p} -\infty$$

as $u \to \infty$. Then we have

$$\begin{aligned} P\left[\max_{\boldsymbol{t}\in g_u\overline{K}} |X_j(\boldsymbol{t})| > h_u\right] &\le P\left[\max_{\boldsymbol{t}\in g_u\overline{K}} X_j(\boldsymbol{t}) > h_u\right] + P\left[\min_{\boldsymbol{t}\in g_u\overline{K}} X_j(\boldsymbol{t}) < -h_u\right] \\ &= 2P\left[\max_{\boldsymbol{t}\in g_u\overline{K}} X_j(\boldsymbol{t}) > h_u\right] \\ &= 2P\left[Z_{X_j,u}/\ell_{X_j,g_u\overline{K}} + \ell_{X_j,g_u\overline{K}} - \sigma_j h_u > 0\right] \\ &\to 0 \quad \text{as } u \to \infty. \end{aligned}$$

The proof for the second part is similar. □

Lemma B.1.7.
$$P\left[M_u(X, g_u K) = 0\right] \overset{u}{\approx} P\left[M_u(X, g_u B_u(K)) = 0\right]$$

Proof. By (B.15) and Lemma 3.2.2, it is easy to see that

$$\frac{D_u(X, g_u B_u(K))}{D_u(X, g_u \overline{K})} = \frac{\lambda_N(B_u(K))}{\lambda_N(\overline{K})} \to 1$$

as $u \to \infty$, where $D_u(X, \cdot)$ is defined in (3.2). Note that the constant C in Lemma 3.2.2 is independent of the choice of search regions. Then

$$\begin{aligned}
&\left|\frac{\mathbb{E}\left[M_u(X, g_u K)\right] - \mathbb{E}\left[M_u(X, g_u B_u(K))\right]}{c\lambda_N(K)}\right| \frac{\mathbb{E}\left[M_u(X, g_u K)\right]}{D_u(X, g_u \overline{K})} \\
&= \left|\frac{\mathbb{E}\left[M_u(X, g_u K)\right] - \mathbb{E}\left[M_u(X, g_u B_u(K))\right]}{D_u(X, g_u \overline{K})}\right| \\
&\leq \left|\frac{\left(\mathbb{E}\left[M_u(X, g_u K)\right] - D_u(X, g_u \overline{K})\right)}{D_u(X, g_u \overline{K})}\right| + \left|\frac{\left(\mathbb{E}\left[M_u(X, g_u B_u(K))\right] - D_u(X, g_u(B_u(K)))\right]}{D_u(X, g_u \overline{K})}\right| \\
&\quad + \left|\frac{D_u(X, g_u(B_u(K)) - D_u(X, g_u \overline{K})}{D_u(X, g_u \overline{K})}\right| \\
&\leq \frac{C\left(D_u(X, g_u \overline{K}) + D_u(X, g_u(B_u(K))\right) u^{-1}}{D_u(X, g_u \overline{K})} + \left|\frac{D_u(X, g_u(B_u(K)) - D_u(X, g_u \overline{K})}{D_u(X, g_u \overline{K})}\right| \\
&\to 0 \quad \text{as } u \to \infty.
\end{aligned}$$

Thus, we have

$$|M_u(X, g_u K) - M_u(X, g_u B_u(K))| \xrightarrow{p} 0$$

as $u \to \infty$, which implies

$$\begin{aligned}
&P\left[M_u(X, g_u K) = 0\right] - P\left[M_u(X, g_u B_u(K)) = 0\right] \\
&\leq P\left[|M_u(X, g_u K) - M_u(X, g_u B_u(K)| > 0.5\right] \\
&\to 0 \quad \text{as } u \to \infty.
\end{aligned}$$

□

Lemma B.1.8. *Let $\{T_u, u \in \mathbb{R}\}$ be a blowing-up system. Suppose $\varepsilon_u \ell_{X,T_u} \to 0$ as $u \to \infty$. Then*

$$P\left[u - \varepsilon_u < \max_{\boldsymbol{t} \in T_u} X(\boldsymbol{t}) \leq u\right] \to 0 \quad \text{as } u \to \infty.$$

Proof. Let $F(x) = e^{-e^{-x}}$. Then $F'(x) = e^{-(e^{-x}+x)}$, $F''(x) = -e^{-(e^{-x}+x)}(1-e^{-x})$, and the Lipschitz constant of $F(x)$ is e^{-1}. Let $F_u(x) = P[(\max_{\boldsymbol{t}\in T_u} X(\boldsymbol{t}) - \ell_{X,T_u})\ell_{X,T_u} \le x]$. By Lemma 3.2.5, we have

$$\lim_{u\to\infty} F_u(x) = F(x)$$

for any continuity point x of F. Since F is continuous on $\mathbb{R}$, by the monotonicity of F_u and F, we have F_u converge to F uniformly. Thus, for any given $\varepsilon > 0$, there exists a finite constant U such that when $u > U$, we have $|F_u(x) - F(x)| < \varepsilon/3$ and $\varepsilon_u \ell_{X,T_u} < e\varepsilon/3$. Then

$$\begin{aligned}
&P\left[u - \varepsilon_u < \max_{\boldsymbol{t}\in T_u} X(\boldsymbol{t}) \le u\right] \\
&= P\left[\left(\max_{\boldsymbol{t}\in T_u} X(\boldsymbol{t}) - \ell_{X,T_u}\right)\ell_{X,T_u} \le (u - \ell_{X,T_u})\ell_{X,T_u}\right] \\
&\quad - P\left[\left(\max_{\boldsymbol{t}\in T_u} X(\boldsymbol{t}) - \ell_{X,T_u}\right)\ell_{X,T_u} \le (u - \varepsilon_u - \ell_{X,T_u})\ell_{X,T_u}\right] \\
&= F_u\left((u - \ell_{X,T_u})\ell_{X,T_u}\right) - F_u\left((u - \varepsilon_u - \ell_{X,T_u})\ell_{X,T_u}\right) \\
&\le 2\varepsilon/3 + \left|F\left((u - \ell_{X,T_u})\ell_{X,T_u}\right) - F\left((u - \varepsilon_u - \ell_{X,T_u})\ell_{X,T_u}\right)\right| \\
&\le 2\varepsilon/3 + e^{-1}\varepsilon_u \ell_{X,T_u} \\
&\le \varepsilon.
\end{aligned}$$

□

Lemma B.1.9.

$$P\left[\max_{\boldsymbol{t}\in g_u B_u(K)} X(\boldsymbol{t}) \le u\right] \overset{u}{\approx} P\left[\max_{\boldsymbol{t}\in (g_u B_u(K))\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right].$$

Proof. Let $h_u = 2N^{-1}d_u^{-1/2}$. Recall that in Definition 3.3.3, we have $h_u/u \to \infty$ as $u \to \infty$.

Then

$$
\begin{aligned}
&\left|P\left[\max_{\boldsymbol{t}\in(g_uB_u(K))\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right] - P\left[\max_{\boldsymbol{t}\in g_uB_u(K)} X(\boldsymbol{t}) \le u\right]\right| \\
&\le P\left[\max_{\boldsymbol{t}\in(g_uB_u(K))\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}, \max_{\boldsymbol{t}\in g_uB_u(K)} X(\boldsymbol{t}) > u\right] \\
&\quad + P\left[u - d_u^{1/2} < \max_{\boldsymbol{t}\in g_uB_u(K)} X(\boldsymbol{t}) \le u\right] \\
&\le P\left[\max_{\boldsymbol{t}\in g_uB_u(K)} \|\nabla X(\boldsymbol{t})\| > 2N^{-1/2}d_u^{-1}d_u^{1/2}\right] + P\left[u - d_u^{1/2} < \max_{\boldsymbol{t}\in g_uB_u(K)} X(\boldsymbol{t}) \le u\right] \\
&\le P\left[\max_{1\le j\le N}\max_{\boldsymbol{t}\in g_uB_u(K)} |X_j(\boldsymbol{t})| > h_u\right] + P\left[u - d_u^{1/2} < \max_{\boldsymbol{t}\in g_uB_u(K)} X(\boldsymbol{t}) \le u\right] \\
&\le \sum_{j=1}^{N} P\left[\max_{\boldsymbol{t}\in g_uB_u(K)} |X_j(\boldsymbol{t})| > h_u\right] + P\left[u - d_u^{1/2} < \max_{\boldsymbol{t}\in g_uB_u(K)} X(\boldsymbol{t}) \le u\right] \\
&\le \sum_{j=1}^{N} P\left[\max_{\boldsymbol{t}\in g_u\overline{K}} |X_j(\boldsymbol{t})| > h_u\right] + P\left[u - d_u^{1/2} < \max_{\boldsymbol{t}\in g_uB_u(K)} X(\boldsymbol{t}) \le u\right] \\
&\to 0 \quad \text{as } u \to \infty,
\end{aligned}
$$

where the second "$\le$" is derived from the mean value theorem and the fact that for any $\boldsymbol{t} \in \mathbb{R}^N$, there exists a grid point $\boldsymbol{t}_u \in G_u$ such that

$$\|\boldsymbol{t} - \boldsymbol{t}_u\| \le d_u N^{1/2}/2,$$

and the last convergence is derived from Lemmas B.1.6 and B.1.8. Note that to use Lemma B.1.8, we still need to show $d_u^{1/2}\ell_{X,g_uB_u(K)} \to 0$ as $u \to \infty$. By Lemma B.1.4, we have

$$\frac{H(\ell_{X,g_uB_u(K)})}{H(u)} \to c_2.$$

Then by Lemma B.1.3, we have

$$
\begin{aligned}
d_u^{1/2}\ell_{X,g_uB_u(K)} &= O\left(d_u^{1/2}u\right) \\
&= o(1),
\end{aligned}
$$

where the last line is from Definition 3.3.3. □

Lemma B.1.10.

$$b_u^N P\left[\max_{\boldsymbol{t}\in(g_uB_{u,\mathbf{0}})\cap G_u} X(\boldsymbol{t}) > u - d_u^{1/2}\right] \overset{u}{\approx} b_u^N P\left[\max_{\boldsymbol{t}\in g_uB_{u,\mathbf{0}}} X(\boldsymbol{t}) > u\right] \to c.$$

Proof. Let $h_u = 2N^{-1}d_u^{-1/2}$. Then

$$\left| P\left[\max_{\boldsymbol{t}\in(g_uB_{u,\mathbf{0}})\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right] - P\left[\max_{\boldsymbol{t}\in g_uB_{u,\mathbf{0}}} X(\boldsymbol{t}) \le u\right]\right|$$

$$\le P\left[\max_{\boldsymbol{t}\in(g_uB_{u,\mathbf{0}})\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}, \max_{\boldsymbol{t}\in g_uB_{u,\mathbf{0}}} X(\boldsymbol{t}) > u\right] + P\left[u - d_u^{1/2} < \max_{\boldsymbol{t}\in g_uB_{u,\mathbf{0}}} X(\boldsymbol{t}) \le u\right]$$

$$\le P\left[\max_{\boldsymbol{t}\in g_uB_{u,\mathbf{0}}} \|\nabla X(\boldsymbol{t})\| > 2N^{-1/2}d_u^{-1}d_u^{1/2}\right] + P\left[u - d_u^{1/2} < \max_{\boldsymbol{t}\in g_uB_{u,\mathbf{0}}} X(\boldsymbol{t}) \le u\right]$$

$$\le P\left[\max_{1\le j\le N}\max_{\boldsymbol{t}\in g_uB_{u,\mathbf{0}}} |X_j(\boldsymbol{t})| > h_u\right] + P\left[u - d_u^{1/2} < \max_{\boldsymbol{t}\in g_uB_{u,\mathbf{0}}} X(\boldsymbol{t}) \le u\right]$$

$$\le \sum_{j=1}^{N} P\left[\max_{\boldsymbol{t}\in g_uB_{u,\mathbf{0}}} |X_j(\boldsymbol{t})| > h_u\right] + P\left[u - d_u^{1/2} < \max_{\boldsymbol{t}\in g_uB_{u,\mathbf{0}}} X(\boldsymbol{t}) \le u\right]$$

$$= 2\sum_{j=1}^{N} P\left[\max_{\boldsymbol{t}\in g_uB_{u,\mathbf{0}}} X_j(\boldsymbol{t}) > h_u\right] + P\left[u - d_u^{1/2} < \max_{\boldsymbol{t}\in g_uB_{u,\mathbf{0}}} X(\boldsymbol{t}) \le u\right].$$

Recall that in Definition 3.3.3, we have

- $h_u/u \to \infty$ as $u \to \infty$
- $g_ub_u^{-1} = O(\exp\{\kappa u^2/2\})$ for
$$\kappa = \frac{1}{2N}\min\left\{1, \kappa_X, \kappa_{X_j}, 1 \le j \le N\right\},$$
where κ_X and $\kappa_{X_j}, 1 \le j \le N$, is defined in Lemma 3.2.8.

Then we have
$$\frac{h_u^N\Psi(h_u)}{u^N\Psi(u)} \to 0, \tag{B.16}$$
where Ψ is the tail distribution function of standard Gaussian distribution (see (A.3)),
$$g_u^Nb_u^{-N} = O(\exp\{\kappa_X u^2/2\}) \text{ and } g_u^Nb_u^{-N} = O(\exp\{\kappa_{X_j} h_u^2/2\}).$$
By Lemmas 3.2.2, 3.2.8, (A.4) and (3.8), we have

$$\begin{aligned}
&b_u^N P\left[\max_{\boldsymbol{t}\in g_uB_{u,\mathbf{0}}} X(\boldsymbol{t}) > u\right] \\
&= b_u^N H_2\lambda_N(g_ub_u^{-1}[0,1]^N)u^N 2^{-N/2}\det\,(\boldsymbol{\Lambda}_X)^{1/2}\,\Psi(u)(1+o(1)) \\
&\to c \text{ as } u \to \infty.
\end{aligned}$$

By Lemmas 3.2.2, 3.2.8, (A.4) and (B.16), we have

$$\begin{aligned}
& b_u^N P\left[\max_{\boldsymbol{t}\in g_u B_{u,\mathbf{0}}} X_j(\boldsymbol{t}) > h_u\right] \\
&= b_u^N P\left[\max_{\boldsymbol{t}\in g_u B_{u,\mathbf{0}}} \sigma_{X_j}^{-1} X_j(\boldsymbol{t}) > \sigma_{X_j}^{-1} h_u\right] \\
&= b_u^N H_2 \lambda_N(g_u b_u^{-1}[0,1]^N)\sigma_{X_j}^{-N} h_u^N 2^{-N/2} \det\left(\boldsymbol{\Lambda}_{X_j/\sigma_{X_j}}\right)^{1/2} \Psi(h_u)(1+o(1)) \\
&= b_u^N H_2 \lambda_N(g_u b_u^{-1}[0,1]^N)\sigma_{X_j}^{-N} u^N 2^{-N/2} \det\left(\boldsymbol{\Lambda}_{X_j/\sigma_{X_j}}\right)^{1/2} \Psi(u)\left(\frac{h_u^N \Psi(h_u)}{u^N \Psi(u)}\right)(1+o(1)) \\
&= H_2 \sigma_{X_j}^{-N} 2^{-N/2} \det\left(\boldsymbol{\Lambda}_{X_j/\sigma_{X_j}}\right)^{1/2} g_u^N u^N \Psi(u)\left(\frac{h_u^N \Psi(h_u)}{u^N \Psi(u)}\right)(1+o(1)) \\
&\le H_2 \sigma_{X_j}^{-N} 2^{-N/2} \det\left(\boldsymbol{\Lambda}_{X_j/\sigma_{X_j}}\right)^{1/2} g_u^N u^{N-1} \phi(u)\left(\frac{h_u^N \Psi(h_u)}{u^N \Psi(u)}\right)(1+o(1)) \\
&\to 0 \quad \text{as } u \to \infty.
\end{aligned}$$

Now the only thing left is to show

$$b_u^N P\left[u - d_u^{1/2} < \max_{\boldsymbol{t}\in g_u B_{u,\mathbf{0}}} X(\boldsymbol{t}) \le u\right] \to 0 \quad \text{as } u \to \infty.$$

Note that by Definition 3.3.3,

$$\begin{aligned}
g_u^N b_u^{-N} &= O(\exp\{N\kappa u^2/2\}) \\
&= O(\exp\{\kappa_X \left(u - d_u^{1/2}\right)^2/2\}).
\end{aligned}$$

Then by Lemma 3.2.8,

$$
\begin{aligned}
& b_u^N P\left[u-d_u^{1/2}<\max_{\boldsymbol{t}\in g_u B_{u,\mathbf{0}}} X(\boldsymbol{t})\le u\right] \\
&= b_u^N P\left[\max_{\boldsymbol{t}\in g_u B_{u,\mathbf{0}}} X(\boldsymbol{t})>u-d_u^{1/2}\right]-b_u^N P\left[\max_{\boldsymbol{t}\in g_u B_{u,\mathbf{0}}} X(\boldsymbol{t})>u\right] \\
&= b_u^N H_2\lambda_N(g_u B_{u,\mathbf{0}})2^{-N/2}\det\left(\boldsymbol{\Lambda}_X^{-1/2}\right)\left(u-d_u^{1/2}\right)^N\Psi\left(u-d_u^{1/2}\right)(1+o(1)) \\
&\quad - b_u^N H_2\lambda_N(g_u B_{u,\mathbf{0}})2^{-N/2}\det\left(\boldsymbol{\Lambda}_X^{-1/2}\right)u^N\Psi(u)(1+o(1)) \qquad \text{(B.17)} \\
&= b_u^N H_2\lambda_N(g_u B_{u,\mathbf{0}})2^{-N/2}\det\left(\boldsymbol{\Lambda}_X^{-1/2}\right)\left(\Psi\left(u-d_u^{1/2}\right)\left(u-d_u^{1/2}\right)^N-\Psi(u)u^N\right) \\
&\quad + o\left(g_u^N\Psi\left(u-d_u^{1/2}\right)u^N\right) \\
&= H_2 g_u^N f_u^N 2^{-N/2}\det\left(\boldsymbol{\Lambda}_X^{-1/2}\right)\left(\Psi\left(u-d_u^{1/2}\right)\left(u-d_u^{1/2}\right)^N-\Psi(u)u^N\right) \\
&\quad + o\left(g_u^N\Psi\left(u-d_u^{1/2}\right)u^N\right).
\end{aligned}
$$

