Local stationarity for spatial data

Vom Fachbereich Mathematik
der Technischen Universität Kaiserslautern
zur Verleihung des akademischen Grades
Doktor der Naturwissenschaften
(Doctor rerum naturalium, Dr. rer. nat.)
genehmigte Dissertation

von

Danilo Pezo Villar

1. Gutachter: Prof. Dr. Jürgen Franke, TU Kaiserslautern
2. Gutachter: Prof. Dr. Rainer Dahlhaus, Universität Heidelberg

Datum der Disputation: 20. Dezember 2017

D386

To my wife Dafne and my son Felipe

Abstract

Following the ideas presented in Dahlhaus (2000) and Dahlhaus and Sahm (2000) for time series, we build a Whittle-type approximation of the Gaussian likelihood for locally stationary random fields. To achieve this goal, we extend a Szegö-type formula, for the multidimensional and local stationary case and secondly we derived a set of matrix approximations using elements of the spectral theory of stochastic processes. The minimization of the Whittle likelihood leads to the so-called Whittle estimator $\widehat{\theta}_T$. For the sake of simplicity we assume known mean (without loss of generality zero mean), and hence $\widehat{\theta}_T$ estimates the parameter vector of the covariance matrix Σ_θ.

We investigate the asymptotic properties of the Whittle estimate, in particular uniform convergence of the likelihoods, and consistency and Gaussianity of the estimator. A main point is a detailed analysis of the asymptotic bias which is considerably more difficult for random fields than for time series. Furthemore, we prove in case of model misspecification that the minimum of our Whittle likelihood still converges, where the limit is the minimum of the Kullback-Leibler information divergence.

Finally, we evaluate the performance of the Whittle estimator through computational simulations and estimation of conditional autoregressive models, and a real data application.

Zusammenfassung

Aufbauend auf den Ansätzen von Dahlhaus (2000) und Dahlhaus and Sahm (2000) für Zeitreihen konstruieren wir eine Approximation vom Whittle-Typ für die Gaußsche Likelihoodfunktion lokal-stationärer Zufallsfelder. Als wesentliche Bausteine verallgemeinern wir einerseits die Szegö-Formel auf den mehrdimensionalen und lokal-stationären Fall und leiten andererseits eine Reihe von Matrixapproximationen unter Verwendung der Spektraltheorie stochastischer Prozesse her. Die Minimierung der Whittle-Likelihood führt dann zum so-genannten Whittle-Schätzer $\widehat{\theta}_T$. Der Einfachheit halber nehmen wir an, dass der Mittelwert bekannt (ohne Beschränkung der Allgemeinheit $= 0$) ist, so dass $\widehat{\theta}_T$ den Parametervektor der Kovarianzmatrix Σ_θ schätzt.

Wir untersuchen die asymptotischen Eigenschaften des Whittle-Schätzers, insbesondere die uniforme Konvergenz der Likelihoodfunktionen sowie die Konsistenz und asymptotische Normalität des Schätzers. Ein Hauptaspekt ist eine detaillierte Untersuchung des asymptotischen Bias, die für Zufallsfelder deutlich problematischer als für Zeitreihen ausfällt. Weiterhin zeigen wir, dass bei Vorliegen eines missspezifizierten Modells das Minimum der Whittle-Likelihood immer noch konvergiert, und zwar gegen das Minimum der Kullback-Leibler Informationsdivergenz.

Schließlich untersuchen wir die praktische Qualität des Whittle-Schätzers durch die Anwendung auf simulierte Daten und einen realen Datensatz.

Acknowledgements

I would like to express my gratitude to Prof. Dr. Jürgen Franke for his invaluable help and encouragement regarding my research along these three years. My gratitude also goes to Prof. Dr. Rainer Dahlhaus for providing with the draft of a paper on the topic, which was unfinished due to gaps in the proofs.

I would like to thank to Prof. Dr. Claudia Redenbach for her support regarding my work, specially for allowing me to participate in the Ph.D. seminar in spite of the different research areas.

My gratitude to the Deutscher Akademischer Austauschdienst (DAAD) for awarding me with the funds that made this project possible, to Ms. Birgitt Skailes for her always prompt answer to every request, and to Dr. Falk Triebsch for supporting us as Ph.D. students as much as his position allows him.

I will never forget my friends and colleagues from the mathematics department and Fraunhofer Institut ITWM who always made me feel at home. Dr. Euna Nyarige, Dr. Rumana Rois, Pak Lo, Yolanda Rodríguez, Martina Sormani, Dr. Prakash Easwaran, Ms. Beate Seagler, Jessica Borsche among others, thanks to you the Uni was like a second home where I could profit from your knowledge and friendship.

I am deeply grateful to Dr. Katja Schladitz, who selflessly offered me help with the language. Thanks because every time we met, I received your advices and comments which enriched my perspectives, thanks for helping me "open my horizons" as you once mentioned it.

I know that this dream made my family in Chile to miss me and pray that everything goes fine. I just thank you mom for all your prayers, for God was always and still is by me side.

I am absolutely sure, that even receiving all the support and advice from the people above, I would not have succeeded without the love and company of my dear wife and son. From the bottom of my heart I thank you, Dafne and Felipe, you are my most precious God's gift.

Contents

Lists

Figures

Tables

Introduction

Stationarity is an ubiquitous concepts in stochastics. By its very definition (unchanged probabilistic structure, roughly speaking), it draws a framework where asymptotic results are possible and meaningful. Nevertheless, in real-world applications such assumption often does not fit well. In many situations the probabilistic properties of data change along time or space in unknown way making the modelling process rather difficult. In other words, increasing data does not necessarily give additional information about the overall properties as it typically does in the stationary world and the reason is simple: the probabilistic structure changes along the index set.

Unfortunately there is no unique natural way how to extend the definition of stationarity in a general and still useful form. In this research we, therefore, have used a special, but wide enough class of processes introduced in Dahlhaus (1997) and Dahlhaus and Sahm (2000) for time series and random fields respectively, which comprises a natural nonstationary extension of many classical (linear) stationary processes used in practice, like autoregressive, moving average, conditional autoregressive and so on, by allowing them to behave only locally as stationary.

The class of locally stationary random field processes contains many of the typical processes for stationary random fields such as CAR or SAR, turning into a very natural nonstationary extension of some stationary process. Thus, a crucial problem in this regard corresponds to the parametric estimation of a locally stationary model. In the stationary setting the Whittle estimator (Whittle (1954), originally for data on the plane but easily extended to any dimension, see Dahlhaus and Künsch (1987)) has turned into a simple and fast method of estimation which takes place in the spectral domain, *i.e.* using the parametric spectral density of the model and the periodogram, which makes calculations faster by means of the FFT algorithm.

Following the ideas presented in Dahlhaus (2000) for time series, we build a Whittle-type approximation (Whittle likelihood) of the Gaussian likelihood for locally stationary random fields. The minimization of the Whittle likelihood leads to the so-called Whittle estimator $\widehat{\theta}_T$. For the sake of simplicity we assume known mean (without loss of generality zero mean), and hence $\widehat{\theta}_T$ estimates the parameter vector of the covariance matrix Σ_θ. We investigate some of its asymptotic properties, namely consistency and Gaussianity. Finally, we evaluate the performance of the estimator through computational simulations and a real data application.

The structure of this work is as follows:

In Chapter 1 we make the notion of locally stationary random field and its relation with linear process precise. Notation and a few conventions to be used are introduced.

In Chapter 2 we construct the Whittle likelihood $\mathcal{L}_T(\theta)$, and thus our Whittle estimator. This is done mainly through an approximation of the Gaussian likelihood using and extension of the Szegö's theorem. The score function of $\mathcal{L}_T(\theta)$ is proved to be biased, we identify the source of this problem and discuss ways to reduce it.

In Chapter 3 we prove consistency and a modified Gaussian law for both, the tapered Whittle estimator and the exact Gaussian estimator. This is achieved by proving asymptotic properties of the likelihoods involved.

In Chapter 4 we present a simulation study considering a 2 dimensional autoregressive process of order one. We estimate the parameter vector using tapered and non tapered Whittle likelihoods, together with different set of parameters and assumptions. We model the classical wheat-yield data set of Mercer and Hall (1911) using a local stationary SAR model.

Finally, we present conclusions and possible directions of future research.

1 Locally Stationary Random Fields

In this chapter we introduce the definition of a locally stationary random field and discuss some implication of it. We begin this exposition giving some basic definitions coming from the stationary setting to motivate and link the locally stationary case. Some assumptions and notation are introduced as well.

1.1 Basic Definitions

Definition 1.1.1 (Random field). *Let $(\Omega, \mathcal{F}, P)$ be a probability space. Given $d, \delta \in \mathbb{N}$, we say that a function X:*

$$X : \Omega \times \mathbb{Z}^d \to \mathbb{R}^\delta$$

is a Random Field *if and only if $\forall t \in \mathbb{Z}^d$, the function:*

$$\begin{aligned} X_t \quad : \quad & \Omega \to \mathbb{R}^\delta \\ & \omega \to X(\omega, t) \end{aligned}$$

is $\mathcal{F} - \mathcal{B}(\mathbb{R}^\delta)$-measurable; where $\mathcal{B}$ denotes the Borel σ-algebra.

We will assume henceforth only univariate random fields, *i.e.* $\delta = 1$. A random field $\{X_t, t \in \mathbb{Z}^d\}$ is described by its finite dimensional distributions

$$F_{X_{t_1},\ldots,X_{t_k}}(x_1, \ldots, x_k) = \mathbb{P}(X_{t_1} \leq x_1, \ldots, X_{t_k} \leq x_k)$$

where the next two consistency conditions must be fulfilled, *i.e.*

Symmetry: $F_{X_{t_1},\ldots,X_{t_k}}(x_1, \ldots, x_k) = F_{X_{\pi 1},\ldots,X_{\pi k}}(x_{\pi 1}, \ldots, x_{\pi k})$, where π is a permutation.

Compatibility: $F_{X_{t_1},\ldots,X_{t_k}}(x_1, \ldots, x_{k-1}) = F_{X_{t_1},\ldots,X_{t_{k-1}},X_{t_k}}(x_1, \ldots, x_{k-1}, \infty)$.

Definition 1.1.2 (Gaussian random field). *A random field $\{X_t, t \in \mathbb{Z}^d\}$ is called Gaussian if $F_{X_{t_1},\ldots,X_{t_k}}$ are multivariate Gaussian distributions for any choice of k and $(t_1, \ldots, t_k) \in \mathbb{Z}^d$.*

Since a multivariate Gaussian distribution is completely specified by its mean μ and covariance matrix Σ, if we restrict the class of random fields to those with constant mean and positive definite Σ, then we speak of a stationary random field.

Definition 1.1.3 (Stationarity). *A random field is called (weakly) stationary if $E X_t = \mu$ and $\Sigma_{r,s} = Cov(X_r, X_s) = Cov(X_{r-s}, X_0)$ for all $t, r, s \in \mathbb{Z}^d$.*

The Cramér spectral representation theorem guarantees that assuming $\mu = 0$ and the summability of the entries $\Sigma_{r,s}$, the stationary process X_t can be represented as

$$X_t = \int_{\Pi^d} A(\lambda) \exp(i\langle\lambda, t\rangle)\, \mathrm{d}\xi(\lambda), \tag{1.1}$$

where $\xi(\lambda)$ is an orthogonal process, $A(\lambda)$ is the transfer function of the process, $\langle\lambda, t\rangle = \sum_{i=1}^{d} \lambda_i t_i$ and $\Pi^d = (-\pi, \pi]^d$. In other words, the process X_t can be represented as a continuous superposition of sinusoids with random amplitudes (for further details see Brillinger (1981)). In the next section we define the locally stationary case by an extension of (1.1).

1.2 Locally Stationary Random Fields

The concept of stationarity describes essentially the situation where the statistical properties of a stochastic process do not change as it moves in the index set (usually time or space). This implies that more observations improve the knowledge about the structure of the overall process. Unfortunately, in the general nonstationary frame this is clearly not the case. In time series, for instance, looking into the future gives not necessarily information about the current state of the process.

To overcome this problem, Dahlhaus (1997) has proposed an asymptotic approach similar to nonparametric regression. To exemplify it let us assume a time-varying AR(1)

$$X_t = g(t;\theta)X_{t-1} + \varepsilon_t, \quad \varepsilon_t \sim_{iid} \mathcal{N}(0, \sigma^2),$$

with $|g(t;\theta)| < 1$ for $t = 1, \ldots, T$ and parameter vector θ. Depending on the functional form of $g(t;\theta)$, this might not behave as stationary for times bigger than T, giving no useful information about its parameters. Therefore, instead of looking into the future when $T \to \infty$, we rescale the function g, *i.e.* $g(t/T)$, obtaining more and more realizations of the local process as T increases. This does not mean to have a higher sampled continuous process, but only an abstract setting where statistical inference over the parameter vector θ is possible. This idea will allow us to find Quasi-ML estimators and their asymptotic properties in the next chapters. Notice that in this setting we do not have a sequence $X_1, \ldots, X_T$ any more, but a triangular array $X_{1,T}, \ldots, X_{T,T}$ indexed by T.

To keep notation simple, we will only consider processes on the cube $D_T = \{1, \ldots, T\}^d$, however the results are still valid for other cuboids, provided that the side length increases proportionally as we consider more and more data points (see Guyon (1995)). The next definition corresponds to the definition of locally stationary random field introduced in Dahlhaus and Sahm (2000).

Definition 1.2.1 (Locally Stationary Random Field)**.** *A sequence of stochastic processes* $(X_{t,T})_{t\in D_T}$ *is called* locally stationary with transfer function A^0 and mean function μ, *if there is a representation*

$$X_{t,T} = \mu(\tfrac{t-1/2}{T}) + \int\limits_{\Pi^d} A^0_{t,T}(\lambda) \exp(i\langle\lambda, t\rangle)\, \mathrm{d}\xi(\lambda), \tag{1.2}$$

such that

(i) $\xi(\lambda)$ *is a stochastic process on* Π^d *with orthogonal increments and* $\overline{\xi(\lambda)} = \xi(-\lambda)$ *for all* $\lambda \in \Pi^d$ *and*

$$Cum\,(\mathrm{d}\xi(\lambda_1), \dots, \mathrm{d}\xi(\lambda_k)) = \eta\Big(\sum_{j=1}^{k} \lambda_j\Big) h_k(\lambda_1, \dots, \lambda_{k-1})\, \mathrm{d}\lambda_1 \cdots \mathrm{d}\lambda_k,$$

where $Cum\,(\dots)$ *denotes the cumulant of* k*th order,* $h_1 = 0$*,* $h_2(\lambda) = 1$*,* $|h_k(\lambda_1, \dots, \lambda_{k-1})| \leq const_k$ *for all* k *and* $\eta(\lambda) = \sum_{j=-\infty}^{+\infty} \delta(\lambda + 2\pi j)$ *represents the Dirac comb.*

(ii) There is a constant K *and a function* $A : [0,1]^d \times \Pi^d \to \mathbb{C}$ *with* $\overline{A(u,\lambda)} = A(u,-\lambda)$ *such that*

$$\sup_{t,\lambda} |A^0_{t,T}(\lambda) - A(\tfrac{t-1/2}{T}, \lambda)| \leq \frac{K}{T}. \tag{1.3}$$

Under some regularity conditions on $A(u,\lambda)$ as a function of the spatial component u defined as $u = \frac{t-1/2}{T}$, it can be proved that the space varying spectrum $f_T(u,\lambda)$ (see Martin and Flandrin (1985) in the time series case) converges in $L^2(\Pi^d)$ to $f(u,\lambda)$. Consequently, $f(u,\lambda) = |A(u,\lambda)|^2$ is called the *varying spectral density of the field.*

We shall shift the points $t \in D_T$ by 1/2 in order to avoid an additional edge effect. Henceforth, we assume that $X_{t,T}$ is Gaussian, *i.e.* $h_k(\lambda) = 0$ for all $k \geq 3$, and for the sake of simplicity μ will be known (without loss of generality $\mu \equiv 0$), nevertheless all the results are easily extendable for the unknown mean case. The condition (1.3) is needed because of two reasons: First, in order to have a tractable mathematical framework we need some degree of smoothness in the spatial component u, which is guaranteed by the regularity of the function $A(u,\lambda)$. Second, an assumption of equality ($A^0_{t,T}(\lambda) = A(\frac{t-1/2}{T}, \lambda)$) would be too restrictive since important families of models, like CAR or AR, would remain excluded. We will return to this point later. Let us discuss the close connection between the representations (1.2)-(1.3) of $X_{t,T}$ and a MA(∞) representation.

Let us define the Fourier coefficients of $A^0_{t,T}(\lambda)$ and $A(u,\lambda)$ as

$$a_{t,T,k} := \int\limits_{\Pi^d} A^0_{t,T}(\lambda) \exp(i\langle\lambda, k\rangle)\, \mathrm{d}\lambda, \tag{1.4}$$

$$a_k(u) := \int\limits_{\Pi^d} A(u,\lambda) \exp(i\langle\lambda, k\rangle)\, \mathrm{d}\lambda, \tag{1.5}$$

respectively, and

$$\varepsilon_t := \int\limits_{\Pi^d} \exp(i\langle\lambda, t\rangle)\,\mathrm{d}\xi(\lambda).$$

The orthogonality of the process $\xi(\lambda)$ implies $\mathrm{E}\,\varepsilon_t = 0$ and $\mathrm{E}\,\varepsilon_{t_1}\varepsilon_{t_2} = (2\pi)^d\delta(t_1 - t_2)$. Inverting expressions (1.4) and (1.5) we obtain

$$\begin{aligned} A^0_{t,T}(\lambda) &= \frac{1}{(2\pi)^d}\sum_{k\in\mathbb{Z}^d} a_{t,T,k}\exp(-i\langle\lambda, k\rangle), \\ A(u,\lambda) &= \frac{1}{(2\pi)^d}\sum_{k\in\mathbb{Z}^d} a_k(u)\exp(-i\langle\lambda, k\rangle), \end{aligned}$$

then

$$\begin{aligned} X_{t,T} &= \int\limits_{\Pi^d} A^0_{t,T}(\lambda)\exp(i\langle\lambda, t\rangle)\,\mathrm{d}\xi(\lambda) \\ &= \frac{1}{(2\pi)^d}\sum_{k\in\mathbb{Z}^d} a_{t,T,k}\int\limits_{\Pi^d}\exp(i\langle\lambda, t-k\rangle)\,\mathrm{d}\xi(\lambda) \\ &= \frac{1}{(2\pi)^d}\sum_{k\in\mathbb{Z}^d} a_{t,T,k}\varepsilon_{t-k}, \qquad (1.6) \end{aligned}$$

i.e. $X_{t,T}$ can be expressed as a linear process. This connection has also been reported in the case of time series (see Dahlhaus (2000)). Inequality (1.3) implies

$$\sup_t |a_{t,T,k} - a_k(\tfrac{t-1/2}{T})| \le (2\pi)^d \sup_{t,\lambda}|A^0_{t,T}(\lambda) - A(\tfrac{t-1/2}{T},\lambda)| = \mathcal{O}(T^{-1}).$$

Conversely, if we assume that the representation (1.6) holds, then we can easily obtain the representation (1.2). On the other hand, if

$$\sup_t \sum_{k\in\mathbb{Z}^d} |a_{t,T,k} - a(\tfrac{t-1/2}{T})| = \mathcal{O}(T^{-1})$$

holds, then

$$\sup_{t,k}|A^0_{t,T}(\lambda) - A(\tfrac{t-1/2}{T},\lambda)| \le \sup_t \frac{1}{(2\pi)^d}\sum_{k\in\mathbb{Z}^d}|a_{t,T,k} - a_k(\tfrac{t-1/2}{T})| = \mathcal{O}(T^{-1}),$$

and hence, the condition (1.3) is fulfilled too. Taking into account this connection between (1.2) and MA(∞) processes, it is easier to understand the necessity of the bound (1.3). We may consider, for instance, a 2D space-varying AR(1), *i.e.*

$$X_{r,s;T} = \alpha(\tfrac{r-1/2}{T}, \tfrac{s-1/2}{T})X_{r-1,s;T} + \beta(\tfrac{r-1/2}{T}, \tfrac{s-1/2}{T})X_{r,s-1;T} + \sigma\varepsilon_{r,s}.$$

After some manipulations we verify from the MA(∞) representation that

$$\begin{aligned} A^0_{t,T}(\lambda) &\neq A(u,\lambda) \\ &= \frac{\sigma}{2\pi(1-\alpha(u_1,u_2)\exp(-i\lambda_1)-\beta(u_1,u_2)\exp(-i\lambda_2))}, \end{aligned} \tag{1.7}$$

but

$$A^0_{t,T}(\lambda) = A(u,\lambda) + \frac{1}{T}B(u,\lambda) + \mathcal{O}(T^{-2}),$$

with $u = (u_1,u_2) = (\frac{r-1/2}{T}, \frac{s-1/2}{T})$, $\lambda = (\lambda_1,\lambda_2)$ and where $B(u,\lambda)$ is usually not zero (see Dahlhaus and Sahm (2000) for a closed-form expression). Making the bound (1.3) stricter, we would need to find the higher order terms in the Taylor expansion in order to fulfil the desired precision, which can be very hard for some families of models. Nevertheless, the simulations conducted in the Chapter 4 showed a good enough performance, considering a transfer function of order one, *i.e.* satisfying (1.3).

Regarding the covariance, straightforward calculations using equation (1.2) yield

$$\Sigma_T(A^0)_{r,s} = \int_{\Pi^d} A^0_{r,T}(\lambda)\overline{A^0_{s,T}(\lambda)}\exp(i\langle\lambda, r-s\rangle)\,\mathrm{d}\lambda.$$

The bound (1.3) implies that $A^0_{t,T}(\lambda) = A(\frac{t-1/2}{T},\lambda) + \mathcal{O}(T^{-1})$. Thus replacing in the last equation we obtain

$$\begin{aligned} \Sigma_T(A^0)_{r,s} &= \mathrm{Cov}\,(X_{r,T}, X_{s,T}) \\ &= \int_{\Pi^d} A(\tfrac{r-1/2}{T},\lambda)\overline{A(\tfrac{s-1/2}{T},\lambda)}\exp(i\langle\lambda, r-s\rangle)\,\mathrm{d}\lambda + \mathcal{O}(T^{-1}) \\ &= \Sigma_T(A)_{r,s} + \mathcal{O}(T^{-1}). \end{aligned} \tag{1.8}$$

The mean value theorem guarantees for some ξ_1 and ξ_2 that

$$\begin{aligned} A(\tfrac{r-1/2}{T},\lambda) &= A(\tfrac{r+s-1}{2T},\lambda) + \tfrac{1}{2T}(r-s)'\nabla_u A(u,\lambda)|_{u=\xi_1}, \\ A(\tfrac{s-1/2}{T},-\lambda) &= A(\tfrac{r+s-1}{2T},-\lambda) + \tfrac{1}{2T}(s-r)'\nabla_u A(u,-\lambda)|_{u=\xi_2}, \end{aligned}$$

which by multiplication and integration yields

$$\begin{aligned} &\int_{\Pi^d} A(\tfrac{r-1/2}{T},\lambda)A(\tfrac{s-1/2}{T},-\lambda)\exp(i\langle\lambda, r-s\rangle)\,\mathrm{d}\lambda = \\ &\qquad\qquad \int_{\Pi^d} |A(\tfrac{r+s-1}{2T},\lambda)|^2\exp(i\langle\lambda, r-s\rangle)\,\mathrm{d}\lambda + \mathcal{O}(T^{-1}). \end{aligned} \tag{1.9}$$

This implies that up to an error term of order $\mathcal{O}(T^{-1})$ (due to the local stationarity) and recalling that $f(u,\lambda) = |A(u,\lambda)|^2$, the right hand side of the equation (1.9) is

$$\int_{\Pi^d} f(\tfrac{r+s-1}{2T},\lambda)\exp(i\langle\lambda, r-s\rangle)\,\mathrm{d}\lambda, \tag{1.10}$$

which corresponds to the natural approximation of the covariances $\Sigma_T(A^0)_{r,s}$. We will return to this point in the Chapter 2.
Following the arguments in the 1D stationary case (see Dzhaparidze (1986)), we may use the matrix

$$\Big\{ \int_{\Pi^d} f(\tfrac{r+s-1}{2T}, \lambda)^{-1} \exp(i\langle \lambda, r-s \rangle) \, \mathrm{d}\lambda \Big\}_{r,s\in D_T} \tag{1.11}$$

as an approximation of $\Sigma_T^{-1}(A^0)$. It will be shown later that by using the matrix $U_T((2\pi)^{-2d} f^{-1})$ where

$$U_T(\phi)_{r,s} = \int_{\Pi^d} \phi(\tfrac{1}{T}[\tfrac{r+s}{2}] - \tfrac{1}{2T}, \lambda) \exp(i\langle \lambda, r-s \rangle) \, \mathrm{d}\lambda, \quad r,s \in D_T \tag{1.12}$$

($[x]$ denotes the smaller integer smaller than or equal to x componentwise for the vector x), can be obtained an easier interpretable formula.

2 Whittle Estimator

In Chapter 1 we introduced the definition of a locally stationary random field (LSRF) following Dahlhaus and Sahm (2000). In this chapter we consider the problem of estimation, *i.e.*, given zero-mean Gaussian observations $X_{t,T}, t \in D_T$ coming from a locally stationary process, we want to fit a parametric model with parameter vector $\theta \in \Theta$. Due to the computational burden implied from the exact Gaussian log-likelihood, a common approach to overcome this problem consists in using the Whittle estimator (Whittle (1954)). For the sake of clarity in the exposition, we start presenting this estimator for the much simpler stationary case. The construction of the Whittle likelihood for LSRFs is much more involved as it makes use of an extension of the Szegö formula and a matrix approximation (see (1.12)). We will show that the score function of our Whittle likelihood has a bias of order $\mathcal{O}(T^{-1})$ whose sources will be identified. Since the bias is transferred to the Whittle estimator (see Proposition 2.3.1), some modification are introduced in the Whittle likelihood in order to decrease the bias order.

2.1 Whittle Estimator for Stationary Random Fields

Let $\Pi^d = (-\pi, \pi]^d$. Suppose a zero-mean stationary random process X_t, $t = (t_1, \ldots, t_d) \in \mathbb{Z}^d$ with spectral density $f(\lambda)$, $\lambda = (\lambda_1, \ldots, \lambda_d) \in \Pi^d$ such that

(a) $\log f \in L^1(\Pi^d)$.

This implies that $\log f(\lambda)$ has a Fourier expansion, and therefore, $f(\lambda)$ may be represented as

$$f(z_1, \ldots, z_d) = \exp\Big(\sum_{k \in \mathbb{Z}^d} a_{k_1,\ldots,k_d} z_1^{k_1} \cdots z_d^{k_d} \Big), \tag{2.1}$$

where $z_j := \exp(i\lambda_j)$ and $k := (k_1, \ldots, k_d)$. Defining

$$\begin{aligned} P(z_1, \ldots, z_d) = \exp\Big(-\Big(\frac{a_{0,\ldots,0}}{2} + \sum_{k_d=1}^{\infty} a_{0,\ldots k_d} z_d^{k_d} + \sum_{k_{d-1}=1}^{\infty} \sum_{k_d=-\infty}^{\infty} a_{0,\ldots,k_{d-1},k_d} z_{d-1}^{k_{d-1}} z_d^{k_d} + \\ \ldots + \sum_{k_1=1}^{\infty} \sum_{k_2=-\infty}^{\infty} \cdots \sum_{k_d=-\infty}^{\infty} a_{k_1,\ldots,k_d} z_1^{k_1} \cdots z_d^{k_d} \Big)\Big) \end{aligned} \tag{2.2}$$

after some algebra, we observe that

$$f(z_1,\dots,z_d) = \frac{1}{P(z_1,\dots,z_d)P(z_1^{-1},\dots,z_d^{-1})} = \frac{1}{|P(z_1,\dots,z_d)|^2}. \tag{2.3}$$

Assuming additionally that

(b) $P(z_1,\dots,z_d)$ has a Fourier expansion,

the model

$$P(B_{t_1},\dots,B_{t_d})X_t = \varepsilon_t, \qquad \sigma^2(\varepsilon) = 1, \tag{2.4}$$

has associated the spectral density function $f(\lambda)$, where $B_{t_j}, j = 1,\dots,d$ are forward shift operators, *i.e.* $B_{t_j}^l X_{t_1,\dots,t_d} = X_{t_1,\dots,t_j+l,\dots,t_d}$.

As it was pointed out in Whittle (1954), conditions (a) and (b) are enough for a process to be found which generates a given set of autocorrelations and in which X_t is expressed as an autoregression upon X_u where $u < t$ in terms of lexicographic order over half-hyperplane.

The next theorem uses this unilateral representation to obtain the Whittle likelihood, which results in a more tractable expression than the exact Gaussian log-likelihood, and consequently it allows to get asymptotic ml-estimators.

Theorem 2.1.1 (Whittle (1954))**.** *Let $X_t, t \in D_T$, be a sequence of observations from a zero-mean stationary Gaussian random field with spectral density function $f_\theta(\lambda)$, $\lambda \in \Pi^d$, $\theta \in \Theta$. Then the joint likelihood is given, apart from boundary-effects, by the expression*

$$p_X(x) = \frac{1}{(2\pi V)^{T^d/2}} \exp\Big(- \frac{T^d}{2(2\pi)^d} \int_{\Pi^d} \frac{I(\lambda)}{f_\theta(\lambda)} \,\mathrm{d}\lambda \Big),$$

where

$$V = \exp\Big(\frac{1}{(2\pi)^d} \int_{\Pi^d} \log f_\theta(\lambda)\,\mathrm{d}\lambda \Big).$$

Taking logarithm, approximated ml-estimates of θ are obtained minimizing

$$\mathcal{L}(\theta) = \frac{1}{(2\pi)^d} \int_{\Pi^d} \Big[\log f_\theta(\lambda) + \frac{I(\lambda)}{f_\theta(\lambda)} \Big] \,\mathrm{d}\lambda, \tag{2.5}$$

where $I(\lambda)$ corresponds to the periodogram, defined by

$$I(\lambda) = \sum_{k\in\mathbb{Z}^d} C_k \exp(-i\langle\lambda, k\rangle),$$

with

$$C_k = \frac{1}{T^d} \sum_{t\in D_T} X_{t_1,\dots,t_d} X_{t_1+k_1,\dots,t_d+k_d}.$$

Proof. Suppose the process has a unilateral representation (2.4). The joint density function of the T^d residuals ε_t is

$$p(\varepsilon) = \frac{1}{(2\pi)^{T^d/2}} \exp\Big(-\frac{1}{2}\sum_{t\in D_T} \varepsilon_t^2\Big).$$

We can write

$$P(B_{t_1},\dots,B_{t_d})X_t = \exp(-a_{0,\dots,0}/2)Q(B_{t_1},\dots,B_{t_d})X_t = \varepsilon_t$$

and therefore $Q(B_{t_1},\dots,B_{t_d})X_t = \exp(a_{0,\dots,0}/2)\varepsilon_t = \varepsilon'_t$ where $\varepsilon'_t \sim \mathcal{N}(0,\exp(a_{0,\dots,0}))$. Therefore, we define the constant $A = \exp(-a_{0,\dots,0}/2)$. Neglecting boundary-effects, we obtain

$$p(\varepsilon) = \frac{A^{T^d}}{(2\pi)^{T^d/2}} \exp\Big(-\frac{1}{2}\sum_{t\in D_T} (P(B_{t_1},\dots,B_{t_d})X_t)^2\Big) \tag{2.6}$$

using (2.3) the argument of the exponential can be inverted as

$$-\frac{1}{2}\sum_{t\in D_T} \frac{1}{f_\theta(B_{t_1},\dots,B_{t_d})} X_t \cdot X_t \tag{2.7}$$

The assumption of unilateral representation implies that $f_\theta(\lambda)$ has a Fourier expansion, thus

$$\frac{1}{f_\theta(B_{t_1},\dots,B_{t_d})}X_t = \sum_{k\in\mathbb{Z}^d} c_{k_1,\dots,k_d} B_{t_1}^{k_1}\cdots B_{t_d}^{k_d} X_t = \sum_{k\in\mathbb{Z}^d} c_{k_1,\dots,k_d} X_{t_1+k_1,\dots,t_d+k_d} \tag{2.8}$$

plugging this in (2.7) and interchanging the order of summation, we obtain

$$-\frac{T^d}{2}\sum_{k\in\mathbb{Z}^d} c_{k_1,\dots,k_d}\Big(\frac{1}{T^d}\sum_{t\in D_T} X_{t_1,\dots,t_d} X_{t_1+k_1,\dots,t_d+k_d}\Big).$$

Neglecting edge-effects, the above expression is approximately

$$\approx -\frac{T^d}{2}\sum_{k\in\mathbb{Z}^d} c_k C_k. \tag{2.9}$$

Using (2.8)

$$\frac{1}{f_\theta(\lambda)} = \sum_{k\in\mathbb{Z}^d} c_k \exp(i\langle\lambda,k\rangle),$$

then

$$\frac{I(\lambda)}{f_\theta(\lambda)} = \sum_{k\in\mathbb{Z}^d}\sum_{k^*\in\mathbb{Z}^d} c_k C_{k^*} \exp(i\langle\lambda, k-k^*\rangle).$$

Integrating in Π^d we obtain

$$\int_{\Pi^d} \frac{I(\lambda)}{f_\theta(\lambda)}\,\mathrm{d}\lambda = (2\pi)^d \sum_{k\in\mathbb{Z}^d} c_k C_k.$$

Consequently, (2.9) is approximately

$$-\frac{T^d}{2(2\pi)^d}\int\limits_{\Pi^d}\frac{I(\lambda)}{f_\theta(\lambda)}\,\mathrm{d}\lambda.$$

Finally, taking logarithm and integrating in (2.1), we obtain in our case

$$a_{0,\dots,0}=\frac{1}{(2\pi)^d}\int\limits_{\Pi^d}\log f_\theta(\lambda)\,\mathrm{d}\lambda. \tag{2.10}$$

Replacing (2.10) in (2.6) and building $-\frac{1}{T^d}\log p(\varepsilon)$ we get (2.5) up to a constant. □

The periodogram involves a bias from the edge effects in the calculation of the covariance estimator. In order to get an efficient, consistent and asymptotic Gaussian estimator of θ, Dahlhaus and Künsch (1987) introduce data tapers. In the next section we use this idea and build the Whittle likelihood using a particular Szegö-type formula for the locally stationary spectral density. This is done through some technical lemmas which will be used in the chapter on asymptotic properties as well.

