This is to certify that the
thesis entitled

ASYMPTOTIC AND COMPUTATIONAL METHODS IN SPATIAL STATISTICS

presented by

Juan Du

has been accepted towards fulfillment
of the requirements for the

Ph.D. degree in Statistics

Major Professor's Signature

6/29/09

Date

ASYMPTOTIC AND COMPUTATIONAL METHODS IN SPATIAL STATISTICS

By

Juan Du

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Statistics

2009

ABSTRACT

ASYMPTOTIC AND COMPUTATIONAL METHODS IN SPATIAL STATISTICS

By

Juan Du

The dissertation consists of two parts. The first part studies connection between fixed-domain asymptotics and the equivalence of Gaussian measures. It is stressed that one of the most important probabilistic tools to establish fixed-domain asymptotics is to use the equivalence of Gaussian measures and related criteria. Two alternative proofs are attempted to show some results about asymptoic optimality of prediction and the equivalence of Gaussian measures based on reproducing kernel Hilbert space, which may have potential power to give preferable conditions for fixed-domain asymptotics in spatial domain without constrains like stationarity of underlying processes.

The second part of the dissertation pertains to the application of covariance tapering to deal with large spatial data sets. When the spatial sample size is extremely large, as for many environmental and ecological studies, operations on the large covariance matrix are numerically challenging. Covariance tapering is a technique to alleviate these numerical challenges. We investigate how tapering affects the asymptotic efficiency of the maximum likelihood estimator (MLE) and establish asymptotic properties, particularly asymptotic distributions of the exact MLEs and tapered MLEs under the fixed-domain asymptotic framework for the Matérn model. We show that under some conditions on the tapering function, the tapered MLE is asymptotically as efficient as the true MLE for the microergodic parameter in the Matérn model. For the general setting, we compare the exact and tapered likelihood and their derivatives in seeking conditions on tapering which may yield no loss of efficiency. The convergence rate of effect of tapering on prediction is also studied. Finally, The computational gain and comparable estimation are illustrated by simulation studies and an application to the US precipitation data for April 1948.

DEDICATION

To my parents - Zihu Du and Wangxian Weng

ACKNOWLEDGMENT

I want to thank all the people who have helped and inspired me during my doctoral study.

Working with my advisors, Professor V. S. Mandrekar and Professor Hao Zhang, was the turning point of my graduate study life. I have heart-felt gratitude for Prof. Mandrekar's guidance of my study at Michigan State University. He accepted me as one of his students when I felt lost and hopeless three years ago. He helped me get through the most difficult time in my life by his integrity, his constant encouragement, his enthusiasm, his great efforts to explain things clearly and simply. What's more, he respected my research interest and made my work with his former student, Prof. Zhang, possible. That has been the greatest help that only an open- minded advisor can offer. He sets an example of a world-class researcher and teacher for me to learn in my entire life.

I want to express my deep and sincere appreciation to Prof. Zhang for his guidance in my research. I was delighted to interact with Prof. Zhang via email or telephone and to have had the chance to follow his direction of research. His insightful ideas and great expertise in both applied and theoretical statistics have inspired me and brought out my genuine interest in spatial statistics. In addition, he has been extremely patient and time-devoted in teaching and training me to be a researcher. Influenced by his righteousness, wise advice, kindness, humility, rigor and passion on research, my research life has become much smoother and rewarding for me. I will continue to pursue my academic goal following his example.

My special gratitude goes to Professor James Stapleton, who has been my mentor in my entire graduate study. With his kind assistance with writing, speaking, teaching and so on, I have become more positive about my career and stepped out of fear. He is always available to help me and cheer me up.

Professor Yimin Xiao, Professor Sarat Dass, and Professor Clifford Weil deserve

enormous thanks as my thesis committee members and advisors. This thesis was compiled using the template provided by Prof. Weil, who has kindly made this heavy work much easier. I am also grateful to Professor Connie Page, Professor Dennis Gilliland, Professor Elijah DiKong, Professor Paul Anderson, Professor Ashton Shortridge, Professor Parthanil Roy and Professor Jennifer Kaplan. I really appreciate their help in all different ways and having their valuable suggestions and comments. In particular, I would like to thank Prof. Page, Prof. Gilliland for hiring me as a consultant at CSTAT. The associated experience broadened my perspective on the practical aspects of statistics.

I am indebted to Professor Mark Meerschaert, Professor Lijian Yang, Professor Vincent Melfi, Ms. Cathy Sparks, Ms. Suzanne Watson and Mr. Eric Segur for their tremendous help and constant support.

Furthermore, Mr and Mrs. Crickmore have always been a constant source of encouragement during the past three years. They are like my spiritual father and mother who drew me close to God. I also wish to thank my sister Zuoya Du and my friend who is from my home town, Shuzhuan Zheng, for their trust and support.

Lastly, and most importantly, I wish to thank my parents, Zihu Du and Wangxian Weng. They bore me, raised me, supported me, taught me, and loved me. To them I dedicate this thesis.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Chapter 1

Introduction and Motivation

1.1 Fixed-domain asymptotics and tapering

With the advancement of technology, large amounts of data are routinely collected over space and/or time in many studies in environmental monitoring, climatology, hydrology and other fields. The large amounts of correlated data present a great challenge to the statistical analysis and may render some traditional statistical approaches impractical. For example, in maximum likelihood or Bayesian inference, the inverse of an $n \times n$ covariance matrix is involved, where the sample size n may be in hundreds of thousands or even larger. Inverting the large covariance matrix repeatedly is a great computational burden if not impractical, and some approximation to the likelihood is necessary.

Covariance tapering is one of the approaches to approximating the covariance matrix and, therefore, the likelihood. Let the second order stationary Gaussian process $X(\mathbf{t}), \mathbf{t} \in \mathbb{R}^d$ have mean 0 and an isotropic covariance function $K(h; \theta, \sigma^2)$, where σ^2 is the variance of the process and θ is the parameter that controls how fast the covariance function decays. Given n observations $\mathbf{X}_n = (X(\mathbf{t}_1), \ldots, X(\mathbf{t}_n))'$, the log-likelihood

is

$$\ell_n(\theta, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log[\det \mathbf{V}_n(\theta, \sigma^2)] - \frac{1}{2}\mathbf{X}_n'[\mathbf{V}_n(\theta, \sigma^2)]^{-1}\mathbf{X}_n, \qquad (1.1.1)$$

where $\mathbf{V}_n(\theta, \sigma^2)$ denotes the covariance matrix of $\mathbf{X}_n$. The idea of tapering is to keep the covariances approximately unchanged at small distance lags and to reduce the covariances to zero at large distances. To implement the idea, let K_{tap} be an isotropic correlation function of compact support; that is, $K_{\text{tap}}(h) = 0$ if $h \geq \gamma$ for some $\gamma > 0$. Then, the tapered covariance function $\tilde{K}$ is the product of K and K_{tap},

$$\tilde{K}(h; \theta, \sigma^2) = K(h; \theta, \sigma^2)K_{\text{tap}}(h), \qquad (1.1.2)$$

and the tapered covariance matrix is a Hadamard product $\tilde{\mathbf{V}}_n = \mathbf{V}_n(\theta, \sigma^2) \circ \mathbf{T}_n$, where $\mathbf{T}_n$ has the (i, j)th element as $K_{\text{tap}}(\|\mathbf{t}_i - \mathbf{t}_j\|)$. The tapered covariance matrix has a high proportion of zero elements and is, therefore, a sparse matrix. Inverting a sparse matrix is much more efficient computationally than inverting a regular matrix of the same dimension [see, e.g., Pissanetzky (1984), Gilbert, et al. (1992) and Davis (2006)]. One would use the tapered covariance function $\tilde{K}$ for spatial interpolation and estimation as if it was the correct covariance function. For example, the tapered maximum likelihood estimator maximizes the corresponding log-likelihood

$$\ell_{\text{n,tap}}(\theta, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log[\det \tilde{\mathbf{V}}_n] - \frac{1}{2}\mathbf{X}_n'\tilde{\mathbf{V}}_n^{-1}\mathbf{X}_n. \qquad (1.1.3)$$

Intuitively, if the taper is sufficiently close to 1 in the neighborhood of the origin, the tapering would not change the behavior of the covariance function near the origin. It has been seen that the behavior of the covariance function near the origin is most important to fixed-domain asymptotics. Stein (1988, 1990a, 1990b, 1999a and 1999b) has established rigorous fixed-domain asymptotic theory for spatial interpolation. Applying the general fixed-domain asymptotic theory, Furrer, Genton and Nychka (2006)

showed that appropriate tapering does not affect the fixed-domain asymptotic mean square error (mse) of prediction for Matérn model. In this direction, we studied further the convergence rate of comparison of mse's of prediction with and without tapering in section 2.3 and 3.2.

Kaufman, et al. (2008) showed that the parameter in the Matérn covariance function, which is consistently estimable under the fixed-domain asymptotic framework, can be estimated consistently by the tapered MLE with θ fixed. However, it has been unknown whether the covariance tapering results in any loss of asymptotic efficiency.

One of the main objectives of this thesis is to establish the asymptotic properties, and particularly the asymptotic distribution of exact and tapered MLEs under the fixed-domain asymptotic framework. We now make a few remarks about why we adopt the fixed-domain asymptotic framework and how to deal with fixed-domain asymptotic problems. When the spatial domain is fixed and bounded, more sample data can be obtained by sampling the domain increasingly densely. This results in the fixed-domain asymptotic framework. It is known that not all parameters in the covariance function are consistently estimable [e.g. Zhang (2004)] under the fixed-domain asymptotic framework.However, there is another asymptotic framework, where more data are sampled by increasing the spatial domain. This is the increasing domain asymptotic framework. Under mild regularity conditions, MLEs for all parameters are consistent and asymptotically normal [see Mardia and Marshall (1984)]. Therefore, asymptotic results are quite different under the two asymptotic frameworks. In any real application, we work with a finite sample and there is a need to know which asymptotic framework is more appropriate to be applied to the finite sample. Zhang and Zimmerman (2005) showed through both theoretical and simulation studies that, for the exponential covariance function, the fixed-domain asymptotic approximation performs at least as well as increasing domain one. Actually, it has been extensively accepted that the fix-domain asymptoics is more relevant to actual spatial data collection. In light of these results, we adopt the fixed-domain asymptotic framework in

this work.

Fixed-domain asymptotic results for estimation are difficult to derive in general because of the increasingly correlated observations and there are only few results in literature [see Stein (1990c), Ying (1991 and 1993), Chen, Simpson and Ying (2000), Zhang (2004), Loh (2005) and Kaufman, et al. (2008)]. Existing asymptotic distributions have been established only for specific models such as the exponential model for covariance functions [see Ying (1991 and 1993) and Chen, Simpson and Ying (2000)] and a particular Matérn model with the smoothness parameter $\nu = 1.5$ [see Loh (2005)]. For the general Matérn model, the fixed-domain asymptotic distribution is not available even when data are observed along a line. In order to evaluate the efficiency of the tapered MLE, we establish the fixed-domain asymptotic distribution of the MLE for the microergodic parameter (Stein (1999b), p.163), which is more important to spatial interpolation, in the general Matérn model [Theorem 2.4.3] under the assumption that data are collected along a line. This result is of interest in its own right, outside the context of tapering.

It is even more difficult to study asymptotic properties of tapered MLE. Indeed, we are not aware of any fixed-domain asymptotic distribution established for tapered MLE. For this reason, we will start with a simple model, the Ornstein-Uhlenbeck process along a line, which is a stationary Gaussian process with zero mean and an exponential covariance function, and has Markovian properties. Due to the Markovian properties, the inverse of the covariance matrix can be given in closed form and is a band matrix. Therefore, for this model, it is not necessary to approximate the likelihood function. However, this simple model serves as a starting point in the study of covariance tapering and provides insight into the more general settings, which we will study subsequently.

Although spatial data are usually collected over a spatial region, there are situations when data are collected along lines. One example is the International H2O Project, where measurements of meteorological data were collected by surface stations

and aircraft along three flight paths that are along straight lines and transect under the varied environmental conditions of the southern Great Plains [see Weckworth et al. (2004), LeMone et al. (2007) and Stassberg et al. (2008)]. Ecological data are sometimes collected along line transects as well.

One of the most important probabilistic tools to establish fixed-domain asymptotics is the equivalence and singularity of Gaussian measures and the related criteria. In the next chapter, we therefore summarize those conditions for the equivalence of Gaussian measures that will be used to study fixed-domain asymptoics later and identify the concordance between the equivalence of Gaussian measures and fixed-domain asymptoics. Even though most of the analytic approaches used so far for fixed-domain asymptotics are through spectral conditions for the equivalence of Gaussian measures, it should be preferable from a practice perspective to characterize the required conditions, such as the taper condition in spatial domain directly. With this in mind, we attempt to employ the conditions for the equivalence of Gaussian measures using reproducing kernel Hilbert space which has no constrains to stationarity of process. Two alternative proofs based on this idea are provided for some results about fixed-domain asymptoic prediction and sufficient conditions for the equivalence of two Gaussian measures.

Chapter 3 focuses on tapering techniques and fixed-domain asymptotic properties of tapered maximum likelihood estimators. We note that the condition on the tail behavior of tapering plays the essential role of maintaining the asymptotic optimality of prediction and consistency as well as efficiency of estimation using tapering. This will be specifically addressed in section 3.2. In Section 3.3 the strong consistency and asymptotic normality of tapered MLE with both parameters jointly maximized for the Ornstein-Uhlenbeck process are presented first. Then, for the microergodic parameter in a Gaussian stationary process having a Matérn covariogram, we establish the asymptotic distribution of both exact and tapered MLEs with one of the parameter chosen fixed. Finally for the general case, in view of the lack of fixed domain-

asymptotic results for MLE of joint maximization under the general setting, we will just study limiting differences between the log-likelihood and tapered log-likelihood and the difference between their derivatives. The results in section 3.4 indicate that tapered MLE and the exact MLE may have the same asymptotic distribution and therefore tapering may yield no loss of efficiency. Finally those theoretical results are supported by simulations study in two dimensional case and this suggests the possible generalization to high dimensional case of the derived asymptotic theorems in previous sections. The computational gain and accuracy of results are also demonstrated by an application of climate data. We put all proofs in the three appendices.

Chapter 2

Fixed-domain asymptotics and the equivalence of two gaussian measures

2.1 Introduction

There are two types of asymptotical framework in spatial statistics. One is the fixed-domain asymptotic framework (Stein(1999b), p.11) or infill asymptotic framework [Cressie(1993), p.350], where more data are taken ever densely in a fixed and bounded domain. The other one is increasing domain asymptotic framework, where more data are sampled by increasing the spatial domain and the minimum distance of nearby points is bounded below. This is the spatial analogue of the asymptotics usually observed in time series. Now, we explain more specifically about why we choose fixed-domain aymptotic framework over increasing domain one in our work. First, fixed domain asymptotics seem to be more relevant to the process of most of the spatial data collections. It is not hard to imagine intuitively that the more and more accurate local weather interpolation or kriging (best linear prediction for spatial process) can often be obtained based on more and more information available from increased number

of weather stations in the same region. Moreover, Stein (1999b) mentioned in his book that, fixed domain framework is the more appropriate asymptotic framework for spatial problems related to local behavior of process like interpolation.

Actually, asymptotic results are quite different under the two asymptotic frameworks, It has been mentioned earlier that not all parameters in the covariance function are consistently estimable [e.g. Zhang (2004)] under the fixed-domain asymptotic framework. Zhang and Zimmerman (2005) argued that MLEs of the microergodic parameters are generally consistent but those of the non-microergodic parameters in general converge in distribution to a non-degenerate distribution. Following Stein (1999, p. 163), we say that a function $h(\phi)$ is microergodic if, for all ϕ and $\tilde{\phi}$ in the parameter space, $h(\phi) \neq h(\tilde{\phi})$ implies that the two measures $\mathrm{P}(\phi)$ and $\mathrm{P}(\tilde{\phi})$ are orthogonal, where $\mathrm{P}(\phi)$ denotes the Gaussian measure corresponding to the parameter ϕ. We refer readers to Stein (1999b) page 163 for the details of the definition of microergodic parameters. In addition, Stein has established asymptotic results that show only the microergodic parameters affect the asymptotic mean square error under the fixed-domain asymptotic framework. In contrast, under increasing domain asymptotic framework, MLEs for all parameters are consistent and asymptotically normal given mild regularity conditions, [see Mardia and Marshall (1984)]. In practice, we only have a finite sample, one has to know which asymptotic framework is more appropriate in order to apply any asymptotic results. Zhang and Zimmerman (2005) provided some guideline on this through both theoretical and numerical studies. Their results show that, for the exponential covariance function, the fixed-domain asymptotic distribution approximates the finite sample distribution at least as well as the increasing domain asymptotic distribution does. More specifically, for microergodic parameters, approximations corresponding to the two frameworks perform about equally well. For the non-microergodic parameters, the fixed-domain asymptotic approximation is preferable.

It was first discerned by Stein in 1985 that the fixed-domain asymptotics was closely

related to the equivalence and singularity of Gaussian measures, which has become the most important mathematical tool to study the kriging [e.g. Stein(1988, 1990a, 1990b, 1999b), Furrer, Genton and Nychka(2006)] and strong consistency of maximum likelyhood estimator (MLE) [e.g. Zhang(2004), Kaufman(2008), Zhang and du(2008)] under fixed-domain aymptotic framework . In this work, we will additionally explore some sufficient conditions for the equivalence of two Gaussian measures and therefore some stronger consequent results for studying the asymptotic efficiency of exact MLE and those after using the tapering technique. In next section, after some review of the basic results for the equivalence and singularity of two Gaussian measures, I will give a detailed proof of the sufficient condition for the the equivalence of two Gaussian measures induced by two random fields observed on a fixed domain, which is a minor extension of Theorem 4 in Yadrenko(1983, P.156), where some steps of the proof was questionable, to the best of my knowledge. The application of the equivalence of Guassian measures to kriging will be covered in section 2.2, where I will provide an alternative proof of Theorem8 in Stein(1999) using Hilbert-Schmidt operator directly. In the last section of this chapter, I will present all the available fixed-domain asymptotic results of exact maximum likelihood estimators for Matérn model, which is widely used for modeling spatial data.

2.2 Equivalence and singularity of Gaussian measures

Two probability measures P_0 and P_1 are equivalent on a measurable space $\{\Omega, \mathcal{F}\}$ if $\mathrm{P}_1(A) = 0$ for any $A \in \mathcal{F}$ implies $\mathrm{P}_0(A) = 0$ and vice versa. It says that if an event occurs with probability one under one measure then it occurs with probability one under the other measure. We usually restrict the event A to the σ-algebra generated by process $\{X(\mathbf{t}), \mathbf{t} \in D\}$. We emphasize this restriction by saying that the two

measures are equivalent on the paths of $\{X(\mathbf{t}), \mathbf{t} \in D\}$, that is, P_0 and P_1 are mutually absolutely continuous on σ- algebra $\mathcal{F}$ generated by $X(\mathbf{t}), \mathbf{t} \in D$. In this work, we assume $X(\mathbf{t})$, $\mathbf{t} \in D$ is a stationary Gaussian process with mean zero unless indicated otherwise, where D is a closed set in $\mathbb{R}^d$. Let $D_n = \{\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_n\}, n = 1, 2, \ldots$ be an increasing sequence of finite subset of D such that $\cup_{n=1}^{\infty} D_n$ is dense in D. Let $\mathrm{P}_0, \mathrm{P}_1$ be equivalent Gaussian measures, denoted by $\mathrm{P}_0 \equiv \mathrm{P}_1$, for $X(\mathbf{t})$, $\mathbf{t} \in D$ with corresponding covariance function $K_0(\mathbf{h})$, $K_1(\mathbf{h})$. Accordingly, $\ell_{i,n}$ represents the log likelihood function of $\mathbf{X}_n = (X(\mathbf{t}_1), \ldots, X(\mathbf{t}_n))'$ when the process is observed in D_n under measure $\mathrm{P}_i, i = 0, 1$. In general, $X(\mathbf{t}_i), i = 1, \ldots, n$ can be assumed to be linear independent bases for both $\mathcal{H}_D(K_0)$ and $\mathcal{H}_D(K_1)$, where by $\mathcal{H}_D(K_i)$ we denote the closure of linear manifold of $X(\mathbf{t}), \mathbf{t} \in D$ with respect to norm given by second moment under K_i (see Stein 1999b, p.115). Moreover, $X(\mathbf{t}_1), X(\mathbf{t}_2), \ldots, X(\mathbf{t}_n)$ can be linearly transformed to $h_{1n}, h_{2n}, \ldots, h_{nn}$ such that for $j, k = 1, \ldots, n$, $\mathrm{E}_0(h_{kn}h_{jn}) = \delta_{kj}$ and $\mathrm{E}_1(h_{kn}h_{jn}) = \sigma_{kn}^2 \delta_{kj}$, where $\mathrm{E}_i(\cdot)$ means taking the expectation under measure P_i. Then, the the equivalence of P_0 and P_1 on $\mathcal{F}$ implies

$$\sum_{j=1}^{n} (1 - \sigma_{kn}^2)^2 < \infty, \quad \sigma_{kn}^2 \asymp 1. \tag{2.2.1}$$

[see Theorem 1 in Ibragimov and Rozanov(1978), p.77]. Here and hereafter, for two functions $a(x), b(x)$ on a domain $\tilde{D}$, we say $a(x) \asymp b(x)$, if there exist positive constants C_0, C_1 such that $C_0 \leq a(x)/b(x) \leq C_1$ for any $x \in \tilde{D}$, and $a(x) \asymp b(x), x \to \infty$, if there exists R large enough such that $a(x) \asymp b(x)$ on $(-\infty, -R) \cup (R, \infty)$. The other important fact (see e.g. Ibragimov and Rozanov(1978) p.67 and p.73) about equivalent measures that will be used in the paper is that if $\mathrm{P}_0 \equiv \mathrm{P}_1$ on $\mathcal{F}$, then Radon-Nikodym derivative or the density of these measures on $\mathcal{F}$ exists, i.e. $p(\omega) = \dfrac{P_1(d\omega)}{P_0(d\omega)} < \infty$ *a.s.*. The conditional expectation of p with respect to σ- algebra $\mathcal{F}_n$ generated by $X(\mathbf{t}_i)$, $\mathbf{t}_i \in D_n$ is the likelihood ratio $\rho_n = \dfrac{p_1(\mathbf{X}_n)}{p_0(\mathbf{X}_n)}$ where p_i is the density function of the

corresponding Gaussian distribution on the space $\mathbb{R}^d$, then

$$\lim_{n\to\infty} \rho_n = \rho \ \ a.s. \text{ and } \lim_{n\to\infty} \mathrm{E}_0(\log \rho_n) = \mathrm{E}_0(\log \rho) < \infty. \tag{2.2.2}$$

Therefore, by using covariance matrix expression $\mathbf{V}_{n,i} = \mathrm{E}_i(\mathbf{X}_n \mathbf{X}_n')$, we have

$$\begin{aligned} \log \rho_n =& \ell_n(\mathbf{X}_n) - \ell_{n,0}(\mathbf{X}_n) \\ =& -\frac{1}{2}\log\frac{\det \mathbf{V}_{1,n}}{\det \mathbf{V}_{0,n}} - \frac{1}{2}\mathbf{X}_n'(\mathbf{V}_{1,n}^{-1} - \mathbf{V}_{0,n}^{-1})\mathbf{X}_n = O(1) \quad a.s.. \end{aligned} \tag{2.2.3}$$

$$\mathrm{E}_0(\log \rho_n) = -\frac{1}{2}\log\frac{\det \mathbf{V}_{1,n}}{\det \mathbf{V}_{0,n}} - \frac{1}{2}\mathrm{E}_0(\mathbf{X}_n'(\mathbf{V}_{1,n}^{-1} - \mathbf{V}_{0,n}^{-1})\mathbf{X}_n) = O(1). \tag{2.2.4}$$

The difference of these two equations gives

$$\mathbf{X}_n'(\mathbf{V}_{1,n}^{-1} - \mathbf{V}_{0,n}^{-1})\mathbf{X}_n - \mathrm{E}_0(\mathbf{X}_n'(\mathbf{V}_{1,n}^{-1} - \mathbf{V}_{0,n}^{-1})\mathbf{X}_n) = O(1) \quad a.s.. \tag{2.2.5}$$

Actually, by using (2.2.1) and Cauchy- Schwarz inequality, we also have

$$\mathrm{E}_0(\mathbf{X}_n'(\mathbf{V}_{1,n}^{-1} - \mathbf{V}_{0,n}^{-1})\mathbf{X}_n) = O(\sqrt{n}), \tag{2.2.6}$$

because

$$\mathrm{E}_0(\mathbf{X}_n'(\mathbf{V}_{1,n}^{-1} - \mathbf{V}_{0,n}^{-1})\mathbf{X}_n) = \mathrm{E}_0(\mathbf{h}_n'(\mathrm{diag}\{\frac{1}{\sigma_{k,n}^2}, k = 1, \ldots, n\} - \mathbf{I}_n)\mathbf{h}_n) \tag{2.2.7}$$

$$= \sum_{k=1}^{n}(\frac{1}{\sigma_{k,n}^2} - 1) \asymp \sum_{k=1}^{n}(1 - \sigma_{k,n}^2) \tag{2.2.8}$$

and $(\sum_{k=1}^{n}(1-\sigma_{k,n}^2))^2 \leq n\sum_{k=1}^{n}(1-\sigma_{k,n}^2)^2$. We will use these results with $\mathbf{V}_{0,n}, \mathbf{V}_{1,n}$ replaced with corresponding ones at different places in the proof of related theorems.

Sometimes, it is more amendable if we work on the isomorphic spectral space of $\mathcal{H}_D(K_i)$ to solve some problems originally in spatial domain. Specifically, let F_i and

f_i be the spectral measure and spectral density of process $X(\mathbf{t}), \mathbf{t} \in D$ corresponding to K_i, then closure of manifold of functions of the form $\exp(i\lambda'\mathbf{t}), \mathbf{t} \in D$ under inner product defined as $\langle \varphi, \psi \rangle_{F_i} = \int \varphi \bar{\psi} dF_i$ gives the isomorphic space of $\mathcal{H}_D(K_i)$, denoted by $L_D^2(F_i)$, by extending the correspondence η between $\sum c_i \exp(i\lambda'\mathbf{t}_i)$ and $\sum c_i X(\mathbf{t}_i)$ to their limits. If we define the operator A on Hilbert space $L_D^2(F_0)$ into space $L_D^2(F_1)$ by $A\varphi(\lambda) = \varphi(\lambda)$ for all $\varphi(\lambda) \in L_D^2(F_0)$, then we have the following well-known theorem:

Theorem 2.2.1. *[see Theorem 4 in Ibragimov and Rozanov(1978), p.80] Under the condition that* $\|\varphi\|_{F_0} \asymp \|\varphi\|_{F_1}$ *for any* $\varphi \in L_D^2(F_0)$, $\mathrm{P}_0 \equiv \mathrm{P}_1$ *if and only if* $\Delta = E - A^*A$ *is Hilbert-Schmidt operator, where* A^* *is the adjoint operator of* A *and* E *is the identity operator.*

Based on this fundamental theorem, the following one-dimensional results Theorem 2.2.2 $\sim$ 2.2.5 given in Ibragimov and Rozanov(1978) have been practical in studying fixed-domain asymptotics for the equivalence of two Gaussian measures.

Consider the case where the spectral measures $F(d\lambda)$ and $F_1(d\lambda)$ have the bounded densities

$$f_0(\lambda) = F(d\lambda)/d\lambda \quad \text{and} \quad f_1(\lambda) = F_1(d\lambda)/d\lambda$$

Theorem 2.2.2. *A necessary and sufficient condition for the equivalence of the Gaussian measures* P_0 *and* P_1 *(with zero mean values) are equivalent on the* σ*-algebra* $\mathcal{F}$ *generated by* $\{X(t), t \in D \subset \mathbb{R}\}$ *if and only if the difference between the covariance functions*

$$b(s,t) = \mathrm{E}_0(X(s)X(t)) - \mathrm{E}_1(X(s)X(t)), \quad s,t \in D$$

can be extendable to a square-integrable function $b(s,t)$ *in the entire plane* $-\infty < s,t < \infty$, *whose Fourier transform*

$$\varphi(\lambda,\nu) = \frac{1}{4\pi^2} \iint e^{i(\lambda s - \nu t)} b(s,t) ds dt$$

satisfies the condition

$$\iint \frac{|\varphi(\lambda,\nu)|^2}{f_0(\lambda)f_1(\nu)}d\lambda\nu < \infty. \tag{2.2.9}$$

Under the condition about spectral density

$$f(\lambda) \geq k(1+\lambda^2)^{-n}, \tag{2.2.10}$$

the space $L_D(F)$ consists of integral analytic functions and therefore condition (2.2.9) can be expressed by only one spectral density function. That is:

Theorem 2.2.3. *Under condition (2.2.10), a necessary and sufficient condition for the equivalence of the Gaussian measures* P_0 *and* P_1 *on the* σ*-algebra* $\mathcal{F}$ *is that the difference between the covariance functions* $b(s,t)$, $s,t \in D$ *which is a finite interval, be extendable to a square-integrable function* $b(s,t), -\infty < s,t < \infty$, *whose Fourier transform satisfies the condition*

$$\iint \frac{|\varphi(\lambda,\nu)|^2}{f_0(\lambda)f_0(\nu)}\, d\lambda\nu < \infty. \tag{2.2.11}$$

Based on the fact that the equivalence of two Gaussian measures under condition (2.2.10) depends only on the high frequency behavior, we have the following strengthened necessary and sufficient condition.

Theorem 2.2.4. *For the spectral density* $f_0(\lambda)$ *of the type given by*

$$\liminf_{\lambda\to\infty} f(\lambda)\,|\lambda|^{2n} > 0,$$

a sufficient and necessary condition for the equivalence of the Gaussian measures P_0 *and* P_1 *on the* σ*-algebra* $\mathcal{F}$ *is that the difference between their covariance functions* $b(s,t), s,t \in D$ *(D is any finite interval), be extendable to a square-integrable function*

$b(s,t), -\infty < s,t < \infty$, *whose Fourier transform* $\varphi(\lambda,\mu)$ *satisfies*

$$\int_{|\lambda|>R}\int_{|\mu|>R} \frac{|\varphi(\lambda,\nu)|^2}{f_0(\lambda)f_0(\nu)}\, d\lambda\nu < \infty. \tag{2.2.12}$$

for any R.

In particular, if the spectral density f_0 is such that

$$f_0(\lambda) \asymp |\varphi(\lambda)|^2 \tag{2.2.13}$$

for sufficiently large $|\lambda|$, where $\varphi(\lambda)$ is the Fourier transform of some square integrable finite function, then there is a sufficient condition which is relatively easy to verify in practice.

Theorem 2.2.5. *For spectral densities of the type given by (2.2.13), a sufficient and necessary condition for the equivalence of the Gaussian measures* P_0 *and* P_1 *on the pathes of* $\{X(t), t \in D\}$, *where* D *is any finite interval, is that the function*

$$h = \frac{f_0(\lambda) - f_1(\lambda)}{f_0(\lambda)}$$

satisfies the condition

$$\int_{|\lambda|>R} |h(\lambda)|^2\, d\lambda < \infty, \tag{2.2.14}$$

for any $R < \infty$.

Another method to find out the sufficient and necessary conditions for the equivalence of Gaussian measures is to use reproducing kernel Hilbert space. This approach has no constrains like stationarity or isotropy on the underlying process. Therefore it could be a potential tool for analyzing nonstationary processes. We will review some important results given by Chatterji and Mandrekar (1978), it is worth noting that those results apply to any dimensional case:

Definition 2.2.6. *Let D be any set and K be a real-valued covariance on $D \times D$. Then K is called a covariance on D if (a) $K(s,t) = K(t,s)$ for any $s,t \in D$ and (b) $\sum_{t,s\in I} a_s a_t K(s,t) \geq 0$ for all finite subsets I of D and $\{a_s, s \in I\}$ of $\mathbb{R}$.*

Definition 2.2.7. *Let D be any set. A Class $\mathcal{K}(K)$ of functions on D forming a Hilbert space is called the reproducing kernel Hilbert space (for short, rkhs) for a covariance K.*

The following theorem gives existence and uniqueness of rkhs.

Theorem 2.2.8. *[Aronszajn(1950)] Let D be any set and K be a real-valued function on $D \times D$. Then there exists a unique Hilbert space $\mathcal{K}(K)$ of functions on D, satisfying*

$$K(\cdot, t) \in \mathcal{K}(K) \quad \text{for each} \quad t \in D;$$
$$(f, K(\cdot, t)) = f(t) \quad \text{for each} \quad f \in \mathcal{K}(K) \quad \text{and} \quad t \in D.$$

To give the sufficient and necessary condition for the equivalence of two Gaussian measures, we need introduce a "product" covariance function which is defined as $K_0 \bigotimes K_1\{[(t_1,t_2),(s_1,s_2)]\} = K_0(t_1,s_1)K_1(t_2,s_2)$, where K_0 is a covariance function on D_1 and K_1 is a covariance function on D_2. We will present the following two theorems in a more general setting. Let D be an index set. Let $\{X(t), t \in D\}$ be a family of real random variables on a measurable space $(\Omega, \mathcal{A})$ and $\mathcal{F} = \sigma\{X(t), t \in D\}$. Suppose P_0, P_1 are two measures on $\mathcal{F}$ such that $\{X(t), t \in D\}$ is a Gaussian process on $(\Omega, \mathcal{F}, \mathrm{P}_i), i = 0, 1$. Denote by $K_0(t,s) = \mathrm{E}_0(X(t) - m_0(t))(X(s) - m_0(s))$ and $K_1(t,s) = \mathrm{E}_1(X(t) - m_1(t))(X(s) - m_1(s))$, where $\mathrm{E}_0(X(t)) = m_0(t)$ and $\mathrm{E}_1(X(t)) = m_1(t)$.

Theorem 2.2.9. *The following are equivalent.*

(a) $\mathrm{P}_0 \equiv \mathrm{P}_1$ *on* $\mathcal{F}$.

(b) $K_0 - K_1 \in \mathcal{K}(K_1 \bigotimes K_0)$ *and* $m_1 - m_0 \in \mathcal{K}(K_0)$.

(c) (i) *There exists* γ_1, γ_2 $(0 < \gamma_1 \leq \gamma_2 < \infty)$ *such that* $\gamma_1 K_0 \ll K_1 \ll \gamma_2 K_0$. (ii) $K_0 - K_1 \in \mathcal{K}(K_0 \bigotimes K_0)$; *and* (iii) $m_0 - m_1 \in \mathcal{K}(K_0)$.

Where $\gamma_1 K_0 \ll K_1$ *means that* $K_1 - \gamma_1 K_0$ *is a covariance.*

By virtue of rkhs, we can obtain the analogue of Theorem 2.2.1 in spatial domain. Let $\mathcal{H}_D(m_i, K_i)$ be as defined before, which is the completion of linear manifold of $\{X(\mathbf{t}), \mathbf{t} \in D\}$ with respect to the inner product given by the corresponding second-order structure (m_i, K_i), $i = 0, 1$ here. Let Λ be the bounded linear operator on $\mathcal{H}_D(m_0, K_0)$ into $\mathcal{H}_D(m_1, K_1)$ such that $\Lambda h_t = h_t$, for any h_t as linear combination of $\{X(\mathbf{t}), \mathbf{t} \in D\}$.