Since $d_u^{1/2}u\to 0$ as $u\to\infty$ (see Definition 3.3.3), by (A.4), we have

$$
\begin{aligned}
\frac{\Psi\left(u-d_u^{1/2}\right)\left(u-d_u^{1/2}\right)^N}{\Psi(u)u^N}-1 &\le \frac{e^{-\left(u-d_u^{1/2}\right)^2/2}}{(1-u^{-2})e^{-u^2/2}}\frac{\left(u-d_u^{1/2}\right)^{N-1}}{u^{N-1}}-1 \\
&= \frac{1}{1-u^{-2}}\frac{\left(u-d_u^{1/2}\right)^{N-1}}{u^{N-1}}e^{d_u^{1/2}u-d_u/2}-1 \\
&= o(1).
\end{aligned}
$$

Note that by Lemma 3.2.2, $g_u^N=O(u^{-N+1}e^{u^2/2})$. Then for the first term in (B.17), we have

$$
\begin{aligned}
& H_2 g_u^N f_u^N 2^{-N/2}\det\left(\boldsymbol{\Lambda}_X^{-1/2}\right)\left(\Psi\left(u-d_u^{1/2}\right)\left(u-d_u^{1/2}\right)^N-\Psi(u)u^N\right) \\
&= H_2 g_u^N f_u^N 2^{-N/2}\det\left(\boldsymbol{\Lambda}_X^{-1/2}\right)\Psi(u)u^N\left(\frac{\Psi\left(u-d_u^{1/2}\right)\left(u-d_u^{1/2}\right)^N}{\Psi(u)u^N}-1\right) \\
&= o\left(g_u^N\Psi(u)u^N\right) \\
&= o\left(u^{-N+1}e^{u^2/2}e^{-u^2/2}u^{N-1}\right) \\
&= o(1).
\end{aligned}
$$

For the second term in (B.17), similarly, we have

$$\begin{aligned} g_u^N \Psi\left(u - d_u^{1/2}\right) u^N &= g_u^N \Psi\left(u - d_u^{1/2}\right)\left(u - d_u^{1/2}\right) u^N \left(u - d_u^{1/2}\right)^{-1} \\ &= O\left(u^{-N+1} e^{u^2/2} e^{-\left(u - d_u^{1/2}\right)^2/2} u^N \left(u - d_u^{1/2}\right)^{-1}\right) \\ &= O\left(u\left(u - d_u^{1/2}\right)^{-1}\right) O\left(e^{d_u^{1/2} u - d_u/2}\right) \\ &= O(1), \end{aligned}$$

and then

$$o\left(g_u^N \Psi\left(u - d_u^{1/2}\right) u^N\right) = o(1).$$

□

Lemma B.1.11.

$$\begin{aligned} &P\left[\bigcap_{\Delta_{u,\boldsymbol{k}} \subset K, \boldsymbol{k} \in \mathbb{Z}^N} \left\{\max_{\boldsymbol{t} \in (g_u B_{u,\boldsymbol{k}}) \cap G_u} X(\boldsymbol{t}) \leq u - d_u^{1/2}\right\}\right] \\ \overset{u}{\approx} &\prod_{\Delta_{u,\boldsymbol{k}} \subset K, \boldsymbol{k} \in \mathbb{Z}^N} P\left[\max_{\boldsymbol{t} \in (g_u B_{u,\boldsymbol{k}}) \cap G_u} X(\boldsymbol{t}) \leq u - d_u^{1/2}\right]. \end{aligned}$$

Proof. By Definition 3.3.3, we have,

$$g_u b_u^{-1}(1 - f_u) \to \infty$$

as $u \to \infty$, and thus,

$$\overline{r}(2 g_u b_u^{-1}(1 - f_u)) \to 0$$

as $u \to \infty$. Then by Lemma 3.2.9 and Definition 3.3.3,

$$P\left[\bigcap_{\Delta_{u,\boldsymbol{k}}\subset K,\boldsymbol{k}\in\mathbb{Z}^N}\left\{\max_{\boldsymbol{t}\in(g_uB_{u,\boldsymbol{k}})\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right\}\right] - \prod_{\Delta_{u,\boldsymbol{k}}\subset K,\boldsymbol{k}\in\mathbb{Z}^N} P\left[\max_{\boldsymbol{t}\in(g_uB_{u,\boldsymbol{k}})\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right]$$

$$\le \frac{1}{2\pi}\binom{v}{2} m^2 \overline{r}\left(2g_ub_u^{-1}(1-f_u)\right)\left(1-\overline{r}^2\left(2g_ub_u^{-1}(1-f_u)\right)\right)^{-1/2} \exp\left\{-\frac{\left(u-d_u^{1/2}\right)^2}{1+\overline{r}\left(2g_ub_u^{-1}(1-f_u)\right)}\right\}$$

$$= O\left((mv)^2\overline{r}\left(2g_ub_u^{-1}(1-f_u)\right)\exp\left\{-\frac{\left(u-d_u^{1/2}\right)^2}{1+\overline{r}\left(2g_ub_u^{-1}(1-f_u)\right)}\right\}\right)$$

$$= O\left(g_u^{2N}d_u^{-2N}\overline{r}(2g_ub_u^{-1}(1-f_u))\exp\left\{-\frac{\left(u-d_u^{1/2}\right)^2}{1+\overline{r}\left(2g_ub_u^{-1}(1-f_u)\right)}\right\}\right)$$

$$= O\left(g_u^{2N}d_u^{-2N}\overline{r}(2g_ub_u^{-1}(1-f_u))\exp\left\{-\frac{u^2}{1+\overline{r}\left(2g_ub_u^{-1}(1-f_u)\right)}\right\}\right)$$

$$= o(1),$$

where, for shortness, we let $m = \#((g_uB_{u,\mathbf{0}}) \cap G_u)$, i.e., the number of grid points in each block, and let $v = n_u(K)$. Here it is easy to see that $mv = O(g_u^N d_u^{-N})$. □

Now we have prepared well for a formal proof of Theorem 3.3.1.

Proof. If $\lambda_N(K) = 0$, then by Lemma 3.2.2,

$$P\left[M_u(X, g_uK) = 0\right] = \exp\left\{-c\lambda_N(K)\right\} = 0.$$

If $\lambda_N(K) > 0$, let $H_u(K) := \{\max_{\boldsymbol{t}\in g_u\overline{K-J_{1,u}(K)}} X(\boldsymbol{t}) \le u\}$, where $J_{1,u}(K)$ is defined in Lemma 3.3.7. Thus

$$P[H_u(K)] \ge P\left[\max_{\boldsymbol{t}\in g_u\overline{J_{2,u}(K)-J_{1,u}(K)}} X(\boldsymbol{t}) \le u\right].$$

Then by Lemmas B.1.5 and 3.3.7, we have

$$\lim_{u\to\infty} P[H_u(K)] = 1. \tag{B.18}$$

By Lemma B.1.7,

$$P\left[M_u(X, g_u K) = 0\right] \overset{u}{\approx} P\left[M_u(X, g_u B_u(K)) = 0\right],$$

where $B_u(K)$ is defined in (3.14). Then by (B.18) and Inequality (3.16), we have

$$\begin{aligned} P\left[M_u(X, g_u B_u(K)) = 0\right] &\overset{u}{\approx} P\left[M_u(X, g_u B_u(K)) = 0 \;\middle|\; H_u(K)\right]. \\ &= P\left[\max_{\boldsymbol{t}\in g_u B_u(K)} X(\boldsymbol{t}) \le u \;\middle|\; H_u(K)\right]. \end{aligned}$$

Again, by (B.18) and Inequality (3.16),

$$P\left[\max_{\boldsymbol{t}\in g_u B_u(K)} X(\boldsymbol{t}) \le u \;\middle|\; H_u(K)\right] \overset{u}{\approx} P\left[\max_{\boldsymbol{t}\in g_u B_u(K)} X(\boldsymbol{t}) \le u\right].$$

By Lemma B.1.9,

$$P\left[\max_{\boldsymbol{t}\in g_u B_u(K)} X(\boldsymbol{t}) \le u\right] \overset{u}{\approx} P\left[\max_{\boldsymbol{t}\in (g_u B_u(K))\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right].$$

Since $B_u(K) = \cup_{\Delta_{u,\boldsymbol{k}}\subset K, \boldsymbol{k}\in\mathbb{Z}^N} B_{u,\boldsymbol{k}}$, we have

$$P\left[\max_{\boldsymbol{t}\in (g_u B_u(K))\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right] = P\left[\bigcap_{\Delta_{u,\boldsymbol{k}}\subset K, \boldsymbol{k}\in\mathbb{Z}^N} \left\{\max_{\boldsymbol{t}\in (g_u B_{u,\boldsymbol{k}})\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right\}\right].$$

By Lemma B.1.11,

$$\begin{aligned} &P\left[\bigcap_{\Delta_{u,\boldsymbol{k}}\subset K, \boldsymbol{k}\in\mathbb{Z}^N} \left\{\max_{\boldsymbol{t}\in (g_u B_{u,\boldsymbol{k}})\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right\}\right] \\ \overset{u}{\approx} &\prod_{\Delta_{u,\boldsymbol{k}}\subset K, \boldsymbol{k}\in\mathbb{Z}^N} P\left[\max_{\boldsymbol{t}\in (g_u B_{u,\boldsymbol{k}})\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right]. \end{aligned}$$

By the stationarity of X and Condition 4 in Definition 3.3.3,

$$\prod_{\Delta_{u,\boldsymbol{k}}\subset K,\boldsymbol{k}\in\mathbb{Z}^N} P\left[\max_{\boldsymbol{t}\in(g_u B_{u,\boldsymbol{k}_i})\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right]$$
$$= \left(P\left[\max_{\boldsymbol{t}\in(g_u B_{u,\boldsymbol{0}})\cap G_u} X(\boldsymbol{t}) \le u - d_u^{1/2}\right]\right)^{n_u(K)}$$
$$= \left(1 - \frac{n_u(K)P\left[\max_{\boldsymbol{t}\in(g_u B_{u,\boldsymbol{0}})\cap G_u} X(\boldsymbol{t}) > u - d_u^{1/2}\right]}{n_u(K)}\right)^{n_u(K)}.$$

By Lemma B.1.10,

$$b_u^N P\left[\max_{\boldsymbol{t}\in(g_u B_{u,\boldsymbol{0}})\cap G_u} X(\boldsymbol{t}) > u - d_u^{1/2}\right] \overset{u}{\approx} b_u^N P\left[\max_{\boldsymbol{t}\in g_u B_{u,\boldsymbol{0}}} X(\boldsymbol{t}) > u\right]$$
$$\to c \quad \text{as } u \to \infty.$$

Then by (3.15),

$$n_u(K)P\left[\max_{\boldsymbol{t}\in(g_u B_{u,\boldsymbol{0}})\cap G_u} X(\boldsymbol{t}) > u - d_u^{1/2}\right] = b_u^{-N} n_u(K) b_u^N P\left[\max_{\boldsymbol{t}\in(g_u B_{u,\boldsymbol{0}})\cap G_u} X(\boldsymbol{t}) > u - d_u^{1/2}\right]$$
$$\to c\lambda_N(K) \quad \text{as } u \to \infty.$$

Finally, combining all of the above implies

$$\lim_{u\to\infty} P\left[M_u(X, g_u K) = 0\right] = \exp\left\{-c\lambda_N(K)\right\}$$

for any $K \subset \mathcal{J}$. □

B.1.4 Proof of Theorem 3.4.5

Theorem B.1.12. *Let the pair (X, g_u) be qualified. Then for any bounded set $S \subset \mathbb{R}^N$ with $\lambda_{N-1}(\partial S) < \infty$, we have*

$$\lim_{u\to\infty} \mathbb{E}\left[M_u(X, g_u S, k)\right] = 0$$

for every $0 \le k \le N-1$.

Proof. This proof is a generalization of the proof of Theorem 6.3.1 in [1]. By Lemma 3.2.2 and $\lambda_{N-1}(\partial S) < \infty$, it suffices to show that as $u \to \infty$,

$$\frac{\mathbb{E}\left[M_u(X, g_u S, N)\right]}{\mathbb{E}\left[\sum_{k=0}^{N-1} M_u(X, g_u S, k) + M_u(X, g_u S, N)\right]} \to 1$$

for any compact $S \subset \mathbb{R}^N$ with $\lambda_{N-1}(\partial S) < \infty$.

Step 1:

Without loss of generality, we can assume that the various first-order partial derivatives are uncorrelated. If this is not the case, it can be achieved by an appropriate orthogonal transformation of the parameter space without changing $\mathbb{E}\left[M_u(X, g_u S, k)\right]$. More specifically, note that X is stationary, let $\boldsymbol{\Lambda}_X$ be the covariance matrix of $\triangledown X(\mathbf{0})$ and let $r(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^N$ be the covariance function of X. Since (X, g_u) is qualified, we have

$$\mathrm{Cov}\left[X_i(\boldsymbol{t}), X_j(\boldsymbol{t})\right] = -r_{ij}(\mathbf{0})$$

for any $1 \leq i, j \leq N$. Since the joint distribution of $\triangledown X(\boldsymbol{t})$ is non-degenerate for any $\boldsymbol{t} \in \mathbb{R}^N$, $\boldsymbol{\Lambda}_X$ is positive-definite and we can find an orthogonal matrix $\boldsymbol{Q}_X \in \mathbb{R}^{N \times N}$ such that

$$\boldsymbol{Q}_X \boldsymbol{\Lambda}_X \boldsymbol{Q}_X^T = \mathrm{diag}\,(\mu_1, \ldots, \mu_N),$$

where μ_i, $1 \leq i \leq N$ are the eigenvalues of $\boldsymbol{\Lambda}_X$. Define

$$X^{\boldsymbol{Q}_X}(\boldsymbol{t}) := X(\boldsymbol{Q}_X \boldsymbol{t})$$

for any $\boldsymbol{t} \in \mathbb{R}^{N \times 1}$. Then $X^{\boldsymbol{Q}_X}(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^{N \times 1}$ is also a centered, stationary random field with covariance function

$$r^{\boldsymbol{Q}_X}(\boldsymbol{t}) = r(\boldsymbol{Q}_X \boldsymbol{t})$$

for $\boldsymbol{t} \in \mathbb{R}^N$. It is easy to check that the pair $(X^{\boldsymbol{Q}_X}, g_u)$ is also qualified, and we have

$$\triangledown X^{\boldsymbol{Q}_X} = \boldsymbol{Q}_X \triangledown X.$$

Thus, the covariance matrix of $\triangledown X^{\boldsymbol{Q}_X}(\boldsymbol{t})$, written as $\boldsymbol{\Lambda}_{\boldsymbol{Q}_X}$, can be calculated by

$$\boldsymbol{\Lambda}_{\boldsymbol{Q}_X} = \mathbb{E}\left[\triangledown X^{\boldsymbol{Q}_X}(\boldsymbol{t}) \triangledown X^{\boldsymbol{Q}_X}(\boldsymbol{t})^T\right] = \boldsymbol{Q}_X \boldsymbol{\Lambda}_X \boldsymbol{Q}_X^T = \mathrm{diag}\,(\mu_1, \ldots, \mu_N),$$

which implies the uncorrelatedness between the various first-order partial derivatives of $X^{\boldsymbol{Q}_X}$. Moreover, note that the function $f^{\boldsymbol{Q}_X}(\boldsymbol{t}) = \boldsymbol{Q}_X \boldsymbol{t}$, $\boldsymbol{t} \in \mathbb{R}^N$ is a homeomorphism.

Thus, for any $0 \leq k \leq N$, $\boldsymbol{t}_0$ is a critical point with index k of $X^{\boldsymbol{Q}_X}$ is equivalent to that $\boldsymbol{Q}_X \boldsymbol{t}_0$ is a critical point with index k of X. For any Lebesgue measurable set $A \subset \mathbb{R}^N$, let

$$f^{\boldsymbol{Q}_X}(A) := \{\boldsymbol{Q}_X \boldsymbol{t} : \boldsymbol{t} \in \mathbb{R}^{N\times 1}\}.$$

Since $\det(\boldsymbol{Q}_X) = 1$, we have

$$\lambda_N\left(f^{\boldsymbol{Q}_X}(A)\right) = \lambda_N(A).$$

Let $\boldsymbol{x}''$ be the vectorization (see the second paragraph in Section 3.2) of the symmetric matrix $(x_{ij})_{1\leq i,j\leq N}$ and let $\phi(\boldsymbol{x}'', \boldsymbol{x}', x)$ be the joint probability density function of $(\nabla^2 X, \nabla X, X)$, where $\nabla^2 X$ is also vectorized in the same way, and we drop the parameter $\boldsymbol{t} \in \mathbb{R}^N$ in this density function since X and its derivatives are all stationary Gaussian random fields. Similarly, let $\phi(x)$, $\phi(\boldsymbol{x}')$, $\phi(\boldsymbol{x}'', x)$ and $\phi(\boldsymbol{x}''|x)$ be the probability density functions of X, ∇X, $(\nabla^2 X, X)$, and $(\nabla^2 X|X)$, respectively. Then by Corollary 3.4.1 and the stationarities of $X^{\boldsymbol{Q}_X}$ and X, we have

$$\begin{aligned}
\mathbb{E}\left[M_u(X, g_u S, k)\right] &= \int_{\boldsymbol{t}\in g_u S} \int_u^\infty \int_{D_k} |\det(\boldsymbol{x}'')| \phi(\boldsymbol{x}'', \boldsymbol{0}, x) d\boldsymbol{x}'' dx d\boldsymbol{t} \\
&= \lambda_N(g_u S) \int_u^\infty \int_{D_k} |\det(\boldsymbol{x}'')| \phi(\boldsymbol{x}'', \boldsymbol{0}, x) d\boldsymbol{x}'' dx \\
&= \lambda_N\left(g_u f^{\boldsymbol{Q}_X}(S)\right) \int_u^\infty \int_{D_k} |\det(\boldsymbol{x}'')| \phi(\boldsymbol{x}'', \boldsymbol{0}, x) d\boldsymbol{x}'' dx \\
&= \int_{\boldsymbol{t}\in g_u f^{\boldsymbol{Q}_X}(S)} \int_u^\infty \int_{D_k} |\det(\boldsymbol{x}'')| \phi(\boldsymbol{x}'', \boldsymbol{0}, x) d\boldsymbol{x}'' dx d\boldsymbol{t} \\
&= \mathbb{E}\left[M_u\left(X^{\boldsymbol{Q}_X}, g_u f^{\boldsymbol{Q}_X}(S), k\right)\right],
\end{aligned}$$

where $D_k \subset \mathbb{R}^{N(N+1)/2}$ is defined in Corollary 3.4.1.