2.2 Whittle Estimator for Locally Stationary Random Fields

Let $X_{t,T}, t\in D_T$ be a zero mean locally stationary Gaussian random field from a parametric class $\Theta\subset\mathbb{R}^p$. We denote $X=(X_{1,T},\dots,X_{T^d,T})$ the vectorization of the random field. Its log-likelihood can be written as

$$-\frac{2l(\theta;X)}{T^d}=\log(2\pi)+\frac{1}{T^d}\log\det\Sigma_\theta+\frac{1}{T^d}X'\Sigma_\theta^{-1}X,\quad \theta\in\Theta,$$

where Σ_θ is the parametric covariance matrix of X. As the constant $\log(2\pi)$ does not contribute to find the minimum, we consider only the essential component, denoted by

$$\mathcal{L}_T^{(ex)}(\theta)=\frac{1}{T^d}[\log\det\Sigma_\theta+X'\Sigma_\theta^{-1}X], \tag{2.11}$$

where (ex) stands for exact Gaussian likelihood. The exact estimator will be denoted by

$$\widetilde{\theta}_T:=\arg\min_{\theta\in\Theta}\mathcal{L}_T^{(ex)}(\theta). \tag{2.12}$$

The calculation of the inverse and determinant in (2.11) implies typically $\mathcal{O}(T^{3d})$ operations which can be computationally burdensome for large random fields. The problem involved in the calculation of (2.12) within a reasonable time has been partially solved by using spectral methods. As shown in the previous section, the minimization of the Whittle likelihood involves the calculation of the periodogram which can be done efficiently by means of the Fast Fourier Transform algorithm, which typically involves only $\mathcal{O}(dT^d\log_2 T)$ operations.

These ideas suggest the use of a Whittle estimator, denoted by

$$\widehat{\theta}_T := \arg\min_{\theta\in\Theta} \mathcal{L}_T(\theta),$$

where $\mathcal{L}_T(\theta)$ corresponds to the Whittle approximation of the exact likelihood (2.11). In this section we build this Whittle approximation. The technical difficulties are higher given the local stationarity. Unfortunately, our approximation leads to a biased estimator. Though the bias vanishes asymptotically, this complicates the finding of an asymptotic Gaussian law as it will be discussed in Chapter 3.

We start by introducing notation, a set of assumptions and some technical lemmas. In order to maintain the discussion streamlined some of the proofs are relegated to the technical section (Section 2.3) at the end of this chapter.

Assumptions:

We define: $\nabla_i = \frac{\partial}{\partial\theta_i}$, $\nabla^2_{ij} = \frac{\partial^2}{\partial\theta_i\partial\theta_j}$ and so on, and

$$k_1(r) = \prod_{j=1}^{d} \frac{1}{(|r_j|+1)^3}, \quad r \in \mathbb{Z}^d. \tag{2.13}$$

In the following, the functions $A = A(u,\lambda)$ and $A_\theta = A_\theta(u,\lambda)$ are defined satisfying the conditions given in (1.3). The expression $\mathcal{L}_T^{(h)}(\theta)$ stands for the discretized version of the tapered Whittle likelihood. Further details are given later.

(A1) Let X_t, $t \in D_T$ be a realization of a locally stationary centered Gaussian random field with transfer function A^0 and fast decaying covariance $Cov(X_{t,T}, X_{s,T}) = \mathcal{O}((|t_i - s_i| + 1)^{-3})$, $i = 1,\dots,d$. We fit a class of locally stationary centered Gaussian process with transfer function A^0_θ, $\theta \in \Theta \subset \mathbb{R}^p$, Θ compact.

(A2) $\theta_0 = \arg\min_{\theta\in\Theta} \mathcal{L}^{(h)}(\theta)$ where $\mathcal{L}^{(h)}(\theta) = \lim_{T\to\infty} \mathcal{L}_T^{(h)}(\theta)$ exists, is unique and lies in the interior of Θ. This is also valid when the taper function h tends to one, *i.e.* a non tapered likelihood $\mathcal{L}(\theta)$.

(A3) The spectral densities $f(u,\lambda) = |A(u,\lambda)|^2$, $f_\theta(u,\lambda) = |A_\theta(u,\lambda)|^2$ are bounded from above and away from zero uniformly in θ, u and λ.

(A4) The Fourier coefficients $(\hat{f}_\theta)_n$ of $f_\theta(u,\lambda)$ are $\mathcal{O}(k_1(|n|))$ in frequency direction and uniformly in u and θ.

(A5) $A(u,\lambda)$ is differentiable with respect to u and λ with uniform continuous derivatives $\frac{\partial^2}{\partial u^2}\frac{\partial^2}{\partial\lambda^2}A(u,\lambda)$. $A_\theta(u,\lambda)$ is differentiable with respect to θ, u and λ with uniformly continuous derivatives $\nabla_{ijk}\frac{\partial^2}{\partial u^2}\frac{\partial^2}{\partial\lambda^2}A_\theta(u,\lambda)$.

(A6) Let $h(u)$ be a taper function such that $h(0) = h'(0) = 0$ and $\sup_u |h'(u)^2|$, $\sup_u h''(u) = \mathcal{O}(\rho^{-2})$, where ρ stands for the proportion of tapered data.

Henceforth the notation ϕ^{-1} stands for the reciprocal of the function ϕ, not for its inverse, unless the contrary is explicitly said. Considering the matrices $\Sigma_T(A)$ and $U_T(\phi)$ introduced in (1.12) and (1.8) respectively, we have the following lemmas.

Lemma 2.2.1. *If A and f fulfil assumptions (A3) and (A5), then the matrices $\Sigma_T(A)$, $\Sigma_T^{-1}(A)$, $U_T(f)$ and $U_T^{-1}(f)$ are bounded with respect to the operator norm for matrices*

(i) $\|\Sigma_T(A)\|_{op} \leq (2\pi)^d \sup_{(u,\lambda)\in[0,1]^d\times\Pi^d} |A(u,\lambda)|^2 + o(1).$

(ii) $\|\Sigma_T^{-1}(A)\|_{op} \leq (2\pi)^{-d} \sup_{(u,\lambda)\in[0,1]^d\times\Pi^d} |A(u,\lambda)|^{-2} + o(1).$

(iii) $\|U_T(f)\|_{op} \leq (2\pi)^d \sup_{(u,\lambda)\in[0,1]^d\times\Pi^d} f(u,\lambda) + o(1).$

(iv) $\|U_T^{-1}(f)\|_{op} \leq (2\pi)^{-d} \sup_{(u,\lambda)\in[0,1]^d\times\Pi^d} f^{-1}(u,\lambda) + o(1).$

Proof. A proof of this result can be found in Dahlhaus (1996), Lemma 4.4 for the univariate locally stationary time series case. The proof for locally stationary random fields is completely analogous and we do not give details here.

□

The next two lemmas form the core of the technical arguments for both, establishing the Whittle estimator and its asymptotic properties. Next, we use the functions A_k and ϕ_k which are analogous to A and ϕ.

Lemma 2.2.2. *Let A_k and ϕ_k, $k = 1, \ldots, n$ fulfil assumptions (A3), (A4) and (A5). Furthermore let*

$$\begin{aligned} C_k &= \Sigma_T(A_k), \qquad \psi_k(u,\lambda) = (2\pi)^d |A_k(u,\lambda)|^2. \\ C_k &= U_T(\phi_k), \qquad \psi_k(u,\lambda) = (2\pi)^d \phi_k(u,\lambda). \end{aligned}$$

Then

$$\frac{1}{T^d} \operatorname{tr}\Big[\prod_{k=1}^n C_k\Big] = \frac{1}{(2\pi)^d} \int_{[0,1]^d} \int_{\Pi^d} \prod_{k=1}^n \psi_k(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-1}). \tag{2.14}$$

Proof. See proof in Section 2.3 on page 31. □

Lemma 2.2.3. *Let A_k and ϕ_k, $k = 1, \ldots, n$ fulfil assumptions (A3), (A4) and (A5). Furthermore, let*

$$\begin{aligned} C_k &= \Sigma_T(A_k), \qquad \psi_k(u,\lambda) = (2\pi)^d |A_k(u,\lambda)|^2 \quad \textit{or} \\ C_k &= \Sigma_T^{-1}(A_k), \qquad \psi_k(u,\lambda) = (2\pi)^{-d} |A_k(u,\lambda)|^{-2} \quad \textit{or} \\ C_k &= U_T(\phi_k), \qquad \psi_k(u,\lambda) = (2\pi)^d \phi_k(u,\lambda) \quad \textit{or} \\ C_k &= U_T^{-1}(\phi_k), \qquad \psi_k(u,\lambda) = (2\pi)^{-d} \phi_k^{-1}(u,\lambda). \end{aligned}$$

Then, in each of these cases

$$\frac{1}{T^d}\operatorname{tr}\Big[\prod_{k=1}^{n} C_k\Big] = \frac{1}{(2\pi)^d}\int\limits_{[0,1]^d}\int\limits_{\Pi^d}\prod_{k=1}^{n}\psi_k(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-1}). \tag{2.15}$$

Proof. See proof in Section 2.3 on page 36. □

Proposition 2.2.1 (Szegö-type Formula)**.** *Suppose assumptions (A3), (A4) and (A5) hold, then*

$$\frac{1}{T^d}\log\det\Sigma_T(A) = \frac{1}{(2\pi)^d}\int\limits_{[0,1]^d}\int\limits_{\Pi^d}\log[(2\pi)^d f(u,\lambda)]\,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-1}) \tag{2.16}$$

Proof. Consider the matrix $U_T(\phi)$ introduced in (1.12), with $\phi(u,\lambda) = f(u,\lambda)$ and $f(u,\lambda) = |A(u,\lambda)|^2$. Using Lemma 2.3.1(i) below we get

$$\begin{aligned}\log\det U_T(f) - \log\det\Sigma_T(A) &= \log\det U_T(f)\Sigma_T^{-1}(A)\\ &\leq \operatorname{tr}(U_T(f)\Sigma_T^{-1}(A) - I).\end{aligned}$$

We divide by T^d and Lemma 2.2.3 yields

$$\frac{1}{T^d}\log\det\Sigma_T(A) = \frac{1}{T^d}\log\det U_T(f) + \mathcal{O}(T^{-1}). \tag{2.17}$$

Notice that $U_T(1)_{r,r} = (2\pi)^d$, $\log\det U_T(1) = T^d\log(2\pi)^d$, and let us consider the function f^x, $\forall x\in[0,1]$. From Lemma 2.3.1(ii) below

$$\begin{aligned}\frac{1}{T^d}\log\det U_T(f) &= \frac{1}{T^d}\int_0^1\frac{\partial}{\partial x}\log\det U_T(f^x)\,\mathrm{d}x + \log(2\pi)^d\\ &= \frac{1}{T^d}\int_0^1\operatorname{tr}(U_T^{-1}(f^x)\frac{\partial}{\partial x}U_T(f^x))\,\mathrm{d}x + \log(2\pi)^d.\end{aligned} \tag{2.18}$$

The dominated convergence theorem yields

$$\frac{\partial}{\partial x}U_T(f^x)_{r,s} = \frac{1}{(2\pi)^d}\int\limits_{\Pi^d}\exp(i\langle\lambda, r-s\rangle)f^x\log f\,\mathrm{d}\lambda = U_T(f^x\log f)_{r,s}. \tag{2.19}$$

Thus, plugging this into (2.18) and using again Lemma 2.2.3 we obtain

$$\begin{aligned}\frac{1}{T^d}\log\det U_T(f) &= \frac{1}{T^d}\int_0^1\operatorname{tr}(U_T^{-1}(f^x)U_T(f^x\log f))\,\mathrm{d}x + \log(2\pi)^d\\ &= \frac{1}{(2\pi)^d}\int\limits_{[0,1]^d}\int\limits_{\Pi^d}\log[(2\pi)^d f(u,\lambda)]\,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-1}).\end{aligned}$$

The result follows using the expression above in (2.17) □

The proof of Lemma 2.2.2 implies that the result above is uniform in θ, when $f(u,\lambda) = f_\theta(u,\lambda)$. Additionally, we need to find an approximation for $\Sigma_T^{-1}(A)$ easier to calculate. This is achieved in the next proposition.

Proposition 2.2.2. *Let $\phi(u,\lambda) = (2\pi)^{-2d} f^{-1}(u,\lambda)$ where (A3), (A4) and (A5) hold, then*

$$\frac{1}{T^d}\|\Sigma_T^{-1}(A) - U_T(\phi)\|_E^2 = \mathcal{O}(T^{-1}), \tag{2.20}$$

where $\|\cdot\|_E$ corresponds to the Euclidean norm (see Definition 2.3.1).

Proof. We make notation shorter by writing $\Sigma_T := \Sigma_T(A)$ and $U_T := U_T(\phi)$

$$\begin{aligned}
\|\Sigma_T^{-1} - U_T\|_E &= \|\Sigma_T^{-1/2}(I - \Sigma_T^{1/2}U_T\Sigma_T^{1/2})\Sigma_T^{-1/2}\|_E \\
&= \operatorname{tr}((I - \Sigma_T^{1/2}U_T\Sigma_T^{1/2})\Sigma_T^{-1}(I - \Sigma_T^{1/2}U_T\Sigma_T^{1/2})^*\Sigma_T^{-1*}) \\
&= \|(I - \Sigma_T^{1/2}U_T\Sigma_T^{1/2})\Sigma_T^{-1}\|_E^2 \\
&\leq \|I - \Sigma_T^{1/2}U_T\Sigma_T^{1/2}\|_E^2\|\Sigma_T^{-1}\|_{op}^2,
\end{aligned}$$

where $*$ denotes conjugate and $\|\cdot\|_{op}$ corresponds to the operator norm. Lemma 2.2.1 ensures $\|\Sigma_T^{-1}\|_{op}^2 = \mathcal{O}(1)$, therefore it is enough to analyse the first norm in the last inequality. After some algebra we obtain

$$\|I - \Sigma_T^{1/2}U_T\Sigma_T^{1/2}\|_E^2 = \operatorname{tr}(I - 2U_T\Sigma_T + U_T\Sigma_T U_T\Sigma_T).$$

Consequently

$$\frac{1}{T^d}\|\Sigma_T^{-1} - U_T\|_E^2 \leq (1 - \frac{2}{T^d}\operatorname{tr}(U_T\Sigma_T) + \frac{1}{T^d}\operatorname{tr}(U_T\Sigma_T U_T\Sigma_T))\mathcal{O}(1).$$

Applying Lemma 2.2.2 two times we get the desired result. □

Before presenting the deduction of the Whittle likelihood we need the next proposition.

Proposition 2.2.3. *Let $g : [0,1]^d \to \mathbb{R}$ be a Lipschitz continuous function with constant L, then*

$$\Big|\frac{1}{T^d}\sum_{t\in D_T} g(\tfrac{t-1/2}{T}) - \int_{[0,1]^d} g(u)\,\mathrm{d}u\Big| = \mathcal{O}(\tfrac{L}{T}), \tag{2.21}$$

where $D_T = \{1,\ldots,T\}^d$.

Proof. There exist a $\xi_t \in [\frac{t-1}{T}, \frac{t}{T}]$ such that the left-hand side of (2.21) can be written as

$$\begin{aligned}
\frac{1}{T^d}\Big|\sum_{t\in D_T}\Big\{g(\tfrac{t-1/2}{T})-g(\xi_t)\Big\}\Big| &\leq \frac{1}{T^d}\sum_{t\in D_T}|g(\tfrac{t-1/2}{T})-g(\xi_t)| \\
&\leq \frac{L}{T^d}\sum_{t\in D_T}\|\tfrac{t-1/2}{T}-\xi_t\| \\
&\leq \frac{L}{T^d}\sum_{t\in D_T}\frac{\sqrt{d}}{T} \\
&= \mathcal{O}(\tfrac{L}{T}).
\end{aligned}$$

□

Proposition 2.2.4 (Whittle Likelihood)**.** *Suppose assumptions (A1), (A3), (A4) and (A5) hold. Then, the Whittle-type approximation of the likelihood (2.11) is*

$$\mathcal{L}_T(\theta)=\frac{1}{(2\pi T)^d}\sum_{t\in D_T}\int_{\Pi^d}\Big\{\log[(2\pi)^d f_\theta(\tfrac{t-1/2}{T},\lambda)]+\frac{J_T(\frac{t-1/2}{T},\lambda)}{f_\theta(\frac{t-1/2}{T},\lambda)}\Big\}\,\mathrm{d}\lambda, \tag{2.22}$$

where

$$J_T\Big(\tfrac{t-1/2}{T},\lambda\Big)=\frac{1}{(2\pi)^d}\sum_{\substack{k\in D_T-D_T:\\ [t\pm k/2]\in D_T}}X_{[t+k/2],T}X_{[t-k/2],T}\exp(i\langle\lambda,k\rangle), \tag{2.23}$$

is called preperiodogram.

Proof. Given Proposition 2.2.3, the error involved in the Szegö-type formula when discretized has order $\mathcal{O}(T^{-1})$. Therefore, assuming a Gaussian model $\Sigma_\theta := \Sigma_T(A_\theta)$ for $\Sigma_T(A)$ we obtain

$$\begin{aligned}
\frac{1}{T^d}\log\det\Sigma_\theta &= \frac{1}{(2\pi)^d}\int_{[0,1]^d}\int_{\Pi^d}\log[(2\pi)^d f_\theta(u,\lambda)]\,\mathrm{d}\lambda\,\mathrm{d}u+\mathcal{O}(T^{-1}) \\
&= \frac{1}{(2\pi T)^d}\sum_{t\in D_T}\int_{\Pi^d}\log[(2\pi)^d f_\theta(\tfrac{t-1/2}{T},\lambda)]\,\mathrm{d}\lambda\,\mathrm{d}u \\
&\quad +\mathcal{O}(T^{-1}).
\end{aligned} \tag{2.24}$$

Proposition 2.2.2 implies the asymptotic equivalence between the matrices Σ_θ^{-1} and $U_T((2\pi)^{-2d}f_\theta^{-1})$. Thus, we can approximate $X'\Sigma_\theta^{-1}X$ by $X'U_T((2\pi)^{-2d}f_\theta^{-1})X$.

$$\begin{aligned}
\frac{1}{T^d}X'U_T((2\pi)^{-2d}f_\theta^{-1})X &= \frac{1}{(2\pi)^{2d}T^d}\sum_{r,s\in D_T}X_{r,T}X_{s,T} \\
&\quad\times\int_{\Pi^d}f_\theta^{-1}\Big(\tfrac{[(r+s)/2]-1/2}{T},\lambda\Big)\exp(i\langle\lambda,r-s\rangle)\,\mathrm{d}\lambda,
\end{aligned}$$

where $[\cdot]$ is the floor function. The change of variable $s \to t$, $r - s \to k$ implies

$$= \frac{1}{(2\pi)^{2d}T^d} \sum_{k\in D_T - D_T} \sum_{t\in (D_T - k)\cap D_T} \int_{\Pi^d} X_{t+k,T} X_{t,T} f_\theta^{-1}\Big(\tfrac{[t+k/2]-1/2}{T}, \lambda\Big) \exp(i\langle\lambda, k\rangle)\,\mathrm{d}\lambda.$$

Making a shift equal to k in the second summation we get rid of the dependency on k and the order of summation can be exchanged

$$= \frac{1}{(2\pi)^{2d}T^d} \sum_{t\in D_T} \sum_{\substack{k\in D_T - D_T:\\ [(t-k)\in D_T]}} \int_{\Pi^d} X_{t,T} X_{t-k,T} f_\theta^{-1}\Big(\tfrac{[t-k/2]-1/2}{T}, \lambda\Big) \exp(i\langle\lambda, k\rangle)\,\mathrm{d}\lambda.$$

Notice the elements involving t are indexed in the sequence $t-k$, $[t-k/2]$ and t, therefore, we can rearrange them in the sequence $[t-k/2]$, t and $[t+k/2]$ using the symmetry of the set $D_T - D_T$. This yields

$$\frac{1}{(2\pi)^{2d}T^d} \sum_{t\in D_T} \sum_{\substack{k\in D_T - D_T:\\ [t\pm k/2]\in D_T}} \int_{\Pi^d} X_{[t+k/2],T} X_{[t-k/2],T} f_\theta^{-1}\Big(\tfrac{t-1/2}{T}, \lambda\Big) \exp(i\langle\lambda, k\rangle)\,\mathrm{d}\lambda,$$

which can be written as

$$\frac{1}{(2\pi T)^d} \sum_{t\in D_T} \int_{\Pi^d} J_T\Big(\tfrac{t-1/2}{T}, \lambda\Big) f_\theta^{-1}\Big(\tfrac{t-1/2}{T}, \lambda\Big)\,\mathrm{d}\lambda$$

where

$$J_T\Big(\tfrac{t-1/2}{T}, \lambda\Big) = \frac{1}{(2\pi)^d} \sum_{\substack{k\in D_T - D_T:\\ [t\pm k/2]\in D_T}} X_{[t+k/2],T} X_{[t-k/2],T} \exp(i\langle\lambda, k\rangle). \tag{2.25}$$

In short we can write

$$\frac{1}{T^d} X' U_T X = \frac{1}{(2\pi T)^d} \sum_{t\in D_T} \int_{\Pi^d} \frac{J_T(\frac{t-1/2}{T}, \lambda)}{f_\theta(\frac{t-1/2}{T}, \lambda)}\,\mathrm{d}\lambda. \tag{2.26}$$

The result follows combining (2.24) and (2.26). □

Summarizing, we obtained the discretized version of the Whittle likelihood for locally stationary Gaussian random fields of dimension d.

$$\mathcal{L}_T(\theta) = \frac{1}{(2\pi T)^d} \sum_{t\in D_T} \int_{\Pi^d} \Big\{ \log[(2\pi)^d f_\theta(\tfrac{t-1/2}{T}, \lambda)] + \frac{J_T(\frac{t-1/2}{T}, \lambda)}{f_\theta(\frac{t-1/2}{T}, \lambda)} \Big\}\,\mathrm{d}\lambda. \tag{2.27}$$

We have assumed the true process to be Gaussian. We might ask what happens if the model fitted f_θ is not the right one. We can address this problem by considering the Kullback-Leibler information divergence.

Proposition 2.2.5. *Let $X = \{X_{t,T}, t \in D_T\}$ be a locally stationary Gaussian random field with density $g(X)$ and spectral density $f(u,\lambda) = |A(u,\lambda)|^2$. Furthermore, we will consider a parametric model with density $g_\theta(X)$ and spectral density $f_\theta(u,\lambda)$. Suppose additionally that the assumptions (A1), (A3), (A4) and (A5) hold. Then, the Kullback-Leibler information divergence corresponds to*

$$D(f_\theta, f) = \frac{1}{2}\mathcal{L}(\theta) + const.$$

where the constant does not depend on θ.

Proof. By definition

$$\begin{aligned}
D(f_\theta, f) &= \lim_{T\to\infty} \frac{1}{T^d} \mathrm{E}_g \log \frac{g}{g_\theta} \\
&= \lim_{T\to\infty} \frac{1}{T^d} \int\limits_{\mathbb{R}^d} \{\log g(X) - \log g_\theta(X)\} g(X)\,\mathrm{d}X \\
&= \lim_{T\to\infty} \frac{1}{2T^d} \int\limits_{\mathbb{R}^d} \{\log\det\Sigma_\theta - \log\det\Sigma(A)\} g(X)\,\mathrm{d}X \\
&\quad + \lim_{T\to\infty} \frac{1}{2T^d} \int\limits_{\mathbb{R}^d} \{X'\Sigma_\theta^{-1}X - X'\Sigma(A)X\} g(X)\,\mathrm{d}X \\
&= \lim_{T\to\infty} \frac{1}{2T^d} \{\log\det\Sigma_\theta - \log\det\Sigma(A)\} \\
&\quad + \frac{1}{2}\lim_{T\to\infty} \frac{1}{T^d} \{\mathrm{E}_g X'\Sigma_\theta^{-1}X - \mathrm{E}_g X'\Sigma^{-1}(A)X\}
\end{aligned}$$

Proposition 2.2.1 implies

$$\begin{aligned}
&= \frac{1}{2}\lim_{T\to\infty} \Big\{ \frac{1}{(2\pi)^d} \int\limits_{[0,1]^d} \int\limits_{\Pi^d} \log \frac{f_\theta(u,\lambda)}{f(u,\lambda)}\,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-1}) \Big\} \\
&\quad + \frac{1}{2}\lim_{T\to\infty} \frac{1}{T^d} \{\mathrm{tr}(\Sigma_\theta^{-1}\Sigma(A)) - \mathrm{tr}(\Sigma^{-1}(A)\Sigma(A))\}
\end{aligned}$$

by using Lemma 2.2.3 we obtain

$$\begin{aligned}
&= \frac{1}{2(2\pi)^d} \int\limits_{[0,1]^d} \int\limits_{\Pi^d} \Big\{ \log \frac{f_\theta(u,\lambda)}{f(u,\lambda)} + \frac{f(u,\lambda)}{f_\theta(u,\lambda)} - 1 \Big\}\,\mathrm{d}\lambda\,\mathrm{d}u \\
&= \frac{1}{2(2\pi)^d} \int\limits_{[0,1]^d} \int\limits_{\Pi^d} \log[(2\pi)^d f_\theta(u,\lambda)] + \frac{f(u,\lambda)}{f_\theta(u,\lambda)}\,\mathrm{d}\lambda\,\mathrm{d}u \\
&\quad - \frac{1}{2(2\pi)^d} \int\limits_{[0,1]^d} \int\limits_{\Pi^d} \{\log[(2\pi)^d f(u,\lambda)] + 1\}\,\mathrm{d}\lambda\,\mathrm{d}u \\
&= \frac{1}{2}\mathcal{L}(\theta) + const.
\end{aligned}$$

□

The parameter θ_0, which minimizes $\mathcal{L}(\theta)$ will also minimize $D(f_\theta, f)$ and hence $f_{\theta_0}(u,\lambda)$ is the best approximation of the true $f(u,\lambda)$ in the sense of the Kullback-Leibler information divergence. In Chapter 4, we present a simulation where the true parameter function is given by an exponential function of location. Assuming an unknown functional form of the parameter function, we use polynomials of different order, *i.e.* we only have an approximating model of the true one (see Figures 4.2 - 4.5). It is worth noticing how the polynomials approximate quite well the exponential as we increase the order, reflecting what the result above tells us.

A non Gaussian case is much more technical and it is not considered here. Nevertheless, in Chapter 4 we consider briefly a simulation with uniform innovations, giving results quite similar to, in some cases even better than, the Gaussian case.

We return now to the result proved in Proposition 2.2.4. Because of rounding, *i.e.* the use of floor functions, we have artificially introduced a bias. Next, we study this and other sources of bias and consider alternatives to reduce it.

Let $S_T(\theta)$ be the score function of the Whittle likelihood $\mathcal{L}_T(\theta)$, *i.e.* the gradient w.r.t. the parameter θ. Departures of its expectation from zero imply that the estimator is biased (see Proposition 2.3.1). Below we calculate the expectation of the score function and identify the sources of bias. In order to avoid cumbersome notation, the variable θ will represent θ_i for $i = 1, \ldots, p$.

Proposition 2.2.6. *Under conditions of Proposition 2.2.4 and considering $f_\theta(u,\lambda) = |A_\theta(u,\lambda)|^2$ as the true model, it holds that*

$$E_\theta[S_T(\theta)] = \mathcal{O}(T^{-1}).$$

Proof.

$$\begin{aligned}
\mathrm{E}_\theta[S_T(\theta)] &= \mathrm{E}_\theta\Big[\frac{\nabla_\theta}{(2\pi T)^d}\sum_{t\in D_T}\int_{\Pi^d}\Big\{\log[(2\pi)^d f_\theta(\tfrac{t-1/2}{T},\lambda)] + \frac{J_T(\frac{t-1/2}{T},\lambda)}{f_\theta(\frac{t-1/2}{T},\lambda)}\Big\}\,\mathrm{d}\lambda\Big]\\
&= \frac{1}{(2\pi T)^d}\,\mathrm{tr}\left[U_T\Big(\tfrac{\nabla_\theta f_\theta}{f_\theta}\Big)\right]\\
&\quad + \frac{1}{(2\pi T)^d}\mathrm{E}_\theta\Big[\sum_{t\in D_T}\int_{\Pi^d} J_T(\tfrac{t-1/2}{T},\lambda)\nabla_\theta f_\theta^{-1}(\tfrac{t-1/2}{T},\lambda)\,\mathrm{d}\lambda\Big].
\end{aligned}$$

The second term can be written as

$$\begin{aligned}
&\frac{1}{(2\pi T)^d}\sum_{t\in D_T}\int_{\Pi^d}\frac{1}{(2\pi)^d}\sum_k \mathrm{E}_\theta[X_{[t+k/2],T}X_{[t-k/2],T}]\exp(i\langle\lambda,k\rangle)\nabla_\theta f_\theta^{-1}(\tfrac{t-1/2}{T},\lambda)\,\mathrm{d}\lambda\\
=\;& \frac{1}{(2\pi T)^d}\sum_{r,s\in D_T}\int_{\Pi^d}\frac{1}{(2\pi)^d}\nabla_\theta f_\theta^{-1}(\tfrac{1}{T}[\tfrac{r+s}{2}] - \tfrac{1}{2T},\lambda)\mathrm{E}_\theta[X_{s,T}X_{r,T}]\exp(i\langle\lambda,r-s\rangle)\,\mathrm{d}\lambda
\end{aligned}$$

$$= \frac{1}{(2\pi)^{2d}T^d} \sum_{r,s\in D_T} \int_{\Pi^d} \nabla_\theta f_\theta^{-1}(\tfrac{1}{T}[\tfrac{r+s}{2}] - \tfrac{1}{2T}, \lambda_1) \exp(i\langle\lambda_1, r-s\rangle)\, d\lambda_1$$
$$\times \int_{\Pi^d} A_\theta(\tfrac{s-1/2}{T}, \lambda_2)\overline{A_\theta(\tfrac{r-1/2}{T}, \lambda_2)} \exp(i\langle\lambda_2, s-r\rangle)\, d\lambda_2$$
$$= \frac{1}{(2\pi T)^d} \sum_{r,s\in D_T} U_T((2\pi)^{-d}\nabla_\theta f_\theta^{-1})_{r,s}\Sigma_T(A_\theta)_{s,r}$$
$$= \frac{1}{(2\pi T)^d} \operatorname{tr}\Big[U_T((2\pi)^{-d}\nabla_\theta f_\theta^{-1})\Sigma_T(A_\theta)\Big].$$

Therefore, the expectation of the score function takes the form

$$\mathrm{E}_\theta[S_T(\theta)] = \frac{1}{(2\pi)^d}\Big\{\frac{1}{T^d}\operatorname{tr}\Big[U_T(\tfrac{\nabla_\theta f_\theta}{f_\theta})\Big] + \frac{1}{T^d}\operatorname{tr}\Big[U_T((2\pi)^{-d}\nabla_\theta f_\theta^{-1})\Sigma_T(A_\theta)\Big]\Big\}. \quad (2.28)$$

The result follows applying Lemma 2.2.2 to the two right-hand side terms in (2.28). □

Proposition 2.3.1 (see Section 2.3) implies that our Whittle estimator has a bias of order $\mathcal{O}(T^{-1})$. We distinguish at least three sources: 1) The skewed definition of the preperiodogram, 2) Non-stationarity and 3) Edge-effects.