Theorem 2.2.10. $\mathrm{P}_0 \equiv \mathrm{P}_1$ *on* $\mathcal{F}$ *if and only if that*

(a) Λ *is one-one bounded, with bounded inverse on* $\mathcal{H}_D(m_0, K_0)$ *into* $\mathcal{H}_D(m_1, K_1)$

(b) $(I - \Lambda^*\Lambda)$ *is Hilbert-Schmidt, where* I *is identity on* $\mathcal{H}_D(m_0, K_0)$.

(c) $m_1 - m_0 \in \mathcal{K}(K_0)$.

Since those conditions for the equivalence of Gaussian measures based rkhs has no constrains on the set D, we can extend Theorem 2.2.3 and Theorem 2.2.5 to high dimensional case. We note that some slightly weaker results presented in Yadrenko(1983, p.154 and p.156), but we believe our proofs are more straightforward and clearer in the context of this thesis.

Let $\{X_{\mathbf{t}}, \mathbf{t} \in D\}$ be centered Gaussian random field, with covariance function K_j and spectral measure F_j with spectral density f_j under corresponding measures $\mathrm{P}_j, j = 0, 1$, where $D \subset \mathbb{R}^d$ is bounded. Let $b(\mathbf{s}, \mathbf{t}) = K_0(\mathbf{s}, \mathbf{t}) - K_1(\mathbf{s}, \mathbf{t})$, $\mathbf{s}, \mathbf{t} \in D$, as is well known that $L_j(s, t) = \int e^{i\boldsymbol{\lambda}'(\mathbf{s}-\mathbf{t})} f_j(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}$. Suppose there exist constants γ_1, γ_2 such that $\gamma_1 K_0 \ll K_1 \ll \gamma_2 K_0$.

Theorem 2.2.11. *A necessary and sufficient condition for the equivalence of the Gaussian measures* P_0 *and* P_1 *is that the covariance difference* $b(\mathbf{s}, \mathbf{t}), \mathbf{s}, \mathbf{t} \in D$ *can be*

extended to a square integrable function $b(\mathbf{s},\mathbf{t})$ *on* $\mathbb{R}^d \times \mathbb{R}^d$ *and the Fourier transform* φ *of which satisfies*

$$\int_{\mathbb{R}^d}\int_{\mathbb{R}^d} \frac{|\varphi(\boldsymbol{\lambda},\boldsymbol{\mu})|}{f_0(\boldsymbol{\lambda})f_0(\boldsymbol{\mu})}\, d\boldsymbol{\lambda}\, d\boldsymbol{\mu} < \infty. \tag{2.2.15}$$

Proof. If follows from the Theorem 2.2.9(c) [Chatterji and Mandrekar(1978), p.183]. that $\mathrm{P}_0 \equiv \mathrm{P}_1$ is equivalent to

$$b(\mathbf{s},\mathbf{t}) \in \mathcal{K}(K_0 \otimes K_0),$$

where the reproducing kernel Hilbert space

$$\mathcal{K}(K_0 \otimes K_0) = \left\{\hat{\varphi},\ \hat{\varphi} = \iint e^{-i(\mathbf{s}'\boldsymbol{\lambda}+\mathbf{t}'\boldsymbol{\nu})}\varphi(\boldsymbol{\lambda},\boldsymbol{\nu})\, dF_0(\boldsymbol{\lambda})\, dF_0(\boldsymbol{\mu}),\ \varphi \in L^2(F_0 \times F_0)\right\},$$

which follows from Theorem 2.2.8. Because

$$K_0 \otimes K_0((\mathbf{s},\mathbf{t}),(\mathbf{s}_1,\mathbf{t}_1)) = K_0(\mathbf{s}-\mathbf{s}_1)K_0(\mathbf{t}-\mathbf{t}_1) = \iint e^{i(\mathbf{t}-\mathbf{t}_1)'\boldsymbol{\lambda}+i(\mathbf{s}-\mathbf{s}_1)'\boldsymbol{\mu}}\, dF(\boldsymbol{\lambda})\, dF(\boldsymbol{\mu}),$$

and for any $\hat{\varphi} \in \mathcal{K}(K_0 \otimes K_0)$, there exists $\varphi \in L^2(F_0 \times F_0)$ such that

$$\begin{aligned}
&\langle \hat{\varphi}, K_0 \otimes K_0((\mathbf{s},\mathbf{t}),\cdot)\rangle \\
&= \iiiint e^{-i(\mathbf{t}_1{}'\boldsymbol{\lambda}+\mathbf{s}_1{}'\boldsymbol{\mu})} e^{-i(\mathbf{t}-\mathbf{t}_1)'\boldsymbol{\lambda}-i(\mathbf{s}-\mathbf{s}_1)'\boldsymbol{\mu}}\varphi(\boldsymbol{\lambda},\boldsymbol{\mu})\, dF(\boldsymbol{\lambda})\, dF(\boldsymbol{\mu})\, d\mathbf{s}_1\, d\mathbf{t}_1 \\
&= \hat{\varphi}(\mathbf{s},\mathbf{t}).
\end{aligned}$$

□

We will employ the following well-known properties of Fourier transform. For any square integrable functions (with respect to Lebesgue measure) $\varphi_j(\boldsymbol{\lambda})$, $\boldsymbol{\lambda} \in \mathbb{R}^d$, there are square integrable functions $a_j(\mathbf{t}), \mathbf{t} \in \mathbb{R}^d$ such that

$$\varphi_j(\boldsymbol{\lambda}) = \int_{\mathbb{R}^d} \exp(-i\boldsymbol{\lambda}'\mathbf{t})a_j(\mathbf{t})\, d\mathbf{t},\ j = 1,2.$$

Furthermore,

$$\varphi_1(\boldsymbol{\lambda})\varphi_2(\boldsymbol{\lambda}) = \int_{\mathbb{R}^d} \exp(-i\boldsymbol{\lambda}'\mathbf{t})(a_1 * a_2)(\mathbf{t})\, d\mathbf{t} \tag{2.2.16}$$

$$\int_{\mathbb{R}^d} \exp(i\boldsymbol{\lambda}'\mathbf{t})\varphi_1(\boldsymbol{\lambda})\varphi_2(\boldsymbol{\lambda})\, d\boldsymbol{\lambda} = (2\pi)^d (a_1 * a_2)(\mathbf{t}), \tag{2.2.17}$$

where all the equalities are in the $L^2(\, d\boldsymbol{\lambda})$ sense, and $a_1 * a_2$ is the convolution, i.e.,

$$a_1 * a_2(\mathbf{t}) = \int_{\mathbb{R}^d} a_1(\mathbf{s}) a_2(\mathbf{t} - \mathbf{s})\, d\mathbf{s}.$$

Theorem 2.2.12. *Suppose*

$$0 < f_0(\boldsymbol{\lambda}) \asymp |\varphi(\boldsymbol{\lambda})|^2, \quad |\boldsymbol{\lambda}| \to \infty, \tag{2.2.18}$$

where φ is the Fourier transform of some square integrable finite function identically zero outside bounded set T. Let

$$h(\boldsymbol{\lambda}) = \frac{f_1(\boldsymbol{\lambda}) - f_0(\boldsymbol{\lambda})}{f_0(\boldsymbol{\lambda})}$$

satisfy the condition, for some $M > 0$

$$\int_{|\boldsymbol{\lambda}|>M} |h(\boldsymbol{\lambda})|^2\, d\boldsymbol{\lambda} < \infty. \tag{2.2.19}$$

Then $\mathrm{P}_0 \equiv \mathrm{P}_1$.

Proof. When the function $h(\boldsymbol{\lambda})$ is square integrable and $f_0(\boldsymbol{\lambda}) \asymp |\varphi(\boldsymbol{\lambda})|^2$, where $\varphi = \int e^{i\mathbf{t}'\boldsymbol{\lambda}} c(\boldsymbol{\lambda})\, d\boldsymbol{\lambda}$ and $c(\boldsymbol{\lambda}) = 0$ when $\boldsymbol{\lambda}$ lies out of bounded set T. Then Plancherel's Theorem (see,e.g. Yosida (1968), p.153) implies there exists square integrable function

$a(\mathbf{t})$ such that $h(\boldsymbol{\lambda}) = \int e^{-i\boldsymbol{\lambda}'\mathbf{t}} a(\mathbf{t})\, d\mathbf{t}$. Furthermore, for $\mathbf{s}, \mathbf{t} \in D \times D$,

$$b(\mathbf{s},\mathbf{t}) = \int e^{i\boldsymbol{\lambda}'(\mathbf{s}-\mathbf{t})}(f_1(\boldsymbol{\lambda}) - f_0(\boldsymbol{\lambda}))\, d\boldsymbol{\lambda} = \int e^{i\boldsymbol{\lambda}'(\mathbf{s}-\mathbf{t})} h(\boldsymbol{\lambda})\, |\varphi(\boldsymbol{\lambda})|^2\, d\boldsymbol{\lambda}$$

By (3.3.83),

$$|\varphi(\boldsymbol{\lambda})|^2 = \int \exp(-i\boldsymbol{\lambda}'\mathbf{t}) \left(\int c(\mathbf{z})\overline{c(\mathbf{z}-\mathbf{t})}\, d\mathbf{z} \right) d\mathbf{t},$$

Applying (3.3.84) to $b(\boldsymbol{\lambda})$ and $|\varphi(\boldsymbol{\lambda})|^2$, we get $(\mathbf{s},\mathbf{t}) \in D \times D$

$$\begin{aligned} b(\mathbf{s},\mathbf{t}) &= (2\pi)^d \int_{\mathbb{R}^d} a(\mathbf{w}) \int_{\mathbb{R}^d} c(\mathbf{z})\overline{c(-(\mathbf{s}-\mathbf{t}-\mathbf{w}-\mathbf{z}))}\, d\mathbf{z} d\mathbf{w} \\ &= (2\pi)^d \int_{\mathbb{R}^d \times \mathbb{R}^d} a(\mathbf{u}-\mathbf{v}) c(\mathbf{s}-\mathbf{u})\overline{c(\mathbf{t}-\mathbf{v})}\, d\mathbf{u}\, d\mathbf{v}. \end{aligned} \tag{2.2.20}$$

The finite functions $c(\mathbf{s}-\mathbf{u}), \overline{c(\mathbf{t}-\mathbf{v})}$ vanish if the variables $\mathbf{u}$ and $\mathbf{v}$ lie outside of a compact set $T' \subset \mathbb{R}^d$, so that

$$b(\mathbf{s},\mathbf{t}) = (2\pi)^d \int_{T' \times T'} a(\mathbf{u}-\mathbf{v}) c(\mathbf{s}-\mathbf{u})\overline{c(\mathbf{t}-\mathbf{v})}\, d\mathbf{u}\, d\mathbf{v}.$$

We can choose an extension $a(\mathbf{u},\mathbf{v})$, $(\mathbf{u},\mathbf{v}) \in \mathbb{R}^d \times \mathbb{R}^d$, of the function $a(\mathbf{u}-\mathbf{v}), \mathbf{u},\mathbf{v} \in T'$, such that the function $a(\mathbf{u},\mathbf{v})$ is square integrable and denote $\psi(\mathbf{u},\mathbf{v})$ the Fourier transform of the function $a(\mathbf{u},\mathbf{v})$. We can define function $b(\mathbf{s},\mathbf{t}), \mathbf{s},\mathbf{t} \in \mathbb{R}^d \times \mathbb{R}^d$,

$$b(\mathbf{s},\mathbf{t}) = (2\pi)^d \int_{\mathbb{R}^d \times \mathbb{R}^d} a(\mathbf{u},\mathbf{v}) c(\mathbf{s}-\mathbf{u})\overline{c(\mathbf{t}-\mathbf{v})}\, d\mathbf{u}\, d\mathbf{v}.$$

which is extension of (3.3.86). Its Fourier transform is $\psi(\mathbf{u},\mathbf{v})\varphi(\mathbf{u})\overline{\varphi}(\mathbf{v})$ and

$$\iint \frac{|\psi(\mathbf{u},\mathbf{v})\varphi(\mathbf{u})\overline{\varphi}(\mathbf{v})|^2}{f_0(\mathbf{u}) f_0(\mathbf{v})}\, d\mathbf{u}\, d\mathbf{v} = \iint |\psi(\mathbf{u},\mathbf{v})|^2\, d\mathbf{u}\, d\mathbf{v} < \infty.$$

It follows from Theorem 1 that $\mathrm{P}_0 \equiv \mathrm{P}_1$ in this case.

Secondly, consider an arbitrary $f(\boldsymbol{\lambda})$ of the type (2.2.18) and $f_1(\boldsymbol{\lambda}) > f_0(\boldsymbol{\lambda})$. Let

$\tilde{f}_0(\boldsymbol{\lambda}) = |\varphi(\boldsymbol{\lambda})|^2$, then $\tilde{f}_1(\boldsymbol{\lambda}) = \tilde{f}_0(\boldsymbol{\lambda}) + (f_1(\boldsymbol{\lambda}) - f_0(\boldsymbol{\lambda}))$. From the result in first case, we know there exists extension of the difference $b(\mathbf{s}, \mathbf{t})$ such that corresponding Fourier transform $\phi(\boldsymbol{\lambda}, \boldsymbol{\mu})$ satisfies

$$\iint \frac{|\phi(\boldsymbol{\lambda}, \boldsymbol{\mu})|^2}{\tilde{f}_0(\boldsymbol{\lambda})\tilde{f}_0(\boldsymbol{\mu})}\, d\boldsymbol{\lambda}\, d\boldsymbol{\mu} < \infty. \tag{2.2.21}$$

Note that

$$b(\mathbf{s}, \mathbf{t}) = \int e^{i\boldsymbol{\lambda}'(\mathbf{s}-\mathbf{t})} \left| \tilde{f}_1(\boldsymbol{\lambda}) - \tilde{f}_0(\boldsymbol{\lambda}) \right| d\boldsymbol{\lambda} = \int e^{i\boldsymbol{\lambda}'(\mathbf{s}-\mathbf{t})} |f_1(\boldsymbol{\lambda}) - f_0(\boldsymbol{\lambda})|\, d\boldsymbol{\lambda}$$

For large enough M_1, there exists γ_3 such that $\tilde{f}_0(\boldsymbol{\lambda}) < \gamma_3 f_0(\boldsymbol{\lambda}), |\boldsymbol{\lambda}| > M_1$, this together with (2.2.21) gives

$$\int_{|\boldsymbol{\lambda}| \geq M_1} \int_{|\boldsymbol{\mu}| \geq M_1} \frac{|\phi(\boldsymbol{\lambda}, \boldsymbol{\mu})|^2}{f_0(\boldsymbol{\lambda}) f_0(\boldsymbol{\mu})}\, d\boldsymbol{\lambda}\, d\boldsymbol{\mu} < \infty,$$

Hence $\mathrm{P}_0 \equiv \mathrm{P}_1$ in this case from Theorem 1.

Finally, for any $f(\boldsymbol{\lambda})$ of the type (2.2.18), let $f_2(\boldsymbol{\lambda}) = f_0(\boldsymbol{\lambda}) + \max\{0, f_1(\boldsymbol{\lambda}) - f_0(\boldsymbol{\lambda})\}$. It is observed that $f_2(\boldsymbol{\lambda}) \geq f_0(\boldsymbol{\lambda}), f_2(\boldsymbol{\lambda}) \geq f_1(\boldsymbol{\lambda})$ and

$$\int_{|\boldsymbol{\lambda}| > M_2} \left(\frac{f_2(\boldsymbol{\lambda}) - f_0(\boldsymbol{\lambda})}{f_0(\boldsymbol{\lambda})} \right)^2 d\boldsymbol{\lambda} < \infty, \qquad \int_{|\boldsymbol{\lambda}| > M_2} \left(\frac{f_2(\boldsymbol{\lambda}) - f_1(\boldsymbol{\lambda})}{f_1(\boldsymbol{\lambda})} \right)^2 d\boldsymbol{\lambda} < \infty,$$

for some $M_2 > 0$. Let P_2 be the measure induced by Gaussian homogenous random field on D with the corresponding spectral density f_2. Based on the results for the second case, we have $\mathrm{P}_0 \equiv \mathrm{P}_2$ and $\mathrm{P}_0 \equiv \mathrm{P}_2$. This implies $\mathrm{P}_0 \equiv \mathrm{P}_1$. The proof is completed. □

In the last part of this section, I will particularly address the the equivalence of Gaussian measures for Matérn model. It means that the underlying process $\{X(\mathbf{t})\}$

possesses the following isotropic Matérn covariance function

$$K(h;\sigma^2,\theta,\nu) = \frac{\sigma^2(\theta h)^\nu}{\Gamma(\nu)2^{\nu-1}}\mathcal{K}_\nu(\theta h), \quad h > 0, \tag{2.2.22}$$

with parameters σ^2, θ and ν, where $\mathcal{K}_\nu$ is the modified Bessel function of order ν [see Abramowitz and Stegun (1967) p.375-376], σ^2 is the covariance parameter, θ is the scale parameter and ν is the smoothness parameter. This covariance structure has been widely used in practice for modeling spatial data. Due to parameter ν, Matérn covariagram has great flexibility to model a large variety of Gaussian processes with different degree of smoothness. In particular, it includes exponential covariance $K(\mathbf{s},\mathbf{t}) = \sigma^2 \exp\{-\theta\,|\mathbf{t}-\mathbf{s}|\}$ as a special case with ν equal $1/2$ and Gaussian covariance function $K(\mathbf{s},\mathbf{t}) = \sigma^2 \exp\{-\theta\,|\mathbf{t}-\mathbf{s}|^2\}$ when $\nu \to \infty$. In addition, Matérn model is mathematically amendable largely because it possesses a rational spectral density as in the following

$$f(\boldsymbol{\lambda};\sigma^2,\theta) = \frac{\Gamma(\nu+d/2)}{\Gamma(\nu)\pi^{d/2}}\frac{\sigma^2\theta^{2\nu}}{(\theta^2+\|\boldsymbol{\lambda}\|^2)^{\nu+d/2}}. \tag{2.2.23}$$

Zhang(2004) showed

Theorem 2.2.13. *Let P_i be probability measure under which $X(\mathbf{t}), \mathbf{t} \in \mathbb{R}^d$ is stationary Gaussian with mean 0 and an isotropic Matérn covariogram (it is a term used in geostatistics for covariance) with a variance σ_i^2, a scale parameter $\alpha_i, i = 1,2$, and the same smoothness parameter ν, where $d = 1, 2$, or 3. For any bounded infinite set $D \subset \mathbb{R}^d$, $\mathrm{P}_1 \equiv \mathrm{P}_2$ on the paths of $X(\mathbf{t}), \mathbf{t} \in D$, if and only if $\sigma_0^2\theta_0^{2\nu} = \sigma_1^2\theta_1^{2\nu}$.*

This theorem implies that for Matérn model, neither σ^2 nor θ is consistently estimable, but the quantity $\sigma^2\theta^{2\nu}$ is (Zhang (2004)). We will apply the results in this section to spatial interpolation in next section and parameter estimation later.

2.3 Equivalence of Gaussian measures and kriging

The best linear unbiased prediction for spatial data is often called kriging, or spatial interpolation. If two covariograms define two equivalent Gaussian measures, the interpolation under the two covariograms will be asymptotically equal. This can be seen from some well established probability results. Blackwell and Dubins (1962) derived the following results. Let two probability measures P_0 and P_1 be equivalent on $X(\mathbf{t}), \mathbf{t} \in D$ and $\mathbf{t}_k, k = 1, \ldots, n$ be the sampling sites in D, and $\mathbf{t}_k, k = n+1, n+2, \ldots,$ the sites in D where spatial interpolation is needed. Write $X_k = X(\mathbf{t}_k)$.

$$\sup |\, \mathrm{P}_1(A|X_1, \ldots, X_n) - \mathrm{P}_0(A|X_1, \ldots, X_n)| \to 0, \text{ as } n \to \infty$$

where the supremum is taken over all $A \in \sigma(X_k, k > n)$. In particular,

$$\begin{aligned} &\sup_{k>n, B} |\, \mathrm{P}_1(X_k \in B|X_1, \ldots, X_n) - \mathrm{P}_0(X_k \in B|X_1, \ldots, X_n)| \\ &\to 0, \text{ as } n \to \infty. \end{aligned} \tag{2.3.1}$$

Therefore, the predictive distribution of X_k for any $k > n$ given $X_1, \ldots, X_n$ would be nearly the same for sufficiently large n. This asymptotic optimality of prediction based on misspecified covariance structure is also studied systematically by Stein (e.g., 1990a,1999b). In the following, we will briefly review some of the results and give an alternative proof for the first basic theorem (Theorem 8 in Stein(1999b))using Hilbert-Schmidt operator.

Let the underlying process $X(\mathbf{t})$ be stationary with mean zero and be observed at locations $\mathbf{t}_k, k = 1, 2 \ldots,$ in a bounded region $D \subset \mathbb{R}^d$. Assume the set of sampling sites $\{\mathbf{t}_k, k = 1, 2, \ldots\}$ is dense in D. For any sample size n and any site $\mathbf{t} \neq \mathbf{t}_k$ for any k, let $\hat{X}_i(\sigma^2, n)$ be the best linear prediction based on $X(\mathbf{t}_1), \ldots, X(\mathbf{t}_n)$ under

the covariance function $K_i(\mathbf{h})$, $i = 0, 1$ and write the corresponding prediction error

$$e_i(\mathbf{t}, n) = X(\mathbf{t}) - \hat{X}_i(\mathbf{t}, n). \tag{2.3.2}$$

In general, for any $h \in \mathcal{H}(K_0)$, let $e_i(h, n) = h - \mathrm{E}(h|X_1, \ldots, X_n)$ be prediction error for h. If we regard P_0 as the measure corresponding to the true covariance function and P_1 the measure corresponding to misspecified covariance function. The following theorem [Theorem8 in Stein(1999b)] states that as long as these two covariance structures specify equivalent Gaussian measures, the misspecified covariance K_1 will still retain the asymptotical optimality of prediction in terms of quality of point prediction and accuracy of assessments of mean square errors. Suppose $(0, K_0)$ and (m_1, K_1) are two possible second-order structures on $\mathcal{H}_\prime$, the linear manifold generated by $\{X(\mathbf{t}), \mathbf{t} \in D\}$. Define $\mathcal{H}(0, K)$, $\mathcal{H}(m, K)$ as in section 2.2. Take $\psi_1, \psi_2, \ldots$ to be the Gram-Schmidt orthogonalization of $h_1, h_2, \ldots$ under $(0, K_0)$ so that $K_0(\psi_j, \psi_k) = \delta_{jk}$. Define operator Λ as in Theorem 2.2.10. Based on this theorem, we will give an alternative proof here.

Theorem 2.3.1. *If* $\mathrm{P}_0 \equiv \mathrm{P}_1$, *let* $\mathcal{H}_{-n}$ *be made up of elements* h *of* $\mathcal{H}(0, K_0)$ *for which* $\mathrm{E}_0\, e_0(\psi, n)^2 > 0$. *Then*

$$\lim_{n\to\infty} \sup_{\psi\in\mathcal{H}_{-n}} \left| \frac{\mathrm{E}_1\, e_0(\psi, n)^2 - \mathrm{E}_0\, e_0(\psi, n)^2}{\mathrm{E}_0\, e_0(\psi, n)^2} \right| = 0 \tag{2.3.3}$$

$$\lim_{n\to\infty} \sup_{\psi\in\mathcal{H}_{-n}} \left| \frac{\mathrm{E}_0\, e_1(\psi, n)^2 - \mathrm{E}_1\, e_1(\psi, n)^2}{\mathrm{E}_1\, e_1(\psi, n)^2} \right| = 0 \tag{2.3.4}$$

$$\lim_{n\to\infty} \sup_{\psi\in\mathcal{H}_{-n}} \frac{\mathrm{E}_0\, e_1(\psi, n)^2 - \mathrm{E}_0\, e_0(\psi, n)^2}{\mathrm{E}_0\, e_0(\psi, n)^2} = 0 \tag{2.3.5}$$

$$\lim_{n\to\infty} \sup_{\psi\in\mathcal{H}_{-n}} \frac{\mathrm{E}_1\, e_0(\psi, n)^2 - \mathrm{E}_1\, e_1(\psi, n)^2}{\mathrm{E}_1\, e_1(\psi, n)^2} = 0 \tag{2.3.6}$$

Proof. It follows from Theorem 2.2.10 that $\mathrm{P}_0 \equiv \mathrm{P}_1$ implies $(I - \Lambda^*\lambda)$ is Hilbert-

Schmidt and $m_1 \in \mathcal{K}(K_0)$. Therefore

$$\sum_{j=1}^{\infty} \|(I - \Lambda^*\Lambda)\psi_j\|_0^2 < \infty \quad \text{and} \quad \sum_{j=1}^{\infty} m_{1j}^2 < \infty. \tag{2.3.7}$$

Where $m_{1j} = \mathrm{E}_1(\psi_j)$. For $\psi \in \mathcal{H}(0, K_0)$,we can write $\psi = \sum_{j=1}^{\infty} c_j\psi_j$,where $\sum c_j^2 < \infty$. Then the error of the best linear prediction for ψ given $\mathcal{H}_n(0, K_0)$ is

$$e_0(\psi, n) = \sum_{j=n+1}^{\infty} c_j\psi_j.$$

If $\mathrm{E}_0\, e_0(\psi, n)^2 > 0$,then as $n \to \infty$

$$\begin{aligned}
&\left|\frac{\mathrm{E}_1\, e_0(\psi,n)^2 - \mathrm{E}_0\, e_0(\psi,n)^2}{\mathrm{E}_0\, e_0(\psi,n)^2}\right| \\
&= \left|\frac{\mathrm{E}_1(\Lambda e_0(\psi,n)^2) - \mathrm{E}_0\, e_0(\psi,n)^2}{\mathrm{E}_0\, e_0(\psi,n)^2}\right| \\
&= \frac{\left|(\Lambda e_0(\psi,n), \Lambda e_0(\psi,n))_{\mathcal{H}(m_1,K_1)} + (\mathrm{E}_1\, e_0(\psi,n))^2 - (e_0(\psi,n), e_0(\psi,n))_{\mathcal{H}(0,K_0)}\right|}{\mathrm{E}_0\, e_0(\psi,n)^2} \\
&= \frac{\left|(\Lambda^*\Lambda e_0(\psi,n), e_0(\psi,n))_{\mathcal{H}(0,K_0)} - (e_0(\psi,n), e_0(\psi,n))_{\mathcal{H}(0,K_0)} + (\sum_{j=n+1}^{\infty} c_j m_j)^2\right|}{\sum_{j=n+1}^{\infty} c_j^2} \\
&\leq \frac{\left|((I - \Lambda^*\Lambda)e_0(\psi,n), e_0(\psi,n))_{\mathcal{H}(0,K_0)}\right| + \sum\limits_{j=n+1}^{\infty} c_j^2 \sum\limits_{j=n+1}^{\infty} (m_1)_j^2}{\sum_{j=n+1}^{\infty} c_j^2}
\end{aligned} \tag{2.3.8}$$

Cauchy-Schwarz inequality indicates that the left hand side of (2.3.8)

$$\begin{aligned}
&\leq \frac{\|(I - \Lambda^*\Lambda)e_0(\psi,n)\|_0\, \|e_0(\psi,n))\|_0}{\sum_{j=n+1}^{\infty} c_j^2} + \sum_{j=n+1}^{\infty} m_{1j}^2 \\
&\leq \frac{\sum_{j=n+1}^{\infty} |c_j|\, \|(I - \Lambda^*\Lambda)\psi_j\|_0 \sqrt{\sum_{j=n+1}^{\infty} c_j^2}}{\sum_{j=n+1}^{\infty} c_j^2} + \sum_{j=n+1}^{\infty} m_1 j^2
\end{aligned} \tag{2.3.9}$$

and this is

$$\begin{aligned} &\leq \frac{\sqrt{\sum_{j=n+1}^{\infty} c_j^2}\sqrt{\sum_{j=n+1}^{\infty} \|(I-\Lambda^*\Lambda)\psi_j\|_0^2}}{\sqrt{\sum_{j=n+1}^{\infty} c_j^2}} + \sum_{j=n+1}^{\infty} m_{1j}^2 \\ &= \sqrt{\sum_{j=n+1}^{\infty} \|(I-\Lambda^*\Lambda)\psi_j\|_0^2} + \sum_{j=n+1}^{\infty} m_{1j}^2 \end{aligned} \tag{2.3.10}$$

The right side of (2.3.9) does not depend on ψ and tends to 0 as $n \to \infty$ by the (2.3.7), so (2.3.3) follows.Switching the roles of K_0 and K_1 yields (2.3.4). Next, since $\mathrm{E}_1 e_1^2 \leq \mathrm{E}_1 e_0^2$

$$\frac{\mathrm{E}_0 e_1(\psi,n)^2}{\mathrm{E}_0 e_0(\psi,n)^2} \leq \frac{\mathrm{E}_0 e_1(\psi,n)^2}{\mathrm{E}_1 e_1(\psi,n)^2} \cdot \frac{\mathrm{E}_1 e_1(\psi,n)^2}{\mathrm{E}_0 e_0(\psi,n)^2}.$$

So (2.3.5) follows from (2.3.3)(2.3.4). Again switching the roles of K_0 and K_1 yields (2.3.6). □

Another sensible measure for how well predictions based on K_1 do when K_0 is the correct covariance function is

$$\frac{\mathrm{E}_0(e_1(\mathbf{t},n) - e_0(\mathbf{t},n))^2}{\mathrm{E}_0 e_0(\mathbf{t},n)^2},$$

i.e., how large the mean squared difference of predictions is relative to correct mean squared error. Because the mean squared error is often calculated in practice, it is also of interest to compare the presumed MSE with the actual MSE by evaluating the ratio $\mathrm{E}_1 e_1(\mathbf{t},n)^2 / \mathrm{E}_0 e_1(\mathbf{t},n)^2$. The following theorem is a special version of Stein (1999b, p.135).

Theorem 2.3.2. *If the two covariance functions K_0 and K_1 define two equivalent Gaussian probability measures, and the set of sampling sites $\{\mathbf{t}_k, k = 1, 2, \dots\}$ is dense*

in D, *where* $D \subset \mathbb{R}^d$ *is bounded, then uniformly in* $\mathbf{t} \in D$ *such that* $\mathrm{E}_0\, e_0(\mathbf{t}, n)^2 > 0$,

$$\lim_{n\to\infty} \frac{\mathrm{E}_0(e_1(\sigma^2, n) - e_0(\mathbf{t}, n))^2}{\mathrm{E}_0\, e_0(\mathbf{t}, n)^2} = 0 \tag{2.3.11}$$

$$\lim_{n\to\infty} \mathrm{E}_1\, e_1(\mathbf{t}, n)^2 / \mathrm{E}_0\, e_1(\mathbf{t}, n)^2 = 1. \tag{2.3.12}$$

Sufficient conditions for equivalent probability measures exist in terms of spectral density. Let $f_i(\boldsymbol{\lambda})$ for $\boldsymbol{\lambda} \in \mathbb{R}^d$ be the spectral density corresponding to the covariance function $K_i(\mathbf{h})$ for $i = 0, 1$. If $f_i(\boldsymbol{\lambda})$'s are isotropic, i.e., depending only on $\|\boldsymbol{\lambda}\|$, it follows from Theorem 2.2.5 that the corresponding Gaussian measures are equivalent if for some $\epsilon > 0$,

$$f_1(\boldsymbol{\lambda})/f_0(\boldsymbol{\lambda}) - 1 = O(\|\boldsymbol{\lambda}\|^{-(d/2+\epsilon)}) \text{ as } \|\boldsymbol{\lambda}\| \to \infty. \tag{2.3.13}$$

Condition (2.2.19) implies that $f_1(\boldsymbol{\lambda})/f_0(\boldsymbol{\lambda}) \to 1$ as $\|\boldsymbol{\lambda}\| \to \infty$ and imposes some rate on the convergence. Condition (2.2.19) is stronger than necessary for (2.3.11) and (2.3.12) to hold, as seen from the next theorem (see Stein 1999b, p.136).

Theorem 2.3.3. *Let the underlying process* $X(\mathbf{t})$ *be Gaussian under probability* P_i *with mean 0 and spectral density* f_i, $i = 0, 1$. *If for some* $\varphi > 1$, $f_0(\boldsymbol{\lambda})\, \|\boldsymbol{\lambda}\|^{\varphi}$ *is bounded away from* 0 *and* ∞ *and*

$$\frac{f_1(\boldsymbol{\lambda})}{f_0(\boldsymbol{\lambda})} \to 1 \text{ as } \|\boldsymbol{\lambda}\| \to \infty,$$

then (2.3.11) and (2.3.12) hold.

If observations are taken on infinite lattice $\delta\mathbb{Z}$ while $\delta \to 0$, the rate imposed on the tendency of $f_1/f_0 \to 1$ will yield the rate of convergence to optimality of predictions in (2.3.11) and (2.3.12) (see Stein 1999a, p.252). As (2.3.2) defined in finite sample case, we let $e_i(\mathbf{t}, \delta)$ be the prediction error of $X(\mathbf{t})$ under measure P_i when the process is observed at an infinite lattice $\delta\mathbb{Z}$.

Theorem 2.3.4. *If* $f_i(\boldsymbol{\lambda})(1 + \|\boldsymbol{\lambda}\|)^{\varphi}$ *is bounded away from 0 and* ∞ $(i = 0, 1)$ *and*

$|f_1(\boldsymbol{\lambda})/f_0(\boldsymbol{\lambda}) - 1| \le A(1 + \|\boldsymbol{\lambda}\|)^{-\gamma}$ *for some positive numbers* φ, γ, A, *then as* $\delta \to 0$

$$\sup_{\sigma^2 \notin \delta\mathbb{Z}^d} \frac{\mathrm{E}_0(e_0(\sigma^2,\delta) - e_1(\sigma^2,\delta))^2}{\mathrm{E}_0\, e_1(\sigma^2,\delta)^2} = O\left(\delta^{\min(\varphi,2\gamma)}(\log\delta^{-1})^{1\{\varphi=2\gamma\}}\right) \tag{2.3.14}$$

$$\sup_{\sigma^2 \notin \delta\mathbb{Z}^d} \left|\frac{\mathrm{E}_1\, e_1(\sigma^2,\delta)^2}{\mathrm{E}_0\, e_1(\sigma^2,\delta)^2} - 1\right| = O\left(\delta^{\min(\varphi,\gamma)}(\log\delta^{-1})^{1\{\varphi=\gamma\}}\right) \tag{2.3.15}$$

Under an additional condition on the spectral densities, Stein(1990a, p.259) also obtained such convergence rates when the observations are unequally spaced on an interval.

2.4 Fixed-domain asymptotics of maximum likelihood estimators

Comparing with increasing domain asymptotics, fixed-domain asymptotic results for estimation are considerably fewer due to lack of analytic tools dealing with increasingly stronger correlations between nearby observations. More specifically, taking more and more data in a fixed domain will give you more and more correlated observations, as opposed to increasing domain asymptotics, where taking samples in an increasingly large domain gives roughly independent observations if correlations decay with distance fast enough. For the simplest model, exponential model, the fixed domain asymptotics of exact MLE has been thoroughly studied by Ying(1991 and 1993), Chen, Simpson, and Ying (2000). For the general Matérn model, Zhang(2004) gives the strong consistency of MLE with θ fixed. However, the fixed-domain asymptotic distribution is not available even when data are observed along a line, therefore we study and establish first the fixed-domain asymptotic distribution of MLE for the microergodic parameter in the general Matérn model. In the following, we will present some of these results.