Step 2:

By (A.9) and (2.8), we have

$$\phi(\boldsymbol{x}'', \boldsymbol{0}, x) = \phi(\boldsymbol{x}'', x)\phi(\boldsymbol{0})$$

for any $x \in \mathbb{R}$ and $\boldsymbol{x}'' \in \mathbb{R}^{N(N+1/2)}$. Then by Corollary 3.4.1 and the stationarity of X, we have

$$
\begin{aligned}
\mathbb{E}\left[M_u(X,g_uS,k)\right] &= \lambda_N(g_uS)\int_u^\infty\int_{D_k}|\mathrm{det}\,(\boldsymbol{x}'')|\phi(\boldsymbol{x}'',\mathbf{0},x)d\boldsymbol{x}''dx\\
&= \lambda_N(g_uS)\int_u^\infty\int_{D_k}|\mathrm{det}\,(\boldsymbol{x}'')|\phi(\boldsymbol{x}'',x)\phi(\mathbf{0})d\boldsymbol{x}''dx\\
&= \lambda_N(g_uS)\int_u^\infty\int_{D_k}|\mathrm{det}\,(\boldsymbol{x}'')|\phi(\boldsymbol{x}''|x)\phi(x)\phi(\mathbf{0})d\boldsymbol{x}''dx\\
&= \lambda_N(g_uS)(2\pi)^{-N/2}\mathrm{det}\,(\boldsymbol{\Lambda}_X)^{-1/2}\int_u^\infty\phi(x)\int_{D_k}|\mathrm{det}\,(\boldsymbol{x}'')|\phi(\boldsymbol{x}''|x)d\boldsymbol{x}''dx
\end{aligned}
$$

for any $0\le k\le N$. Note that $\mu_j, 1\le j\le N$ is the eigenvalue of $\boldsymbol{\Lambda}_X$ on the j-th column. For $1\le i,j\le N$, we make the change of variables

$$v_{ij}=x_{ij}+x\delta_{ij}\mu_j,$$

where δ_{ij} is the Kronecker delta. Then by (A.6),

$$\int_u^\infty\phi(x)\int_{D_k}|\mathrm{det}\,(\boldsymbol{x}'')|\phi(\boldsymbol{x}''|x)d\boldsymbol{x}''dx=(-1)^k\int_u^\infty\phi(x)\int_{D_{k,x}}\mathrm{det}\,(\boldsymbol{v}-x\boldsymbol{\Lambda}_X)\phi^*(\boldsymbol{v})d\boldsymbol{v}dx,$$

where $D_{k,x}\subset\mathbb{R}^{N(N+1)/2}$ is the region over which $\boldsymbol{v}-x\boldsymbol{\Lambda}_X$ has exactly k negative eigenvalues, and $\phi^*(\boldsymbol{v})$ is a zero-mean Gaussian density independent of u and k. We can write

$$\mathrm{det}\,(\boldsymbol{v}-x\boldsymbol{\Lambda}_X)=\sum_{\ell=0}^N b_\ell(\boldsymbol{v})x^\ell,$$

where b_ℓ, $0\le\ell\le N$ are multivariate polynomial functions of v_{ij}, and particularly $b_N(\boldsymbol{v})=(-1)^N\mathrm{det}\,(\boldsymbol{\Lambda}_X)$. Furthermore, for any $0\le k\le N$ and $x>u$, we can write

$$(-1)^k\int_{D_{k,x}}\mathrm{det}\,(\boldsymbol{v}-x\boldsymbol{\Lambda}_X)\phi^*(\boldsymbol{v})d\boldsymbol{v}=\sum_{\ell=0}^N c_{k,x,\ell}x^\ell,$$

where

$$c_{k,x,\ell}=(-1)^k\int_{D_{k,x}}b_\ell(\boldsymbol{v})\phi^*(\boldsymbol{v})d\boldsymbol{v}$$

for $0 \leq \ell \leq N$. Therefore

$$\begin{aligned}
&\frac{\mathbb{E}\left[M_u(X, g_uS, N)\right]}{\mathbb{E}\left[\sum_{k=0}^{N-1} M_u(X, g_uS, k) + M_u(X, g_uS, N)\right]} \\
&= \frac{\int_u^\infty \phi(x)(-1)^N \int_{D_{N,x}} \det(\boldsymbol{v} - x\boldsymbol{\Lambda}_X)\phi^*(\boldsymbol{v})d\boldsymbol{v}dx}{\int_u^\infty \phi(x) \sum_{k=0}^N (-1)^k \int_{D_{k,x}} \det(\boldsymbol{v} - x\boldsymbol{\Lambda}_X)\phi^*(\boldsymbol{v})d\boldsymbol{v}dx} \\
&= \frac{\int_u^\infty \phi(x) \sum_{\ell=0}^N c_{N,x,\ell} x^\ell dx}{\int_u^\infty \phi(x) \sum_{k=0}^N \sum_{\ell=0}^N c_{k,x,\ell} x^\ell dx} \\
&= \frac{\sum_{\ell=0}^N \int_u^\infty c_{N,x,\ell}\phi(x) x^\ell dx}{\sum_{\ell=0}^N \sum_{k=0}^N \int_u^\infty c_{k,x,\ell}\phi(x) x^\ell dx}.
\end{aligned}$$

Since b_ℓ, $0 \leq \ell \leq N$ are multivariate polynomial functions of v_{ij}, we have for any $0 \leq k, \ell \leq N$ and $x > u$,

$$\begin{aligned}
|c_{k,x,\ell}| &= \left| \int_{D_{k,x}} b_\ell(\boldsymbol{v})\phi^*(\boldsymbol{v})d\boldsymbol{v} \right| \\
&\leq \int_{\mathbb{R}^{N(N+1/2)}} |b_\ell(\boldsymbol{v})|\phi^*(\boldsymbol{v})d\boldsymbol{v} \\
&\leq \max_{1\leq j\leq N} \int_{\mathbb{R}^{N(N+1/2)}} |b_j(\boldsymbol{v})|\phi^*(\boldsymbol{v})d\boldsymbol{v} \\
&< \infty,
\end{aligned}$$

and especially,

$$\begin{aligned}
|c_{N,x,N}| &= \left| \int_{D_{N,x}} b_N(\boldsymbol{v})\phi^*(\boldsymbol{v})d\boldsymbol{v} \right| \\
&= |\det(\boldsymbol{\Lambda}_X)| \left| \int_{D_{N,x}} \phi^*(\boldsymbol{v})d\boldsymbol{v} \right| \\
&\to |\det(\boldsymbol{\Lambda}_X)| \left| \int_{\mathbb{R}^{N(N+1)/2}} \phi^*(\boldsymbol{v})d\boldsymbol{v} \right| \quad \text{as } x \to \infty,
\end{aligned}$$

where the last limit is given by Lemma 3.2.10. Therefore all of $|c_{k,x,\ell}|$ are uniformly bounded from above and all of $|c_{N,x,N}|$ are bounded from below for x large enough. This implies

$$\lim_{x\to\infty} \left| \frac{c_{N,x,N} x^N}{c_{k,x,\ell} x^\ell} \right| = \infty \tag{B.19}$$

for any $0 \leq k \leq N$ and $0 \leq \ell \leq N-1$. Then by L'Hospital's rule and (B.19),

$$
\begin{aligned}
&\lim_{u\to\infty} \frac{\mathbb{E}\left[M_u(X, g_u S, N)\right]}{\mathbb{E}\left[\sum_{k=0}^{N-1} M_u(X, g_u S, k) + M_u(X, g_u S, N)\right]} \\
&= \lim_{u\to\infty} \frac{\sum_{\ell=0}^{N} \int_u^\infty c_{N,x,\ell}\phi(x)x^\ell dx}{\sum_{\ell=0}^{N}\sum_{k=0}^{N} \int_u^\infty c_{k,x,\ell}\phi(x)x^\ell dx} \\
&= \lim_{u\to\infty} \frac{\sum_{\ell=0}^{N} c_{N,u,\ell}\phi(u)u^\ell}{\sum_{\ell=0}^{N}\sum_{k=0}^{N} c_{k,u,\ell}\phi(u)u^\ell} \\
&= \lim_{u\to\infty} \frac{c_{N,u,N}}{\sum_{k=0}^{N} c_{k,u,N}} \\
&= \lim_{u\to\infty} \frac{(-1)^N \int_{D_{N,u}} b_N(\boldsymbol{v})\phi^*(\boldsymbol{v})d\boldsymbol{v}}{\sum_{k=0}^{N}(-1)^k \int_{D_{k,u}} b_N(\boldsymbol{v})\phi^*(\boldsymbol{v})d\boldsymbol{v}} \\
&= \lim_{u\to\infty} \frac{(-1)^N \int_{D_{N,u}} (-1)^N \det\left(\boldsymbol{\Lambda}_X\right)\phi^*(\boldsymbol{v})d\boldsymbol{v}}{\sum_{k=0}^{N}(-1)^k \int_{D_{k,u}} (-1)^N \det\left(\boldsymbol{\Lambda}_X\right)\phi^*(\boldsymbol{v})d\boldsymbol{v}} \\
&= \lim_{u\to\infty} \frac{\int_{D_{N,u}} \phi^*(\boldsymbol{v})d\boldsymbol{v}}{\sum_{k=0}^{N}(-1)^{N+k} \int_{D_{k,u}} \phi^*(\boldsymbol{v})d\boldsymbol{v}} \\
&\geq \lim_{u\to\infty} \frac{\int_{D_{N,u}} \phi^*(\boldsymbol{v})d\boldsymbol{v}}{\int_{\mathbb{R}^{N(N+1)/2}} \phi^*(\boldsymbol{v})d\boldsymbol{v}} \\
&= 1,
\end{aligned}
$$

where the last limit is again given by Lemma 3.2.10. □

Appendix C

Appendix for Chapter 4

C.1 Derivatives of the Covariance Function

Let $\rho^{(i)}(\|\boldsymbol{t}\|^2)$, $\boldsymbol{t} \in \mathbb{R}$ and $1 \le i \le 6$ be the i-th derivative (if exists) of ρ at $\|\boldsymbol{t}\|^2$. Let $R_{i_1,\ldots,i_m}(\boldsymbol{t})$ be the partial derivative of $R(\boldsymbol{t})$ at $\boldsymbol{t} \in \mathbb{R}$ along the directions $i_1, \ldots, i_m$ for $i_1, \ldots, i_m \in \{1, 2, \ldots, N\}$. Let Π_n be the set of permutations on $\{1, \ldots, n\}$ for any $n \ge 1$. Then we can get the expressions of all partial derivatives of $R(\boldsymbol{t})$ up to the sixth order in terms of $\rho^{(i)}$ as follows. For any $i_1, \ldots, i_6 \in \{1, 2, \ldots, N\}$ and $\boldsymbol{t} \in \mathbb{R}^N$,

$$R_{i_1}(\boldsymbol{t}) = 2t_{i_1}\rho^{(1)}(\|\boldsymbol{t}\|^2), \tag{C.1}$$

$$R_{i_1 i_2}(\boldsymbol{t}) = 2\rho^{(1)}(\|\boldsymbol{t}\|^2)\delta_{i_1,i_2} + 4t_{i_1}t_{i_2}\rho^{(2)}(\|\boldsymbol{t}\|^2), \tag{C.2}$$

$$R_{i_1 i_2 i_3}(\boldsymbol{t}) = 4(t_{i_3}\delta_{i_1,i_2} + t_{i_1}\delta_{i_2,i_3} + t_{i_2}\delta_{i_1,i_3})\rho^{(2)}(\|\boldsymbol{t}\|^2) + 8t_{i_1}t_{i_2}t_{i_3}\rho^{(3)}(\|\boldsymbol{t}\|^2), \tag{C.3}$$

$$\begin{aligned}
&R_{i_1 i_2 i_3 i_4}(\boldsymbol{t}) \\
&= 4\left(\delta_{i_1,i_2}\delta_{i_3,i_4} + \delta_{i_2,i_3}\delta_{i_1,i_4} + \delta_{i_1,i_3}\delta_{i_2,i_4}\right)\rho^{(2)}(\|\boldsymbol{t}\|^2) \\
&\quad + 8\left(t_{i_3}t_{i_4}\delta_{i_1,i_2} + t_{i_1}t_{i_4}\delta_{i_2,i_3} + t_{i_2}t_{i_4}\delta_{i_1,i_3} + t_{i_2}t_{i_3}\delta_{i_1,i_4} + t_{i_1}t_{i_3}\delta_{i_2,i_4} + t_{i_1}t_{i_2}\delta_{i_3,i_4}\right)\rho^{(3)}(\|\boldsymbol{t}\|^2) \\
&\quad + 16t_{i_1}t_{i_2}t_{i_3}t_{i_4}\rho^{(4)}(\|\boldsymbol{t}\|^2),
\end{aligned} \tag{C.4}$$

and

$$R_{i_1 i_2 i_3 i_4 i_5 i_6}(\boldsymbol{t}) = \rho^{(3)}(\|\boldsymbol{t}\|^2) \sum_{(p_1 \dots, p_5) \in \Pi_5} \delta_{i_{p_1}, i_6} \delta_{i_{p_2}, i_{p_3}} \delta_{i_{p_4}, i_{p_5}} + f(\boldsymbol{t}), \tag{C.5}$$

where $f(\boldsymbol{t})$ is a real-valued function of $\boldsymbol{t} \in \mathbb{R}^N$ such that $f(\boldsymbol{0}) = 0$.

C.2 Discussion on the Conditions in Definition 4.3.2

By Appendix A.1, for any $z \in \mathbb{R}$ and any compact set $T \subset \mathbb{R}^N \setminus \{\boldsymbol{0}_N\}$ with $\lambda_{N-1}(\partial T) < \infty$, $\{(X(\boldsymbol{t})|\nabla X(\boldsymbol{0}) = \boldsymbol{0}_N, X(\boldsymbol{0}) = z), \boldsymbol{t} \in T\}$ is a Gaussian random field. In this section, we will explain why the conditions in Definition 4.3.2 allow us to apply Lemma 2.1.10 to this Gaussian random field when $\rho(x)$ is four times continuously differentiable on T.

In the remainder of this section, fix a compact set $T \subset \mathbb{R}^N \setminus \{\boldsymbol{0}_N\}$ with $\lambda_{N-1}(\partial T) < \infty$. It suffices to show

$$\max_{1 \le i,j \le N} |C_{ij}(\boldsymbol{t}, \boldsymbol{t}) + C_{ij}(\boldsymbol{s}, \boldsymbol{s}) - 2C_{ij}(\boldsymbol{s}, \boldsymbol{t})| \le K |\log(\|\boldsymbol{t} - \boldsymbol{s}\|)|^{-(1+\alpha)}$$

for some finite $K > 0$, some $\alpha > 0$, and all $\boldsymbol{s}, \boldsymbol{t} \in T$ such that $\|\boldsymbol{t} - \boldsymbol{s}\|$ is small enough, where for any $i, j \in \{1, \dots, N\}$ and $\boldsymbol{s}, \boldsymbol{t} \in T$,

$$C_{ij}(\boldsymbol{s}, \boldsymbol{t}) := \mathrm{Cov}\left[X_{ij}(\boldsymbol{s}), X_{ij}(\boldsymbol{t}) | \nabla X(\boldsymbol{0}) = \boldsymbol{0}_N, X(\boldsymbol{0}) = z\right].$$

Since $x \le (-\log(x))^{-2}$ for any $0 < x < 1$, we have $\|\boldsymbol{t}\| < (-\log(\|\boldsymbol{t}\|))^{-2}$ when $0 < \|\boldsymbol{t}\| < 1$. Thus, it suffices to show that there exist constants $\beta > 0$ and $0 < \varepsilon < 1$ such that for any $\boldsymbol{s}, \boldsymbol{t} \in T$ satisfying $\|\boldsymbol{t} - \boldsymbol{s}\| < \varepsilon$,

$$\max_{1 \le i,j \le N} |C_{ij}(\boldsymbol{t}, \boldsymbol{t}) + C_{ij}(\boldsymbol{s}, \boldsymbol{s}) - 2C_{ij}(\boldsymbol{s}, \boldsymbol{t})| \le \beta \|\boldsymbol{t} - \boldsymbol{s}\|.$$

Fix $\boldsymbol{s}, \boldsymbol{t} \in T$, $z \in \mathbb{R}$, and $1 \le i, j \le N$. Let $\boldsymbol{Y}_1 := (X_{ij}(\boldsymbol{s}), X_{ij}(\boldsymbol{t}))^T$ and $\boldsymbol{Y}_2 := (\nabla X(\boldsymbol{0}), X(\boldsymbol{0}))^T$. By Lemma 2.1.3, $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ are both centered and Gaussian. Consider Gaussian 2-vector $(\boldsymbol{Y}_1 | \boldsymbol{Y}_2 = (\boldsymbol{0}_N, z)^T)$. By (A.7), we have

$$\boldsymbol{\Sigma}_{1|2} = \boldsymbol{\Sigma}_{11} - \boldsymbol{\Sigma}_{12} \boldsymbol{\Sigma}_{22}^{-1} \boldsymbol{\Sigma}_{12}^T, \tag{C.6}$$

where $\boldsymbol{\Sigma}_{1|2}$ is the covariance matrix of $(\boldsymbol{Y}_1 | \boldsymbol{Y}_2 = (\boldsymbol{0}_N, z)^T)$, $\boldsymbol{\Sigma}_{11}$ is the covariance matrix of $\boldsymbol{Y}_1$, $\boldsymbol{\Sigma}_{22}$ is the covariance matrix of $\boldsymbol{Y}_2$, and $\boldsymbol{\Sigma}_{12}$ is the cross-covariance matrix of $\boldsymbol{Y}_1$ and $\boldsymbol{Y}_2$ (i.e., $\boldsymbol{\Sigma}_{12} = \mathbb{E}[\boldsymbol{Y}_1 \boldsymbol{Y}_2^T]$). By (2.7) and (2.8), we have

$$\boldsymbol{\Sigma}_{11} = \begin{pmatrix} R_{ijij}(\boldsymbol{0}) & R_{ijij}(\boldsymbol{t} - \boldsymbol{s}) \\ R_{ijij}(\boldsymbol{t} - \boldsymbol{s}) & R_{ijij}(\boldsymbol{0}) \end{pmatrix}, \quad \boldsymbol{\Sigma}_{22}^{-1} = \begin{pmatrix} -\frac{1}{2\rho^{(1)}(0)} \boldsymbol{I}_N & \boldsymbol{0}_{N \times 1} \\ \boldsymbol{0}_{1 \times N} & 1 \end{pmatrix},$$

and

$$\boldsymbol{\Sigma}_{12} = \begin{pmatrix} -R_{ij1}(\boldsymbol{s}) & \cdots & -R_{ijN}(\boldsymbol{s}) & R_{ij}(\boldsymbol{s}) \\ -R_{ij1}(\boldsymbol{t}) & \cdots & -R_{ijN}(\boldsymbol{t}) & R_{ij}(\boldsymbol{t}) \end{pmatrix},$$

where $R_{i_1 \dots i_p}$, $p = 2, 3, 4$ and $1 \leq i_1, \dots, i_p \leq N$ are the same as in Section C.1 (one should note that by Condition (2) in Definition 4.3.2 and (i) of Lemma 2.1.4), $R_{i_1 i_2 i_3 i_4}(\boldsymbol{t})$, $1 \leq i_1 \dots i_4 \leq N$ exist for any $\boldsymbol{t} \in \mathbb{R}^N$). Then,

$$\boldsymbol{\Sigma}_{12}\boldsymbol{\Sigma}_{22}^{-1}\boldsymbol{\Sigma}_{12}^T = \begin{pmatrix} a & b \\ b & c \end{pmatrix}$$

where

$$a := R_{ij}(\boldsymbol{s})^2 - \frac{1}{2\rho^{(1)}(0)}\left(\sum_{k=1}^N R_{ijk}(\boldsymbol{s})^2\right), \quad c := R_{ij}(\boldsymbol{t})^2 - \frac{1}{2\rho^{(1)}(0)}\left(\sum_{k=1}^N R_{ijk}(\boldsymbol{t})^2\right)$$

and

$$b := R_{ij}(\boldsymbol{s})R_{ij}(\boldsymbol{t}) - \frac{1}{2\rho^{(1)}(0)}\left(\sum_{k=1}^N R_{ijk}(\boldsymbol{s})R_{ijk}(\boldsymbol{t})\right).$$