1) Skewed preperiodogram:

From equations (1.9) and (1.10), the natural approximation matrix of $\Sigma_T(A_\theta)$ corresponds to

$$\widehat{U}_T(f)_{r,s} = \int_{\Pi^d} f_\theta(\tfrac{r+s-1}{2T}, \lambda) \exp(i\langle\lambda, r-s\rangle)\, d\lambda, \quad (2.29)$$

where $f_\theta(u,\lambda) = |A_\theta(u,\lambda)|^2$. The reason not to use (2.29) but $U_T(f)$ lies in that the resulting formula for the preperiodogram is closely related to the Wigner-Wille spectrum, and therefore it can be interpreted as a natural generalization of the spectrum for non-stationary processes (see Martin and Flandrin (1985)). This nice relationship is not possible using (2.29), in fact the likelihood approximation in this case turns out to be

$$\widehat{\mathcal{L}}_T(\theta) = \frac{1}{(2\pi T)^d}\Big\{\operatorname{tr}\Big[\widehat{U}_T(\log[(2\pi)^d f_\theta])\Big] + X'\widehat{U}_T((2\pi)^{-d} f_\theta^{-1})X\Big\}, \quad (2.30)$$

where technically, the second term cannot be written in terms of a preperiodogram.

However, the use of U_T introduces a bias due to the floor functions involved. To overcome this problem, we can modify slightly the preperiodogram in the following form

$$J_T(\tfrac{t-1/2}{T}, \lambda) := \frac{1}{(2\pi)^d}\sum_k \Big(\tfrac{X_{[t-k/2],T}X_{[t+k/2],T} + X_{[t-k/2]^*,T}X_{[t+k/2]^*,T}}{2}\Big) \quad (2.31)$$
$$\times \exp(i\langle\lambda, k\rangle),$$

where $[\cdot]^*$ denotes the ceiling function. In order to verify the magnitude of the bias reduction we need to analyse the difference between the score expectations, using $\widehat{\mathcal{L}}_T(\theta)$ and the likelihood $\mathcal{L}_T(\theta)$ derived from using (2.31).

Note that the first term in brackets in (2.30) involves the elements of the diagonal of $\widehat{U}_T$ and therefore no difference exist when compared with U_T, hence we stick to the difference of the second terms.

Using (2.31) it is straightforward to verify that

$$\frac{1}{(2\pi T)^d}\sum_{t\in D_T}\int_{\Pi^d}\frac{J_T(\frac{t-1/2}{T},\lambda)}{f_\theta(\frac{t-1/2}{T},\lambda)}\,\mathrm{d}\lambda = \frac{1}{(2\pi)^{2d}T^d}\sum_{r,s\in D_T}X_{r,T}X_{s,T}$$
$$\times\int_{\Pi^d}\frac{1}{2}\Big(f_\theta^{-1}(\tfrac{1}{T}[\tfrac{r+s}{2}]-\tfrac{1}{2T},\lambda)+f_\theta^{-1}(\tfrac{1}{T}[\tfrac{r+s}{2}]^*-\tfrac{1}{2T},\lambda)\Big)\exp(i\langle\lambda,r-s\rangle)\,\mathrm{d}\lambda.$$

For the sake of simplicity, let us denote this expression as

$$\tfrac{1}{(2\pi T)^d}X'\widetilde{U}_T((2\pi)^{-d}f_\theta^{-1})X. \tag{2.32}$$

Thus, using same arguments as above

$$\frac{1}{(2\pi T)^d}\Big|\mathrm{E}_\theta\nabla_\theta\Big[X'\widehat{U}_T((2\pi)^{-d}f_\theta^{-1})X - X'\widetilde{U}_T((2\pi)^{-d}f_\theta^{-1})X\Big]\Big|$$
$$= \frac{1}{(2\pi)^{2d}T^d}\Big|\sum_{r,s\in D_T}\int_{\Pi^d}\nabla_\theta\Big\{f_\theta^{-1}(\tfrac{r+s-1}{2T},\lambda_1)-\frac{1}{2}(f_\theta^{-1}(\tfrac{1}{T}[\tfrac{r+s}{2}]-\tfrac{1}{2T},\lambda_1)$$
$$+f_\theta^{-1}(\tfrac{1}{T}[\tfrac{r+s}{2}]^*-\tfrac{1}{2T},\lambda_1))\Big\}\exp(i\langle\lambda_1,r-s\rangle)\,\mathrm{d}\lambda_1$$
$$\times\int_{\Pi^d}A_\theta(\tfrac{s-1/2}{T},\lambda_2)\overline{A_\theta(\tfrac{r-1/2}{T},\lambda_2)}\exp(i\langle\lambda_2,s-r\rangle)\,\mathrm{d}\lambda_2\Big|.$$

Due to the Lipschitz continuity of $\nabla_\theta f_\theta(u,\lambda)$ in u we can make the following approximation for each θ

$$\nabla_\theta f_\theta^{-1}(\tfrac{1}{T}[\tfrac{r+s}{2}]-\tfrac{1}{2T},\lambda_1)=\nabla_\theta f_\theta^{-1}(\tfrac{r+s-1}{2T},\lambda_1)+\vec{v}_1\cdot\tfrac{L}{T}\mathcal{O}(\|\tfrac{r+s}{2T}-\tfrac{1}{T}[\tfrac{r+s}{2}]\|)$$

where $\vec{v}_1$ is a unit vector and L is the Lipschitz constant. The same can be done around $\frac{1}{T}[\frac{r+s}{2}]^*-\frac{1}{2T}$ with a given $\vec{v}_2$. Thus, the term

$$\nabla_\theta f_\theta^{-1}(\tfrac{r+s-1}{2T},\lambda_1)-\frac{1}{2}\Big(\nabla_\theta f_\theta^{-1}(\tfrac{1}{T}[\tfrac{r+s}{2}]-\tfrac{1}{2T},\lambda_1)+\nabla_\theta f_\theta^{-1}(\tfrac{1}{T}[\tfrac{r+s}{2}]^*-\tfrac{1}{2T},\lambda_1)\Big)$$

can be approximated as

$$-\frac{L}{2T}\Big\{\vec{v}_1\mathcal{O}(\|\tfrac{r+s}{2T}-\tfrac{1}{T}[\tfrac{r+s}{2}]\|)+\vec{v}_2\mathcal{O}(\|\tfrac{r+s}{2T}-\tfrac{1}{T}[\tfrac{r+s}{2}]^*\|)\Big\}=\mathcal{O}(T^{-2}).$$

Obviously the estimator based on $\widehat{U}_T(f)$ has no bias associated to floor functions in the preperiodogram, thus the difference of bias of the score functions just calculated implies a reduction of bias from $\mathcal{O}(T^{-1})$ to $\mathcal{O}(T^{-2})$ by incorporating the local average

$$\frac{1}{2}(X_{[t-k/2],T}X_{[t+k/2],T} + X_{[t-k/2]^*,T}X_{[t+k/2]^*,T}). \tag{2.33}$$

As pointed out in Dahlhaus (2000) while the periodogram corresponds to the Fourier transform of the covariance estimator of lag k over the whole field, the preperiodogram $J_T(\frac{t-1/2}{T}, \lambda)$ uses (2.33) as some kind of local estimator of the covariance of lag k at point $t - 1/2$.

2) Non-stationarity:

The second source of bias comes from the approximation $A^0_{t,T}(\lambda) \approx A(\frac{t-1/2}{T}, \lambda)$ where the error is at most $\mathcal{O}(T^{-1})$ (see (1.3)). Assuming equality in this approximation would imply leaving out some important models as those used in Chapter 4 from our considerations, and therefore this bias cannot be avoided. However, if it is possible to get a second order approximation $B(\frac{t-1/2}{T}, \lambda)$, i.e.

$$A^0_{t,T}(\lambda) \approx A(\tfrac{t-1/2}{T}, \lambda) + \frac{1}{T}B(\tfrac{t-1/2}{T}, \lambda),$$

then it would be possible to reduce the bias to an order $\mathcal{O}(T^{-2})$. Let us look closer how.

The contribution to this bias will be given through the matrix $\Sigma_T(A)$ by the approximation of the transfer function A. This contribution to the overall bias arises in the term containing $\Sigma_T(A)$ in equation (2.28). To measure how big it is, notice that Lemma 2.2.2 implies

$$\begin{aligned}&\frac{1}{T^d}\big|\operatorname{tr}(\widetilde{U}_T((2\pi)^{-d}\nabla_\theta f_\theta^{-1})\Sigma_T(A_\theta)) - \operatorname{tr}(\widetilde{U}_T((2\pi)^{-d}\nabla_\theta f_\theta^{-1})\widetilde{\Sigma}_T(A_\theta))\big| \\ &\quad = \frac{2}{T}\int_{[0,1]^d}\int_{\Pi^d} Re(A(u,\lambda)\overline{B(u,\lambda)})\nabla_\theta f_\theta^{-1}(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-2}),\end{aligned}$$

where $A_\theta(u,\lambda) = A(u,\lambda) + \frac{1}{T}B(u,\lambda) + \mathcal{O}(T^{-2})$ and $A_\theta(u,\lambda) = A(u,\lambda)$ correspond to the transfer functions associated to $\Sigma_T(A_\theta)$ and $\widetilde{\Sigma}_T(A_\theta)$ respectively, Re corresponds to the real part and $\widetilde{U}_T$ corresponds to the matrix used in (2.32) which incorporates the bias correction for skewness. Repeating the calculations but considering our second order approximation $A_\theta(u,\lambda) = A(u,\lambda) + \frac{1}{T}B(u,\lambda)$ associated to the matrix $\widetilde{\Sigma}_T(A_\theta)$, the result turns out to be

$$\begin{aligned}&\frac{1}{T^d}\big|\operatorname{tr}(\widetilde{U}_T((2\pi)^{-d}\nabla_\theta f_\theta^{-1})\Sigma_T(A_\theta)) - \operatorname{tr}(\widetilde{U}_T((2\pi)^{-d}\nabla_\theta f_\theta^{-1})\widetilde{\Sigma}_T(A_\theta))\big| \\ &\quad = \mathcal{O}(T^{-2})\int_{[0,1]^d}\int_{\Pi^d} 2Re(A(u,\lambda))\nabla_\theta f_\theta^{-1}(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u,\end{aligned}$$

which shows our assertion.

3) Edge-effects:

Close to the edges, the preperiodogram (2.23) involves few observations in comparison to, for instance, the center of the field, causing an important bias. In one dimension, the edge effect vanishes as the number of observations increases. In higher dimensions the proportion of points at the edges with respect to the whole field has order $\mathcal{O}(T^{-1})$. This implies a biased preperiodogram with bias of the same order. Below we use ideas of Dahlhaus and Künsch (1987) by introducing taper functions to overcome this problem.

Definition 2.2.1 (Taper function)**.** *A twice differentiable function* $\tilde{h}_\rho : [0,1] \to [0,1]$ *is called* taper *of proportion* ρ *if* $\tilde{h}_\rho(0) = \tilde{h}_\rho(1) = \tilde{h}'_\rho(0) = \tilde{h}'_\rho(1) = 0$*,* $\tilde{h}_\rho(x) = 1$ *for* $x \in [\rho/2, 1-\rho/2]$ *and monotone on* $[0, 1/2]$ *and* $[1/2, 1]$

In dimension d, the standardized $h_\rho(u)$ taper is, then, defined as

$$h_\rho(u) = \Big(\int_0^1 \tilde{h}^2_\rho(x)\,\mathrm{d}x \Big)^{-d/2} \prod_{j=1}^{d} \tilde{h}_\rho(u_j).$$

One common example of a taper function that will be used in Chapter 4 is the Tukey-Hanning taper defined as follows: Let $\tilde{h}_c(u) = \frac{1}{2}(1-\cos(\pi u))$,

$$\tilde{h}_\rho(u) = \begin{cases} \tilde{h}_c(2u/\rho) & 0 \le u < \rho/2 \\ 1 & \rho/2 \le u \le 1/2 \\ \tilde{h}_\rho(1-u) & 1/2 < u \le 1 \end{cases} \tag{2.34}$$

This case exemplify how a taper function downweighs the data at the edges, making the data contribution from this source less important.

Remark 2.2.1. *A couple of results to be used later are*

1) *Using Cauchy-Schwarz inequality it can be easily shown that*

$$\int_0^1 h_\rho(u)\,\mathrm{d}u \le 1 \le \int_0^1 h^4_\rho(u)\,\mathrm{d}u, \tag{2.35}$$

2) *For a constant* $k \in (0,1)$

$$\int_{[0,1]} \tilde{h}_\rho(u)\,\mathrm{d}u = 1 - (1-k)\rho.$$

Consequently, the normalizing term is approximately

$$\left(\int_0^1 \tilde{h}_\rho^2(u)\,\mathrm{d}u\right)^{-d/2} \approx 1 + \frac{d}{2}(1-k)\rho,$$

which implies

$$\int_{[0,1]^d} h_\rho(u)\,\mathrm{d}u \approx (1 - \frac{d}{2}(1-k)\rho)(1 + \frac{d}{2}(1+K)\rho) = 1 + \mathcal{O}(\rho), \tag{2.36}$$

with $K > k$.

From now on, we drop the index ρ in $h_\rho(u)$ to shorten notation. Considering the data $X_{t,T},\ t \in D_T$, its tapered version will be denoted as

$$X_{t,T}^{(h)} = \left(\frac{1}{T^d}\sum_{t\in D_T}\prod_{i=1}^d \tilde{h}_\rho^2\left(\tfrac{t_i-1/2}{T}\right)\right)^{-1/2}\left(\prod_{i=1}^d \tilde{h}_\rho\left(\tfrac{t_i-1/2}{T}\right)\right)X_{t,T} = h(\tfrac{t-1/2}{T})X_{t,T}.$$

Replacing $X_{t,T}$ in (2.25) by its tapered version $X_{t,T}^{(h)}$ we obtain the tapered preperiodogram $J_T^{(h)}$. Notice that when $\rho = 0$ we recover the classical non-tapered case.

In the following proposition, we modify the expectation of the score function by introducing a taper. A bias reduction to an order $\mathcal{O}(T^{-2})$ is proved. To avoid additional bias sources (skew preperiodogram), we use $\widehat{U}_T$ as an approximating matrix. The aim is measuring only the source that comes up from edge-effects.

Proposition 2.2.7. *Let $f_\theta(u,\lambda)$ fulfil (A3), (A4), (A5) and $h(u)$ fulfil (A6). Furthermore, we assume $\rho = \mathcal{O}(T^{-\beta})$, with a suitable $\beta \geq 0$ then*

$$\frac{1}{T^d}\left|\operatorname{tr}\left[\widehat{U}_T(h^2\tfrac{\nabla_\theta f_\theta}{f_\theta})\right] + \operatorname{tr}\left[\widehat{U}_T((2\pi)^{-d}\nabla_\theta f_\theta^{-1})\widehat{U}_T(h^2 f_\theta)\right]\right| = \mathcal{O}(T^{-2}).$$

Proof. For a shorter notation we abbreviate: $g(u,\lambda) = (2\pi)^{-d}\nabla_\theta f_\theta^{-1}(u,\lambda)$ and $f(u,\lambda) = f_\theta(u,\lambda) = |A_\theta(u,\lambda)|^2$ and consequently $f_n(u)$ and $g_n(u)$ denote the n-th Fourier coefficient of $f(u,\lambda)$ and $g(u,\lambda)$, respectively.

$$\operatorname{tr}(\widehat{U}_T(h^2 f)\widehat{U}_T(g)) = \sum_{r,s\in D_T} h^2(\tfrac{r+s-1}{2T}) f_{r-s}(\tfrac{r+s-1}{2T})\overline{g_{r-s}(\tfrac{r+s-1}{2T})}.$$

We use the following change of variable: $n = r-s$ and $m = (r+s)/2$. Each component of r and s are integers. This must be reflected on the domain of n and m. After some calculations, we can summarize this with the conditions $m \equiv \pm\frac{n}{2}$ mod 1. Notice that each component of r and s lies between 1 and T, hence there exist a second condition for n and m, and this is $1+\frac{|n|}{2} \leq m \leq T-\frac{|n|}{2}$, which is valid component-wise. The last summation can be written as

$$\sum_{n\in D_T-D_T} \sum_{\substack{1+\frac{|n|}{2}\leq m\leq T-\frac{|n|}{2} \\ m\equiv\pm\frac{n}{2} \bmod 1}} h^2(\tfrac{m}{T}-\tfrac{1}{2T}) f_n(\tfrac{m}{T}-\tfrac{1}{2T}) \overline{g_n(\tfrac{m}{T}-\tfrac{1}{2T})}. \tag{2.37}$$

Shifting and rearranging terms, the summation is

$$T^d \Big(\sum_{n\in D_T-D_T} \frac{1}{T^d} \sum_{\substack{1\leq m\leq T-|n| \\ m\equiv\pm\frac{n}{2} \bmod 1}} h^2(\tfrac{|n|}{2T}-\tfrac{1}{2T}+(1-\tfrac{|n|}{T})\tfrac{m}{T-|n|}) f_n(\tfrac{|n|}{2T}-\tfrac{1}{2T}+(1-\tfrac{|n|}{T})\tfrac{m}{T-|n|})$$
$$\times\, g_{-n}(\tfrac{|n|}{2T}-\tfrac{1}{2T}+(1-\tfrac{|n|}{T})\tfrac{m}{T-|n|})\Big).$$

For the sake of ease of notation, the indices in the inner summation represent d sums (each component of the vector m). Notice that this is a Riemann sum. The domain of the outer summation can be split into two sets, namely

$$\begin{aligned} C_1 &: \{n\in D_T-D_T : |n_i|\geq \delta T^\alpha, \text{for at least one } i=1,\dots,d,\ \delta>0,\ 0<\alpha<1\} \\ C_2 &: \{n\in D_T-D_T : |n_i|<\delta T^\alpha, \forall i\}, \end{aligned}$$

then the last sum can be written as

$$T^d \Big(\sum_{n\in C_1} \sum_{\substack{1\leq m\leq T-|n| \\ m\equiv\pm\frac{n}{2} \bmod 1}} b_n(\tfrac{m}{T}+\tfrac{|n|}{2T}-\tfrac{1}{2T})\frac{1}{T^d} + \sum_{n\in C_2} \sum_{\substack{1\leq m\leq T-|n| \\ m\equiv\pm\frac{n}{2} \bmod 1}} b_n(\tfrac{m}{T}+\tfrac{|n|}{2T}-\tfrac{1}{2T})\frac{1}{T^d}\Big),$$

where we have abbreviated the function $h^2(\cdot)f_n(\cdot)g_{-n}(\cdot)$ by $b_n(\cdot)$. We analyse both summations separately. The Proposition 2.3.2 and the remark after imply for the first summation

$$\sum_{n\in C_1} \sum_{\substack{1\leq m\leq T-|n| \\ m\equiv\pm\frac{n}{2} \bmod 1}} b_n(\tfrac{m}{T}+\tfrac{|n|}{2T}-\tfrac{1}{2T})\frac{1}{T^d} = \sum_{n\in C_1} \Bigg\{ \int\limits_{[\frac{|n|}{2T}-\frac{1}{2T}, 1-\frac{|n|}{2T}-\frac{1}{2T}]} b_n(u)\,\mathrm{d}u \tag{2.38}$$
$$+\mathcal{O}(T^{-1}) \int\limits_{[\frac{|n|}{2T}-\frac{1}{2T}, 1-\frac{|n|}{2T}-\frac{1}{2T}]} \sum_{i=1}^d \frac{\partial b_n}{\partial u_i}\,\mathrm{d}u_1\cdots\mathrm{d}u_d + \mathcal{O}(\frac{L_n}{T^2}) \Bigg\}.$$

Since $L_n = \frac{2M}{\sqrt{d}} k_1(n)$ for a $M>0$,

$$\sum_{n\in C_1} \mathcal{O}(\frac{k_1(n)}{T^2}) = \mathcal{O}(T^{-2}).$$

With a bit more effort, this rate might be improved, but it is not necessary. The assumption over the rate of decay of $f_n(u)$ implies that $b_n(u)=\mathcal{O}(k_1^2(n))$ uniformly in u, thus

$$\mathcal{O}(T^{-1}) \sum_{n\in C_1} \mathcal{O}(k_1^2(n)) \approx C\mathcal{O}(T^{-1}) \sum_{n\in C_1} \prod_{j=1}^d \frac{1}{(1+|n_j|)^6}.$$

Without loss of generality we may consider only the points $n \in D_T - D_T$ such that $|n_1| \geq \delta T^\alpha$. Then the right term of the last summation turns out to be

$$\begin{aligned}
&= dC\mathcal{O}(T^{-1}) \sum_{|n_1|\geq \delta T^\alpha} \sum_{|n^{(1)}|\in D_T - D_T} \prod_{j=1}^{d} \frac{1}{(1+|n_j|)^6} \\
&= dC\mathcal{O}(T^{-1}) \sum_{|n_1|\geq \delta T^\alpha} \frac{1}{(1+|n_1|)^6} \sum_{|n^{(1)}|\in D_T - D_T} \prod_{j=2}^{d} \frac{1}{(1+|n_j|)^6} \\
&\leq dC\mathcal{O}(T^{-1}) \cdot T \cdot \frac{1}{(1+\delta T^\alpha)^6} \cdot \mathcal{O}(1) \\
&= \mathcal{O}(T^{-6\alpha}),
\end{aligned}$$

where $n^{(1)}$ corresponds to n with the first component being omitted. If $1/3 \leq \alpha \leq 1$, then the desired rate (at least -2) is achieved. The first summation turns out to be

$$\sum_{n\in C_1} \sum_{\substack{1\leq m\leq T-|n| \\ m\equiv \pm\frac{n}{2} \bmod 1}} b_n(\tfrac{m}{T} + \tfrac{|n|}{2T} - \tfrac{1}{2T}) \frac{1}{T^d} = \sum_{n\in C_1} \int\limits_{[\frac{|n|}{2T} - \frac{1}{2T}, 1 - \frac{|n|}{2T} - \frac{1}{2T}]} b_n(u)\, \mathrm{d}u + \mathcal{O}(T^{-2}). \quad (2.39)$$

The arguments needed in order to bound the summation over C_2 are similar to the previous case. Let $\Omega_i = [\frac{|n^{(i)}|}{2T} - \frac{1}{2T}, 1 - \frac{|n^{(i)}|}{2T} - \frac{1}{2T}]$, $i = 0, \ldots, d$, and Ω_0 be the original domain. Without loss of generality we obtain

$$\int\limits_{\Omega_0} \frac{\partial b_n}{\partial u_1}\, \mathrm{d}u = \int\limits_{\Omega_1} \{b_n(1 - \tfrac{|n|}{2T} - \tfrac{1}{2T}, u_2, \ldots, u_d) - b_n(\tfrac{|n|}{2T} - \tfrac{1}{2T}, u_2, \ldots, u_d)\}\, \mathrm{d}u_2 \cdots \mathrm{d}u_d. \quad (2.40)$$

Since $|f_n(u) g_{-n}(u)| = \mathcal{O}(k_1^2(n))$, (2.40) is bounded by

$$D\{\int\limits_{\Omega_1} h^2(1 - \tfrac{|n|}{2T} - \tfrac{1}{2T}, u_2, \ldots, u_d)\, \mathrm{d}u_2 \cdots \mathrm{d}u_d + \int\limits_{\Omega_1} h^2(\tfrac{|n|}{2T} - \tfrac{1}{2T}, u_2, \ldots, u_d)\, \mathrm{d}u_2 \cdots \mathrm{d}u_d\}, \quad (2.41)$$

with $D > 0$. The second integral can be decomposed as

$$h^2(\tfrac{|n_1|}{2T} - \tfrac{1}{2T}) \underbrace{\int\limits_{\Omega_1} h^2(u_2, \ldots, u_d)\, \mathrm{d}u_2 \ldots \mathrm{d}u_d}_{\mathcal{O}(1)}.$$

Since $|n_1| < \delta T^\alpha$, $\frac{|n_1|}{2T} - \frac{1}{2T} = \mathcal{O}(T^{\alpha-1})$. A Taylor expansion of $h(u)$ about $u = 0$ yields $h(u) \approx h''(u)u^2/2$. From assumption (A6), $h''(u) = \mathcal{O}(\rho^{-2})$ and hence $h(u) = \mathcal{O}(\rho^{-2}u^2)$. We have assumed that $\rho = \mathcal{O}(T^{-\beta})$ obtaining

$$h^2(\tfrac{|n_1|}{2T} - \tfrac{1}{2T}) = \mathcal{O}(T^{4(\alpha+\beta-1)}), \quad (2.42)$$

the same is valid for the first integral in (2.41). In order to fulfil the minimum rate of -2 the exponent must fulfil $\alpha + \beta \leq 1/2$. Since $1/3 \leq \alpha \leq 1$ we choose $\alpha = 1/3$ and $\beta = 1/6$. Thus, we obtain

$$\sum_{n \in C_2} \sum_{\substack{1 \leq m \leq T - |n| \\ m \equiv \pm \frac{n}{2} \bmod 1}} b_n\left(\tfrac{m}{T} + \tfrac{|n|}{2T} - \tfrac{1}{2T}\right) \frac{1}{T^d} = \sum_{n \in C_2} \int_{[\frac{|n|}{2T} - \frac{1}{2T}, 1 - \frac{|n|}{2T} - \frac{1}{2T}]} b_n(u)\, \mathrm{d}u + \mathcal{O}(T^{-2}). \quad (2.43)$$

The equations (2.39) and (2.43) imply that (2.37) is equivalent to

$$T^d \Big[\sum_{n \in D_T - D_T} \int_{[\frac{|n|}{2T}, 1 - \frac{|n|}{2T}]} h^2(u) f_n(u) g_{-n}(u)\, \mathrm{d}u + \mathcal{O}(T^{-2}) \Big],$$

where the domain has been shifted by $1/2T$ producing an error at most of order $\mathcal{O}(T^{-2})$ as well. The last term can be decomposed as

$$\begin{aligned} \sum_{n \in D_T - D_T} \int_{[\frac{|n|}{2T}, 1 - \frac{|n|}{2T}]} h^2(u) f_n(u) g_{-n}(u)\, \mathrm{d}u &= \sum_{n \in \mathbb{Z}^d} \int_{[0,1]^d} h^2(u) f_n(u) g_{-n}(u)\, \mathrm{d}u \\ &\quad - \underbrace{\sum_{n \in D_T - D_T} \int_{[0,1]^d - [\frac{|n|}{2T}, 1 - \frac{|n|}{2T}]} h^2(u) f_n(u) g_{-n}(u)\, \mathrm{d}u}_{R_1} \\ &\quad - \underbrace{\sum_{n \notin D_T - D_T} \int_{[0,1]^d} h^2(u) f_n(u) g_{-n}(u)\, \mathrm{d}u}_{R_2}. \end{aligned}$$

The first summation on the right-hand side turns out to be

$$\begin{aligned} &\sum_{n \in \mathbb{Z}^d} \int_{[0,1]^d} h^2(u) f_n(u) g_{-n}(u)\, \mathrm{d}u \\ &\quad = \int_{[0,1]^d} h^2(u) \sum_{n \in \mathbb{Z}^d} \int_{\Pi^d} \int_{\Pi^d} f_\theta(u, \lambda_1) (2\pi)^{-d} \nabla_\theta f_\theta^{-1}(u, \lambda_2) \exp(i \langle \lambda_1 - \lambda_2, n \rangle)\, \mathrm{d}\lambda_1\, \mathrm{d}\lambda_2 \\ &\quad = \int_{[0,1]^d} h^2(u) \int_{\Pi^d} \int_{\Pi^d} f_\theta(u, \lambda_1) (2\pi)^{-d} \nabla_\theta f_\theta^{-1}(u, \lambda_2) \sum_{n \in \mathbb{Z}^d} \exp(i \langle \lambda_1 - \lambda_2, n \rangle)\, \mathrm{d}\lambda_1\, \mathrm{d}\lambda_2 \\ &\quad = \int_{[0,1]^d} h^2(u) \int_{\Pi^d} \int_{\Pi^d} f_\theta(u, \lambda_1) (2\pi)^{-d} \nabla_\theta f_\theta^{-1}(u, \lambda_2) (2\pi)^d \delta(\lambda_1 - \lambda_2)\, \mathrm{d}\lambda_1\, \mathrm{d}\lambda_2 \\ &\quad = \int_{[0,1]^d} \int_{\Pi^d} h^2(u) f_\theta(u, \lambda) \nabla_\theta f_\theta^{-1}(u, \lambda)\, \mathrm{d}\lambda \\ &\quad = \int_{[0,1]^d} \int_{\Pi^d} -h^2(u) \frac{\nabla_\theta f_\theta(u, \lambda)}{f_\theta(u, \lambda)}\, \mathrm{d}\lambda \end{aligned}$$

$$= \frac{1}{T^d} \operatorname{tr}\Big[\widehat{U}_T(-h^2 \tfrac{\nabla_\theta f_\theta}{f_\theta})\Big].$$

From assumption (A4) we obtain

$$\begin{aligned} |R_2| &\leq \sum_{n \notin D_T - D_T} \int_{[\frac{|n|}{2T}, 1 - \frac{|n|}{2T}]} h^2(u)|f_n(u)||g_{-n}(u)| \, \mathrm{d}u \\ &\leq \sum_{n \notin D_T - D_T} \int_{[0,1]^d} h^2(u)|f_n(u)||g_{-n}(u)| \, \mathrm{d}u \\ &= \sum_{n \notin D_T - D_T} \mathcal{O}(k_1^2(n)) \\ &= \mathcal{O}(T^{-5d}). \end{aligned}$$

Regarding R_1

$$\begin{aligned} |R_1| &\leq \sum_{n \in D_T - D_T} \sum_{j=1}^{d} \int_{[0,1]^{d-1}} \Big| \int_{[0,1] - [\frac{|n_j|}{2T}, 1 - \frac{|n_j|}{2T}]} b_n(u^{(j)}, u_j) \, \mathrm{d}u_j \Big| \, \mathrm{d}u^{(j)} \\ &\leq 2d \sum_{n \in D_T - D_T} \int_{[0,1]^{d-1}} \Big| \int_{[0, \frac{|n_j|}{2T}]} b_n(u^{(1)}, u_1) \, \mathrm{d}u_1 \Big| \, \mathrm{d}u^{(1)}, \end{aligned}$$

where $u^{(j)}$ represents a vector $u \in [0,1]^d$ with the jth component omitted, and where w.l.o.g. we have assumed that $u^{(1)}$ contributes mostly to the sum. Consequently, we carry out a second order Taylor expansion of the variable u_1 around 0. Considering the assumption (A6) regarding the taper function we obtain

$$b_n(u^{(1)}, u_1) = \frac{u_1^2}{2} \frac{\partial^2 b_n}{\partial u_1}(u^{(1)}, \xi)$$

$$\begin{aligned} &\leq 2d \sum_{n \in D_T - D_T} \int_{[0,1]^{d-1}} \Big| \int_{[0, \frac{|n_1|}{2T}]} \sup_{\xi \in [0, \frac{n_1}{2T}]} \Big| \frac{\xi^2}{2} \frac{\partial^2 b_n}{\partial u_1}(u^{(1)}, \xi) \Big| \, \mathrm{d}u_1 \Big| \, \mathrm{d}u^{(1)} \\ &= 2d \sum_{n \in D_T - D_T} \int_{[0,1]^{d-1}} \frac{|n_1|}{2T} \sup_{\xi \in [0, \frac{n_1}{2T}]} \Big| \frac{\xi^2}{2} \frac{\partial^2 b_n}{\partial u_1}(u^{(1)}, \xi) \Big| \, \mathrm{d}u^{(1)} \\ &\leq 2d \sum_{n \in D_T - D_T} \frac{|n_1|^3}{16T^3} \sup_{u \in [0,1]^d} \Big| \frac{\partial^2 b_n}{\partial u^2}(u) \Big| \\ &= \mathcal{O}(T^{-3}) \sum_{n \in D_T - D_T} |n_1|^3 \sup_{u \in [0,1]^d} \Big| \frac{\partial^2 b_n}{\partial u^2}(u) \Big| \\ &= \mathcal{O}(T^{-3}) \mathcal{O}(\rho^{-2}) \underbrace{\sum_{n \in D_T - D_T} |n_1|^3 \sup_{u \in [0,1]^d} \Big| \frac{\partial^2}{\partial u_1}(f_n(u) g_n(u)) \Big|}_{\mathcal{O}(1)} \end{aligned}$$

$$= \mathcal{O}(T^{-3}\rho^{-2}), \tag{2.44}$$

where we have used the conditions (A4) and (A5) on $f(u,\lambda)$ and $g(u,\lambda)$ to prove the convergence of the sum. Recall that we have assumed $\beta = 1/6$ and $\rho = \mathcal{O}(T^{-\beta})$, thus the last rate is $\mathcal{O}(T^{-3+2\beta})$ which is smaller than $\mathcal{O}(T^{-2})$, which proves our assertion. □

Using Lemmas 2.2.2 and 2.3.4 the variance of the score function for tapered data is

$$\begin{aligned}
\mathrm{Var}\,(S_T^{(h)}(\theta)) &= \frac{1}{(2\pi T)^{2d}} \mathrm{Var}\,\{X'\widetilde{U}_T^{(h)}((2\pi)^{-d}\nabla_\theta f_\theta^{-1})X\} \\
&= \frac{2}{(2\pi T)^{2d}} \mathrm{tr}(\Sigma_T(A_\theta)\widetilde{U}_T^{(h)}((2\pi)^{-d}\nabla_\theta f_\theta^{-1})\Sigma_T(A_\theta)\widetilde{U}_T^{(h)}((2\pi)^{-d}\nabla_\theta f_\theta^{-1})) \\
&= \frac{2}{(2\pi T)^{d}} \Big(\int_{[0,1]^d} \int_{\Pi^d} h^4(u) \Big\{ \frac{\nabla_\theta f_\theta}{f_\theta} \Big\}^2 \mathrm{d}\lambda\, \mathrm{d}u + \mathcal{O}(T^{-1}) \Big).
\end{aligned}$$

If $\int_{\Pi^d} f_\theta(u,\lambda)\,\mathrm{d}\lambda \approx g_\theta$, *i.e.* the process $X_{t,T}$ has approximately the same variance (total power) on each point $\frac{t-1/2}{T}$, then from (2.35) we obtain

$$\mathrm{Var}\,(S_T^{(h)}(\theta))/\mathrm{Var}\,(S_T(\theta)) \approx \int_{[0,1]^d} h_\rho^4(u)\,\mathrm{d}u \approx 1 + \frac{9}{4}\rho d,$$

when $T \to \infty$. Thus, in order to achieve an asymptotically efficient estimator we need to chose $\rho = o(1)$, but not so fast such that the bias term in (2.44) tends to zero as well. In Chapter 4 we have used ρ of order $T^{-1/6}$ for simulations and the real data application.

