Asymptotic analysis of ML-covariance parameter estimators based on covariance approximations

Reinhard Furrer and Michael Hediger

Department of Mathematics, University of Zurich
e-mail: reinhard.furrer@math.uzh.ch*;* michael.hediger@math.uzh.ch

Abstract: Given a zero-mean Gaussian random field with a covariance function that belongs to a parametric family of covariance functions, we introduce a new notion of likelihood approximations, termed truncated-likelihood functions. Truncated-likelihood functions are based on direct functional approximations of the presumed family of covariance functions. For compactly supported covariance functions, within an increasing-domain asymptotic framework, we provide sufficient conditions under which consistency and asymptotic normality of estimators based on truncated-likelihood functions are preserved. We apply our result to the family of generalized Wendland covariance functions and discuss several examples of Wendland approximations. For families of covariance functions that are not compactly supported, we combine our results with the covariance tapering approach and show that ML estimators, based on truncated-tapered likelihood functions, asymptotically minimize the Kullback-Leibler divergence, when the taper range is fixed.

MSC2020 subject classifications: Primary 60G15, 62F12; secondary 41A99.
Keywords and phrases: Gaussian random fields, compactly supported covariance functions, likelihood approximations, consistency, asymptotic normality, covariance tapering.

1. Introduction

1.1. On infill- and increasing-domain asymptotics

Maximum likelihood (ML) estimators for covariance parameters are highly popular in inference for random fields. Aiming towards asymptotic properties of such estimators, one needs to specify how the observation points and the associated sampling domain behave as the number of observation points increases. Two well-studied asymptotic frameworks are referred to as infill-domain asymptotics (also termed fixed-domain asymptotics) and increasing-domain asymptotics (see [13, p. 100], for an introduction of terms). In infill-domain asymptotics, observation points are sampled within a bounded sampling domain, whereas in increasing-domain asymptotics, the sampling domain grows as the number of observation points increases. When referring to infill- and increasing-domain

asymptotics, one often places additional assumptions on the minimum distance between any two distinct observation points. In increasing-domain asymptotics, the latter distance is often assumed to be bounded away from zero, while in infill-domain asymptotics, one frequently assumes that distinct observation points can be sampled arbitrarily close to each other (see for example [37]). There is a fair amount of literature which demonstrates that asymptotic properties of ML estimators for covariance parameters can be quite different under the two mentioned asymptotic frameworks (see [37] or more lately [6]). For example, it is known that some covariance parameters can not be estimated consistently under an infill-domain asymptotic framework ([34, 36]), whereas they can be estimated consistently, under given regularity conditions, within an increasing-domain asymptotic framework ([25, 4]). It is worth noting that in infill-domain asymptotics, these results can depend on the dimension d of the Euclidean space $\mathbb{R}^d$, where the random field is assumed to be observed. For example, when the true covariance function belongs to the Matérn family ([26]), and smoothness parameters are given, it is shown in [36], that for $d = 1, 2, 3$, the scale and variance parameters can not be estimated consistently via an ML approach in an infill-domain asymptotic framework. The case where $d = 4$ is still open, but for $d \geq 5$, it is shown in [2] that under infill-domain asymptotics, all covariance parameters of the Matérn family can be estimated consistently using an ML approach.

1.2. Compactly supported covariance functions

In recent years, the dataset sizes have steadily increased such that statistical analyses on random fields can become quite expensive in terms of computational resources (see for example [15] for a recent discussion). One prominent issue with large datasets is the large size of covariance matrices, constructed upon applying an underlying covariance function to given data. However, in certain fields of application, observed correlations are assumed to vanish beyond a certain cut-off distance (see [18, pp. 750–751], and references therein, for an example in meteorology or also [10] and [19]). On the other hand, in the context of real valued random fields, it is common practice to multiply a presumed covariance function with a known positive-definite and compactly supported covariance function, called the covariance taper. The resulting compactly supported covariance function is referred to as the tapered covariance function. For an introduction to covariance tapering we refer to [17]. The use of compactly supported covariance functions can thus be of great importance for some fields of application. Not only do they potentially reflect the nature of the underlying covariance structure, but also, their application can lead to sparse covariance matrices. The latter are helpful in terms of the high computational costs in the context of large datasets. An excellent introduction to the construction of compactly supported covariance functions, associated to stationary and isotropic Gaussian random fields, is given in [21]. Additional results are available in [35], [28] and [11].

1.3. Motivation

The parametric family of generalized Wendland covariance functions represents one example of a family of compactly supported covariance functions which allows, similar to the Matérn family, for a continuous parametrization of smoothness (in the mean square sense) of the underlying random field. Its origin is due to Wendland ([32]) and an early adaptation for statistical applications was given by Gneiting ([20]). In its general form (see [21] and [28] for special cases) the generalized Wendland covariance function with smoothness parameters ν and κ, variance parameter σ^2 and range parameter β is given by

$$\phi(t) := \frac{\sigma^2}{B(2\kappa, \nu+1)\beta^{2\kappa+\nu}} \int_t^\beta w(w^2 - t^2)^{\kappa-1}(\beta - w)^\nu dw, \tag{1}$$

if $t \in [0, \beta)$ and is zero otherwise. In the above display, B is the beta function. For technical details about valid parameter values, we refer to [9] or Section 6 of the present article. Clearly, in comparison with closed-form covariance functions, computing (1) is cumbersome, as it involves numerical integration. Depending on the support β and a set of locations $s_1, \ldots, s_n \in \mathbb{R}^d$, the $n \times n$ covariance matrix $\Sigma_{i,j} = \phi(\|s_i - s_j\|)$ requires at most $n(n-1)/2$ calculations of (1). One strategy, which facilitates computing Σ, is to reduce the number of times (1) must be calculated. As an illustration, we give three examples which involve approximations $\tilde{\phi}_i$, $i = 1, 2, 3$, of ϕ (respectively approximations $\widetilde{\Sigma}_i$ of Σ):

- ($\tilde{\phi}_1$) Truncation of the support
- ($\tilde{\phi}_2$) linear interpolation
- ($\tilde{\phi}_3$) addition of a nugget effect

For $\tilde{\phi}_1$, we truncate ϕ to obtain $\tilde{\phi}_1$ which has a smaller support compared to ϕ. This becomes especially interesting, when the original function ϕ tails off slowly (high degree of differentiability at the origin). As a result, $\widetilde{\Sigma}_1$ will be more sparse compared to Σ. Example $\tilde{\phi}_2$ is to predefine the numbers at which (1) is calculated. This is achieved by introducing a partition $0 < t_1 < \ldots < t_N = \beta$ of the support of ϕ. Then, $\tilde{\phi}_2$ results in N calculations of ϕ. This defines a closed form approximation of ϕ. Notice that $t_1, \ldots, t_N$ do not need to be equispaced. Finally, $\tilde{\phi}_3$ can be interpreted as a tuning option for a given approximation $\tilde{\phi}_*$ of ϕ:

$$\tilde{\phi}_3(t) := \begin{cases} \tilde{\phi}_*(t) + \delta, & t = 0, \\ \tilde{\phi}_*(t), & t \neq 0, \end{cases} \quad \delta \geq 0.$$

With regard to practical usage, this form of approximation increases numerical stability. Further, it allows for more flexibility in practice, where the number of observations n is given and $\widetilde{\Sigma}_*$ based on $\tilde{\phi}_*$ might not be positive-definite.

Following up the above examples, we picture an approximation $\tilde{\phi}$ of ϕ (respectively approximation $\widetilde{\Sigma}$ of Σ). Several questions arise:

- What are conditions on $\tilde{\phi}$ to ensure that $\widetilde{\Sigma}$ is asymptotically (as $n \to \infty$) equivalent to Σ and eventually (for n large enough) remains positive-definite?

- In terms of ML estimators for covariance parameters, how shall a log-likelihood approximation based on $\tilde{\phi}$ be defined?
- Under which conditions on $\tilde{\phi}$ are ML estimators based on $\tilde{\phi}$ consistent and asymptotically normal?

In the more general setting of a given parametric family of covariance functions, the present study gives a concrete context, where the latter questions are answered by introducing the notion of truncated-ML estimators.

1.4. Framework and contribution

Truncated-ML estimators for covariance parameters are based on truncated-likelihood functions. The latter are defined upon parametric families of sequences of functions, which approximate a presumed family of covariance functions on a common domain. Colloquially we will call these parametric sequences of functions covariance approximations. The respective matrices, constructed upon applying covariance approximations to a given collection of observation points, will be termed covariance matrix approximations. We will allow for covariance matrix approximations that are not necessarily positive semi-definite. Therefore, truncated-likelihood functions are more general than existing likelihood approximations methods such as low-rank, Vecchia, or covariance tapering approaches (see [22] for a summary of commonly used methods).