Let the second order stationary Gaussian process $X(\mathbf{t}), \mathbf{t} \in \mathbb{R}^d$ have mean 0 and

an isotropic covariance function $K(h;\theta,\sigma^2)$, where σ^2 is the variance of the process and θ is the parameter that controls how fast the covariance function decays. Given n observations $\mathbf{X}_n = (X(\mathbf{t}_1),\ldots,X(\mathbf{t}_n))'$, the log-likelihood is

$$\ell_n(\theta,\sigma^2) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log[\det \mathbf{V}_n(\theta,\sigma^2)] - \frac{1}{2}\mathbf{X}_n'[\mathbf{V}_n(\theta,\sigma^2)]^{-1}\mathbf{X}_n, \qquad (2.4.1)$$

where $\mathbf{V}_n(\theta,\sigma^2)$ denotes the covariance matrix of $\mathbf{X}_n$. The maximum likelihood estimator (MLE) of (θ,σ^2) maximizes the likelihood function $\ell_n(\theta,\sigma^2)$, i.e. $(\hat{\theta},\hat{\sigma}^2) =$ ArgMax $\ell_n(\theta,\sigma^2)$. In this work, we will focus on Matérn model, that is, the covariance function considered in (2.4.1) is Matérn defined as in (3.3.14). As we mentioned earlier, the product $\sigma^2\theta^{2\nu}$ is shown to be consistently estimable, although both parameters σ^2 and θ are not if spatial sampling domain is bounded [see e.g. Zhang (2004)]. It is actually more important to estimate $\sigma^2\theta^{2\nu}$ well for spatial interpolation [see, e.g. Zhang (2004), Stein(1999b)]. For the exponential model which is the special case with $\nu = 1/2$, the underlying process is known as Ornstein-Ulenbeck process. Because this process possesses some nice properties such as the Markovian properties, the fix-domain asymptotic analysis is relatively more tractable [Ying (1991), Chen et al. (2000)]. Ying (1991) gives the following fixed-domain strong consistency and asymptotic normality for the MLE of the consistently estimable parameter $\theta\sigma^2$ for exponential model as following:

Theorem 2.4.1. *Let the underlying process $\{X(t), t \in [0,1]\}$ be Gaussian with mean 0 and an exponential covariance function $K(h) = \sigma^2\exp(-\theta h)$. The domain of the maximization in (θ,σ^2) is $J = [a,\infty] \times [w,v]$ or $[a,b] \times [w,\infty]$, $0 < a < b < \infty$ and $0 < w < v < \infty$. Then, with probability one the maximum likelihood estimator $(\hat{\theta}_{n,tap},\hat{\sigma}^2_{n,tap})$ exists for all large n and as $n \to \infty$*

$$\hat{\theta}_n\hat{\sigma}_n^2 \longrightarrow \theta_0\sigma_0^2 \qquad a.s., \qquad (2.4.2)$$

$$\sqrt{n}(\hat{\theta}_n\hat{\sigma}_n^2 - \theta_0\sigma_0^2) \xrightarrow{d} N(0, 2(\theta_0\sigma_0^2)^2). \qquad (2.4.3)$$

For general Matérn model, there is no such properties like Markovian property strictly speaking for the underlying process. So the fixed-domain asymptotic properties for $(\hat{\theta}, \hat{\sigma}^2)$ by joint maximizing both parameters are not available yet. However, Zhang (2004) proved that when ν is known and θ is fixed at any value $\theta_1 > 0$, the maximizer of likelihood (2.4.1) $\hat{\sigma}_n^2$ ensures the following strong consistency of an estimator of the consistently estimable parameter $\theta\sigma^2$.

Theorem 2.4.2. *Let the underlying process* $\{X(\mathbf{t}), \mathbf{t} \in \mathbb{R}^d\}$, $d = 1, 2, or\ 3$, *be second order stationary Gaussian with mean 0 and possess an isotropic Matérn covariogram* 3.3.14 *with the unknown parameter values* σ_0^2, θ_0 *and a known* ν. *Let* D_n, $n = 1, 2, \ldots$ *be an increasing sequence of finite subsets of* $\mathbb{R}^d$ *such that* $\bigcup_{n=1}^{\infty} D_n$ *is bounded and infinite, and* $L_n(\sigma^2, \theta)$ *be likelihood function when the process is observed at locations in* D_n. *For any fixed* $\theta_1 > 0$, *let* $\hat{\sigma}_n^2$ *maximize* $L_n(\sigma^2, \theta_1)$, *Then* $\hat{\sigma}^2\theta_1^{2\nu} \to \sigma_0^2\theta_0^{2\nu}$ *with* P_0 *probability 1, where* P_0 *is the Gaussian measure defined by the Matérn covariogram corresponding to parameter values* σ_0^2, θ *and* ν.

We showed the asymptotic normality of this estimator.

Theorem 2.4.3. *[Du, Zhang and Mandrekar(2009)] With the same notation and assumptions as in previous Theorem 2.4.2, for any fixed* θ_1,

$$\sqrt{n}(\hat{\sigma}_n^2\theta_1^{2\nu} - \sigma_0^2\theta_0^{2\nu}) \xrightarrow{d} N(0, 2(\sigma_0^2\theta_0^{2\nu})^2). \tag{2.4.4}$$

This theorem is proved based on those theoretical results related to the equivalence of Gaussian measures. The detailed proof will be provided in next chapter.

Chapter 3

Covariance tapering and fixed-domain properties of tapered maximum likelihood estimators

3.1 Introduction

As introduced in the very beginning, applying some traditional statistical approachs, such as best linear unbiased prediction or kriging, the maximum likelihood estimation or the Bayesian inferences, to large spatial data sets are often computationally infeasible because of cubic order matrix algorithms on matrix inverse or determinant involved. To obviate these computational hurdles, a natural idea is to make the covariances exactly zero after certain distance so that the resulting matrix has a high proportion of zero entries and is therefore a sparse matrix. Operations on sparse matrices take up less computer memories and run faster. However, this has to be done in a way such that the resulting matrix is still positive definite. Covariance tapering assures that the tapered covariance matrix is positive definite while retaining most of the information. Technically this method is to taper the covariance function to zero beyond a certain range by directly multiplying a positive definite but compactly

supported function, that results in the so called tapered covariance matrix which can be efficiently handled by well-established sparse matrix algorithms. The tapered covariogram is of the form

$$\tilde{K}(h;\boldsymbol{\theta}) = K(h;\boldsymbol{\theta})K_{\text{tap}}(h), \tag{3.1.1}$$

where $K(\boldsymbol{h},\boldsymbol{\theta})$ is the covariance function of the underlying process that depends on a vector of parameter $\boldsymbol{\theta}$ and $K_{\text{tap}}(h)$ is the taper, a known correlation function that is 0 after a threshold distance. Some examples of taper are spherical, Wendland tapers (Wendland, 1995, 1998). See also Wu (1995), Gneiting (2002) and Mitra et al. (2003).

One would then use $\tilde{K}(\boldsymbol{h},\boldsymbol{\theta})$ in estimation and interpolation as if it was the correct covariance function. For example, we would maximize the following tapered log likelihood function for Gaussian observations to obtain an estimate for $\boldsymbol{\theta}$,

$$\ell_{\text{n,tap}}(\theta,\sigma^2) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log[\det \tilde{\mathbf{V}}_n] - \frac{1}{2}\mathbf{X}_n'\tilde{\mathbf{V}}_n^{-1}\mathbf{X}_n. \tag{3.1.2}$$

where the tapered covariance matrix is a Hadamard product $\tilde{\mathbf{V}}_n = \mathbf{V}_n(\theta,\sigma^2) \circ \mathbf{T}_n$, and $\mathbf{T}_n$ has the (i,j)th element as $K_{\text{tap}}(\|\mathbf{t}_i - \mathbf{t}_j\|)$. $\mathbf{X}_n$ is the vector of observations. Maximizing (3.3.4) is computationally feasible for extremely large datasets but results in a pseudo-likelihood estimator whose properties need to be studied. Kaufman (2008) established consistency of the pseudo-likelihood estimator for Matérn class. Very recently, Du, Zhang and Mandrekar (2009) gave some general conditions to ensure that tapering does not affect the efficiency of the maximum likelihood estimator. For spatial interpolation, Furrer Genton and Nychika (2006) showed that under some regularity conditions, tapering produces asymptotically optimal prediction.

All the conditions that result in no effect on consistency, asymptotic efficiency of estimation or asymptotic optimality of prediction put constrains on the tail behavior of spectral density of tapering function. Roughly speaking, this can be accounted for

by noting that spectral density is the Fourier transform of covariance function. So the faster the spectral density of taper decays, the smoother the tapering function is at origan and it hardly changes the degree of differentiability of the underlying process, which actually plays a central role in any fixed-domain asymptotic results. Therefore, section 3.2 is devoted to address the tail properties of tapering spectral density and effect of tapering on spatial interpolation. The effect on estimation is in section 3.3, we will focus on studying the asymptotic distribution of tapered MLE since it is unknown if the covariance tapering causes any loss of asymptotic efficiency before. The content of these two sections is from the joint papers [Du, Zhang and Mandrekar (2009)], which has been accepted by Annals of Statistics, and [Zhang and Du(2008)] with my advisors. In the absence of asymptotic distributions for the true MLE in the general case, we compare the log likelihood function and the tapered log likelihood function, and their derivatives in section 3.4. The simulation study and an application to the climate data to show accuracy and computational are presented in the last section.

3.2 Tail properties of tapering spectral density

If $K(\boldsymbol{h})$ is the covariance function of $X(\mathbf{t})$, and $\tilde{K}(\boldsymbol{h}) = K(\boldsymbol{h})K_{\text{tap}}(\boldsymbol{h})$ the tapered covariance function, denote the spectral density corresponding to $K(\boldsymbol{h})$, $\tilde{K}(\boldsymbol{h})$ and $K_{tap}(\boldsymbol{h})$ by $f_0(\boldsymbol{\lambda}), f_1(\boldsymbol{\lambda})$ and $f_{tap}(\boldsymbol{\lambda})$ for $\boldsymbol{\lambda} \in \mathbb{R}^d$, respectively. We have the following relationship from a well-known fact about the Fourier transformation,

$$f_1(\boldsymbol{\lambda}) = \int f_0(\boldsymbol{\lambda} - \mathbf{x}) f_{\text{tap}}(\mathbf{x})\, d\mathbf{x}. \tag{3.2.1}$$

Based on an equivalent equation as (3.2.1), Furrer Genton and Nychika (2006) have shown the following result.

Theorem 3.2.1. *Let* $f_0(\boldsymbol{\lambda}) \propto \sigma^2\theta^{2\nu}/(\theta^2 + \|\boldsymbol{\lambda}\|^2)^{\nu+d/2}$ *for* $\boldsymbol{\lambda} \in \mathbb{R}^d$ *be the spectral density function of a Matérn covariance function. Let the spectral density* $f_{tap}(\boldsymbol{\lambda})$ *of*

the taper satisfies the following taper condition:

$$0 < f_{tap}(\boldsymbol{\lambda}) < M(1 + \|\boldsymbol{\lambda}\|^2)^{-\nu - d/2 - \epsilon} \tag{3.2.2}$$

for some $\epsilon \geq 0$ *and* $M > 0$. *Then* $f_1(\boldsymbol{\lambda})/f_0(\boldsymbol{\lambda})$ *has a finite limit as* $\|\boldsymbol{\lambda}\| \to \infty$ *and this limit equals 1 if* $\epsilon > 0$.

This theorem and Theorem 2.3.3 imply that if $f_{\text{tap}}(\boldsymbol{\lambda})$ has a lower tail than $f_0(\boldsymbol{\lambda})$, then predictions under both models will be nearly the same when a large sample is obtained from a bounded region.

We can give a stronger result which combined with Theorem 2.3.4 gives the convergence rate of mse's based on a tapered covariance structure, this indicates the degree of maintaining the optimality of prediction using appropriate tapering. It also provides the condition for the two measures corresponding to the two spectral densities, i.e. tapered and untapered, to be equivalent in one dimension case.

Theorem 3.2.2. *If condition (3.2.15) holds for some* $\epsilon > d/2 + \gamma/2$, *where* $0 < \gamma < 1$, *the following holds.*

(i)

$$|f_1(\boldsymbol{\lambda})/f_0(\boldsymbol{\lambda}) - 1| \leq A(1 + \|\boldsymbol{\lambda}\|)^{-\gamma} \tag{3.2.3}$$

$$C_1 < f_i(\boldsymbol{\lambda})(1 + \|\boldsymbol{\lambda}\|)^{2\nu + d} < C_2 \qquad (i = 0, 1) \tag{3.2.4}$$

for some constants $A, C_1, C_2 > 0$.

(ii) The two Gaussian measures corresponding to the two spectral densities are equivalent for $d = 1$ and $1 > \gamma > 1/2$.

Proof. We can assume $\theta = 1$ without loss of any generality because the results will be the same except for the magnitude of the constants A and C_i, $i = 1, 2$. Then in view of isotropy, $f_0(\boldsymbol{\lambda})$ can be written as $f_0^*(\|\boldsymbol{\lambda}\|) = M_1/(1 + \|\boldsymbol{\lambda}\|^2)^{\nu + d/2}$, where M_1 is a positive constant [see (3.3.14)]. By continuity and $f_0(\boldsymbol{\lambda}) > 0$, to show (3.2.3) it

suffices to show that there exists $A > 0$ such that when $\|\boldsymbol{\lambda}\|$ large enough,

$$\frac{|f_1(\boldsymbol{\lambda}) - f_0(\boldsymbol{\lambda})|(1 + \|\boldsymbol{\lambda}\|)^\gamma}{f_0(\boldsymbol{\lambda})} \leq A. \tag{3.2.5}$$

Indeed, since tapered covariance function $C(\cdot)$ is also isotropic, from (3.2.1) it follows that

$$f_1(\boldsymbol{\lambda}) = f_1^*(\|\boldsymbol{\lambda}\|) = \int_{\mathbb{R}^d} f_0^*(\|\boldsymbol{\lambda} - \boldsymbol{x}\|) f_{\text{tap}}^*(\|\boldsymbol{x}\|)\, d\boldsymbol{x} \tag{3.2.6}$$

where $f_{\text{tap}}(\boldsymbol{x}) = f_{tap}^*(\|\boldsymbol{x}\|)$. In addition, if we set $\boldsymbol{x} = r\mathbf{u}$, $\boldsymbol{\lambda} = \omega\boldsymbol{v}$ with $\|\mathbf{u}\| = \|\boldsymbol{v}\| = 1$, where where $\mathbf{u}$, $\boldsymbol{v}$ represent the direction of $\mathbf{x}$ and $\boldsymbol{\lambda}$ respectively, the RHS of (3.2.6) becomes

$$M_2 \int_{\partial B_d} \int_0^\infty f_0^*(\|r\mathbf{u} - \omega\boldsymbol{v}\|) f_{\text{tap}}^*(r) r^{d-1} dr dU(\mathbf{u}), \tag{3.2.7}$$

where $M_2 = 2\pi^{d/2}/\Gamma(n/2)$ which is the surface area of a d- dimensional unit sphere, ∂B_d is the surface of this unit sphere and U is the uniform probability measure on ∂B_d. On the other hand, note that f_{tap} is the spectral density of the correlation function $K_{tap}(\boldsymbol{h})$, so f_0 can be rewritten in an analogous form

$$f_0(\boldsymbol{\lambda}) = f_0^*(\omega) = M_2 \int_{\partial B_d} \int_0^\infty f_0^*(\|\omega\boldsymbol{v}\|) f_{\text{tap}} * (r) r^{d-1} dr dU(\mathbf{u}).$$

and the LHS of (3.3.86) is therefore bounded by

$$M_2(1+\omega)^\gamma \int_{\partial B_d} \int_0^\infty \frac{\left|f_0^*(\|r\mathbf{u} - \omega\boldsymbol{v}\|) - f_0^*(\|\omega\boldsymbol{v}\|)\right|}{f_0^*(\omega)} f_{\text{tap}}^*(r) r^{d-1} dr dU(\mathbf{u}). \tag{3.2.8}$$

Furthermore, for ω sufficiently large, we will bound the inner integral over intervals $[0, \omega - \Delta]$, $[\omega - \Delta, \omega + \Delta]$, $[\omega + \Delta, \infty)$ by some quantities independent of $\mathbf{u}$, where $\Delta = O(\omega^\delta)$ for $\delta = (d + 2\nu + \gamma)/(d + 2\nu + 1)$, clearly $0 < \delta < 1$. Then the assessment of the whole iterated integral in (3.2.8) becomes straightforward because the outer integral is one.

First, note that the Mean Value Theorem implies that there exits ξ between ω and

$\|r\mathbf{u} - \omega\boldsymbol{v}\|$ such that

$$f_0^*(\|r\mathbf{u} - \omega\boldsymbol{v}\|) - f_0^*(\|\omega\boldsymbol{v}\|) = f_0^{*\prime}(\xi)(\|r\mathbf{u} - \omega\boldsymbol{v}\| - \|\omega\boldsymbol{v}\|)$$

and $\left|f_0^{*\prime}(x)\right| = [2M_1(\nu + d/2)x]/[(1 + x^2)^{\nu+d/2+1}]$ which is decreasing in $x > 0$. Therefore, for $r \in [0, \omega - \Delta]$ or $[\omega + \Delta, \infty)$, we have $\|r\mathbf{u} - \omega\boldsymbol{v}\| \geq |\omega - r| \geq \Delta$. This together with $\omega > \Delta$ entails $\xi > \Delta$, thus

$$|f_0^*(\|r\mathbf{u} - \omega\boldsymbol{v}\|) - f_0^*(\|\omega\boldsymbol{v}\|)| \leq \left|f_0^{*\prime}(\Delta)\right| \|r\mathbf{u}\| = \frac{M_3 \Delta r}{(1 + \Delta^2)^{\nu+d/2+1}}, \tag{3.2.9}$$

where M_3 denotes $2M_1(\nu + d/2)$. For $r \in [\omega - \Delta, \omega + \Delta]$, we have $\|r\mathbf{u} - \omega\boldsymbol{v}\| \geq |\omega - r|$ and $\omega > \Delta > |\omega - r|$, which indicates $\xi > |\omega - r|$, so

$$|f_0^*(\|r\mathbf{u} - \omega\boldsymbol{v}\|) - f_0^*(\|\omega\boldsymbol{v}\|)| \leq \left|f_0^{*\prime}(|\omega - r|)\right| \|r\mathbf{u}\| = \frac{M_3|\omega - r|r}{(1 + |\omega - r|^2)^{\nu+d/2+1}} \tag{3.2.10}$$

From (3.2.9) and taper condition (3.2.15), it follows that

$$\begin{aligned}
&\int_{|r-\omega|>\Delta} \frac{\left|f_0^*(\|r\mathbf{u} - \omega\boldsymbol{v}\|) - f_0^*(\|\omega\boldsymbol{v}\|)\right|}{f_0^*(\omega)} f_{\text{tap}}^*(r) r^{d-1}\, dr \\
&\leq \frac{M_3 \Delta (1 + \omega^2)^{\nu+d/2}}{M_1(1 + \Delta^2)^{\nu+d/2+1}} \int_{|r-\omega|>\Delta} r g^*(r) r^{d-1}\, dr \\
&\leq \frac{M_3 M \Delta (1 + \omega^2)^{\nu+d/2}}{M_1(1 + \Delta^2)^{\nu+d/2+1}} \left(\int_0^{\omega-\Delta} \frac{r^d}{(1+r^2)^{v+d/2+\epsilon}}\, dr + \int_{\omega+\Delta}^{\infty} \frac{r^d}{(1+r^2)^{v+d/2+\epsilon}}\, dr \right)
\end{aligned} \tag{3.2.11}$$

As $\epsilon > d/2 \geq 1/2 - \nu$, the first integral in (3.2.11) is finite and the second integral tends to 0 as $\omega \to \infty$. To control the last inner integral part, we need to use the monotonicity of $r^d/[(1 + r^2)^{v+d/2+\epsilon}]$ which is decreasing in r for $\epsilon > d/2$. In view of this, (3.2.10) and taper condition (3.2.15), we have

$$\int_{\omega-\Delta}^{\omega+\Delta} \frac{\left|f_0^*(\|r\mathbf{u} - \omega\boldsymbol{v}\|) - f_0^*(\|\omega\boldsymbol{v}\|)\right|}{f_0^*(\omega)} f_{\text{tap}}^*(r) r^{d-1}\, dr$$

$$\leq \frac{M_3 M(1+\omega^2)^{\nu+d/2}}{M_1} \int_{\omega-\Delta}^{\omega+\Delta} \frac{|r-\omega|\, r^d}{(1+|r-\omega|^2)^{\nu+d/2+1}(1+r^2)^{\nu+d/2+\epsilon}}\, dr$$

$$\leq \frac{M_3 M(\omega-\Delta)^d(1+\omega^2)^{\nu+d/2}}{M_1(1+(\omega-\Delta)^2)^{\nu+d/2+\epsilon}} \int_{\omega-\Delta}^{\omega+\Delta} \frac{|r-\omega|}{(1+|r-\omega|^2)^{\nu+d/2+1}}\, dr, \qquad (3.2.12)$$

where the finiteness of the last integral is easy to be verified via changing of variables. Since the total mass of outer integral in (3.2.8) is one, combining (3.2.8) with (3.2.11) and (3.2.12) gives

$$\limsup_{\omega\to\infty} M_2(1+\omega)^{\gamma} \int_{\partial B_d} \int_0^{\infty} \frac{\left| f_0^*(\|r\mathbf{u}-\omega\boldsymbol{v}\|) - f_0^*(\|\omega\boldsymbol{v}\|) \right|}{f_0^*(\omega)} f_{\mathrm{tap}}^*(r) r^{d-1} dr dU(\mathbf{u}) < \infty,$$

which implies (3.3.86) for $\|\boldsymbol{\lambda}\|$ sufficiently large and the proof of (3.2.3) is done.

To show (3.2.4), first it is noted that there exist C_1^* and C_2^* positive such that

$$C_1^* < f_0(\boldsymbol{\lambda})(1+\|\boldsymbol{\lambda}\|)^{2\nu+d} < C_2^*. \qquad (3.2.13)$$

In addition, (3.2.3) entails

$$f_1(\boldsymbol{\lambda}) \leq f_0(\boldsymbol{\lambda})(1+A(1+\|\boldsymbol{\lambda}\|)^{-\gamma}) \leq f_0(\boldsymbol{\lambda})(1+A).$$

Therefore

$$f_1(\boldsymbol{\lambda})(1+\|\boldsymbol{\lambda}\|)^{2\nu+d} < C_2^*(1+A). \qquad (3.2.14)$$

On the other hand, (3.2.6) and (3.2.13) imply

$$(1+\|\boldsymbol{\lambda}\|)^{2\nu+d} f_1(\boldsymbol{\lambda}) = (1+\|\boldsymbol{\lambda}\|)^{2\nu+d} \int_{\mathbb{R}^d} f_1(\boldsymbol{\lambda}-\boldsymbol{x}) f_{\mathrm{tap}}(\boldsymbol{x})\, d\boldsymbol{x} > C_1^* \int_{\mathbb{R}^d} f_{\mathrm{tap}}(\boldsymbol{x})\, d\boldsymbol{x} = C_1^*,$$

which together with (3.2.14) completes the proof of (3.2.4) with $C_1 = C_1^*$, $C_2 = C_2^*(1+A)$, and **(i)** is proved. Finally **(ii)** follows readily from combining (3.2.3) with (2.3.13). □

In practice, the compactly supported radial basis functions constructed by Wend-

land (1995; 1998) are often used as the tapering function after parameterization. These Wendland tapers are of the form

$$K_{w,d,k}(\boldsymbol{h};\theta) = M_{d,k}p_{d,k}(\frac{\|\boldsymbol{h}\|}{\theta})1_{\{0\le x\le\theta\}},$$

where $p_{d,k}$ is a polynomial of degree $[d/2]+3k+1$. Specially $K_{w,2,1}(h;\theta) = (1-\frac{h}{\theta})_+^4(1+4\frac{h}{\theta})$ and $K_{w,2,2}(h;\theta) = (1-\frac{h}{\theta})_+^6(1+6\frac{h}{\theta}+\frac{35h^2}{3\theta^2})$, $\theta>0, h>0$ ($x_+ = \max\{0,x\}$) are the two wendland tapers discussed in Furrer Genton and Nychika (2006), their one dimensional spectral densities denoted by g_1, g_2 have the following tail properties: $\lambda^4 g_1(\lambda)\to 120/(\pi\gamma^3)$ and $\lambda^6 g_2(\lambda)\to 17920/(\pi\gamma^5)$, as $\lambda\to\infty$. Therefore, according to **(ii)** in Theorem 2.3.4, tapering with $K_{W,2,1}$ or $K_{W,2,2}$ will yield the equivalent Gaussian measure if the exponential model ($\nu = 1/2$) is studied, for instance. It was shown (Wendland, 1998, p.8) that $g_{d,k}(\boldsymbol{\lambda}) \le M(1+\|\boldsymbol{\lambda}\|^2)^{-d/2-k-1/2}$ with $g_{d,k}$ denoting the spectral density of $K_{W,d,k}$ in $\mathbb{R}^d$, so we can see $K_{W,d,k}$ satisfies taper condition (3.2.15) with $\epsilon > (d+\gamma)/2$ whenever $k > (d+\gamma)/2+\nu-1/2$.

By assuming faster decay of spectral density of tapering function than (3.2.15), Kaufman(2009) showed the following equivalence of two Gaussian measures denoted by P_0, P_1 corresponding to exact and tapered Matérn covariance functions.

Theorem 3.2.3. *Let K_0 be the Matérn covariance function on $\mathbb{R}^d$, $d\le 3$, with parameters $\sigma^2, theta, \nu$, and let f_{tap} be the spectral density of tapering function K_{tap}. Suppose there exist $\epsilon > \max\{d/4, 1-\nu\}$ and $M_\epsilon < \infty$ such that*

$$0 < f_{tap}(\boldsymbol{\lambda}) < M(1+\|\boldsymbol{\lambda}\|^2)^{-\nu-d/2-\epsilon} \tag{3.2.15}$$

Then $\mathrm{P}_0 \equiv \mathrm{P}_1$ on the paths of $\{X(\mathbf{t}), \mathbf{t}\in D\}$, for any bounded subset $D\subset\mathbb{R}^d$.

Actually tapering condition (3.2.15) with $\epsilon > \max\{d/4, 1-\nu\}$ implies that the difference of the spectral density function of exact and tapered covariance functions

is of the following order,

$$\frac{f_0(\boldsymbol{\lambda}) - f_1(\boldsymbol{\lambda})}{f_1(\boldsymbol{\lambda})} = O(\|\boldsymbol{\lambda}\|^{-r})$$

with $r > 1/2$. The condition (2.2.14) in Theorem 2.2.5 is satisfied and therefore the Gaussian measure specified by tapered covariance function is equivalent to the original one. This theorem will lead to the strong consistency of tapered MLE under the tapering condition above. Now if we strengthen the tapering condition (3.2.15) for $d = 1$ by letting $\epsilon > \max\{1/2, 1 - \nu\}$, we get

$$\frac{f_0(\boldsymbol{\lambda}) - f_1(\boldsymbol{\lambda})}{f_1(\boldsymbol{\lambda})} = O(\|\boldsymbol{\lambda}\|^{-r})$$

with $r > 1$. This is stronger result than equivalence of Gaussian measures and it will enable tapering to retain the asymptotic efficiency. The detailed proof of this result will be given in the next section.

3.3 Fixed-domain asymptotic properties of tapered maximum likelihood estimators

3.3.1 Strong consistency of tapered MLE for Matérn model

Assume that the underlying process $X(\mathbf{t}), \mathbf{t} \in \mathbb{R}^d$ is stationary Gaussian with mean 0 and a Matérn covariogram as defined in section 2.2. Namely,

$$K(x;\ \sigma^2, \theta, \nu) = \frac{\sigma^2(\theta x)^\nu}{\Gamma(\nu)2^{\nu-1}} K_\nu(\theta x),\ x \geq 0. \tag{3.3.1}$$

Let the process be observed at n locations $\mathbf{t}_k$, $k = 1, \ldots, n$ and ν is known. Write $\boldsymbol{X}_n = (X(\mathbf{t}_1), \ldots, X(\mathbf{t}_n))'$. Then the maximum likelihood estimator following log

likelihood function

$$\ell_n(\theta, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log[\det \mathbf{V}_n(\theta, \sigma^2)] - \frac{1}{2}\mathbf{X}_n'[\mathbf{V}_n(\theta, \sigma^2)]^{-1}\mathbf{X}_n, \qquad (3.3.2)$$

where $\boldsymbol{V}_n(\theta, \sigma^2)$ is the covariance matrix whose (i, j)th element is $K(\|\mathbf{t}_i - \mathbf{t}_j\|, \sigma^2, \theta, \nu)$. The maximization has to be carried out using some iterative procedure such as the Newton-Raphson method or Fisher's scoring method. Then the covariance matrix V_n has to be inverted repeatedly. When n is large, the inversion can be quite slow if not impractical. To overcome this, we replace the covariance function with a tapered covariance function

$$\tilde{K}(h; \theta, \sigma^2) = K(h; \theta, \sigma^2)K_{\text{tap}}(h), \qquad (3.3.3)$$

for some correlation function having a finite support, i.e., $K_{\text{tap}}(\boldsymbol{x}) = 0$ if $\|\boldsymbol{x}\| \geq \gamma$ for some $\gamma > 0$. The covariance matrix corresponding to the tapered covariance function K is the Hadamard product $\boldsymbol{V}_n = \boldsymbol{V}_n \circ \boldsymbol{T}_n$ where $\boldsymbol{T}_n = (K_{\text{tap}}(\|\mathbf{t}_i - \mathbf{t}_j\|))_{i,j=1}^n$. We would then maximize the following function, which will henceforth be call the tapered log likelihood function,

$$\ell_{\text{n,tap}}(\theta, \sigma^2) = -\frac{n}{2}\log 2\pi - \frac{1}{2}\log[\det \tilde{\mathbf{V}}_n] - \frac{1}{2}\mathbf{X}_n'\tilde{\mathbf{V}}_n^{-1}\mathbf{X}_n. \qquad (3.3.4)$$

How does this tapering affect the estimation? We will resort to the fixed-domain asymptotic theory to find an answer. As we mentioned in first chapter, It is known that not all parameters in the covariance function are consistently estimable (Ying, 1991; Zhang, 2004, e.g.) under the fixed-domain asymptotic framework.

First we note that Zhang (2004) showed that two Matérn covariance functions with the same smoothness parameter define two equivalent Gaussian measures if they have the same product $\sigma^2\theta^{2\nu}$ on the paths of $\{X(\mathbf{t}), \mathbf{t} \in D\}$ where D is bounded (Zhang, 2004, Theorem 2). Consequently, the parameters σ^2 and θ cannot be estimated consistently if the spatial sampling domain is bounded, regardless of estimation method.

However, Zhang (2004) showed that $\sigma^2\theta^{2\nu}$ is consistently estimable and constructed a consistent estimator.

Following Zhang's approach, we will construct an consistent estimator of the product $\sigma^2\theta^{2\nu}$ using the tapering covariance function. We assume ν is known for technical reason. Then the tapered covariance matrix $\boldsymbol{V}_n = \sigma^2\mathbf{R}_n(\theta) \circ \boldsymbol{T}_n$, where $\mathbf{R}_n(\theta)$ is the correlation matrix and depends only on θ. Fix θ at a known value θ_1 and maximize $\ell_{\mathrm{n,tap}}(\sigma^2, \theta_1)$ which equals

$$\begin{aligned} &-\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\tilde{\boldsymbol{V}}_n| - \frac{1}{2}\boldsymbol{X}_n'\tilde{\boldsymbol{V}}_n^{-1}\boldsymbol{X}_n \\ &= -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log|\mathbf{R}_n \circ \boldsymbol{T}_n| - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}\boldsymbol{X}_n'(\mathbf{R}_n(\theta_1) \circ \boldsymbol{T}_n)^{-1}\boldsymbol{X}_n \end{aligned} \quad (3.3.5)$$

Hence

$$\hat{\sigma}^2_{\mathrm{tap}}(\theta_1) = \mathrm{ArgMax}\ell_{\mathrm{n,tap}}(\sigma^2, \theta_1) = \frac{1}{n}\boldsymbol{X}_n'(\mathbf{R}_n(\theta_1) \circ \boldsymbol{T}_n)^{-1}\boldsymbol{X}_n.$$

Theorem 3.3.1. *[Zhang, et al.(2008)]*

Assume the underlying process on $\mathbb{R}^d, d \leq 3$ is Gaussian with mean 0 and a Matérn covariance function (3.3.1). If the taper is such that the tapered covariance function $\tilde{K}(\cdot, \sigma^2, \theta, \nu)$ and the untapered covariance function $K(\cdot, \sigma^2, \theta, \nu)$ define two equivalent measures for $X(\mathbf{t}), \mathbf{t} \in D$, then as $n \to \infty$

$$\hat{\sigma}^2_{tap}(\theta_1)\theta_1^{2\nu} \to \sigma_0^2\theta_0^{2\nu} \qquad a.s.$$

where σ_0^2 and θ_0 denote the true value of the parameters.

Combing this theorem and the equivalence result (3.2.3) under tapering given by Kaufman yields the following strong consistency theorem. To simplify the notation, write $\hat{\sigma}^2_{\mathrm{tap}}$ for $\hat{\sigma}^2_{\mathrm{tap}}(\theta_1)$.

Theorem 3.3.2 (Kaufman, et al.(2008)). *Let the underlying process $\{X(\mathbf{t}), \mathbf{t} \in \mathbb{R}^d\}$, $d = 1, 2, or\ 3$, be second order stationary Gaussian with mean 0 and possess an*

isotropic Matérn covariogram 3.3.14 *with the unknown parameter values* σ_0^2, θ_0 *and a known* ν. *Let* D_n, $n = 1, 2, \ldots$ *be an increasing sequence of finite subsets of* $\mathbb{R}^d$ *such that* $\bigcup_{n=1}^{\infty} D_n$ *is bounded and infinite, and* $\ell_{n,tap}(\sigma^2, \theta)$ *be likelihood function when the process is observed at locations in* D_n. *With* K_{tap} *satisfying the taper condition (3.2.15) with* $\epsilon > \max\{d/4, 1-\nu\}$ *For any fixed* $\theta_1 > 0$, *let* $\hat{\sigma}_n^2$ *maximize* $\ell_{n,tap}(\sigma^2, \theta_1)$, *Then* $\hat{\sigma}_{n,tap}^2 \theta_1^{2\nu} \to \sigma_0^2 \theta_0^{2\nu}$ *with* P_0 *probability 1, where* P_0 *is the Gaussian measure defined by the Matérn covariogram corresponding to parameter values* σ_0^2, θ *and* ν.

The previous theorem says that $\hat{\sigma}_{\text{tap}}^2 \theta_1^{2\nu}$ is a strongly consistent estimator for the microergodic parameter $\sigma_0^2 \theta_0^{2\nu}$ for any fixed θ_1. Does the choice of θ_1 affect the asymptotic distribution and therefore the asymptotic efficiency of the estimator? This is a hard question to answer because the fixed-domain asymptotic distribution of the true MLE has only been established for some simple models and has not been explicitly given for general cases. It would be a harder problem to find the explicit infill asymptotic distribution for the tapered MLE. Du, Zhang and Mandrekar (2009) considered this issue which will be presented in the rest of this section.

The main results for the Ornstein-Uhlenbeck process are presented in subsection 3.3.2. For the microergodic parameter in the Ornstein-Uhlenbeck process, we establish the asymptotic distribution of tapered MLE. In subsection 3.3.3, we present the main results for a Gaussian stationary process having a Matérn covariogram. We put all proofs in the two appendices.