The last proposition (and previous bias sources) implies that the Whittle likelihood must be modified in order to achieve the above mentioned bias reductions. Therefore, the resulting Whittle likelihood turns out to be

$$\mathcal{L}_T^{(h)}(\theta) = \frac{1}{(2\pi T)^d} \sum_{t \in D_T} \int_{\Pi^d} \Big\{ h^2(\tfrac{t-1/2}{T}) \log[(2\pi)^d f_\theta(\tfrac{t-1/2}{T},\lambda)] + \frac{J_T^{(h)}(\frac{t-1/2}{T},\lambda)}{f_\theta(\frac{t-1/2}{T},\lambda)} \Big\}\,\mathrm{d}\lambda,$$

where the tapered preperiodogram corrected for skewness is

$$\begin{aligned}
J_T^{(h)}(\tfrac{t-1/2}{T},\lambda) := \frac{1}{(2\pi)^d} \sum_k \Big(& \frac{X^{(h)}_{[t-k/2],T}X^{(h)}_{[t+k/2],T} + X^{(h)}_{[t-k/2]^*,T}X^{(h)}_{[t+k/2]^*,T}}{2} \Big) \\
& \times \exp(i\langle\lambda, k\rangle).
\end{aligned}$$

2.3 Appendix: Technical Results

In this section we present technical results needed for proving the propositions and lemmas of this and the next chapter. Some results (Lemmas 2.3.1, 2.3.4 and Proposition 2.3.2) are classical and their proofs can be found in the standard literature. Others correspond to the proofs of technical lemmas needed specifically for our work.

Definition 2.3.1. *Let A be an $n \times n$ matrix. We define the operator (spectral) and Euclidean norm respectively*

$$\begin{aligned} \|A\|_{op} &= \sup_{x\in\mathbb{C}^n} \frac{|Ax|}{|x|} = \sup_{x\in\mathbb{C}^n} \Big(\frac{x^*A^*Ax}{x^*x}\Big)^{1/2} \\ &= [\textit{maximum characteristic root of } A^*A]^{1/2}. \end{aligned}$$

where A^* denotes the conjugate transpose of A, and

$$\|A\|_E = [\mathrm{tr}(AA^*)]^{1/2}.$$

Lemma 2.3.1. *Let A, B be $n \times n$ matrices, and Id denotes the $n \times n$ unit matrix. Then*

(i) $\log\det A \le \mathrm{tr}(A - Id)$ *for a positive $n \times n$ matrix A*

(ii) $\frac{\partial}{\partial x}\log\det A = \mathrm{tr}(A^{-1}\frac{\partial}{\partial x}A)$ *if all elements of A are differentiable functions of $x \in \mathbb{R}$.*

(iii) $\|A\|^2_{op} \le (\sup_i \sum_{j=1}^n |a_{ij}|)(\sup_j \sum_{i=1}^n |a_{ij}|)$

(iv) $|\,\mathrm{tr}(AB)| \le \|A\|_E\|B\|_E$

(v) $\|AB\|_{op} \le \|A\|_{op}\|B\|_E$

(vi) $\|AB\|_E \le \|A\|_E\|B\|_{op}$

(vii) $\|AB\|_{op} \le \|A\|_{op}\|B\|_{op}$

Proof. (i) Let $\{\lambda_i\}_{i=1}^n$ be the eigenvalues of A. Since $A > 0$, $\lambda_i > 0\ \forall i$, then from the concavity of logarithm

$$\log\det A = \log\prod_{i=1}^n \lambda_i = \sum_{i=1}^n \log\lambda_i \le \sum_{i=1}^n(\lambda_i - 1) = \mathrm{tr}(A) - \mathrm{tr}(Id).$$

(ii) The Jacobi's formula yields $\frac{\mathrm{d}}{\mathrm{d}x}\log\det A = \frac{1}{\det A}\mathrm{tr}(\mathrm{adj}(A)\frac{\mathrm{d}A}{\mathrm{d}x})$. Thus, the result follows considering that $\frac{1}{\det A} = A^{-1}\,\mathrm{adj}^{-1}(A)$.

(iii) Given λ and x eigenvalue and eigenvector of A respectively, $|\lambda|\|x\| = \|\lambda x\| = \|Ax\| \le \|A\|\|x\|$, hence $\rho(A) \le \|A\|$, with $\rho(\cdot)$ spectral radius. Finally $\|A\|^2_{op} = \rho(A^*A) \le \|A^*A\|_1 \le \|A^*\|_1\|A\|_1 = \|A\|_\infty\|A\|_1$.

(iv) Using Cauchy-Schwarz inequality

$$\mathrm{tr}^2(AB) = (\sum_{i,j=1}^n a_{ij}b_{ji})^2 \le (\sum_{i,j=1}^n a_{ij}^2)(\sum_{i,j=1}^n b_{ji}^2) = \mathrm{tr}(AA^*)\,\mathrm{tr}(BB^*) = \|A\|^2_E\|B\|^2_E.$$

(v) We denote $\|A\|_{op} = [\lambda_A^{(n)}]^{1/2}$ the maximum eigenvalue of A^*A, then

$$\det(A^*AB^*B) = \det(A^*A)\det(B^*B) = \prod_{i=1}^{n} \lambda_A^{(i)}\lambda_B^{(i)},$$

hence

$$\|AB\|_{op}^2 = \lambda_{AB}^{(n)} = \lambda_A^{(n)}\lambda_B^{(n)} = \|A\|_{op}^2\lambda_B^{(n)}.$$

The result follows noting that $\lambda_B^{(n)} \leq \sum_{i=1}^{n} \lambda_B^{(i)} = \text{tr}(BB^*) = \|B\|_E^2$

(vi) By definition

$$\|AB\|_E = [\text{tr}(ABB^*A^*)]^{1/2} = \text{tr}(A^*ABB^*)^{1/2} = (\sum_{i=1}^{n} \lambda_i^{A^*ABB^*})^{1/2}.$$

The right-hand side of the last equation can be bounded as

$$(\sum_{i=1}^{n} \lambda_i^{A^*A}\lambda_i^{BB^*})^{1/2} \leq (\sum_{i=1}^{n} \lambda_i^{A^*A} \max_{1\leq i\leq n} \lambda_i^{BB^*})^{1/2} = \|A\|_E\|B\|_{op}.$$

(vii) Take x with $\|x\|_2 = 1$ then notice

$$\|ABx\|_2 \leq \|A(Bx)\|_2 \leq \|A\|_2\|Bx\|_2 \leq \|A\|_2\|B\|_2\|x\|_2.$$

The result follows taking supremum over $\|x\|_2 = 1$.

□

In (2.13) we defined the function $k_1(r)$. Additionally, we define $k_2(r, s, T)$ as

$$k_2(r, s, T) = \prod_{j=1}^{d} \Big(\frac{1}{r_j^3 s_j^3} + \frac{1}{(T + 1 - r_j)^3(T + 1 - s_j)^3}\Big), \tag{2.45}$$

where $r, s \in D_T,\ T \in \mathbb{N}$.

Lemma 2.3.2. *The functions k_1 and k_2 satisfy the following relations*

(i) $\frac{1}{T^d} \sum_{r\in D_T} \mathcal{O}(k_2(r, r, T)) = \mathcal{O}(T^{-d})$

(ii) $\sum_{t\in D_T} k_1(|r - t|)k_1(|t - s|) = \mathcal{O}(k_1(|r - s|))$

(iii) $\sum_{t\in D_T} k_2(r, t, T)k_1(|t - s|) = \mathcal{O}(k_2(r, s, T))$

(iv) $\sum_{t\in D_T} k_1(|r - t|)k_1(|t - s|) = \mathcal{O}(k_2(r, s, T))$

Proof. (i) It is enough to notice that

$$\frac{1}{T^d} \sum_{r\in D_T} \mathcal{O}(k_2(r, r, T)) = \frac{1}{T^d}\mathcal{O}\Big(\Big[\sum_{r_j}(\frac{1}{r_j^6} + \frac{1}{(T+1-r_j)^6})\Big]^d\Big) = T^{-d}\mathcal{O}(1)$$

(ii) The triangular inequality yields

$$\sum_{t\in D_T} k_1(|r-t|)k_1(|t-s|) \le \sum_{t\in D_T}\prod_{j=1}^{d}\frac{1}{(|r_j-t_j||t_j-s_j|+|r_j-s_j|+1)^3}$$
$$=\sum_{t\in D_T}\prod_{j=1}^{d}\frac{1}{(|r_j-s_j|+1)^3}\prod_{j=1}^{d}\left(\frac{|r_j-s_j|+1}{|r_j-t_j||t_j-s_j|+|r_j-s_j|+1}\right)^3$$
$$=k_1(|r-s|)\sum_{t_1}\left(\frac{|r_1-s_1|+1}{|r_1-t_1||t_1-s_1|+|r_1-s_1|+1}\right)^3\times$$
$$\cdots\times\sum_{t_d}\left(\frac{|r_d-s_d|+1}{|r_d-t_d||t_d-s_d|+|r_d-s_d|+1}\right)^3$$

The result follows noticing that each sum is convergent for any T.

(iii) The sum can be written like

$$k_2(r,s,T)\sum_{t\in D_T}\prod_{j=1}^{d}\left(\frac{s_j^3(T+1-r_j)^3(T+1-s_j)^3}{t_j^3(|t_j-s_j|+1)^3(r_j^3s_j^3+(T+1-r_j)^3(T+1-s_j)^3)}\right.$$
$$\left.+\frac{r_j^3s_j^3(T+1-s_j)^3}{(T+1-t_j)^3(|t_j-s_j|+1)^3(r_j^3s_j^3+(T+1-r_j)^3(T+1-s_j)^3)}\right)$$

which is equal to $k_2(r,s,T)\mathcal{O}(1)$ since by comparison test with $\sum(1/t^3)$ can be verified that each of the d sums are bounded.

(iv) The result is obtained analogously to the previous case by comparing with $\sum(1/t^6)$.

□

Proof of Lemma 2.2.2. The key to this result are the following two approximations for the involved matrices:

$$\Big[\prod_{k=1}^{n} C_k\Big]_{r,s} = \mathcal{O}(k_1(|r-s|)) \tag{2.46}$$

and

$$\Big[\prod_{k=1}^{n} C_k\Big]_{r,s} = \Big(\widetilde{U}_T\Big(\prod_{k=1}^{n}\psi_k\Big)\Big)_{r,s} + \mathcal{O}(T^{-1}) + \mathcal{O}(k_2(r,s,T)) \tag{2.47}$$

$\forall r,s\in D_T$, where $\widetilde{U}_T$ is defined as

$$\Big(\widetilde{U}_T(\phi)\Big)_{r,s} = \frac{1}{(2\pi)^d}\int_{\Pi^d}\phi(\tfrac{r+s-1}{2T},\lambda)\exp(i\langle\lambda,r-s\rangle)\,\mathrm{d}\lambda$$

From (2.47) the assertion follows with:

$$\frac{1}{T^d}\sum_{r\in D_T}\Big(\prod_{k=1}^{n}C_k\Big)_{r,r} = \frac{1}{T^d}\sum_{r\in D_T}\Big(\widetilde{U}_T(\prod_{k=1}^{n}\psi_k)\Big)_{r,r}$$

$$+\frac{1}{T^d}\sum_{r\in D_T}\mathcal{O}(T^{-1})+\frac{1}{T^d}\sum_{r\in D_T}\mathcal{O}(k_2(r,r,T)).$$

Using Lemma 2.3.2 (i)

$$\begin{aligned}\frac{1}{T^d}\sum_{r\in D_T}\Big(\prod_{k=1}^{n}C_k\Big)_{r,r} &= \frac{1}{T^d}\sum_{r\in D_T}\Big(\widetilde{U}_T\Big(\prod_{k=1}^{n}\psi_k\Big)\Big)_{r,r}+\mathcal{O}(T^{-1})\\ &= \frac{1}{T^d}\Big[\sum_{r\in D_T}\frac{1}{(2\pi)^d}\int\limits_{\Pi^d}\prod_{k=1}^{n}\psi_k\Big(\tfrac{r-1/2}{T},\lambda\Big)\,\mathrm{d}\lambda\Big]+\mathcal{O}(T^{-1})\\ &= \frac{1}{(2\pi)^d}\int\limits_{[0,1]^d}\int\limits_{\Pi^d}\prod_{k=1}^{n}\psi_k(u,k)\,\mathrm{d}\lambda\,\mathrm{d}u+\mathcal{O}(T^{-1}).\end{aligned}$$

The properties (2.46) and (2.47) are proved by induction over n. Let us assume that $C_n=\Sigma_T(A_n)$.
For $n=1$ we have by assumption (A1) on the covariances

$$\begin{aligned}(C_1)_{r,s} &= \int\limits_{\Pi^d}A^0_{r,T}(\lambda)\overline{A^0_{s,T}(\lambda)}\exp(i\langle\lambda,r-s\rangle)\,\mathrm{d}\lambda\\ &= Cov(X_{r,T},X_{s,T})\\ &= \mathcal{O}(k_1(|r-s|)).\end{aligned}$$

Furthermore, the regularity imposed in (A5) implies

$$\begin{aligned}(C_1)_{r,s} &= \int\limits_{\Pi^d}A^0_{r,T}(\lambda)\overline{A^0_{s,T}(\lambda)}\exp(i\langle\lambda,r-s\rangle)\,\mathrm{d}\lambda\\ &= \int\limits_{\Pi^d}A\big(\tfrac{r-1/2}{T},\lambda\big)\overline{A\big(\tfrac{s-1/2}{T},\lambda\big)}\exp(i\langle\lambda,r-s\rangle)\,\mathrm{d}\lambda+\mathcal{O}(T^{-1})\\ &= \int\limits_{\Pi^d}\big|A\big(\tfrac{r+s-1}{2T},\lambda\big)\big|^2\exp(i\langle\lambda,r-s\rangle)\,\mathrm{d}\lambda+\mathcal{O}(T^{-1})\\ &= \Big(\widetilde{U}_T((2\pi)^d|A|^2)\Big)_{r,s}+\mathcal{O}(T^{-1}),\end{aligned}$$

using (1.3) and (1.9). We assume the properties (2.46) and (2.47) are proved for all $n\le m-1$ and let $n=m$

$$\Big(\prod_{k=1}^{n}C_k\Big)_{r,s}=\Big(\Big(\prod_{k=1}^{n-1}C_k\Big)C_n\Big)_{r,s},$$

which is equivalent to multiplying the rth row of $(\prod_{k=1}^{n-1}C_k)$ and the sth column of C_n. Using the induction hypothesis, we know that the elements of the rth row of

$(\prod_{k=1}^{n-1} C_k)$ have orders: $\mathcal{O}(k_1(|r-s_j|))$ with s_j running over D_T and the orders of C_n are $\mathcal{O}(k_1(|r_j-s|))$ with r_j running over D_T therefore

$$\begin{aligned}\Big(\Big(\prod_{k=1}^{n-1} C_k\Big)C_n\Big)_{r,s} &= \sum_{j=1}^{T^d}\Big(\prod_{k=1}^{n-1} C_k\Big)_{r,j}(C_n)_{j,s} \\ &= \sum_{t\in D_T}\Big(\prod_{k=1}^{n-1} C_k\Big)_{r,t}(C_n)_{t,s} \\ &= \sum_{t\in D_T} k_1(|r-t|)k_1(|t-s|) \\ &= \mathcal{O}(k_1(|r-s|)),\end{aligned}$$

by using Lemma 2.3.2 (ii). On the other hand

$$\begin{aligned}(\prod_{k=1}^{n} C_k)_{r,s} &= \sum_{t\in D_T}(\prod_{k=1}^{n-1} C_k)_{r,t}(C_n)_{t,s} \\ &= \sum_{t\in D_T}\{(\widetilde{U}_T(\prod_{k=1}^{n-1}\psi_k))_{r,t} + \mathcal{O}(T^{-1}) + \mathcal{O}(k_2(r,t,T))\} \\ &\quad \times\Sigma_T(A_n)_{t,s}. \end{aligned} \tag{2.48}$$

Note that

$$\begin{aligned}\sum_{t\in D_T}\Sigma_T(A_n)_{t,s} &= \sum_{t\in D_T}\mathcal{O}(k_1(|t-s|)) \\ &= M\sum_{t\in D_T}\prod_{j=1}^{d}\tfrac{1}{(|t_j-s_j|+1)^3} \\ &= M\sum_{t_1}\tfrac{1}{(|t_1-s_1|+1)^3}\cdots\sum_{t_d}\tfrac{1}{(|t_d-s_d|+1)^3} \\ &= \mathcal{O}(1),\end{aligned}$$

for some $M>0$. From Lemma 2.3.2 (iii) and the induction hypothesis $\Sigma_T(A_n)_{t,s} = \mathcal{O}(k_1|t-s|)$, (2.48) yields

$$\begin{aligned}&= \sum_{t\in D_T}(\widetilde{U}_T(\prod_{k=1}^{n-1}\psi_k))_{r,t}\Sigma_T(A_n)_{t,s} + \mathcal{O}(T^{-1})\mathcal{O}(1) + \mathcal{O}(k_2(r,s,T)) \\ &= \sum_{t\in D_T}\frac{1}{(2\pi)^d}\int_{\Pi^d}\prod_{k=1}^{n-1}\psi_k(\tfrac{r+t-1}{2T},\lambda)\exp(i\langle\lambda,r-t\rangle)\,\mathrm{d}\lambda \\ &\times\int_{\Pi^d}A^0_{t,T}(\lambda)\overline{A^0_{s,T}(\lambda)}\exp(i\langle\lambda,t-s\rangle)\,\mathrm{d}\lambda + \mathcal{O}(T^{-1}) + \mathcal{O}(k_2(r,s,T))\end{aligned}$$

We let $g(u,\lambda) = \prod_{k=1}^{n-1} \psi_k(u,\lambda)$ and $h(u,\lambda) = \psi_n(u,\lambda)$ and let g_n and h_n the n*th* coefficient of the Fourier expansion of $g(u,\lambda)$ and $h(u,\lambda)$ respectively in frequency direction. Then, the last expression is

$$= \frac{1}{(2\pi)^d} \sum_{t\in D_T} g_{r-t}(\tfrac{r+t-1}{2T}) \int_{\Pi^d} \exp(i\langle\lambda, t-s\rangle) A^0_{t,T}(\lambda)\overline{A^0_{s,T}(\lambda)}\,\mathrm{d}\lambda$$
$$+ \mathcal{O}(T^{-1}) + \mathcal{O}(k_2(r,s,T))$$

From (1.3) we replace $A^0_{t,T}(\lambda)$ and $A^0_{s,T}(\lambda)$ by the respective estimates and get

$$= \frac{1}{(2\pi)^d} \sum_{t\in D_T} g_{r-t}(\tfrac{r+t-1}{2T}) \int\limits_{\Pi^d} \exp(i\langle\lambda, t-s\rangle) A(\tfrac{2t-1}{2T},\lambda)\overline{A(\tfrac{2s-1}{2T},\lambda)}\,\mathrm{d}\lambda$$
$$+ \mathcal{O}(T^{-1}) + \mathcal{O}(k_2(r,s,T)),$$

and using (1.9) we have

$$= \frac{1}{(2\pi)^d} \sum_{t\in D_T} g_{r-t}(\tfrac{r+t-1}{2T}) \int\limits_{\Pi^d} \exp(i\langle\lambda, t-s\rangle) \left|A(\tfrac{t+s-1}{2T},\lambda)\right|^2 \mathrm{d}\lambda \tag{2.49}$$
$$+ \mathcal{O}(T^{-1}) + \mathcal{O}(k_2(r,s,T))$$
$$= \frac{1}{(2\pi)^{2d}} \sum_{t\in D_T} g_{r-t}(\tfrac{r+t-1}{2T}) h_{t-s}(\tfrac{t+s-1}{2T}) + \mathcal{O}(T^{-1}) + \mathcal{O}(k_2(r,s,T)). \tag{2.50}$$

(A4) and Lemma 2.2.1 (iv) imply for the sum above

$$\sum_{t\in D_T} g_{r-t}(\tfrac{r+t-1}{2T}) h_{t-s}(\tfrac{t+s-1}{2T}) = \sum_{t\in D_T} \{g_{r-t}(\tfrac{r+s-1}{2T}) + \mathcal{O}(k_1(|r-t|))\}$$
$$\times \{h_{t-s}(\tfrac{r+s-1}{2T}) + \mathcal{O}(k_1(|t-s|))\}$$
$$= \sum_{t\in D_T} g_{r-t}(\tfrac{r+s-1}{2T}) h_{t-s}(\tfrac{r+s-1}{2T}) + \mathcal{O}(k_2(r,s,T))$$

We denote $D^0_T := \{-T,\ldots,T\}^d$. Hence (2.49) can be written as

$$\frac{1}{(2\pi)^{2d}} \sum_{t\in\mathbb{Z}^d} g_{r-t}(\tfrac{r+s-1}{2T}) h_{t-s}(\tfrac{r+s-1}{2T}) - \frac{1}{(2\pi)^{2d}} \sum_{t\notin D_T} \mathcal{O}(k_1(|r-t|)k_1(|t-s|))$$
$$+\mathcal{O}(T^{-1}) + \mathcal{O}(k_2(r,s,T))$$
$$= \frac{1}{(2\pi)^{2d}} \sum_{t\in\mathbb{Z}^d} g_{r-t}(\tfrac{r+s-1}{2T}) h_{t-s}(\tfrac{r+s-1}{2T}) - \mathcal{O}(k_2(r,s,T)) + \mathcal{O}(T^{-1}) + \mathcal{O}(k_2(r,s,T))$$
$$= \lim_{T\to\infty} \frac{1}{(2\pi)^{2d}} \sum_{t\in D^0_T} g_{r-t}(\tfrac{r+s-1}{2T}) h_{t-s}(\tfrac{r+s-1}{2T}) + \mathcal{O}(T^{-1}) + \mathcal{O}(k_2(r,s,T))$$
$$= \lim_{T\to\infty} \frac{1}{(2\pi)^{2d}} \sum_{t\in D^0_T} \int\limits_{\Pi^d}\int\limits_{\Pi^d} g(\tfrac{r+s-1}{2T},\lambda_1) h(\tfrac{r+s-1}{2T},\lambda_2)$$

$$
\begin{aligned}
&\times \exp(i\langle\lambda_1, r-t\rangle + i\langle\lambda_2, t-s\rangle)\, \mathrm{d}\lambda_1\, \mathrm{d}\lambda_2 + \mathcal{O}(T^{-1}) + \mathcal{O}(k_2(r,s,T)) \\
= \; & \lim_{T\to\infty} \frac{1}{(2\pi)^{2d}} \int\limits_{\Pi^d}\int\limits_{\Pi^d} g(\tfrac{r+s-1}{2T},\lambda_1) h(\tfrac{r+s-1}{2T},\lambda_2) \exp(i\langle\lambda_1, r\rangle - i\langle\lambda_2, s\rangle) \\
&\times \sum_{t\in D_T^0} \exp(i\langle\lambda_2-\lambda_1, t\rangle)\, \mathrm{d}\lambda_1\, \mathrm{d}\lambda_2 + \mathcal{O}(T^{-1}) + \mathcal{O}(k_2(r,s,T)) \qquad (2.51)
\end{aligned}
$$

The sum corresponds to the Fejér kernel in d dimensions. Letting $T \to \infty$ we obtain the Dirac delta function. Thus, the last equation turns out be

$$
\begin{aligned}
= \; & \frac{1}{(2\pi)^{2d}} \int\limits_{\Pi^d}\int\limits_{\Pi^d} g(\tfrac{r+s-1}{2T},\lambda_1) h(\tfrac{r+s-1}{2T},\lambda_2) \\
&\times \exp(i\langle\lambda_1, r\rangle - i\langle\lambda_2, s\rangle)(2\pi)^d \delta(\lambda_2-\lambda_1)\, \mathrm{d}\lambda_1\, \mathrm{d}\lambda_2 + \mathcal{O}(T^{-1}) + \mathcal{O}(k_2(r,s,T)) \\
= \; & \frac{1}{(2\pi)^{d}} \int\limits_{\Pi^d} g(\tfrac{r+s-1}{2T},\lambda) h(\tfrac{r+s-1}{2T},\lambda) \exp(i\langle\lambda, r-s\rangle)\, \mathrm{d}\lambda + \mathcal{O}(T^{-1}) + \mathcal{O}(k_2(r,s,T)) \\
= \; & \frac{1}{(2\pi)^{d}} \int\limits_{\Pi^d} (gh)(\tfrac{r+s-1}{2T},\lambda) \exp(i\langle\lambda, r-s\rangle)\, \mathrm{d}\lambda + \mathcal{O}(T^{-1}) + \mathcal{O}(k_2(r,s,T)) \\
= \; & \Big(\widetilde{U}_T(\prod_{k=1}^{n} \psi_k)\Big)_{r,s} + \mathcal{O}(T^{-1}) + \mathcal{O}(k_2(r,s,T))
\end{aligned}
$$

Finally, the proof in the case $C_n = U_T(\phi)$ turns out to be simpler. We prove the case for $n = 1$. The general case follows analogously to what we have done with $C_n = \Sigma_T(A_n)$.

Since $\phi(u,\lambda)$ satisfies (A4) we obtain immediately $(C_1)_{r,s} = \mathcal{O}(k_1(|r-s|))$. The same assumption implies Lipschitz continuity in u, thus

$$
|\phi(\tfrac{1}{T}[\tfrac{r+s}{2}] - \tfrac{1}{2T},\lambda) - \phi(\tfrac{r+s-1}{2T},\lambda)| \le \frac{L}{T}|[\tfrac{r+s}{2}] - \tfrac{r+s}{2}| = \mathcal{O}(T^{-1})
$$

therefore

$$
\begin{aligned}
(C_1)_{r,s} &= \int\limits_{\Pi^d} \phi(\tfrac{1}{T}[\tfrac{r+s}{2}] - \tfrac{1}{2T},\lambda) \exp(i\langle\lambda, r-s\rangle)\, \mathrm{d}\lambda \\
&= \int\limits_{\Pi^d} \phi(\tfrac{r+s-1}{2T},\lambda) \exp(i\langle\lambda, r-s\rangle)\, \mathrm{d}\lambda + \mathcal{O}(T^{-1}) \\
&= \big(\widetilde{U}_T((2\pi)^d\phi)\big)_{r,s} + \mathcal{O}(T^{-1})
\end{aligned}
$$

□

Lemma 2.3.3. *If $A(u,\lambda)$ and $f(u,\lambda)$ fulfil assumptions (A3), (A4), (A5), then*

$$
\frac{1}{T^d}\|U_T((2\pi)^{-2d}|A|^{-2}) - \Sigma_T^{-1}(A)\|_E^2 \;=\; \mathcal{O}(T^{-1}) \qquad (2.52)
$$

$$\begin{aligned} \frac{1}{T^d}\|U_T((2\pi)^{-2d}f^{-1}) - U_T^{-1}(f)\|_E^2 &= \mathcal{O}(T^{-1}) && (2.53)\\ \frac{1}{T^d}\|U_T(|A|^2) - \Sigma_T(A)\|_E^2 &= \mathcal{O}(T^{-1}) && (2.54) \end{aligned}$$

Proof. (i) Lemma 2.3.1(i) and lemma 2.2.1 yield

$$\begin{aligned} \|U_T((2\pi)^{-2d}|A|^{-2}) - \Sigma_T^{-1}(A)\|_E^2 &\leq \|U_T((2\pi)^{-2d}|A|^{-2})\Sigma_T(A) - Id\|_E^2 \\ &\qquad \times\|\Sigma_T^{-1}(A)\|_{op}^2 \\ &= \|U_T((2\pi)^{-2d}|A|^{-2})\Sigma_T(A) - Id\|_E^2\mathcal{O}(1) \end{aligned}$$

By using Lemma 2.2.2, the last norm above can be bounded as

$$\begin{aligned} \|U_T((2\pi)^{-2d}|A|^{-2})\Sigma_T(A) - Id\|_E^2 &= \operatorname{tr}(U_T^2((2\pi)^{-2d}|A|^{-2})\Sigma_T^2(A)) \\ &\quad - 2\operatorname{tr}(U_T((2\pi)^{-2d}|A|^{-2})\Sigma_T(A)) + \operatorname{tr}(Id) \\ &= T^d + \mathcal{O}(T^{d-1}) - 2T^d + \mathcal{O}(T^{d-1}) + T^d \\ &= \mathcal{O}(T^{d-1}) \end{aligned}$$

(ii) The result follows using same arguments as in (i)

(iii) We applied two times the factorization in (i) obtaining

$$\begin{aligned} \|U_T(|A|^2) - \Sigma_T(A)\|_E^2 &\leq \|U_T(|A|^2)\|_{op}^2\|\Sigma_T(A)\|_{op}^2\|U_T^{-1}(|A|^2) - \Sigma_T^{-1}(A)\|_E^2 \\ &= \mathcal{O}(1)\|U_T^{-1}(|A|^2) - \Sigma_T^{-1}(A)\|_E^2 \\ &\leq \mathcal{O}(1)\|U_T^{-1}(|A|^2) - U_T((2\pi)^{-2d}|A|^{-2})\|_E^2 \\ &\qquad +\mathcal{O}(1)\|U_T((2\pi)^{-2d}|A|^{-2}) - \Sigma_T^{-1}(A)\|_E^2, \end{aligned}$$

therefore from (i) and (ii), we obtain the desired result.