We work in an increasing-domain asymptotic framework, where collections of observation points are realizations of finite collections of a randomly perturbed regular grid (see also [4]). We consider a stationary Gaussian random field, with a zero-mean function and a true unknown covariance function that belongs to a given parametric family of covariance functions. If the presumed family of covariance functions is compactly supported, we provide sufficient conditions under which truncated-ML estimators and (regular) ML estimators for covariance parameters are consistent and asymptotically normal. Some conditions imposed on families of covariance functions are identical to the conditions that were already considered in [4]. The main difference is that we work with compactly supported covariance functions. Therefore, it is possible to simplify some of the conditions that were set up in [4]. As for statistical applications, we apply these results to the family of generalized Wendland covariance functions. In contrast to the infill-domain asymptotic framework considered in [9], we show that under the studied increasing-domain asymptotic framework, under some conditions on the parameter space, (regular) ML estimators for variance and range parameters are consistent and asymptotically normal. Further, we show that the same asymptotic results are recovered for truncated-ML estimators, based on various generalized Wendland approximations, such as truncations, linear interpolations and added nugget effects.

Additionally, we provide an extension to families of covariance functions which are not compactly supported. We combine our results with the covariance tapering approach. That is, we study covariance taper approximations and their asymptotic influence on the conditional Kullback-Leibler divergence of the

misspecified distribution from the true distribution (see also [5]). We show that the latter divergence is minimized by truncated-tapered ML estimators.

1.5. Structure of the article

The rest of the article is organized as follows. Section 2 establishes the context. We introduce some primary notation, define the sampling domain and the random field itself. In Section 3 we introduce regularity conditions on covariance functions and approximations. In Section 4 we present intermediate asymptotic results on covariance matrices and approximations. Section 5 contains our main results: We introduce truncated-ML estimators and present results on consistency and asymptotic normality. In Section 6, we apply our results to the family of generalized Wendland covariance functions and discuss several examples of generalized Wendland approximations. Then, in the context of non-compactly supported covariance functions, Section 7 contains results on the asymptotic influence of taper approximations on the Kullback-Leibler divergence. Section 8 gives an outlook and some final comments. The Appendix is split into three parts. Covariance approximations for isotropic random fields are discussed in Appendix A. Appendix B contains additional supporting results, whereas all the proofs are left for Appendix C.

2. Context

2.1. Primary notation

The set $\mathbb{N}_+$ and $\mathbb{R}_+$ shall represent the set of positive integers and non-negative real numbers, respectively. For $d \in \mathbb{N}_+$, we use the notation $\mathrm{B}(x; r)$ ($\mathrm{B}[x; r]$) for the open (closed) ball of radius $r > 0$ with center $x \in \mathbb{R}^d$. Given $n \in \mathbb{N}_+$, for some set $A \subset \mathbb{R}^n$, we write $\mathfrak{B}(A)$ for the Borel σ-algebra on A.

For a vector $(w_1, \ldots, w_d) = w \in \mathbb{R}^d$, we write $\|w\| = \left(w_1^2 + \cdots + w_d^2\right)^{1/2}$ for the Euclidean norm of w on $\mathbb{R}^d$. In the case of $d = 1$ we use the notation $|\cdot|$ for the Euclidean norm. For two vectors $w, w' \in \mathbb{R}^d$, $\langle w, w' \rangle = w^\mathrm{t} w' = \sum_{i=1}^d w_i w_i'$ represents the inner product that induces $\|\cdot\|$ on $\mathbb{R}^d$. Given $D \subset \mathbb{R}^d$, we write $\mathcal{B}_\mathrm{C}(D; S)$ for the space of real valued, uniformly bounded functions on D, having compact support $S \subset D$. If $f \in \mathcal{B}_\mathrm{C}(D; S)$ and f is also continuous, we use the notation $\mathcal{C}_\mathrm{C}(D; S)$ instead of $\mathcal{B}_\mathrm{C}(D; S)$. For $f \in \mathcal{C}_\mathrm{C}(D; S)$ we write $\|f\|_\infty = \sup\{|f(h)| : h \in D\}$ for the uniform norm on $\mathcal{C}_\mathrm{C}(D; S)$. For vectors $w \in \mathbb{R}^d$, $|w|_\infty = \max_{i=1,\ldots,d} |w_i|$ denotes the uniform norm on $\mathbb{R}^d$.

For a real $n \times n$ matrix A, $\|A\|_2 = \max_{\{z: z^\mathrm{t} z = 1\}} \langle z, A^\mathrm{t} A z \rangle^{1/2}$ denotes the spectral norm of A. We write $A \succ 0$ ($A \prec 0$) to indicate that A is positive-definite (negative-definite). Further, $\lambda_1(A) \geq \cdots \geq \lambda_n(A)$ denote the n real eigenvalues of a matrix $A \in \mathrm{S}_{n \times n}(\mathbb{R})$, where $\mathrm{S}_{n \times n}(\mathbb{R})$ represents the space of real symmetric $n \times n$ matrices.

We use the notation $\nabla f(x) = \left(\frac{\partial f}{\partial x_1}(x), \ldots, \frac{\partial f}{\partial x_p}(x)\right)$ for the gradient of f at x, where $x \mapsto f(x)$ is any differentiable, real valued function, defined on

some $E \subset \mathbb{R}^p$. Further, for a vector valued, differentiable function $g(x) = (g_1(x), \ldots, g_m(x))$, with values in $\mathbb{R}^m$, defined on some $U \subset \mathbb{R}^p$, we write $J_g(x)_{l,k} = \frac{\partial g_l}{\partial x_k}(x)$, $1 \leq l \leq m$, $1 \leq k \leq p$, for the Jacobi-matrix of g at x.

A mapping Y from a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to a measure space $(E, \mathcal{A})$ will be called a random element if it is $\mathcal{F}/\mathcal{A}$ measurable. If we write that $Y \colon (\Omega, \mathcal{F}) \to (E, \mathcal{A})$ is measurable, we mean that it is $\mathcal{F}/\mathcal{A}$ measurable. If $(Y_n)_{n \in \mathbb{N}_+}$ denotes a sequence of random elements, where for any $n \in \mathbb{N}_+$, Y_n is a mapping form a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ to a measure space $(E, \mathcal{A})$, we use the notation

$$Y_n \xrightarrow[n \to \infty]{\mathbb{P}} Y \text{ and } Y_n \xrightarrow[n \to \infty]{\mathrm{d}} \mathcal{L},$$

to indicate convergence of $(Y_n)_{n \in \mathbb{N}_+}$ to a random element Y in probability and in distribution, respectively. Note that for convergence in distribution, the introduced notation indicates that the limit Y has law $\mathcal{L}$ on $(E, \mathcal{A})$. A sequence of estimators $(\hat{\theta}_n)_{n \in \mathbb{N}_+}$ for $\theta_0 \in \mathbb{R}^p$ will be referred to as consistent if it converges in probability to θ_0. Finally, $\mathcal{N}(\mu, \Sigma)$ indicates a multivariate normal distribution with mean vector μ and covariance matrix Σ.

2.2. Random sampling scheme

On a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, we consider a real valued Gaussian random function Z, which has sample functions on $\mathbb{R}^d$. We assume that Z is stationary (homogeneous) with zero-mean function and covariance function $c_{\theta_0}(s)$, $s \in \mathbb{R}^d$, where $\theta_0 \in \Theta$, with $\Theta \subset \mathbb{R}^p$, compact and convex. Thus, we consider a real valued random field $\{Z_s \colon s \in \mathbb{R}^d\}$, which has true and unknown covariance function c_{θ_0} that belongs to a family of covariance functions $\{c_\theta \colon \theta \in \Theta\}$.

Let $\mathcal{Q} := [-1, 1]^d$ and $X \colon \Omega \to \mathcal{Q}^{\mathbb{N}_+}$ be a stochastic process, defined on the same probability space $(\Omega, \mathcal{F}, \mathbb{P})$, but independent of Z. We assume that the sequence $(X_i)_{i \in \mathbb{N}_+}$ is a sequence of independent random vectors with common law on $\mathcal{Q}$, which has a strictly positive probability density function on $\mathcal{Q}$ (see also Remark 2.1). Given $\tau \in [0, 1/2)$ and a sequence of deterministic points $(v_i)_{i \in \mathbb{N}_+}$, with $v_i \in \mathbb{N}_+^d$, we define a randomly perturbed regular grid S as the process