3.3.2 The fixed-domain asymptotics of tapered MLE for Exponential Model

We assume the underlying process $X(t), t \in [0, 1]$ is Gaussian that has a mean 0 and an isotropic exponential covariogram $K(h) = \sigma^2 \exp(-\theta h)$. Such a process is known as the Ornstein-Uhlenbeck process, which has a Markovian property that will be exploited in our proof.

The exponential isotropic covariance function is one of the most commonly used models for spatial data analysis. It follows from Ying (1991) and Zhang (2004) that both σ^2 and θ are not consistently estimable under the fixed-domain asymptotic framework, but the product $\sigma^2\theta$ is. Applying the fixed-domain asymptotic theory for spatial interpolation, Zhang (2004) showed that it is only this product, and not the individual parameters σ^2 and θ, that asymptotically affects the interpolation. Therefore, it is important to estimate this product well. In this section, we establish the asymptotic properties of the tapered MLE of this product. For simplicity of argument, we will maximize the likelihood function over $(\theta, \sigma^2) \in J = [a, b] \times [w, v]$ for some constants $0 < a \leq b$ and $0 < w \leq v$ and do not require that J contains the true parameter value (θ_0, σ_0^2). However, we do assume that $\theta_0\sigma_0^2 \in \{\theta\sigma^2, (\theta, \sigma^2) \in J\}$; that is, there exists a pair (θ, σ^2) in J such that $\theta\sigma^2 = \theta_0\sigma_0^2$.

The following two assumptions are made throughout this section:

(A1) The process is observed at points $t_{k,n} \in [0, 1], k = 1, \ldots, n$, with $0 \leq t_{1,n} < t_{2,n} < \cdots < t_{n,n} \leq 1$, and suppose that $n\Delta_{k,n}$ is bounded away from 0 and ∞, where $\Delta_{k,n} = t_{k,n} - t_{k-1,n}, k = 2, \ldots, n$. We also assume that $t_{n,n} \to 1$ and $t_{1,n} \to 0$ as $n \to \infty$.

(A2) $K_{\text{tap}}(h; \gamma)$ is an isotropic correlation function such that $K_{\text{tap}}(h; \gamma) = 0$ if $h \geq \gamma$, where $\gamma \in (0, 1)$ is a constant. Moreover, $K_{tap}(h; \gamma)$ has a bounded second derivative in $h \in (0,1)$ and $K'_{\text{tap}}(h; \gamma) = ch + o(h)$ as $h \to 0+$ for some constant c.

A taper can be any correlation function with compact support, and such correlation functions have been studied in literature [see Wu (1995), Wendland (1995 and 1998) and Gneiting (1999 and 2002)]. We believe that a large number of compactly supported correlation functions satisfy assumption (A2). Particularly, a Wendland taper is a truncated polynomial and, therefore, satisfies (A2) if the degree of the polynomial is greater than 3.

We also note that the assumption in (A2) that K_{tap} has a bounded second deriv-

ative in $h \in (0,1)$ can be weakened, so that d^2K_{tap}/dh^2 exists at any $h \in (0,\gamma)$ as long as the first derivative exists everywhere in $(0,1)$. The weakened condition will necessarily make the proof longer and, therefore, is not considered in this paper.

Before we state the main results of this section, we need to introduce some notations that will be used throughout this paper. For sequences of real positive numbers a_n and a sequence of real or random numbers b_n that may depend on model parameters, $b_n = O_u(a_n)$ if, for any n, $P(|b_n| \leq Ma_n) = 1$, for some $0 < M < \infty$, which does not depend on parameters but could be random. That is, b_n/a_n is bounded uniformly in the parameters. Similarly, we write $b_n = o_u(a_n)$ to mean that, with a probability 1, b_n/a_n converges to 0 uniformly in parameters. The following theorem compares the tapered log-likelihood function with the untapered one, and their derivatives. This theorem is essential to the establishment of the asymptotic properties of the tapered MLE to be given in the subsequent theorem.

Theorem 3.3.3. *Under the assumptions* (A1) *and* (A2), *uniformly in* $(\theta, \sigma^2) \in J$ *and with* P_0*-probability 1,*

$$\ell_{n,tap}(\theta,\sigma^2) = \ell_n(\theta,\sigma^2) + o_u(n^{1/2}), \tag{3.3.6}$$

$$\frac{\partial}{\partial\theta}\ell_{n,tap}(\theta,\sigma^2) = \frac{\partial}{\partial\theta}\ell_n(\theta,\sigma^2) + o_u(n^{1/2}), \tag{3.3.7}$$

where P_0 *is the probability measure corresponding to the true parameter values* σ_0^2, θ_0.

The next theorem establishes the strong consistency and the asymptotic distribution of the tapered MLE. Comparing the asymptotic distribution of MLE of $\sigma_0^2\theta_0$ in Ying (1993) and that in the following theorem, we see that the tapered MLE is asymptotically equally efficient.

Theorem 3.3.4. *Assume* (A1) *and* (A2) *hold, and let* $(\hat{\theta}_{n,tap}, \hat{\sigma}^2_{n,tap})$ *maximize the*

tapered likelihood function over $(\theta, \sigma^2) \in J$. *Then, as* $n \to \infty$,

$$P_0(\lim_{n\to\infty} \hat{\theta}_{n,tap}\hat{\sigma}^2_{n,tap} = \theta_0\sigma_0^2) = 1, \tag{3.3.8}$$

$$\sqrt{n}(\hat{\theta}_{n,tap}\hat{\sigma}^2_{n,tap} - \theta_0\sigma_0^2) \xrightarrow{d} N(0, 2(\theta_0\sigma_0^2)^2), \tag{3.3.9}$$

where P_0 *is the probability measure corresponding to the true parameter values* σ_0^2, θ_0.

If one of the parameters is fixed at any chosen value, we can easily get the following corollary. This will be the type of the estimator which we are able to deal with for general case.

Corollary 3.3.5. *In particular, let* $\sigma_2^2 > 0$ *and* $\theta_1 > 0$ *be predetermined constants and define* $\hat{\theta}_{n,tap,1}$, $\hat{\sigma}^2_{n,tap,2}$ *as solutions of the maximization problems*

$$\ell_{n,tap}(\hat{\theta}_{n,tap,2}, \sigma_2^2) = \sup_{\theta\in[a,b]} \ell_{n,tap}(\theta, \sigma_2^2), \tag{3.3.10}$$

and

$$\ell_{n,tap}(\theta_1, \hat{\sigma}^2_{n,tap,1}) = \sup_{\sigma^2\in[u,v]} \ell_{n,tap}(\theta_1, \sigma^2). \tag{3.3.11}$$

Then $\hat{\theta}_{n,tap,2} \to \theta_2 \triangleq \theta_0\sigma_0^2/\sigma_2^2$ *a.s. and* $\hat{\sigma}^2_{n,tap,1} \to \sigma_1^2 \triangleq \theta_0\sigma_0^2/\theta_1$ *a.s.. Moreover, as* $n \to \infty$

$$\sqrt{n}(\hat{\theta}_{n,tap,2} - \theta_2) \xrightarrow{d} N(0, 2\theta_2^2), \tag{3.3.12}$$

$$\sqrt{n}(\hat{\sigma}^2_{n,tap,1} - \sigma_1^2) \xrightarrow{d} N(0, 2\sigma_1^4). \tag{3.3.13}$$

3.3.3 Fixed-domain asymptotic distribution of MLE and tapered MLE for general Matérn model

In this section, we will focus on studying the asymptotics of tapered MLE for a general Matérn model. We assume the underlying process is stationary with mean 0 and the

following isotropic Matérn covariogram

$$K(h;\sigma^2,\theta,\nu)=\frac{\sigma^2(\theta h)^\nu}{\Gamma(\nu)2^{\nu-1}}\mathcal{K}_\nu(\theta h),\qquad h>0, \tag{3.3.14}$$

with unknown σ^2,θ and known ν, where $\mathcal{K}_\nu$ is the modified Bessel function of order ν [see Abramowitz and Stegun (1967) p.375-376], σ^2 is the covariance parameter, θ is the scale parameter and ν is the smoothness parameter. Further, assume that the process is observed at n sites $t_1,t_2,\ldots,t_n$ in a bounded interval $D\subset\mathbb{R}$, and write $\mathbf{X}_n=(X(t_1),\ldots,X(t_n))'$. Zhang (2004) noted that neither σ^2 nor θ is consistently estimable under the fixed-domain asymptotic framework, but the quantity $\sigma^2\theta^{2\nu}$ is consistently estimable. Furthermore, this consistently estimable quantity is more important to prediction than the parameters σ^2 and θ.

The primary focus of this section is to establish the asymptotic distribution of the estimators for $\sigma^2\theta^{2\nu}$. This is a more difficult problem than in the exponential case, and we cope with it by considering an easy version of the problem. Following Zhang (2004), we fix θ at an arbitrarily chosen value θ_1 and consider the following estimators:

$$\hat{\sigma}^2_n=\text{ArgMax }\ell_n(\theta_1,\sigma^2), \tag{3.3.15}$$

$$\hat{\sigma}^2_{n,tap}=\text{ArgMax }\ell_{n,tap}(\theta_1,\sigma^2), \tag{3.3.16}$$

where $\ell_n(\theta_1,\sigma^2)$ and $l_{n,tap}(\theta_1,\sigma^2)$ are the log-likelihood function and the tapered log-likelihood function, respectively.

We make the following assumption on the spectral density of the taper $K_{\text{tap}}(h)$. Similar conditions were used in Furrer et al. (2006) and Kaufman et al.(2008). Our condition here is stronger, and it is necessary for our approach to deriving the asymptotic distribution of tapered MLE:

(A3) The spectral density of the taper, denoted by $f_{\text{tap}}(\lambda)$, satisfies for some constant

$\epsilon > \max\{1/2, 1-\nu\}$ and $0 < M < \infty$

$$f_{\text{tap}}(\lambda) \leq \frac{M}{(1+\lambda^2)^{\nu+1/2+\epsilon}}. \tag{3.3.17}$$

We note that taper condition (3.3.17) is satisfied by some well-known tapers. For example, Wendland tapers (1995, 1998) have isotropic spectral densities that are continuous and satisfy $g_{d,k}(\lambda) \leq M(1+\lambda^2)^{-d/2-k-1/2}$ for some constant M, where d is the dimension of the domain ($d = 1$ in this work). Therefore, condition (3.3.17) is satisfied if $k > \max\{1/2, \nu\}$. Furrer, Genton and Nychka(2006) gave explicit tail limits for two Wendland tapers $K_1(h;\gamma) = (1-\frac{h}{\gamma})_+^4(1+4\frac{h}{\gamma})$, $\gamma > 0$ and $K_2(h;\gamma) = (1-\frac{h}{\gamma})_+^6(1+6\frac{h}{\gamma}+\frac{35h^2}{3\gamma^2})$, $\gamma > 0$ ($x_+ = \max\{0, x\}$), and showed that $\lambda^4 g_1(\lambda) \to 120/(\pi\gamma^3)$ and $\lambda^6 g_2(\lambda) \to 17920/(\pi\gamma^5)$, as $\lambda \to \infty$, where g_i is the spectral density of K_i ($i = 1, 2$). Therefore, condition (3.3.17) holds if $\nu < 1$ for taper K_1 and $\nu < 2$ for taper K_2.

One important probabilistic tool we will extensively use is the equivalence of probability measures. The assumption (A3) implies that the tapered covariance function specifies a Gaussian measure that is equivalent to the Gaussian measure specified by the true covariance function [Kaufman et al. (2008)]. It readily follows that $\hat{\sigma}^2_{n,tap}\theta_1^{2\nu}$ is a strongly consistent estimator of $\sigma_0^2\theta_0^{2\nu}$ [e.g., Kaufman, et al. (2008)].

The main results in this section are the following three theorems. The next theorem is a general result about two equivalent Gaussian measures and is not restricted to the case of Matérn model or covariance tapering. It will be used to prove the other two theorems.

Theorem 3.3.6. *Let $X(t), t \in \mathbb{R}$ be a stationary Gaussian process having mean zero and an isotropic covariogram K_j and a spectral density f_j under measure $\mathrm{P}_j, j = 0, 1$. Assume the process is observed at $t_1, t_2, \cdots$ in a bounded interval D, and let $\mathbf{X}_n = (X(t_1), \ldots X(t_n))'$. If*

$$0 < f_0(\lambda) \asymp |\lambda|^{-r_1}, \quad \lambda \to \infty \qquad \textit{for some} \qquad r_1 > 1, \tag{3.3.18}$$

and

$$h(\lambda) = \frac{f_1(\lambda)}{f_0(\lambda)} - 1 = O(|\lambda|^{-r_2}), \quad \lambda \to \infty \quad \textit{for some} \quad r_2 > 1, \tag{3.3.19}$$

then

$$\mathrm{E}_0(\mathbf{X}_n'(\mathbf{V}_{1,n}^{-1} - \mathbf{V}_{0,n}^{-1})\mathbf{X}_n) = O(1), \tag{3.3.20}$$

where $V_{j,n}$ is the covariance matrix of $\mathbf{X}_n$ given by the covariogram K_j, $j = 0, 1$, and E_0 is the expectation with respect to P_0.

We note that condition (3.3.19) is stronger than that to ensure the equivalence the equivalence of the two Gaussian measures corresponding to the two spectral densities f_0 and f_1. Indeed, under condition (3.3.18), the two Gaussian measures are equivalent if (3.3.19) holds for some $r_2 > 1/2$. However, equivalence alone can not imply (3.3.20), and we need stronger conditions than the equivalence of two measures. We will show later that condition (A3) implies that (3.3.19) holds, for some $r_2 > 1$, if f_0 and f_1 represent the spectral densities of the true and tapered covariograms, respectively.

Theorem 3.3.7. *Suppose condition (A3) is satisfied, and the underlying process is stationary Gaussian having a mean 0 and a Matérn covariance function, and the sampling locations $\{t_1, t_2, \dots\}$ are from a bounded interval. Then, for any fixed $\theta_1 > 0$, with P_0-probability 1, uniformly in $\sigma^2 \in [w, v]$,*

$$\ell_{n,tap}(\theta_1, \sigma^2) = \ell_n(\theta_1, \sigma^2) + O_u(1) \tag{3.3.21}$$

$$\frac{\partial}{\partial \sigma^2}\ell_{n,tap}(\theta_1, \sigma^2) = \frac{\partial}{\partial \sigma^2}\ell_n(\theta_1, \sigma^2) + O_u(1), \tag{3.3.22}$$

where P_0 is the probability measure corresponding to the true parameter values $\sigma_0^2, \theta_0, \nu$.

Next, we give the asymptotic distributions for both exact MLE and tapered MLE of the consistently estimable quantity $\sigma^2\theta^{2\nu}$.

Theorem 3.3.8. *Assume that the underlying process is stationary Gaussian having a mean 0 and a Matérn covariance function with a known smoothness parameter ν, and the sampling locations $\{t_1, t_2, \dots\}$ are from a bounded interval:*

(i) *For any fixed θ_1,*

$$\sqrt{n}(\hat{\sigma}_n^2\theta_1^{2\nu} - \sigma_0^2\theta_0^{2\nu}) \xrightarrow{d} N(0, 2(\sigma_0^2\theta_0^{2\nu})^2). \tag{3.3.23}$$

(ii) *In addition, if the taper satisfies condition (A3),*

$$\sqrt{n}(\hat{\sigma}_{n,tap}^2\theta_1^{2\nu} - \sigma_0^2\theta_0^{2\nu}) \xrightarrow{d} N(0, 2(\sigma_0^2\theta_0^{2\nu})^2). \tag{3.3.24}$$

Theorem 3.3.8 implies that the covariance tapering does not reduce the asymptotic efficiency for the Matérn model under condition (3.3.17). In this paper, we are not able to show that (3.3.24) remains true if θ_1 is replaced by the MLE of θ. Therefore, the results in this theorem are not as strong as those in Theorem 3.3.4. More work will need to be done to extend Theorem 3.3.4 to the general Matérn case.

We note that asymptotic distributional results about the microergodic parameter $\sigma^2\theta^{2\nu}$ in the general Matérn class have not appeared in literature. Theorem 3.3.8 (i) is the first of such results, and its proof requires a novel approach.

3.3.4 Discussion

There are some open problems for future research. First, for the Matérn model, the estimator of $\sigma^2\theta^{2\nu}$ is constructed by fixing θ at an arbitrary value. For a finite sample, common practice is to also estimate θ. It is an interesting question to see if Theorem 3.3.8 still holds for the MLE $\hat{\sigma}^2\hat{\theta}^{2\nu}$ and the tapered MLE $\hat{\sigma}_{n,tap}^2\hat{\theta}_{n,tap}^{2\nu}$. Our conjecture is that Theorem 5 can be extended to this case.

The main results in Sections 3.3.2 and 3.3.3 are for the processes with one dimensional index. It is a more interesting problem to study the high dimensional case. However, our techniques in Section 3.3.3 cannot be directly extended to obtain analogous asymptotic distribution in the high dimensional case. For example, for a d-dimensional process, we would need (3.3.19) to hold for some $r_2 > d$ in order for the proof to carry through. Unfortunately, for the Matérn model, (3.3.19) dose not hold for any $r_2 > 2$. The high dimensional case calls for new techniques for establishing asymptotic distributions. A referee suggested letting the bandwidth γ vary and go to 0 as n increases to ∞. This is a natural scheme in the fixed-domain asymptotic framework. We believe that everything in Section 3.3.2 carries through if the bandwidth goes to 0 not too fast. When the bandwidth of the taper depends on n, it is not obvious if our techniques in Section 3.3.3 still apply, because the properties of equivalence of probability measures are no longer directly applicable.

3.3.5 Appendix 1. Proofs for Section 3.3.2

In the sequel, we often suppress n in the subscripts. For example, write $t_k = t_{k,n}$, $\Delta_k = \Delta_{k,n}$. We will need three lemmas for the proofs of the theorems in Section 3.3.2.

Lemma 3.3.9. *Let $X(t)$ be the Gaussian Ornstein-Uhlenbeck process, and assume (A1) holds. Denote* $\mathrm{E}(X(t_i)|X(t_j), j \neq i) = -\sum_{j\neq i} b_{ij,n}(\theta)X(t_j), 1 \leq i \leq n$, $\mathrm{Var}(X(t_i)|X(t_j), j \neq i) = d_{i,n}(\theta, \sigma^2)$, *which is written as d_i for short. Then, for $1<i<n$,*

$$b_{ii-1,n}(\theta) = -\frac{e^{-\theta\Delta_i}(1-e^{-2\theta\Delta_{i+1}})}{1-e^{-2\theta(\Delta_i+\Delta_{i+1})}},\ b_{ii+1,n}(\theta) = -\frac{e^{-\theta\Delta_{i+1}}(1-e^{-2\theta\Delta_i})}{1-e^{-2\theta(\Delta_i+\Delta_{i+1})}}, \quad (3.3.25)$$

$$b_{12,n}(\theta) = -e^{-\theta\Delta_2},\ b_{nn-1,n}(\theta) = -e^{-\theta\Delta_n},\ b_{ij,n}(\theta) = 0 \quad \textit{for} \quad |i-j|>1. \quad (3.3.26)$$

In addition, uniformly in $(\theta, \sigma^2) \in J$, $1 < i < n$, $1 \le j \le n$,

$$d_1 = 2\sigma^2\theta\Delta_2 + O_u(\frac{1}{n^2}),\ d_n = 2\sigma^2\theta\Delta_n + O_u(\frac{1}{n^2}),\ d_i = \frac{2\sigma^2\theta\Delta_i\Delta_{i+1}}{\Delta_i + \Delta_{i+1}} + O_u(\frac{1}{n^2}), \quad (3.3.27)$$

$$b'_{ij,n}(\theta) = O_u(\frac{1}{n^2}),\ b'_{1j,n}(\theta) = O_u(\frac{1}{n}),\ b'_{nj,n}(\theta) = O_u(\frac{1}{n}),\ \frac{\partial}{\partial\theta} d_j^{-1} = O_u(n). \quad (3.3.28)$$

Proof. Note that $\mathrm{E}(X(t_i)|X(t_j), j \neq i) = -\sum_{j\neq i} b_{ij,n}(\theta)X(t_j), 1 \le i \le n$ if and only if

$$Cov(X(t_i) + \sum_{k\neq i} b_{ik,n}(\theta)X(t_k), X(t_j)) = 0, \text{ for any } j \neq i, j = 1, \ldots, n.$$

We therefore prove (3.3.25) and (3.3.26) by verifying that

$$\mathrm{Cov}\left(X(t_i) + b_{i,i-1}X(t_{i-1}) + b_{i,i+1}X(t_{i+1}), X(t_j)\right) = 0, \text{ for any } j \neq i, \quad (3.3.29)$$

where we let $b_{10} = b_{n,n+1} = 0$. For $i = 1$ or n, (3.3.29) readily follows the stationarity and the Markovian property of the Ornstein-Uhlenbeck process. For $1 < i < n$, (3.3.29) holds, because, if $j \ge i + 1$, the LHS of (3.3.29) equals

$$\sigma^2 e^{-\theta(t_j - t_{i+1})}\left(e^{-\theta\Delta_{i+1}} - \frac{e^{-\theta\Delta_i}(1 - e^{-2\theta\Delta_{i+1}})}{1 - e^{-2\theta(\Delta_i+\Delta_{i+1})}} e^{-\theta(\Delta_i+\Delta_{i+1})} - \frac{e^{-\theta\Delta_{i+1}}(1 - e^{-2\theta\Delta_i})}{1 - e^{-2\theta(\Delta_i+\Delta_{i+1})}}\right),$$

which is zero. We can get the similar expression when $j \le i - 1$. Therefore, (3.3.29) is proved. Since $d_i = \mathrm{E}(X(t_i) + b_{i,i-1}X(t_{i-1}) + b_{i,i+1}X(t_{i+1}))^2$, straightforward calculation yields

$$d_1 = \sigma^2(1 - e^{-2\theta\Delta_2}),\ d_n = \sigma^2(1 - e^{-2\theta\Delta_n}),\ d_i = \sigma^2\frac{(1 - e^{-2\theta\Delta_i})(1 - e^{-2\theta\Delta_{i+1}})}{1 - e^{-2\theta(\Delta_i+\Delta_{i+1})}}, 1 < i < n.$$

Then, (3.3.27) follows the Taylor expansion. To establish the properties of the derivatives in (3.3.28), we repeatedly use the Taylor expansion. Here we only provide a proof for $b'_{ij,n}(\theta) = O_u(1/n^2)$ for $1 < i < n$, since all other derivatives can be proved

similarly. Since $b_{ij} = 0$ for $|i-j| > 1$, we only need to consider $j = i-1$ or $i+1$. Write the derivative

$$b'_{ii-1,n}(\theta) = A/(1-e^{-2\theta(\Delta_i+\Delta_{i+1})})^2, \text{ where}$$

$$\begin{aligned}
A =&(\Delta_i e^{-\theta\Delta_i} - (\Delta_i + 2\Delta_{i+1})e^{-\theta\Delta_i - 2\theta\Delta_{i+1}})(1-e^{-2\theta(\Delta_i+\Delta_{i+1})}) \\
&- (-e^{-\theta\Delta_i} + e^{-\theta\Delta_i - 2\theta\Delta_{i+1}})2(\Delta_i+\Delta_{i+1})e^{-2\theta(\Delta_i+\Delta_{i+1})} \\
=&\Delta_i e^{-\theta\Delta_i} - (\Delta_i + 2\Delta_{i+1})\big(e^{-\theta\Delta_i - 2\theta\Delta_{i+1}} - e^{-3\theta\Delta_i - 2\theta\Delta_{i+1}}\big) \\
&- \Delta_i e^{-3\theta\Delta_i - 4\theta\Delta_{i+1}}. \qquad (3.3.30)
\end{aligned}$$

Note that

$$\left|\frac{1}{1-e^{-2\theta(\Delta_i+\Delta_{i+1})}} - \frac{1}{2\theta(\Delta_i+\Delta_{i+1})}\right|$$

is uniformly bounded and $n(\Delta_i + \Delta_{i+1})$ is bounded away from 0 and ∞ by Assumption (A1). Hence, $1/(1-e^{-2\theta(\Delta_i+\Delta_{i+1})}) = O(1/n)$, and it suffices to show that A is $O_u(1/n^4)$. Using, again, the fact that $\Delta_i = O_u(1/n)$ and applying the Taylor expansion, we get

$$\begin{aligned}
&\Delta_i e^{-\theta\Delta_i} = \Delta_i - \theta{\Delta_i}^2 + (1/2)\,\theta^2{\Delta_i}^3 + O_u(1/n^4), \\
&- (\Delta_i + 2\Delta_{i+1})\big(e^{-\theta\Delta_i - 2\theta\Delta_{i+1}} - e^{-3\theta\Delta_i - 2\theta\Delta_{i+1}}\big) \\
&= -2\,\theta{\Delta_i}^2 + 4\,\theta^2{\Delta_i}^3 + 12\,\theta^2{\Delta_i}^2\Delta_{i+1} - 4\,\theta\Delta_i\Delta_{i+1} + 8\,\theta^2\Delta_i{\Delta_{i+1}}^2 + O_u(1/n^4), \\
&- \Delta_i e^{-3\theta\Delta_i - 4\theta\Delta_{i+1}} = \\
&- \Delta_i + 3\,\theta{\Delta_i}^2 + 4\theta\Delta_i\Delta_{i+1} - \frac{9}{2}\theta^2{\Delta_i}^3 - 12\theta^2{\Delta_i}^2\Delta_{i+1} - 8\theta^2\Delta_i{\Delta_{i+1}}^2 + O_u(1/n^4).
\end{aligned}$$

All the terms except $O_u(1/n^4)$ are canceled out. Therefore, $A = O_u(1/n^4)$ and $b'_{ii-1,n}(\theta) = O_u(1/n^2)$. Similarly, we can show $b'_{ii+1,n}(\theta) = O_u(1/n^2)$. □

We now introduce the following notations. Let $\tilde{\mathbf{O}}_n$ denote a matrix of which the elements are $O_u(1/n)$ except those in the first and last rows, which are uniformly

bounded; that is, $O_u(1)$. Denote, by $\breve{\mathbf{O}}_n$, the matrix whose (i,j)th element is $O_u(1)$ if $i = 1$ or n or $i = j$, and is $O_u(1/n)$ otherwise. Therefore,

$$\tilde{\mathbf{O}}_n = \begin{pmatrix} O_u(1) & \dots & O_u(1) \\ O_u(1/n) & \dots & O_u(1/n) \\ \dots & \dots & \dots \\ O_u(1/n) & \dots & O_u(1/n) \\ O_u(1) & \dots & O_u(1) \end{pmatrix}, \quad \breve{\mathbf{O}}_n = \tilde{\mathbf{O}}_n + \begin{pmatrix} O_u(1) & & \\ & \ddots & \\ & & O_u(1) \end{pmatrix}.$$

Lemma 3.3.10. *Under assumptions (A1) and (A2), uniformly in $\theta \in [a,b]$:*

(i) $\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n) = \mathbf{I}_n + \tilde{\mathbf{O}}_n, \quad \mathbf{V_n^{-1}}\frac{\partial \mathbf{V}_n}{\partial \theta} = \breve{\mathbf{O}}_n,$

(ii) $\frac{\partial}{\partial \theta}(\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n)) = \tilde{\mathbf{O}}_n, \quad \frac{\partial}{\partial \theta}(\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial \theta}) = \breve{\mathbf{O}}_n,$

(iii) $1 < \det(\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n)) = O_u(1), \quad (\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n))^{-1} = \mathbf{I_n} + \tilde{\mathbf{O}}_n,$

where $\mathbf{I_n}$ is the $n \times n$ identity matrix.

From the definitions of $\tilde{\mathbf{O}}_n$ and $\breve{\mathbf{O}}_n$, we have

$$\tilde{\mathbf{O}}_n\breve{\mathbf{O}}_n = \tilde{\mathbf{O}}_n, \quad \breve{\mathbf{O}}_n\tilde{\mathbf{O}}_n = \tilde{\mathbf{O}}_n, \quad \tilde{\mathbf{O}}_n\tilde{\mathbf{O}}_n = \tilde{\mathbf{O}}_n. \tag{3.3.31}$$

Then, Lemma 2 (i) and (ii) and the following well known fact [see, e.g., Graybill (1983) pp.357-358]:

$$\frac{\partial}{\partial \theta}\mathbf{V}_n^{-1} = -\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial \theta}\mathbf{V}_n^{-1}. \tag{3.3.32}$$

imply

$$\frac{\partial}{\partial \theta}(\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n))^{-1} = \tilde{\mathbf{O}}_n. \tag{3.3.33}$$

Proof. We can assume $\sigma^2 = 1$ without loss of any generality, because all quantities in the lemma do not depend on σ^2. We will repeatedly use Lemma 1 and particularly

the fact that $\mathbf{V}_n^{-1}$ is a band matrix. The proof involves tedious computation, and we will keep a balance between brevity and clarity.

Several quantities in the Lemma are of the form $\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{Q})$, where $\mathbf{Q}$ is an $n \times n$ matrix whose (i,j)th element is $\varrho(t_i - t_j)$ for some even function $\varrho(t)$ that has a bounded second derivative on $[-1,0) \cup (0,1]$. If the limits of the derivative $\varrho'(0+) = \lim_{t \to +} \varrho'(t)$ and $\varrho'(0-) = \lim_{t \to 0-} \varrho'(t)$ exist and are finite, we show now

$$\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{Q}) = \varrho(0)\mathbf{I}_n + (\varrho'(0+) - \varrho'(0-))\text{diag}\{O_u(1), \ldots, O_u(1)\} + \tilde{\mathbf{O}}_n, \quad (3.3.34)$$

where $\text{diag}(O_u(1), \ldots, O_u(1))$ denotes a diagonal $n \times n$ matrix with bounded elements. There are immediate corollaries from (3.3.34). First, it implies $\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{Q}) = \breve{\mathbf{O}}_n$. Second, by taking $\varrho(t) = -|t|$, we get $\mathbf{V}_n^{-1}(\partial \mathbf{V}_n / \partial\theta) = \breve{\mathbf{O}}_n$. Lastly, if $\varrho(t) = K_{\text{tap}}(|t|)$, then $\varrho'(0+) = \varrho'(0-)$ and $\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n) = \mathbf{I}_n + \tilde{\mathbf{O}}_n$ because $K_{\text{tap}}(0) = 1$.

To prove (3.3.34) we need the following well known result [e.g., Ripley (1981) p.89]:

$$\mathbf{V}_n^{-1}(\theta, \sigma^2) = \mathbf{D}_n^{-1}(\theta, \sigma^2)\mathbf{B}_n(\theta), \quad (3.3.35)$$

where $\mathbf{B}_n(\theta) = (b_{ij,n}(\theta))_{1 \le i,j \le n}$ and $\mathbf{D_n}(\theta, \sigma^2) = \text{diag}\{d_i(\theta, \sigma^2), i = 1, \ldots, n\}$, in which $b_{ij,n}(\theta)$, $d_i(\theta, \sigma^2)$ are defined as in Lemma 3.3.9 and $b_{ii,n}(\theta) = 1$. let ω_{ij} denote the (i,j)th element of $\mathbf{V}_n^{-1}(\mathbf{V_n} \circ \mathbf{Q})$. Hereafter, for the ease of notation, we will suppress the dependence of any quantity on n and parameters [e.g. $b_{ij,n}(\theta) = b_{ij}$, $d_{i,n}(\theta, \sigma^2) = d_i$], wherever confusion does not arise, throughout the rest of the paper. Write $b_{ij} = 0$ if $j < 1$ or $j > n$ and $t_0 = t_1, t_{n+1} = t_n$. Since $b_{ij} = 0$ if $|i - j| > 1$,

$$\omega_{ij} = d_i^{-1} \sum_{k=i-1}^{i+1} b_{ik} K(|t_k - t_j|)\varrho(t_k - t_j). \quad (3.3.36)$$

For any $i > j$, and $k = i-1$ or $i+1$, we have $t_k - t_j \ge 0$. Hence, the Taylor Theorem

implies

$$\varrho(t_k - t_j) = \varrho(t_i - t_j) + \varrho'(t_i - t_j)(t_k - t_i) + \varrho''(t_i - t_j + \xi(t_k - t_j))(t_k - t_i)^2/2,$$

for some $\xi \in (0,1)$. Since ϱ has a bounded second derivative on $(0,1)$, and $t_k - t_i = O(1/n)$, we have

$$\varrho(t_k - t_j) = \varrho(t_i - t_j) + \varrho'(t_i - t_j)(t_k - t_i) + O(1/n^2). \tag{3.3.37}$$

Here then,

$$\begin{aligned} \omega_{ij} = d_i^{-1}\varrho(t_i - t_j) \sum_{k=i-1}^{i+1} b_{ik}K(t_k - t_j) \\ + d_i^{-1}\varrho'(t_i - t_j) \sum_{k=i-1}^{i+1} K(t_k - t_j)(t_k - t_i)b_{ik} + O_u(1/n), \end{aligned} \tag{3.3.38}$$

where we have used $d_i^{-1} = O_u(n)$. Note that $d_i^{-1}\sum_{k=i-1}^{i+1} b_{ik}K(t_k - t_j)$ is the (i,j) element of $\mathbf{D}_n^{-1}\mathbf{B}_n\mathbf{V}_n = \mathbf{I}_n$. Hence, the first summand in (3.3.38) equals $\varrho(0)1_{\{i=j\}}$.

Similar to the establishment of (3.3.37), we can show

$$K(t_k - t_j) = K(t_i - t_j) + K'(t_i - t_j)(t_k - t_i) + O_u(1/n^2).$$

It follows that, for $i > j$,

$$\omega_{ij} = d_i^{-1}\varrho'(t_i - t_j)K(t_i - t_j) \sum_{k=i-1}^{i+1} (t_k - t_i)b_{ik} + O_u(1/n), \tag{3.3.39}$$

By utilizing the explicit expressions of b_{ij} given in Lemma 1, we can show

$$\sum_{k=i-1}^{i+1} (t_k - t_i)b_{ik} = b_{i,i+1}\Delta_{i+1} - b_{i,i-1}\Delta_i = \begin{cases} O_u(1/n^2) & \text{if } 1 < i < n \\ O_u(1/n) & \text{if } i = 1 \text{ or } n. \end{cases} \tag{3.3.40}$$

Then, for $i > j$

$$\omega_{ij} = \begin{cases} O_u(1) & \text{if } i = 1 \text{ or } n \\ O_u(1/n) & \text{if } 1 < i < n. \end{cases} \tag{3.3.41}$$

In view of the fact that ϱ is an even function, we can show, similarly, that (3.3.41) holds for $i < j$.