□

Proof of Lemma 2.2.3. From Lemma 2.2.2 we get

$$\begin{aligned} &\Big|\frac{1}{T^d}\operatorname{tr}\Big[\prod_{k=1}^n C_k\Big] - \frac{1}{(2\pi)^d}\int\limits_{[0,1]^d}\int\limits_{\Pi^d}\prod_{k=1}^n \psi_k(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u\Big| \\ &\leq \frac{1}{T^d}\Big|\operatorname{tr}\Big[\prod_{k=1}^n C_k\Big] - \operatorname{tr}\Big[\prod_{k=1}^n U_T(\phi_k)\Big]\Big| \\ &\quad +\Big|\frac{1}{T^d}\operatorname{tr}\Big[\prod_{k=1}^n U_T(\phi_k)\Big] - \frac{1}{(2\pi)^d}\int\limits_{[0,1]^d}\int\limits_{\Pi^d}\prod_{k=1}^n \psi_k(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u\Big| \\ &= \frac{1}{T^d}\Big|\operatorname{tr}\Big[\prod_{k=1}^n C_k\Big] - \operatorname{tr}\Big[\prod_{k=1}^n U_T(\phi_k)\Big]\Big| + \mathcal{O}(T^{-1}) \end{aligned}$$

We define the set

$$K := \{k; 1 \leq k \leq n \text{ such that } C_k \in \{\Sigma_T(A_k)^{-1}, U_T(\phi_k)^{-1}\}\}$$

We prove that

$$\frac{1}{T^d}\Big|\operatorname{tr}\Big[\prod_{k=1}^{n} C_k\Big] - \operatorname{tr}\Big[\prod_{k=1}^{n} U_T(\phi_k)\Big]\Big| = \mathcal{O}(T^{-1}) \tag{2.55}$$

by induction over $|K|$ (cardinality of K). If $|K| = 0$, then $C_k \in \{\Sigma_T(A_k), U_T(\phi_k)\}$ for all $1 \leq k \leq n$; therefore Lemma 2.2.2 applies and shows (2.55). Now we assume that (2.55) is proved for all $|K| \leq m$ and let $|K| = m + 1$. Without loss of generality we may assume $n \in K$, furthermore, let us assume $C_n = \Sigma_T^{-1}(A_n)$.

$$\begin{aligned}\frac{1}{T^d}\Big|\operatorname{tr}\Big[\prod_{k=1}^{n} C_k\Big] - \operatorname{tr}\Big[\prod_{k=1}^{n} U_T(\phi_k)\Big]\Big| &\leq \frac{1}{T^d}\Big|\operatorname{tr}\Big[\prod_{k=1}^{n} C_k\Big] - \operatorname{tr}\Big[\Big(\prod_{k=1}^{n-1} C_k\Big)U_T(\phi_n)\Big]\Big| \\ &+\frac{1}{T^d}\Big|\operatorname{tr}\Big[\Big(\prod_{k=1}^{n-1} C_k\Big)U_T(\phi_n)\Big] - \operatorname{tr}\Big[\prod_{k=1}^{n} U_T(\phi_k)\Big]\Big|,\end{aligned}$$

where the second term above has order $\mathcal{O}(T^{-1})$ by the induction hypothesis. The first term can be bounded as

$$\begin{aligned}&\frac{1}{T^d}\Big|\operatorname{tr}\Big[\prod_{k=1}^{n} C_k\Big] - 2\operatorname{tr}\Big[\Big(\prod_{k=1}^{n-1} C_k\Big)U_T(\phi_n)\Big] + \operatorname{tr}\Big[\Big(\prod_{k=1}^{n-1} C_k\Big)U_T(\phi_n)C_n^{-1}U_T(\phi_n)\Big]\Big| \\ &+\frac{1}{T^d}\Big|\operatorname{tr}\Big[\Big(\prod_{k=1}^{n-1} C_k\Big)U_T(\phi_n)\Big] - \operatorname{tr}\Big[\Big(\prod_{k=1}^{n-1} C_k\Big)U_T(\phi_n)C_n^{-1}U_T(\phi_n)\Big]\Big|.\end{aligned}$$

Regarding the second term, we can approximate the matrix $\Sigma_T^{-1}(A_n)$ by $U_T(\phi_n)$ where $\phi_n(u,\lambda) = (2\pi)^{-2d}|A_n(u,\lambda)|^{-2}$ (see Lemma 2.3.3). Thus, the induction hypothesis implies

$$\begin{aligned}\frac{1}{T^d}\operatorname{tr}\Big[\Big(\prod_{k=1}^{n-1} C_k\Big)U_T(\phi_n)\Big] &= \frac{1}{(2\pi)^{2d}}\int_{[0,1]^d}\int_{\Pi^d}\prod_{k=1}^{n-1}\psi_k(u,\lambda)|A_n(u,\lambda)|^{-2}\,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-1}) \\ &= \frac{1}{T^d}\operatorname{tr}\Big[\Big(\prod_{k=1}^{n-1} C_k\Big)U_T(\phi_n)C_n^{-1}U_T(\phi_n)\Big],\end{aligned}$$

and hence this second term has order $\mathcal{O}(T^{-1})$ such that

$$\begin{aligned}&= \frac{1}{T^d}\Big|\operatorname{tr}\Big[\Big(\prod_{k=1}^{n-1} C_k\Big)(C_n - 2U_T(\phi_n) + U_T(\phi_n)C_n^{-1}U_T(\phi_n))\Big]\Big| + \mathcal{O}(T^{-1}) \\ &= \frac{1}{T^d}\Big|\operatorname{tr}\Big[\Big(\prod_{k=1}^{n-1} C_k\Big)(C_n - U_T(\phi_n))C_n^{-1}(C_n - U_T(\phi_n))\Big]\Big| + \mathcal{O}(T^{-1}).\end{aligned}$$

Using Lemmas 2.3.1 (iv), (vi), (vii), 2.2.1 and 2.3.3 we can bound the last expression by

$$\begin{aligned}
&\le \frac{1}{T^d}\|(C_n - U_T(\phi_n))\Big(\prod_{k=1}^{n-1} C_k\Big)\|_E \|(C_n - U_T(\phi_n))C_n^{-1}\|_E \\
&\le \frac{1}{T^d}\|C_n - U_T(\phi_n)\|_E \Big\|\prod_{k=1}^{n-1} C_k\Big\|_{op} \|C_n - U_T(\phi_n)\|_E \|C_n^{-1}\|_{op} \\
&\le \frac{1}{T^d}\Big(\prod_{k=1}^{n-1} \|C_k\|_{op}\Big)\|C_n^{-1}\|_{op}\|C_n - U_T(\phi_n)\|_E^2 \\
&= \mathcal{O}(1)\mathcal{O}(T^{-1}).
\end{aligned}$$

In case $C_n = U_T^{-1}(\phi_n)$ we may use Lemma 2.3.3 to approximate the inverse. The rest of the proof follows analogously. □

Proposition 2.3.1. *Suppose (A1)-(A5) hold, then*

$$\mathcal{O}(bias\ \widehat{\theta}_T) = \mathcal{O}(bias\ S_T(\theta)).$$

where $\widehat{\theta}_T := \arg\min_{\theta\in\Theta} \mathcal{L}_T(\theta)$, $\mathcal{L}_T(\theta)$ *is the Whittle likelihood and* $S_T(\theta)$ *is its score function.*

Proof. By Taylor expansion

$$0 = \nabla_\theta \mathcal{L}_T(\widehat{\theta}) = \nabla_\theta \mathcal{L}_T(\theta_0) + \nabla_\theta^2 \mathcal{L}_T(\overline{\theta}_T)(\widehat{\theta}_T - \theta_0) \tag{2.56}$$

where $\overline{\theta}_T$ lies between θ_0 and $\widehat{\theta}_T$, *i.e.* $|\overline{\theta}_T - \theta_0| \le |\widehat{\theta}_T - \theta_0|$ and therefore the assumptions, together with Theorem 3.1.2, imply $\widehat{\theta}_T \xrightarrow{p} \theta_0$ then $\overline{\theta}_T \xrightarrow{p} \theta_0$. From (2.56) we obtain

$$\begin{aligned}
\widehat{\theta}_T - \theta_0 &= -\{\nabla_\theta^2 \mathcal{L}_T(\overline{\theta}_T)\}^{-1}\nabla_\theta \mathcal{L}_T(\theta_0), \\
\mathrm{E}_{\theta_0}\widehat{\theta}_T - \theta_0 &= -\mathrm{E}_{\theta_0}\{\nabla_\theta^2 \mathcal{L}_T(\overline{\theta}_T)\}^{-1}\nabla_\theta \mathcal{L}_T(\theta_0),
\end{aligned}$$

by Theorem 3.1.1, $\nabla_\theta^2 \mathcal{L}_T(\overline{\theta}_T) \xrightarrow{p} \Gamma_1^{-1}$, such that we decompose

$$= -\Gamma_1^{-1}\mathrm{E}_{\theta_0}\nabla_\theta \mathcal{L}_T(\theta_0) - \mathrm{E}_{\theta_0}\{[\nabla_\theta^2 \mathcal{L}_T(\overline{\theta}_T)]^{-1} - \Gamma_1^{-1}\}\nabla_\theta \mathcal{L}_T(\theta_0),$$

and we get by Slutsky's theorem

$$= -\Gamma_1^{-1}\mathrm{E}_{\theta_0}\nabla_\theta \mathcal{L}_T(\theta_0) - o(1)\mathrm{E}_{\theta_0}\nabla_\theta \mathcal{L}_T(\theta_0).$$

Therefore,

$$\mathcal{O}(bias\ \widehat{\theta}_T) = \mathcal{O}(bias\ \nabla_\theta \mathcal{L}_T(\theta_0)).$$

□

Lemma 2.3.4. *Let A and B be two symmetric matrices and X a Gaussian vector with $EX = \mu$ and $Cov\,X = \Sigma = CC'$ then*

(i) $E(X'AX) = \mathrm{tr}(A\Sigma) + \mu' A\mu$.

If additionally $C'AC \neq 0$ and $C'BC \neq 0$ then

(ii) $Cov(X'AX, X'BX) = 2\,\mathrm{tr}(A\Sigma B\Sigma) + 4\mu' A\Sigma B\mu$

Proof. (i) Notice that $X'AX$ is a 1×1 matrix and hence

$$\begin{aligned} \mathrm{E}\,(X'AX) = \mathrm{tr}\,\mathrm{E}\,(X'AX) &= \mathrm{E}\ \mathrm{tr}(AXX') \\ &= \mathrm{tr}\,A(\Sigma + \mu\mu') \\ &= \mathrm{tr}\,A\Sigma + \mathrm{tr}(A\mu\mu') \\ &= \mathrm{tr}\,A\Sigma + \mathrm{tr}\,\mu' A\mu \\ &= \mathrm{tr}\,A\Sigma + \mu' A\mu. \end{aligned}$$

(ii) This proof is much more involved than (i) and it is not given here (see Theorem 3.2d.3 in Mathai and Provost (1992))

□

Proposition 2.3.2. *Let $m : [a,b]^d \to \mathbb{R}$ differentiable with gradient ∇m Lipschitz continuous, i.e. for some $L > 0$*

$$\|\nabla m(x) - \nabla m(y)\| \leq L\|x - y\|, \qquad x, y \in [a,b]^d$$

then

$$\Big| \int\limits_{(a,b]^d} m(x)\,\mathrm{d}x - \Big(\frac{b-a}{T}\Big)^d \sum_{k_1,\ldots,k_d=1}^{T} m(x_{k_1},\ldots,x_{k_d}) \Big| = \mathcal{O}(T^{-1}) \int\limits_{(a,b]^d} \sum_{i=1}^{d} \frac{\partial m}{\partial x_i}\,\mathrm{d}x + \mathcal{O}(\frac{L}{T^2}),$$

where we have used $x = (x_1,\ldots,x_d)$ and $\mathrm{d}x = \mathrm{d}x_1 \cdots \mathrm{d}x_d$.

Proof. Let $(x_{1k_1}, x_{2k_2}, \ldots, x_{dk_d})$, $1 \leq k_1,\ldots,k_d \leq T$ be an equidistant grid in $(a,b]^d$ and let $I_{k_1\cdots k_d} = (x_{k_1-1}, x_{k_1}] \times \cdots \times (x_{k_d-1}, x_{k_d}]$

$$x_{k_j} = a + k_j \frac{(b-a)}{T}, \qquad \forall\ j = 1,\ldots,d$$

The term $I_{k_1\cdots k_d}$, $1 \leq k_1,\ldots,k_d \leq T$ denotes a partition of $(a,b]^d$ which coincides Lebesgue a.e. with $(a,b]^d$. We want to approximate $\int_{(a,b]^d} m(x)dx$ by a Riemann sum:

$$\begin{aligned} &\Big(\frac{b-a}{T}\Big)^d \sum_{k_1,\ldots,k_d=1}^{T} m(x_{k_1},\ldots,x_{k_d}) \\ &\qquad = \sum_{k_1,\ldots,k_d=1}^{T} \int\limits_{I_{k_1\cdots k_d}} m(x_{k_1},\ldots,x_{k_d})\,\mathrm{d}x_1 \cdots \mathrm{d}x_d. \end{aligned} \tag{2.57}$$

By mean-value theorem and the Lipschitz continuity of ∇m

$$m(x) \leq m(y) + \nabla m(y)'(x-y) + \frac{L}{2}\|x-y\|^2$$

then

$$\Big|\int_{I_k}(m(x)-m(x_k))\,\mathrm{d}x\Big| \leq \Big|\int_{I_k}\nabla m(x_k)'(x-x_k)\,\mathrm{d}x\Big| + \frac{L}{2}\int_{I_k}\|x-x_k\|^2\,\mathrm{d}x. \quad (2.58)$$

With the change of variable $u = x - x_k$ ($u_i = x_i - x_{k_i}$), (2.58) can be bounded as

$$\Big|(-1)^d \int\limits_{(0,\frac{b-a}{T}]^d} \nabla m(x_k)'u\,\mathrm{d}u\Big| + \frac{dL(b-a)^2}{2T^2}\int\limits_{I_k}\mathrm{d}x.$$

Aditionally

$$\int\limits_{(0,\frac{b-a}{T}]^d} \nabla m(x_k)'u\,\mathrm{d}u = \frac{1}{2}\Big(\frac{b-a}{T}\Big)^{d+1}\sum_{i=1}^{d}\frac{\partial m(x_k)}{\partial x_i}.$$

Putting all this together

$$\Big|\int\limits_{I_k}(m(x)-m(x_k))\,\mathrm{d}x\Big| \leq \frac{1}{2}\Big(\frac{b-a}{T}\Big)^{d+1}\sum_{i=1}^{d}\frac{\partial m(x_k)}{\partial x_i} + \frac{dL}{2}\Big(\frac{b-a}{T}\Big)^2\int\limits_{I_k}\mathrm{d}x,$$

and using (2.57), we sum over $k = (k_1, \ldots, k_d)$

$$\Big|\int\limits_{(a,b]^d} m(x)\,\mathrm{d}x - \Big(\frac{b-a}{T}\Big)^d \sum_{k_1,\ldots,k_d=1}^{T} m(x_{k_1},\ldots,x_{k_d})\Big| \leq \frac{1}{2}\Big(\frac{b-a}{T}\Big)^{d+1}\sum_k\sum_{i=1}^{d}\frac{\partial m(x_k)}{\partial x_i} + \frac{dL}{2}\Big(\frac{b-a}{T}\Big)^2(b-a)^d.$$

By Lipschitz continuity of m, we have

$$\Big(\frac{b-a}{T}\Big)^d\sum_k\sum_{i=1}^{d}\frac{\partial m(x_k)}{\partial x_i} = \int\limits_{(a,b]^d}\sum_{i=1}^{d}\frac{\partial m}{\partial x_i}\,\mathrm{d}x + \mathcal{O}(\frac{L}{T}).$$

Finally,

$$\Big|\int\limits_{(a,b]^d} m(x)\,\mathrm{d}x - \Big(\frac{b-a}{T}\Big)^d \sum_{k_1,\ldots,k_d=1}^{T} m(x_{k_1},\ldots,x_{k_d})\Big| = \mathcal{O}(T^{-1})\int\limits_{(a,b]^d}\sum_{i=1}^{d}\frac{\partial m}{\partial x_i}\,\mathrm{d}x + \mathcal{O}(\frac{L}{T^2}).$$

□

Remark: If additionally $|m_n(x)| \leq Mk_1(n)$ is uniform in x with $M > 0$, then $\|\nabla m_n(x)\| \leq Mk_1(n)$. This implies, for a Lipschitz constant L_n, that $\|\nabla m_n(x) - \nabla m_n(y)\| \leq L_n\|x-y\| \leq L_n\sqrt{d}$. Besides, $\|\nabla m_n(x) - \nabla m_n(y)\| \leq \|\nabla m_n(x)\| + \|\nabla m_n(y)\| \leq 2Mk_1(n)$, we equal both bounds obtaining $L_n = 2Mk_1(n)/\sqrt{d}$.

3 Asymptotic Properties

In this chapter we derive asymptotic properties of the tapered Whittle estimator

$$\widehat{\theta}_T^{(h)} := \arg\min_{\theta\in\Theta} \mathcal{L}_T^{(h)}(\theta),$$

where

$$\mathcal{L}_T^{(h)}(\theta) := \frac{1}{(2\pi T)^d} \sum_{t\in D_T} \int_{\Pi^d} \Big\{ h^2(\tfrac{t-1/2}{T}) \log[(2\pi)^d f_\theta(\tfrac{t-1/2}{T},\lambda)] + \frac{J_T^{(h)}(\frac{t-1/2}{T},\lambda)}{f_\theta(\frac{t-1/2}{T},\lambda)} \Big\} \,\mathrm{d}\lambda,$$

with $J_T^{(h)}$ corrected for skewness, and of the estimator

$$\widetilde{\theta}_T := \arg\min_{\theta\in\Theta} \mathcal{L}_T^{(ex)}(\theta),$$

where the exact Gaussian log-likelihood is given by

$$\mathcal{L}_T^{(ex)}(\theta) := \frac{1}{T^d} \log\det\Sigma_\theta + \frac{1}{T^d} X'\Sigma_\theta^{-1} X.$$

We prove consistency of these estimators, *i.e.*

$$\widehat{\theta}_T^{(h)} \xrightarrow{p} \theta_0 \qquad \text{and} \qquad \widetilde{\theta}_T \xrightarrow{p} \theta_0,$$

where

$$\theta_0 := \arg\min_{\theta\in\Theta} \mathcal{L}^{(h)}(\theta)$$

with

$$\mathcal{L}^{(h)}(\theta) := \frac{1}{(2\pi)^d} \int_{[0,1]^d} h^2(u) \int_{\Pi^d} \Big\{ \log[(2\pi)^d f_\theta(u,\lambda)] + \frac{f(u,\lambda)}{f_\theta(u,\lambda)} \Big\} \,\mathrm{d}\lambda \,\mathrm{d}u,$$

and Gaussianity (in the case of the exact estimator, $\widetilde{\theta}_T$, the convergence holds assuming $h(u) = 1$ in $\mathcal{L}^{(h)}(\theta)$) . We keep the convention of Chapter 2 by dropping ρ from h_ρ, *i.e* $h_\rho(u) = h(u)$. Regarding the first property, we prove equicontinuity and uniform convergence of the likelihoods and their derivatives (up to order two). In order to prove the second property we use cumulants. We conclude this chapter commenting on some aspects of the Gaussian law in the presence of bias. We use ideas presented in Dahlhaus (2000).

3.1 Consistency

Definition 3.1.1 (Equicontinuity)**.** *A sequence of random variables $Z_T(\theta)$, $\theta \in \Theta$ is* equicontinuous in probability*, if for each $\eta > 0$ and $\epsilon > 0$ there exists a $\delta > 0$ such that*

$$\lim_{T\to\infty} \mathbb{P}(\sup_{|\theta_1-\theta_2|\le\delta} |Z_T(\theta_1) - Z_T(\theta_2)| > \eta) < \epsilon.$$

Lemma 3.1.1. *Suppose assumptions (A1), (A3), (A4) and (A5) hold, then $\mathcal{L}_T^{(h)}(\theta)$, $\nabla\mathcal{L}_T^{(h)}(\theta)$, $\nabla^2\mathcal{L}_T^{(h)}(\theta)$, $\mathcal{L}_T^{(ex)}(\theta)$, $\nabla\mathcal{L}_T^{(ex)}(\theta)$ and $\nabla^2\mathcal{L}_T^{(ex)}(\theta)$ are equicontinuous in probability.*

Proof. The mean value theorem guarantees that for $\bar{\theta} = \alpha\theta_1 + (1-\alpha)\theta_2$, for some $\alpha \in [0,1]$

$$\mathcal{L}_T(\theta_2) - \mathcal{L}_T(\theta_1) = (\theta_2 - \theta_1)'\nabla\mathcal{L}_T(\bar{\theta}), \tag{3.1}$$

where $\mathcal{L}_T(\theta)$ corresponds to the tapered Whittle or the exact likelihood. What needs to be shown is that the gradient $\nabla\mathcal{L}_T$ is bounded in probability, which is equivalent to prove it for each partial derivative ∇_i, $i = 1, \ldots, p$.
Let us start examining the tapered Whittle likelihood case. Straightforward calculations yield

$$\nabla_i\mathcal{L}_T^{(h)}(\theta) = \frac{1}{(2\pi T)^d} \operatorname{tr} U_T(h^2 f_\theta^{-1}\nabla_i f_\theta) + \frac{1}{(2\pi T)^d} X'\widetilde{U}_T^{(h)}((2\pi)^{-d}\nabla_i f_\theta^{-1})X. \tag{3.2}$$

The uniform continuity and non-randomness of f_θ guarantee that the first term in the right-hand side of (3.2) exists and it is bounded in probability. Regarding the second term we obtain

$$\frac{1}{(2\pi T)^d} X'\widetilde{U}_T^{(h)}((2\pi)^{-d}\nabla_i f_\theta^{-1})X \le \frac{1}{(2\pi T)^d} X'X \|\widetilde{U}_T^{(h)}((2\pi)^{-d}\nabla_i f_\theta^{-1})\|_{op}.$$

Using Lemma 2.2.1 we get

$$\begin{aligned} \|\widetilde{U}_T^{(h)}((2\pi)^{-d}\nabla_i f_\theta^{-1})\|_{op} &= \|\widetilde{U}_T((2\pi)^{-d}h^2\nabla_i f_\theta^{-1})\|_{op} \\ &\le \sup_{(u,\lambda)\in[0,1]^d\times\Pi^d} |h^2(u)\nabla_i f_\theta^{-1}(u,\lambda)| + o(1). \end{aligned}$$

The boundedness follows from the differentiability assumptions with respect to θ.
Regarding the quantity $\frac{1}{(2\pi T)^d}X'X$, we note that

$$\begin{aligned} \frac{1}{(2\pi T)^d}\mathrm{E}\, X'X &= \frac{1}{(2\pi T)^d} \operatorname{tr}\mathrm{E}\, X'X \\ &= \frac{1}{(2\pi T)^d} \operatorname{tr}\Sigma \\ &= \int_{[0,1]^d}\int_{\Pi^d} f(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-1}), \end{aligned} \tag{3.3}$$

and for the variance

$$\begin{aligned}\mathrm{Var}\,(\frac{1}{(2\pi T)^d}X'X) &= \frac{1}{(2\pi T)^{2d}}\mathrm{Var}\,(X'X)\\ &= \frac{1}{(2\pi T)^{2d}}\,\mathrm{tr}(\Sigma^2)\\ &= \frac{1}{(2\pi T)^d}\Big[\int\limits_{[0,1]^d}\int\limits_{\Pi^d} f^2(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-1})\Big]. \qquad (3.4)\end{aligned}$$

where we have used Lemma 2.3.4. Summarizing, the expectation is bounded and the variance converge to zero, therefore $\frac{1}{(2\pi T)^d}X'X$ is bounded in probability (see Proposition 3.1.1). Taking supremum in (3.1) we obtain

$$\sup_{|\theta_1-\theta_2|\le\delta} |\mathcal{L}_T(\theta_2) - \mathcal{L}_T(\theta_1)| \le \delta\Big(K_1 + \frac{K_2}{T^d}X'X\Big), \qquad (3.5)$$

with K_1, K_2 constants. Thus, the equicontinuity of $\mathcal{L}_T^{(h)}$ is proved. For proving equicontinuity of the exact likelihood and their first derivatives (including first derivative of the Whittle likelihood) the argument is basically the same, we only have to verify the boundedness of the derivative on the right-hand side of (3.1).
Using Lemma 2.3.1 (ii), the derivative of the exact likelihood with respect to θ_i is

$$\nabla_i\mathcal{L}_T^{(ex)}(\theta) = \frac{1}{T^d}\Big[\,\mathrm{tr}\{\Sigma_\theta^{-1}\nabla_i\Sigma_\theta\} + X'\nabla_i\Sigma_\theta^{-1}X\Big].$$

Regarding the first term, from Lemma 2.2.3 this is equal to

$$\frac{1}{T^d}\,\mathrm{tr}\,\Big\{\Sigma_\theta^{-1}\Sigma_\theta(\nabla_i f_\theta)\Big\} = \frac{1}{(2\pi)^d}\int\limits_{[0,1]^d}\int\limits_{\Pi^d} f_\theta^{-1}\nabla_i f_\theta\,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-1}),$$

which is bounded because of the regularity assumptions. The second term can be bounded as

$$\frac{1}{T^d}X'\nabla_i\Sigma_\theta^{-1}X \le \frac{1}{T^d}X'X\|\nabla_i\Sigma_\theta^{-1}\|_{op}.$$

Using Lemma 2.3.1 (vii), Lemma 2.2.2, Lemma 2.2.1 the norm on the right-hand side can be written as

$$\begin{aligned}\|\nabla_i\Sigma_\theta^{-1}\|_{op} = \|-\Sigma_\theta^{-1}\nabla_i\Sigma_\theta\Sigma_\theta^{-1}\|_{op} &= \|\Sigma_\theta^{-1}\Sigma_\theta(\nabla_i f_\theta)\Sigma_\theta^{-1}\|_{op}\\ &\le \|\Sigma_\theta^{-1}\|_{op}^2\|\Sigma_\theta(\nabla_i f_\theta)\|_{op}\\ &< \infty.\end{aligned}$$

Hence the equicontinuity follows from (3.3) and (3.4). The results of equicontinuity for $\nabla\mathcal{L}_T^{(h)}(\theta)$ and $\nabla\mathcal{L}_T^{(ex)}(\theta)$ follows exactly in the same way given the second order differentiablity and uniform continuity conditions assumed.
In order to prove the equicontinuity of $\nabla_{ij}^2\mathcal{L}_T^{(h)}(\theta)$ we follow a different approach since

we have assumed only second order continuous differentiability. It is enough to verify the equicontinuity of each term separately. Differentiating in (3.2) we obtain

$$\nabla^2_{ij}\mathcal{L}_T^{(h)}(\theta) = \frac{1}{(2\pi T)^d}\Big[\operatorname{tr} U_T(h^2\nabla_i f_\theta^{-1}\nabla_j f_\theta) + \operatorname{tr} U_T(h^2 f_\theta^{-1}\nabla^2_{ij} f_\theta) + X' U_T^{(h)}((2\pi)^{-d}\nabla^2_{ij} f_\theta^{-1})X\Big]. \quad (3.6)$$

The first two terms in brackets are deterministic and uniformly continuous in θ, hence equicontinuous in probability. We study now the third term. Its equicontinuity would imply the equicontinuity of the second derivative.

$$\frac{1}{(2\pi T)^d}|X'U_T^{(h)}((2\pi)^{-d}\nabla^2_{ij} f_{\theta_1}^{-1})X - X'U_T^{(h)}((2\pi)^{-d}\nabla^2_{ij} f_{\theta_2}^{-1})X| \le \frac{1}{(2\pi T)^d}X'X \times \|U_T^{(h)}((2\pi)^{-d}\nabla^2_{ij} f_{\theta_1}^{-1}) - U_T^{(h)}((2\pi)^{-d}\nabla^2_{ij} f_{\theta_2}^{-1})\|_{op}.$$

About the normed term, the Lemma 2.2.1 implies

$$\|U_T((2\pi)^{-d}h^2\nabla^2_{ij}(f_{\theta_1}^{-1} - f_{\theta_2}^{-1}))\|_{op} \le \sup_{(u,\lambda)\in[0,1]^d\times\Pi^d}|h^2\nabla^2_{ij}(f_{\theta_1}^{-1} - f_{\theta_2}^{-1})| + o(1).$$

The regularity assumptions with respect to θ imply that for $\epsilon > 0$, there exists a T_0 and δ such that

$$\begin{aligned}\sup_{|\theta_1-\theta_2|<\delta}&\|U_T((2\pi)^{-d}h^2\nabla^2_{ij}(f_{\theta_1}^{-1} - f_{\theta_2}^{-1}))\|_{op}\\ &\le \sup_{|\theta_1-\theta_2|<\delta}\ \sup_{(u,\lambda)\in[0,1]^d\times\Pi^d}|h^2\nabla^2_{ij}(f_{\theta_1}^{-1} - f_{\theta_2}^{-1})| + o(1)\\ &< \epsilon \qquad \forall T \ge T_0\end{aligned}$$

since the term $o(1)$ depends on T (see Lemma 2.2.1). We have proved before that $X'X/(2\pi T)^d$ is bounded in probability which implies finally the equicontinuity of the third term, and hence of $\nabla^2_{ij}\mathcal{L}_T^{(h)}(\theta)$.