$$\{S_i := v_i + \tau X_i \colon i \in \mathbb{N}_+\}, \tag{2}$$

where we assume that for all $I \in \mathbb{N}_+$, $\{v_i, 1 \leq i \leq I^d\} = \{1, \ldots, I\}^d$. Therefore, for any $\omega \in \Omega$, $S(\omega)$ is a sequence on $\mathbb{N}_+$, with image $S[\mathbb{N}_+](\omega) \subset \prod_{i=1}^\infty (v_i + \tau \mathcal{Q}) =: \mathcal{G}$ and any first I^d coordinates are in $\{1, \ldots, I\}^d + \tau \mathcal{Q}$ (see also Figure 1). At this point we remark that if nothing is mentioned, the parameter $\tau \in [0, 1/2)$ and the sequence $(v_i)_{i \in \mathbb{N}_+}$ shall be fixed. Let $X_{(n)} := (X_1, \ldots, X_n)$ and $S_{(n)} := (S_1, \ldots, S_n)$ denote finite collections of X and S, respectively. We use the notation $x_{(n)} := (x_1, \ldots, x_n)$ for a vector that contains the first n entries of a given sequence in $\mathcal{Q}^{\mathbb{N}_+}$. Correspondingly, given $\tau \in [0, 1/2)$, $v_1, \ldots, v_n$ and $x_{(n)} \in \mathcal{Q}^n$, we write $s_{(n)} := (s_1, \ldots, s_n)$, $s_i = v_i + \tau x_i$, for n perturbed grid locations in $\mathcal{G}_n := \prod_{i=1}^n (v_i + \tau \mathcal{Q})$.
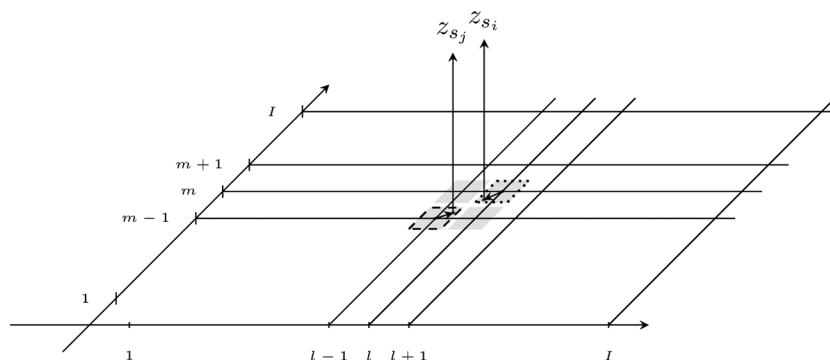
FIG 1. *For $\tau = 0.4$ and $I \in \mathbb{N}_+$, a random field Z is observed at two realizations s_i and s_j of $S_i = v_i + \tau X_i$ and $S_j = v_j + \tau X_j$, $i \neq j$, $1 \leq i, j \leq I^2$. Dotted and dashed lines mark the borders of the ranges of S_i and S_j, respectively.*

On $(\Omega, \mathcal{F}, \mathbb{P})$, we define the random vector

$$\omega \mapsto Z_{(n)}(\omega) := \big(Z_{S_1(\omega)}(\omega), \ldots, Z_{S_n(\omega)}(\omega)\big) = (z_{s_1}, \ldots, z_{s_n}) =: z_{(n)}, \quad (3)$$

which denotes Z observed at a finite collection of S. The situation, where a Gaussian random field is assumed to be observed at a randomly perturbed regular grid, with parameter τ and deterministic points $(v_i)_{i \in \mathbb{N}_+}$, as introduced above, is also considered in [4]. Given $\theta \in \Theta$, we let $\Sigma_\theta(s_{(n)}) := [c_\theta(s_i - s_j)]_{1 \leq i,j \leq n}$ denote the non-random $n \times n$ covariance matrix based on an arbitrary $s_{(n)} \in \mathcal{G}_n$. On $(\Omega, \mathcal{F}, \mathbb{P})$, we write

$$\omega \mapsto \Sigma_{n,\theta}(\omega) := \Sigma_\theta(S_{(n)}(\omega)), \quad \theta \in \Theta,$$

for the $n \times n$ random covariance matrix based on a finite collection $S_{(n)}$ of S.

Remark 2.1. Some technical remarks are worth pointing out. We assume that the random function $Z(s, \omega) := Z_s(\omega)$, $s \in \mathbb{R}^d$, is measurable as a function from the measure space $(\mathbb{R}^d \times \Omega, \mathfrak{B}(\mathbb{R}^d) \otimes \mathcal{F})$ to $(\mathbb{R}, \mathfrak{B}(\mathbb{R}))$. That is to say that Z is (jointly) measurable. This condition makes sure that the components $\omega \mapsto Z_{S_i(\omega)}(\omega) = Z(S_i(\omega), \omega)$, $i = 1, \ldots, n$, of (3) are $\mathcal{F}/\mathfrak{B}(\mathbb{R})$ measurable as the composition of the measurable functions $\omega \mapsto (S_i(\omega), \omega)$ and $(s, \omega) \mapsto Z(s, \omega)$. Thus, the random vector $Z_{(n)}$ is well defined. Since Z and S are independent, it is readily seen that the conditional distribution of $Z_{(n)}$ given $S_{(n)} = s_{(n)}$ is Gaussian, with characteristic function $\exp(-(1/2)a^{\mathrm{t}}\Sigma_{n,\theta_0}(\omega)a)$, $a \in \mathbb{R}^n$. In addition, we note that for fixed $\omega \in \Omega$, $S[\mathbb{N}_+](\omega)$ is not bounded and if we define $\Delta_\tau := 1 - 2\tau$, we are given some fixed $\Delta_\tau > 0$, which is independent of $n \in \mathbb{N}_+$ and $\theta \in \Theta$, such that

$$\inf_{n \in \mathbb{N}_+} \inf_{\substack{1 \leq i,j \leq n \\ i \neq j}} \|s_i - s_j\| \geq \Delta_\tau. \quad (4)$$

Hence, we are in an increasing-domain asymptotic framework where the minimum distance between any two distinct observation points is bounded away

from zero. The assumption that for any given $i \in \mathbb{N}_+$, X_i has strictly positive probability density function on $\mathcal{Q}$, is purely technical (see also the proof of Theorem 5.2). As it can be seen from the mentioned proof, if $\tau = 0$, the assumption becomes redundant.

3. Regularity conditions on covariance functions and covariance approximations

3.1. Regularity conditions on the family of covariance functions

Assumption 3.1 (Regularity conditions on c_θ).

(1) There exist real constants C, $L < \infty$, which are independent of $\theta \in \Theta$, such that $c_\theta \in \mathcal{B}_C(\mathbb{R}^d; S_\theta)$, with $S_\theta \subset \mathrm{B}[0; C]$ and $\|c_\theta\|_\infty \leq L$.

(2) For any $s \in \mathbb{R}^d$, the first, second and third order partial derivatives of $\theta \mapsto c_\theta(s)$ exist. In addition, for any $q = 1, 2, 3$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, $\frac{\partial^q c_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}} \in \mathcal{B}_C(\mathbb{R}^d; S_\theta(i_1, \ldots, i_q))$, where there exist constants C', $L' < \infty$, which are independent of $\theta \in \Theta$, such that $S_\theta(i_1, \ldots, i_q) \subset \mathrm{B}[0; C']$ and $\left\|\frac{\partial^q c_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}\right\|_\infty \leq L'$.

(3) Fourier inversion holds, that is for any $\theta \in \Theta$

$$c_\theta(s) = \int_{\mathbb{R}^d} \hat{c}_\theta(f) \, \mathrm{e}^{\mathrm{i}\langle f, s \rangle} \, \mathrm{d}f,$$

with $\Theta \times \mathbb{R}^d \ni (\theta, f) \mapsto \hat{c}_\theta(f)$ continuous and strictly positive.

Remark 3.1. Note that (1) and (2) of Assumption 3.1 are different to the conditions assumed in [4] (compare also to Condition 3.2 imposed in [5], or Condition 4 stated in [7]). In [4] it is assumed that a given covariance function k_θ is not only bounded on $\mathbb{R}^d$, but also it decays sufficiently fast in the Euclidean norm on $\mathbb{R}^d$. Explicitly, it is assumed in Condition 2.1 of [4] that there exists a finite constant A, which is independent of $\theta \in \Theta$, such that for any $s \in \mathbb{R}^d$, $|k_\theta(s)| \leq A/(1 + \|s\|^{d+1})$. This polynomial decay condition on k_θ can be interpreted as a summability condition on the entries of the respective covariance matrices $K_\theta(s_{(n)})_{i,j} := k_\theta(s_i - s_j)$, which guaranties that the maximal eigenvalues of $K_\theta(s_{(n)})$ are uniformly bounded in $n \in \mathbb{N}_+$, $s_{(n)} \in \mathcal{G}_n$ and $\theta \in \Theta$ (see Lemmas D.1 and D.5 in [4]). Note that the exponent $d + 1$ can be replaced by $d + \alpha$, with $\alpha > 0$ some fixed constant (see also (6) in [6]). In the present study we show that under the assumption of a minimal spacing between any two distinct observation points, if c_θ has compact support on $\mathbb{R}^d$, the number of possible observation points, which are covered by the support of c_θ, must be bounded uniformly in $n \in \mathbb{N}_+$, $s_{(n)} \in \mathcal{G}_n$ and $\theta \in \Theta$ (see Lemma B.1). This, together with the condition that c_θ is also uniformly bounded on Θ and $\mathbb{R}^d$, will be sufficient to conclude that the maximal eigenvalues of $\Sigma_\theta(s_{(n)})$ are uniformly bounded in $n \in \mathbb{N}_+$, $s_{(n)} \in \mathcal{G}_n$ and $\theta \in \Theta$ (see Lemmas 4.1 and B.3). Similar remarks can be made with regard to the conditions imposed on the

partial derivatives of c_θ with respect to θ (see Lemma B.5). In addition, (3) of Assumption 3.1 is also imposed in [4] (compare also to [8] and [7]). It guarantees that the minimal eigenvalues of $\Sigma_\theta(s_{(n)})$ are bounded from below, uniformly in $n \in \mathbb{N}_+$, $s_{(n)} \in \mathcal{G}_n$ and $\theta \in \Theta$ (see Lemmas 4.1 and B.3). Finally, we remark that within the framework of compactly supported covariance functions, the given conditions are very minimal and can be considered as classical in the context of ML estimation. Especially, if one is not interested in the asymptotic distribution, and rather seeks conditions under which ML estimators are consistent (with regard to a concrete example, we refer to Remark 6.2).