By Condition (3) in Definition 4.3.2, $\rho^{(5)}(x)$ is bounded on $x \in [0, \delta_\rho^2]$. Then there exists a finite constant $C_{T,1} > 0$ such that for any $\boldsymbol{s}, \boldsymbol{t} \in T$ satisfying $\|\boldsymbol{t} - \boldsymbol{s}\| \leq \min(\delta_\rho, 1)$ and any $1 \leq i, j \leq N$,

$$2(R_{ijij}(\boldsymbol{0}) - R_{ijij}(\boldsymbol{t} - \boldsymbol{s})) \leq C_{T,1}\|\boldsymbol{t} - \boldsymbol{s}\|.$$

Since $R_{ijk\ell}(\boldsymbol{t})$, $1 \leq i, j, k, \ell \leq N$ are all continuous on $\boldsymbol{t} \in T$ and T is compact, there exists a finite constant $C_{T,2} > 0$ such that for any $\boldsymbol{s}, \boldsymbol{t} \in T$ and $1 \leq i, j, k \leq N$,

$$|R_{ij}(\boldsymbol{s}) - R_{ij}(\boldsymbol{t})| \leq C_{T,2}\|\boldsymbol{t} - \boldsymbol{s}\| \text{ and } |R_{ijk}(\boldsymbol{s}) - R_{ijk}(\boldsymbol{t})| \leq C_{T,2}\|\boldsymbol{t} - \boldsymbol{s}\|.$$

Note that by (4.1), $\rho^{(1)}(0) < 0$. Then

$$\begin{aligned} a + c - 2b &= (R_{ij}(\boldsymbol{s}) - R_{ij}(\boldsymbol{t}))^2 - \frac{1}{2\rho^{(1)}(0)}\left(\sum_{k=1}^N (R_{ijk}(\boldsymbol{s}) - R_{ijk}(\boldsymbol{t}))^2\right) \\ &\leq \left(1 - \frac{N}{2\rho^{(1)}(0)}\right) C_{T,2}^2 \|\boldsymbol{t} - \boldsymbol{s}\|^2. \end{aligned}$$

Finally, for any $\boldsymbol{s}, \boldsymbol{t} \in \mathbb{R}^N$ satisfying $\|\boldsymbol{t}-\boldsymbol{s}\| \le \min(\delta_\rho, 1)$, we have

$$\begin{aligned}
&\max_{1\le i,j\le N} |C_{ij}(\boldsymbol{t},\boldsymbol{t}) + C_{ij}(\boldsymbol{s},\boldsymbol{s}) - 2C_{ij}(\boldsymbol{s},\boldsymbol{t})| \\
&\le C_{T,1}\|\boldsymbol{t}-\boldsymbol{s}\| + \left(1 - \frac{N}{2\rho^{(1)}(0)}\right) C_{T,2}^2 \|\boldsymbol{t}-\boldsymbol{s}\|^2 \\
&\le \left(C_{T,1} + \left(1 - \frac{N}{2\rho^{(1)}(0)}\right) C_{T,2}^2\right) \|\boldsymbol{t}-\boldsymbol{s}\|,
\end{aligned}$$

and hence proved.

C.3 Proof of Lemma 4.3.5

Fix $N \ge 2$. For any $r > 0$ and $\boldsymbol{t} \in \mathbb{R}^N$, let $B(\boldsymbol{t}, r)$ be the N-dimensional open ball centered at $\boldsymbol{t}$ with radius r. Since $\rho(x)$ is four times continuously differentiable on $x \in [0, \delta_\rho^2]$ (see δ_ρ in Definition 4.3.2), all the partial derivatives of $R(\boldsymbol{t})$ up to fourth order are continuous on $\boldsymbol{t} \in B(\boldsymbol{0}_N, \delta_\rho)$. Recall that $L = N(N+1)/2+2$. For convenience, we will not distinguish the $N \times N$ matrix $\nabla^2 X$ from its usual vectorization in notations, but one can easily distinguish them from a given context. For any $\boldsymbol{t} \in B(\boldsymbol{0}_N, \delta_\rho) \setminus \{\boldsymbol{0}_N\}$, denote

1. $\boldsymbol{V}_{11}(\boldsymbol{t}) \in \mathbb{R}^{L\times L}$: the covariance matrix of the random L-vector $(\nabla^2 X(\boldsymbol{t}), X(\boldsymbol{t}), X(\boldsymbol{0}))$;
2. $\boldsymbol{V}_{12}(\boldsymbol{t}) \in \mathbb{R}^{L\times 2N}$: the covariance matrix between random vectors
$$(\nabla^2 X(\boldsymbol{t}), X(\boldsymbol{t}), X(\boldsymbol{0})) \quad \text{and} \quad (\nabla X(\boldsymbol{t}), \nabla X(\boldsymbol{0}))\,;$$
3. $\boldsymbol{V}_{22}(\boldsymbol{t}) \in \mathbb{R}^{2N\times 2N}$: the covariance matrix of the random $(2N)$-vector
$$(\nabla X(\boldsymbol{t}), \nabla X(\boldsymbol{0}))\,.$$

Since $\boldsymbol{\Sigma}(\boldsymbol{t})$, $\boldsymbol{t} \in \mathbb{R}^N \setminus \{\boldsymbol{0}_N\}$ is the covariance matrix of the random L-vector

$$\left(\nabla^2 X(\boldsymbol{t}), X(\boldsymbol{t}), X(\boldsymbol{0}) | \nabla X(\boldsymbol{t}) = \nabla X(\boldsymbol{0}) = \boldsymbol{0}_N\right),$$

by (A.7), we have for any $\boldsymbol{t} \in B(\boldsymbol{0}_N, \delta_\rho) \setminus \{\boldsymbol{0}_N\}$,

$$\boldsymbol{\Sigma}(\boldsymbol{t}) = \boldsymbol{V}_{11}(\boldsymbol{t}) - \boldsymbol{V}_{12}(\boldsymbol{t})\boldsymbol{V}_{22}^{-1}(\boldsymbol{t})\boldsymbol{V}_{12}^T(\boldsymbol{t}).$$

C.3.1 The Blocked Covariance Matrix

Let $N \geq 2$. For convenience, we adopt the following notations for $\boldsymbol{t} \in B(\mathbf{0}_N, \delta_\rho) \setminus \{\mathbf{0}_N\}$:

1. $\boldsymbol{G}_{00}(\boldsymbol{t}) := \mathrm{Cov}\left[X(\mathbf{0}), X(\boldsymbol{t})\right] = R(\boldsymbol{t})$;
2. $\boldsymbol{G}_{01}(\boldsymbol{t}) := \mathrm{Cov}\left[X(\mathbf{0}), \nabla X(\boldsymbol{t})\right] = -\mathrm{Cov}\left[\nabla X(\mathbf{0}), X(\boldsymbol{t})\right] = (R_1(\boldsymbol{t}), \ldots, R_N(\boldsymbol{t}))$;
3. $\boldsymbol{G}_{20}(\boldsymbol{t}) := \mathrm{Cov}\left[\nabla^2 X(\mathbf{0}), X(\boldsymbol{t})\right] = \mathrm{Cov}\left[X(\mathbf{0}), \nabla^2 X(\boldsymbol{t})\right] \in \mathbb{R}^{(L-2)\times 1}$;
4. $\boldsymbol{G}_{21}(\boldsymbol{t}) := \mathrm{Cov}\left[\nabla^2 X(\mathbf{0}), \nabla X(\boldsymbol{t})\right] = -\mathrm{Cov}\left[\nabla X(\mathbf{0}), \nabla^2 X(\boldsymbol{t})\right] \in \mathbb{R}^{(L-2)\times N}$;
5. $\boldsymbol{G}_{22}(\boldsymbol{t}) := \mathrm{Cov}\left[\nabla^2 X(\mathbf{0}), \nabla^2 X(\boldsymbol{t})\right] \in \mathbb{R}^{(L-2)\times(L-2)}$.

Their relationships with the covariance function R are given by Lemma 2.1.4. In particular, by (C.1)-(C.4), we have

1. $\boldsymbol{G}_{00}(\mathbf{0}) := \mathrm{Cov}\left[X(\mathbf{0}), X(\mathbf{0})\right] = R(\mathbf{0})$;
2. $\boldsymbol{G}_{01}(\mathbf{0}) := \mathrm{Cov}\left[X(\mathbf{0}), \nabla X(\mathbf{0})\right] = (R_1(\mathbf{0}), \ldots, R_N(\mathbf{0})) = \mathbf{0}_{1\times N}$;
3. $\boldsymbol{G}_{20}(\mathbf{0}) := \mathrm{Cov}\left[\nabla^2 X(\mathbf{0}), X(\mathbf{0})\right] \in \mathbb{R}^{(L-2)\times 1}$;
4. $\boldsymbol{G}_{21}(\mathbf{0}) := \mathrm{Cov}\left[\nabla^2 X(\mathbf{0}), \nabla X(\mathbf{0})\right] = \mathbf{0}_{(L-2)\times N}$;
5. $\boldsymbol{G}_{22}(\mathbf{0}) := \mathrm{Cov}\left[\nabla^2 X(\mathbf{0}), \nabla^2 X(\mathbf{0})\right] \in \mathbb{R}^{(L-2)\times(L-2)}$.

Immediately, we have

$$\boldsymbol{V}_{11}(\boldsymbol{t}) = \begin{pmatrix} \boldsymbol{G}_{22}(\mathbf{0}) & \boldsymbol{G}_{20}(\mathbf{0}) & \boldsymbol{G}_{20}(\boldsymbol{t}) \\ \boldsymbol{G}_{20}^T(\mathbf{0}) & \boldsymbol{G}_{00}(\mathbf{0}) & \boldsymbol{G}_{00}(\boldsymbol{t}) \\ \boldsymbol{G}_{20}^T(\boldsymbol{t}) & \boldsymbol{G}_{00}^T(\boldsymbol{t}) & \boldsymbol{G}_{00}(\mathbf{0}) \end{pmatrix} \tag{C.7}$$

and

$$\boldsymbol{V}_{12}(\boldsymbol{t}) = \begin{pmatrix} -\boldsymbol{G}_{21}(\mathbf{0}) & -\boldsymbol{G}_{21}(\boldsymbol{t}) \\ -\boldsymbol{G}_{01}(\mathbf{0}) & -\boldsymbol{G}_{01}(\boldsymbol{t}) \\ \boldsymbol{G}_{01}(\boldsymbol{t}) & -\boldsymbol{G}_{01}(\mathbf{0}) \end{pmatrix} = \begin{pmatrix} \mathbf{0}_{(L-2)\times N} & -\boldsymbol{G}_{21}(\boldsymbol{t}) \\ \mathbf{0}_{1\times N} & -\boldsymbol{G}_{01}(\boldsymbol{t}) \\ \boldsymbol{G}_{01}(\boldsymbol{t}) & \mathbf{0}_{1\times N} \end{pmatrix}. \tag{C.8}$$

As for $\boldsymbol{V}_{22}^{-1}(\boldsymbol{t})$, by (C.2), we have

$$\mathrm{Cov}\left[\nabla X(\mathbf{0}), \nabla X(\mathbf{0})\right] = -2\rho^{(1)}(0)\boldsymbol{I}_N$$

and

$$\mathrm{Cov}\left[\nabla X(\boldsymbol{t}), \nabla X(\boldsymbol{0})\right] = -2\rho^{(1)}(\|\boldsymbol{t}\|^2)\boldsymbol{I}_N - 4\rho^{(2)}(\|\boldsymbol{t}\|^2)\boldsymbol{t}\boldsymbol{t}^T.$$

Since $\rho^{(1)}(0) < 0$ (Remark 4.3.4), we can define

$$k_1(\boldsymbol{t}) := \frac{\rho^{(1)}(\|\boldsymbol{t}\|^2)}{\rho^{(1)}(0)} \text{ and } k_2(\boldsymbol{t}) := \frac{2\rho^{(2)}(\|\boldsymbol{t}\|^2)}{\rho^{(1)}(0)},$$

for any $\boldsymbol{t} \in B(\boldsymbol{0}_N, \delta_\rho) \setminus \{\boldsymbol{0}_N\}$. Thus,

$$\boldsymbol{V}_{22}(\boldsymbol{t}) = -2\rho^{(1)}(0) \begin{pmatrix} \boldsymbol{I}_N & k_1(\boldsymbol{t})\boldsymbol{I}_N + k_2(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T \\ k_1(\boldsymbol{t})\boldsymbol{I}_N + k_2(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T & \boldsymbol{I}_N \end{pmatrix}$$

and then

$$\boldsymbol{V}_{22}^{-1}(\boldsymbol{t}) = -\frac{1}{2\rho^{(1)}(0)} \begin{pmatrix} \boldsymbol{I}_N & k_1(\boldsymbol{t})\boldsymbol{I}_N + k_2(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T \\ k_1(\boldsymbol{t})\boldsymbol{I}_N + k_2(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T & \boldsymbol{I}_N \end{pmatrix}^{-1}.$$

To further calculate $\boldsymbol{V}_{22}^{-1}(\boldsymbol{t})$, we need the following two facts which can be easily checked:

1. for any symmetric matrix $\boldsymbol{B} \in \mathbb{R}^{N\times N}$ such that $\boldsymbol{I}_N - \boldsymbol{B}^2$ is invertible,

$$\begin{pmatrix} \boldsymbol{I}_N & \boldsymbol{B} \\ \boldsymbol{B} & \boldsymbol{I}_N \end{pmatrix}^{-1} = \begin{pmatrix} (\boldsymbol{I}_N - \boldsymbol{B}^2)^{-1} & -\boldsymbol{B}(\boldsymbol{I}_N - \boldsymbol{B}^2)^{-1} \\ -(\boldsymbol{I}_N - \boldsymbol{B}^2)^{-1}\boldsymbol{B} & (\boldsymbol{I}_N - \boldsymbol{B}^2)^{-1} \end{pmatrix};$$

2. (The Sherman–Morrison Formula) for any $\boldsymbol{w}, \boldsymbol{v} \in \mathbb{R}^N$ such that $1 + \boldsymbol{v}^T\boldsymbol{w} \neq 0$,

$$(\boldsymbol{I}_N + \boldsymbol{w}\boldsymbol{v}^T)^{-1} = \boldsymbol{I}_N - \frac{\boldsymbol{w}\boldsymbol{v}^T}{1 + \boldsymbol{v}^T\boldsymbol{w}}.$$

By taking $\boldsymbol{B} = k_1(\boldsymbol{t})\boldsymbol{I}_N + k_2(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T$ in Fact 1, we get

$$\boldsymbol{I}_N - \boldsymbol{B}^2 = \left(1 - k_1^2(\boldsymbol{t})\right)\left(\boldsymbol{I}_N - k_3(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\right), \tag{C.9}$$

where by letting $k_*(\boldsymbol{t}) := k_1(\boldsymbol{t}) + k_2(\boldsymbol{t})\|\boldsymbol{t}\|^2$ for any $\boldsymbol{t} \in B(\boldsymbol{0}_N, \delta_\rho) \setminus \{\boldsymbol{0}_N\}$,

$$k_3(\boldsymbol{t}) := \frac{2k_1(\boldsymbol{t})k_2(\boldsymbol{t}) + k_2^2(\boldsymbol{t})\|\boldsymbol{t}\|^2}{1 - k_1^2(\boldsymbol{t})} = \frac{k_2(\boldsymbol{t})(k_1(\boldsymbol{t}) + k_*(\boldsymbol{t}))}{1 - k_1^2(\boldsymbol{t})}.$$

By (4.4),

$$1 - k_1^2(\boldsymbol{t}) = \frac{(\rho^{(1)}(0))^2 - (\rho^{(1)}(\|\boldsymbol{t}\|^2))^2}{(\rho^{(1)}(0))^2} > 0 \text{ for any } \boldsymbol{t} \in B(\boldsymbol{0}_N, \delta_\rho) \setminus \{\boldsymbol{0}_N\}. \tag{C.10}$$

Thus, $k_3(\boldsymbol{t})$ is well-defined for any $\boldsymbol{t} \in B(\mathbf{0}_N, \delta_\rho) \setminus \{\mathbf{0}_N\}$. To apply Fact 2 to the right-hand side of (C.9), we still need to show $1 - k_3(\boldsymbol{t})\|\boldsymbol{t}\|^2 \neq 0$ for any $\boldsymbol{t} \in B(\mathbf{0}_N, \delta_\rho) \setminus \{\mathbf{0}_N\}$. One can easily check that

$$\left(1 - k_3(\boldsymbol{t})\|\boldsymbol{t}\|^2\right)\left(1 - k_1^2(\boldsymbol{t})\right) = 1 - k_*^2(\boldsymbol{t}).$$

Taking $\boldsymbol{t}' := (t_1', t_2', \ldots, t_N') = (0, \ldots, 0, \|\boldsymbol{t}\|)$ and by (4.3), we have

$$|k_*(\boldsymbol{t})| = |k_*(\boldsymbol{t}')| = \left|k_1(\boldsymbol{t}') + \|\boldsymbol{t}\|^2 k_2(\boldsymbol{t}')\right| = \left|\frac{\rho^{(1)}(\|\boldsymbol{t}'\|^2) + 2t_N'^2 \rho^{(2)}(\|\boldsymbol{t}'\|^2)}{\rho^{(1)}(0)}\right| < 1.$$