Now, let us consider w_{ii}. First, note that

$$\varrho(t_{i-1} - t_i) = \varrho(0) + \varrho'(0-)(t_{i-1} - t_i) + O(1/n^2), \tag{3.3.42}$$

$$\varrho(t_{i+1} - t_i) = \varrho(0) + \varrho'(0+)(t_{i+1} - t_i) + O(1/n^2). \tag{3.3.43}$$

Since $d_i^{-1} \sum_{k=i-1}^{i+1} b_{ik} K(t_k - t_i) = 1$,

$$\begin{aligned} \omega_{ii} &= d_i^{-1} \sum_{k=i-1}^{i+1} b_{ik} K(t_k - t_i) \varrho(t_k - t_i) \\ &= \varrho(0) + d_i^{-1} \{ b_{i,i-1} K(t_{i-1} - t_i) \varrho'(0-)(t_{i-1} - t_i) \\ &\quad + b_{i,i+1} K(t_{i+1} - t_i) \varrho'(0+)(t_{i+1} - t_i) \} + O_u(1/n^2). \end{aligned}$$

Since $K(h) = K(0) + K'(0)h + o_u(h)$ as $h \to 0$,

$$\omega_{ii} = \varrho(0) + K(0) d_i^{-1} \{ b_{i,i-1} \varrho'(0-)(t_{i-1} - t_i) + b_{i,i+1} \varrho'(0+)(t_{i+1} - t_i) \} + O_u(1/n^2),$$

which can be rewritten as

$$\begin{aligned} \omega_{ii} =& \varrho(0) + \varrho'(0-) K(0) d_i^{-1} \sum_{k=i-1}^{i+1} (t_k - t_i) b_{ik} \\ &+ [\varrho'(0+) - \varrho'(0-)] K(0) d_i^{-1} b_{i,i+1} \Delta_{i+1} + O_u(1/n^2). \end{aligned} \tag{3.3.44}$$

Then, (3.3.34) follows from (3.3.40), (3.3.41) and (3.3.44). **(i)** is therefore proved.

To prove (ii), note that (3.3.60) and Lemma 3.3.10(i) imply,,

$$\begin{aligned}\frac{\partial}{\partial\theta}(\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n)) &= -\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta}\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n) + \mathbf{V}_n^{-1}\left(\frac{\partial \mathbf{V}_n}{\partial\theta} \circ \mathbf{T}_n\right) \\ &= -\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta}(\mathbf{I}_n + \tilde{\mathbf{O}}_n) + \mathbf{V}_n^{-1}\left(\frac{\partial \mathbf{V}_n}{\partial\theta} \circ \mathbf{T}_n\right) \\ &= \mathbf{V}_n^{-1}\left(\frac{\partial \mathbf{V}_n}{\partial\theta} \circ (\mathbf{T}_n - \mathbf{J}_n)\right) + \tilde{\mathbf{O}}_n,\end{aligned}$$

which is clearly $\tilde{\mathbf{O}}_n$ from (3.3.34) by taking $\varrho(t) = -|t|\,(K_{\text{tap}}(|t|) - 1)$ that is differentiable at 0, where $\mathbf{J}_n$ is a matrix of all 1's.

Next, we will show $\frac{\partial}{\partial\theta}(\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta}) = \check{\mathbf{O}}_n$ similarly. Write

$$\frac{\partial}{\partial\theta}(\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta}) = -\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta}\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta} + \mathbf{V}_n^{-1}\frac{\partial^2 \mathbf{V}_n}{\partial\theta^2}. \tag{3.3.45}$$

By **(i)**, the first term on the RHS of (3.3.45) is $\check{\mathbf{O}}_n\check{\mathbf{O}}_n = \check{\mathbf{O}}_n$, and the second term is $\tilde{\mathbf{O}}_n$, because $\mathbf{V}_n^{-1}\frac{\partial^2 \mathbf{V}_n}{\partial\theta^2} = \mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{Q})$ with $\varrho(t) = t^2$, which has a continuous second derivative so that (3.3.34) applies. This completes the proof of Lemma 3.3.10 **(ii)**.

Let $\mathbf{A}_n = \mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n)$ and a_{ij} denote the (i,j)th element of $\mathbf{A}_n$. We now apply a series of column operations, so that $\mathbf{A}_n$ becomes $\mathbf{I}_n + \mathbf{\Omega}_n$ and each of the operations retains the determinant of $\mathbf{A}_n$, where $\mathbf{\Omega}_n$ is a matrix whose elements are bounded by M/n for some constant M not depending on θ; that is, $\mathbf{\Omega}_n(i,j) \leq M/n$. We have shown that $\mathbf{A}_n = \mathbf{I}_n + \tilde{\mathbf{O}}_n$, where elements of $\tilde{\mathbf{O}}_n$ are $O_u(1/n)$ except those that are on the first and last rows that are bounded. We can subtract from the jth column the first column multiplied by the $(1,j)$th element of $\mathbf{A}_n$, $2 < j < n$. Then, all elements in the first row are $O_u(1/n)$, except the $(1,1)$th element, which is $1 + O_u(1/n)$ and remains unchanged throughout the operations. Similarly, we can reduce the elements in the last row to $O_u(1/n)$ except the last (n,n)th element. Applying the Hadamard inequality [Bellman (1970) p.130], we can show there exists some constant M such

that

$$\det(\mathbf{A}_n) = \det(\mathbf{I}_n + \mathbf{\Omega}_n) \leq \left((1 + M/n)^2 + (n-1)(M/n)^2\right)^{n/2},$$

which is bounded.

To show $\mathbf{A}_n^{-1} = \mathbf{I}_n + \tilde{\mathbf{O}}_n$, we first note that by Oppenheim's inequality [Mirsky (1955) p.421], which yields the inequality for the determinant of Hadamard product of positive definite matrices, $\det(\mathbf{V}_n \circ \mathbf{T}_n) > \det(\mathbf{V}_n) \prod_{1 \leq i \leq n} t_{ii}$ where t_{ii} is the diagonal element of $\mathbf{T}_n$ and it is one. Therefore, $\det(\mathbf{A}_n) > 1$. We only need to show that the (i,j)th cofactor

$$A_{ij} = \det(\mathbf{A}_n) 1_{\{i=j\}} + O_u(1/n) + 1_{\{j=1 \text{ or } j=n\}} O_u(1).$$

Similar to proving $\det(\mathbf{A}_n) = O_u(1)$, we can show all the $(n-1)$ by $(n-1)$ cofactors are also $O_u(1)$. In addition, $A_{ij} = O_u(1/n)$ for $1 < j < n, i \neq j$ since it has one row of elements $O_u(1/n)$ and replacing that row with $O_u(1)$ would yield a bounded determinant. To complete proof of the Lemma, it remains to show $A_{ii} = \det(\mathbf{A}_n) + O_u(1/n), 1 < i < n$, which is true by Laplace expansion $\det(\mathbf{A}_n) = (1 + O_u(1/n))A_{ii} + \sum_{j:j \neq i} a_{ij} A_{ij}$ and observing that $\sum_{j:j \neq i} a_{ij} A_{ij} = O_u(1/n)$, for $1 < i < n$. □

Lemma 3.3.11. *For any $\theta \in [a,b]$, let $S_n(\theta)$, $n = 1, 2, \ldots$, be a sequence of random variables such that $\mathrm{E}(S_n(\theta)) = O_u((\log n)^r)$, $\mathrm{E}[S_n(\theta) - \mathrm{E}\, S_n(\theta)]^6 = O_u((\log n)^r)$ uniformly in θ for some constant $r > 0$. Assume that, with probability one, $S_n(\theta)$ is differentiable with respect to θ and $S_n'(\theta) = O_u(n^2 (\log n)^r)$ uniformly in θ. Then,*

$$\sup_{\theta \in [a,b]} |S_n(\theta)| = o(n^{1/2}) \quad a.s.$$

Proof. Let $a = \theta_0 < \theta_1 < \cdots < \theta_{M_n} = b$ partition $[a,b]$ into intervals of equal length, where M_n is the integer part of $n^{3/2+\alpha}$ for some $0 < \alpha < 1/14$. Then,

$$\sup_{\theta \in [a,b]} |S_n(\theta)| \leq \max_{1 \leq k \leq M_n} |S_n(\theta_k)| + \max_{1 \leq k \leq M_n} \sup_{\theta \in [\theta_{k-1}, \theta_k]} |S_n(\theta_k) - S_n(\theta)|. \quad (3.3.46)$$

Because there exists constant $C > 0$ such that the sixth central moment of $S_n(\theta)$ is uniformly bounded by $C(\log n)^r$,

$$\begin{aligned} \mathrm{P}\left(\max_{1\le k\le M_n} |S_n(\theta_k) - \mathrm{E}(S_n(\theta_k))| \ge n^{\frac{1}{2}-\alpha}\right) &\le \sum_{k=1}^{M_n} \mathrm{P}\left(|S_n(\theta_k) - \mathrm{E}(S_n(\theta_k))| \ge n^{\frac{1}{2}-\alpha}\right) \\ &\le M_n \frac{C(\log n)^r}{n^{3-6\alpha}} = \frac{C(\log n)^r}{n^{3/2-7\alpha}}. \end{aligned}$$

Since $3/2 - 7\alpha > 1$ with $\alpha < 1/14$, it follows from Borel-Cantelli lemma that

$$\max_{1\le k\le M_n} |S_n(\theta_k) - \mathrm{E}(S_n(\theta_k))| = O(n^{1/2-\alpha}) \quad a.s.$$

Since $\mathrm{E}(S_n(\theta)) = O((\log n)^r)$ uniformly in θ, and

$$\max_{1\le k\le M_n} |S_n(\theta_k)| \le \max_{1\le k\le M_n} |S_n(\theta_k) - \mathrm{E}(S_n(\theta_k))| + \max_{1\le k\le M_n} |\mathrm{E}(S_n(\theta_k))|,$$

then

$$\max_{1\le k\le M_n} |S_n(\theta_k)| = O(n^{1/2-\alpha}) \quad a.s. \tag{3.3.47}$$

On the other hand, for $\theta \in [\theta_{k-1}, \theta_k]$,

$$|S_n(\theta_k) - S_n(\theta)| \le \sup_{\theta\in[a,b]} |S_n'(\theta)|(\theta_k - \theta_{k-1}) = O_u(n^{1/2-\alpha}(\log n)^r) \ a.s.$$

Therefore,

$$\max_{1\le k\le M_n} \sup_{\theta\in[\theta_{k-1},\theta_k]} |S_n(\theta_k) - S_n(\theta)| = o(n^{1/2}) \quad a.s. \tag{3.3.48}$$

The proof is completed by combining (3.3.46), (3.3.47) and (3.3.48). □

Proof of Theorem 3.3.3. Recall that the tapered and untapered log likelihoods are given by (3.3.4) and (3.3.2), respectively. The proof of (3.3.6) consists of direct comparisons of the log determinants and the two quadratic forms. First, Lemma 3.3.10

(iii) implies that

$$\log[\det(\mathbf{V}_n \circ \mathbf{T}_n)] = \log[\det(\mathbf{V}_n)] + O_u(1). \tag{3.3.49}$$

Define $\mathbf{H}_n(\theta) = (\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n))^{-1} - \mathbf{I}_n$. Then, $\mathbf{H}_n = \tilde{\mathbf{O}}_n$ by Lemma 3.3.10 **(iii)**. Because

$$(\mathbf{V}_n \circ \mathbf{T}_n)^{-1} = \mathbf{V}_n^{-1} + \mathbf{H}_n\mathbf{V}_n^{-1}, \tag{3.3.50}$$

$$\mathbf{X}_n'(\mathbf{V}_n \circ \mathbf{T}_n)^{-1}\mathbf{X}_n = \mathbf{X}_n'\mathbf{V}_n^{-1}\mathbf{X}_n + \mathbf{X}_n'\mathbf{H}_n\mathbf{V}_n^{-1}\mathbf{X}_n. \tag{3.3.51}$$

Proof of (3.3.6) would be completed if, uniformly in (θ, σ^2), with probability 1,

$$\mathbf{X}_n'\mathbf{H}_n\mathbf{V}_n^{-1}\mathbf{X}_n = o_u(n^{1/2}). \tag{3.3.52}$$

We will apply Lemma 3 to prove (3.3.52). Define

$$S_n(\theta) = \sigma^2\mathbf{X}_n'\mathbf{H}_n\mathbf{V}_n^{-1}\mathbf{X}_n,$$

and note that $S_n(\theta)$ depends on θ but not on σ^2. In view of symmetry of $\mathbf{H}_n\mathbf{V}_n^{-1}$ by (3.3.50), we can write

$$E_0 S_n(\theta) = \sigma^2 \operatorname{trace}\{\mathbf{H}_n\mathbf{V}_n^{-1}\mathbf{V}_{n,0}\}, \tag{3.3.53}$$

where hereafter in the proof the expectation is evaluated under the true parameter σ_0^2 and θ_0, and $\mathbf{V}_{n,0} = \mathbf{V}_n(\theta_0, \sigma_0^2)$. The rth cumulant of $\mathbf{X}_n'\mathbf{H}_n\mathbf{V}_n^{-1}\mathbf{X}_n$

$$\kappa_r = 2^{r-1}(r-1)!\operatorname{trace}\{\mathbf{H}_n\mathbf{V}_n^{-1}\mathbf{V}_{n,0}\}^r, \tag{3.3.54}$$

$r = 1, 2, \ldots$ [see Searle (1971), Theorem 1, p.55].

Next, we show that

$$\mathbf{V}_n^{-1}(\theta, \sigma^2)\mathbf{V}_n(\theta_0, \sigma_0^2) = O_u(1)\mathbf{I}_n + \tilde{\mathbf{O}}_n. \tag{3.3.55}$$

Then, it follows from (3.3.53)-(3.3.55) and (3.3.31) that the first moment and the sixth central moment of $S_n(\theta)$ are uniformly bounded, because the sixth central moment of $S_n(\theta)$ is $\kappa_6 + 15\kappa_4\kappa_2 + 10\kappa_3^2 + 15\kappa_2^2$, which is uniformly bounded because all of the four cumulants involved are uniformly bounded.

By using notations introduced in proof of Lemma 3.3.9, we now give explicit expression for the elements of $\mathbf{V}_n^{-1}(\theta, \sigma^2)$ based on Lemma 3.3.9 and (3.3.35).

For brevity, we drop the parameters in the matrices and write $\mathbf{B}_{n,0} = \mathbf{B}_n(\theta_0)$ and $\mathbf{D}_{n,0} = \mathbf{D}_n(\theta_0, \sigma_0^2)$. Decompose $\mathbf{V}_n^{-1}$ into

$$\mathbf{V}_n^{-1} = \mathbf{D}_n^{-1}\mathbf{B}_{n,0} + \mathbf{D}_n^{-1}(\mathbf{B}_{n,0} - \mathbf{B}_n) = \mathbf{A}_1 + \mathbf{A}_2.$$

Then,

$$\mathbf{A}_1\mathbf{V}_{n,0} = \mathbf{D}_n^{-1}\mathbf{B}_{n,0}\mathbf{B}_{n,0}^{-1}\mathbf{D}_{n,0} = \mathbf{D}_n^{-1}\mathbf{D}_{n,0}$$

and the diagonals of $\mathbf{D}_n^{-1}\mathbf{D}_{n,0}$ converge uniformly to $(\sigma_0^2\theta_0)/(\sigma^2\theta)$ by (3.3.27). Therefore,

$$\mathbf{A}_1\mathbf{V}_{n,0} = \mathrm{diag}(O_u(1), \ldots, O_u(1)). \tag{3.3.56}$$

In addition,

$$\mathbf{A}_2\mathbf{V}_{n,0} = \tilde{\mathbf{O}}_n \quad \text{uniformly in } \theta \in [a, b]. \tag{3.3.57}$$

Indeed, the absolute value of the (i, j) th element of $\mathbf{A}_2\mathbf{V}_{n,0}$, $1 < i < n$ is

$$\begin{aligned} &\left| \sum_{k=1}^{n} d_i^{-1}(b_{ik,n}(\theta) - b_{ik,n}(\theta_0))\sigma_0^2 e^{-\theta_0|t_k - t_j|} \right| \\ &\leq d_i^{-1}\sigma_0^2 \sum_{|k-i|\leq 1} |b_{ik,n}(\theta) - b_{ik,n}(\theta_0)| = O_u(\frac{1}{n}), \end{aligned} \tag{3.3.58}$$

where the last equality follows from (3.3.27), (3.3.28) and the Taylor Theorem. Similarly, we can show the elements on the first and last rows are $O_u(1)$. Hence, (3.3.55) follows from (3.3.56), (3.3.57) immediately. Lastly, note that $\frac{\partial}{\partial\theta}S_n(\theta) = O_u(n^2)$ by Lemma 3.3.10. The conditions of Lemma 3.3.11 are satisfied. Therefore,

$$\sup_{\theta\in[a,b]} S_n(\theta) = o(n^{1/2}), \tag{3.3.59}$$

which implies (3.3.52).

We have now proved (3.3.6). (3.3.7) can be proved similarly, and the remaining proof will be brief. By using the following facts in matrix algebra

$$\begin{aligned}
\frac{\partial}{\partial\theta}\log[\det(\mathbf{V}_n)] &= \operatorname{trace}\{\mathbf{V}_n^{-1}\frac{\partial\mathbf{V}_n}{\partial\theta}\},\\
\frac{\partial}{\partial\theta}\mathbf{V}_n^{-1} &= -\mathbf{V}_n^{-1}\frac{\partial\mathbf{V}_n}{\partial\theta}\mathbf{V}_n^{-1}, &(3.3.60)\\
\frac{\partial}{\partial\theta}(\mathbf{V}_n\circ\mathbf{T}_n) &= \frac{\partial\mathbf{V}_n}{\partial\theta}\circ\mathbf{T}_n, &(3.3.61)
\end{aligned}$$

The derivatives of the log likelihood functions can be written as

$$\begin{aligned}
\frac{\partial}{\partial\theta}\ell_n(\theta,\sigma^2) &= -\operatorname{trace}\{\mathbf{V}_n^{-1}\frac{\partial\mathbf{V}_n}{\partial\theta}\} + \mathbf{X}_n'\mathbf{V}_n^{-1}\frac{\partial\mathbf{V}_n}{\partial\theta}\mathbf{V}_n^{-1}\mathbf{X}_n, &(3.3.62)\\
\frac{\partial}{\partial\theta}\ell_{\mathrm{n,tap}}(\theta,\sigma^2) &= -\operatorname{trace}\{(\mathbf{V}_n\circ\mathbf{T}_n)^{-1}(\frac{\partial\mathbf{V}_n}{\partial\theta}\circ\mathbf{T}_n)\}\\
&\quad + \mathbf{X}_n'(\mathbf{V}_n\circ\mathbf{T}_n)^{-1}(\frac{\partial\mathbf{V}_n}{\partial\theta}\circ\mathbf{T}_n)(\mathbf{V}_n\circ\mathbf{T}_n)^{-1}\mathbf{X}_n. &(3.3.63)
\end{aligned}$$

We first show that the two traces differ by $O_u(1)$. Write $\mathbf{A}_n = \mathbf{V}_n^{-1}(\mathbf{V}_n\circ\mathbf{T}_n)$. It is straightforward to verify

$$(\mathbf{V}_n\circ\mathbf{T}_n)^{-1}(\frac{\partial\mathbf{V}_n}{\partial\theta}\circ\mathbf{T}_n) = \mathbf{A}_n^{-1}\mathbf{V}_n^{-1}\frac{\partial\mathbf{V}_n}{\partial\theta}\mathbf{A}_n + \mathbf{A}_n^{-1}\frac{\partial\mathbf{A}_n}{\partial\theta}.$$

Then,

$$\text{trace}\{(\mathbf{V}_n \circ \mathbf{T}_n)^{-1}(\frac{\partial \mathbf{V}_n}{\partial \theta} \circ \mathbf{T}_n)\} = \text{trace}\{\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial \theta}\} + \text{trace}(\mathbf{A}_n^{-1}\frac{\partial \mathbf{A}_n}{\partial \theta}), \quad (3.3.64)$$

where the second trace in the RHS is clearly uniformly bounded by Lemma 3.3.10. Similarly, we can write

$$(\mathbf{V}_n \circ \mathbf{T}_n)^{-1}(\frac{\partial \mathbf{V}_n}{\partial \theta} \circ \mathbf{T}_n)(\mathbf{V}_n \circ \mathbf{T}_n)^{-1} = \mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial \theta}\mathbf{V}_n^{-1} + \mathbf{W}_n\mathbf{V}_n^{-1}$$

for some matrix $\mathbf{W}_n$, which is $\tilde{\mathbf{O}}_n$. Again note that $\mathbf{W}_n\mathbf{V}_n^{-1}$ is symmetric, using the exact same same technique for deriving (3.3.52), we can show

$$\mathbf{X}_n'\mathbf{W}_n\mathbf{V}_n^{-1}\mathbf{X}_n = o_u(n^{1/2}). \quad (3.3.65)$$

The proof is complete. □

Proof of Theorem 3.3.4. First, for (3.3.8), it suffices to show that, for any $\epsilon > 0$,

$$P_0\left(\inf_{\{(\theta,\sigma^2)\in J, \left|\theta\sigma^2 - \tilde{\theta}\tilde{\sigma^2}\right| \geq \epsilon\}} \{\ell_{\text{n,tap}}(\tilde{\theta}, \tilde{\sigma^2}) - \ell_{\text{n,tap}}(\theta, \sigma^2)\} \longrightarrow \infty\right) = 1, \quad (3.3.66)$$

where $(\tilde{\theta}, \tilde{\sigma^2}) \in J$ can be any fixed vector such that $\tilde{\theta}\tilde{\sigma^2} = \theta_0\sigma_0^2$.

Ying (1991) has shown (3.3.66) for the log likelihood function $\ell_n(\theta, \sigma^2)$. More specifically, Ying (1991) showed that, uniformly in $(\theta, \sigma^2) \in J$ and $\left|\theta\sigma^2 - \tilde{\theta}\tilde{\sigma^2}\right| \geq \epsilon$, with probability 1,

$$\ell_n(\tilde{\theta}, \tilde{\sigma^2}) - ell_n(\theta, \sigma^2) \geq \eta n + O_u(n^{1/2+\alpha}) \text{ for any } \alpha > 0,$$

[see the proof of Theorem 1 in Ying (1991) p. 289]. Then, (3.3.66) follows because of (3.3.6) in Theorem 1.

Similarly, we can show (3.3.9) by using (3.3.7) and some asymptotic results in Ying (1991). We can write [see (3.10) and (3.11) in Ying (1991) p. 291]

$$\frac{\partial}{\partial\theta}\ell_n(\theta,\sigma^2) = \frac{\sigma_0^2\theta_0}{\sigma^2\theta^2}\sum_{k=2}^{n} W_{k,n}^2 - \frac{n}{\theta} + O_u(1),$$

where

$$W_{k,n} = \frac{X(t_k) - e^{-\theta_0\Delta_k}X(t_{k-1})}{\sigma_0\sqrt{1-e^{-2\theta_0\Delta_k}}}.$$

Note that $W_{k,n}$ depends only on the true parameters and are i.i.d. N(0, 1) for $k = 1,\ldots,n$.

Then, for any $(\theta,\sigma^2) \in J$, we have

$$\sigma^2\theta^2\frac{\partial}{\partial\theta}\ell_{\mathrm{n,tap}}(\theta,\sigma^2) = \sigma_0^2\theta_0\sum_{k=2}^{n}(W_{k,n}^2 - 1) - n(\sigma^2\theta - \sigma_0^2\theta_0) + o_u(n^{1/2})$$

by Theorem 1. In particular, for $(\theta,\sigma^2) = (\hat{\theta}_{n,tap},\hat{\sigma}_{n,tap}^2)$, the left hand side is zero. Therefore, we obtain

$$0 = \theta_0\sigma_0^2\sum_{k=2}^{n}(W_{k,n}^2 - 1) - n(\hat{\theta}_{\mathrm{n,tap}}\hat{\sigma}_{\mathrm{n,tap}}^2 - \theta_0\sigma_0^2) + o_u(n^{1/2}).$$

Since $W_{k,n}^2$, $k = 1,\ldots,n$, are i.i.d. χ_1^2, we have

$$\begin{aligned}\sqrt{n}(\hat{\theta}_{\mathrm{n,tap}}\hat{\sigma}_{\mathrm{n,tap}}^2 - \theta_0\sigma_0^2) &= \theta_0\sigma_0^2 n^{-1/2}\sum_{k=2}^{n}(W^{0}{}_{k,n}^2 - 1) + o_u(1)\\ &\xrightarrow{d} N(0, 2(\theta_0\sigma_0^2)^2).\end{aligned}$$

The proof is complete. □

3.3.6 Appendix 2. Proofs for Section 3.3.3

We will employ some known properties of equivalent Gaussian measures reviewed in Chapter 2 [e.g. (2.2.2)$\sim$ (2.2.5)] and will refer to Ibragimov and Rozanov (1978) frequently. Before we proceed with the proof of the main results in Section 3, we will establish the following lemmas.

Lemma 3.3.12. *Let $f_1(\lambda)$ be the spectral density corresponding to isotropic Matérn covariogram $K(h;\sigma_1^2,\theta_1)$ and $\tilde{f}_1(\lambda)$ be the spectral density corresponding to the tapered covariance function $\tilde{K}(h;\sigma_1^2,\theta_1) = K(h;\sigma_1^2,\theta_1)K_{tap}(h)$. Under condition (A3), there exists $r > 1$ such that*

$$\frac{\tilde{f}_1(\lambda) - f_1(\lambda)}{f_1(\lambda)} = O(|\lambda|^{-r}) \qquad as \quad |\lambda| \to \infty. \tag{3.3.67}$$

Proof. Using the fact that Fourier transform of product of two functions is the convolution of their Fourier transforms, we have

$$\tilde{f}_1(\lambda) = \int_{\mathbb{R}} f_1(x) f_{\text{tap}}(\lambda - x)dx, \tag{3.3.68}$$

where f_{tap} is the spectral density corresponding to K_{tap}. It is seen that $\tilde{f}_1(\lambda)/f_1(\lambda)$ does not depend on σ_1^2 so that we can assume without loss of generality that $\sigma_1^2 = 1$. It suffices to consider the case that $\lambda > 0$, because $\tilde{f}_1(\lambda)$ is symmetric about $\lambda = 0$. Using $\int_{\mathbb{R}} f_{\text{tap}}(\lambda - x)dx = 1$ and breaking down these integrals over intervals $(-\infty, \lambda - \lambda^k] \cup [\lambda + \lambda^k, +\infty)$ and $(\lambda - \lambda^k, \lambda + \lambda^k)$ for any $k \in (0,1)$, we have

$$\begin{aligned}\frac{\tilde{f}_1(\lambda)}{f_1(\lambda)} - 1 = \frac{\int_{|\lambda-x|\geq\lambda^k} f_1(x) f_{\text{tap}}(\lambda - x)\, dx}{f_1(\lambda)} - \int_{|\lambda-x|\geq\lambda^k} f_{\text{tap}}(\lambda - x)dx \\ + \frac{\int_{|\lambda-x|<\lambda^k} (f_1(x) - f_1(\lambda)) f_{\text{tap}}(\lambda - x)\, dx}{f_1(\lambda)} = T_1 + T_2 + T_3.\end{aligned}$$

By condition (A3), we have

$$|T_1| \leq \frac{M}{(1+\lambda^{2k})^{\nu+\frac{1}{2}+\epsilon}} \frac{1}{f_1(\lambda)} \int f_1(x)dx.$$

The Matérn spectral density has a closed form

$$f_1(\lambda) = \frac{c\sigma_1^2\theta_1^{2\nu}}{(\theta_1^2+|\lambda|^2)^{\nu+1/2}} \text{ for } c = \frac{\Gamma(\nu+1/2)}{\Gamma(\nu)\pi^{1/2}}. \tag{3.3.69}$$

Here, we set $\sigma_1^2 = 1$. In addition, $\int f_1(\lambda)d\lambda = \sigma_1^2 = 1$ is the variance. Then,

$$|T_1| \leq \frac{M(\theta_1^2+\lambda^2)^{\nu+\frac{1}{2}}}{c\theta_1^{2\nu}(1+\lambda^{2k})^{\nu+\frac{1}{2}+\epsilon}}. \tag{3.3.70}$$

Similarly,

$$|T_2| \leq \frac{M}{(\nu+\epsilon)\lambda^{2k(v+\epsilon)}}. \tag{3.3.71}$$

Since $\epsilon > 1/2$ and $v+\epsilon > 1$ by Condition (A3), we can choose k to be sufficiently close to 1, so that both T_1 and T_2 are $O(\lambda^{-r})$ for some $r > 1$.

To bound T_3, write, for some ξ between λ and x,

$$f_1(x) - f_1(\lambda) = f_1'(\lambda)(x-\lambda) + f_1''(\xi)\frac{(x-\lambda)^2}{2}.$$

Then,

$$\begin{aligned} T_3 =& \frac{1}{f_1(\lambda)}[f_1'(\lambda)\int_{|x-\lambda|<\lambda^k}(x-\lambda)f_{\text{tap}}(\lambda-x)\,dx \\ &+ \int_{|\lambda-x|<\lambda^k} f_1''(\xi)\frac{(x-\lambda)^2}{2}f_{\text{tap}}(\lambda-x)\,dx]. \end{aligned}$$

The first term is 0 because the integrand is odd. For the second term, note

$$0 < f_1''(\xi) = \frac{c\,\theta_1^{2\nu}(2\nu+1)}{(\theta_1^2+\xi^2)^{\nu+3/2}}\left(\frac{(2\nu+3)\xi^2}{\theta_1^2+\xi^2} - 1\right) \leq \frac{c\,\theta_1^{2\nu}(2\nu+3)^2}{(\theta_1^2+\xi^2)^{\nu+3/2}}.$$

Therefore, if λ is sufficiently large, for ξ lying between x and λ, where x is in the interval $|x-\lambda| < \lambda^k$,

$$0 < f_1''(\xi) \leq \frac{c\theta_1^{2\nu}(2\nu+3)^2}{(\theta_1^2+(\lambda-\lambda^k)^2)^{\nu+3/2}} \leq \frac{2c\theta_1^{2\nu}(2\nu+3)^2}{\lambda^{2\nu+3}}.$$

Then,

$$0 < T_3 < \frac{(2\nu+3)^2(\theta_1^2+\lambda^2)^{\nu+1/2}}{\lambda^{2\nu+3}} \int (x-\lambda)^2 f_{\text{tap}}(x-\lambda)dx.$$

Condition (A3) implies that $x^2 f_{\text{tap}}(x)$ is integrable. Then, $T_3 = O(\lambda^{-2})$. The proof is complete. □

Lemma 3.3.13. *For any real number $r > 0$, there exists $\xi_r(\lambda)$ such that*

$$\xi_r(\lambda) = \int c_r(t)\exp(-i\lambda t)\,dt, \quad 0 < |\xi_r(\lambda)|^2 \asymp |\lambda|^{-r}, \quad \lambda \to \infty, \tag{3.3.72}$$

where $c_r(t)$ is square integrable and has a compact support.

Proof. We only need to show the case $0 < r \leq 1$, because the product of any functions of the type given by (3.3.72) belongs to this type, due to the fact that the Fourier transform of convolution coincides with the product of Fourier transforms. Let

$$\xi_r(\lambda) = \int_{-1}^{1} e^{-i\lambda t}\,|t|^{r/2-1}\,dt = 2\int_0^1 \cos(\lambda t)t^{r/2-1}\,dt.$$

We will show that $\xi_r(\lambda)$ satisfies (3.3.72). We only need to prove it for $\lambda \geq 0$, because $\xi_r(\lambda)$ is symmetric about $\lambda = 0$ and $\xi_r(0) > 0$. Let $u = \lambda t$. We can write

$$\xi_r(\lambda) = 2\lambda^{-r/2}\int_0^{\lambda} \cos(u)u^{r/2-1}du.$$

Then, $\xi_r(\lambda)^2 \asymp |\lambda|^{-r}$ as $\lambda \to +\infty$, because $\cos(u)u^{r/2-1}$ is integrable for $0 < r \leq 1$.

Next, we will show $\xi_r(\lambda) > 0$ for any $\lambda > 0$. It suffices to show

$$y(\lambda) = \int_0^\lambda \cos(u)u^{-\delta}\,du > 0, \tag{3.3.73}$$

where $\delta = 1 - r/2 \in [1/2, 1)$. Note that $y'(\lambda) = \cos(\lambda)\lambda^{-\delta}$ and $y''(\lambda) = -\sin(\lambda)\lambda^{-\delta} - \delta\cos(\lambda)\lambda^{-\delta-1}$. Therefore, the minimum points are $\{2k\pi + 3\pi/2, k = 0, 1, \dots\}$. So, we only need to show $y(2k\pi + 3\pi/2) > 0$, $k = 0, 1, \dots$ by induction. First, using monotonicity of $\cos(u)$, we have

$$\begin{aligned} y(\frac{3\pi}{2}) &= \int_0^{\pi/4} \cos(u)u^{-\delta}du + \int_{\pi/4}^{\pi/2} \cos(u)u^{-\delta}du + \int_{\pi/2}^{\pi} \cos(u)u^{-\delta}du + \int_{\pi}^{3\pi/2} \cos(u)u^{-\delta}du. \\ &\geq \cos(\pi/4)\frac{(\pi/4)^{1-\delta}}{1-\delta} + \frac{1}{(\pi/2)^\delta}(1 - \frac{\sqrt{2}}{2}) - \frac{1}{(\pi/2)^\delta} - \frac{1}{\pi^\delta} \\ &\geq \frac{\sqrt{2}}{2}\sqrt{\pi} + \frac{2}{\pi}(1 - \frac{\sqrt{2}}{2}) - \sqrt{\frac{2}{\pi}} - \frac{1}{\sqrt{\pi}} > 0. \end{aligned} \tag{3.3.74}$$

Next, suppose $y(2(k-1)\pi + 3\pi/2) > 0$, for $k \geq 1$, then

$$\begin{aligned} &y(2k\pi + 3\pi/2) \\ &= y(2(k-1)\pi + 3\pi/2) + \int_{2k\pi-\pi/2}^{2k\pi+\pi/2} \cos(u)u^{-\delta}du + \int_{2k\pi+\pi/2}^{2k\pi+3\pi/2} \cos(u)u^{-\delta}du \\ &= y(2(k-1)\pi + 3\pi/2) + \int_{2k\pi-\pi/2}^{2k\pi+\pi/2} \cos(u)u^{-\delta}du - \int_{2k\pi-\pi/2}^{2k\pi+\pi/2} \cos(u)(u+\pi)^{-\delta}du \\ &= y(2(k-1)\pi + 3\pi/2) + \int_{2k\pi-\pi/2}^{2k\pi+\pi/2} \cos(u)\left(u^{-\delta} - (u+\pi)^{-\delta}\right)du. \end{aligned}$$

The integral is positive because the integrand is positive. This completes the proof of Lemma 3.3.13. □

Proof of Theorem 3.3.6. Write the Cholesky decomposition of $\mathbf{V}_{0,n} = \mathbf{L}\mathbf{L}'$ for some lower triangular matrix $\mathbf{L}$. Let $\mathbf{Q}$ be an orthogonal matrix such that

$$\mathbf{Q}\mathbf{L}^{-1}\mathbf{V}_{1,n}\mathbf{L}'^{-1}\mathbf{Q}' = \mathrm{diag}\{\sigma_{1,n}^2, \dots, \sigma_{n,n}^2\}.$$

Then,

$$\mathbf{Q}\mathbf{L}'\mathbf{V}_{1,n}^{-1}\mathbf{L}\mathbf{Q}' = \text{diag}\{1/\sigma_{1,n}^2, \ldots, 1/\sigma_{n,n}^2\}.$$

Taking the trace of both sides, we have

$$\text{trace}(\mathbf{V}_{0,n}\mathbf{V}_{1,n}^{-1}) = \sum_{i=1}^{n} 1/\sigma_{k,n}^2.$$

Hence,

$$\mathrm{E}_0(\mathbf{X}_n'(\mathbf{V}_{1,n}^{-1} - \mathbf{V}_{0,n}^{-1})\mathbf{X}_n) = \text{trace}(\mathbf{V}_{0,n}\mathbf{V}_{1,n}^{-1}) - n = \sum_{k=1}^{n}(\frac{1}{\sigma_{k,n}^2} - 1).$$

Let $\mathbf{e}_n = \mathbf{Q}\mathbf{L}^{-1}\mathbf{X}_n$. Obviously,

$$\mathrm{E}_0\, \mathbf{e}_n\mathbf{e}_n' = \mathbf{I}_n, \quad \mathrm{E}_1\, \mathbf{e}_n\mathbf{e}_n' = \text{diag}\{\sigma_{1,n}^2, \ldots, \sigma_{n,n}^2\}. \tag{3.3.75}$$

(3.3.20) follows if, for any orthogonal sequence $\{\eta_k, k = 1, 2, \ldots\}$ in the Hilbert space $L_D^2(d\,\mathrm{P}_0)$ spanned by $X(t), t \in D$ under the covariance inner product corresponding to P_0, there exists a constant $M > 0$ independent of $\eta_k, k = 1, 2, \ldots$, such that

$$\sum_{k=1}^{\infty} \left| \frac{1}{\mathrm{E}_1\, \eta_k^2} - 1 \right| < M. \tag{3.3.76}$$

One important technique to prove (3.3.76) is to write, for any $s, t \in D$,

$$E_1 X(t)X(s) - E_0 X(t)X(s) = \int_{\mathbb{R}} \int_{\mathbb{R}} e^{i(\lambda s - \mu t)} \Phi(\lambda, \mu)\, d\lambda\, d\mu, \tag{3.3.77}$$

where $\Phi(\lambda, \mu)$ is square integrable with respect to Lebesgue measure on $\mathbb{R}^2$. For any bounded region D, the existence of such a function Φ and, therefore, the equivalence of P_0 and P_1, are shown in Ibragimov and Rozanov (1978) p.104, Theorem 17, under the assumption that the function $h(\lambda)$ in (3.3.19) is square integrable. However, we will show, under the assumption of this lemma (3.3.19), which requires integrability

of $h(\lambda)$ being, that Φ takes a particular form

$$\Phi(\lambda,\mu) = \overline{\Phi_1(\lambda)}\Phi_2(\mu)\int_T e^{-i(\lambda-\mu)\omega}\,d\omega \tag{3.3.78}$$

for some functions $\Phi_j(\lambda), \lambda \in \mathbb{R}$ such that $\int |\Phi_j(\lambda)|^2/f_0(\lambda)\,d\lambda < \infty$, $j = 1,2$, and a compact interval T that is solely determined by r_1 and r_2. This particular form is central to the proof, and we will establish it at the end of this proof. We now proceed by assuming it is true.