3.2. Regularity conditions on the family covariance approximations

Given $\theta \in \Theta$, we let $(\tilde{c}_{m,\theta})_{m \in \mathbb{N}_+}$ denote a sequence of real valued functions defined on $\mathbb{R}^d$. The families $\left\{(\tilde{c}_{m,\theta})_{m \in \mathbb{N}_+} : \theta \in \Theta\right\}$ can be put under the following assumption.

Assumption 3.2 (Regularity conditions on $\tilde{c}_{m,\theta}$).

(1) For any $\theta \in \Theta$ and $m \in \mathbb{N}_+$, the function $\tilde{c}_{m,\theta} \colon \left(\mathbb{R}^d, \mathfrak{B}(\mathbb{R}^d)\right) \to \left(\mathbb{R}, \mathfrak{B}(\mathbb{R})\right)$ is measurable and such that $\tilde{c}_{m,\theta}(s) = \tilde{c}_{m,\theta}(-s)$ for any $s \in \mathbb{R}^d$.

(2) For any $m \in \mathbb{N}_+$, $\tilde{c}_{m,\theta}$ satisfies (1) of Assumption 3.1, where respective constants $\widetilde{C}$ and $\widetilde{L}$ can be further chosen independently of $m \in \mathbb{N}_+$.

(3) $\sup_{\theta \in \Theta} \|\tilde{c}_{m,\theta} - c_\theta\|_\infty \xrightarrow{m \to \infty} 0$.

(4) For any $m \in \mathbb{N}_+$, $\tilde{c}_{m,\theta}$ satisfies (2) of Assumption 3.1, where respective constants $\widetilde{C}'$ and $\widetilde{L}'$ can be further chosen independently of $m \in \mathbb{N}_+$.

(5) For any $q = 1, 2, 3$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, we have that

$$\sup_{\theta \in \Theta} \left\| \frac{\partial^q \tilde{c}_{m,\theta}}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}} - \frac{\partial^q c_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}} \right\|_\infty \xrightarrow{m \to \infty} 0.$$

To make the notation easier, we write $(\tilde{c}_{m,\theta}) \coloneqq (\tilde{c}_{m,\theta})_{m \in \mathbb{N}_+}$. In the following, we formally introduce covariance matrix approximations (random and non-random versions). To do so, let $r \colon \mathbb{N}_+ \to \mathbb{N}_+$ be such that $r(n) \to \infty$ as $n \to \infty$. Given $s_{(n)} \in \mathcal{G}_n$, we let $\widetilde{\Sigma}_\theta(s_{(n)}) \coloneqq [\tilde{c}_{r(n),\theta}(s_i - s_j)]_{1 \le i,j \le n}$ denote the non-random $n \times n$ matrix based on a given family $\left\{(\tilde{c}_{m,\theta}) \colon \theta \in \Theta\right\}$. Then, on $(\Omega, \mathcal{F}, \mathbb{P})$, if $\left\{(\tilde{c}_{m,\theta}) \colon \theta \in \Theta\right\}$ is a family of Borel measurable sequences of functions, we write

$$\omega \mapsto \widetilde{\Sigma}_{n,\theta}(\omega) \coloneqq \widetilde{\Sigma}_\theta(S_{(n)}(\omega)),$$

for the $n \times n$ random matrix based on a finite collection $S_{(n)}$ of S. Colloquially we will use the term covariance approximation when we refer to a given family $\left\{(\tilde{c}_{m,\theta}) \colon \theta \in \Theta\right\}$, which can approximate a family of covariance functions $\{c_\theta \colon \theta \in \Theta\}$ in the sense of Assumption 3.2. In these terms $\{c_\theta \colon \theta \in \Theta\}$ itself is a covariance approximation. The expression covariance matrix approximation will be used for both, $\widetilde{\Sigma}_\theta(s_{(n)})$ and its random version $\widetilde{\Sigma}_{n,\theta}$. Similar, we use the expression covariance matrix for both, $\Sigma_\theta(s_{(n)})$ and $\Sigma_{n,\theta}$.

Remark 3.2. (1), (2) and (4) of Assumption 3.2 are natural extensions of (1) and (2) of Assumption 3.1. Notice that the measurability condition imposed in (1) of Assumption 3.2 makes sure that $\widetilde{\Sigma}_{n,\theta}$ is $\mathcal{F}/\mathfrak{B}(\mathbb{R}^{n^2})$ measurable. Condition (3) of Assumption 3.2 specifies in which sense a family $\{(\tilde{c}_{m,\theta}) \colon \theta \in \Theta\}$ approximates the family $\{c_\theta \colon \theta \in \Theta\}$. We require that $(\tilde{c}_{m,\theta})$ converges uniformly on $\mathbb{R}^d$ to c_θ, where the convergence is also uniform on the parameter space Θ. In fact, we will show (see Lemmas B.3 and 4.1) that the uniform convergence of $(\tilde{c}_{m,\theta})$ to c_θ, together with the condition that the families $\{c_\theta \colon \theta \in \Theta\}$ and $\{(\tilde{c}_{m,\theta}) \colon \theta \in \Theta\}$ have uniformly bounded compact support, are, among others, sufficient criteria to proof that the matrices $\Sigma_\theta(s_{(n)})$ and $\widetilde{\Sigma}_\theta(s_{(n)})$ are asymptotically (as $n \to \infty$) equivalent, uniformly on Θ and $\mathcal{G}$. Condition (5) of Assumption 3.2 will allow us to conclude that a similar result holds true for the first, second and third order partial derivatives (with respect to θ) of $\widetilde{\Sigma}_\theta(s_{(n)})$ and $\Sigma_\theta(s_{(n)})$. For concrete examples of covariance approximations, where the conditions of Assumption 3.2 are verified, we refer to Section 6.

4. Uniform asymptotic equivalence of covariance matrices and covariance matrix approximations

This section presents intermediate results on covariance matrices and approximations. In particular, Lemma 4.1 gives precise conditions under which $\widetilde{\Sigma}_{n,\theta}$ eventually (for n large enough) remains positive-definite with $\mathbb{P}$ probability one.

Lemma 4.1. *Assume that the family $\{c_\theta \colon \theta \in \Theta\}$ satisfies (1) and (3) of Assumption 3.1. Consider $\{(\tilde{c}_{m,\theta}) \colon \theta \in \Theta\}$ that satisfies (1), (2) and (3) of Assumption 3.2. Then, we have that $\mathbb{P}$ a.s.*

$$\sup_{n \in \mathbb{N}_+} \sup_{\theta \in \Theta} \|\Sigma_{n,\theta}\|_2 < \infty \ \text{and} \ \sup_{n \in \mathbb{N}_+} \sup_{\theta \in \Theta} \|\widetilde{\Sigma}_{n,\theta}\|_2 < \infty.$$

In particular we can conclude that $\mathbb{P}$ a.s.

$$\sup_{\theta \in \Theta} \|\Sigma_{n,\theta} - \widetilde{\Sigma}_{n,\theta}\|_2 \xrightarrow{n \to \infty} 0.$$

Further, it is true that $\mathbb{P}$ a.s.

$$\inf_{n \in \mathbb{N}_+} \inf_{\theta \in \Theta} \lambda_n(\Sigma_{n,\theta}) > 0,$$

and there exists $N \in \mathbb{N}_+$ such that $\mathbb{P}$ a.s.

$$\inf_{n \geq N} \inf_{\theta \in \Theta} \lambda_n(\widetilde{\Sigma}_{n,\theta}) > 0.$$

5. Truncated-ML estimators

Given a square matrix A, we define $\det_+(A)$ to be the product of the strictly positive eigenvalues of A. If all of the eigenvalues are less or equal to zero,

$\det_+(A) = 1$. Further, we use the notation A^+ for the pseudoinverse of A (sometimes called Moore-Penrose inverse). For the given collection $\{c_\theta : \theta \in \Theta\}$, we define, on $(\Omega, \mathcal{F}, \mathbb{P})$, for any $n \in \mathbb{N}_+$ and $\theta \in \Theta$, the random variable

$$l_n(\theta) := \frac{1}{n} \log \left(\det_+(\Sigma_{n,\theta}) \right) + \frac{1}{n} \langle Z_{(n)}, \Sigma_{n,\theta}^+ Z_{(n)} \rangle. \tag{5}$$

Given $\omega \in \Omega$, $\theta \mapsto l_n(\theta)(\omega)$ shall be called the *truncated-modified log-likelihood function* based on $\{c_\theta : \theta \in \Theta\}$. A sequence of estimators $(\hat{\theta}_n(c))_{n \in \mathbb{N}_+}$, defined on $(\Omega, \mathcal{F}, \mathbb{P})$, will be called a sequence of *truncated-ML estimators* for θ_0 based on $\{c_\theta : \theta \in \Theta\}$, if for any $n \in \mathbb{N}_+$,

$$\hat{\theta}_n(c) \in \operatorname*{argmin}_{\theta \in \Theta} l_n(\theta).$$