The arguments to prove equicontinuity of $\nabla^2_{ij}\mathcal{L}_T^{(ex)}(\theta)$ are the same as before, we only need to show the boundedness of the difference of derivatives evaluated in θ_1 and θ_2. It is easy to verify that

$$\nabla_j\mathcal{L}_T^{(ex)}(\theta) = \frac{1}{T^d}\Big[\operatorname{tr}\Big\{\Sigma_\theta^{-1}\Sigma_\theta(\nabla_j f_\theta)\Big\} - X'\Sigma_\theta^{-1}\Sigma_\theta(\nabla_j f_\theta)\Sigma_\theta^{-1}X\Big] \quad (3.7)$$

Applying Lemma 2.2.3 to the derivative of the first term with respect to θ_i we obtain

$$\nabla_i\frac{1}{T^d}\operatorname{tr}\Big(\Sigma_\theta^{-1}\Sigma_\theta(\nabla_j f_\theta)\Big) = \frac{1}{(2\pi)^d}\int_{[0,1]^d}\int_{\Pi^d}(\nabla_i f_\theta^{-1}\nabla_j f_\theta + f_\theta^{-1}\nabla^2_{ij} f_\theta)\,\mathrm{d}\lambda\,\mathrm{d}u.$$

Again, its equicontinuity is a consequence of the regularity conditions imposed on f_θ. The derivative of the second term yields

$$\nabla_i(X'\Sigma_\theta^{-1}\Sigma_\theta(\nabla_j f_\theta)\Sigma_\theta^{-1}X) = X'\Big(-\Sigma_\theta^{-1}\Sigma_\theta(\nabla_i f_\theta)\Sigma_\theta^{-1}\Sigma_\theta(\nabla_j f_\theta)\Sigma_\theta^{-1}$$

$$+\Sigma_\theta^{-1}\Sigma_\theta(\nabla^2_{ij}f_\theta)\Sigma_\theta^{-1} - \Sigma_\theta^{-1}\Sigma_\theta(\nabla_j f_\theta)\Sigma_\theta^{-1}\Sigma_\theta(\nabla_i f_\theta)\Sigma_\theta^{-1}\Big) X.$$

We denote this expression as $D\mathcal{L}(\theta)$. Thus,

$$\begin{aligned}\frac{1}{T^d}|D\mathcal{L}(\theta_1) - D\mathcal{L}(\theta_2)| \leq \frac{1}{T^d}X'X \Big\{ &\Big\| \Sigma_{\theta_1}^{-1}\Sigma_\theta(\nabla_i f_{\theta_1})\Sigma_{\theta_1}^{-1}\Sigma_\theta(\nabla_j f_{\theta_1})\Sigma_{\theta_1}^{-1} \\ &-\Sigma_{\theta_2}^{-1}\Sigma_\theta(\nabla_i f_{\theta_2})\Sigma_{\theta_2}\Sigma_\theta(\nabla_j f_{\theta_2})\Sigma_{\theta_2}^{-1}\Big\|_{op} \\ &+\|\Sigma_{\theta_1}^{-1}\Sigma_\theta(\nabla^2_{ij}f_{\theta_1})\Sigma_{\theta_1}^{-1} - \Sigma_{\theta_2}^{-1}\Sigma_\theta(\nabla^2_{ij}f_{\theta_2})\Sigma_{\theta_2}^{-1}\|_{op} \\ &+\Big\|\Sigma_{\theta_1}^{-1}\Sigma_\theta(\nabla_j f_{\theta_1})\Sigma_{\theta_1}^{-1}\Sigma_\theta(\nabla_i f_{\theta_1})\Sigma_{\theta_1}^{-1} \\ &-\Sigma_{\theta_2}^{-1}\Sigma_\theta(\nabla_j f_{\theta_2})\Sigma_{\theta_2}^{-1}\Sigma_\theta(\nabla_i f_{\theta_2})\Sigma_{\theta_2}^{-1}\Big\|_{op}\Big\}.\end{aligned}$$

The expression in curly brackets turns out to be bounded by applying several times the triangular inequality, Lemma 2.2.1 and the regularity assumptions. Finally, the result follows from the boundedness of $|X'X|/(2\pi T)^d$ in probability. □

Proposition 3.1.1. *If $E X_n \to \mu$ and $Var X_n \to 0$, then $X_n \xrightarrow{p} \mu$*

Proof. Let $\epsilon > 0$. Chebyshev's inequality yields

$$\mathbb{P}(|X_n - \mu| > \epsilon) \leq \frac{\operatorname{Var} X_n}{\epsilon^2} + \frac{(\operatorname{E} X_n - \mu)^2}{\epsilon^2},$$

which shows the assertion when $n \to \infty$. □

Theorem 3.1.1. *Suppose assumptions (A1), (A3), (A4) and (A5) hold, then we have for $k = 0, 1, 2$*

(i) $\sup_{\theta\in\Theta} |\nabla^k(\mathcal{L}_T^{(h)}(\theta) - \mathcal{L}^{(h)}(\theta))| \xrightarrow{p} 0$

(ii) $\sup_{\theta\in\Theta} |\nabla^k(\mathcal{L}_T^{(ex)}(\theta) - \mathcal{L}^{(1)}(\theta))| \xrightarrow{p} 0$

(iii) If $h(u) = h_T(u) \to 1$ for $u \in [0,1]^d$, then

$$\sup_{\theta\in\Theta} |\nabla^k(\mathcal{L}_T^{(ex)}(\theta) - \mathcal{L}^{(h)}(\theta))| \xrightarrow{p} 0$$

Proof. In the following we use Proposition 3.1.1 to prove the pointwise consistency of the likelihoods and their derivatives. We have already proved in Lemma 3.1.1 the equicontinuity of the likelihoods and their derivatives which in addition yields the desired result. Summarizing, we need to show that the expectation of every non-deterministic likelihood converges to its deterministic counterpart and that their respective variance vanishes for $T \to \infty$. The same steps are valid for the derivatives.

(i) Let $k = 0$.

Recall that

$$\mathcal{L}_T^{(h)}(\theta) = \frac{1}{(2\pi T)^d} \sum_{t\in D_T} \int_{\Pi^d} \Big\{ h^2(\tfrac{t-1/2}{T}) \log[(2\pi)^d f_\theta(\tfrac{t-1/2}{T},\lambda)] + \frac{J_T^{(h)}(\frac{t-1/2}{T},\lambda)}{f_\theta(\frac{t-1/2}{T},\lambda)} \Big\} \,\mathrm{d}\lambda$$

$$\mathcal{L}^{(h)}(\theta) = \frac{1}{(2\pi)^d} \int_{[0,1]^d} h^2(u) \int_{\Pi^d} \Big\{ \log[(2\pi)^d f_\theta(u,\lambda)] + \frac{f(u,\lambda)}{f_\theta(u,\lambda)} \Big\} \,\mathrm{d}\lambda\,\mathrm{d}u$$

The term $\mathcal{L}_T^{(h)}(\theta) - \mathcal{L}^{(h)}(\theta)$ consists of two differences, the logarithmic and preperiodogram terms. Assumption (A3), (A5) and Proposition (2.2.3) guarantee the first term to have a rate $\mathcal{O}(T^{-1})$. From Chapter 2 we know that

$$\frac{1}{(2\pi T)^d} \sum_{t\in D_T} \int_{\Pi^d} \frac{J_T^{(h)}(\frac{t-1/2}{T},\lambda)}{f_\theta(\frac{t-1/2}{T},\lambda)} \,\mathrm{d}\lambda = \frac{1}{(2\pi T)^d} X' \widetilde{U}_T^{(h)}((2\pi)^{-d} f_\theta^{-1}) X.$$

The expectation of this term is

$$\begin{aligned}\frac{1}{(2\pi T)^d} \mathrm{E}\, X' \widetilde{U}_T^{(h)}((2\pi)^{-d} f_\theta^{-1}) X &= \frac{1}{(2\pi T)^d} \operatorname{tr} \widetilde{U}_T^{(h)}((2\pi)^{-d} f_\theta^{-1}) \Sigma(A) \\ &= \frac{1}{(2\pi)^d} \int_{[0,1]^d} \int_{\Pi^d} h^2(u) \frac{f(u,\lambda)}{f_\theta(u,\lambda)} \,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-1}),\end{aligned}$$

where we have used Lemma 2.2.2. Consequently, $\mathrm{E}\,\mathcal{L}_T^{(h)}(\theta) = \mathcal{L}^{(h)}(\theta) + \mathcal{O}(T^{-1})$. In order to calculate the variance it is enough to invoke Lemmas 2.3.4 and 2.2.2 as follows

$$\begin{aligned}\operatorname{Var}\Big[\frac{1}{(2\pi T)^d} X' \widetilde{U}_T^{(h)}((2\pi)^{-d} f_\theta^{-1}) X\Big] &= \frac{2}{(2\pi T)^{2d}} \operatorname{tr}\Big[\Sigma(A) \widetilde{U}_T^{(h)}((2\pi)^{-d} f_\theta^{-1})\Big]^2 \\ &= \frac{2}{(2\pi T)^d} \Big[\int_{[0,1]^d} \int_{\Pi^d} h^4(u) \frac{f^2(u,\lambda)}{f_\theta^2(u,\lambda)} \,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-1}) \Big].\end{aligned}$$

Since the first part of $\mathcal{L}_T^{(h)}(\theta)$ is deterministic, its variance turns out to be

$$\operatorname{Var} \mathcal{L}_T^{(h)}(\theta) = \mathcal{O}(T^{-d}).$$

The regularity conditions imposed on $f(u,\lambda)$ and $f_\theta(u,\lambda)$ guarantee the uniform continuity of $\mathcal{L}^{(h)}(\theta)$. Moreover this term is deterministic and hence equicontinuous, which concludes our proof for $k = 0$.

For $k = 1$ we take derivatives of $\mathcal{L}_T^{(h)}(\theta)$ and expectations, then

$$\mathrm{E}\,\nabla_i \mathcal{L}_T^{(h)}(\theta) = \frac{1}{(2\pi T)^d} \{ \operatorname{tr} U_T^{(h)}(\tfrac{\nabla_i f_\theta}{f_\theta}) + \operatorname{tr} \widetilde{U}_T^{(h)}((2\pi)^{-d} \nabla_i f_\theta^{-1}) \Sigma(A) \}$$

$$= \frac{1}{(2\pi)^d}\Big[\int\limits_{[0,1]^d}\int\limits_{\Pi^d} h^2(u)\frac{\nabla_i f_\theta}{f_\theta}\,\mathrm{d}\lambda\,\mathrm{d}u + \int\limits_{[0,1]^d}\int\limits_{\Pi^d} h^2(u)\nabla_i \frac{f(u,\lambda)}{f_\theta(u,\lambda)}\,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-1})\Big]$$
$$= \nabla_i \mathcal{L}^{(h)}(\theta) + \mathcal{O}(T^{-1}).$$

Using the same arguments as for the case $k = 0$, we obtain

$$\begin{aligned}\operatorname{Var} \nabla_i \mathcal{L}_T^{(h)}(\theta) &= \frac{1}{(2\pi T)^d}\Big[\int\limits_{[0,1]^d}\int\limits_{\Pi^d}\{h^2(u)f(u,\lambda)\nabla_i f_\theta^{-1}\}^2\,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-1})\Big] \\ &= \mathcal{O}(T^{-d}).\end{aligned}$$

Regarding the case $k = 2$ we calculate first the second derivatives with respect to θ_i, θ_j.

$$\begin{aligned}\nabla_{ij}^2 \mathcal{L}_T^{(h)}(\theta) = \frac{1}{(2\pi T)^d}\Big[&\operatorname{tr} U_T(h^2 \nabla_i f_\theta^{-1} \nabla_j f_\theta^{-1}) + \operatorname{tr} U_T(h^2 f_\theta^{-1} \nabla_{ij} f_\theta) \\ &+ X' U_T(h^2 (2\pi)^{-d} \nabla_{ij}^2 f_\theta^{-1}) X\Big].\end{aligned}$$

We calculate the expectation only of the third term in brackets (the other two are deterministic) obtaining

$$\begin{aligned}\frac{1}{(2\pi T)^d}\mathrm{E}\, X' U_T(h^2 (2\pi)^{-d} \nabla_{ij}^2 f_\theta^{-1}) X &= \frac{1}{(2\pi T)^d}\operatorname{tr} U_T(h^2 (2\pi)^{-d} \nabla_{ij}^2 f_\theta^{-1}) \Sigma(A) \\ &= \frac{1}{(2\pi)^d}\int\limits_{[0,1]^d}\int\limits_{\Pi^d} h^2(u) f(u,\lambda) \nabla_{ij}^2 f_\theta^{-1}(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u \\ &\quad + \mathcal{O}(T^{-1}).\end{aligned}$$

Thus, $\mathrm{E}\,\nabla_{ij}^2 \mathcal{L}_T^{(h)}(\theta) = \nabla_{ij}^2 \mathcal{L}^{(h)}(\theta) + \mathcal{O}(T^{-1})$.
Finally, we calculate the variance of the third term analogously to the case $k = 0$ obtaining

$$\begin{aligned}\operatorname{Var} \nabla_{ij}^2 \mathcal{L}_T^{(h)}(\theta) &= \frac{1}{(2\pi T)^d}\Big[\int\limits_{[0,1]^d}\int\limits_{\Pi^d} h^4(u)\{f(u,\lambda)\nabla_{ij}^2 f_\theta(u,\lambda)\}^2\,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-1})\Big] \\ &= \mathcal{O}(T^{-d}).\end{aligned}$$

(ii) Let $k = 0$.

Recall that

$$\begin{aligned}\mathcal{L}_T^{(ex)}(\theta) &= \frac{1}{T^d}[\log\det \Sigma_\theta + X' \Sigma_\theta^{-1} X] \\ \mathcal{L}^{(1)}(\theta) &= \frac{1}{(2\pi)^d}\int\limits_{[0,1]^d}\int\limits_{\Pi^d}\Big\{\log[(2\pi)^d f_\theta(u,\lambda)] + \frac{f(u,\lambda)}{f_\theta(u,\lambda)}\Big\}\,\mathrm{d}\lambda\,\mathrm{d}u\end{aligned}$$

The Proposition 2.2.1 yields

$$\frac{1}{T^d}\log\det\Sigma_\theta - \frac{1}{(2\pi)^d}\int\limits_{[0,1]^d}\int\limits_{\Pi^d}\log[(2\pi)^d f_\theta(u,\lambda)]\,\mathrm{d}\lambda\,\mathrm{d}u = \mathcal{O}(T^{-1}). \tag{3.8}$$

Consider the matrix $U_T((2\pi)^{-2d}f_\theta^{-1})$. Using Lemma 2.2.2, we note that

$$\begin{aligned}\mathrm{E}\,\frac{1}{T^d}X'U_T((2\pi)^{-2d}f_\theta^{-1})X &= \frac{1}{T^d}\operatorname{tr}U_T((2\pi)^{-2d}f_\theta^{-1})\Sigma(A)\\ &= \frac{1}{(2\pi)^d}\int\limits_{[0,1]^d}\int\limits_{\Pi^d}\frac{f(u,\lambda)}{f_\theta(u,\lambda)}\,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-1}). \qquad (3.9)\end{aligned}$$

Combining expressions (3.8) and (3.9), the following expectation yields

$$\begin{aligned}\mathrm{E}\left(\mathcal{L}_T^{(ex)}(\theta) - \mathcal{L}^{(1)}(\theta)\right) &= \mathrm{E}\,\frac{1}{T^d}X'\{\Sigma_\theta^{-1} - U_T((2\pi)^{-2d}f_\theta^{-1})\}X + \mathcal{O}(T^{-1})\\ &= \frac{1}{T^d}\operatorname{tr}\Sigma_\theta^{-1}\Sigma(A) - \frac{1}{T^d}\operatorname{tr}U_T((2\pi)^{-2d}f_\theta^{-1})\Sigma(A) + \mathcal{O}(T^{-1})\\ &= \mathcal{O}(T^{-1}).\end{aligned}$$

where we have used Lemmas 2.2.2 and 2.2.3 in the second line. Thus, $\mathrm{E}\,\mathcal{L}_T^{(ex)}(\theta) = \mathcal{L}^{(1)}(\theta) + \mathcal{O}(T^{-1})$. Analogously, the variance turns out to be

$$\begin{aligned}\mathrm{Var}\left(\mathcal{L}_T^{(ex)}(\theta) - \mathcal{L}^{(1)}(\theta)\right) &= \frac{1}{T^{2d}}\mathrm{Var}\,X'\{\Sigma_\theta^{-1} - U_T((2\pi)^{-2d}f_\theta^{-1})\}X\\ &= \frac{1}{T^{2d}}\operatorname{tr}\left(\{\Sigma_\theta^{-1} - U_T((2\pi)^{-2d}f_\theta^{-1})\}\Sigma(A)\right)^2\\ &= \mathcal{O}(T^{-d-1}).\end{aligned}$$

We move on to the case $k=1$. Lemma 2.3.1 (ii) yields an expression for the derivative of a determinant, thus

$$\nabla_i\mathcal{L}_T^{(ex)} = \frac{1}{T^d}[\operatorname{tr}\Sigma_\theta^{-1}\nabla_i\Sigma_\theta + X'\nabla_i\Sigma_\theta^{-1}X].$$

Applying Lemma 2.2.3, the first term in brackets yields

$$\begin{aligned}\frac{1}{T^d}\operatorname{tr}\{\Sigma_\theta^{-1}\nabla_i\Sigma_\theta\} &= \frac{1}{T^d}\operatorname{tr}\{\Sigma_\theta^{-1}\Sigma_\theta(\nabla_i f_\theta)\}\\ &= \frac{1}{(2\pi)^d}\int\limits_{[0,1]^d}\int\limits_{\Pi^d} f_\theta^{-1}(u,\lambda)\nabla_i f_\theta(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u + \mathcal{O}(T^{-1}),\end{aligned}$$

and so

$$\nabla_i\mathcal{L}_T^{(ex)}(\theta) - \nabla_i\mathcal{L}^{(1)}(\theta) = \frac{1}{T^d}X'\nabla_i\Sigma_\theta^{-1}X - \frac{1}{(2\pi)^d}\int\limits_{[0,1]^d}\int\limits_{\Pi^d} f(u,\lambda)\nabla_i f_\theta^{-1}(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u$$

$$+\mathcal{O}(T^{-1}).$$

The non-deterministic term in the right-hand side of this equation has expectation given by

$$\begin{aligned} \mathrm{E}\,\frac{1}{T^d}X'\nabla_i\Sigma_\theta^{-1}X &= -\frac{1}{T^d}\,\mathrm{tr}\,\Sigma_\theta^{-1}\Sigma_\theta(\nabla_i f_\theta)\Sigma_\theta^{-1}\Sigma(A) \\ &= -\frac{1}{(2\pi)^d}\int\limits_{[0,1]^d}\int\limits_{\Pi^d} f_\theta^{-2}(u,\lambda)\nabla_i f_\theta(u,\lambda)f(u,\lambda)\,\mathrm{d}\lambda\,\mathrm{d}u+\mathcal{O}(T^{-1}). \end{aligned}$$

Note that $\nabla_i f_\theta^{-1} = -f_\theta^{-2}\nabla_i f_\theta$, therefore $\mathrm{E}\,\nabla_i\mathcal{L}_T^{(ex)}(\theta) = \nabla_i\mathcal{L}^{(1)}(\theta)+\mathcal{O}(T^{-1})$. Finally we calculate the variance

$$\begin{aligned} \mathrm{Var}\Big(\nabla_i\mathcal{L}_T^{(ex)}(\theta)-\nabla_i\mathcal{L}^{(1)}(\theta)\Big) &= \frac{1}{T^{2d}}\mathrm{Var}\,X'\nabla_i\Sigma_\theta^{-1}X \\ &= \frac{1}{T^{2d}}\,\mathrm{tr}\Big(\Sigma_\theta^{-1}\Sigma_\theta(\nabla_i f_\theta)\Sigma_\theta^{-1}\Sigma(A)\Big)^2 \\ &= \frac{1}{T^d}\Big[\frac{1}{(2\pi)^d}\int\limits_{[0,1]^d}\int\limits_{\Pi^d} f^2(u,\lambda)f_\theta^{-4}(u,\lambda)(\nabla_i f_\theta(u,\lambda))^2\,\mathrm{d}\lambda\,\mathrm{d}u+\mathcal{O}(T^{-1})\Big] \\ &= \mathcal{O}(T^{-d}). \end{aligned}$$

As before, the equicontinuity of $\mathcal{L}_T^{(ex)}$ and $\mathcal{L}^{(1)}$ imply the uniform consistency. The result for $k=2$ follows in the same way.

(iii) Triangular inequality yields for $k=0,1,2$

$$|\nabla^k(\mathcal{L}_T^{(ex)}(\theta)-\mathcal{L}^{(h)}(\theta))| \le |\nabla^k(\mathcal{L}_T^{(ex)}(\theta)-\mathcal{L}^{(1)}(\theta))|+|\nabla^k(\mathcal{L}^{(1)}(\theta)-\mathcal{L}^{(h)}(\theta))|. \quad (3.10)$$

The item (ii) above yields the convergence in probability of the first term on the right-hand side of (3.10). Particularly for $k=0$, the second term can be bounded as follows

$$|\mathcal{L}^{(h)}-\mathcal{L}^{(1)}| \le \frac{1}{(2\pi)^d}\sup_{u\in[0,1]^d}\int\limits_{\Pi^d}\Big|\log[(2\pi)^d f_\theta(u,\lambda)]+\frac{f(u,\lambda)}{f_\theta(u,\lambda)}\Big|\,\mathrm{d}\lambda\int\limits_{[0,1]^d}|h(u)-1|\,\mathrm{d}u$$

From the assumptions on $f(u,\lambda)$ and $f_\theta(u,\lambda)$ the first integral is bounded. In order to reduce the variance and control the bias, in Chapter 2 we assumed a relation between ρ and T, namely $\rho=\mathcal{O}(T^{-\beta})$, $\beta\in(0,1/6]$. From (2.36) follows that

$$\int\limits_{[0,1]^d}|h(u)-1|\,\mathrm{d}u=\mathcal{O}(\rho),$$

which implies $\mathcal{L}^{(h)}=\mathcal{L}^{(1)}+\mathcal{O}(T^{-\beta})$, and consequently $\mathrm{E}\,\mathcal{L}_T^{(ex)}=\mathcal{L}^{(h)}+\mathcal{O}(T^{-\beta})$. The result follows from the convergence of the variance of $\mathcal{L}_T^{(ex)}$ to 0 proved in (ii). The arguments for the cases $k=1,2$ are the same. □

From Lemma 3.1.1 and Theorem 3.1.1, the consistency follows straightforwardly.

Theorem 3.1.2 (Consistency). *We assume the conditions (A1), (A2), (A3), (A4) and (A5) hold. Additionally, assume that the taper function* $h(u) = h_T(u)$ *converge to one, then*

$$\widehat{\theta}_T^{(h)} \xrightarrow{p} \theta_0 \qquad \text{and} \qquad \widetilde{\theta}_T \xrightarrow{p} \theta_0.$$

Proof. To shorten notation let $\mathcal{L}(\theta) := \mathcal{L}^{(1)}(\theta)$. It is straightforward to note that

$$\begin{aligned} \mathcal{L}(\widehat{\theta}_T^{(h)}) - \mathcal{L}(\theta_0) &\geq 0 \\ \mathcal{L}_T^{(h)}(\theta_0) - \mathcal{L}_T^{(h)}(\widehat{\theta}_T^{(h)}) &\geq 0. \end{aligned}$$

Therefore,

$$|\mathcal{L}(\widehat{\theta}_T^{(h)}) - \mathcal{L}(\theta_0)| \leq |\mathcal{L}(\widehat{\theta}_T^{(h)}) - \mathcal{L}_T^{(h)}(\widehat{\theta}_T^{(h)})| + |\mathcal{L}_T^{(h)}(\theta_0) - \mathcal{L}(\theta_0)|. \tag{3.11}$$

Since $h(u) = h_T(u) \to 1$ as $T \to \infty$, Theorem 3.1.1 guarantees that each term on the right hand side (3.11) converges to zero. Therefore, the result follows from the compactness assumption on Θ. The consistency of $\widetilde{\theta}_T$ follows analogously. □

3.2 Gaussianity

In the last section we have proved the consistency of $\widehat{\theta}_T^{(h)}$ and $\widetilde{\theta}_T$. Now we move on to verify the Gaussianity of these estimators. In Chapter 2 we proved that the score function of $\mathcal{L}_T^{(h)}(\theta)$ has a bias of order $\mathcal{O}(T^{-1})$ which implies that our estimator $\widehat{\theta}_T^{(h)}$ has also a bias of the same order. Furthermore, the bias has three different sources, two of which can be reduced to $\mathcal{O}(T^{-2})$ by a simple change in the preperiodogram and by introducing taper functions. However, the bias produced by the approximation (1.3) persists but it might be reduced by considering a more precise approximation of $A^0_{t,T}(\lambda)$, *e.g.* $A(u,\lambda) + \frac{1}{T}B(u,\lambda)$, as exemplified in Chapter 1. Unfortunately, getting an expression for $B(u,\lambda)$ can be hard in practice and so an according bias correction would be difficult to implement. Taking this bias into account we prove a version of a Gaussian law, namely

$$T^{d/2}(\widehat{\theta}_T^{(h)} - \mathrm{E}\,\widehat{\theta}_T^{(h)}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Gamma_h^{-1} V_h \Gamma_h^{-1}),$$

where we center around the expectation instead of the limit parameter. The chosen technique for proving Gaussianity involves cumulants. In what follows we present this definition, some properties and a couple of results which play a crucial role in our proof.

Definition 3.2.1 (Cumulant). *The* rth *order joint cumulant,* $Cum(Y_1, \ldots, Y_r)$, *is given by the coefficient of* $i^r t_1 \cdots t_r$ *in the Taylor series expansion of* $log(\phi_Y(t_1, \ldots, t_r))$ *about the origin, where*

$$\phi_Y(t_1, \ldots, t_r) = E \exp\Big(i \sum_{j=1}^{r} Y_j t_j\Big)$$

corresponds to the joint characteristic function of $(Y_1, \ldots, Y_r)$.

It is a well known fact that the joint characteristic function of a Gaussian vector with parameters μ and Σ is

$$\phi_Y(t) = \exp(it'\mu - \tfrac{1}{2}t'\Sigma t),$$

therefore, from the definition, the joint cumulants of order higher than 2 vanish. The opposite claim is also true, *i.e.* the only distribution with vanishing higher order joint cumulants is the Gaussian (see Jammalamadaka et al. (2006)).
We list, without proof, some of the properties to be used (for details see Brillinger (1981), p. 19)

Properties 3.2.1. *Let* μ, $a_1, \ldots, a_l$ *be constants.*

(i) $Cum(a_1Y_1, \ldots, a_lY_l) = a_1 \cdots a_l Cum(Y_1, \ldots, Y_l)$.

(ii) For the random vector $(Z_1, Y_1, \ldots, Y_l)$,

$$Cum(Y_1 + Z_1, Y_2, \ldots, Y_l) = Cum(Y_1, Y_2, \ldots, Y_l) + Cum(Z_1, Y_2, \ldots, Y_l)$$

.

(iii) For $l = 2, 3, \ldots$, *it holds that* $Cum(Y_1 + \mu, \ldots, Y_l) = Cum(Y_1, \ldots, Y_l)$.

(iv) $Cum(Y_1, \ldots, Y_l)$ *is symmetric in its arguments.*

Definition 3.2.2 (hook, communicate, indecomposable, Brillinger (1981))**.** *Consider a (not necessarily rectangular) two-way table*

$$\begin{matrix} (1,1) & \cdots & (1,J_1) \\ \vdots & & \vdots \\ (I,1) & \cdots & (I,J_I) \end{matrix}$$

and a partition $P_1 \cup \cdots \cup P_M$ *of its entries. We shall say that sets* $P_{m'}, P_{m''}$*, of the partition,* ***hook*** *if there exist* $(i_1, j_1) \in P_{m'}$ *and* $(i_2, j_2) \in P_{m''}$ *such that* $i_1 = i_2$. *We shall say that the sets* $P_{m'}$ *and* $P_{m''}$ ***communicate*** *if there exists a sequence of sets* $P_{m_1} = P_{m'}$, $P_{m_2}, \ldots, P_{m_N} = P_{m''}$ *such that* P_{m_n} *and* $P_{m_{n+1}}$ *hook for* $n = 1, 2, \ldots, N-1$. *A partition is said to be* ***indecomposable*** *if all sets communicate.*

Theorem 3.2.1. *Consider a two-way array of random variables* X_{ij}; $j = 1, \ldots, J_i$; $i = 1, \ldots, I$. *Consider the* I *random variables*

$$Y_i = \prod_{j=1}^{J_i} X_{ij}, \quad i = 1, \ldots, I.$$

The joint cumulant $Cum(Y_1, \ldots, Y_I)$ *is then given by*

$$\sum_{\nu} Cum(X_{ij};\ (i,j) \in \nu_1) \cdots Cum(X_{ij};\ (i,j) \in \nu_p),$$

where the summation is over all indecomposable partitions $\nu = \nu_1 \cup \cdots \cup \nu_p$.

Proof. See proof of Theorem 2.3.2 in Brillinger (1981), pag. 392-393. □

Theorem 3.2.2 (Gaussian Law). *Suppose assumptions (A1), (A2), (A3), (A4) and (A5) hold, then we have*

(i) $T^{d/2}(\widehat{\theta}_T^{(h)} - E\widehat{\theta}_T^{(h)}) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Gamma_h^{-1} V_h \Gamma_h^{-1})$,

(ii) $T^{d/2}(\widetilde{\theta}_T - E\widetilde{\theta}_T) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Gamma_1^{-1} V_1 \Gamma_1^{-1})$,

where

$$(V_h)_{ij} = \frac{2}{(2\pi)^d} \int\limits_{[0,1]^d} \int\limits_{\Pi^d} h^4(u) f^2(u,\lambda) \nabla_i f_{\theta_0}^{-1}(u,\lambda) \nabla_j f_{\theta_0}^{-1}(u,\lambda)\, \mathrm{d}\lambda\, \mathrm{d}u,$$

$$(\Gamma_h)_{ij} = \frac{1}{(2\pi)^d} \int\limits_{[0,1]^d} \int\limits_{\Pi^d} h^2(u) \Big\{ (f - f_{\theta_0}) \nabla^2_{ij} f_{\theta_0}^{-1} + f_{\theta_0}^2 \nabla_i f_{\theta_0}^{-1} \nabla_j f_{\theta_0}^{-1} \Big\}\, \mathrm{d}\lambda\, \mathrm{d}u.$$

V_1 and Γ_1 corresponds to the nontapered versions of V_h and Γ_h respectively.

Proof. The mean value theorem guarantees

$$\nabla_i \mathcal{L}_T^{(h)}(\widehat{\theta}_T^{(h)}) - \nabla_i \mathcal{L}_T^{(h)}(\theta_0) = \{\nabla^2 \mathcal{L}_T^{(h)}(\theta_T^{(i)})(\widehat{\theta}_T^{(h)} - \theta_0)\}_i, \tag{3.12}$$

for some $\theta_T^{(i)}$ with $|\theta_T^{(i)} - \theta_0| \leq |\widehat{\theta}_T^{(h)} - \theta_0|$, $i = 1, \ldots, p$. Notice that if $\widehat{\theta}_T^{(h)}$ lies in the interior of Θ, then $\nabla \mathcal{L}_T^{(h)}(\widehat{\theta}_T^{(h)}) = 0$. On the other hand, if $\widehat{\theta}_T^{(h)}$ lies on the boundary of Θ, then for some $\delta > 0$, $|\widehat{\theta}_T^{(h)} - \theta_0| \geq \delta$, hence $\mathbb{P}(T^{d/2}|\nabla \mathcal{L}_T^{(h)}(\widehat{\theta}_T^{(h)})| \geq \epsilon) \leq \mathbb{P}(|\widehat{\theta}_T^{(h)} - \theta_0| \geq \delta) \to 0$ when T increases for all $\epsilon > 0$. Thus the term $\nabla_i \mathcal{L}_T^{(h)}(\widehat{\theta}_T^{(h)}) = 0$ for a large enough T. We obtain the result by proving

(i) $\nabla^2 \mathcal{L}_T^{(h)}(\theta_T^{(i)}) - \nabla^2 \mathcal{L}_T^{(h)}(\theta_0) \xrightarrow{p} 0$,

(ii) $\nabla^2 \mathcal{L}_T^{(h)}(\theta_0) \xrightarrow{p} \nabla^2 \mathcal{L}^{(h)}(\theta_0)$ and $\Gamma_h = \nabla^2 \mathcal{L}^{(h)}(\theta_0)$,

(iii) $T^{d/2}(\nabla \mathcal{L}_T^{(h)}(\theta_0) - \mathrm{E}\, \nabla \mathcal{L}_T^{(h)}(\theta_0)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, V_h)$.