Let $dZ_0(\lambda)$ denote the stochastic orthogonal measure so that $X(t)$ has the spectral representation under measure P_0; that is, $X(t) = \int \exp(-i\lambda t)dZ_0(\lambda)$. Then, for any $\eta \in L_D^2(d\mathrm{P}_0)$, there is a function $\phi(\lambda)$ such that $\eta = \int \phi(\lambda)dZ_0(\lambda)$ and $E_0\eta^2 = \int |\phi(\lambda)|^2 f_0(\lambda)\,d\lambda$. We first show

$$E_1\eta^2 = \int |\phi(\lambda)|^2 f_1(\lambda)\,d\lambda \tag{3.3.79}$$

$$E_1\eta^2 - E_0\eta^2 = \iint \overline{\phi(\lambda)}\phi(\mu)\Phi(\lambda,\mu)\,d\lambda\,d\mu. \tag{3.3.80}$$

Indeed, the two equations hold for $\eta = X(t) = \int \exp(-i\lambda t)dZ_0(\lambda)$ for any $t \in D$ [assuming (3.3.77) is true]. Consequently, they hold for any linear combination

$$\eta = \sum_{j=1}^{J} c_j X(t_j) = \int \phi(\lambda)dZ_0(\lambda),$$

for any J and $t_1,\ldots,t_J \in D$, where $\phi(\lambda) = \sum_{j=1}^{J} c_j e^{-i\lambda t_j}$.

For any $\eta \in L_D^2(d\mathrm{P}_0)$, we can find a sequence of finite linear combinations of $X(t), t \in D$, say, η_m, $m = 1,2,\ldots$, such that $\lim_{m\to\infty} E_0(\eta - \eta_m)^2 = 0$. If $\eta_m = \int \phi_m(\lambda)dZ_0(\lambda)$, we have

$$E_0(\eta - \eta_m)^2 = \int |\phi(\lambda) - \phi_m(\lambda)|^2 f_0(\lambda)\,d\lambda \to 0. \tag{3.3.81}$$

Then,

$$\int |\phi(\lambda) - \phi_m(\lambda)|^2 f_1(\lambda)\, d\lambda = \int |\phi(\lambda) - \phi_m(\lambda)|^2 f_0(\lambda)(1 + h(\lambda))\, d\lambda \to 0,$$

because $h = (f_1 - f_0)/f_0$ is bounded. It follows that η_m converges in $L^2(dP_1)$ norm to some variable $\tilde{\eta}$ because $E_1(\eta_\ell - \eta_m)^2 = \int |\phi_\ell(\lambda) - \phi_m(\lambda)|^2 f_1(\lambda)\, d\lambda \to 0$ as $\ell, m \to \infty$. Then,

$$E_1\tilde{\eta}^2 = \lim_{m\to\infty} E_1\eta_m^2 = \lim_{m\to\infty} \int |\phi_m(\lambda)|^2 f_1(\lambda)\, d\lambda = \int |\phi(\lambda)|^2 f_1(\lambda)\, d\lambda.$$

Since L_2 convergence implies convergence in probability, we have $\eta_m \to \tilde{\eta}$ in probability P_1. On the other hand, $\eta_m \to \eta$ in probability P_0 and, consequently, in probability P_1, due to the equivalence of the two probabilities. Then, we must have $\mathrm{P}_1(\eta = \tilde{\eta}) = 1$ and $E_1\eta^2 = E_1\tilde{\eta}^2$. We have proved (3.3.79). To show (3.3.80), note that

$$\begin{aligned}
&\left| \iint \overline{\phi_m(\lambda)}\phi_m(\mu)\Phi(\lambda,\mu)\, d\lambda\, d\mu - \iint \overline{\phi(\lambda)}\phi(\mu)\Phi(\lambda,\mu)\, d\lambda\, d\mu \right| \\
&\qquad \le \iint \left| (\overline{\phi_m(\lambda)} - \overline{\phi(\lambda)})\phi_m(\mu) \right| |\Phi(\lambda,\mu)|\, d\lambda\, d\mu \\
&\qquad\quad + \iint \left| (\phi_m(\mu) - \phi(\mu))\overline{\phi(\lambda)} \right| |\Phi(\lambda,\mu)|\, d\lambda\, d\mu,
\end{aligned} \tag{3.3.82}$$

where the first term tends to zero, because Cauchy-Schwarz inequality implies that its square is bounded by

$$|T|^2 \int \left| \overline{\phi_m(\lambda)} - \overline{\phi(\lambda)} \right|^2 f_0(\lambda)\, d\lambda \int |\phi_m(\mu)|^2 f_0(\mu)\, d\mu \int \frac{|\Phi_1(\lambda)|^2}{f_0(\lambda)}\, d\lambda \int \frac{|\Phi_2(\mu)|^2}{f_0(\mu)}\, d\mu$$
$$\to 0,$$

by (3.3.81) and square integrability of $\Phi_1(\lambda)/\sqrt{f_0(\lambda)}$, where and hereafter $|T|$ stands for the length of finite interval T. Similarly, we can show the second term in (3.3.82) also tends to zero. Therefore, (3.3.80) is now proved by taking the limit of $E_1\eta_m^2 - E_0\eta_m^2$ and $\iint \overline{\phi_m(\lambda)}\phi_m(\mu)\Phi(\lambda,\mu)\, d\lambda\, d\mu$.

Applying (3.3.80) to the orthonormal sequence $\eta_k = \int \phi_k(\lambda) dZ_0(\lambda)$, $k = 1, 2, \ldots$, we have

$$\begin{aligned} E_1\eta_k^2 - 1 &= \iint \overline{\phi_k(\lambda)}\phi_k(\mu) \int_T e^{-i(\lambda-\mu)\omega}\, d\omega \overline{\Phi_1(\lambda)}\Phi_2(\mu)\, d\lambda\, d\mu \\ &= \int_T \overline{A_{1,k}(\omega)} A_{2,k}(\omega)\, d\omega, \end{aligned}$$

where $A_{j,k}(\omega) = \int \phi_k(\lambda) \exp(i\lambda\omega)\Phi_j(\lambda)\, d\lambda$, j=1,2. Because (3.3.19) implies $P_0 \equiv P_1$, there exists constant $C > 0$ such that $E_1\eta_k^2 > C$ [see Ibragimov and Rozanov (1978) Page 104, Theorem 17 and page 76, (2.8)], and, therefore,

$$|1/E_1\eta_k^2 - 1| \le |E_1\eta_k^2 - 1|/C \le (1/2C) \sum_{j=1}^{2} \int_T |A_{j,k}(\omega)|^2\, d\omega.$$

In view that $A_{j,k}(\omega)$ is the inner product of the two integrable functions $\phi_k(\lambda) f_0(\lambda)^{1/2}$ and $\exp(i\lambda\omega)\Phi_j(\lambda)/f_0(\lambda)^{1/2}$ in $L^2(\, d\lambda)$, and that $\phi_k(\lambda) f_0(\lambda)^{1/2}$, $k = 1, 2, \ldots$, is an orthonormal sequence in $L^2(\, d\lambda)$ [because $E_0\eta_\ell\eta_k = \int \phi_\ell(\lambda)\phi_k(\lambda) f_0(\lambda)\, d\lambda$], we have, by Bessel's inequality,

$$\sum_{k=1}^{\infty} |A_{j,k}(\omega)|^2 \le \int |\Phi_j(\lambda)|^2/f_0(\lambda)\, d\lambda < \infty.$$

It follows that

$$\begin{aligned} \sum_{k=1}^{\infty} |E_1\eta_k^2 - 1| &\le (1/2C) \sum_{j=1}^{2} \int_T \sum_{k=1}^{\infty} |A_{j,k}(\omega)|^2\, d\omega \\ &\le (|T|/2C) \sum_{j=1}^{2} \int |\Phi_j(\lambda)|^2/f_0(\lambda)\, d\lambda < \infty. \end{aligned}$$

We just need to show (3.3.77) and (3.3.78) to complete the proof. We will employ the following well-known properties of Fourier transform. For any square integrable functions (with respect to Lebesgue measure) $\varphi_j(\boldsymbol{\lambda})$, $\boldsymbol{\lambda} \in \mathbb{R}^d$, there are square integrable

functions $a_j(\mathbf{t}), \mathbf{t} \in \mathbb{R}^d$ such that

$$\varphi_j(\boldsymbol{\lambda}) = \int_{\mathbb{R}^d} \exp(-i\boldsymbol{\lambda}'\mathbf{t}) a_j(\mathbf{t})\, d\mathbf{t}, \ \ j = 1, 2.$$

Furthermore,

$$\varphi_1(\boldsymbol{\lambda})\varphi_2(\boldsymbol{\lambda}) = \int_{\mathbb{R}^d} \exp(-i\boldsymbol{\lambda}'\mathbf{t})(a_1 * a_2)(\mathbf{t})\, d\mathbf{t} \tag{3.3.83}$$

$$\int_{\mathbb{R}^d} \exp(i\boldsymbol{\lambda}'\mathbf{t})\varphi_1(\boldsymbol{\lambda})\varphi_2(\boldsymbol{\lambda})\, d\boldsymbol{\lambda} = (2\pi)^d (a_1 * a_2)(\mathbf{t}), \tag{3.3.84}$$

where all the equalities are in the $L^2(\, d\boldsymbol{\lambda})$ sense, and $a_1 * a_2$ is the convolution; that is,

$$a_1 * a_2(\mathbf{t}) = \int_{\mathbb{R}^d} a_1(\mathbf{s}) a_2(\mathbf{t} - \mathbf{s})\, d\mathbf{s}.$$

By Lemma 3.3.13, there exists a continuous and square integrable function $\xi_j(\lambda)$ $(j = 1, 2)$ such that

$$\xi_j(\lambda) = \int c_j(t) \exp(-i\lambda t) dt, \quad 0 < |\xi_j(\lambda)|^2 \asymp |\lambda|^{-r_j}, \quad \lambda \to \infty, \tag{3.3.85}$$

for some square integrable function $c_j(t)$ that has a compact support [i.e., $c_j(t)$ is 0 outside a compact set].

Let $\xi(\lambda) = (f_0(\lambda) - f_1(\lambda))/|\xi_1(\lambda)|^2$. Then, $\xi(\lambda)$ is square integrable by the assumption of the theorem and the properties of $\xi_1(\lambda)$. Therefore, we can write, for some square integrable function $c(t)$, $\xi(\lambda) = \int \exp(-i\lambda t) c(t) dt$. Furthermore, for all s, t,

$$\begin{aligned} E_0 X(s)X(t) - E_1 X(s)X(t) &= \int e^{i\lambda(s-t)} (f_0(\lambda) - f_1(\lambda))\, d\lambda \\ &= \int e^{i\lambda(s-t)} \xi(\lambda) \, |\xi_1(\lambda)|^2 \, d\lambda, \end{aligned} \tag{3.3.86}$$

which we will denote by $b(s,t)$. By (3.3.83),

$$|\xi_1(\lambda)|^2 = \int \exp(-i\lambda t) \left(\int c_1(z)\overline{c_1(z-t)}dz \right) dt.$$

Applying (3.3.84) to $\xi(\lambda)$ and $|\xi_1(\lambda)|^2$, we get

$$\begin{aligned} b(s,t) =& 2\pi \int_{\mathbb{R}} c(w) \int_{\mathbb{R}} c_1(z)\overline{c_1(-(s-t-w-z))}dzdw \\ =& 2\pi \int_{\mathbb{R}^2} c(u-v)c_1(s-u)\overline{c_1(t-v)}dudv, \end{aligned} \tag{3.3.87}$$

which holds for all $s,t \in \mathbb{R}$. If we restrict s,t to the compact set D, the integral (3.3.87) is an integral over a compact set, say, $\Delta \times \Delta$. This is because c_1 is 0 outside a compact interval.

Next, we write $c(t)$ as a convolution of two functions. For this purpose, we write $\xi(\lambda) = \xi_2(\lambda)\xi_3(\lambda)$. Then, $\xi_3(\lambda)$ so defined is square integrable from assumptions and (3.3.85) and, therefore, can be written as

$$\xi_3(\lambda) = \int \exp(-i\lambda t)c_3(t)dt.$$

Then, $c = c_2 * c_3$ and, consequently,

$$c(u-v) = \int c_2(x)c_3(u-v-x)dx = \int c_2(u-\omega)c_3(\omega-v)\,d\omega. \tag{3.3.88}$$

Since we are only interested in $b(s,t)$ for $s,t \in D$ and, consequently, only interested in $c(u-v)$ for $u,v \in \Delta$, we will restrict both u,v to the interval Δ, so that the second interval in (3.3.88) is an integral on a finite interval, say, T, because c_2 has a compact support. Define the bivariate function

$$a(u,v) = \int_T c_2(u-\omega)c_3(\omega-v)\,d\omega, \quad u,v \in \mathbb{R},$$

which is square integrable because

$$|a(u,v)|^2 \le |T| \int_T |c_2(u-\omega)|^2 |c_3(\omega - v)|^2 \, d\omega$$

and both c_2 and c_3 are square integrable. In addition, for $u, v \in \Delta$, we have, from (3.3.88),

$$a(u,v) = c(u-v).$$

We therefore have shown that, for $s, t \in D$,

$$b(s,t) = 2\pi \int_{\mathbb{R}^2} a(u,v) c_1(s-u) \overline{c_1(t-v)} du dv.$$

Note that the integral is a convolution of functions of (u,v). Applying (3.3.84), we get

$$2\pi b(s,t) = \int \exp(i(\lambda s + \mu t)) \varphi_1(\lambda,\mu) \varphi_2(\lambda,\mu) \, d\lambda \, d\mu,$$

where

$$\varphi_1(\lambda,\mu) = \int_{\mathbb{R}^2} a(u,v) e^{-i(u\lambda + v\mu)} du dv, \quad \varphi_2(\lambda,\mu) = \int_{\mathbb{R}^2} c_1(u) \overline{c_1(v)} e^{-i(u\lambda+v\mu)} du dv.$$

Clearly,

$$\varphi_2(\lambda,\mu) = \xi_1(\lambda) \overline{\xi_1(-\mu)}.$$

Now,

$$\begin{aligned}
\varphi_1(\lambda,\mu) &= \int_{\mathbb{R}^2} a(u,v) e^{-i(u\lambda+v\mu)} du dv \\
&= \int_T \int_{\mathbb{R}^2} c_2(u-\omega) c_3(\omega - v) e^{-i(u\lambda+v\mu)} du dv \, d\omega \\
&= \int_T \int_{\mathbb{R}^2} c_2(x) c_3(-y) e^{-i((x+\omega)\lambda + (y+\omega)\mu)} dx dy d\omega \\
&= \int_T \left(\int_{\mathbb{R}^2} c_2(x) e^{-ix\lambda} c_3(-y) e^{-iy\mu} dx dy \right) e^{-i(\lambda+\mu)\omega} \, d\omega \\
&= \xi_2(\lambda) \xi_3(-\mu) \int_T e^{-i(\lambda+\mu)\omega} \, d\omega.
\end{aligned}$$

Hence,

$$b(s,t) = \frac{1}{2\pi}\int_{\mathbb{R}^2} e^{i(\lambda s+\mu t)}\xi_1(\lambda)\xi_2(\lambda)\overline{\xi_1(-\mu)}\xi_3(-\mu)\int_T e^{-i(\lambda+\mu)\omega}\,d\omega\,d\lambda\,d\mu$$
$$= \frac{1}{2\pi}\int_{\mathbb{R}^2} e^{i(\lambda s-\mu t)}\overline{\Phi_1(\lambda)}\Phi_2(\mu)\int_T e^{-i(\lambda-\mu)\omega}\,d\omega\,d\lambda\,d\mu$$

for $\Phi_1(\lambda) = \overline{\xi_1(\lambda)\xi_2(\lambda)}$, and $\Phi_2(\mu) = \overline{\xi_1(\mu)}\xi_3(\mu)$. Clearly, $\int |\Phi_j(\lambda)|^2/f_0(\lambda)\,d\lambda < \infty$ by the assumptions of this theorem, (3.3.85) and the square-integrability of ξ_2 and ξ_3. The proof is complete. □

Proof of Theorem 3.3.7. Let σ_1^2 be such that $\sigma_1^2\theta_1^{2\nu} = \sigma_0^2\theta_0^{2\nu}$, and let P_j be the probability measure under which the process has a Matérn covariogram with parameters (θ_j, σ_j^2) for $j = 0, 1$. Then, $\mathrm{P}_0 \equiv \mathrm{P}_1$ by Theorem 2 in Zhang (2004). Consequently, we only need to show that (3.3.21) and (3.3.22) hold, almost surely, with respect to P_1.

Let $f_1(\lambda) = f_1(\lambda; \theta_1, \sigma_1^2)$ be the spectral density under measure P_1 and $f_2(\lambda)$ the corresponding tapered spectral density as as $\tilde{f}_1(\lambda)$ defined in Lemma 3.3.12, from which we see that, for some constant $c > 0$,

$$\int_{|\lambda|>c}\left|\frac{f_2(\lambda) - f_1(\lambda)}{f_1(\lambda)}\right|^2 d\lambda < \infty,$$

which is a sufficient condition for the equivalence of the measures P_1 and P_2 where P_2 is the measure corresponding the tapered spectral density f_2.

Let $\mathbf{V}_{j,n}$, $j = 1, 2$, be the covariance matrix corresponding to the spectral density f_j that depends on σ_1^2 and θ_1 and does not depend on σ^2. For any σ^2, we have

$$\ell_{\mathrm{n,tap}}(\theta_1, \sigma^2) - \ell_n(\theta_1, \sigma^2) = -\log(\det \mathbf{V}_{2,n}/\det \mathbf{V}_{1,n}) - \frac{\sigma_1^2}{\sigma^2}\mathbf{X}_n'(\mathbf{V}_{2,n}^{-1} - \mathbf{V}_{1,n}^{-1})\mathbf{X}_n. \tag{3.3.89}$$

Split it into three additive terms as follows:

$$[-\log\frac{\det \mathbf{V}_{3,n}}{\det \mathbf{V}_{2,n}} - \mathrm{E}_1(\mathbf{X}_n'(\mathbf{V}_{3,n}^{-1} - \mathbf{V}_{2,n}^{-1})\mathbf{X}_n)] + (1 - \frac{\sigma_1^2}{\sigma^2})\,\mathrm{E}_1(\mathbf{X}_n'(\mathbf{V}_{3,n}^{-1} - \mathbf{V}_{2,n}^{-1})\mathbf{X}_n)$$
$$- \frac{\sigma_1^2}{\sigma^2}[\mathbf{X}_n'(\mathbf{V}_{3,n}^{-1} - \mathbf{V}_{2,n}^{-1})\mathbf{X}_n - \mathrm{E}_1(\mathbf{X}_n'(\mathbf{V}_{3,n}^{-1} - \mathbf{V}_{2,n}^{-1})\mathbf{X}_n)] = I_1 + I_2 - I_3.$$

Because $\mathrm{P}_1 \equiv \mathrm{P}_2$, the first term is bounded as we discussed previously in (2.2.4). Similarly, by (2.2.5), the third term I_3 is bounded uniformly in $\sigma^2 \in [w, v]$, almost surely. The second term I_2 is also bounded uniformly in $\sigma^2 \in [w, v]$ because, by Theorem 3.3.6,

$$\mathrm{E}_1(\mathbf{X}_n'(\mathbf{V}_{3,n}^{-1} - \mathbf{V}_{2,n}^{-1})\mathbf{X}_n) = O(1). \tag{3.3.90}$$

Therefore (3.3.21) is proved. To show (3.3.22), first observe

$$\frac{\partial}{\partial\sigma^2}\ell_{\mathrm{n,tap}}(\theta_1, \sigma^2) - \frac{\partial}{\partial\sigma^2}\ell_n(\theta_1, \sigma^2) = -\frac{\sigma_1^2}{\sigma^4}(\mathbf{X}_n'\mathbf{V}_{3,n}^{-1}\mathbf{X}_n - \mathbf{X}_n'\mathbf{V}_{2,n}^{-1}\mathbf{X}_n), \tag{3.3.91}$$

which can be rewritten as

$$\frac{\sigma_1^2}{\sigma^4}[\mathbf{X}_n'(\mathbf{V}_{3,n}^{-1} - \mathbf{V}_{2,n}^{-1})\mathbf{X}_n - \mathrm{E}_1(\mathbf{X}_n'(\mathbf{V}_{3,n}^{-1} - \mathbf{V}_{2,n}^{-1})\mathbf{X}_n)] - \frac{\sigma_1^2}{\sigma^4}\,\mathrm{E}_1(\mathbf{X}_n'(\mathbf{V}_{3,n}^{-1} - \mathbf{V}_{2,n}^{-1})\mathbf{X}_n). \tag{3.3.92}$$

Then, (3.3.22) immediately follows (2.2.5) and Theorem 3. □

Proof of Theorem 3.3.8. As $\sigma_1^2 = \sigma_0^2(\theta_0/\theta_1)^{2\nu}$, we only need to show

$$\sqrt{n}\left(\frac{\hat{\sigma}_n^2}{\sigma_1^2} - 1\right) \xrightarrow{d} N(0, 2), \tag{3.3.93}$$

$$\sqrt{n}(\frac{\hat{\sigma}_{\mathrm{n,tap}}^2}{\sigma_1^2} - 1) \xrightarrow{d} N(0, 2). \tag{3.3.94}$$

Let $\mathbf{V}_{j,n}$ be the covariance matrix of $\mathbf{X}_n$ corresponding to parameter values (θ_i, σ_i^2), $j = 0, 1$. Write $\mathbf{V}_{1,n} = \sigma_1^2\mathbf{R}_{1,n}$, where $\mathbf{R}_{1,n}$ is the related correlation matrix. First,

we note that $\hat{\sigma}_n^2$ has a closed form express

$$\hat{\sigma}_n^2 = \frac{1}{n}\mathbf{X}_n'\mathbf{R}_{1,n}^{-1}\mathbf{X}_n \tag{3.3.95}$$

that can be derived straightforwardly from the maximization. Then,

$$\begin{aligned}\sqrt{n}\left(\frac{\hat{\sigma}_n^2}{\sigma_1^2} - 1\right) &= \sqrt{n}\left(\frac{\mathbf{X}_n'\mathbf{R}_{1,n}^{-1}\mathbf{X}_n}{\sigma_1^2 n} - 1\right) \\ &= (1/\sqrt{n})(\mathbf{X}_n'(\mathbf{V}_{1,n}^{-1} - \mathbf{V}_{0,n}^{-1})\mathbf{X}_n) + \sqrt{n}\left(\frac{\mathbf{X}_n'\mathbf{V}_{0,n}^{-1}\mathbf{X}_n}{n} - 1\right). \end{aligned} \tag{3.3.96}$$

Since $\mathbf{V}_{0,n}^{-1/2}\mathbf{X}_n$ consists of i.i.d. $N(0,1)$ variables, $\mathbf{X}_n'\mathbf{V}_{0,n}^{-1}\mathbf{X}_n$ is the sum of i.i.d variables having a χ_1^2 distribution. The central limit theorem implies that the second term in (3.3.96) converges in distribution to $N(0,2)$.

(3.3.23) in Theorem 3.3.8 follows if the first term is shown to be bounded almost surely with respect to P_0. In view of (2.2.5), if suffices to show that

$$\mathrm{E}_0(\mathbf{X}_n'(\mathbf{V}_{1,n}^{-1} - \mathbf{V}_{0,n}^{-1})\mathbf{X}_n) = O(1).$$

To this end, we only need to verify that conditions of Theorem 3.3.6 are satisfied. The Matérn spectral density (3.3.69) satisfies, as $\lambda \to \infty$,

$$0 < f(\lambda; \theta_i, \sigma_i^2) \sim |\lambda|^{-(2\nu+1)}. \tag{3.3.97}$$

Moreover, in view of $\sigma_0^2\theta_0^{2\nu} = \sigma_1^2\theta_1^{2\nu}$,

$$h(\lambda) = \frac{f_1(\lambda)}{f_0(\lambda)} - 1 = \left(\frac{\theta_0^2 + \lambda^2}{\theta_1^2 + \lambda^2}\right)^{\nu+1/2} - 1 = \left(1 + \frac{\theta_0^2 - \theta_1^2}{\theta_1^2 + \lambda^2}\right)^{\nu+1/2} - 1, \tag{3.3.98}$$

where $f_i(\lambda)$ stands for $f(\lambda; \theta_i, \sigma_i^2)$, $i = 0, 1$. Using the Taylor expansion, we can get

$$h(\lambda) \sim |\lambda|^{-2}.$$

Hence, (3.3.23) is proved.

Next, we derive the asymptotic distribution of the tapered MLE $\hat{\sigma}^2_{\mathrm{n,tap}}$. Similar to $\hat{\sigma}^2_n$, the tapered MLE $\hat{\sigma}^2_{\mathrm{n,tap}}$ takes the closed form

$$\hat{\sigma}^2_{n,tap} = \frac{1}{n}\mathbf{X}'_n\tilde{\mathbf{R}}^{-1}_{1,n}\mathbf{X}_n,$$

where $\tilde{\mathbf{R}}_{1,n}$ is the tapered correlation matrix corresponding to $\mathbf{R}_{1,n}$. It follows from (2.2.5) 3.3.90 with $\sigma^2_1 = 1$ that

$$\mathbf{X}'_n\tilde{\mathbf{R}}^{-1}_{1,n}\mathbf{X}_n - \mathbf{X}'_n\mathbf{R}^{-1}_{1,n}\mathbf{X}_n = O(1). \ \ a.s.$$

Then, with probability 1

$$\hat{\sigma}^2_{n,tap} = \frac{1}{n}\mathbf{X}'_n\mathbf{R}^{-1}_{1,n}\mathbf{X}_n + O(1/n) = \hat{\sigma}^2_n + O(1/n).$$

It follows immediately that $\hat{\sigma}^2_{n,tap}$ and $\hat{\sigma}^2_n$ have the same asymptotic distribution. The proof is complete. □

3.4 Comparison between tapered and exact likelihood functions

3.4.1 Main results of comparison

In previous section, the derived asymptotic distribution of tapered MLE for general Matérn model is only a partial extension of the results in Exponential case, because one of the parameters θ is chosen to be fixed. However in practice, if you have no prior knowledge about either of the parameters, what is commonly done is to jointly maximize both parameters like we did in Exponential case. Actually the exponential model has suggested some insight into the generalization to the case when the

covariance function is not exponential. However, explicit fixed-domain asymptotic distributions of exact MLE with all parameters maximized simultaneously or with high dimensional index in the general case are not yet available. It would be a harder problem to establish the asymptotic distribution of the tapered MLE. In light of this, we will focus on the comparisons between the log likelihood function and the tapered log likelihood function, and between their derivatives. We will establish results similar to Theorem 3.3.3. Consequently, the tapered MLE and MLE share the same asymptotic distribution under some regularity conditions. Therefore tapering would not reduce the asymptotic efficiency.

Consider the stationary Gaussian process $X(t), t \in [0,1]$ with covariance function $K(|s-t|) = \sigma^2 \rho_\theta(|s-t|)$, $s,t \in [0,1]$ where σ^2 is the variance, and ρ_θ is the correlation function that depends on a parameter θ. For simplicity of argument, an equally spaced sampling scheme is used throughout the rest of this section, i.e. $X(t)$ is observed at n points $t_i = i/n, i = 1, \ldots, n$. As in the previous section, write

$$\mathrm{E}(X(t_i)|X(t_j), j \neq i) = -\sum_{j \neq i} b_{ij,n}(\theta) X(t_j),$$

$$d_{i,n}(\theta, \sigma^2) = \mathrm{Var}(X(t_i)|X(t_j), j \neq i),\ i = 1, \ldots, n,$$

where $b_{ii,n} = 1$. The notations $\mathbf{D}_n$, $\mathbf{B}_n$, $\mathbf{T}_n$ are also used for the general case, these quantities have the same meaning as before. For any sequence of integers $a_n < n/4$, let $\tilde{\mathbf{O}}_n(a_n)$ denote a matrix of which the elements are uniformly $O_u(1/n)$ in the middle $n - 2a_n$ rows and the elements in the first and last a_n rows are uniformly bounded. Denote by $\breve{\mathbf{O}}_n(a_n)$ the matrix whose (i,j)th element is uniformly $O_u(1)$ if $1 \leq i \leq a_n$ or $n - a_n < i \leq n$ or $|i-j| < a_n$, and is uniformly $O_u(1/n)$ otherwise. We make the following assumptions on the coefficients $b_{ij,n}(\theta)$ and on the prediction variances $d_{i,n}(\theta, \sigma^2)$.

(C1) As function of $\theta \in [a,b]$, $b_{ij,n}(\theta)$ is uniformly bounded in θ and in (i,j,n). For

some sequence of positive integers k_n such that $k_n = O((\log\log n)^{1/8})$,

$$\sum_{j:|j-i|>k_n} \left(|b_{ij,n}(\theta)| \bigvee \left| b'_{ij,n}(\theta) \right| \right) = O_u(\frac{k_n^3}{n^2}), \text{ for any } i, \text{ and} \tag{3.4.1}$$

$$\left| \sum_{j:|j-i|\le k_n} b_{ij,n}(\theta) \right| \bigvee \left| b'_{ij,n}(\theta) \right| = \begin{cases} O_u(k_n^3/n^2) & \text{if } k_n < i \le n - k_n \\ O_u(k_n^3/n) & \text{otherwise}, \end{cases} \tag{3.4.2}$$

where $x \bigvee y = \max\{x, y\}$.

(C2) $d_{i,n}(\theta,\sigma^2)^{-1}=O_u(n)$, $d_{i,n}(\theta_0,\sigma_0^2)/d_{i,n}(\theta,\sigma^2)=O_u(1)$ and $\partial d_{i,n}^{-1}(\theta,\sigma^2)/\partial\theta=O_u(n)$.

(C3) $\rho_\theta(h)$ is twice differentiable function in (θ, h) with $\rho_\theta(h)$, $\frac{\partial}{\partial\theta}\rho_\theta(h)$, $\frac{\partial^2}{\partial\theta^2}\rho_\theta(h)$ having bounded second derivatives in h and these bounds are independent of θ. Moreover, $\lim_{h\to 0+} \rho'_\theta(h) > 0$ for any θ.

Lemma 3.4.1. *Under conditions* (A2) *and* (C1)-(C3), *the following holds for some constant* $r \ge 0$.

(I) $\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n) = \mathbf{I_n} + (\log n)^r \tilde{\mathbf{O}}_n$, $\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta} = (\log n)^r \breve{\mathbf{O}}_n$.

(II) $\frac{\partial}{\partial\theta}(\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n)) = (\log n)^r \tilde{\mathbf{O}}_n$, $\frac{\partial}{\partial\theta}(\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta}) = (\log n)^r \breve{\mathbf{O}}_n$.

(III) $\det(\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n)) = O_u((\log n)^r)$, $(\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n))^{-1} = \mathbf{I_n} + (\log n)^r \tilde{\mathbf{O}}_n$.

where $\tilde{\mathbf{O}}_n$ *denotes* $\tilde{\mathbf{O}}_n(a_n)$ *and* $\breve{\mathbf{O}}_n$ *denotes* $\breve{\mathbf{O}}_n(a_n)$ *for some* $a_n = O(k_n)$.

This lemma is analogous to Lemma 3.3.10 and conditions (C1)-(C3) are satisfied in the exponential case.

Theorem 3.4.2. *Under the conditions in Lemma 3.4.1, (3.3.6) and (3.3.7) hold uniformly in* $(\theta, \sigma^2) \in J$.

Proof. The development of this theorem follows closely the approach used for Theorem 3.3.3. As in proof of Theorem 3.3.3, to show (3.3.6) we compare the correspond-

ing log determinants and quadratic forms of $\ell_{\mathrm{n,tap}}(\theta,\sigma^2)$ and $\ell_n(\theta,\sigma^2)$. Since the first equality in Lemma 3.4.1 (III) gives

$$\log[\det(\mathbf{V}_n \circ \mathbf{T}_n)] = \log[\det(\mathbf{V}_n)] + O_u(\log\log n),$$

along the same line comparing the quadratic forms in proof of Theorem 3.3.3, we only need to show (3.3.52), i.e. $\mathbf{X}_n'\mathbf{H}_n\mathbf{V}_n^{-1}\mathbf{X}_n = o_u(n^{1/2})$ *a.s.* to obtain (3.3.6). Where recall that $\mathbf{H}_n = (\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n))^{-1} - \mathbf{I}_n$, then $\mathbf{H}_n = (\log n)^r\tilde{\mathbf{O}}_n$ by (III) in Lemma 3.4.1.