Similarly, on $(\Omega, \mathcal{F}, \mathbb{P})$, for a given collection of sequences of real valued functions $\{(\tilde{c}_{m,\theta}) : \theta \in \Theta\}$, we introduce, for any $n \in \mathbb{N}_+$ and $\theta \in \Theta$, the random variable

$$\tilde{l}_n(\theta) := \frac{1}{n} \log \left(\det_+(\widetilde{\Sigma}_{n,\theta}) \right) + \frac{1}{n} \langle Z_{(n)}, \widetilde{\Sigma}_{n,\theta}^+ Z_{(n)} \rangle. \tag{6}$$

Then, for $\omega \in \Omega$, the function $\theta \mapsto \tilde{l}_n(\theta)(\omega)$ denotes the truncated-modified log-likelihood function based on $\{(\tilde{c}_{m,\theta}) : \theta \in \Theta\}$. A sequence of estimators $(\hat{\theta}_n(\tilde{c}))_{n \in \mathbb{N}_+}$, defined on $(\Omega, \mathcal{F}, \mathbb{P})$, will be called a sequence of truncated-ML estimators for θ_0 based on $\{(\tilde{c}_{m,\theta}) : \theta \in \Theta\}$, if for any $n \in \mathbb{N}_+$

$$\hat{\theta}_n(\tilde{c}) \in \operatorname*{argmin}_{\theta \in \Theta} \tilde{l}_n(\theta). \tag{7}$$

At this point is important to note that for a given $\omega \in \Omega$, it is in general not true that $l_n(\theta)(\omega)$ and $\tilde{l}_n(\theta)(\omega)$ are continuous in θ for any $n \in \mathbb{N}_+$. Nevertheless, a consequence of Lemma 4.1 is the following proposition:

Proposition 5.1. *Assume that the family* $\{c_\theta : \theta \in \Theta\}$ *satisfies (1) and (3) of Assumption 3.1. Consider* $\{(\tilde{c}_{m,\theta}) : \theta \in \Theta\}$ *that satisfies (1), (2) and (3) of Assumption 3.2. Then, we have that for any* $n \in \mathbb{N}_+$, $\mathbb{P}$ *a.s.,*

$$l_n(\theta) = \frac{1}{n} \log \left(\det(\Sigma_{n,\theta}) \right) + \frac{1}{n} \langle Z_{(n)}, \Sigma_{n,\theta}^{-1} Z_{(n)} \rangle.$$

Further there exists $N \in \mathbb{N}_+$ *such that for any* $n \geq N$, $\mathbb{P}$ *a.s.,*

$$\tilde{l}_n(\theta) = \frac{1}{n} \log \left(\det(\widetilde{\Sigma}_{n,\theta}) \right) + \frac{1}{n} \langle Z_{(n)}, \widetilde{\Sigma}_{n,\theta}^{-1} Z_{(n)} \rangle,$$

and we have that

$$\sup_{\theta \in \Theta} \left| l_n(\theta) - \tilde{l}_n(\theta) \right| \xrightarrow[n \to \infty]{\mathbb{P}} 0.$$

Using Proposition 5.1, we notice that if, for any $s \in \mathbb{R}^d$ and $m \in \mathbb{N}_+$, both $\theta \mapsto c_\theta(s)$ and $\theta \mapsto \tilde{c}_{m,\theta}(s)$ are k times differentiable, we have that $\theta \mapsto l_n(\theta)(\omega)$ and $\theta \mapsto \tilde{l}_n(\theta)(\omega)$ are k times differentiable for n large enough, respectively.

For the rest of the article, if we refer to truncated-ML estimators (without mentioning further whether estimators are based on families of covariance functions or approximations), we refer to both, truncated-ML estimators based on families of covariance functions and approximations. The same is applied for the notion of truncated-modified log-likelihood functions based on either covariance functions or approximations. However, if $\{c_\theta \colon \theta \in \Theta\}$ satisfies the assumptions of Proposition 5.1, a sequence of truncated-ML estimators $\big(\hat{\theta}_n(c)\big)_{n \in \mathbb{N}_+}$ shall be simply called a sequence of ML estimators for θ_0. Similarly, we will simply refer to a modified log-likelihood function when the given family $\{c_\theta \colon \theta \in \Theta\}$ is under the assumptions of Proposition 5.1.

Remark 5.1. The introduction of truncated-modified log-likelihood functions is not standard. Modified refers to the fact that the log-likelihood for the Gaussian density function of a random vector $(Z_{s_1}, \ldots, Z_{s_n})$ is scaled by $-2/n$. This is common practice in the literature about ML estimators for covariance parameters under an increasing-domain asymptotic framework (see for instance [4, 5] and also [7]). The matrices $\Sigma_{n,\theta}(\omega)$ and $\widetilde{\Sigma}_{n,\theta}(\omega)$ are not necessarily positive-definite. In particular, $\widetilde{\Sigma}_{n,\theta}(\omega)$ can be negative-definite. If the matrices $\Sigma_{n,\theta}(\omega)$ and $\widetilde{\Sigma}_{n,\theta}(\omega)$ are not positive-definite, we truncate the log-likelihood by a pseudo-determinant and -inverse to obtain the functions $\theta \mapsto l_n(\theta)(\omega)$ and $\theta \mapsto \tilde{l}_n(\theta)(\omega)$. Hence, the use of the expression "truncated".

Remark 5.2. As it was mentioned in Remark 2.2 of [5], for given $\omega \in \Omega$, we allow the functions $\theta \mapsto l_n(\theta)(\omega)$ and $\theta \mapsto \tilde{l}_n(\theta)(\omega)$ to have more than one minimizer. In which case the asymptotic results given in Section 5.1 hold true for any given sequence of truncated-ML estimators. With regard to the existence of a minimizer we refer to Remark 2.1 in [4].

5.1. Consistency and asymptotic normality of truncated-ML estimators

The main results of this section are that under suitable conditions on the families of covariance functions and approximations, truncated-ML estimators for covariance parameters are not only consistent (Theorem 5.2 and Corollary 5.3) but also asymptotically normal (Theorem 5.4 and Corollary 5.5). In particular, we will make use of the conditions presented in Assumptions 3.1 and 3.2. However, in the context of random fields that are observed at randomly perturbed regular grid locations as defined in (2), we will further make use of the following two technical conditions that were also imposed in [4]. Associated to the common range $\mathcal{Q}$, of the process X, we define the set $D_\tau := \bigcup_{z \in \mathbb{Z}^d \setminus \{0\}} (z + \tau U_\mathcal{Q})$, where $U_\mathcal{Q} := \{u_1 - u_2 \colon u_1 \in \mathcal{Q}, u_2 \in \mathcal{Q}\}$ denotes the set of differences between two points in $\mathcal{Q}$.

Assumption 5.1 (Asymptotic identifiability around θ_0). For $\tau = 0$, there does not exists $\theta \neq \theta_0$ such that $c_\theta(z) = c_{\theta_0}(z)$ for all $z \in \mathbb{Z}^d$. If $\tau \neq 0$, there does not exists $\theta \neq \theta_0$ such that $s \mapsto c_\theta(s) - c_{\theta_0}(s)$ is zero a.e. with respect to the Lebesgue measure on D_τ and $c_\theta(0) = c_{\theta_0}(0)$.

Assumption 5.2 (Local identifiability around θ_0). For $\tau = 0$, there does not exists $\mathbb{R}^p \setminus \{0\} \ni \alpha = (\alpha_1, \dots, \alpha_p)$ such that $\sum_{k=1}^p \alpha_k \frac{\partial c_{\theta_0}}{\partial \theta_k}(z) = 0$ for all $z \in \mathbb{Z}^d$. For $\tau \neq 0$, there does not exists $\mathbb{R}^p \setminus \{0\} \ni \alpha = (\alpha_1, \dots, \alpha_p)$ such that $s \mapsto \sum_{k=1}^p \alpha_k \frac{\partial c_{\theta_0}}{\partial \theta_k}(s)$ is zero a.e. with respect to the Lebesgue measure on D_τ and $\sum_{k=1}^p \alpha_k \frac{\partial c_{\theta_0}}{\partial \theta_k}(0) = 0$.

Theorem 5.2. *Let* $\big(\hat{\theta}_n(\tilde{c})\big)_{n \in \mathbb{N}_+}$ *be a sequence of truncated-ML estimators for* θ_0 *based on* $\big\{(\tilde{c}_{m,\theta}) \colon \theta \in \Theta\big\}$. *Assume that* $\{c_\theta \colon \theta \in \Theta\}$ *satisfies Assumption 3.1 (regarding (2), $q = 1$ and the continuity of first order partial derivatives is sufficient) and Assumption 5.1. Suppose further that* $\big\{(\tilde{c}_{m,\theta}) \colon \theta \in \Theta\big\}$ *satisfies Assumption 3.2 (regarding (4) and (5), $q = 1$ and the continuity of first order partial derivatives is sufficient). Then, we have that*

$$\hat{\theta}_n(\tilde{c}) \xrightarrow[n \to \infty]{\mathbb{P}} \theta_0.$$

The following corollary is immediate.