Thus, for any $\boldsymbol{t} \in B(\mathbf{0}_N, \delta_\rho) \setminus \{\mathbf{0}_N\}$,

$$1 - k_3(\boldsymbol{t})\|\boldsymbol{t}\|^2 = \frac{1 - k_*^2(\boldsymbol{t})}{1 - k_1^2(\boldsymbol{t})} > 0. \tag{C.11}$$

Then by Fact 2, we have

$$\begin{aligned}(\boldsymbol{I}_N - \boldsymbol{B}^2)^{-1} &= \frac{1}{1 - k_1^2(\boldsymbol{t})}\left(\boldsymbol{I}_N + \frac{k_3(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T}{1 - k_3(\boldsymbol{t})\|\boldsymbol{t}\|^2}\right) \\ &= \frac{1}{1 - k_1^2(\boldsymbol{t})}\left(\boldsymbol{I}_N + k_4(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\right),\end{aligned}$$

where

$$k_4(\boldsymbol{t}) := \frac{k_3(\boldsymbol{t})}{1 - k_3(\boldsymbol{t})\|\boldsymbol{t}\|^2} = \frac{\left(1 - k_1^2(\boldsymbol{t})\right)k_3(\boldsymbol{t})}{1 - k_*^2(\boldsymbol{t})} = \frac{k_2(\boldsymbol{t})(k_1(\boldsymbol{t}) + k_*(\boldsymbol{t}))}{1 - k_*^2(\boldsymbol{t})}.$$

Then

$$\begin{aligned}&(\boldsymbol{I}_N - \boldsymbol{B}^2)^{-1}\boldsymbol{B} = \boldsymbol{B}(\boldsymbol{I}_N - \boldsymbol{B}^2)^{-1} \\ &= \frac{1}{1 - k_1^2(\boldsymbol{t})}\left(\boldsymbol{I}_N + k_4(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\right)\left(k_1(\boldsymbol{t})\boldsymbol{I}_N + k_2(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\right) \\ &= \frac{1}{1 - k_1^2(\boldsymbol{t})}\left(k_1(\boldsymbol{t})\boldsymbol{I}_N + \left(k_2(\boldsymbol{t}) + k_1(\boldsymbol{t})k_4(\boldsymbol{t}) + k_2(\boldsymbol{t})k_4(\boldsymbol{t})\|\boldsymbol{t}\|^2\right)\boldsymbol{t}\boldsymbol{t}^T\right) \\ &= \frac{1}{1 - k_1^2(\boldsymbol{t})}\left(k_1(\boldsymbol{t})\boldsymbol{I}_N + k_5(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\right),\end{aligned}$$

where

$$\begin{aligned}k_5(\boldsymbol{t}) :&= k_2(\boldsymbol{t}) + k_1(\boldsymbol{t})k_4(\boldsymbol{t}) + k_2(\boldsymbol{t})k_4(\boldsymbol{t})\|\boldsymbol{t}\|^2 \\ &= k_2(\boldsymbol{t}) + k_*(\boldsymbol{t})k_4(\boldsymbol{t}) \\ &= \frac{k_2(\boldsymbol{t})(1 + k_1(\boldsymbol{t})k_*(\boldsymbol{t}))}{1 - k_*^2(\boldsymbol{t})}.\end{aligned}$$

It is easy to see $k_i(\boldsymbol{t})$, $i = 1, 2, \ldots, 5$ are all well-defined for any $\boldsymbol{t} \in B(\mathbf{0}_N, \delta_\rho) \setminus \{\mathbf{0}_N\}$ (see δ_ρ in Definition 4.3.2). In summary, for any $\boldsymbol{t} \in B(\mathbf{0}_N, \delta_\rho) \setminus \{\mathbf{0}_N\}$, we have

$$\begin{aligned}
&-2\rho^{(1)}(0)\boldsymbol{V}_{22}^{-1}(\boldsymbol{t}) \\
&= \begin{pmatrix} \frac{1}{1-k_1^2(\boldsymbol{t})}\boldsymbol{I}_N & -\frac{k_1(\boldsymbol{t})}{1-k_1^2(\boldsymbol{t})}\boldsymbol{I}_N \\ -\frac{k_1(\boldsymbol{t})}{1-k_1^2(\boldsymbol{t})}\boldsymbol{I}_N & \frac{1}{1-k_1^2(\boldsymbol{t})}\boldsymbol{I}_N \end{pmatrix} + \begin{pmatrix} \frac{k_4(\boldsymbol{t})}{1-k_1^2(\boldsymbol{t})}\boldsymbol{t}\boldsymbol{t}^T & -\frac{k_5(\boldsymbol{t})}{1-k_1^2(\boldsymbol{t})}\boldsymbol{t}\boldsymbol{t}^T \\ -\frac{k_5(\boldsymbol{t})}{1-k_1^2(\boldsymbol{t})}\boldsymbol{t}\boldsymbol{t}^T & \frac{k_4(\boldsymbol{t})}{1-k_1^2(\boldsymbol{t})}\boldsymbol{t}\boldsymbol{t}^T \end{pmatrix} \\
&= \frac{1}{1-k_1^2(\boldsymbol{t})}\begin{pmatrix} \boldsymbol{I}_N & -k_1(\boldsymbol{t})\boldsymbol{I}_N \\ -k_1(\boldsymbol{t})\boldsymbol{I}_N & \boldsymbol{I}_N \end{pmatrix} + \frac{1}{1-k_1^2(\boldsymbol{t})}\begin{pmatrix} k_4(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T & -k_5(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T \\ -k_5(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T & k_4(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T \end{pmatrix},
\end{aligned} \tag{C.12}$$

where

$$k_1(\boldsymbol{t}) = \frac{\rho^{(1)}(\|\boldsymbol{t}\|^2)}{\rho^{(1)}(0)}, \quad k_2(\boldsymbol{t}) = \frac{2\rho^{(2)}(\|\boldsymbol{t}\|^2)}{\rho^{(1)}(0)}, \quad k_*(\boldsymbol{t}) = k_1(\boldsymbol{t}) + k_2(\boldsymbol{t})\|\boldsymbol{t}\|^2,$$

$$k_4(\boldsymbol{t}) = \frac{k_2(\boldsymbol{t})(k_1(\boldsymbol{t}) + k_*(\boldsymbol{t}))}{1-k_*^2(\boldsymbol{t})} \text{ and } k_5(\boldsymbol{t}) = \frac{k_2(\boldsymbol{t})(1 + k_1(\boldsymbol{t})k_*(\boldsymbol{t}))}{1-k_*^2(\boldsymbol{t})}.$$

Here one should note that

$$\begin{aligned}
k_5(\boldsymbol{t}) - k_4(\boldsymbol{t}) &= \frac{k_2(\boldsymbol{t})(1-k_1(\boldsymbol{t}))(1-k_*(\boldsymbol{t}))}{1-k_*^2(\boldsymbol{t})} \\
&= \frac{k_2(\boldsymbol{t})(1-k_1(\boldsymbol{t}))}{1+k_*(\boldsymbol{t})} \\
&\to 0 \text{ as } \|\boldsymbol{t}\| \to 0.
\end{aligned} \tag{C.13}$$

In consequence, by (C.7), (C.8) and (C.12),

$$\begin{aligned}
&-2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))\boldsymbol{V}_{12}(\boldsymbol{t})\boldsymbol{V}_{22}^{-1}(\boldsymbol{t})\boldsymbol{V}_{12}^T(\boldsymbol{t}) \\
&= \begin{pmatrix} \mathbf{0}_{(L-2)\times N} & -\boldsymbol{G}_{21}(\boldsymbol{t}) \\ \mathbf{0}_{1\times N} & -\boldsymbol{G}_{01}(\boldsymbol{t}) \\ \boldsymbol{G}_{01}(\boldsymbol{t}) & \mathbf{0}_{1\times N} \end{pmatrix} \begin{pmatrix} \boldsymbol{I}_N & -k_1(\boldsymbol{t})\boldsymbol{I}_N \\ -k_1(\boldsymbol{t})\boldsymbol{I}_N & \boldsymbol{I}_N \end{pmatrix} \begin{pmatrix} \mathbf{0}_{N\times(L-2)} & \mathbf{0}_{N\times 1} & \boldsymbol{G}_{01}^T(\boldsymbol{t}) \\ -\boldsymbol{G}_{21}^T(\boldsymbol{t}) & -\boldsymbol{G}_{01}^T(\boldsymbol{t}) & \mathbf{0}_{N\times 1} \end{pmatrix} \\
&\quad + \begin{pmatrix} \mathbf{0}_{(L-2)\times N} & -\boldsymbol{G}_{21}(\boldsymbol{t}) \\ \mathbf{0}_{1\times N} & -\boldsymbol{G}_{01}(\boldsymbol{t}) \\ \boldsymbol{G}_{01}(\boldsymbol{t}) & \mathbf{0}_{1\times N} \end{pmatrix} \begin{pmatrix} k_4(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T & -k_5(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T \\ -k_5(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T & k_4(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T \end{pmatrix} \begin{pmatrix} \mathbf{0}_{N\times(L-2)} & \mathbf{0}_{N\times 1} & \boldsymbol{G}_{01}^T(\boldsymbol{t}) \\ -\boldsymbol{G}_{21}^T(\boldsymbol{t}) & -\boldsymbol{G}_{01}^T(\boldsymbol{t}) & \mathbf{0}_{N\times 1} \end{pmatrix} \\
&= \begin{pmatrix} \boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{21}^T(\boldsymbol{t}) & \boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) & k_1(\boldsymbol{t})\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) \\ \boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{21}^T(\boldsymbol{t}) & \boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) & k_1(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) \\ k_1(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{21}^T(\boldsymbol{t}) & k_1(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) & \boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) \end{pmatrix} \\
&\quad + \begin{pmatrix} k_4(\boldsymbol{t})\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{21}^T(\boldsymbol{t}) & k_4(\boldsymbol{t})\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) & k_5(\boldsymbol{t})\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) \\ k_4(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{21}^T(\boldsymbol{t}) & k_4(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) & k_5(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) \\ k_5(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{21}^T(\boldsymbol{t}) & k_5(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) & k_4(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) \end{pmatrix},
\end{aligned}$$

and then the blocked version of $\boldsymbol{\Sigma}(\boldsymbol{t})$ is given by

$$\boldsymbol{\Sigma}(\boldsymbol{t}) = \boldsymbol{V}_{11}(\boldsymbol{t}) - \boldsymbol{V}_{12}(\boldsymbol{t})\boldsymbol{V}_{22}^{-1}(\boldsymbol{t})\boldsymbol{V}_{12}^T(\boldsymbol{t}) = \boldsymbol{M}_0(\boldsymbol{t}) + \frac{1}{2\rho^{(1)}(0)\left(1 - k_1^2(\boldsymbol{t})\right)}(\boldsymbol{M}_1(\boldsymbol{t}) + \boldsymbol{M}_2(\boldsymbol{t})), \tag{C.14}$$

where

$$\boldsymbol{M}_0(\boldsymbol{t}) := \begin{pmatrix} \boldsymbol{G}_{22}(\boldsymbol{0}) & \boldsymbol{G}_{20}(\boldsymbol{0}) & \boldsymbol{G}_{20}(\boldsymbol{t}) \\ \boldsymbol{G}_{20}^T(\boldsymbol{0}) & \boldsymbol{G}_{00}(\boldsymbol{0}) & \boldsymbol{G}_{00}(\boldsymbol{t}) \\ \boldsymbol{G}_{20}^T(\boldsymbol{t}) & \boldsymbol{G}_{00}^T(\boldsymbol{t}) & \boldsymbol{G}_{00}(\boldsymbol{0}) \end{pmatrix},$$

$$\boldsymbol{M}_1(\boldsymbol{t}) := \begin{pmatrix} \boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{21}^T(\boldsymbol{t}) & \boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) & k_1(\boldsymbol{t})\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) \\ \boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{21}^T(\boldsymbol{t}) & \boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) & k_1(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) \\ k_1(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{21}^T(\boldsymbol{t}) & k_1(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) & \boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) \end{pmatrix}$$

and

$$\boldsymbol{M}_2(\boldsymbol{t}) := \begin{pmatrix} k_4(\boldsymbol{t})\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{21}^T(\boldsymbol{t}) & k_4(\boldsymbol{t})\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) & k_5(\boldsymbol{t})\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) \\ k_4(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{21}^T(\boldsymbol{t}) & k_4(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) & k_5(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) \\ k_5(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{21}^T(\boldsymbol{t}) & k_5(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) & k_4(\boldsymbol{t})\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) \end{pmatrix}.$$

Fix a direction $\boldsymbol{u} \in \mathbb{S}^{N-1}$, then by (C.14), each element in the covariance matrix $\boldsymbol{\Sigma}(\boldsymbol{u}r)$ is a continuous function of $r \in [0, \delta_\rho]$.

In the following sections, we will calculate the asymptotic expansion of $\boldsymbol{\Sigma}(\boldsymbol{t})$ as $\|\boldsymbol{t}\| \to 0$. To make the following proofs better organized, we will first calculate the asymptotic expansions associated with coefficients $k_i(\boldsymbol{t})$, $i = 1, \ldots, 5$ as $\|\boldsymbol{t}\| \to 0$. Then the calculation will be performed separately for each of the three parts in (C.14):

1. the main part

$$\begin{aligned} &\boldsymbol{\Sigma}(\boldsymbol{t})[1:(L-2), 1:(L-2)] \\ &= \boldsymbol{G}_{22}(\boldsymbol{0}) + \frac{1}{2\rho^{(1)}(0)(1 - k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{21}^T(\boldsymbol{t}) + \frac{k_4(\boldsymbol{t})}{2\rho^{(1)}(0)(1 - k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{21}^T(\boldsymbol{t}); \end{aligned}$$

2. the side part

$$\begin{aligned} &\boldsymbol{\Sigma}(\boldsymbol{t})[1:(L-2), L-1] \\ &= \boldsymbol{G}_{20}(\boldsymbol{0}) + \frac{1}{2\rho^{(1)}(0)(1 - k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) + \frac{k_4(\boldsymbol{t})}{2\rho^{(1)}(0)(1 - k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) \end{aligned}$$

   and

$$\begin{aligned} &\boldsymbol{\Sigma}(\boldsymbol{t})[1:(L-2), L] \\ &= \boldsymbol{G}_{20}(\boldsymbol{t}) + \frac{k_1(\boldsymbol{t})}{2\rho^{(1)}(0)(1 - k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) + \frac{k_5(\boldsymbol{t})}{2\rho^{(1)}(0)(1 - k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}); \end{aligned}$$

3. the corner part

$$\begin{aligned}&\boldsymbol{\Sigma}(\boldsymbol{t})[L-1,L-1]=\boldsymbol{\Sigma}(\boldsymbol{t})[L,L]\\=\ &\boldsymbol{G}_{00}(\mathbf{0})+\frac{1}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t})+\frac{k_4(\boldsymbol{t})}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t})\end{aligned}$$

and

$$\begin{aligned}&\boldsymbol{\Sigma}(\boldsymbol{t})[L-1,L]=\boldsymbol{\Sigma}(\boldsymbol{t})[L,L-1]\\=\ &\boldsymbol{G}_{00}(\boldsymbol{t})+\frac{k_1(\boldsymbol{t})}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t})+\frac{k_5(\boldsymbol{t})}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}).\end{aligned}$$

C.3.2 Asymptotic Expansions of Coefficients

Note that by Condition (3) in Definition 4.3.2, we have

$$\rho(x)=\rho(0)+\rho^{(1)}(0)x+\frac{1}{2}\rho^{(2)}(0)x^2+\frac{1}{6}\rho^{(3)}(0)x^3+o(x^3)\text{ as }x\downarrow 0.$$

In this part, we would like to use the above expansion to expand

(i) $\frac{\|\boldsymbol{t}\|^2}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}$,

(ii) $\frac{k_4(\boldsymbol{t})\|\boldsymbol{t}\|^2}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}$,

(iii) $\frac{k_1(\boldsymbol{t})\|\boldsymbol{t}\|^2}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}$,

(iv) $\frac{k_5(\boldsymbol{t})\|\boldsymbol{t}\|^2}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}$.