Throughout the rest of the paper, r represents some positive constant that is independent of parameters and may differ from time to time, but all the values it stands for are bounded above. First, by definition it is straightforward to verify the analogues to (3.3.31) and (3.3.33) with $\tilde{\mathbf{O}}_n, \breve{\mathbf{O}}_n$ replaced by $(\log n)^r\tilde{\mathbf{O}}_n(a_n)$, $(\log n)^r\breve{\mathbf{O}}_n(a_n)$ respectively, still hold in this general case. This together with (3.4.1), (3.4.2) and the general expression (3.3.35) entails $\frac{\partial}{\partial\theta}S_n(\theta) = O_u(n^2(\log n)^r)$ almost surely, where $S_n(\theta) = \sigma^2\mathbf{X}_n'\mathbf{H}_n\mathbf{V}_n^{-1}\mathbf{X}_n$, and it is noted that at most $2k_n + 1$ elements on each column of $\mathbf{V}_n^{-1}$ are $O_u(n)$ and all others are $O_u(k_n^3/n)$. Consequently, (3.3.52) follows from Lemma 3 if we can show

$$\mathrm{E}(S_n) = O_u((\log n)^r) \quad \text{and} \quad \mathrm{E}(S_n - \mathrm{E}(S_n))^6 = O_u((\log n)^r). \tag{3.4.3}$$

Indeed, thanks to (3.3.53) and (3.3.54), (3.4.3) directly results from the analogue of (3.3.55), namely

$$\mathbf{V}_n^{-1}(\theta,\sigma^2)\mathbf{V}_n(\theta_0,\sigma_0^2) = O_u(1)\mathbf{I}_n + (\log n)^r\tilde{\mathbf{O}}_n. \tag{3.4.4}$$

which follows if (3.3.56) and

$$\mathbf{A}_2\mathbf{V}_{n,0} = (\log n)^r\tilde{\mathbf{O}}_n \quad \text{uniformly in } \theta. \tag{3.4.5}$$

hold based on the same proof to show (3.3.55) in previous subsection 3.3.5. Since (3.3.56) is a easy consequence of condition **b)**, we only need to show (3.4.5) henceforth. Consider the absolute values of the (i,j) th element of $\mathbf{A_2V}_{n,0}$, $k_n < i \leq n - k_n$, which under condition **a)** equals

$$\left|\sum_{k=1}^{n} d_i^{-1}(b_{ik,n}(\theta) - b_{ik,n}(\theta_0))\sigma_0^2\rho_{\theta_0}(\frac{|k-j|}{n})\right|$$

$$\leq d_i^{-1}\sigma_0^2 \sum_{k:|k-i|\leq k_n} |b_{ik,n}(\theta)-b_{ik,n}(\theta_0)| + d_i^{-1}\sigma_0^2 \sum_{k:|k-i|>k_n} |b_{ik,n}(\theta)-b_{ik,n}(\theta_0)| = O_u(\frac{(\log\log n)^{\frac{1}{4}}}{n}).$$

Because $d_i^{-1} = O_u(n)$ and $b_{ik,n}(\theta)-b_{ik,n}(\theta_0) = O_u((\log\log n)^{1/8}/n^2)$ for $k_n < i \leq n-k_n$ by (3.4.2) and Mean Value Theorem. Similarly, it also follows from (3.4.2) that the elements on the first and last k_n rows are $O_u((\log\log n)^{1/4})$, thus (3.4.5) and therefore (3.3.6) follows. By the same reasoning as before, to show (3.3.7), we compare the corresponding traces and quadratic forms in (3.3.62) and (3.3.63) in the same way as in proof of Theorem 3.3.3. Imitating the proof of (3.3.64) and (3.3.65) in section 3.3.5, one can show Lemma 3.4.1 along with (3.4.4) implies

$$\text{trace}\{(\mathbf{V}_n \circ \mathbf{T}_n)^{-1}(\frac{\partial \mathbf{V}_n}{\partial\theta} \circ \mathbf{T}_n)\} = \text{trace}\{\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta}\} + O_u((\log n)^r),$$

and (3.3.65) is still true in this general case. Furthermore, those two results give (3.3.7) immediately and the proof of Theorem 3.4.2 is completed. □

Next we can give a condition in terms of spectral density function for comparison between exact and tapered likelihood. Even though for two Matérn spectral densities, (3.3.19) cannot hold for any $r_2 > 2$. However it might be hold for one Matérn spectral density and another spectral density of some covariance structure. Suggested by this idea, we have the following theorem of closeness of tapered and untapered likelihood.

Theorem 3.4.3. *If the underlying process is stationary Gaussian $X(\mathbf{t}), \mathbf{t} \in \mathbb{R}^d, d \leq 3$ having a mean 0 and a Matérn covariance function. Let $f_1(\boldsymbol{\lambda})$ be the spectral density*

corresponding to isotropic Matérn covariogram $K(h;\sigma_1^2,\theta_1)$ *and* $\tilde{f}_1(\lambda)$ *be the spectral density corresponding to the tapered covariance function* $\tilde{K}(h;\sigma_1^2,\theta_1) = K(h;\sigma_1^2,\theta_1)\cdot K_{tap}(h)$. *Suppose there exists* $r > d$ *such that*

$$\frac{\tilde{f}_1(\lambda) - f_1(\lambda)}{f_1(\lambda)} = O(|\lambda|^{-r}) \qquad as \quad |\lambda| \to \infty. \tag{3.4.6}$$

And the sampling locations $\{t_1, t_2, \dots\}$ *are from a bounded set* $D \subset \mathbb{R}^d$. *Then, for any fixed* $\theta_1 > 0$, *with* P_0*-probability 1, uniformly in* $\sigma^2 \in [w, v]$,

$$\begin{aligned}
\ell_{n,tap}(\theta_1,\sigma^2) =& \ell_n(\theta_1,\sigma^2) + O_u(1),\\
\frac{\partial}{\partial\sigma^2}\ell_{n,tap}(\theta_1,\sigma^2) =& \frac{\partial}{\partial\sigma^2}\ell_n(\theta_1,\sigma^2) + O_u(1),
\end{aligned}$$

where P_0 *is the probability measure corresponding to the true parameter values* $\sigma_0^2, \theta_0, \nu$.

This theorem can be proved by following the similar proof to show Thoerem 3.3.7 and Theorem 3.3.6, if we note that (3.4.6) implies (3.3.20).

3.4.2 Discussion

In previous section, we investigated how the covariance tapering affects the asymptotic efficiency of maximum likelihood estimators. We started out with the Ornstein-Uhlenbeck process that has an exponential covariance function. For this particular model, the inverse of the covariance matrix is a banded matrix. We would expect that, for this model, it would be easy to establish the asymptotic properties of tapered MLE. It turns out that even in this simple case, it is quite involved to derive the asymptotic distribution of the tapered MLE.

We also considered the general case when the stochastic process on the real line has a covariance function that may not be exponential. We gave conditions on the coefficients of drop-one prediction under which the tapered MLE and true MLE will have the same asymptotic distribution. One of the condition requires that the coefficient

decays rapidly as the sampling site moves away from the prediction site. Similar conditions can be given for a spatial process on $\mathbb{R}^d$ for $d > 1$ and the results in Subsection 3.4.1 can all be extended to the high dimensional case.

We did not put more conditions on the taper in the general case than that in the exponential model. It is an interesting problem to establish Theorem 3.4.2 by weakening the conditions (C1) and (C2) but restricting the taper to a class that satisfies certain properties. Thoerem 3.4.3 suggests that when the spectral density of the taper has a lower tail than the spectral density of the covariance function of the underlying process, it is then possible to establish Theorem 3.4.2.

3.4.3 Appendix 3: Proof of Lemma in subsection 3.4.1

Proof of Lemma 3.4.1. Let n be large enough such that $[n\gamma] > 2k_n$. As in the proof of Lemma 3.3.10, we assume $\sigma^2 = 1$ without loss of generality. In addition, we assume all quantities with indices out of range $[1, n]$ are 0, say $b_{ij} = 0$ if $j < 1$ or $j > n$. Let $\varrho(t)$ and $g(t)$ be bounded even function on $\mathbb{R}$ that may depend on θ and have bounded second derivative on $[-1, 0) \cup (0, 1]$. Define $\mathbf{Q}$ and $\mathbf{G}$ to be the matrices whose (i, j)th element is $\varrho(t_i - t_j)$ and $g(t_i - t_j)$ respectively. We will show

$$\mathbf{V}_n^{-1}(\mathbf{G} \circ \mathbf{Q}) = \mathbf{L}_1 + \mathbf{L}_2 + (\log\log n)^{3/8}\tilde{\mathbf{O}}_n(0) \tag{3.4.7}$$

where $\mathbf{L}_k, k = 1, 2$ stands for a matrix whose (i,j)th element is denoted by $\tau_{ij,k}$ and

$$\tau_{ij,1} = d_i^{-1}\varrho(\frac{i-j}{n}) \sum_{\ell=i-k_n}^{i+k_n} b_{i\ell} g(\frac{\ell - j}{n}), \quad \tau_{ij,2} = 1_{\{|i-j|<k_n\}}\varrho'_+(\frac{|i-j|}{n})O_u((\log\log n)^{\frac{1}{4}})$$

where $\varrho'_+(t) = \lim_{h\to 0+}\varrho'(t+h)$. Then if we take $g(t) = \rho_\theta(|t|)$, $\varrho(t) = K_{\text{tap}}(|t|)$, $\mathbf{V}_n^{-1}(\mathbf{G}\circ\mathbf{Q}) = \mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n)$. In this case $\tau_{ij,1} = 1_{\{i=j\}}$, because $\tau_{ij,1}$ is $K_{\text{tap}}(|i-j|/n)$ multiplied by the (i, j)th element of $\mathbf{D}_n^{-1}\mathbf{B}_n\mathbf{V}_n = \mathbf{I}_n$ and $K_{tap}(0) = 1$. In addition, the assumption (A2) implies for $0 < |i - j| < k_n$, we have $\varrho'(|i - j|) = c\,|i - j| + o_u(k_n) =$

$O_u((\log\log n)^{1/8}/n)$ and $\varrho'_+(0) = \varrho'(0+) = 0$. Therefore $\tau_{ij,2} = O_u((\log\log n)^{3/8}/n)$ and from (3.4.7) it follows that

$$\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n) = \mathbf{I}_n + (\log\log n)^{3/8}\tilde{\mathbf{O}}_n(k_n). \tag{3.4.8}$$

Next, let $\mathbf{G} = \mathbf{J}_n$, where $\mathbf{J}_n$ is the matrix of all 1s, and $\varrho(t) = \dfrac{\partial}{\partial\theta}\rho_\theta(|t|)$, then $\mathbf{V}_n^{-1}(\mathbf{G} \circ \mathbf{Q}) = \mathbf{V}_n^{-1}\dfrac{\partial \mathbf{V}_n}{\partial\theta}$. By (3.4.2) and $d_i^{-1} = O_u(n)$, we have

$$\tau_{ij,1} = d_i^{-1}\varrho(\frac{i-j}{n})\sum_{\ell=i-k_n}^{i+k_n} b_{i\ell} = \begin{cases} O_u(\log\log n)^{3/8}/n), & \text{if } k_n < i \le n - k_n \\ O_u((\log\log n)^{3/8}), & \text{otherwise.} \end{cases} \tag{3.4.9}$$

which in conjunction with (3.4.7) implies

$$\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta} = (\log\log n)^{3/8}\check{\mathbf{O}}_n(k_n). \tag{3.4.10}$$

and (I) is obtained if we can complete the proof of (3.4.7). Actually (3.4.7) can be shown in the similar fashion to show (3.3.34) in the previous Lemma. The detailed proof is given in the following:

First similar to (3.3.37), we have when $|\ell - i| \le k_n$,

$$\varrho(\frac{|\ell-j|}{n}) = \varrho(\frac{|i-j|}{n}) + \varrho'_+(\frac{|i-j|}{n})(\frac{|\ell-j| - |i-j|}{n}) + O_u(\frac{(\log\log n)^{\frac{1}{4}}}{n^2}). \tag{3.4.11}$$

Let ξ_{ij} denote the (i,j)th element of $\mathbf{V}_n^{-1}(\mathbf{G}\circ\mathbf{Q})$. Since $\sum_{\ell:|\ell-i|>k_n}|b_{i\ell}| = O_u(k_n^3/n^2)$ and $\varrho(|\ell-j|/n) = \varrho((l-j)/n)$, it follows from (3.3.35) and (3.4.11) that

$$\xi_{ij} = d_i^{-1}\sum_{\ell=i-k_n}^{i+k_n} b_{i\ell}g(\frac{\ell-j}{n})\varrho(\frac{\ell-j}{n}) + O_u(\frac{k_n^3}{n}) = \tau_{ij,1} + \tilde{\tau}_{ij} + O_u(\frac{(\log\log n)^{3/8}}{n}). \tag{3.4.12}$$

where

$$\tilde{\tau}_{ij} = d_i^{-1}\varrho'_+(\frac{|i-j|}{n})\sum_{\ell=i-k_n}^{i+k_n} b_{i\ell}\frac{|\ell-j| - |i-j|}{n}g(\frac{\ell-j}{n}). \tag{3.4.13}$$

Consider the case that $|i-j| \geq k_n$, then when $i - k_n \leq \ell \leq i + k_n$, $\ell - j$ and $i - j$ have the same sign. So by Taylor Theorem,

$$g(\frac{\ell - j}{n}) = g(\frac{i-j}{n}) + g'(\frac{i-j}{n})(\frac{\ell - i}{n}) + O_u(\frac{(\log\log n)^{\frac{1}{4}}}{n^2}).$$

Plugging this into (3.4.13) gives

$$\tilde{\tau}_{ij} = d_i^{-1}\varrho'_+(\frac{|i-j|}{n})g(\frac{i-j}{n}) \sum_{\ell=i-k_n}^{i+k_n} b_{i\ell}\frac{\ell - i}{n} + O_u((\log\log n)^{\frac{3}{8}}/n^2). \tag{3.4.14}$$

We will show the following:

$$\sum_{\ell=i-k_n}^{i+k_n} b_{i\ell}\frac{\ell-i}{n} = \begin{cases} O_u((\log\log n)^{\frac{3}{8}}/n^2), & \text{if } k_n < i \leq n - k_n; \\ O_u((\log\log n)^{\frac{1}{4}}/n), & \text{otherwise.} \end{cases} \tag{3.4.15}$$

Before proving (3.4.15), we note that (3.4.15), (3.4.14) and $d_i^{-1} = O_u(n)$ yield, if $|i-j| \geq k_n$

$$\tilde{\tau}_{ij} = \begin{cases} O_u((\log\log n)^{\frac{3}{8}}/n), & \text{if } k_n < i \leq n - k_n; \\ O_u((\log\log n)^{1/4}), & \text{if } i \leq k_n \text{ or } i > n - k_n. \end{cases} \tag{3.4.16}$$

On the other hand, by condition (C1), we always have $\sum_{\ell=i-k_n}^{i+k_n} b_{i\ell}\frac{|\ell-j| - |i-j|}{n} g(\frac{\ell - j}{n})$ $= O_u((\log\log n)^{1/4}/n)$, observing that $||\ell - j| - |i-j|| \leq |\ell - i|$. So in view of (3.4.13), if $|i-j| < k_n$,

$$\tilde{\tau}_{ij} = \varrho'_+(\frac{|i-j|}{n})O_u((\log\log n)^{1/4}).$$

This together with (3.4.12) and (3.4.16) completes the proof of (3.4.7) if we can show (3.4.15) now.

We now set to prove (3.4.15). Let us only consider the case when $k_n < i \leq n - k_n$

because the other case is obvious under condition (C1). Note that when $k_n < i \leq n - k_n$,

$$\sum_{\ell=i-k_n}^{i+k_n} b_{i\ell}\frac{\ell-i}{n} = \sum_{k=1}^{k_n}(b_{i,i+k}-b_{i,i-k})\frac{k}{n}.$$

It suffices to show

$$\sum_{k=1}^{k_n}(b_{i,i+k}-b_{i,i-k})\frac{k}{n} = O_u((\log\log n)^{\frac{3}{8}}/n^2) \tag{3.4.17}$$

Because $\mathbf{D}_n^{-1}\mathbf{B}_n\mathbf{V}_n = \mathbf{I}_n$ and σ^2 is set to be 1,

$$\sum_{\ell=1}^{n} b_{i\ell}\rho_\theta(|t_\ell - t_j|) = 0 \quad \text{if } i \neq j. \tag{3.4.18}$$

Then we have $\sum_{\ell=1}^{n} b_{il}\rho_\theta(|t_\ell - t_{i\pm k_n}|) = 0$, which implies

$$\begin{cases} \sum\limits_{1\leq k\leq k_n}(b_{i,i-k}\rho_\theta(t_{i-k}-t_{i-k_n})+b_{i,i+k}\rho_\theta(t_{i+k}-t_{i-k_n})) = -\rho_\theta(t_i-t_{i-k_n})+O_u(k_n^3/n^2), \\ \sum\limits_{1\leq k\leq k_n}(b_{i,i-k}\rho_\theta(t_{i+k_n}-t_{i-k})+b_{i,i+k}\rho_\theta(t_{i+k_n}-t_{i+k})) = -\rho_\theta(t_{i+k_n}-t_i)+O_u(k_n^3/n^2). \end{cases} \tag{3.4.19}$$

where the remainders are $O_u(k_n^3/n^2)$ because they are bounded by $\sum_{j:|i-j|>k_n}|b_{ij}|$. Since $t_i = i/n$, $t_{i\pm k} - t_{i-k_n} = t_{i+k_n} - t_{i\mp k} = (k_n \pm k)/n, k = 1, \ldots, k_n$. Taking the difference of the equalities in (3.4.19), we get

$$\sum_{1\leq k\leq k_n}\left(\rho_\theta(\frac{k_n+k}{n})-\rho_\theta(\frac{k_n-k}{n})\right)(b_{i,i+k}-b_{i,i-k}) = O_u(k_n^3/n^2). \tag{3.4.20}$$

By condition (C3) and Taylor Theorem,

$$\rho_\theta(\frac{k_n+k}{n})-\rho_\theta(\frac{k_n-k}{n}) = \rho'_\theta(\frac{k_n-k}{n})\frac{2k}{n} + O_u(\frac{k^2}{n}) = \rho'_+(0)\frac{2k}{n} + O_u(\frac{k_n^2}{n}), \tag{3.4.21}$$

where $\rho'_+(0) = \lim_{t\to 0+}\rho'_\theta(t)$ and we suppressed θ. The last equality follows from Mean Value Theorem for $\rho'_\theta(h)$ at 0. Combining (3.4.21) with (3.4.20) gives

$$\sum_{1\le k\le k_n} \rho'_+(0)\frac{2k}{n}(b_{i,i+k}-b_{i,i-k}) = O_u(k_n^3/n^2), \tag{3.4.22}$$

and (3.4.17) follows by noting that $\rho'_+(0) > 0$ for every $\theta \in [a,b]$ and therefore has a uniform positive bound from below. Hence, (3.4.7) is established and proof (I) is completed.

In what follows we will consider $\frac{\partial}{\partial\theta}(\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n))$ and $\frac{\partial}{\partial\theta}(\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta})$ in the similar way as in the proof of Lemma 3.3.10. First, consider $\frac{\partial}{\partial\theta}(\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n))$, from (3.3.60) it can be written as

$$-\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta}\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n) + \mathbf{V}_n^{-1}\left(\frac{\partial \mathbf{V}_n}{\partial\theta} \circ \mathbf{T}_n\right). \tag{3.4.23}$$

This together with (3.4.8) imply that $\frac{\partial}{\partial\theta}(\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n))$ equals

$$-\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta} - \mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta}(\log\log n)^{3/8}\tilde{\mathbf{O}}_n(k_n) + \mathbf{V}_n^{-1}\left(\frac{\partial \mathbf{V}_n}{\partial\theta} \circ \mathbf{T}_n\right), \tag{3.4.24}$$

where the second summand is $(\log\log n)^{7/8}\tilde{\mathbf{O}}_n(2k_n)$ because of (3.4.10) and the fact that $\check{\mathbf{O}}_n(k_n)\tilde{\mathbf{O}}_n(k_n) = (\log\log n)^{1/8}\tilde{\mathbf{O}}_n(2k_n)$. The last summand can be rewritten as

$$\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta} + \mathbf{V}_n^{-1}\left(\frac{\partial \mathbf{V}_n}{\partial\theta} \circ (\mathbf{T}_n - \mathbf{J}_n)\right). \tag{3.4.25}$$

From (3.4.24) and (3.4.25), we see that

$$\frac{\partial}{\partial\theta}(\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n)) = (\log\log n)^{7/8}\tilde{\mathbf{O}}_n(2k_n), \tag{3.4.26}$$

if we can show

$$\mathbf{V}_n^{-1}\left(\frac{\partial \mathbf{V}_n}{\partial\theta} \circ (\mathbf{T}_n - \mathbf{J}_n)\right) = (\log\log n)^{1/2}\tilde{\mathbf{O}}_n(k_n). \tag{3.4.27}$$

Recall (3.4.7) with $g(t) = \frac{\partial}{\partial\theta}\rho_\theta(|t|)$ and $\varrho(t) = K_{\text{tap}}(|t|) - 1$ so that $\mathbf{V}_n^{-1}(\mathbf{G} \circ \mathbf{Q}) = \mathbf{V}_n^{-1}\left(\frac{\partial \mathbf{V}_n}{\partial\theta} \circ (\mathbf{T}_n - \mathbf{J}_n)\right)$. Note that for $i \neq j$,

$$\sum_{\ell=i-k_n}^{i+k_n} b_{i\ell}\frac{\partial}{\partial\theta}\rho_\theta(\frac{|\ell - j|}{n}) = \begin{cases} O_u((\log\log n)^{1/2}/n^2), & \text{if } k_n < i \leq n - k_n \\ O_u((\log\log n)^{1/2}/n), & \text{otherwise.} \end{cases} \tag{3.4.28}$$

which follows from taking derivative of $\sum_{\ell=i-k_n}^{i+k_n} b_{i\ell}(\theta)\rho_\theta(|\ell - j|/n)$ in θ, because (3.4.2) entails

$$\sum_{\ell=i-k_n}^{i+k_n} b'_{il}(\theta)\rho_\theta(\frac{|\ell - j|}{n}) = \begin{cases} O_u((\log\log n)^{1/2}/n^2), & \text{if } k_n < i \leq n - k_n \\ O_u((\log\log n)^{1/2}/n), & \text{otherwise.} \end{cases}$$

and (3.4.18) implies $\frac{\partial}{\partial\theta}\sum_{\ell=i-k_n}^{i+k_n} b_{i\ell}(\theta)\rho_\theta(t_\ell - t_j) = O_u((\log\log)^{3/8}/n^2)$, if $i \neq j$, under condition (3.4.1). Therefore (3.4.28) associated with $\varrho(0) = 0$ gives

$$\mathbf{L}_1 = (\log\log n)^{1/2}\tilde{\mathbf{O}}_n(k_n).$$

On the other hand, note that $\varrho'(t) = \frac{d}{dt}K_{\text{tap}}(|t|)$, then for $|i - j| < k_n$, $\varrho'_+(|i - j|) = O((\log\log n)^{1/8}/n)$ and therefore $\tau_{ij,2}$=$O_u((\log\log n)^{3/8}/n)$ by the same corresponding reasoning in proof of (3.4.8). Consequently, (3.4.27) follows from (3.4.7) and (3.4.26) is proved.

In order to investigate $\frac{\partial}{\partial\theta}(\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta})$, we employ the expression (3.3.45). Note that the first term on the RHS of (3.3.45) is $(\log\log n)^{7/8}\breve{\mathbf{O}}_n(2k_n)$ because of (3.4.10) and $\breve{\mathbf{O}}_n(k_n)\breve{\mathbf{O}}_n(k_n) = (\log\log n)^{1/8}\breve{\mathbf{O}}_n(2k_n)$. The second term

$$\mathbf{V}_n^{-1}\frac{\partial^2 \mathbf{V}_n}{\partial\theta^2} = \mathbf{V}_n^{-1}(\mathbf{G} \circ \mathbf{Q}),$$

with $\mathbf{G} = \mathbf{J}_n$ and $\mathbf{Q} = \frac{\partial^2 \mathbf{V}_n}{\partial\theta^2}$. Then in this case $\varrho(t) = \frac{\partial^2}{\partial\theta^2}\rho_\theta(|t|)$ which is uniformly

bounded, so (3.4.9) holds and from (3.4.7) it follows that

$$\mathbf{V}_n^{-1}\frac{\partial^2 \mathbf{V}_n}{\partial\theta^2} = (\log\log n)^{3/8}\breve{\mathbf{O}}_n(k_n).$$

Combining these two terms gives

$$\frac{\partial}{\partial\theta}(\mathbf{V}_n^{-1}\frac{\partial \mathbf{V}_n}{\partial\theta}) = (\log\log n)^{7/8}\breve{\mathbf{O}}_n(2k_n).$$

Now, it remains to show (III) to finish the proof of this lemma. Actually, in view of the element of $\mathbf{A}_n = \mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n)$ specified in (3.4.8) and (3.4.7), the determinant and inverse of $\mathbf{A}_n$ can be bounded by closely following the approach dealing with **(iii)** in Lemma 3.3.10, so a brief proof will be given in the following. First, we can turn $\mathbf{A}_n$ into $\mathbf{I}_n + (\log\log n)^{1/2}\tilde{\mathbf{O}}(0)$ through a sequence of column operations. Then it follows from Hadamard inequality and invariance under column operations that there exists constant $\tilde{r}$ such that

$$\det(\mathbf{V}_n^{-1}(\mathbf{V}_n \circ \mathbf{T}_n)) \leq (1 + \frac{2\tilde{r}\log\log n}{n})^{n/2},$$

which is $O_u((\log n)^{\tilde{r}})$. Similarly, we can show all the $n-1 \times n-1$ cofactors are also $O_u((\log n)^{\tilde{r}})$. Moreover, $\det(\mathbf{A}_n) > 1$ still holds in this general case by the same reason as in the proof of Lemma 3.3.10, therefore the proof of (III) with $r = 2\tilde{r}$ is completed by using Laplacian expansion to calculate inverse of matrix in the conventional way and the whole lemma is proved by taking $a_n = k_n$ in (I), (III) and $a_n = 2k_n$ in (II). □

3.5 Simulation and model fitting of climate data

3.5.1 Simulation study

We are aware that most of the theoretical results in this chapter is about process with one-dimensional index. So we conducted the simulation to assess the validity of the asymptotic results obtained for multidimensional case and study the finite sample performance of tapered MLE.

Because both Zhang(2004) and Kaufman(2008) have pointed out in their work that even the general results apply only to estimators with fixed θ, but joint estimation of σ^2 and θ is what we usually do in practice. Kaufman(2005) did simulation to show that the joint maximizer $\hat{\sigma}^2_{\text{n,tap}}\hat{\theta}^{2\nu}_{\text{n,tap}}$ performs better than $\hat{\sigma}^2_{\text{n,tap}}\hat{\theta}^{2\nu}_{1}$ unless the chosen θ_1 happens to be the true value or very close by. Therefore, we adopted joint maximization using 1000 independent realizations from Gaussian process with mean zero and Matérn covariogram (3.3.14) with parameter values $\nu = 0.8$, $\sigma = 1$ and $\theta = 5$ at each of increasing sets of locations within unit square. Wendland 1 tapering function $K_1(h;\gamma) = (1-\frac{h}{\gamma})^4_+(1+4\frac{h}{\gamma})$, $\gamma > 0$ will be used according to condition (A3). The Figure 3.1 shows all the original, tapering and tapered covariance functions.

For the first part of the simulation, we calculated MLE, tapered MLE, and the approximated 95% confidence interval of $\eta = \sigma^2\theta^{2\nu}$ ($= 13.13$), which is $(\hat{\sigma}^2\hat{\theta}^{2\nu} \pm 1.96 * \sqrt{2}\hat{\sigma}^2\hat{\theta}^{2\nu}/\sqrt{n})$ based on asymptotic results (3.3.24) with $\gamma = 0.6, 0.3, 0.2, 0.1$ respectively. To obtain those irregularly spaced points, we first generated grid points with increment 0.1 on each side in unit square, then add some more closely spaced points within $[0, 0.2]^2$ with increment 0.03, then some more with increment 0.01, 0.02 and so on. Because it has been demonstrated in Stein [(1999b), p.223] that having some more closely spaced sample points will dramatically improve the estimation of variogram. Following this idea, we generated five set of sample points with increasing sample sizes $169, 298, 385, 751, 1087$. Figure 3.2 shows the sample location with sample size 169.

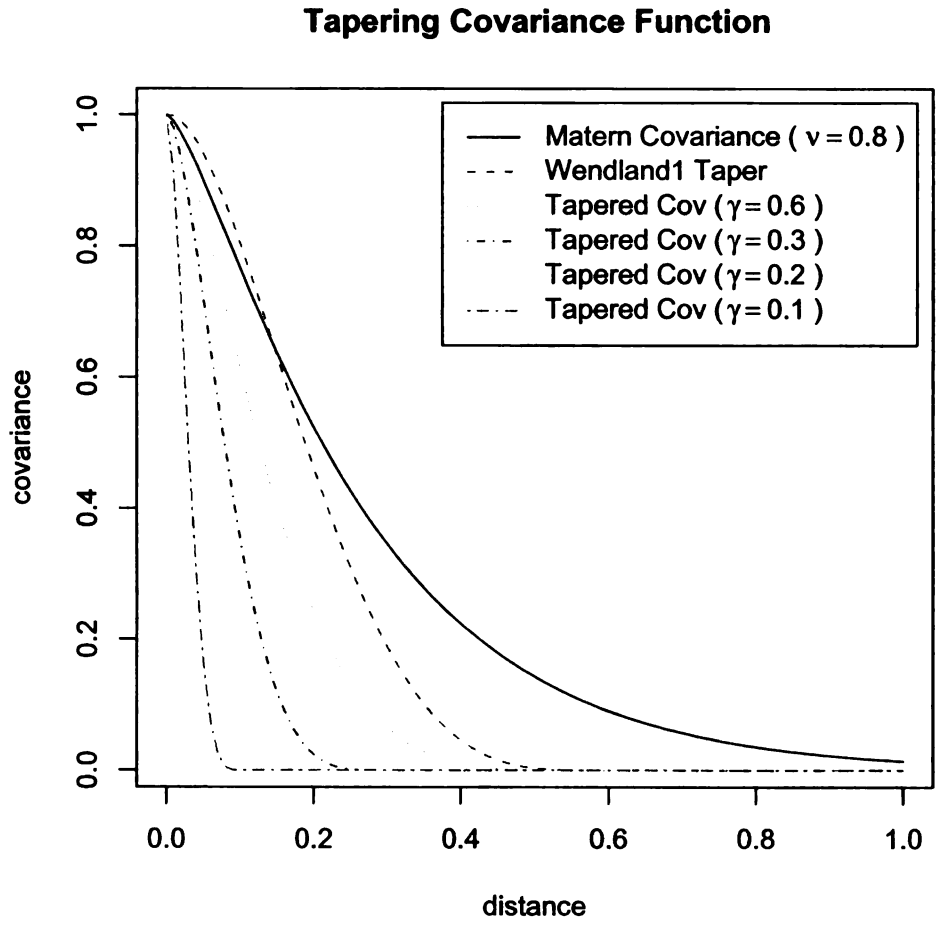

Figure 3.1: Matérn and tapered covariograms.

Boxplot of MLE, tapered MLE with $\gamma = 0.6, 0.3$ for θ, σ^2 and $\sigma^2\theta^{2\nu}$ are shown in Figures 3.3, 3.5 and 3.7 with increasing numbers of sample size. The corresponding histograms are depicted in Figures 3.4, 3.6 and 3.8. In each figure, the plot of exact MLE serves as baseline. From these boxplots and histograms we can see the estimates of θ and σ^2, untapered or tapered, are skewed and quite biased no matter how large the sample size is, but the estimates of $\sigma^2\theta^{2\nu}$ are more and more centered at true value and have less and less variability with increasing sample size. It again suggests that neither σ^2 nor θ is consistently estimable, but the product $\sigma^2\theta^{2\nu}$ is. Moreover, the distributions for both MLE and tapered MLE are more and more normal with

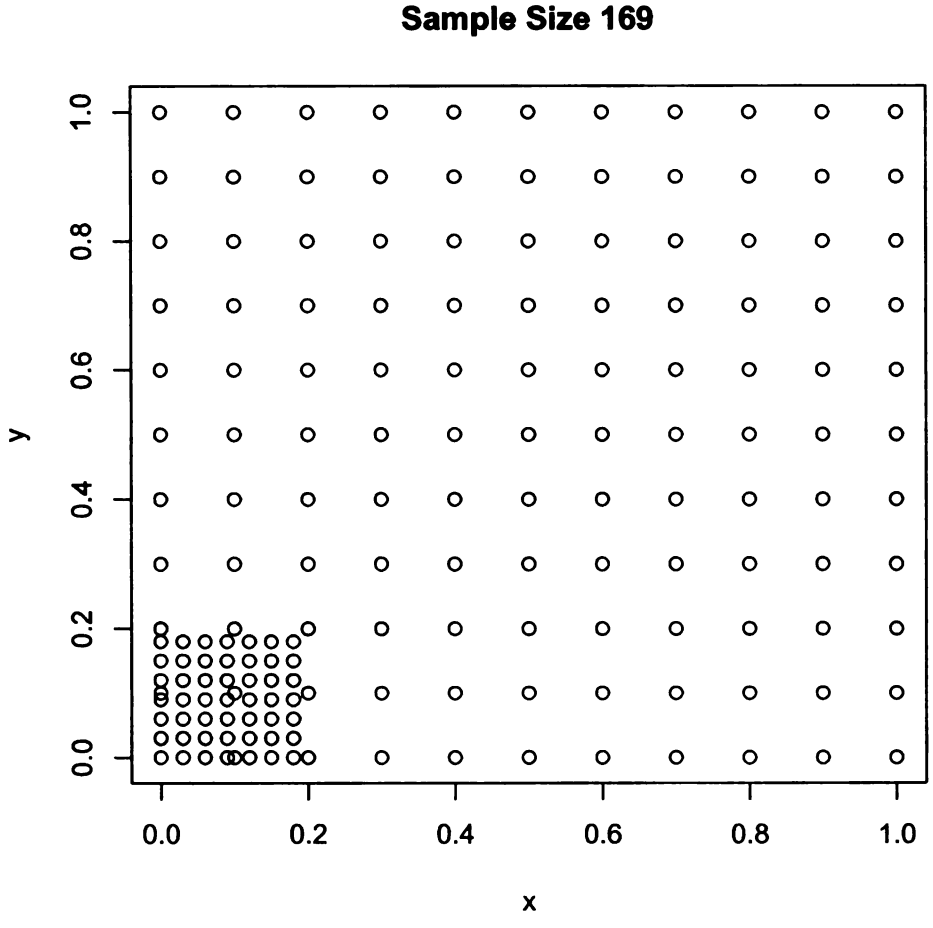

Figure 3.2: Sample data set locations of size 196.

decreasing variance, this agrees with the asymptotic normality result given by Theorem 3.3.8. The distribution for tapered MLE with $\gamma = 0.6$ looks more similar to the one given by exact MLE than tapered MLE with $\gamma = 0.3$ but it takes more time to compute though. It is not surprising, since larger tapering range will keep more information and therefore gives more accurate results, however the tapered covariance matrix will be less sparse and therefore more computational time is needed.

Actually, our simulation study shows that as long as the number of observations keeps increasing, the distribution of tapered MLE of $\sigma^2\theta^{2\nu}$ will be more and more symmetric around true value and normal with decreasing variance even when the degree of tapering is more severe with $\gamma = 0.2$. As showed in Figure 3.9. When n=751, the distributions for $\hat{\sigma}^2\hat{\theta}^{2\nu}$ with different tapering ranges 0.3 or 0.2 are roughly the same.

This table 3.1 lists average standard deviation, the relative coverage frequency (rcf) and the average length of the intervals (al) of 95% confidence interval constructed using the result in Theorem 3.3.8. Even though rcf's by using tapering technique are lower than the nominal level 95% for sample size 169 and confidence intervals are inflated more than exact MLE, they are dramatically improved and tend to perform similar to the MLE as sample size increases. When sample size reaches 385, all the relative coverage frequencies are quite close to the nominal level 95% and interval lengths by tapering are close to the MLE based one. These findings support our conjecture that the convergence results in Theorem 3.3.4 and Theorem 3.3.8 should still hold for multidimensional case under certain conditions.