Corollary 5.3. *Suppose that* $\{c_\theta \colon \theta \in \Theta\}$ *satisfies Assumption 3.1 (regarding (2), $q = 1$ and the continuity of first order partial derivatives is sufficient) and Assumption 5.1. Then, we can conclude that a sequence of ML estimators* $\big(\hat{\theta}_n(c)\big)_{n \in \mathbb{N}_+}$ *for* θ_0 *is consistent.*

Before we present the results about asymptotic normality, it is helpful to consider some additional notation. Let $K \in \mathbb{N}_+$, such that for any $\omega \in \Omega$, the sequences of functions

$$\big(l_{n,K}(\theta)(\omega)\big)_{n \in \mathbb{N}_+} := \big(l_{n+K-1}(\theta)(\omega)\big)_{n \in \mathbb{N}_+} \tag{8}$$

and

$$\big(\tilde{l}_{n,K}(\theta)(\omega)\big)_{n \in \mathbb{N}_+} := \big(\tilde{l}_{n+K-1}(\theta)(\omega)\big)_{n \in \mathbb{N}_+} \tag{9}$$

are differentiable with respect to θ. Note that if $\{c_\theta \colon \theta \in \Theta\}$ satisfies Assumption 3.1 and the collection $\big\{(\tilde{c}_{m,\theta}) \colon \theta \in \Theta\big\}$ satisfies Assumption 3.2, then we know about the existence of such a K under application of Proposition 5.1. For the given $K \in \mathbb{N}_+$, on $(\Omega, \mathcal{F}, \mathbb{P})$, we introduce the sequence of random functions

$$\big\{\big(\underbrace{\omega \mapsto G_{n,K}(\omega, \theta)}_{=: G_{n,K}(\theta)}\big)_{n \in \mathbb{N}_+} \colon \theta \in \Theta\big\},$$

where for $n \in \mathbb{N}_+$ and $\theta \in \Theta$, the random vector $G_{n,K}(\theta)$ has components $G_{j,n,K}(\theta)$, $j = 1, \dots, p$, with

$$G_{j,n,K}(\theta) = \frac{\partial l_{n,K}}{\partial \theta_j}(\theta) - \mathbb{E}\left[\frac{\partial l_{n,K}}{\partial \theta_j}(\theta) \;\middle|\; S_{(n)}\right],$$

and thus

$$G_{n,K}(\theta) = \nabla l_{n,K}(\theta) - \mathbb{E}\big[\nabla l_{n,K}(\theta) \mid S_{(n)}\big]. \tag{10}$$

Similarly, on $(\Omega, \mathcal{F}, \mathbb{P})$, we introduce the sequence of random functions

$$\big\{ \big(\underbrace{\omega \mapsto \widetilde{G}_{n,K}(\omega, \theta)}_{=:\widetilde{G}_{n,K}(\theta)} \big)_{n \in \mathbb{N}_+} : \theta \in \Theta \big\},$$

where for any $n \in \mathbb{N}_+$ and $\theta \in \Theta$, the components of $\widetilde{G}_{n,K}(\theta)$ are given by

$$\widetilde{G}_{j,n,K}(\theta) = \frac{\partial \widetilde{l}_{n,K}}{\partial \theta_j}(\theta) - \mathbb{E}\bigg[\frac{\partial \widetilde{l}_{n,K}}{\partial \theta_j}(\theta) \,\bigg|\, S_{(n)}\bigg], \quad j = 1, \ldots, p,$$

and thus

$$\widetilde{G}_{n,K}(\theta) = \nabla \widetilde{l}_{n,K}(\theta) - \mathbb{E}\big[\nabla \widetilde{l}_{n,K}(\theta) \mid S_{(n)}\big]. \tag{11}$$

If the collection $\{c_\theta : \theta \in \Theta\}$ satisfies Assumption 3.1, we simply write, for any $n \in \mathbb{N}_+$,

$$J_{G_n}(\theta_0) := J_{G_{n,1}}(\theta_0),$$

for the random Jacobi-matrix of $\theta \mapsto G_{n,1}(\theta)$ evaluated at θ_0.

Theorem 5.4. *Let $\big(\hat{\theta}_n(\tilde{c})\big)_{n \in \mathbb{N}_+}$ be an sequence of truncated-ML estimators for θ_0 based on $\big\{(\tilde{c}_{m,\theta}) : \theta \in \Theta\big\}$. Suppose that $\{c_\theta : \theta \in \Theta\}$ satisfies Assumptions 3.1, 5.1 and 5.2. Suppose further that $\big\{(\tilde{c}_{m,\theta}) : \theta \in \Theta\big\}$ satisfies Assumption 3.2. Then, we have that*

$$n^{1/2}\big(\hat{\theta}_n(\tilde{c}) - \theta_0\big) \xrightarrow[n \to \infty]{\mathrm{d}} \mathcal{N}(0, \Lambda^{-1}), \tag{12}$$

where $\mathrm{S}_{p \times p} \ni \Lambda \succ 0$ is deterministic and such that

$$\frac{1}{2}J_{G_n}(\theta_0) \xrightarrow[n \to \infty]{\mathbb{P}} \Lambda \xleftarrow[n \to \infty]{\mathbb{P}} \frac{1}{2}J_{\widetilde{G}_{n,N}}(\theta_0),$$

with $N \in \mathbb{N}_+$ as in Proposition 5.1.

Corollary 5.5. *Suppose that $\{c_\theta : \theta \in \Theta\}$ satisfies Assumptions 3.1, 5.1 and 5.2. Then, we can conclude that a sequence of ML estimators $\big(\hat{\theta}_n(c)\big)_{n \in \mathbb{N}_+}$ for θ_0 is such that*

$$n^{1/2}(\hat{\theta}_n(c) - \theta_0) \xrightarrow[n \to \infty]{\mathrm{d}} \mathcal{N}(0, \Lambda^{-1}),$$

with Λ as in Theorem 5.4.

Remark 5.3. Under Assumptions 3.1, for any $K \in \mathbb{N}_+$, $\mathbb{E}\big[\nabla l_{n,K}(\theta) \mid S_{(n)}\big] = 0$ with $\mathbb{P}$ probability one. However, even if $\{c_\theta : \theta \in \Theta\}$ is under Assumptions 3.1 and $\big\{(\tilde{c}_{m,\theta}) : \theta \in \Theta\big\}$ is under Assumptions 3.2, it is not in general true that $\mathbb{P}$ a.s. $\mathbb{E}\big[\nabla \widetilde{l}_{n,N}(\theta) \mid S_{(n)}\big] = 0$, where N is as in Proposition 5.1. Notice further that under Assumptions 3.1, for $\omega \in \Omega$, $-(n/2)J_{G_n}(\theta_0)(\omega)$ represents the second derivative of the log-likelihood $\theta \mapsto -(n/2)l_n(\theta)(\omega)$ based on $\{c_\theta : \theta \in \Theta\}$.

6. Example of application: generalized Wendland functions

In this section we work in the same setting as in Section 2.2, but we additionally assume that Z is isotropic. Explicitly, for the given family of covariance functions $\{c_\theta : \theta \in \Theta\}$, we assume that there exists a parametric family $\{\varphi_\theta : \theta \in \Theta\}$ such that for any $\theta \in \Theta$, $s \in \mathbb{R}^d$, $c_\theta(s) = \varphi_\theta(\|s\|)$. The family $\{\varphi_\theta : \theta \in \Theta\}$ is called the radial version of $\{c_\theta : \theta \in \Theta\}$. We can recycle the notation of Section 3 and easily translate Assumptions 3.1 and 3.2 by considering families of approximations $\{(\tilde{\varphi}_{m,\theta}) : \theta \in \Theta\}$ for $\{\varphi_\theta : \theta \in \Theta\}$ on $\mathbb{R}_+$. This allows us to readily recover the results of Sections 4 and 5 for isotropic random fields. For the details we refer to Assumptions A.1 and A.2, as well as Theorems A.1 and A.2 in Appendix A.

In terms of an explicit family of radial covariance functions, we reconsider the generalized Wendland covariance function which we have already introduced in (1) of Section 1.3. Let $\Theta \ni \theta := (\sigma^2, \beta)$, where $\Theta := [\sigma^2_{\min}, \sigma^2_{\max}] \times [\beta_{\min}, \beta_{\max}]$, with $0 < \sigma^2_{\min} < \sigma^2_{\max} < \infty$ and $1 - 2\tau < \beta_{\min} < \beta_{\max} < \infty$. We assume that the covariance function of the random field Z is given by $\phi_{\theta_0}(\|s\|)$, $s \in \mathbb{R}^d$, $\theta_0 \in \Theta$, where ϕ_{θ_0} belongs to the family $\{\phi_\theta : \theta \in \Theta\}$ which is defined by

$$\phi_\theta(t) := \sigma^2 \phi_{\nu,\kappa}\left(\frac{t}{\beta}\right), \quad t \in [0,\infty), \tag{13}$$

where

$$\phi_{\nu,\kappa}(r) := \begin{cases} \frac{1}{\mathrm{B}(2\kappa, \nu+1)} \int_r^1 u(u^2 - r^2)^{\kappa-1}(1-u)^\nu du, & r \in [0,1), \\ 0, & r \in [1,\infty), \end{cases}$$

compare to (1) of Section 1.3. We treat κ and ν as given but such that $\kappa > 0$ and $\nu \geq (d+1)/2 + \kappa$. Notice that the latter restriction on κ and ν makes sure that for any $\theta \in \Theta$, ϕ_θ belongs to the class Φ_d, the class of real valued and continuous functions, defined on $\mathbb{R}_+$, which are strictly positive at the origin and such that for any finite collection of points in $\mathbb{R}^d$, evaluation at the Euclidean norm of pairwise differences between points of the collection results in a nonnegative definite matrix (see for example [21]). Actually, in the latter reference it is argued that for $\kappa > 0$, $\phi_{\nu,\kappa} \in \Phi_d$ if and only if $\nu \geq (d+1)/2 + \kappa$. For the respective family defined on $\mathbb{R}^d$, we use the notation $w_\theta(s) := \phi_\theta(\|s\|)$.