From the consistency of $\widehat{\theta}_T^{(h)}$ we obtain that $\theta_T^{(i)} \xrightarrow{p} \theta_0$, thus, the result (i) follows from the equicontinuity of $\nabla^2 \mathcal{L}_T^{(h)}(\theta)$ proved in Lemma 3.1.1. The item (ii) is consequence of the uniform convergence of $\mathcal{L}_T^{(h)}(\theta)$ proved in Proposition 3.1.1. The matrix Γ_h is obtained by simply differentiating $\mathcal{L}^{(h)}(\theta)$ and rearranging terms. In order to prove (iii) we use the method of cumulants. The difference in (iii) has zero expectation. Thus, let us examine the covariance term.

$$\begin{aligned} &\mathrm{Cov}\,[T^{d/2}(\nabla_i \mathcal{L}_T^{(h)}(\theta_0) - \mathrm{E}\, \nabla_i \mathcal{L}_T^{(h)}(\theta_0)), T^{d/2}(\nabla_j \mathcal{L}_T^{(h)}(\theta_0) - \mathrm{E}\, \nabla_j \mathcal{L}_T^{(h)}(\theta_0))] \\ &\quad = T^d \mathrm{Cov}\,[\nabla_i \mathcal{L}_T^{(h)}(\theta_0), \nabla_j \mathcal{L}_T^{(h)}(\theta_0)]. \end{aligned}$$

Recall that from the definition of $\mathcal{L}_T^{(h)}$ and (2.32), $\nabla_i \mathcal{L}_T^{(h)}(\theta_0)$ can be written as

$$\nabla_i \mathcal{L}_T^{(h)}(\theta_0) = \frac{1}{(2\pi T)^d} \operatorname{tr} U_T(h^2 f_{\theta_0} \nabla_i f_{\theta_0}) + \frac{1}{(2\pi T)^d} X' \widetilde{U}_T^{(h)}((2\pi)^{-d} \nabla_i f_{\theta_0}^{-1}) X.$$

Then the covariance turns out to be

$$\begin{aligned} & T^d \mathrm{Cov}\, [\frac{1}{(2\pi T)^d} X' \widetilde{U}_T^{(h)}((2\pi)^{-d} \nabla_i f_{\theta_0}^{-1}) X, \frac{1}{(2\pi T)^d} X' \widetilde{U}_T^{(h)}((2\pi)^{-d} \nabla_j f_{\theta_0}^{-1}) X] \\ = \;& \frac{2}{(2\pi)^{2d}} \frac{1}{T^d} \operatorname{tr} [\widetilde{U}_T^{(h)}((2\pi)^{-d} \nabla_i f_{\theta_0}^{-1}) \Sigma(A) \widetilde{U}_T^{(h)}((2\pi)^{-d} \nabla_j f_{\theta_0}^{-1}) \Sigma(A)] \\ = \;& \frac{2}{(2\pi)^d} \int\limits_{[0,1]^d} \int\limits_{\Pi^d} h^4(u) f^2(u,\lambda) \nabla_i f_{\theta_0}^{-1}(u,\lambda) \nabla_j f_{\theta_0}^{-1}(u,\lambda)\, \mathrm{d}\lambda \mathrm{d}u + \mathcal{O}(T^{-1}) \end{aligned}$$

where we have used Lemmas 2.3.4 and 2.2.2. We examine now the joint cumulants of order higher than 2. Using Properties 3.2.1 (i) and (iii) note

$$\begin{aligned} & \mathrm{Cum}\, (T^{d/2}(\nabla_{i_1} \mathcal{L}_T^{(h)}(\theta_0) - \mathrm{E}\, \nabla_{i_1} \mathcal{L}_T^{(h)}(\theta_0)), \ldots, T^{d/2}(\nabla_{i_l} \mathcal{L}_T^{(h)}(\theta_0) - \mathrm{E}\, \nabla_{i_l} \mathcal{L}_T^{(h)}(\theta_0))) \\ & \quad = T^{dl/2} \mathrm{Cum}\, (\nabla_{i_1} \mathcal{L}_T^{(h)}(\theta_0), \ldots, \nabla_{i_l} \mathcal{L}_T^{(h)}(\theta_0)). \end{aligned} \tag{3.13}$$

Since $\nabla_i \mathcal{L}_T^{(h)}(\theta_0) = \frac{1}{(2\pi T)^d} X' \widetilde{U}_T^{(h)}((2\pi)^{-d} \nabla_i f_{\theta_0}^{-1}) X + C$, with C a constant, the Property 3.2.1 (iii) implies that (3.13) is equal to

$$\begin{aligned} & T^{dl/2} \mathrm{Cum}\, \Big(\frac{1}{(2\pi T)^d} X' \widetilde{U}_T^{(h)}((2\pi)^{-d} \nabla_{i_1} f_{\theta_0}^{-1}) X, \ldots, \frac{1}{(2\pi T)^d} X' \widetilde{U}_T^{(h)}((2\pi)^{-d} \nabla_{i_l} f_{\theta_0}^{-1}) X \Big) \\ & = \frac{1}{(2\pi)^{dl}} T^{-dl/2} \mathrm{Cum}\, (X' \widetilde{U}_T^{(h)}((2\pi)^{-d} \nabla_{i_1} f_{\theta_0}^{-1}) X, \ldots, X' \widetilde{U}_T^{(h)}((2\pi)^{-d} \nabla_{i_l} f_{\theta_0}^{-1}) X). \end{aligned}$$

Notice that each term in the cumulant corresponds to a quadratic form $X' A_j X$, $j = 1, \ldots, l$. Using Theorem 3.2.1 and Properties 3.2.1 we can reduce the cumulant to an expression involving traces of product of matrices, which can be bounded easily by means of Lemma 2.2.2.

In order to make the argument clear enough, let us consider the following example: Suppose we want to calculate $\mathrm{Cum}\, (X'AX, X'BX, X'CX)$ where each matrix, A, B and C, is $(n \times n)$-symmetric and $X \sim \mathcal{N}_n(0, \Sigma)$ like in our case. The Properties 3.2.1 yield

$$\mathrm{Cum}\, (X'AX, X'BX, X'CX) = \sum_{i,j} \sum_{k,l} \sum_{m,n} a_{ij} b_{kl} c_{mn} \mathrm{Cum}\, (X_i X_j, X_k X_l, X_m X_n),$$

where a_{ij}, b_{kl} and c_{mn} are the entries of the matrices A, B and C respectively. When using Theorem 3.2.1 over $\mathrm{Cum}\, (X_i X_j, X_k X_l, X_m X_n)$ we observe that many indecomposable partitions of the table

$$\begin{array}{cc} X_i & X_j \\ X_k & X_l \\ X_m & X_n \end{array}$$

generate vanishing cumulants, namely all $\mathrm{Cum}\,(X_1, \ldots, X_l)$ for $l > 2$ (since the X_i are Gaussian) and for all $i = 1, \ldots, n$, $\mathrm{Cum}\,(X_i) = \mathrm{E}\, X_i = 0$. Thus,

$$\begin{aligned}&\mathrm{Cum}\,(X_iX_j, X_kX_l, X_mX_n)\\ &\quad = \sum_{\nu} \mathrm{Cum}\,(X_{ij};\ (i,j) \in \nu_1)\mathrm{Cum}\,(X_{kl};\ (k,l) \in \nu_2)\mathrm{Cum}\,(X_{mn};\ (m,n) \in \nu_3)\end{aligned}$$

where each set ν_1, ν_2, ν_3 of the partition ν contains only two indices. Writing this by extension we obtain

$$\begin{aligned}
& \mathrm{Cum}\,(X'AX, X'BX, X'CX) \\
= & \sum_{i,j}\sum_{k,l}\sum_{m,n} a_{ij}b_{kl}c_{mn}\mathrm{Cum}\,(X_i, X_k)\mathrm{Cum}\,(X_j, X_m)\mathrm{Cum}\,(X_l, X_n) \\
+ & \sum_{i,j}\sum_{k,l}\sum_{m,n} a_{ij}b_{kl}c_{mn}\mathrm{Cum}\,(X_i, X_k)\mathrm{Cum}\,(X_j, X_n)\mathrm{Cum}\,(X_l, X_m) \\
+ & \sum_{i,j}\sum_{k,l}\sum_{m,n} a_{ij}b_{kl}c_{mn}\mathrm{Cum}\,(X_i, X_l)\mathrm{Cum}\,(X_j, X_m)\mathrm{Cum}\,(X_k, X_n) \\
+ & \sum_{i,j}\sum_{k,l}\sum_{m,n} a_{ij}b_{kl}c_{mn}\mathrm{Cum}\,(X_i, X_l)\mathrm{Cum}\,(X_j, X_n)\mathrm{Cum}\,(X_k, X_m) \\
+ & \sum_{i,j}\sum_{k,l}\sum_{m,n} a_{ij}b_{kl}c_{mn}\mathrm{Cum}\,(X_i, X_m)\mathrm{Cum}\,(X_j, X_k)\mathrm{Cum}\,(X_l, X_n) \\
+ & \sum_{i,j}\sum_{k,l}\sum_{m,n} a_{ij}b_{kl}c_{mn}\mathrm{Cum}\,(X_i, X_m)\mathrm{Cum}\,(X_j, X_l)\mathrm{Cum}\,(X_k, X_n) \\
+ & \sum_{i,j}\sum_{k,l}\sum_{m,n} a_{ij}b_{kl}c_{mn}\mathrm{Cum}\,(X_i, X_n)\mathrm{Cum}\,(X_j, X_k)\mathrm{Cum}\,(X_l, X_m) \\
+ & \sum_{i,j}\sum_{k,l}\sum_{m,n} a_{ij}b_{kl}c_{mn}\mathrm{Cum}\,(X_i, X_n)\mathrm{Cum}\,(X_j, X_l)\mathrm{Cum}\,(X_k, X_m) \\
= & \sum_{i,j}\sum_{k,l}\sum_{m,n} a_{ij}\mathrm{Cum}\,(X_j, X_m)c_{mn}\mathrm{Cum}\,(X_n, X_l)b_{lk}\mathrm{Cum}\,(X_k, X_i) \\
+ & \sum_{i,j}\sum_{k,l}\sum_{m,n} a_{ij}\mathrm{Cum}\,(X_j, X_n)c_{nm}\mathrm{Cum}\,(X_m, X_l)b_{lk}\mathrm{Cum}\,(X_k, X_i) \\
+ & \sum_{i,j}\sum_{k,l}\sum_{m,n} a_{ij}\mathrm{Cum}\,(X_j, X_m)c_{mn}\mathrm{Cum}\,(X_n, X_k)b_{kl}\mathrm{Cum}\,(X_l, X_i) \\
+ & \sum_{i,j}\sum_{k,l}\sum_{m,n} a_{ij}\mathrm{Cum}\,(X_j, X_n)c_{nm}\mathrm{Cum}\,(X_m, X_k)b_{kl}\mathrm{Cum}\,(X_l, X_i) \\
+ & \sum_{i,j}\sum_{k,l}\sum_{m,n} a_{ij}\mathrm{Cum}\,(X_j, X_k)b_{kl}\mathrm{Cum}\,(X_l, X_n)c_{nm}\mathrm{Cum}\,(X_m, X_i) \\
+ & \sum_{i,j}\sum_{k,l}\sum_{m,n} a_{ij}\mathrm{Cum}\,(X_j, X_l)b_{lk}\mathrm{Cum}\,(X_k, X_n)c_{nm}\mathrm{Cum}\,(X_m, X_i) \\
+ & \sum_{i,j}\sum_{k,l}\sum_{m,n} a_{ij}\mathrm{Cum}\,(X_j, X_k)b_{kl}\mathrm{Cum}\,(X_l, X_m)c_{mn}\mathrm{Cum}\,(X_n, X_i) \\
+ & \sum_{i,j}\sum_{k,l}\sum_{m,n} a_{ij}\mathrm{Cum}\,(X_j, X_l)b_{lk}\mathrm{Cum}\,(X_k, X_m)c_{mn}\mathrm{Cum}\,(X_n, X_i)
\end{aligned}$$

$$= 4\,\mathrm{tr}(A\Sigma C\Sigma B\Sigma) + 4\,\mathrm{tr}(A\Sigma B\Sigma C\Sigma),$$

where we have used the symmetry of the matrices, Properties 3.2.1 (iv), the properties of trace of products of matrices and that $\mathrm{Cum}\,(X_i, X_j) = \mathrm{Cov}\,(X_i, X_j)$. The number of indecomposable partitions can easily be proved to be $2^{l-1}(l-1)!$ with l being the number of quadratic forms. Using this fact in our problem, the lth joint cumulant of the quadratic forms takes the form

$$T^{dl/2}\mathrm{Cum}\,(\nabla_{i_1}\mathcal{L}_T^{(h)}(\theta_0), \ldots, \nabla_{i_l}\mathcal{L}_T^{(h)}(\theta_0))$$
$$= T^{-dl/2}\sum_{\nu}\sum_{j_k\in\sigma(i_1,\ldots,i_l)} \mathrm{tr}\prod_{k=1}^{l}\widetilde{U}_T^{(h)}((2\pi)^{-d}\nabla_{j_k}f_{\theta_0}^{-1})\Sigma(A),$$

where ν runs over all indecomposable partitions of the table

$$\begin{matrix} X_{i_1} & X_{j_1} \\ \vdots & \vdots \\ X_{i_l} & X_{j_l} \end{matrix}$$

and σ denotes in this case the set of permutations of a sequence. Consequently this is

$$= T^{-dl/2+d}\sum_{\nu}\sum_{j_k\in\sigma(i_1,\ldots,i_l)} \frac{1}{T^d}\,\mathrm{tr}\prod_{k=1}^{l}\widetilde{U}_T^{(h)}((2\pi)^{-d}\nabla_{j_k}f_{\theta_0}^{-1})\Sigma(A),$$

which is bounded given Lemma 2.2.2. Therefore the whole expression for the cumulant has order $\mathcal{O}(T^{-dl/2+d})$, *i.e.* the cumulants vanish for $l > 2$ when $T \to \infty$, which implies finally the Gaussianity in (iii).
From the Taylor expansion in (3.12) we have

$$\widehat{\theta}_T^{(h)} = \theta_0 - \Gamma_{h,T}^{-1}\nabla\mathcal{L}_T^{(h)}(\theta_0),$$
$$\mathrm{E}\,\widehat{\theta}_T^{(h)} = \theta_0 - \mathrm{E}\,\Gamma_{h,T}^{-1}\nabla\mathcal{L}_T^{(h)}(\theta_0),$$

where $\Gamma_{h,T}^{-1} := \{\nabla^2\mathcal{L}_T^{(h)}(\theta_T^{(*)})\}^{-1}$. Thus,

$$T^{d/2}(\widehat{\theta}_T^{(h)} - \mathrm{E}\,\widehat{\theta}_T^{(h)}) = -T^{d/2}(\Gamma_{h,T}^{-1}\nabla\mathcal{L}_T^{(h)}(\theta_0) - \mathrm{E}\,\Gamma_{h,T}^{-1}\nabla\mathcal{L}_T^{(h)}(\theta_0)). \tag{3.14}$$

Step (ii) from above implies that $\Gamma_{h,T}^{-1} \xrightarrow{p} \Gamma_h^{-1}$. Finally, the desired result follows applying Slutski's theorem and the Gaussianity proved above, in the right-hand side of (3.14), *i.e.*

$$T^{d/2}\Gamma_h^{-1}(\nabla\mathcal{L}_T^{(h)}(\theta_0) - \mathrm{E}\,\nabla\mathcal{L}_T^{(h)}(\theta_0)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \Gamma_h^{-1}V_h\Gamma_h^{-1}). \tag{3.15}$$

The Gaussianity of $\widetilde{\theta}_T$ follows by similar steps. The presence of Σ_θ^{-1} in the expression for $\nabla\mathcal{L}_T^{(ex)}(\theta)$ implies the use of Lemma 2.2.3 in order to calculate the cumulants. Further steps in the proof are similar to the one above. □

Remark 3.2.1. *Notice that the Gaussian distribution of* $\widehat{\theta}_T^{(h)} - E\widehat{\theta}_T^{(h)}$ *depends on the true parameter* θ_0. *In practice we may approximate this by plugging our estimator* $\widehat{\theta}_T^{(h)}$ *in* θ_0 *which, given the asymptotic vanishing bias, yields reasonable estimates of the Gaussian parameters.*

The crucial problem to obtain a Gaussian law centered in the true parameter vector θ_0 is the bias of $\widehat{\theta}_T^{(h)}$. As we have seen, its sources are the skew periodogram, edge effects and local stationarity. We were able to tackle the first two, yet the third source cannot be remove because of the very definition of a local stationary process. Nevertheless, in Dahlhaus and Sahm (2000) the authors propose to find a second order approximation of the transfer function, namely $A_{t,T}^0(\lambda) = A(u,\lambda) + \frac{1}{T}B(u,\lambda) + \mathcal{O}(T^{-2})$. This would imply a biased score function of order $\mathcal{O}(T^{-2})$ and therefore an asymptotic Gaussian law for the dimensions $1, 2$ and 3. Unfortunately, a second order approximation is rather cumbersome to find as well as to implement, which contradicts the spirit of our estimator.

4 Simulations and a Real Data Application

In this chapter we present the results of some simulations and an application of our method to the wheat-yield dataset of Mercer and Hall (1911). We divided the simulations in two steps: stationary and locally stationary. In the first case we simulated data using a backward substitution formula for the case of a 2D AR-process. The estimation was carried out using the ideas in Dahlhaus and Künsch (1987). In the second case the data was simulated using the infill asymptotic property and the estimation was carried out through our method. The results turned out to be quit good despite the small sample size in some cases. Concerning the real data application, we estimated a non-causal model using tapered and non-tapered estimators. The results in the non-tapered stationary case confirmed those of Whittle (1954), while the tapered estimator yielded results slightly different, as it was expected. Finally, we implemented the Whittle estimator for this model and fitted a model with affine functions as the corresponding parameters.

4.1 Simulations

4.1.1 Stationary Case

We consider a 2D causal AR(1)-process on the lattice $D_T = \{1, \dots, T\}^2$ given by

$$\begin{aligned} X_{s,t} &= \alpha X_{s-1,t} + \beta X_{s,t-1} + \varepsilon_{s,t}, \\ \varepsilon_{s,t} &\sim \mathcal{N}(0,1). \end{aligned} \tag{4.1}$$

In terms of backward operators, the model (4.1) can be written as

$$(1 - \alpha T_s^{-1} - \beta T_t^{-1}) X_{s,t} = \varepsilon_{s,t}.$$

Using (2.4) and (2.3) we obtain the spectral density of the process

$$f_\theta(\lambda) = \frac{1}{(2\pi)^2(1 + \alpha^2 + \beta^2 - 2\alpha\cos(\lambda_1) - 2\beta\cos(\lambda_2) + 2\alpha\beta\cos(\lambda_2 - \lambda_1))}, \tag{4.2}$$

where $\lambda = (\lambda_1, \lambda_2)$ and $\theta = (\alpha, \beta, 1)$. Performing n steps of backward substitution in model (4.1) we obtain

$$X_{s,t} = (\alpha T_s^{-1} + \beta T_t^{-1})^n X_{s,t} + \sum_{j=0}^{n-1} (\alpha T_s^{-1} + \beta T_t^{-1})^j \varepsilon_{s,t}.$$

If we assume $E|X_{s-n+k,t-k}| < M < \infty$, $\forall k, n$ with $0 \le k \le n$, a sufficient condition for stationarity to hold is $|\alpha| + |\beta| < 1$, and hence, the first term of the right hand side can be neglected for n big enough. Consequently, the process can be written in its causal form

$$X_{s,t} \approx \sum_{j=0}^{n-1} \sum_{k=0}^{j} \binom{j}{k} \alpha^k \beta^{j-k} \varepsilon_{s-k,t-j+k}.$$

Since every $X_{s,t}$ in the lattice is the result of a big enough linear combination of innovations, the last equation gives us a way to simulate the process (4.1) choosing the parameters α and β which satisfy the stationarity condition[1]. Regarding how big n must be (cut criterion), we choose the necessary n such that terms below $10^{-6}E|X_{s,t}|$ are negligible, *e.g.* $n = -6/\log_{10}(|\alpha| + |\beta|)$.

The Table 4.1 shows the estimation results for the model (4.1) varying T and considering tapered and non-tapered data. The taper used is Tukey-Hanning defined in (2.34). Notice that this taper function fulfils assumption (A6), therefore our theoretical results are fulfilled if we use this taper.

Table 4.1: Parameter estimation for model (4.1) under stationarity assumption using simulated data

	Parameters		
Method	$\hat{\alpha}$	$\hat{\beta}$	T
Simulated	-0.100	-0.200	-
Non-tapered	-0.097	-0.174	10
Tapered ($\rho = 0.0681$)	-0.097	-0.174	10
Non-tapered	-0.106	-0.196	50
Tapered ($\rho = 0.0521$)	-0.108	-0.197	50
Non-tapered	-0.102	-0.207	100
Tapered ($\rho = 0.0464$)	-0.105	-0.209	100
Non-tapered	-0.106	-0.197	500
Tapered ($\rho = 0.0355$)	-0.106	-0.194	500
Non-tapered	-0.108	-0.197	1000
Tapered ($\rho = 0.0316$)	-0.110	-0.197	1000

In this simulation, it has been considered parameters α and β away from the unit root. Notice that even with a small sample size ($T^2 = 100$) the estimation is quite close to the true value. The taper-estimator performance turns out to be similar to non-tapered case. The parameter ρ has been chosen, for both stationary and locally stationary cases, as $\rho = 0.1 \cdot T^{-1/6}$.

[1] The simulation code for the stationary random field case was written in C++, while in the locally stationary case and all the estimation procedures, the language used was R (R Core Team (2016)). We used the Newton-type method of minimization “nlminb” with box constraints from the package **optimx** in R. Graphs were made in Matlab and R.

Table 4.2: Parameter estimation for model (4.1) with $|\alpha| + |\beta| \approx 1$ using simulated data

	Parameters		
Method	$\hat{\alpha}$	$\hat{\beta}$	T
Simulated	-0.480	0.500	-
Non-tapered	-0.467	0.511	100
Tapered ($\rho = 0.0464$)	-0.476	0.501	100
Non-tapered	-0.473	0.504	500
Tapered ($\rho = 0.0355$)	-0.476	0.501	500
Simulated	-0.490	0.500	-
Non-tapered	-0.471	0.500	50
Tapered ($\rho = 0.0521$)	-0.473	0.498	50

In the second experiment, we have simulated AR processes close to the unit root ($|\alpha| + |\beta| \approx 1$). In this case the necessary n to fulfil our cut criterion becomes very large (roughly $n > 1000$). Because of computational restrictions, we just simulated a few examples. As we can observe in Table 4.2, the estimation performed equally well despite the long memory of the data.

We also carried out a number of possible combinations of parameters α and β yielding all similar good results as those presented above.

4.1.2 Locally Stationary Case

We move now to the locally stationary case. Consider a 2D AR(1)-process on the lattice $D_T = \{1, \dots, T\}^2$ given by

$$\begin{aligned} X_{s,t} &= \alpha\Big(\tfrac{s-1/2}{T}, \tfrac{t-1/2}{T}\Big)X_{s-1,t} + \beta\Big(\tfrac{s-1/2}{T}, \tfrac{t-1/2}{T}\Big)X_{s,t-1} + \varepsilon_{s,t}, \\ \varepsilon_{s,t} &\sim \mathcal{N}(0, \sigma^2(\tfrac{s-1/2}{T}, \tfrac{t-1/2}{T})). \end{aligned} \tag{4.3}$$

Since the parameters are supposed to change spatially, the method to simulate random fields considered above turns out to be extremely burdensome and not computationally feasible in many interesting cases. To tackle this problem, we propose a heuristic to generate locally stationary random fields with the desired structure using strongly the infill asymptotic fact, namely, more and more data are obtained with similar local probabilistic structure.

We detail this idea as follows: under model (4.3) over a $T \times T$ square, the representativeness of a point $X_{s,t},\ s, t \neq 1$ will depend entirely on the edges due to the 2D causality of the model. Thus, the key point here consists in generating good representatives of the process on the two boundaries. In order to achieve this, consider two zero-vectors of length $N = 10^5$ to form the left and bottom margins of a square. Next, we simulate the process (4.3) to fill a L-shaped space with thickness equal to 2000 observations both, from the left and from the bottom. The N and thickness considered, allow the parameters to change smoothly and get rid of the influence of the zero-boundary conditions as well. Considering different intervals between observations, say 50, 100, 250 or any other small enough divisor of N we obtain representative data for the left and bottom boundaries of a squared random field (of size $T \times T$ equal to 2000×2000, 1000×1000, 400×400, respectively) generated trough model (4.3). In Algorithm 1 we present a pseudo code used to simulate model (4.3).

Algorithm 1 Locally stationary random field (LS-RF) generator

procedure GENERATOR($X_{s,t}$) ▷ LS-RF generator in $D_T (:= \{1, \dots, T\}^2)$
Require: T, r, N ▷ RF size, thickness and "burning" length, resp.
Require: $\alpha(\cdot,\cdot)$, $\beta(\cdot,\cdot)$ and $\sigma(\cdot,\cdot)$ ▷ define parameter functions
Require: $\varepsilon_{s,t}$ ▷ standard gaussian generator
Ensure: $T|N$ and $N/T \leq r$
  $Y_{0,t} \leftarrow 0$ and $Y_{s,0} \leftarrow 0$ ▷ $\forall\ s,t \in \{1,\dots,r\}$
  for $(s,t) \in \{1,\dots,r\}^2$ **do**
    $Y_{s,t} \leftarrow \alpha\Big(\frac{s-1/2}{T}, \frac{t-1/2}{T}\Big) Y_{s-1,t} + \beta\Big(\frac{s-1/2}{T}, \frac{t-1/2}{T}\Big) Y_{s,t-1} + \sigma\Big(\frac{s-1/2}{T}, \frac{t-1/2}{T}\Big) \varepsilon_{s,t}$
  end for
  $Y_{0,t} \leftarrow 0$ ▷ $\forall t \in \{r+1,\dots,N\}$
  for $(s,t) \in \{1,\dots,r\} \times \{r+1,\dots,N\}$ **do**
    $Y_{s,t} \leftarrow \alpha\Big(\frac{s-1/2}{T}, \frac{t-1/2}{T}\Big) Y_{s-1,t} + \beta\Big(\frac{s-1/2}{T}, \frac{t-1/2}{T}\Big) Y_{s,t-1} + \sigma\Big(\frac{s-1/2}{T}, \frac{t-1/2}{T}\Big) \varepsilon_{s,t}$
  end for
  $Y_{s,0} \leftarrow 0$ ▷ $\forall s \in \{r+1,\dots,N\}$
  for $(s,t) \in \{r+1,\dots,N\} \times \{1,\dots,r\}$ **do**
    $Y_{s,t} \leftarrow \alpha\Big(\frac{s-1/2}{T}, \frac{t-1/2}{T}\Big) Y_{s-1,t} + \beta\Big(\frac{s-1/2}{T}, \frac{t-1/2}{T}\Big) Y_{s,t-1} + \sigma\Big(\frac{s-1/2}{T}, \frac{t-1/2}{T}\Big) \varepsilon_{s,t}$
  end for
  $k \leftarrow N/T$
  $X_{i,0} \leftarrow Y_{i\cdot k,k}$ ▷ $i = 1,\dots,T$
  $X_{0,j} \leftarrow Y_{k,j\cdot k}$ ▷ $j = 1,\dots,T$
  for $(s,t) \in D_T$ **do**
    $X_{s,t} \leftarrow \alpha\Big(\frac{s-1/2}{T}, \frac{t-1/2}{T}\Big) X_{s-1,t} + \beta\Big(\frac{s-1/2}{T}, \frac{t-1/2}{T}\Big) X_{s,t-1} + \sigma\Big(\frac{s-1/2}{T}, \frac{t-1/2}{T}\Big) \varepsilon_{s,t}$
  end for
  return $X_{s,t}$ ▷ The LS Random Field
end procedure

This algorithm is computationally easy but requires causality. In case of noncausal processes, we would have to modify the algorithm or rely on standard Markov chain Monte Carlo methods. In our simulations we have imposed $\sigma^2(u) = 1,\ \forall u \in [0,1]^2$ since in that case $\int_{\Pi^2} \log f_\theta(u,\lambda)\, \mathrm{d}\lambda = 0$ for each u (for this case it is enough to notice that $a_{0,\dots,0} = 0$ in the proof of Theorem (2.1.1)) and thus, we just need to minimize (2.26), making estimation computationally faster. Additionally we can write this expression in the following form

$$\begin{aligned}
\mathcal{L}_T(\theta) &= \frac{1}{(2\pi T)^2} \sum_{t\in D_T} \int_{\Pi^2} \frac{J_T\Big(\frac{t-1/2}{T}, \lambda\Big)}{f_\theta\Big(\frac{t-1/2}{T}, \lambda\Big)} \,\mathrm{d}\lambda \\
&= \frac{1}{(2\pi T)^2} \sum_{t\in D_T} \frac{1}{(2\pi)^2} \sum_{\substack{k\in D_T - D_T:\\ [t\pm k/2]\in D_T\\ [t\pm k/2]^*\in D_T}} \Big(\tfrac{X_{[t+k/2],T}X_{[t-k/2],T}+X_{[t+k/2]^*,T}X_{[t-k/2]^*,T}}{2}\Big) \\
&\quad \times \int_{\Pi^2} \exp(i\langle \lambda, k\rangle) f_\theta^{-1}\Big(\frac{t-1/2}{T}, \lambda\Big)\,\mathrm{d}\lambda.
\end{aligned}$$

The assumption of local stationarity (see Definition (1.3)) implies that there exist a transfer function with distance at most $\mathcal{O}(T^{-1})$ respect to the true one $A^0_{t,T}$. Thus, in our case, the first order time-varying spectral density approximation $f_\theta(u,\lambda)$ for the model (4.1) turns out to be the function (4.2) by considering the parameters as functions of location (see equation (1.7) in Chapter 1). Using the orthogonality of $\exp(i\langle\lambda,k\rangle)$ in Π^2 it is easy to show that

$$\begin{aligned}
\mathcal{L}_T(\theta) = \frac{1}{(2\pi T)^2} \sum_{t\in D_T} \Big\{ &[1+\alpha^2\Big(\tfrac{t-1/2}{T}\Big)+\beta^2\Big(\tfrac{t-1/2}{T}\Big)]X^2_{t,T} \\
&-\alpha\Big(\tfrac{t-1/2}{T}\Big)\Big(X_{[t+\binom{1/2}{0}],T}X_{[t-\binom{1/2}{0}],T}+X_{[t+\binom{1/2}{0}]^*,T}X_{[t-\binom{1/2}{0}]^*,T}\Big) \\
&-\beta\Big(\tfrac{t-1/2}{T}\Big)\Big(X_{[t+\binom{0}{1/2}],T}X_{[t-\binom{0}{1/2}],T}+X_{[t+\binom{0}{1/2}]^*,T}X_{[t-\binom{0}{1/2}]^*,T}\Big) \\
&+\alpha\Big(\tfrac{t-1/2}{T}\Big)\beta\Big(\tfrac{t-1/2}{T}\Big)\Big(X_{[t+\binom{-1/2}{1/2}],T}X_{[t-\binom{-1/2}{1/2}],T} \\
&+X_{[t+\binom{-1/2}{1/2}]^*,T}X_{[t-\binom{-1/2}{1/2}]^*,T}\Big)\Big\},
\end{aligned} \tag{4.4}$$

which reduce substantially the number of calculations due to the elimination of redundant integrals. Strictly speaking, the terms with ceiling functions $[\cdot]^*$ are not present in the original derivations, but we incorporate them to avoid possible biases from the use of floor functions. This expression allows to obtain explicitly the equations of estimation. We denote

$$\begin{aligned}
V_t &:= X_{[t+\binom{1/2}{0}],T}X_{[t-\binom{1/2}{0}],T}+X_{[t+\binom{1/2}{0}]^*,T}X_{[t-\binom{1/2}{0}]^*,T} \\
W_t &:= X_{[t+\binom{0}{1/2}],T}X_{[t-\binom{0}{1/2}],T}+X_{[t+\binom{0}{1/2}]^*,T}X_{[t-\binom{0}{1/2}]^*,T} \\
Z_t &:= X_{[t+\binom{-1/2}{1/2}],T}X_{[t-\binom{-1/2}{1/2}],T}+X_{[t+\binom{-1/2}{1/2}]^*,T}X_{[t-\binom{-1/2}{1/2}]^*,T}
\end{aligned}$$

thus we can replace α and β can be replaced with parametric functions (a possible case might be equations (4.5) and (4.6) below). We will present the simplest case when α and β are assumed constant. Higher order polynomials involve more equations (in fact equations (4.5) and (4.6) would imply a 6×6 system) and/or the resolution of non-linear systems.