For (i), note that

$$\begin{aligned}\left(\rho^{(1)}(0)+\rho^{(1)}(\|\boldsymbol{t}\|^2)\right)^{-1}&=\left(\rho^{(1)}(0)+\rho^{(1)}(0)\right)^{-1}-\left(\rho^{(1)}(0)+\rho^{(1)}(0)\right)^{-2}\rho^{(2)}(0)\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\\&=\left(2\rho^{(1)}(0)\right)^{-1}-\left(2\rho^{(1)}(0)\right)^{-2}\rho^{(2)}(0)\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\end{aligned}$$

and

$$\begin{aligned}\|\boldsymbol{t}\|^2\left(\rho^{(1)}(0)-\rho^{(1)}(\|\boldsymbol{t}\|^2)\right)^{-1}&=-\|\boldsymbol{t}\|^2\left(\rho^{(2)}(0)\|\boldsymbol{t}\|^2+\frac{1}{2}\rho^{(3)}(0)\|\boldsymbol{t}\|^4+o(\|\boldsymbol{t}\|^4)\right)^{-1}\\&=-\left(\rho^{(2)}(0)+\frac{1}{2}\rho^{(3)}(0)\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\right)^{-1}\\&=-\rho^{(2)}(0)^{-1}+\frac{1}{2}\rho^{(2)}(0)^{-2}\rho^{(3)}(0)\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2).\end{aligned}$$

Then we have

$$\begin{aligned}
&\frac{\|\boldsymbol{t}\|^2}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}\\
&=\frac{1}{2}\rho^{(1)}(0)\|\boldsymbol{t}\|^2\left(\rho^{(1)}(0)^2-\rho^{(1)}(\|\boldsymbol{t}\|^2)^2\right)^{-1}\\
&=\frac{1}{2}\rho^{(1)}(0)\left(\rho^{(1)}(0)+\rho^{(1)}(\|\boldsymbol{t}\|^2)\right)^{-1}\|\boldsymbol{t}\|^2\left(\rho^{(1)}(0)-\rho^{(1)}(\|\boldsymbol{t}\|^2)\right)^{-1}\\
&=\frac{1}{2}\rho^{(1)}(0)\left(\left(2\rho^{(1)}(0)\right)^{-1}-\left(2\rho^{(1)}(0)\right)^{-2}\rho^{(2)}(0)\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\right)\\
&\quad\left(-\rho^{(2)}(0)^{-1}+\frac{1}{2}\rho^{(2)}(0)^{-2}\rho^{(3)}(0)\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\right)\\
&=-\frac{1}{4}\rho^{(2)}(0)^{-1}+\frac{1}{8}\left(\rho^{(1)}(0)^{-1}+\rho^{(2)}(0)^{-2}\rho^{(3)}(0)\right)\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\\
&=:a_0+b_0\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2),
\end{aligned}\tag{C.15}$$

where

$$a_0=-\frac{1}{4}\rho^{(2)}(0)^{-1}\text{ and }b_0=\frac{1}{8}\left(\rho^{(1)}(0)^{-1}+\rho^{(2)}(0)^{-2}\rho^{(3)}(0)\right).$$

For (ii), we first note that

$$\begin{aligned}
k_*(\boldsymbol{t})&=k_1(\boldsymbol{t})+k_2(\boldsymbol{t})\|\boldsymbol{t}\|^2\\
&=\rho^{(1)}(0)^{-1}\left(\rho^{(1)}(\|\boldsymbol{t}\|^2)+2\rho^{(2)}(\|\boldsymbol{t}\|^2)\|\boldsymbol{t}\|^2\right)\\
&=\rho^{(1)}(0)^{-1}\left(\rho^{(1)}(0)+\rho^{(2)}(0)\|\boldsymbol{t}\|^2+\frac{1}{2}\rho^{(3)}(0)\|\boldsymbol{t}\|^4+2\rho^{(2)}(0)\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^4)\right)\\
&=\rho^{(1)}(0)^{-1}\left(\rho^{(1)}(0)+3\rho^{(2)}(0)\|\boldsymbol{t}\|^2+\frac{5}{2}\rho^{(3)}(0)\|\boldsymbol{t}\|^4+o(\|\boldsymbol{t}\|^4)\right)\\
&=1+3\rho^{(1)}(0)^{-1}\rho^{(2)}(0)\|\boldsymbol{t}\|^2+\frac{5}{2}\rho^{(1)}(0)^{-1}\rho^{(3)}(0)\|\boldsymbol{t}\|^4+o(\|\boldsymbol{t}\|^4),
\end{aligned}\tag{C.16}$$

which is followed by

$$k_1(\boldsymbol{t})+k_*(\boldsymbol{t})=2+4\rho^{(1)}(0)^{-1}\rho^{(2)}(0)\|\boldsymbol{t}\|^2+3\rho^{(1)}(0)^{-1}\rho^{(3)}(0)\|\boldsymbol{t}\|^4+o(\|\boldsymbol{t}\|^4)$$

and

$$\begin{aligned}
&1-k_*^2(\boldsymbol{t})\\
&=1-\left(1+3\rho^{(1)}(0)^{-1}\rho^{(2)}(0)\|\boldsymbol{t}\|^2+\frac{5}{2}\rho^{(1)}(0)^{-1}\rho^{(3)}(0)\|\boldsymbol{t}\|^4+o(\|\boldsymbol{t}\|^4)\right)^2\\
&=-6\rho^{(1)}(0)^{-1}\rho^{(2)}(0)\|\boldsymbol{t}\|^2-\left(5\rho^{(1)}(0)^{-1}\rho^{(3)}(0)+9\rho^{(1)}(0)^{-2}\rho^{(2)}(0)^2\right)\|\boldsymbol{t}\|^4+o(\|\boldsymbol{t}\|^4),
\end{aligned}$$

and thus

$$
\begin{aligned}
&\|\boldsymbol{t}\|^2(1-k_*^2(\boldsymbol{t}))^{-1} \\
=&-\left(6\rho^{(1)}(0)^{-1}\rho^{(2)}(0)+\left(5\rho^{(1)}(0)^{-1}\rho^{(3)}(0)+9\rho^{(1)}(0)^{-2}\rho^{(2)}(0)^2\right)\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\right)^{-1} \\
=&-\left(6\rho^{(1)}(0)^{-1}\rho^{(2)}(0)\right)^{-1} \\
&+\left(6\rho^{(1)}(0)^{-1}\rho^{(2)}(0)\right)^{-2}\left(5\rho^{(1)}(0)^{-1}\rho^{(3)}(0)+9\rho^{(1)}(0)^{-2}\rho^{(2)}(0)^2\right)\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2) \\
=&-\frac{1}{6}\rho^{(1)}(0)\rho^{(2)}(0)^{-1}+\left(\frac{5}{36}\rho^{(1)}(0)\rho^{(2)}(0)^{-2}\rho^{(3)}(0)+\frac{1}{4}\right)\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2).
\end{aligned}
\tag{C.17}
$$

Then we have

$$
\begin{aligned}
&\|\boldsymbol{t}\|^2k_4(\boldsymbol{t}) \\
=&\|\boldsymbol{t}\|^2k_2(\boldsymbol{t})\frac{k_1(\boldsymbol{t})+k_*(\boldsymbol{t})}{1-k_*^2(\boldsymbol{t})} \\
=&k_2(\boldsymbol{t})(k_1(\boldsymbol{t})+k_*(\boldsymbol{t})) \\
&\left(-\frac{1}{6}\rho^{(1)}(0)\rho^{(2)}(0)^{-1}+\left(\frac{5}{36}\rho^{(1)}(0)\rho^{(2)}(0)^{-2}\rho^{(3)}(0)+\frac{1}{4}\right)\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\right) \\
=&2\rho^{(1)}(0)^{-1}\left(\rho^{(2)}(0)+\rho^{(3)}(0)\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\right) \\
&\left(2+4\rho^{(1)}(0)^{-1}\rho^{(2)}(0)\|\boldsymbol{t}\|^2+3\rho^{(1)}(0)^{-1}\rho^{(3)}(0)\|\boldsymbol{t}\|^4+o(\|\boldsymbol{t}\|^4)\right) \\
&\left(-\frac{1}{6}\rho^{(1)}(0)\rho^{(2)}(0)^{-1}+\left(\frac{5}{36}\rho^{(1)}(0)\rho^{(2)}(0)^{-2}\rho^{(3)}(0)+\frac{1}{4}\right)\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\right) \\
=&-\frac{2}{3}+\left(-\frac{1}{3}\rho^{(1)}(0)^{-1}\rho^{(2)}(0)-\frac{1}{9}\rho^{(2)}(0)^{-1}\rho^{(3)}(0)\right)\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2).
\end{aligned}
\tag{C.18}
$$

Therefore, by (C.15) and (C.18),

$$
\begin{aligned}
&\frac{\|\boldsymbol{t}\|^4k_4(\boldsymbol{t})}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))} \\
=&\left(a_0+b_0\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\right) \\
&\left(-\frac{2}{3}+\left(-\frac{1}{3}\rho^{(1)}(0)^{-1}\rho^{(2)}(0)-\frac{1}{9}\rho^{(2)}(0)^{-1}\rho^{(3)}(0)\right)\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\right) \\
=:&\,a_0'+b_0'\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2),
\end{aligned}
\tag{C.19}
$$

where

$$a_0'=-\frac{2}{3}a_0=\frac{1}{6}\rho^{(2)}(0)^{-1}$$

and

$$b_0' = a_0\left(-\frac{1}{3}\rho^{(1)}(0)^{-1}\rho^{(2)}(0) - \frac{1}{9}\rho^{(2)}(0)^{-1}\rho^{(3)}(0)\right) - \frac{2}{3}b_0 = -\frac{1}{18}\rho^{(2)}(0)^{-2}\rho^{(3)}(0).$$

For (iii), by (C.15),

$$\begin{aligned}
&\frac{\|\boldsymbol{t}\|^2 k_1(\boldsymbol{t})}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))} \\
&= \left(1+\rho^{(1)}(0)^{-1}\rho^{(2)}(0)\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right)\left(a_0 + b_0\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right) \\
&= a_0 + \left(a_0\rho^{(1)}(0)^{-1}\rho^{(2)}(0) + b_0\right)\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2) \\
&=: a_0'' + b_0''\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2),
\end{aligned} \tag{C.20}$$

where

$$a_0'' = a_0 = -\frac{1}{4}\rho^{(2)}(0)^{-1}$$

and

$$\begin{aligned}
b_0'' &= -\frac{1}{4}\rho^{(2)}(0)^{-1}\rho^{(1)}(0)^{-1}\rho^{(2)}(0) + \frac{1}{8}\left(\rho^{(1)}(0)^{-1} + \rho^{(2)}(0)^{-2}\rho^{(3)}(0)\right) \\
&= \frac{1}{8}\left(-\rho^{(1)}(0)^{-1} + \rho^{(2)}(0)^{-2}\rho^{(3)}(0)\right).
\end{aligned}$$

For (iv), by (C.13), (C.15) and (C.19),

$$\begin{aligned}
\frac{\|\boldsymbol{t}\|^4 k_5(\boldsymbol{t})}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))} &= \frac{\|\boldsymbol{t}\|^4 k_4(\boldsymbol{t})}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))} + \frac{\|\boldsymbol{t}\|^4 (k_5(\boldsymbol{t}) - k_4(\boldsymbol{t}))}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))} \\
&= \frac{\|\boldsymbol{t}\|^4 k_4(\boldsymbol{t})}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))} + o(\|\boldsymbol{t}\|^2) \\
&= a_0' + b_0'\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2),
\end{aligned} \tag{C.21}$$

C.3.3 Asymptotic Expansions of the Main Part

In this part, we would like to get the asymptotic expansion of

$$\begin{aligned}
&\boldsymbol{\Sigma}(\boldsymbol{t})[1:(L-2), 1:(L-2)] \\
&= \boldsymbol{G}_{22}(\boldsymbol{0}) + \frac{1}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{21}^T(\boldsymbol{t}) + \frac{k_4(\boldsymbol{t})}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{21}^T(\boldsymbol{t})
\end{aligned} \tag{C.22}$$

as $\|\boldsymbol{t}\| \to 0$. To this end, we still need to expand

(i) $\boldsymbol{G}_{21}(\boldsymbol{t})_{(i_1+j_1(j_1-1)/2)}\left(\boldsymbol{G}_{21}(\boldsymbol{t})_{(i_2+j_2(j_2-1)/2)}\right)^T$

(ii) $\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{21}^T(\boldsymbol{t})\,[i_1+j_1(j_1-1)/2, i_2+j_2(j_2-1)/2]$

for any integers $1 \le i_1 \le j_1 \le N$ and $1 \le i_2 \le j_2 \le N$.

Note that $\boldsymbol{G}_{21}(\boldsymbol{t}) = -\mathrm{Cov}\left[\nabla^2 X(\boldsymbol{t}), \nabla X(\boldsymbol{0})\right]$ can be written in the form:

$$\boldsymbol{G}_{21}(\boldsymbol{t}) = \begin{pmatrix} \boldsymbol{B}_{11}(\boldsymbol{t}) & \boldsymbol{B}_{12}(\boldsymbol{t}) & \cdots & \boldsymbol{B}_{1N}(\boldsymbol{t}) \\ \boldsymbol{B}_{21}(\boldsymbol{t}) & \boldsymbol{B}_{22}(\boldsymbol{t}) & \cdots & \boldsymbol{B}_{2N}(\boldsymbol{t}) \\ \vdots & \vdots & \cdots & \vdots \\ \boldsymbol{B}_{N1}(\boldsymbol{t}) & \boldsymbol{B}_{N2}(\boldsymbol{t}) & \cdots & \boldsymbol{B}_{NN}(\boldsymbol{t}) \end{pmatrix},$$

where by (C.3),

$$\begin{aligned}\boldsymbol{B}_{jk}(\boldsymbol{t}) :&= \begin{pmatrix} R_{1jk}(\boldsymbol{t}) \\ R_{2jk}(\boldsymbol{t}) \\ \vdots \\ R_{jjk}(\boldsymbol{t}) \end{pmatrix} \\ &= 4\rho^{(2)}(\|\boldsymbol{t}\|^2)\left(t_k \begin{pmatrix} \delta_{1,j} \\ \delta_{2,j} \\ \vdots \\ \delta_{j,j} \end{pmatrix} + \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_j \end{pmatrix} \delta_{j,k} + t_j \begin{pmatrix} \delta_{1,k} \\ \delta_{2,k} \\ \vdots \\ \delta_{j,k} \end{pmatrix} \right) + 8\rho^{(3)}(\|\boldsymbol{t}\|^2) t_j t_k \begin{pmatrix} t_1 \\ t_2 \\ \vdots \\ t_j \end{pmatrix}\end{aligned}$$

for any integers $1 \le j, k \le N$. For any integers $1 \le i \le j \le N$ and $1 \le k \le N$,

$$\boldsymbol{G}_{21}(\boldsymbol{t})\left[i + \frac{j(j-1)}{2}, k\right] = \boldsymbol{B}_{jk}(\boldsymbol{t})[i] = R_{ijk}(\boldsymbol{t}). \tag{C.23}$$

Let $\boldsymbol{v}_k = (\delta_{1,k}, \dots, \delta_{N,k})^T$, $k = 1, 2, \dots, N$. Then by (C.3) and (C.23),

$$\begin{aligned}\left(\boldsymbol{G}_{21}(\boldsymbol{t})_{\left(i+\frac{j(j-1)}{2}\right)}\right)^T = \begin{pmatrix} R_{ij1}(\boldsymbol{t}) \\ R_{ij2}(\boldsymbol{t}) \\ \vdots \\ R_{ijN}(\boldsymbol{t}) \end{pmatrix} &= 4\rho^{(2)}(\|\boldsymbol{t}\|^2)(\delta_{i,j}\boldsymbol{t} + t_i\boldsymbol{v}_j + t_j\boldsymbol{v}_i) + 8\rho^{(3)}(\|\boldsymbol{t}\|^2)t_i t_j \boldsymbol{t} \\ &= 4\left(\rho^{(2)}(0) + \rho^{(3)}(0)\|\boldsymbol{t}\|^2\right)(\delta_{i,j}\boldsymbol{t} + t_i\boldsymbol{v}_j + t_j\boldsymbol{v}_i) \\ &\quad + 8\rho^{(3)}(0)t_i t_j \boldsymbol{t} + o(\|\boldsymbol{t}\|^3).\end{aligned} \tag{C.24}$$

Note that for any integers $1 \le i, j \le N$, since $\boldsymbol{v}_j^T \boldsymbol{t} = t_j$ and $\boldsymbol{v}_i^T \boldsymbol{v}_j = \delta_{i,j}$, we have

$$(\delta_{i,j}\boldsymbol{t}^T + t_i\boldsymbol{v}_j^T + t_j\boldsymbol{v}_i^T)\boldsymbol{t} = \delta_{i,j}\|\boldsymbol{t}\|^2 + 2t_it_j.$$

Then for any integers $1 \le i_1 \le j_1 \le N$ and $1 \le i_2 \le j_2 \le N$, we have

$$\begin{aligned}
&\boldsymbol{G}_{21}(\boldsymbol{t})_{(i_1+j_1(j_1-1)/2)}\left(\boldsymbol{G}_{21}(\boldsymbol{t})_{(i_2+j_2(j_2-1)/2)}\right)^T \\
&= 16\left(\rho^{(2)}(0) + \rho^{(3)}(0)\|\boldsymbol{t}\|^2\right)^2 (\delta_{i_1,j_1}\boldsymbol{t}^T + t_{i_1}\boldsymbol{v}_{j_1}^T + t_{j_1}\boldsymbol{v}_{i_1}^T)(\delta_{i_2,j_2}\boldsymbol{t} + t_{i_2}\boldsymbol{v}_{j_2} + t_{j_2}\boldsymbol{v}_{i_2}) \\
&\quad + 32\rho^{(3)}(0)t_{i_2}t_{j_2}\left(\rho^{(2)}(0) + \rho^{(3)}(0)\|\boldsymbol{t}\|^2\right)(\delta_{i_1,j_1}\boldsymbol{t}^T + t_{i_1}\boldsymbol{v}_{j_1}^T + t_{j_1}\boldsymbol{v}_{i_1}^T)\boldsymbol{t} \\
&\quad + 32\rho^{(3)}(0)t_{i_1}t_{j_1}\left(\rho^{(2)}(0) + \rho^{(3)}(0)\|\boldsymbol{t}\|^2\right)(\delta_{i_2,j_2}\boldsymbol{t}^T + t_{i_2}\boldsymbol{v}_{j_2}^T + t_{j_2}\boldsymbol{v}_{i_2}^T)\boldsymbol{t} + o(\|\boldsymbol{t}\|^4) \\
&= 16\left(\rho^{(2)}(0)^2 + 2\rho^{(2)}(0)\rho^{(3)}(0)\|\boldsymbol{t}\|^2\right)\left(\delta_{i_1,j_1}\|\boldsymbol{t}\|^2 + 2t_{i_1}t_{j_1}\right)\delta_{i_2,j_2} \\
&\quad + 16\left(\rho^{(2)}(0)^2 + 2\rho^{(2)}(0)\rho^{(3)}(0)\|\boldsymbol{t}\|^2\right)(\delta_{i_1,j_1}t_{i_2}t_{j_2} + \delta_{j_1,j_2}t_{i_1}t_{i_2} + \delta_{i_1,j_2}t_{j_1}t_{i_2}) \\
&\quad + 16\left(\rho^{(2)}(0)^2 + 2\rho^{(2)}(0)\rho^{(3)}(0)\|\boldsymbol{t}\|^2\right)(\delta_{i_1,j_1}t_{i_2}t_{j_2} + \delta_{i_2,j_1}t_{i_1}t_{j_2} + \delta_{i_1,i_2}t_{j_1}t_{j_2}) \\
&\quad + 32\rho^{(3)}(0)\left(\rho^{(2)}(0) + \rho^{(3)}(0)\|\boldsymbol{t}\|^2\right)\left(\delta_{i_1,j_1}t_{i_2}t_{j_2}\|\boldsymbol{t}\|^2 + \delta_{i_2,j_2}t_{i_1}t_{j_1}\|\boldsymbol{t}\|^2 + 4t_{i_1}t_{j_1}t_{i_2}t_{j_2}\right) \\
&\quad + o(\|\boldsymbol{t}\|^4) \\
&=: a_1\|\boldsymbol{t}\|^2 + b_1\|\boldsymbol{t}\|^4 + o(\|\boldsymbol{t}\|^4),
\end{aligned} \tag{C.25}$$

where

$$\begin{aligned}
a_1 = 16\rho^{(2)}(0)^2\big(&\delta_{i_1,j_1}\delta_{i_2,j_2} + 2\delta_{i_2,j_2}u_{i_1}u_{j_1} + 2\delta_{i_1,j_1}u_{i_2}u_{j_2} \\
&+ \delta_{j_1,j_2}u_{i_1}u_{i_2} + \delta_{i_1,j_2}u_{j_1}u_{i_2} + \delta_{i_2,j_1}u_{i_1}u_{j_2} + \delta_{i_1,i_2}u_{j_1}u_{j_2}\big)
\end{aligned}$$

and

$$\begin{aligned}
b_1 &= 2\rho^{(2)}(0)^{-1}\rho^{(3)}(0)a_1 + 32\rho^{(3)}(0)\rho^{(2)}(0)\left(\delta_{i_1,j_1}u_{i_2}u_{j_2} + \delta_{i_2,j_2}u_{i_1}u_{j_1} + 4u_{i_1}u_{j_1}u_{i_2}u_{j_2}\right) \\
&= 32\rho^{(3)}(0)\rho^{(2)}(0)\big(\delta_{i_1,j_1}\delta_{i_2,j_2} + 3\delta_{i_2,j_2}u_{i_1}u_{j_1} + 3\delta_{i_1,j_1}u_{i_2}u_{j_2} \\
&\quad + \delta_{j_1,j_2}u_{i_1}u_{i_2} + \delta_{i_1,j_2}u_{j_1}u_{i_2} + \delta_{i_2,j_1}u_{i_1}u_{j_2} + \delta_{i_1,i_2}u_{j_1}u_{j_2} + 4u_{i_1}u_{j_1}u_{i_2}u_{j_2}\big).
\end{aligned}$$