Remark 6.1. The restriction $\beta_{\min} > 1 - 2\tau$ is imposed to proof that the family $\{w_\theta : \theta \in \Theta\}$ satisfies Assumptions 5.1 and 5.2 (see the proof of Propositions 6.2). This is not surprising, as $1 - 2\tau$ defines the minimal spacing between pairs of distinct observation points of the randomly perturbed regular grid, defined in (2) of Section 2.2. Further, as we have noted that $\phi_{\nu,\kappa} \in \Phi_d$ if and only if $\nu \geq (d+1)/2 + \kappa$, the two smoothness parameters ν and κ can not be estimated without further constraints.

Proposition 6.1. *Let $\kappa > 4$. Then, the family $\{\phi_\theta : \theta \in \Theta\}$ satisfies Assumption A.1, where for any $\theta \in \Theta$ and for any $q = 1, 2, 3$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, the functions $t \mapsto \phi_\theta(t)$ and $t \mapsto \frac{\partial^q \phi_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}(t)$ are continuous on $\mathbb{R}_+$.*

Proposition 6.2. *Let $\kappa > 2$. Then, the family $\{w_\theta \colon \theta \in \Theta\}$ satisfies Assumptions 5.1 and 5.2.*

Using Propositions 6.1 and 6.2, under application of Theorems A.1 and A.2 (recall also Corollaries 5.3 and 5.5), we obtain the following result:

Proposition 6.3. *Let $\kappa > 4$. A sequence $\big(\hat{\theta}_n(\phi)\big)_{n \in \mathbb{N}_+}$ of ML estimators for θ_0 based on $\{\phi_\theta \colon \theta \in \Theta\}$ is consistent. Further there exists a non-random symmetric $p \times p$ matrix $\Lambda \succ 0$ such that*

$$n^{1/2}\big(\hat{\theta}_n(\phi) - \theta_0\big) \xrightarrow[n \to \infty]{\mathrm{d}} \mathcal{N}(0, \Lambda^{-1}).$$

Remark 6.2. It is worth to note that the restriction $\kappa > 4$ is only needed for the asymptotic distribution of ML estimators, respectively truncated-ML estimators. In particular, in Proposition 6.1, if one only demands conditions involving first order partial derivatives of ϕ_θ, with respect to θ, $\kappa > 2$ is sufficient. With regard to consistency of the estimator $\big(\hat{\theta}_n(\phi)\big)_{n \in \mathbb{N}_+}$ in Proposition 6.3, $\kappa > 2$ is sufficient as well. The same applies for the truncated-ML estimators considered in Examples 6.1, 6.2, 6.3 and 6.4. Keeping in mind the differentiability conditions imposed in Assumption A.1, the given restrictions on κ are not surprising (compare also to [9], within the infill-domain asymptotic framework).

We discuss four examples of generalized Wendland approximations.

Example 6.1 (Truncation of ϕ_θ). Let $\{\phi_\theta \colon \theta \in \Theta\}$ be as in Proposition 6.1. Let $\big\{(\mathfrak{T}_{m,\theta}) \colon \theta \in \Theta\big\}$ be defined as follows: For $\theta \in \Theta$ and $m \in \mathbb{N}_+$, we set,

$$\mathfrak{T}_{m,\theta}(t) := \phi_\theta \mathbb{1}_{[0, C_m]}(t), \quad t \in \mathbb{R}_+, \ C_m \to \infty \text{ as } m \to \infty.$$

Proposition 6.4. *A sequence $\big(\hat{\theta}_n(\mathfrak{T})\big)_{n \in \mathbb{N}_+}$ of truncated-ML estimators for θ_0 based on $\big\{(\mathfrak{T}_{m,\theta}) \colon \theta \in \Theta\big\}$ is consistent and we have that*

$$n^{1/2}\big(\hat{\theta}_n(\mathfrak{T}) - \theta_0\big) \xrightarrow[n \to \infty]{\mathrm{d}} \mathcal{N}(0, \Lambda^{-1}),$$

where Λ is defined as in Proposition 6.3.

In the following we let $M < \infty$ denote a real constant, which is independent of $\beta \in [\beta_{\min}, \beta_{\max}]$ such that $\beta_{\max} \leq M$.

Example 6.2 (Trimmed Bernstein polynomials). Let $\{\phi_\theta \colon \theta \in \Theta\}$ be as in Proposition 6.1. We consider a family $\big\{(\mathfrak{P}_{m,\theta}) \colon \theta \in \Theta\big\}$ defined as follows: For $\theta \in \Theta$ and $m \in \mathbb{N}_+$, we set for $t \in \mathbb{R}_+$,

$$\mathfrak{P}_{m,\theta}(t) := \begin{cases} B_{m,\theta}(t; b_m), & t \leq M, \\ 0, & t > M, \end{cases}$$

with

$$B_{m,\theta}(t; b_m) = \sum_{k=0}^{m} \phi_\theta\left(b_m \frac{k}{m}\right) \binom{m}{k} \left(\frac{t}{b_m}\right)^k \left(1 - \frac{t}{b_m}\right)^{m-k},$$

the Bernstein polynomial of the function ϕ_θ on $[0, b_m)$, where $b_m \xrightarrow{m \to \infty} \infty$ and we assume that $b_m = o(m)$. Thus, for any $0 \leq k \leq m$,

$$b_m \frac{k+1}{m} - b_m \frac{k}{m} \xrightarrow{m \to \infty} 0,$$

the distance between adjacent points converge to zero as m approaches infinity. See also [12] for an introduction of Bernstein polynomials on unbounded intervals.

Proposition 6.5. *The family* $\{(\mathfrak{P}_{m,\theta}) \colon \theta \in \Theta\}$ *satisfies Assumption A.2.*

Using Propositions 6.1, 6.2 and 6.5, under application of Theorems A.1 and A.2, we have proven the following result:

Proposition 6.6. *A sequence* $(\hat{\theta}_n(\mathfrak{P}))_{n \in \mathbb{N}_+}$ *of truncated-ML estimators for* θ_0 *based on* $\{(\mathfrak{P}_{m,\theta}) \colon \theta \in \Theta\}$ *is consistent and we have that*

$$n^{1/2}(\hat{\theta}_n(\mathfrak{P}) - \theta_0) \xrightarrow[n \to \infty]{\mathrm{d}} \mathcal{N}(0, \Lambda^{-1}),$$

where Λ *is defined as in Proposition 6.3.*

Example 6.3 (Linear interpolation). Let $\{\phi_\theta \colon \theta \in \Theta\}$ be as in Proposition 6.1. For a given $m \in \mathbb{N}_+$, we consider a partition of the interval $[0, M]$, $0 = t_0 \leq t_1 \leq \cdots \leq t_{N_m} = M$, where $N_m \xrightarrow{m \to \infty} \infty$ and for $0 \leq k \leq N_m$, $t_{k+1}^m - t_k^m \xrightarrow{m \to \infty} 0$. Then, we define the family $\{(\mathfrak{L}_{m,\theta}) \colon \theta \in \Theta\}$ as follows: For $\theta \in \Theta$ and $m \in \mathbb{N}_+$, we set for $t \in \mathbb{R}_+$,

$$\mathfrak{L}_{m,\theta}(t) := \begin{cases} I_{m,\theta}(t; N_m), & t \leq M, \\ 0, & t > M, \end{cases}$$

where

$$I_{m,\theta}(t; N_m) = \begin{cases} \phi_\theta(t_k^m) + \frac{\phi_\theta(t_{k+1}^m) - \phi_\theta(t_k^m)}{t_{k+1}^m - t_k^m}(t - t_k^m), & t \in [t_k^m, t_{k+1}^m], \\ 0, & t \notin [t_k^m, t_{k+1}^m]. \end{cases}$$

Thus, for a given $m \in \mathbb{N}_+$, $\mathfrak{L}_{m,\theta}$ represents a linear interpolation of the function ϕ_θ on the interval $[0, M]$.