Differentiating with respect to α and β we obtain

$$\begin{aligned}\frac{\partial \mathcal{L}_T(\theta)}{\partial \alpha} &= \frac{1}{T^2}\sum_{t\in D_t}\{2\alpha X_{t,T}^2 - V_t + \beta Z_t\},\\ \frac{\partial \mathcal{L}_T(\theta)}{\partial \beta} &= \frac{1}{T^2}\sum_{t\in D_t}\{2\beta X_{t,T}^2 - W_t + \alpha Z_t\}.\end{aligned}$$

The Cauchy-Schwarz inequality implies that the determinant of this system is always positive, which implies unique solution. The positive determinant together with the fact that $\frac{\partial^2 \mathcal{L}_T(\theta)}{\partial \alpha^2} = \frac{\partial^2 \mathcal{L}_T(\theta)}{\partial \beta^2} > 0$, imply that Hessian matrix is positive definite and hence there exist a unique minimum, namely

$$\begin{aligned}\hat{\alpha} &= \frac{(\sum_{t\in D_T^*} Z_t)(\sum_{t\in D_t^*} W_t) - (\sum_{t\in D_T} X_{t,T}^2)(\sum_{t\in D_t^*} V_t)}{(\sum_{t\in D_T^*} Z_t)^2 - (\sum_{t\in D_T} X_{t,T}^2)^2},\\ \hat{\beta} &= \frac{(\sum_{t\in D_T^*} Z_t)(\sum_{t\in D_t^*} V_t) - (\sum_{t\in D_T} X_{t,T}^2)(\sum_{t\in D_t^*} W_t)}{(\sum_{t\in D_T^*} Z_t)^2 - (\sum_{t\in D_T} X_{t,T}^2)^2}.\end{aligned}$$

Here D_T^* represent a subset of D_T where the given sum makes sense. Since we assumed $\int_{\Pi^2} \log f_\theta(u,\lambda)\,\mathrm{d}\lambda = 0$, the use of data taper consist simply in passing pre-processed tapered data to the above expressions.

We present as follows the results of two simulations. The first one involve functions

$$\alpha(s,t) = \alpha_0 + \alpha_1 s + \alpha_2 t, \tag{4.5}$$
$$\beta(s,t) = \beta_0 + \beta_1 s + \beta_2 t. \tag{4.6}$$

The results are presented in Table 4.3.

In general, the estimations did not perform quite well for small sizes (50 and 100). This might be attributed to the inherent bias of the method. Following partly what done in the stationary setting, we tapered the data using $\rho = \mathcal{O}(T^{-1/6})$. We notice that the tapered results are asymptotically equivalent to the now tapered cases. The reason lies in our simulation algorithm, which has a big enough "burning" space, *i.e.* the influence of the edges dies out before it is considered in the final random field data set, whom we conduct the simulations with.

Table 4.3: Parameter estimation for model (4.1) with α and β given by (4.5) and (4.6), respectively, and $\varepsilon_{s,t} \sim \mathcal{N}(0,1)$.

Method	Parameters						T
	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	
Simulated	0.100	−0.200	0.300	0.400	0.100	−0.400	-
Non-tapered	0.118	−0.288	0.346	0.233	0.253	−0.242	50
Tapered ($\rho = 0.0521$)	0.142	−0.303	0.334	0.288	0.200	−0.273	50
Non-tapered	0.050	−0.160	0.341	0.380	0.096	−0.365	100
Tapered ($\rho = 0.0464$)	0.065	−0.167	0.321	0.402	0.087	−0.389	100
Non-tapered	0.105	−0.184	0.271	0.382	0.109	−0.375	250
Tapered ($\rho = 0.0398$)	0.112	−0.194	0.270	0.391	0.107	−0.390	250
Non-tapered	0.099	−0.199	0.302	0.396	0.102	−0.399	500
Tapered ($\rho = 0.0355$)	0.100	−0.202	0.304	0.400	0.101	−0.404	500
Non-tapered	0.095	−0.196	0.302	0.400	0.097	−0.398	1000
Tapered ($\rho = 0.0316$)	0.096	−0.199	0.304	0.403	0.095	−0.400	1000
Non-tapered	0.099	−0.197	0.299	0.400	0.099	−0.400	2500
Tapered ($\rho = 0.0271$)	0.099	−0.198	0.299	0.401	0.099	−0.401	2500

Table 4.4: Parameter estimation for model (4.1) with α and β given by (4.5) and (4.6), respectively, and $\varepsilon_{s,t} \sim Unif(-\sqrt{3}, \sqrt{3})$.

Method	Parameters						T
	$\hat{\alpha}_0$	$\hat{\alpha}_1$	$\hat{\alpha}_2$	$\hat{\beta}_0$	$\hat{\beta}_1$	$\hat{\beta}_2$	
Simulated	0.100	−0.200	0.300	0.400	0.100	−0.400	-
Non-tapered	0.099	−0.224	0.258	0.340	0.190	−0.420	50
Tapered ($\rho = 0.0521$)	0.117	−0.251	0.245	0.351	0.201	−0.442	50
Non-tapered	0.091	−0.192	0.308	0.347	0.161	−0.395	100
Tapered ($\rho = 0.0464$)	0.106	−0.208	0.301	0.363	0.148	−0.406	100
Non-tapered	0.094	−0.211	0.322	0.393	0.102	−0.398	250
Tapered ($\rho = 0.0398$)	0.094	−0.216	0.330	0.405	0.097	−0.411	250
Non-tapered	0.097	−0.195	0.296	0.404	0.099	−0.409	500
Tapered ($\rho = 0.0355$)	0.099	−0.198	0.296	0.409	0.098	−0.414	500
Non-tapered	0.099	−0.201	0.300	0.399	0.100	−0.399	1000
Tapered ($\rho = 0.0316$)	0.100	−0.202	0.300	0.402	0.099	−0.401	1000
Non-tapered	0.099	−0.200	0.302	0.400	0.099	−0.399	2500
Tapered ($\rho = 0.0271$)	0.099	−0.200	0.302	0.401	0.099	−0.401	2500

In order to evaluate the behaviour of our estimation procedure regarding departures from the Gaussianity assumption, we have run simulations of model (4.3) with functions $\alpha(s,t)$, $\beta(s,t)$ given by equations (4.5) and (4.6) respectively, same parameter values as the previous experiment, but with $\varepsilon_{s,t} \sim Unif(-\sqrt{3}, \sqrt{3})$. The choice of the limits of the distribution obeys to the fact that $\mathrm{E}\,\varepsilon_{s,t} = 0$, $\mathrm{Var}\,\varepsilon_{s,t} = 1$ which facilitates the comparison with the previous zero-one Gaussian experiment. Interestingly, the results are quite similar to the Gaussian case, even a few simulations outperformed it. The results are depicted in Table 4.4.

In the second case the parameter functions are given by

$$\begin{aligned}\alpha(s,t) &= \exp[-4\{(s-0.5)^2+(t-0.5)^2\}]-0.4 \quad s,t\in[0,1],\\ \beta(s,t) &= 0.3.\end{aligned}$$

In Figure 4.1 is depicted the function α.

For this particular experiment we have assumed we do not know the functional form of α and β, so we have fitted polynomials in two variables.

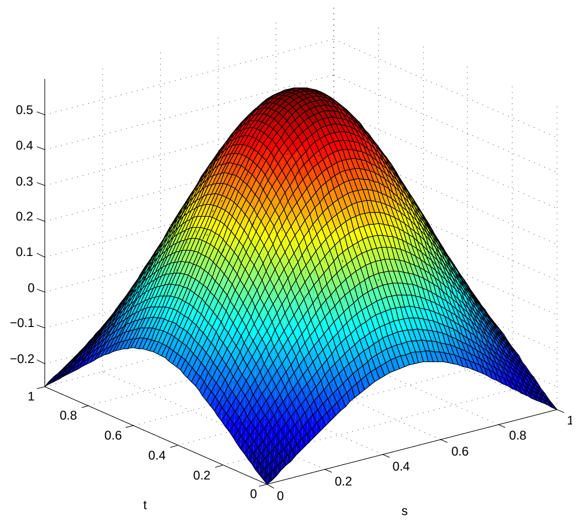
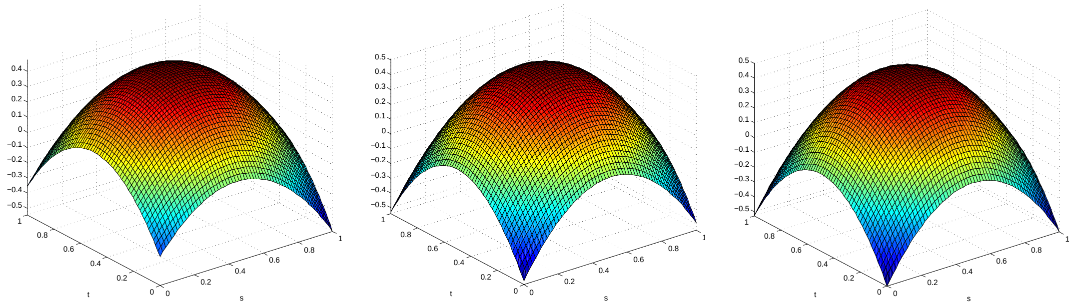


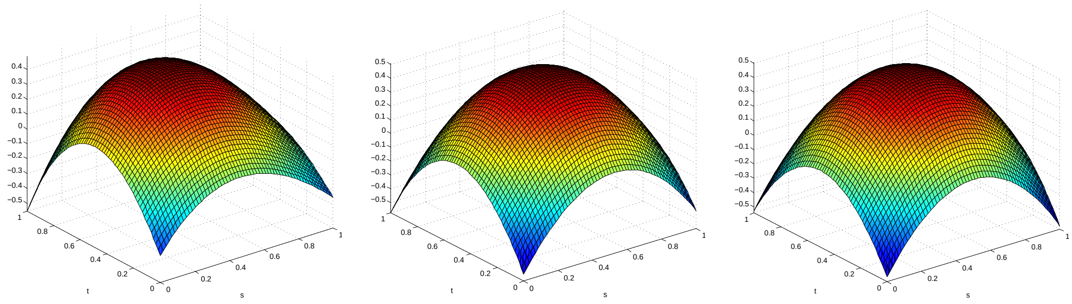
Figure 4.1: Function $\alpha(s,t)$

In Figures 4.2 - 4.5 we present results for two pairs of functions assumed for both α and β. The results are quite good for points away from the borders and for almost all sizes. As it was expected, the estimation of the function α tend to misbehave at the edges (especially in the corners), but this improves when increasing the polynomial order and/or the field size T (not shown). In order to build a model selection criteria more research has to be done.



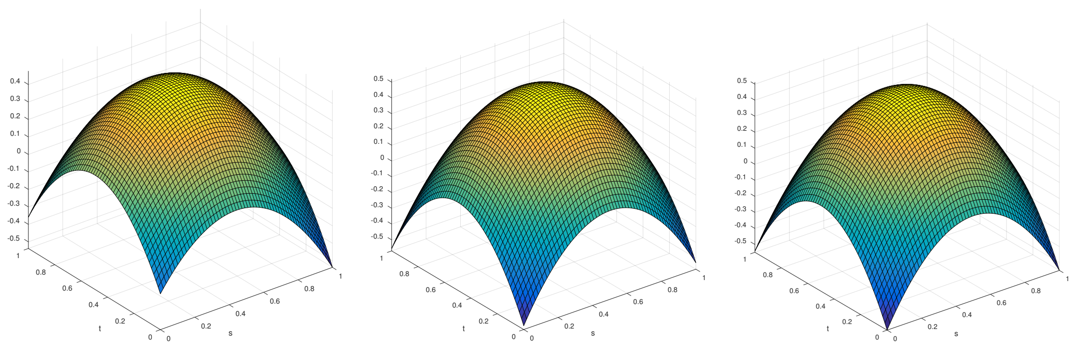
(a) $T = 50$ (b) $T = 250$ (c) $T = 1000$

Figure 4.2: Non-tapered data, $\hat{\alpha}$ quadratic, $\hat{\beta}$ constant



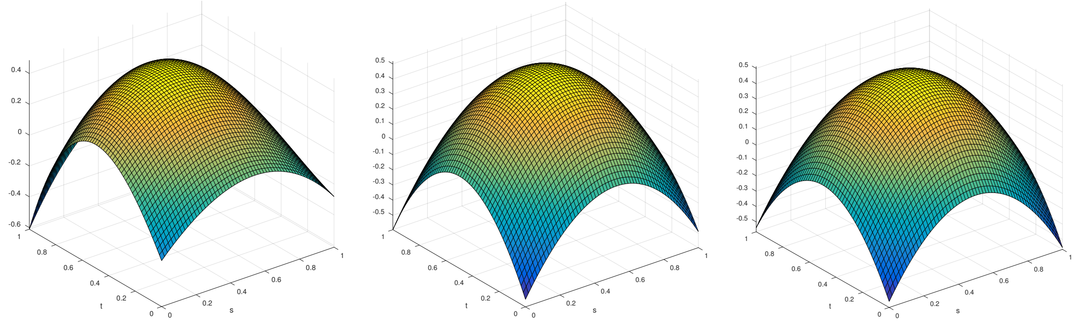
(a) $T = 50$ (b) $T = 250$ (c) $T = 1000$

Figure 4.3: Non-tapered data, $\hat{\alpha}$ cubic, $\hat{\beta}$ linear



(a) $T = 50$ (b) $T = 250$ (c) $T = 1000$

Figure 4.4: Tapered data, $\hat{\alpha}$ quadratic, $\hat{\beta}$ constant



(a) $T = 50$ (b) $T = 250$ (c) $T = 1000$

Figure 4.5: Tapered data, $\hat{\alpha}$ cubic, $\hat{\beta}$ linear

The estimation of β in the different cases performed also good. We present in Table 4.5 the results for such function.

Table 4.5: Parameter estimation for function $\hat{\beta}$

Data type	$T = 50$	$T = 250$	$T = 1000$
Non-tapered	0.299	0.303	0.302
Tapered	0.312	0.304	0.302
Non-tapered	$0.143 + 0.192s + 0.113t$	$0.285 + 0.022s + 0.013t$	$0.303 - 0.001s - 0.001t$
Tapered	$0.197 + 0.127s + 0.099t$	$0.292 + 0.017s + 0.006t$	$0.305 - 0.006s - 0.001t$

The values of ρ are 0.0521, 0.0398 and 0.0316 for $T = 50$, $T = 250$ and $T = 1000$ respectively. The estimators were carried out with different values of ρ as well (order $\mathcal{O}(T^{-1/3})$, not shown), being the results rather similar.

It is worth noticing that even if we fit a polynomial of higher order, *e.g.* a first order polynomial for the lower half of Table 4.5, as long as the size increases, the parameters for s and t become negligible. This confirms the theoretical results about Kullback-Leibler information divergence found in Chapter 2.

4.2 Real Data Application

To exemplify our method using a real data set, we have considered the wheat-yields of Mercer and Hall (1911). This consists of a uniform trial over a lattice of 20×25 (one acre area each one of them) where the production was measured in pounds. To have a sense of the data we display its level plot in Figure 4.6, where rows run west-east and columns north-south. Out of the spatial correlation graph depicted in Figure 4.7 can be observed a vertical correlation stronger than the horizontal one. We proposed to fit the following model:

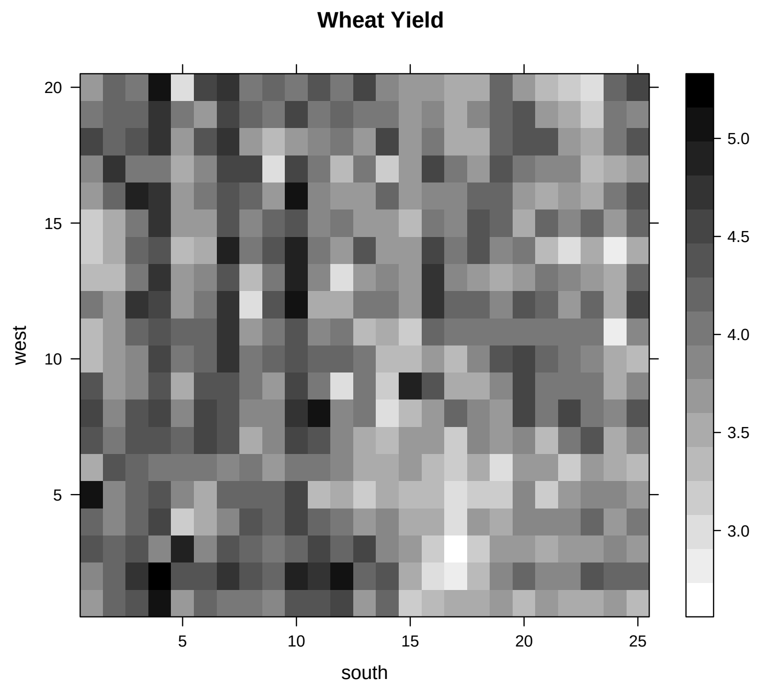


Figure 4.6: Mercer and Hall (1911) (own elaboration)

Let $X_{s,t}$ be a 2-dimensional stationary SAR process on the lattice $[1, T_1] \times [1, T_2]$, where $T_1 = 25$ and $T_2 = 20$

$$\begin{aligned} X_{s,t} &= \alpha(X_{s-1,t} + X_{s+1,t}) + \beta(X_{s,t-1} + X_{s,t+1}) + \varepsilon_{s,t}, \\ \varepsilon_{s,t} &\sim \mathcal{N}(0, \sigma^2). \end{aligned} \tag{4.7}$$

where its spectral density function is

$$f_\theta(\lambda) = \frac{\sigma^2}{(2\pi)^2(1 - 2\alpha\cos(\lambda_1) - 2\beta\cos(\lambda_2))^2}. \tag{4.8}$$

Beside the fact that this model seems to be a realistic description of what we see in Figure 4.7, it also allows us to compare with the results of Whittle (1954), who performs the same fit, but using a different estimation method.

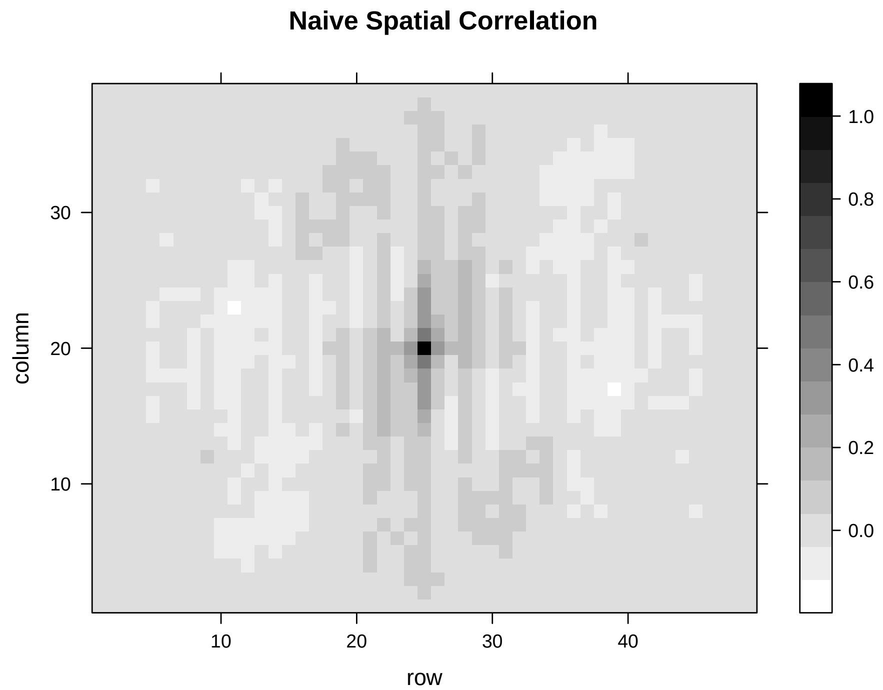


Figure 4.7: Spatial Correlation Wheat-Yield data (own elaboration). Regarding what the "naive" term means, see Dahlhaus and Künsch (1987) and the distinctions made therein.

For the sake of completeness, we present the estimation of parameters first assuming stationarity with non-tapered data (Whittle's original method) and with tapered-data (using Tukey-Hanning taper function (2.34)). In a second step we present the estimation in the locally stationary case with both, non-tapered and tapered data.

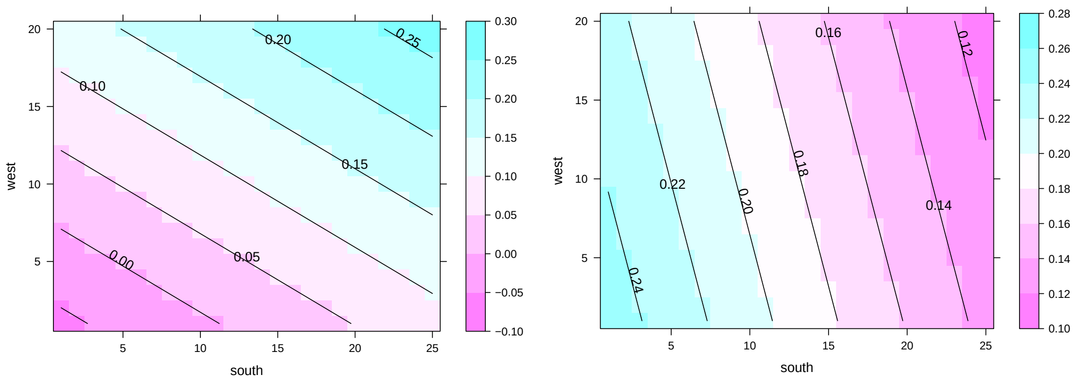
Table 4.6: Parameter estimation of model (4.7) under stationarity assumption using wheat-yield data set

	Parameters		
Method	$\hat{\alpha}$	$\hat{\beta}$	$\hat{\sigma}^2$
Whittle (1954)	0.102	0.213	-
Non-tapered	0.097	0.211	0.136
Tapered ($\rho = 0.0596$)	0.098	0.217	0.132

Whittle (1954) did not give the value of the variance of the innovations, nevertheless Casáis (2006) performed similar estimations using this model and obtained $\hat{\sigma}^2 = 0.125$, which is close to our results. In order to build the estimator, Whittle (1954) neglects edge-effects. On the other hand, Dahlhaus and Künsch (1987) proved that using data tapers leads to small-bias estimators with smaller variance than the naive estimators (among other properties). Nevertheless, we are not in a position to favour one of these two values (or even more considering different values of ρ), since the sample size turns out to be very small.

Following Whittle (1954) we have assumed the data has constant and known mean, which is subtracted at the beginning of the estimation procedure. However, there exist some evidence in favour of a space-varying mean (see Dreesman and Tutz (2001) for a discussion). In our research we have devoted ourselves to the local stationarity in the covariance structure of discrete spatial models. We let this area for discussion elsewhere.

Finally we present results from the fitting of a locally stationary process for the wheat yield data. The model assumed has the form (4.7) with varying parameters in α and β. Since the sample size is considerably small we let aside the possibility of varying variance and estimate just the constant value. The functional forms assumed are (4.5) and (4.6). In Figures 4.8a and 4.8b are depicted the level plots of the estimations.



(a) $\hat{\alpha}(s,t)$ (b) $\hat{\beta}(s,t)$

Figure 4.8: Level plots of estimated functions α and β using the non-tapered wheat yield data

The estimated functions are $\hat{\alpha}(s,t) = -0.076 + 0.147s + 0.197t$ and $\hat{\beta}(s,t) = 0.256 - 0.121s - 0.025t$, while the variance was estimated as $\hat{\sigma}^2 = 0.138$. Using Theorem 3.2.2 we have estimated confidence intervals for each parameter. We estimate $f(u,\lambda)$ using the preperiodogram and plug the estimated values for α, β and σ in the corresponding derivative functions. With a level of 95% only $\widehat{\beta}_0$ and $\widehat{\sigma}$ turned out to be significative (the intervals are $[0.032, 0.480]$ and $[0.339, 0.405]$ respectively). Considering the simulations presented above, we have to be carefull with the current results due to the rather small sample size. We conclude that in the case of horizontal association, measured by parameter function α, it becomes stronger as we move in north-east direction over the field. On the other hand, the vertical association, measured by parameter function β, it becomes stronger as we move in direction east-west being the changes along the level curves in vertical direction almost negligible. We also conducted estimations with tapered data obtaining essentially the same results for different values of ρ (not shown).

4.3 Discussion

The stationary simulations showed to perform better as the sample size T increases. Though not all depicted, we ran additional simulation whose posterior estimation yielded better results for decreasing ρ. According to our results, the chosen parameter values for each model played no role regarding estimation (see Table 4.2), but they actually did in terms of computer data generation; those close to the unit root turned out to be quit burdensome computationally. In the locally stationary case we faced the problem of generating data with the desired properties. To tackle this problem, we proposed a fast and simple algorithm which make use of the infill asymptotic assumption. The estimation results confirmed the consistency of the data generating algorithm. The above two observations for the stationary case regarding T and ρ hold also for the local stationary case. Nevertheless it is clear that reasonable results start for bigger values of T (say between 100 and 250), considering that in this case the number of parameters increase as well. The use of a different distribution (in our case uniform) did not represent any major change in the estimated parameter in comparison with the Gaussian case. The estimations for the parametric non linear function $\hat{\alpha}(s,t)$ turned out to be also reasonable just by simple inspection, we do not have the necessary tools to compare functions with different polynomial order. We let a model selection criteria as a matter of future research.

The results obtained from the stationary estimation for the non tapered wheat-yield data, turned out to be very similar with respect to those given in Whittle (1954), thus our method is consistent. The model assumed was a non causal SAR (eq. 4.7). In the locally stationary case, we just estimated affine functions for parameters α and β and constant σ^2. Because of the small sample size data, we treat the results carefully. With a Gaussian law might be obtained an estimation of the error, and hence, weighing up rightly conclusions from the results.

Conclusion

In this work we have developed an estimator for centered Gaussian locally stationary random fields processes. We have applied, and in some cases extended the ideas of Prof. Dr. Rainer Dahlhaus (Dahlhaus (1996), Dahlhaus (2000)) who proposed a Whittle-type approximation of the Gaussian likelihood for locally stationary time series. This approach has motivated our research of the random field case taking as starting point what has been presented in Dahlhaus and Sahm (2000). The Whittle likelihood was obtained thanks to an extension of the Szegö's theorem and an approximation of the matrix $\Sigma_T^{-1}(A_\theta)$. The score function of the Whittle likelihood proved to be biased, which implies a bias of same order for the Whittle estimator $\widehat{\theta}_T$. The bias sources were identified (skewed preperiodogram, local stationarity and edge-effects), studied and some solutions were proposed. Yet, the bias coming from the local stationarity remains. In order to get a Gaussian law we centered the estimator in $\mathrm{E}\,\widehat{\theta}_T^{(h)}$, obtaining a useful law in any dimension. Consistency of the estimator and uniform convergence of likelihoods were proved as well.

Finally, we conducted some simulations under different sets of parameters, different field sizes, with tapered and non tapered data, considering linear and non linear parameter functions and different distributions of innovations. In all cases the results performed reasonably well, excepting for those with small amount of data. We fitted stationary and locally stationary SAR models to the well known wheat-yield data set of Mercer and Hall (1911). Our estimation of the stationary case (following Dahlhaus and Künsch (1987)) turned out to be quite similar to what is found in the literature (see Whittle (1954)). From the results of the non stationary case we learned that as we move in the south west - north east and east - west directions the horizontal and vertical association respectively get stronger.

Possible directions of future research are:

(a) Determine a model selection criteria.

(b) Incorporate a local stationary mean.

(c) Generalize this findings to higher dimension random fields ($\delta > 1$, see Definition 1.1.1).

(d) Build a test of non stationarity for random fields using this approach (see Paparoditis et al. (2009) for an analogous in the time series context).

References

Brillinger, D. (1981). Time Series: Data Analysis and Theory. Holden-Day, Inc.

Casáis, R. M. C. (2006). Contributions to spectral spatial statistics. Univ Santiago de Compostela.

Dahlhaus, R. (1996). On the Kullback-Leibler information divergence of locally stationary processes. Stochastic Processes and their Applications *62*, 139–168.

Dahlhaus, R. (1997). Fitting Time Series models to Nonstationary processes. The Annals of Statistics *25*, 1–37.

Dahlhaus, R. (2000). A Likelihood Approximation for Locally Stationary Processes. The Annals of Statistics *28*, 1762–1794.

Dahlhaus, R. and Künsch, H. (1987). Edge effects and efficient parameter estimation for stationary random fields. Biometrika *74*, 877–882.

Dahlhaus, R. and Sahm, M. (2000). Likelihood Methods for Nonstationary Time Series and Random Fields. Resenhas do Instituto de Matemática e Estatística da Universidade de Sao Paulo *4*, 457–477.

Dreesman, J. M. and Tutz, G. (2001). Non-Stationary Conditional Models for Spatial Data Based on Varying Coefficients. Journal of the Royal Statistical Society: Series D (The Statistician) *50*, 1–15.

Dzhaparidze, K. (1986). Parameter Estimation and Hypothesis Testing in Spectral Analysis of Stationary Time Series. Springer Series in Statistics, Springer New York.

Guyon, X. (1995). Random fields on a network: modeling, statistics, and applications. Springer Science & Business Media.

Jammalamadaka, S. R., Rao, T. S. and Terdik, G. (2006). Higher Order Cumulants of Random Vectors and Applications to Statistical Inference and Time Series. Sankhya: The Indian Journal of Statistics (2003-2007) *68*, 326–356.

Martin, W. and Flandrin, P. (1985). Wigner-Ville spectral analysis of nonstationary processes. IEEE Transactions on Acoustics, Speech, and Signal Processing *33*, 1461–1470.

Mathai, A. M. and Provost, S. B. (1992). Quadratic forms in random variables: theory and applications. Dekker.

Mercer, W. and Hall, A. (1911). The experimental error of field trials. The Journal of Agricultural Science *4*, 107–132.

Paparoditis, E. et al. (2009). Testing temporal constancy of the spectral structure of a time series. Bernoulli *15*, 1190–1221.

R Core Team (2016). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna, Austria.

Whittle, P. (1954). On Stationary Processes in the Plane. Biometrika *41*, 434–449.

Lebenslauf

Persönliche Daten

Name:	Danilo Alberto Pezo Villar

Ausbildung

2001–2007,2011	Mathematikingenieur, mit Hauptfach Angewandte Statistik, Fachbereich Mathematik, Universidad Técnica Federico Santa María, Chile.
2008–2010,2013	M.Sc. Mathematik, Fachbereich Mathematik, Universidad Técnica Federico Santa María, Chile.
2010–2012	B.A. Pädagogik, Universidad de Playa Ancha, Chile.
2014–2017	Doktorand, Fachbereich Mathematik, Technische Universität Kaiserslautern, Deutschland.

Berufserfahrungen

2004–2007	Aushilfkraft in verschidenen Fächern, Fachbereich Mathematik, Universidad Técnica Federico Santa María, Chile.
2008–2013	Dozent, Fachbereich Mathematik, Universidad Técnica Federico Santa María, Chile.
2011–2014	Wissentschaftliche Mitarbeiter, Centro Interdisciplinario de Neurociencia de Valaparaíso, Chile.
2015–2017	Wissentschaftliche Hilfkraft, Fraunhofer-Institut für Techno- und Wirtschaftsmathematik ITWM, Deutschland.

Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig verfasst und keine anderen als die angegebenen Hilfsmittel benutzt habe. Die Stellen, die anderen Werken dem Wortlaut oder dem Sinn nach entnommen wurden, habe ich durch die Angabe der Quelle, auch der benutzten Sekundärliteratur, als Entlehnung kenntlich gemacht.

Kaiserslautern, den 09. Oktober 2017

Danilo Pezo Villar