By (C.24), we also have

$$\begin{aligned}
&\boldsymbol{G}_{21}(\boldsymbol{t})_{(i_1+j_1(j_1-1)/2)}\boldsymbol{t} = (\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t})_{(i_1+j_1(j_1-1)/2)} \\
&= \left(4\rho^{(2)}(\|\boldsymbol{t}\|^2)(\delta_{i,j}\boldsymbol{t}^T + t_i\boldsymbol{v}_j^T + t_j\boldsymbol{v}_i^T) + 8\rho^{(3)}(\|\boldsymbol{t}\|^2)t_it_j\boldsymbol{t}^T\right)\boldsymbol{t} \\
&= 4\rho^{(2)}(\|\boldsymbol{t}\|^2)(\delta_{i,j}\|\boldsymbol{t}\|^2 + 2t_it_j) + 8\rho^{(3)}(\|\boldsymbol{t}\|^2)t_it_j\|\boldsymbol{t}\|^2 \\
&= 4\left(\rho^{(2)}(0) + \rho^{(3)}(0)\|\boldsymbol{t}\|^2\right)(\delta_{i,j}\|\boldsymbol{t}\|^2 + 2t_it_j) + 8\rho^{(3)}(0)t_it_j\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^4),
\end{aligned} \tag{C.26}$$

and thus

$$
\begin{aligned}
&\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{21}^T(\boldsymbol{t})\,[i_1+j_1(j_1-1)/2, i_2+j_2(j_2-1)/2] \\
&= 16\left(\rho^{(2)}(0)+\rho^{(3)}(0)\|\boldsymbol{t}\|^2\right)^2(\delta_{i_1,j_1}\|\boldsymbol{t}\|^2+2t_{i_1}t_{j_1})(\delta_{i_2,j_2}\|\boldsymbol{t}\|^2+2t_{i_2}t_{j_2}) \\
&\quad + 4\left(\rho^{(2)}(0)+\rho^{(3)}(0)\|\boldsymbol{t}\|^2\right)(\delta_{i_1,j_1}\|\boldsymbol{t}\|^2+2t_{i_1}t_{j_1})8\rho^{(3)}(0)t_{i_2}t_{j_2}\|\boldsymbol{t}\|^2 \\
&\quad + 4\left(\rho^{(2)}(0)+\rho^{(3)}(0)\|\boldsymbol{t}\|^2\right)(\delta_{i_2,j_2}\|\boldsymbol{t}\|^2+2t_{i_2}t_{j_2})8\rho^{(3)}(0)t_{i_1}t_{j_1}\|\boldsymbol{t}\|^2 \\
&\quad + 8\rho^{(3)}(0)t_{i_1}t_{j_1}\|\boldsymbol{t}\|^2 8\rho^{(3)}(0)t_{i_2}t_{j_2}\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^6) \\
&= 16\left(\rho^{(2)}(0)^2+2\rho^{(2)}(0)\rho^{(3)}(0)\|\boldsymbol{t}\|^2+\rho^{(3)}(0)^2\|\boldsymbol{t}\|^4\right)(\delta_{i_1,j_1}\|\boldsymbol{t}\|^2+2t_{i_1}t_{j_1})(\delta_{i_2,j_2}\|\boldsymbol{t}\|^2+2t_{i_2}t_{j_2}) \\
&\quad + 32\rho^{(3)}(0)\left(\rho^{(2)}(0)+\rho^{(3)}(0)\|\boldsymbol{t}\|^2\right)(\delta_{i_1,j_1}\|\boldsymbol{t}\|^2+2t_{i_1}t_{j_1})t_{i_2}t_{j_2}\|\boldsymbol{t}\|^2 \\
&\quad + 32\rho^{(3)}(0)\left(\rho^{(2)}(0)+\rho^{(3)}(0)\|\boldsymbol{t}\|^2\right)(\delta_{i_2,j_2}\|\boldsymbol{t}\|^2+2t_{i_2}t_{j_2})t_{i_1}t_{j_1}\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^6) \\
&=: a_2\|\boldsymbol{t}\|^4 + b_2\|\boldsymbol{t}\|^6 + o(\|\boldsymbol{t}\|^6),
\end{aligned}
\tag{C.27}
$$

where

$$
a_2 = 16\rho^{(2)}(0)^2(\delta_{i_1,j_1}+2u_{i_1}u_{j_1})(\delta_{i_2,j_2}+2u_{i_2}u_{j_2})
$$

and

$$
\begin{aligned}
b_2 &= 32\rho^{(2)}(0)\rho^{(3)}(0)(\delta_{i_1,j_1}+2u_{i_1}u_{j_1})(\delta_{i_2,j_2}+2u_{i_2}u_{j_2}) \\
&\quad + 32\rho^{(2)}(0)\rho^{(3)}(0)(\delta_{i_1,j_1}u_{i_2}u_{j_2}+\delta_{i_2,j_2}u_{i_1}u_{j_1}+4u_{i_1}u_{j_1}u_{i_2}u_{j_2}) \\
&= 32\rho^{(2)}(0)\rho^{(3)}(0)\left(\delta_{i_1,j_1}\delta_{i_2,j_2}+3\delta_{i_1,j_1}u_{i_2}u_{j_2}+3\delta_{i_2,j_2}u_{i_1}u_{j_1}+8u_{i_1}u_{j_1}u_{i_2}u_{j_2}\right).
\end{aligned}
$$

By (C.4),

$$
\begin{aligned}
\boldsymbol{G}_{22}(\boldsymbol{0})\,[i_1+j_1(j_1-1)/2, i_2+j_2(j_2-1)/2] &= R_{i1j1i2j2}(\boldsymbol{0}) \\
&= 4\rho^{(2)}(0)\left(\delta_{i_1,j_1}\delta_{i_2,j_2}+\delta_{i_2,j_1}\delta_{i_1,j_2}+\delta_{i_1,i_2}\delta_{j_1,j_2}\right).
\end{aligned}
\tag{C.28}
$$

Finally, by (C.15), (C.19), (C.22), (C.25), (C.27) and (C.28) we have for integers $1 \le i_1 \le j_1 \le N$ and $1 \le i_2 \le j_2 \le N$,

$$
\begin{aligned}
&\boldsymbol{\Sigma}(\boldsymbol{t})\,[i_1+j_1(j_1-1)/2, i_2+j_2(j_2-1)/2] \\
&= 4\rho^{(2)}(0)\left(\delta_{i_1,j_1}\delta_{i_2,j_2}+\delta_{i_2,j_1}\delta_{i_1,j_2}+\delta_{i_1,i_2}\delta_{j_1,j_2}\right) \\
&\quad + \|\boldsymbol{t}\|^{-2}\left(a_0+b_0\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\right)\left(a_1\|\boldsymbol{t}\|^2+b_1\|\boldsymbol{t}\|^4+o(\|\boldsymbol{t}\|^4)\right) \\
&\quad + \|\boldsymbol{t}\|^{-4}\left(a_0'+b_0'\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\right)\left(a_2\|\boldsymbol{t}\|^4+b_2\|\boldsymbol{t}\|^6+o(\|\boldsymbol{t}\|^6)\right) \\
&= 4\rho^{(2)}(0)\left(\delta_{i_1,j_1}\delta_{i_2,j_2}+\delta_{i_2,j_1}\delta_{i_1,j_2}+\delta_{i_1,i_2}\delta_{j_1,j_2}\right) \\
&\quad + \left(a_0+b_0\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\right)\left(a_1+b_1\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\right) \\
&\quad + \left(a_0'+b_0'\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\right)\left(a_2+b_2\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2)\right) \\
&=: a_3+b_3\|\boldsymbol{t}\|^2+o(\|\boldsymbol{t}\|^2),
\end{aligned}
$$

where

$$\begin{aligned} a_3 &= 4\rho^{(2)}(0)\left(\delta_{i_1,j_1}\delta_{i_2,j_2} + \delta_{i_2,j_1}\delta_{i_1,j_2} + \delta_{i_1,i_2}\delta_{j_1,j_2}\right) + a_0 a_1 + a_0' a_2 \\ &= 4\rho^{(2)}(0)(\delta_{i_2,j_1}\delta_{i_1,j_2} + \delta_{i_1,i_2}\delta_{j_1,j_2} - \delta_{j_1,j_2}u_{i_1}u_{i_2} - \delta_{i_1,j_2}u_{j_1}u_{i_2} \\ &\quad - \delta_{i_2,j_1}u_{i_1}u_{j_2} - \delta_{i_1,i_2}u_{j_1}u_{j_2} + 2u_{i_1}u_{j_1}u_{i_2}u_{j_2}) \\ &\quad + \frac{8}{3}\rho^{(2)}(0)(\delta_{i_1,j_1} - u_{i_1}u_{j_1})(\delta_{i_2,j_2} - u_{i_2}u_{j_2}), \end{aligned}$$

and by letting $\alpha := \rho^{(1)}(0)^{-1}\rho^{(2)}(0)^2$ and $\beta := \rho^{(3)}(0)$,

$$\begin{aligned} b_3 &= a_0 b_1 + b_0 a_1 + a_0' b_2 + b_0' a_2 \\ &= \left(2\alpha - \frac{14}{9}\beta\right)\delta_{i_1,j_1}\delta_{i_2,j_2} + \left(4\alpha - \frac{52}{9}\beta\right)\delta_{i_2,j_2}u_{i_1}u_{j_1} + \left(4\alpha - \frac{52}{9}\beta\right)\delta_{i_1,j_1}u_{i_2}u_{j_2} \\ &\quad + (2\alpha - 6\beta)\,\delta_{j_1,j_2}u_{i_1}u_{i_2} + (2\alpha - 6\beta)\,\delta_{i_1,j_2}u_{j_1}u_{i_2} + (2\alpha - 6\beta)\,\delta_{i_2,j_1}u_{i_1}u_{j_2} \\ &\quad + (2\alpha - 6\beta)\,\delta_{i_1,i_2}u_{j_1}u_{j_2} + \frac{64}{9}\beta u_{i_1}u_{j_1}u_{i_2}u_{j_2}, \end{aligned}$$

which can be verified by lengthy but straightforward calculation.

C.3.4 Asymptotic Expansions of the Side Parts

In this part, we would like to get the asymptotic expansion of

$$\begin{aligned} &\boldsymbol{\Sigma}(\boldsymbol{t})[1:(L-2), L-1] \\ &= \boldsymbol{G}_{20}(\boldsymbol{0}) + \frac{1}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) + \frac{k_4(\boldsymbol{t})}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) \end{aligned} \tag{C.29}$$

and

$$\begin{aligned} &\boldsymbol{\Sigma}(\boldsymbol{t})[1:(L-2), L] \\ &= \boldsymbol{G}_{20}(\boldsymbol{t}) + \frac{k_1(\boldsymbol{t})}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) + \frac{k_5(\boldsymbol{t})}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) \end{aligned} \tag{C.30}$$

as $\|\boldsymbol{t}\| \to 0$. To this end, we still need to expand

(i) $\boldsymbol{G}_{20}(\boldsymbol{t})[i + j(j-1)/2]$,

(ii) $(\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}))[i + j(j-1)/2]$,

(iii) $(\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}))[i + j(j-1)/2]$

for any integers $1 \le i \le j \le N$.

By (C.2), we have

$$\boldsymbol{G}_{20}(\mathbf{0})[i + j(j-1)/2] = R_{ij}(\mathbf{0}) = 2\rho^{(1)}(0)\delta_{i,j} \tag{C.31}$$

and

$$\begin{aligned}
&\boldsymbol{G}_{20}(\boldsymbol{t})[i + j(j-1)/2] \\
&= R_{ij}(\boldsymbol{t}) \\
&= 2\rho^{(1)}(\|\boldsymbol{t}\|^2)\delta_{i,j} + 4t_it_j\rho^{(2)}(\|\boldsymbol{t}\|^2) \\
&= 2\delta_{i,j}\left(\rho^{(1)}(0) + \rho^{(2)}(0)\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right) + 4u_iu_j\|\boldsymbol{t}\|^2\left(\rho^{(2)}(0) + \rho^{(3)}(0)\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right) \\
&= 2\delta_{i,j}\rho^{(1)}(0) + (2\delta_{i,j} + 4u_iu_j)\,\rho^{(2)}(0)\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2).
\end{aligned} \tag{C.32}$$

By (C.1),

$$\boldsymbol{G}_{01}(\boldsymbol{t}) = (R_1(\boldsymbol{t}), \ldots, R_N(\boldsymbol{t})) = 2\rho^{(1)}(\|\boldsymbol{t}\|^2)\boldsymbol{t}^T. \tag{C.33}$$

Then by (C.26) and (C.33),

$$\begin{aligned}
&\left(\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t})\right)[i + j(j-1)/2] \\
&= 2\rho^{(1)}(\|\boldsymbol{t}\|^2)\boldsymbol{G}_{21}(\boldsymbol{t})_{(i+j(j-1)/2)}\boldsymbol{t} \\
&= 2\rho^{(1)}(\|\boldsymbol{t}\|^2)\left(4\rho^{(2)}(\|\boldsymbol{t}\|^2)(\delta_{i,j}\|\boldsymbol{t}\|^2 + 2t_it_j) + 8\rho^{(3)}(\|\boldsymbol{t}\|^2)t_it_j\|\boldsymbol{t}\|^2\right) \\
&= 2\left(\rho^{(1)}(0) + \rho^{(2)}(0)\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right)\left(4\rho^{(2)}(\|\boldsymbol{t}\|^2)(\delta_{i,j}\|\boldsymbol{t}\|^2 + 2t_it_j) + 8\rho^{(3)}(\|\boldsymbol{t}\|^2)t_it_j\|\boldsymbol{t}\|^2\right) \\
&= 8\left(\rho^{(1)}(0)\rho^{(2)}(0) + \left(\rho^{(2)}(0)^2 + \rho^{(1)}(0)\rho^{(3)}(0)\right)\|\boldsymbol{t}\|^2\right)(\delta_{i,j} + 2u_iu_j)\|\boldsymbol{t}\|^2 \\
&\quad + 16\rho^{(1)}(0)\rho^{(3)}(0)u_iu_j\|\boldsymbol{t}\|^4 + o(\|\boldsymbol{t}\|^4) \\
&=: \widetilde{a}_1\|\boldsymbol{t}\|^2 + \widetilde{b}_1\|\boldsymbol{t}\|^4 + o(\|\boldsymbol{t}\|^4),
\end{aligned} \tag{C.34}$$

where

$$\widetilde{a}_1 = 8\rho^{(1)}(0)\rho^{(2)}(0)(\delta_{i,j} + 2u_iu_j)$$

and

$$\widetilde{b}_1 = 8\left(\rho^{(2)}(0)^2 + \rho^{(1)}(0)\rho^{(3)}(0)\right)(\delta_{i,j} + 2u_iu_j) + 16\rho^{(1)}(0)\rho^{(3)}(0)u_iu_j.$$

By (C.33),

$$\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) = 2\rho^{(1)}(\|\boldsymbol{t}\|^2)\|\boldsymbol{t}\|^2. \tag{C.35}$$

Then we have

$$\begin{aligned}
\left(\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t})\right)[i + j(j-1)/2] &= 2\rho^{(1)}(\|\boldsymbol{t}\|^2)\boldsymbol{G}_{21}(\boldsymbol{t})_{(i+j(j-1)/2)}\boldsymbol{t}\|\boldsymbol{t}\|^2 \\
&=: \widetilde{a}_2\|\boldsymbol{t}\|^4 + \widetilde{b}_2\|\boldsymbol{t}\|^6 + o(\|\boldsymbol{t}\|^6),
\end{aligned} \tag{C.36}$$

where by (C.34),

$$\widetilde{a}_2 = \widetilde{a}_1 \quad \text{and} \quad \widetilde{b}_2 = \widetilde{b}_1.$$