We have seen that the distribution of tapered MLE is more and more comparable with that of exact MLE with increasing sample size albeit the tapering rage is moderate small. This comparable finite sample behavior of tapered MLE as opposed to that of exact MLE is also shown by using even more severe degree of tapering with taper range 0.1 given that the sample size is large enough, see the case when $n = 1087$ as suggested in the Figure 3.10. This justifies our theoretical result in Theorem 3.3.8,

which gives asymptotic normality of the tapered estimate under the condition about the tail behavior of tapering function, but has no constrains on the tapering range. That is, theoretically speaking, the choice of tapering range has no impact on the asymptotic efficiency.

For the second part of simulation, we recorded the estimates and timing for samples with sample size ranging from 1000 to 8000 to illustrate the comparable estimation as well as computational gain. In this part, the way to locate the sample points is to generate grid points in unit square with increment 0.005 and those with increment 0.0025 within $[0, 0.2]^2$, then randomly choose ranging from 1000 to 8000 points out of them as sample locations.

Aside from hardly reducing the efficiency of the estimation the computational efficiency is also achieved. See Figure 3.11, the red one depicts the time needed to derive the traditional MLE and blue one for tapered MLE, the red one goes up much faster than the blue one. So when the sample size gets larger, the time saving is more and more appealing. But the estimates almost stay the same when sample size is more than 2000. We can see from Table 3.2 when the sample size is 7000, it takes almost 9 times longer to get the exact MLE. (7 m vs. 1 h 6 m using department server with CPU 2.8 Ghz and 8.00 GB RAM), but the difference of the estimates is only 0.02.

All the computations in this section were conducted using open-source software R, with special courtesy to "Spam" created by Furrer.

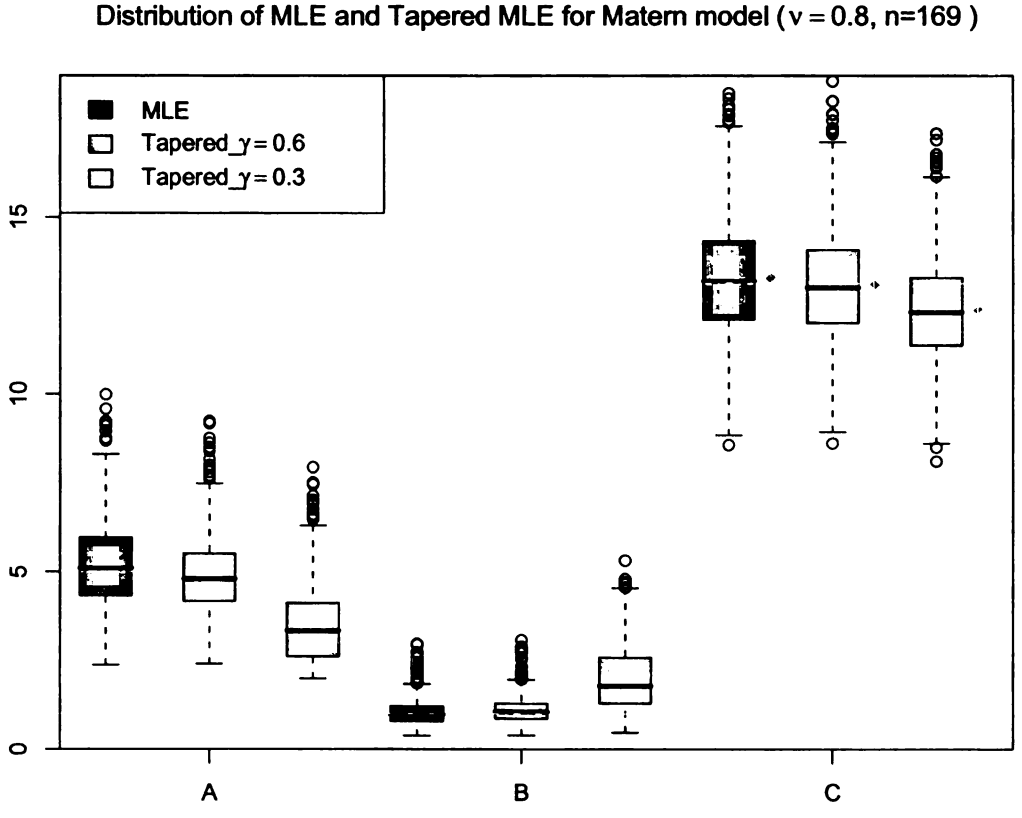

Figure 3.3: The boxplot of MLE and tapered MLE (where A= $\hat{\theta}$, B= $\hat{\sigma}^2$, C= $\hat{\sigma}^2\hat{\theta}^{2\nu}$) with sample size 196.

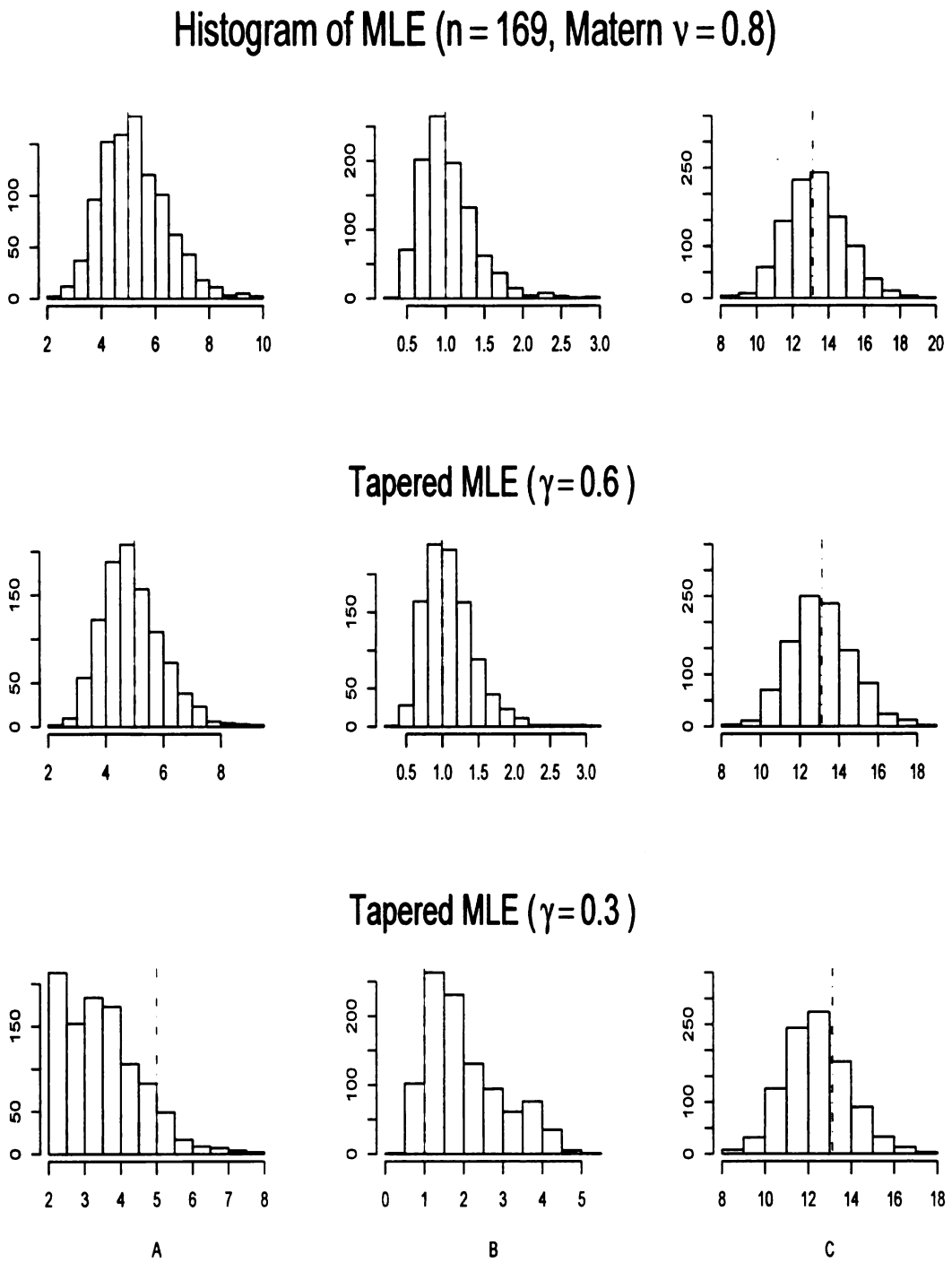

Figure 3.4: The histogram of MLE and tapered MLE (where A= $\hat{\theta}$, B= $\hat{\sigma}^2$, C= $\hat{\sigma}^2\hat{\theta}^{2\nu}$) with sample size 196.

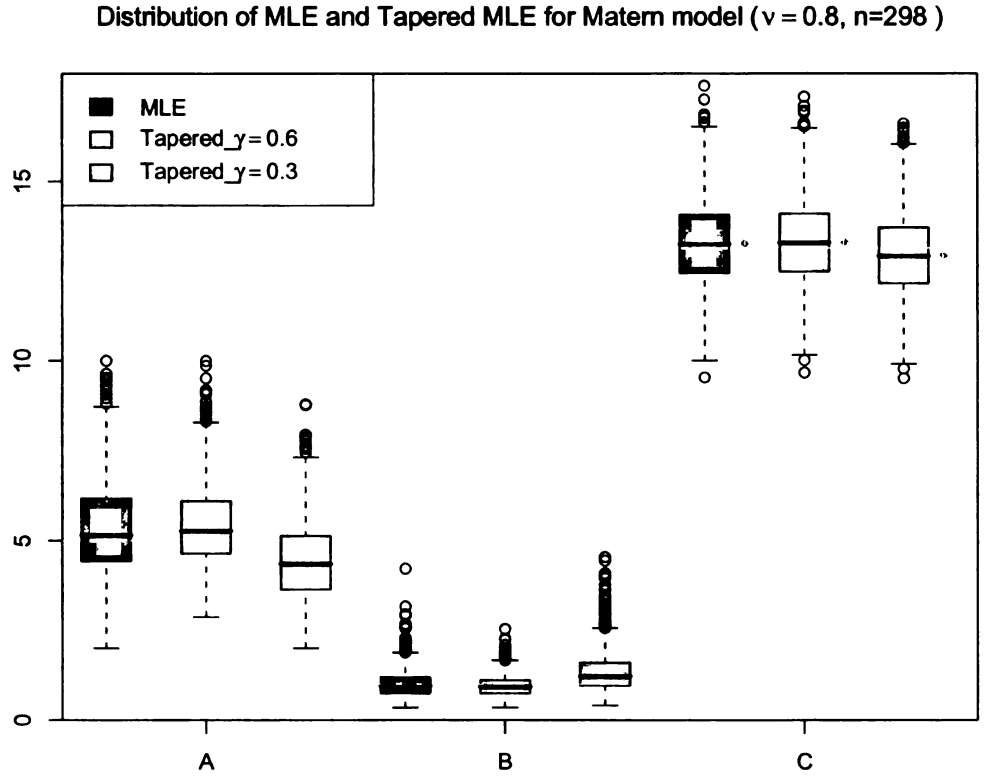

Figure 3.5: The boxplot of MLE and tapered MLE (where A= $\hat{\theta}$, B= $\hat{\sigma}^2$, C= $\hat{\sigma}^2\hat{\theta}^{2\nu}$) with sample size 298.

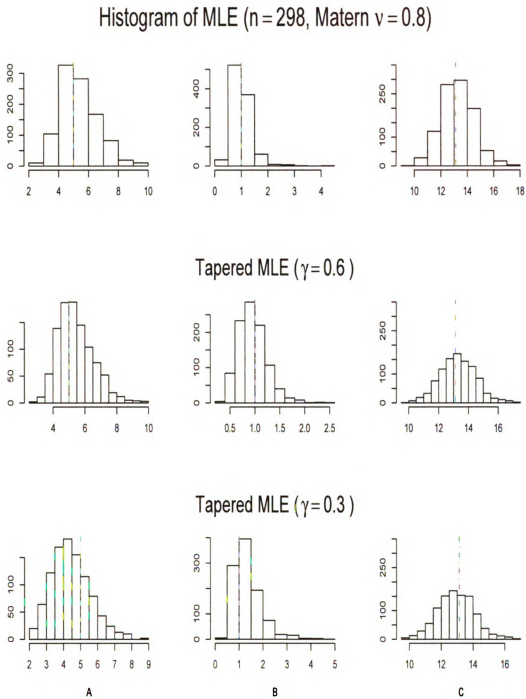

Figure 3.6: The histogram of MLE and tapered MLE (where A= $\hat{\theta}$, B= $\hat{\sigma}^2$, C= $\hat{\sigma}^2\hat{\theta}^{2\nu}$) with sample size 298.

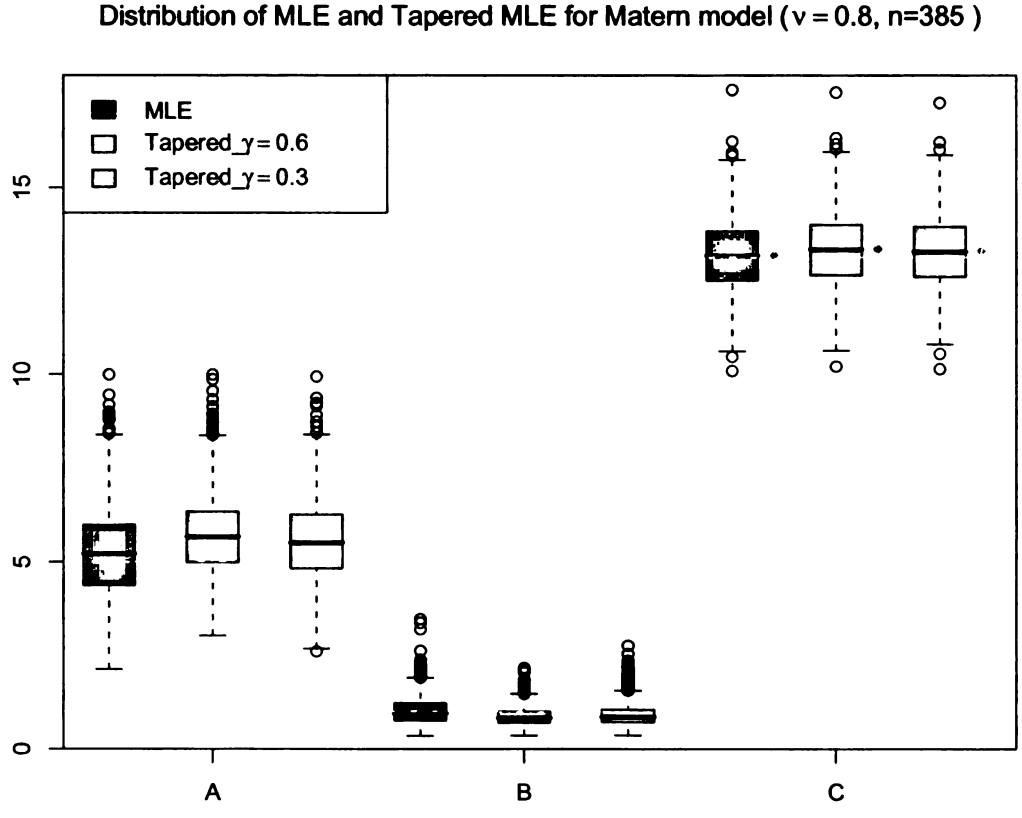

Figure 3.7: The boxplot of MLE and tapered MLE (where A= $\hat{\theta}$, B= $\hat{\sigma}^2$, C= $\hat{\sigma}^2\hat{\theta}^{2\nu}$) with sample size 385.

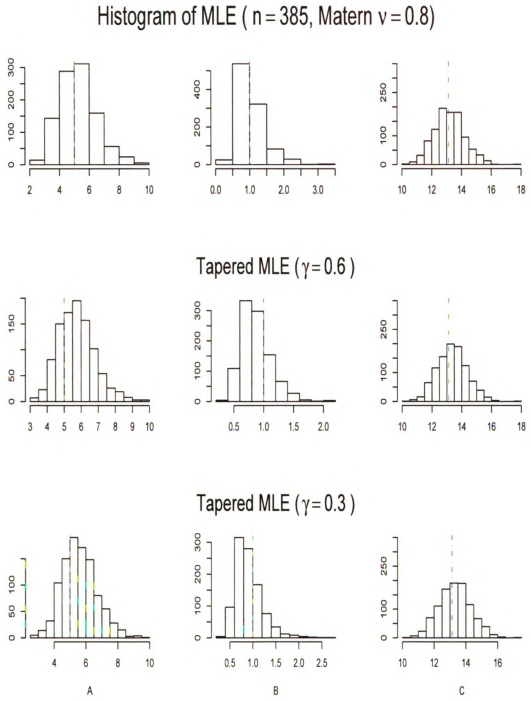

Figure 3.8: The histogram of MLE and tapered MLE (where A= $\hat{\theta}$, B= $\hat{\sigma}^2$, C= $\hat{\sigma}^2\hat{\theta}^{2\nu}$) with sample size 385.

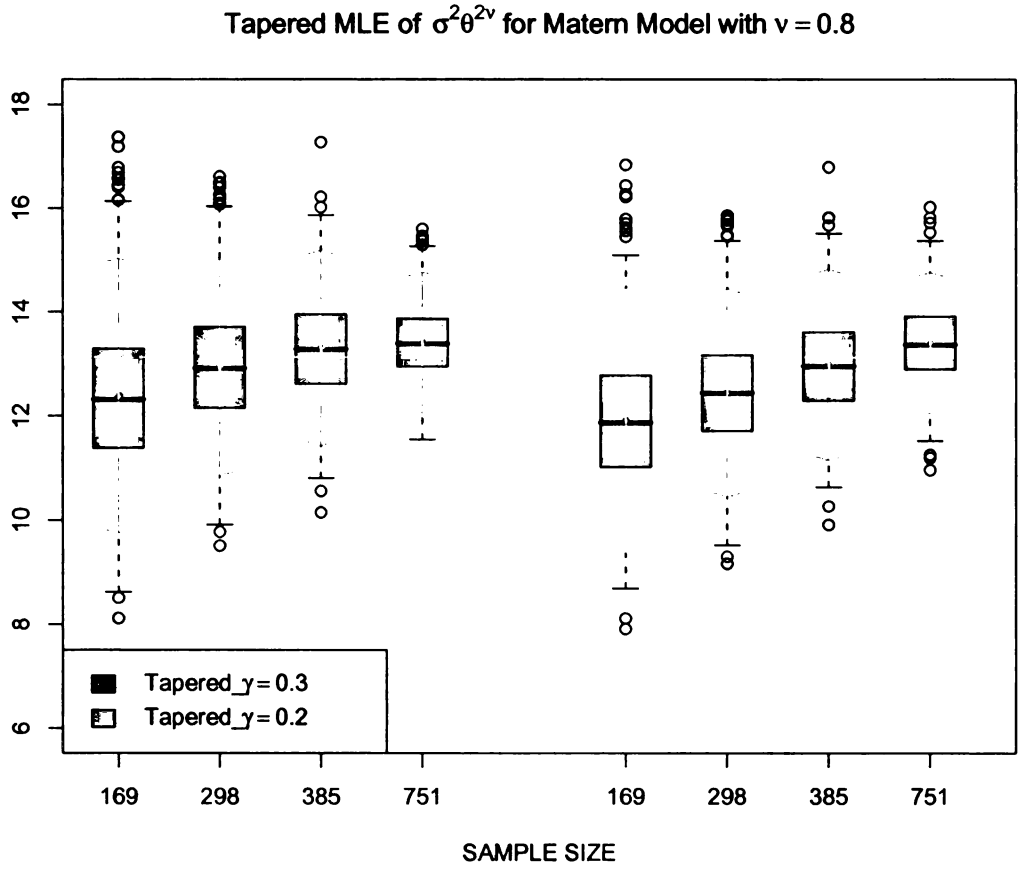

Figure 3.9: The boxplots of tapered MLE with different tapering range and increasing sample size.

Table 3.1: Average standard deviation(asd), relative coverage frequency(rcf), average length of intervals(al) of 95% CI of $\sigma^2\theta^{2\nu}$ for Matérn model with $\nu = 0.8$.

n	169	298	385	751
MLE rcf	0.93	0.92	0.93	0.94
al	5.67	4.26	3.73	2.65
asd	1.45	1.09	0.95	0.68
$\gamma = 0.6$ rcf	0.91	0.92	0.94	0.94
al	5.59	4.27	3.77	2.69
asd	1.43	1.09	0.96	0.69
$\gamma = 0.3$ rcf	0.85	0.91	0.94	0.93
al	5.28	4.15	3.76	2.71
asd	1.35	1.06	0.96	0.69
$\gamma = 0.2$ rcf	0.79	0.85	0.93	0.93
al	5.08	4.00	3.67	2.71
asd	1.30	1.02	0.94	0.69

Table 3.2: Timing and estimation for tapered MLE with tapering range $\gamma = 0.1$ using computer with CPU 2.8 Ghz and 8.00 GB RAM

n	1000	2000	3000	4000	5000	6000	7000	8000
MLE	14.25	12.46	13.26	13.06	13.01	13.66	13.17	13.47
s	41.77	208.92	507.83	966.31	1499.67	2590.83	3954.36	5584.57
TMLE	13.76	12.40	13.18	12.97	12.99	13.72	13.19	13.52
s	8.37	33.61	79.83	115.29	189.07	295.11	442.87	624.83

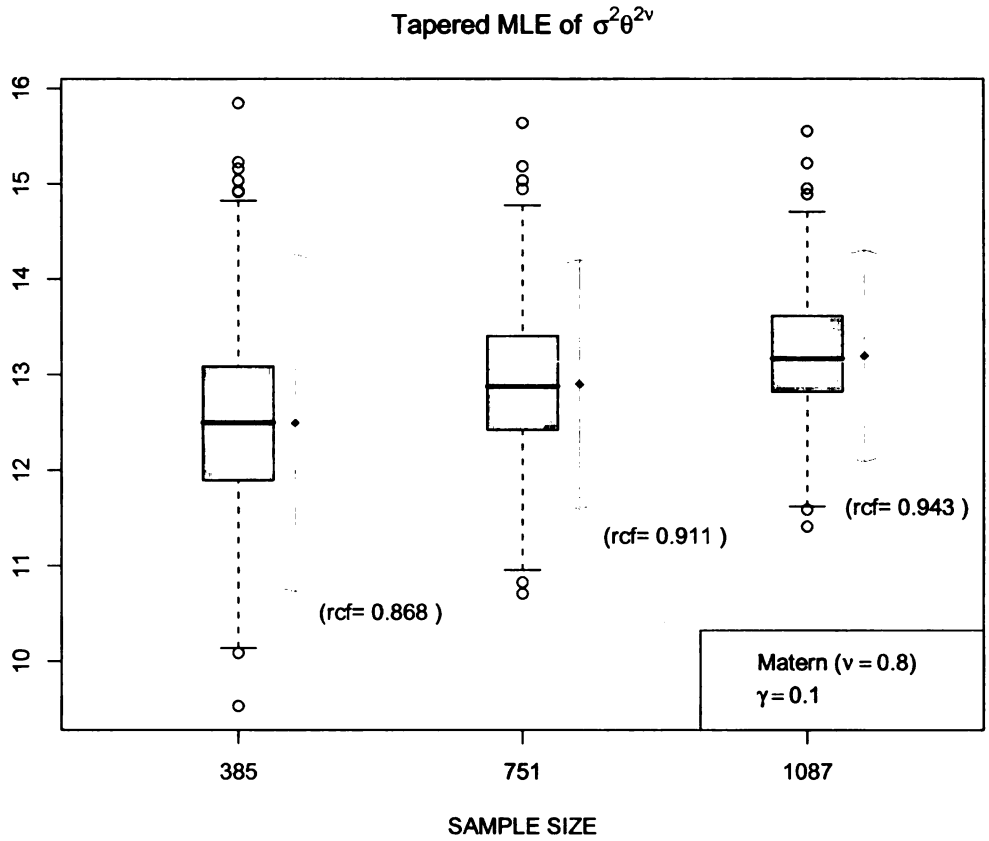

Figure 3.10: The boxplots of tapered MLE with severe degree of tapering range.

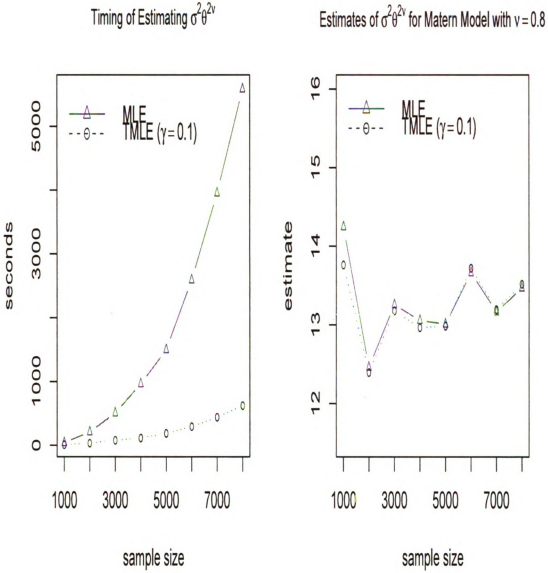

Figure 3.11: Timing and estimation for tapered MLE for Matérn model.

3.5.2 Fitting a Matérn model to climate data

One of the applications of tapering is to fit a model to a large climate data set, which has been common based on satellite observing or remote sensing systems. The tapering technique is often important to fill in the missing value or refine the irregular data observations to standard gridded version using kriging, because the kriging or estimation involving repeatedly inversion of large matrix can be computationally too heavy to be feasible. To demonstrate how tapering technique obviate these hurdles without sacrificing the accuracy in estimation, we chose to use the climate data set analyzed in Furrer, Genton and Nychka(2006). Actually the whole data base consists of monthly average temperature and total precipitation from 1895 to 1997 at 5,900 weather stations (Johns, et al.(2003)). It is accessible via http://www.image.ucar.edu/GSP/Data/US.monthly.met by the University Corporation for Atmospheric Research (UCAR).

In order to have exact MLE to compare against and make the data set closer to Gaussian and stationary, we studied the precipitation anomaly field of April 1948 recorded at 5909 stations (See coverage map Figure 3.12) as Furrer, Genton and Nychka(2006) did for kriging, where however they assume the parameters are all known. It is noted that the "anomaly" field here means the the raw monthly total precipitation recorded in the conterminous US for April 1948 were taken square root and standardized based on long run mean and standardization to meet the model assumptions (Fuentes et al., 1998, 2005). We fit a Matérn with $\nu = 0.3$ and calculate MLE and tapered MLE of consistently estimable parameter $\sigma^2\theta^{2\nu}$, tapering still with Wandland 1 taper with $\gamma = 50$ miles. From Table 3.3, we note that the tapered estimate 0.0158487 is very close to MLE 0.0158899, so is the standard error and 95% confidence interval (0.0153168, 0.0164630) vs. (0.0152771, 0.0164204) based on asymptotical result Theorem 3.3.8. However, the time to estimate exact MLE is almost 8 times longer (3.5 m vs. 28.3 m). Hence the accuracy and computational gain of the tapered MLE

Table 3.3: Fitting Matérn model with $\nu = 0.3$ and estimate consistently estimable parameter $\sigma^2\theta^{2\nu}$, tapering with $\gamma = 50$ miles.

	estimate	SD	95% CI	length	time (s)
MLE	0.0158487	0.0002917	(0.0152771, 0.0164204)	0.0011432	1698.99
TMLE	0.0158899	0.0002924	(0.0153168 , 0.0164630)	0.0011462	212.95

are obtained by applying the tapering method to large datasets. The more sizable computational gain will be archived if we work on even larger datasets.

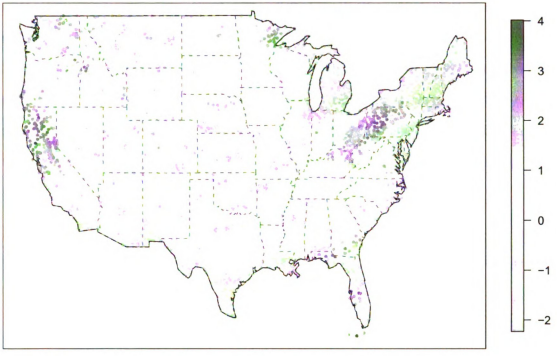

Figure 3.12: The Precipitation Anomaly Field of April 1948.

3.6 Future Work

There are some open problems for future research. First, for the Matérn model, the estimator of $\sigma^2\theta^{2\nu}$ is constructed by fixing θ at an arbitrary value. For a finite sample, common practice is to also estimate θ. It is an interesting question to see if Theorem 3.3.1 and Theorem 3.3.8 still hold for the MLE and tapered MLE by jointly maximizing

these two parameters. Our conjecture is that Theorem 3.3.8 can be extended to this case.

Second, the theoretical results in this work are for the processes with one dimensional index. It is a more practical problem to study the high dimensional case. This is our on-going research.

Third, in our study, we fixed the tapering range. We can consider another kind of tapering regime, which is to let the tapering range depend on the sample size and tend to zero with increasing sample. This will give a lot faster calculation, but we still need to find out theoretical justification of this type of tapering.

Finally, huge datasets arise most frequently when spatial locations are observed repeatedly over time. So we believe covariance tapering is then potentially more powerful to deal with the large spatio-temporal data. This is another direction of my current research.

Bibliography

Abramowitz, M. and Stegun, I. (1967). *Handbook of Mathematical Functions.* U.S. Government Printing Office, Washington, D.C. (eds.).

Aronszajn(1950). Theory of reproducing kernels. *Transactions of the American Mathematical Society*, 68:337–404, 1950.

Blackwell, D. and Dubins, L. (1962). Merging of opinions with increasing information. *Annals of Mathematical Statistics*, **33**, 882–886.

Chatterji, S.D., Mandrekar, V.S., (1978). Equivalence and singularity of Gaussian measures and applications, *Probabilistic Analysis and Related Topics* (Ed. A.T. Barucha Ried), **1**, AP, 169–197,

Chen, H.-S., Simpson, D.G., Ying, Z., (2000). Infill asymptotics for a stochastic process model with measurement error. Statist. Sinica 10, 141–156.

Davis, T. A. (2006). *Direct Methods for Sparse Linear Systems.* Philadelphia: Society for Industrial and Applied Mathematics.

Du, J., Zhang, H., and Mandrekar, V. (2007). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *Accepted by the Annals of Statistics* .

Furrer, R., Genton, M. G., and Nychika, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics*, **15**(3), 502–523.

Gilbert, J. R., Moler, C. and Schreiber, R. (1992). Sparse matrices in MATLAB: Design and implementation. *SIAM J. Matrix Anal. Appl*, **13**, 333–356.

Gneiting, T. (1999). Radial positive definite functions generated by Euclids hat. *J. Multivar. Anal.* **69**, 88–119.

Gneiting, T. (2002). Compactly supported correlation functions. *Journal of Multivariate Analysis*, **83**(2), 493–508.

Graybill, Franklin A. (1983). *Matrices with Applications in Statistics* (Second Edition). Wadsworth, Belmont, California.

IR(1978)) Ibragimov, I. A. and Rozanov, Yu. A. (1978). *Gaussian Random Processes.* Springer-Verlag, New York.

Kaufman, C. (2005). Covariance Tapering for Likelihood-based Estimation in Large Spatial Datasets. *Ph.D Thesis Proposal*, 1–33.

Kaufman, C., Schervish, M. and Nychka, D. (2008). Covariance tapering for likelihood-based estimation in large spatial datasets. *Journal of the American Statistical Association.* (In press).

LeMone, M.A., Chen, F., Alfieri, J.G., Cuenca, R., Hagimoto, Y., Blanken, P.D., Niyogi, D., Kang, S., Davis, K. and Grossman, R. (2007). NCAR/CU surface vegetation observation network during the international H2O Project 2002 field campaign. *Bulletin of the American Meteorological Society*, **88**, 65–81.

Loh, W.-L. (2005). Fixed-domain asymptotics for a subclass of Matérn-type Gaussian random fields. *The Annals of Statistics*, **33**, 2344–2394.

Mardia,K. V. and Marshall,R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71**, 135–46.

Mirsky, L. (1955). *An Introduction to Linear Algebra.* Oxford: Oxford University Press.

Mitra, S., Gneiting, T., and Sasvári, Z. (2003). Polynomial covariance functions on intervals. *Bernoulli*, **9**(2), 229–241.

Pissanetzky, S. (1984). *Sparse Matrix Technology.* Academic Press.

Ripley, B. D. (1981). *Spatial Statistics.* Wiley, New York.

Rudin, W, (1966). *Real and Complex Analysis*, McGraw Hill.

Stassberg, D., LeMone, M.A., T. Warner, T., and Alfieri, J. G. (2008). Comparison of observed 10 m wind speeds to those based on Monin-Obukhov similarity theory using aircraft and surface data from the International H2O Project. *Mon. Wea. Rev.* **136**, 964–972.

Searle, S.R. (1971). *Linear Models.* Wiley, New York.

Stein, M. L. (1988). Asymptotically efficient prediction of a random field with a misspecified covariance function. *The Annals of Statistics* **16**, 55–63.

Stein, M. L. (1990a). Uniform asymptotic optimality of linear predictions of a random field using an incorrect second-order structure. *Annals of Statistics*, **18**, 850–872.

Stein, M. L. (1990b). Bounds on the efficiency of linear predictions using an incorrect covariance function. *The Annals of Statistics* **18**, 1116–1138.

Stein, M. L. (1990c) A comparison of generalized cross validation and modified maximum likelihood for estimating the parameters of a stochastic process. *The Annals of Statistics* **18**, 1139–1157.

Stein, M. L. (1999). Predicting random frields with increasing dense observations. *Annals of Statistics*, **9**, 242–273.

Stein, M. L. (1999). *Interpolation of Spatial Data: Some Theory for Kriging.* Springer, New York.

Weckworth, T.M., Parsons, D.B., Koch, S.E., J.A. Moore, M.A. LeMone, B.B. Demoz, C. Flamant, B. Geerts, J. Wang, W.F. Feltz (2004). An overview of the international H2O project (IHOP_2002) and some preliminary highlights. *Bull. Amer. Meteor. Soc.* **85**, 253–277.

Wendland, H. (1995). Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics*, **4**, 289–396.

Wendland, H. (1998). Error estimates for interpolation by compactly supported radial basis functions of minimal degree. *Advances in Computational Mathematics*, **93**, 258–272.

Wu, Z. M. (1995). Compactly supported positive definite radial functions. *Advances in Computational Mathematics*, **4**, 283–292.

Ying, Z. (1991). Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *Journal of Multivariate Analysis*, **36**, 280–296.

Ying, Z. (1993). Maximum likelihood estimation of parameters under a spatial sampling scheme. *Ann. Statist.* **21**, 1567–1590.

Zhang, H. (2004). Inconsistent estimation and asymptotically equivalent interpolations in model-based geostatistics. *Journal of the American Statistical Association*, **99**, 250–261.

Zhang, H and Du, J (2008). Covariance Tapering in Spatial Statistics. *in Positive definite functions: From Schoenberg to space-time challenges*, (eds) J. Mateu and E. Porcu. Gráficas Castañ, Spain, 181–198. (ISBN: 978-84-612-8282-1)

Zhang, H. and Zimmerman, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika*, **92**, 921–936.