Now combining (C.15), (C.19), (C.29), (C.31), (C.34) and (C.36) implies for any integers $1 \le i \le j \le N$,

$$\begin{aligned}
&\boldsymbol{\Sigma}(\boldsymbol{t})[i + j(j-1)/2, L-1] \\
&= 2\rho^{(1)}(0)\delta_{i,j} + \left(a_0 + b_0\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right)\left(\widetilde{a}_1 + \widetilde{b}_1\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right) \\
&\quad + \left(a_0' + b_0'\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right)\left(\widetilde{a}_2 + \widetilde{b}_2\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right) \\
&=: \widetilde{a}_3 + \widetilde{b}_3\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2),
\end{aligned}$$

where

$$\begin{aligned}
\widetilde{a}_3 &= 2\rho^{(1)}(0)\delta_{i,j} + a_0\widetilde{a}_1 + a_0'\widetilde{a}_2 \\
&= 2\rho^{(1)}(0)\delta_{i,j} + 8\rho^{(1)}(0)\rho^{(2)}(0)(\delta_{i,j} + 2u_iu_j)\left(-\frac{1}{4}\rho^{(2)}(0)^{-1} + \frac{1}{6}\rho^{(2)}(0)^{-1}\right) \\
&= \frac{4}{3}\rho^{(1)}(0)(\delta_{i,j} - u_iu_j),
\end{aligned}$$

and by $\widetilde{a}_1 = \widetilde{a}_2$ and $\widetilde{b}_1 = \widetilde{b}_2$,

$$\begin{aligned}
\widetilde{b}_3 &= a_0\widetilde{b}_1 + b_0\widetilde{a}_1 + a_0'\widetilde{b}_2 + b_0'\widetilde{a}_2 \\
&= \left(-\frac{1}{4}\rho^{(2)}(0)^{-1} + \frac{1}{6}\rho^{(2)}(0)^{-1}\right) \\
&\quad \left(8\left(\rho^{(2)}(0)^2 + \rho^{(1)}(0)\rho^{(3)}(0)\right)(\delta_{i,j} + 2u_iu_j) + 16\rho^{(1)}(0)\rho^{(3)}(0)u_iu_j\right) \\
&\quad + \left(\frac{1}{8}\left(\rho^{(1)}(0)^{-1} + \rho^{(2)}(0)^{-2}\rho^{(3)}(0)\right) - \frac{1}{18}\rho^{(2)}(0)^{-2}\rho^{(3)}(0)\right) \\
&\quad \left(8\rho^{(1)}(0)\rho^{(2)}(0)(\delta_{i,j} + 2u_iu_j)\right) \\
&= \left(\frac{1}{3}\alpha' - \frac{1}{9}\beta'\right)\delta_{i,j} + \left(\frac{2}{3}\alpha' - \frac{14}{9}\beta'\right)u_iu_j,
\end{aligned}$$

where $\alpha' := \rho^{(2)}(0)$ and $\beta' := \rho^{(1)}(0)\rho^{(2)}(0)^{-1}\rho^{(3)}(0)$. Also combining (C.20), (C.21), (C.30), (C.32), (C.34) and (C.36) implies for any integers $1 \le i \le j \le N$,

$$\begin{aligned}\boldsymbol{\Sigma}(\boldsymbol{t})[i+j(j-1)/2, L] &= 2\rho^{(1)}(0)\delta_{i,j} + (2\delta_{i,j} + 4u_i u_j)\,\rho^{(2)}(0)\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2) \\ &\quad + \left(a_0'' + b_0''\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right)\left(\widetilde{a}_1 + \widetilde{b}_1\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right) \\ &\quad + \left(a_0' + b_0'\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right)\left(\widetilde{a}_2 + \widetilde{b}_2\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right) \\ &=: \widetilde{a}_4 + \widetilde{b}_4\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2),\end{aligned}$$

where by $a_0 = a_0''$,

$$\widetilde{a}_4 = 2\rho^{(1)}(0)\delta_{i,j} + a_0''\widetilde{a}_1 + a_0'\widetilde{a}_2 = 2\rho^{(1)}(0)\delta_{i,j} + a_0\widetilde{a}_1 + a_0'\widetilde{a}_2 = \widetilde{a}_3,$$

and by $(2\delta_{i,j} + 4u_i u_j)\rho^{(2)}(0) + b_0''\widetilde{a}_1 = b_0\widetilde{a}_1$,

$$\widetilde{b}_4 = (2\delta_{i,j} + 4u_i u_j)\,\rho^{(2)}(0) + a_0''\widetilde{b}_1 + b_0''\widetilde{a}_1 + a_0'\widetilde{b}_2 + b_0'\widetilde{a}_2 = a_0\widetilde{b}_1 + b_0\widetilde{a}_1 + a_0'\widetilde{b}_2 + b_0'\widetilde{a}_2 = \widetilde{b}_3.$$

C.3.5 Asymptotic Expansions of the Corner Part

In this part, we would like to get the asymptotic expansion of

$$\begin{aligned}&\boldsymbol{\Sigma}(\boldsymbol{t})[L-1, L-1] = \boldsymbol{\Sigma}(\boldsymbol{t})[L, L] \\ &= \boldsymbol{G}_{00}(\boldsymbol{0}) + \frac{1}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) + \frac{k_4(\boldsymbol{t})}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t})\end{aligned} \tag{C.37}$$

and

$$\begin{aligned}&\boldsymbol{\Sigma}(\boldsymbol{t})[L-1, L] = \boldsymbol{\Sigma}(\boldsymbol{t})[L, L-1] \\ &= \boldsymbol{G}_{00}(\boldsymbol{t}) + \frac{k_1(\boldsymbol{t})}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) + \frac{k_5(\boldsymbol{t})}{2\rho^{(1)}(0)(1-k_1^2(\boldsymbol{t}))}\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t})\end{aligned} \tag{C.38}$$

as $\|\boldsymbol{t}\| \to 0$. To this end, we still need to expand

(i) $\boldsymbol{G}_{00}(\boldsymbol{t})$,

(ii) $\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t})$,

(iii) $\boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t})$.

For (i), it is easy to see

$$\boldsymbol{G}_{00}(\boldsymbol{t}) = R(\boldsymbol{t}) = \rho(0) + \rho^{(1)}(0)\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2) \text{ and } \boldsymbol{G}_{00}(\boldsymbol{0}) = R(\boldsymbol{0}) = \rho(0). \tag{C.39}$$

For (ii), by (C.33), we have

$$\begin{aligned} \boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t}) &= 4\left(\rho^{(1)}(0) + \rho^{(2)}(0)\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right)^2 \|\boldsymbol{t}\|^2 \\ &= 4\rho^{(1)}(0)^2\|\boldsymbol{t}\|^2 + 8\rho^{(1)}(0)\rho^{(2)}(0)\|\boldsymbol{t}\|^4 + o(\|\boldsymbol{t}\|^4) \\ &=: \widehat{a}_1\|\boldsymbol{t}\|^2 + \widehat{b}_1\|\boldsymbol{t}\|^4 + o(\|\boldsymbol{t}\|^4), \end{aligned} \tag{C.40}$$

where

$$\widehat{a}_1 = 4\rho^{(1)}(0)^2 \text{ and } \widehat{b}_1 = 8\rho^{(1)}(0)\rho^{(2)}(0),$$

By (C.35),

$$\begin{aligned} \boldsymbol{G}_{01}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t}) &= 4\rho^{(1)}(0)^2\|\boldsymbol{t}\|^4 + 8\rho^{(1)}(0)\rho^{(2)}(0)\|\boldsymbol{t}\|^6 + o(\|\boldsymbol{t}\|^6) \\ &=: \widehat{a}_2\|\boldsymbol{t}\|^4 + \widehat{b}_2\|\boldsymbol{t}\|^6 + o(\|\boldsymbol{t}\|^6), \end{aligned} \tag{C.41}$$

where

$$\widehat{a}_2 = 4\rho^{(1)}(0)^2 \text{ and } \widehat{b}_2 = 8\rho^{(1)}(0)\rho^{(2)}(0).$$

Therefore, combining (C.15), (C.19), (C.37), (C.39), (C.40) and (C.41) implies

$$\begin{aligned} &\boldsymbol{\Sigma}(\boldsymbol{t})[L-1, L-1] = \boldsymbol{\Sigma}(\boldsymbol{t})[L, L] \\ &= \rho(0) + \left(a_0 + b_0\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right)\left(\widehat{a}_1 + \widehat{b}_1\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right) \\ &\quad + \left(a_0' + b_0'\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right)\left(\widehat{a}_2 + \widehat{b}_2\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right) \\ &=: \widehat{a}_3 + \widehat{b}_3\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2), \end{aligned}$$

where

$$\begin{aligned} \widehat{a}_3 &= \rho(0) + a_0\widehat{a}_1 + a_0'\widehat{a}_2 \\ &= \rho(0) + \left(-\frac{1}{4}\rho^{(2)}(0)^{-1} + \frac{1}{6}\rho^{(2)}(0)^{-1}\right) 4\rho^{(1)}(0)^2 \\ &= \rho(0) - \frac{1}{3}\rho^{(1)}(0)^2\rho^{(2)}(0)^{-1} \end{aligned}$$

and by $\widehat{a}_1 = \widehat{a}_2$ and $\widehat{b}_1 = \widehat{b}_2$,

$$\begin{aligned}
\widehat{b}_3 &= a_0\widehat{b}_1 + b_0\widehat{a}_1 + a_0'\widehat{b}_2 + b_0'\widehat{a}_2 \\
&= (a_0 + a_0')\widehat{b}_1 + (b_0 + b_0')\widehat{a}_1 \\
&= \left(-\frac{1}{4}\rho^{(2)}(0)^{-1} + \frac{1}{6}\rho^{(2)}(0)^{-1}\right) 8\rho^{(1)}(0)\rho^{(2)}(0) \\
&\quad + \left(\frac{1}{8}\left(\rho^{(1)}(0)^{-1} + \rho^{(2)}(0)^{-2}\rho^{(3)}(0)\right) - \frac{1}{18}\rho^{(2)}(0)^{-2}\rho^{(3)}(0)\right) 4\rho^{(1)}(0)^2 \\
&= -\frac{1}{6}\rho^{(1)}(0) + \frac{5}{18}\rho^{(1)}(0)^2\rho^{(2)}(0)^{-2}\rho^{(3)}(0).
\end{aligned}$$

Also, combining (C.20), (C.21), (C.38), (C.39), (C.40) and (C.41) implies

$$\begin{aligned}
&\boldsymbol{\Sigma}(\boldsymbol{t})[L-1, L] = \boldsymbol{\Sigma}(\boldsymbol{t})[L, L-1] \\
&= \rho(0) + \rho^{(1)}(0)\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2) \\
&\quad + \left(a_0'' + b_0''\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right)\left(\widehat{a}_1 + \widehat{b}_1\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right) \\
&\quad + \left(a_0' + b_0'\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right)\left(\widehat{a}_2 + \widehat{b}_2\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2)\right) \\
&=: \widehat{a}_4 + \widehat{b}_4\|\boldsymbol{t}\|^2 + o(\|\boldsymbol{t}\|^2),
\end{aligned}$$

where by $a_0 = a_0''$,

$$\widehat{a}_4 = \rho(0) + a_0''\widehat{a}_1 + a_0'\widehat{a}_2 = \rho(0) + a_0\widehat{a}_1 + a_0'\widehat{a}_2 = \widehat{a}_3,$$

and by $\rho^{(1)}(0) + b_0''\widehat{a}_1 = b_0\widehat{a}_1$,

$$\widehat{b}_4 = \rho^{(1)}(0) + a_0''\widehat{b}_1 + b_0''\widehat{a}_1 + a_0'''\widehat{b}_2 + b_0'''\widehat{a}_2 = a_0\widehat{b}_1 + b_0\widehat{a}_1 + a_0'\widehat{b}_2 + b_0'\widehat{a}_2 = \widehat{b}_3.$$

Hence the proof of Lemma 4.3.5 is completed.

C.4 Features of the Side Parts

Fix $N \geq 2$. Let X be qualified. Suppose that X also satisfies (4.51) for some $\tilde{\delta}_\rho > 0$. Recall in Section 4.5.2, $\boldsymbol{u}_0 = (0, \dots, 0, 1) \in \mathbb{R}^N$. In this section, we would like to calculate the side parts (C.29) and (C.30) when $\boldsymbol{t}$ has the form $\boldsymbol{t} := \boldsymbol{u}_0 r$ for any $r \in [0, \tilde{\delta}_\rho]$. Recall that by (C.31)

$$\boldsymbol{G}_{20}(\boldsymbol{0})[i + j(j-1)/2] = 2\rho^{(1)}(0)\delta_{i,j}, \tag{C.42}$$

by (C.32)

$$\boldsymbol{G}_{20}(\boldsymbol{t})[i+j(j-1)/2] = 2\rho^{(1)}(\|\boldsymbol{t}\|^2)\delta_{i,j} + 4t_it_j\rho^{(2)}(\|\boldsymbol{t}\|^2), \tag{C.43}$$

by (C.34)

$$\begin{aligned} &\left(\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t})\right)[i+j(j-1)/2] \\ &= 2\rho^{(1)}(\|\boldsymbol{t}\|^2)\left(4\rho^{(2)}(\|\boldsymbol{t}\|^2)(\delta_{i,j}\|\boldsymbol{t}\|^2 + 2t_it_j) + 8\rho^{(3)}(\|\boldsymbol{t}\|^2)t_it_j\|\boldsymbol{t}\|^2\right), \end{aligned} \tag{C.44}$$

and by (C.36)

$$\begin{aligned} &\left(\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t})\right)[i+j(j-1)/2] \\ &= 2\rho^{(1)}(\|\boldsymbol{t}\|^2)\|\boldsymbol{t}\|^2\left(4\rho^{(2)}(\|\boldsymbol{t}\|^2)(\delta_{i,j}\|\boldsymbol{t}\|^2 + 2t_it_j) + 8\rho^{(3)}(\|\boldsymbol{t}\|^2)t_it_j\|\boldsymbol{t}\|^2\right) \end{aligned} \tag{C.45}$$

for any $\boldsymbol{t} \in B(\mathbf{0}_N, \delta_\rho) \setminus \{\mathbf{0}_N\}$ and integers $1 \le i \le j \le N$. Then by taking $\boldsymbol{t} := \boldsymbol{u}_0 r$ into (C.42)-(C.45), one can easily check

(i) for any integers $1 \le i < j \le N$

$$\boldsymbol{G}_{20}(\mathbf{0})[i+j(j-1)/2] = 0,$$

$$\boldsymbol{G}_{20}(\boldsymbol{t})[i+j(j-1)/2] = 0,$$

$$\left(\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t})\right)[i+j(j-1)/2] = 0,$$

and

$$\left(\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t})\right)[i+j(j-1)/2] = 0;$$

(ii) for any integer $1 \le k \le N-1$

$$\boldsymbol{G}_{20}(\mathbf{0})[k+k(k-1)/2] = 2\rho^{(1)}(0),$$

$$\boldsymbol{G}_{20}(\boldsymbol{t})[k+k(k-1)/2] = 2\rho^{(1)}(r^2),$$

$$\left(\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t})\right)[k+k(k-1)/2] = 8\rho^{(1)}(r^2)\rho^{(2)}(r^2)r^2,$$

and

$$\left(\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t})\right)[k+k(k-1)/2] = 8\rho^{(1)}(r^2)\rho^{(2)}(r^2)r^4;$$

(iii) when $k = N$

$$\boldsymbol{G}_{20}(\mathbf{0})[k+k(k-1)/2] = 2\rho^{(1)}(0),$$

$$\boldsymbol{G}_{20}(\boldsymbol{t})[k+k(k-1)/2] = 2\rho^{(1)}(r^2) + 4r^2\rho^{(2)}(r^2),$$

$$\left(\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{G}_{01}^T(\boldsymbol{t})\right)[k+k(k-1)/2] = 24\rho^{(1)}(r^2)\rho^{(2)}(r^2)r^2 + 16\rho^{(1)}(r^2)\rho^{(3)}(r^2)r^4,$$

and

$$\left(\boldsymbol{G}_{21}(\boldsymbol{t})\boldsymbol{t}\boldsymbol{t}^T\boldsymbol{G}_{01}^T(\boldsymbol{t})\right)[k+k(k-1)/2] = 24\rho^{(1)}(r^2)\rho^{(2)}(r^2)r^4 + 16\rho^{(1)}(r^2)\rho^{(3)}(r^2)r^6.$$

Then by (C.29) and (C.30), we have for $k = 1, 2, \dots, N-1$

$$\begin{aligned}
&\mathbf{\Sigma}(\boldsymbol{u}_0 r)[k + k(k-1)/2, L-1] \\
&= 2\rho^{(1)}(0) + \frac{1}{2\rho^{(1)}(0)(1 - k_1^2(\boldsymbol{u}_0 r))} 8\rho^{(1)}(r^2)\rho^{(2)}(r^2)r^2 \\
&\quad + \frac{k_4(\boldsymbol{u}_0 r)}{2\rho^{(1)}(0)(1 - k_1^2(\boldsymbol{u}_0 r))} 8\rho^{(1)}(r^2)\rho^{(2)}(r^2)r^4 \\
&= 2\rho^{(1)}(0) + 8\rho^{(1)}(r^2)r^2\rho^{(2)}(r^2)\frac{1 + k_4(\boldsymbol{u}_0 r)r^2}{2\rho^{(1)}(0)(1 - k_1^2(\boldsymbol{u}_0 r))} \\
&= 2\rho^{(1)}(0)\left(1 + k_1(\boldsymbol{u}_0 r)k_2(\boldsymbol{u}_0 r)r^2\frac{1 + k_4(\boldsymbol{u}_0 r)r^2}{1 - k_1^2(\boldsymbol{u}_0 r)}\right) \\
&= 2\rho^{(1)}(0)\left(1 + \frac{k_1(\boldsymbol{u}_0 r)k_2(\boldsymbol{u}_0 r)r^2}{1 - k_*^2(\boldsymbol{u}_0 r)}\right) \\
&= 2\rho^{(1)}(0)\left(\frac{1 - k_1(\boldsymbol{u}_0 r)k_*(\boldsymbol{u}_0 r) - k_2^2(\boldsymbol{u}_0 r)r^4}{1 - k_*^2(\boldsymbol{u}_0 r)}\right)
\end{aligned} \tag{C.46}$$

and

$$\begin{aligned}
&\mathbf{\Sigma}(\boldsymbol{u}_0 r)[k + k(k-1)/2, L] \\
&= 2\rho^{(1)}(r^2) + \frac{k_1(\boldsymbol{u}_0 r)}{2\rho^{(1)}(0)(1 - k_1^2(\boldsymbol{u}_0 r))} 8\rho^{(1)}(r^2)\rho^{(2)}(r^2)r^2 \\
&\quad + \frac{k_5(\boldsymbol{u}_0 r)}{2\rho^{(1)}(0)(1 - k_1^2(\boldsymbol{u}_0 r))} 8\rho^{(1)}(r^2)\rho^{(2)}(r^2)r^4 \\
&= 2\rho^{(1)}(r^2) + 8\rho^{(1)}(r^2)r^2\rho^{(2)}(r^2)\frac{k_1(\boldsymbol{u}_0 r) + k_5(\boldsymbol{u}_0 r)r^2}{2\rho^{(1)}(0)(1 - k_1^2(\boldsymbol{u}_0 r))} \\
&= 2\rho^{(1)}(0)k_1(\boldsymbol{u}_0 r)\left(1 + k_2(\boldsymbol{u}_0 r)r^2\frac{k_1(\boldsymbol{u}_0 r) + k_5(\boldsymbol{u}_0 r)r^2}{1 - k_1^2(\boldsymbol{u}_0 r)}\right) \\
&= 2\rho^{(1)}(0)k_1(\boldsymbol{u}_0 r)\left(1 + \frac{k_2(\boldsymbol{u}_0 r)k_*(\boldsymbol{u}_0 r)r^2}{1 - k_*^2(\boldsymbol{u}_0 r)}\right) \\
&= 2\rho^{(1)}(0)k_1(\boldsymbol{u}_0 r)\left(\frac{1 - k_1(\boldsymbol{u}_0 r)k_*(\boldsymbol{u}_0 r)}{1 - k_*^2(\boldsymbol{u}_0 r)}\right).
\end{aligned} \tag{C.47}$$