Proposition 6.7. *The family* $\{(\mathfrak{L}_{m,\theta}) \colon \theta \in \Theta\}$ *satisfies Assumption A.2.*

Using Propositions 6.1, 6.2 and 6.7, under application of Theorems A.1 and A.2, we have further proven the following result:

Proposition 6.8. *A sequence* $(\hat{\theta}_n(\mathfrak{L}))_{n \in \mathbb{N}_+}$ *of truncated-ML estimators for* θ_0 *based on* $\{(\mathfrak{L}_{m,\theta}) \colon \theta \in \Theta\}$ *is consistent and we have that*

$$n^{1/2}(\hat{\theta}_n(\mathfrak{L}) - \theta_0) \xrightarrow[n \to \infty]{\mathrm{d}} \mathcal{N}(0, \Lambda^{-1}),$$

where Λ *is defined as in Proposition 6.3.*

Example 6.4 (Vanishing nugget effect). Let $\{\phi_\theta : \theta \in \Theta\}$ be as in Proposition 6.1 and consider a family $\big\{\big(\tilde{\phi}_{m,\theta}\big) : \theta \in \Theta\big\}$ that satisfies Assumption A.2. Then, define for any $\theta \in \Theta$ and $m \in \mathbb{N}_+$, the function

$$\mathfrak{S}_{m,\theta}(t) := \begin{cases} \tilde{\phi}_{m,\theta}(t) + \delta(m), & t = 0, \\ \tilde{\phi}_{m,\theta}(t), & t \neq 0, \end{cases} \tag{14}$$

where $(\delta(m))_{m \in \mathbb{N}_+}$ is independent of $\theta \in \Theta$ and $t \in \mathbb{R}_+$ and such that $\delta(m) \to 0$, as $m \to \infty$. Note that since the family $\{\phi_\theta : \theta \in \Theta\}$ satisfies Assumption A.2, we could also choose $\tilde{\phi}_{m,\theta} \equiv \phi_\theta$ in (14).

Proposition 6.9. *A sequence $\big(\hat{\theta}_n\big(\mathfrak{S}\big)\big)_{n \in \mathbb{N}_+}$ of truncated-ML estimators for θ_0 based on $\big\{\big(\mathfrak{S}_{m,\theta}\big) : \theta \in \Theta\big\}$ is consistent and we have that*

$$n^{1/2}\big(\hat{\theta}_n(\mathfrak{S}) - \theta_0\big) \xrightarrow[n \to \infty]{\mathrm{d}} \mathcal{N}(0, \Lambda^{-1}),$$

where Λ is defined as in Proposition 6.3.

Remark 6.3. As it was already mentioned in the introduction, computing (13) is costly. However, if κ is a positive integer, closed form solutions of (13) exist. More specifically, if $\kappa = k \in \mathbb{N}_+$, then

$$\phi_{\nu,k}(r) = A_{\nu+k}(r)P_k(r),$$

where P_k is a polynomial of order k and $A_{\nu+k}$ the Askey function ([3]) of order $\nu + k$,

$$A_{\nu+k}(r) = \begin{cases} (1-r)^{\nu+k}, & 0 \leq r < 1, \\ 0, & r \geq 1. \end{cases}$$

In addition, if $\kappa \in (\mathbb{N}_+ - 1/2)$, a positive half-integer, it is shown in [28] that further closed form solutions of (13), involving polynomial, logarithmic and square root terms, exist. Thus, in the specific example of generalized Wendland covariance functions, covariance approximations will facilitate computing (13) when $\kappa \notin \mathbb{N}_+ \cup (\mathbb{N}_+ - 1/2)$.

7. Covariance taper approximations: beyond compactly supported covariance functions

Asymptotic properties of (regular) tapered-ML estimators were addressed in both the infill- and increasing-domain asymptotic framework (see [23, 14, 30] and [16]). The direct functional approximation approach studied here can be combined with covariance tapering. Given observations of S, it is known that under weak assumptions on the presumed covariance function, ML estimators based on tapered covariance functions (tapered-ML estimators) preserve consistency (see [16], in particular Corollary 2 in the increasing-domain framework).

However, this is the case for covariance tapers that have a compact support which is not fixed, but rather grows to the entire $\mathbb{R}^d$ as the number of observations from S increases. Within an increasing-domain asymptotic framework, given a fixed compact support of the covariance taper, one can in general not expect tapered-ML estimators to be consistent. Still, under suitable conditions, tapered-ML estimators asymptotically minimize the Kullback-Leibler divergence (see for instance Theorem 3.3 in [5]). Given the theory developed here, we can readily recover the same result for truncated-tapered ML estimators, ML estimators based on tapered covariance function, where the covariance taper is replaced with a functional approximation of it. To be more formal, let us remain in the setting of Section 2, but assume that Z has true and unknown covariance function k_{θ_0}, $\theta_0 \in \Theta$, which belongs to a family $\{k_\theta \colon \theta \in \Theta\}$ which satisfies:

- For any $s \in \mathbb{R}^d$, $\theta \mapsto k_\theta(s)$ is continuously differentiable
- There exist constants $A < \infty$ and $\alpha > 0$ such that for all $i = 1, \ldots, p$, for all $s \in \mathbb{R}^d$ and for all $\theta \in \Theta$, $|k_\theta(s)| \leq A/(1 + \|s\|^{d+\alpha})$ and $\left|\frac{\partial k_\theta}{\partial \theta_i}(s)\right| \leq A/(1 + \|s\|^{d+\alpha})$
- $\{k_\theta \colon \theta \in \Theta\}$ satisfies (3) of Assumption 3.1.

The given assumptions are very weak and satisfied for instance for the Matérn family (see also Condition 2.1 in [4] or Remark 3.1). Then, we consider a fixed covariance taper $s \mapsto t_{\theta_0'}(s)$, $\theta_0' \in \Theta'$, $\Theta' \subset \mathbb{R}^l$, compact and convex. We assume that $t_{\theta_0'}$ belongs to a family of tapers $\{t_{\theta'} \colon \theta' \in \Theta'\}$ that satisfies Assumption 3.1 (regarding (2), $q = 1$ and the continuity of first order partial derivatives is sufficient). As we have seen in Section 6 (Proposition 6.1), we may choose, with $\theta_0' = (\beta_0, 1)$, $\kappa > 2$, $\nu \geq (d+1)/2 + \kappa$, a generalized Wendland taper (see also Remark 6.2). In the given context it is more convenient to write $t_{\beta_0} \coloneqq t_{\theta_0'}$, where β_0 is the taper range, that is $t_{\beta_0}(s) = 0$ for $\|s\| \geq \beta_0$. Based on a finite collection $S_{(n)}$ of S, on $(\Omega, \mathcal{F}, \mathbb{P})$, we then define the tapered $n \times n$ covariance matrix $R_{n,\theta_{i,j}} \coloneqq k_\theta(S_i - S_j) t_{\beta_0}(S_i - S_j)$, $1 \leq i, j \leq n$. Additionally, we consider a covariance matrix approximation

$$\widetilde{R}_{n,\theta_{i,j}} = k_\theta(S_i - S_j) \tilde{t}_{r(n),\theta_0'}(S_i - S_j), \quad 1 \leq i, j \leq n, \ r(n) \to \infty \text{ as } n \to \infty,$$

of $R_{n,\theta}$, where $(\tilde{t}_{m,\theta_0'})$ is a sequence of functions that belongs to a family of taper approximations $\{(\tilde{t}_{m,\theta'}) \colon \theta' \in \Theta'\}$, for which Assumption 3.2 applies. Again, we write $\tilde{t}_{m,\theta_0'} \coloneqq \tilde{t}_{m,\beta_0}$, $m \in \mathbb{N}_+$, to highlight the fixed range parameter. We note that the results of Lemma 4.1 and Proposition 5.1 remain true with $\Sigma_{n,\theta}$ and $\widetilde{\Sigma}_{n,\theta}$ replaced with $R_{n,\theta}$ and $\widetilde{R}_{n,\theta}$, respectively. We know (see Remark 2.1) that the conditional distribution of $Z_{(n)}$ given $S_{(n)}$ is given by the random variable $\omega \mapsto \mathcal{N}(0, K_{n,\theta_0}(\omega))$. On the other hand, we can assume a misspecified distribution $\omega \mapsto \mathcal{N}(0, R_{n,\theta}(\omega))$, where the true covariance matrix is replaced with the tapered covariance matrix $R_{n,\theta}(\omega)$, $\theta \in \Theta$. Then, we define the scaled (see [5]) conditional Kullback-Leibler divergence of $\mathcal{N}(0, R_{n,\theta})$ from $\mathcal{N}(0, K_{n,\theta_0})$,

$$d_{n,\theta} \coloneqq \frac{1}{n} \log \left(\det(R_{n,\theta} K_{n,\theta_0}^{-1}) \right) + \frac{1}{n} \operatorname{tr}(K_{n,\theta_0} R_{n,\theta}^{-1}) - 1.$$

The distribution $\mathcal{N}(0, R_{n,\theta})$ shall be called a regular taper miss-specified distribution. If we choose $n \geq N$ (N as in Proposition 5.1), we can even further misspecify the distribution of $Z_{(n)}$ given $S_{(n)}$ by replacing $R_{n,\theta}$ with $\widetilde{R}_{n,\theta}$ in $\mathcal{N}(0, R_{n,\theta})$. This gives rise to the scaled conditional Kullback-Leibler divergence of $\mathcal{N}(0, \widetilde{R}_{n,\theta})$ from $\mathcal{N}(0, K_{n,\theta_0})$,

$$\tilde{d}_{n,\theta} := \frac{1}{n} \log \big(\det(\widetilde{R}_{n,\theta} K_{n,\theta_0}^{-1}) \big) + \frac{1}{n} \operatorname{tr}(K_{n,\theta_0} \widetilde{R}_{n,\theta}^{-1}) - 1.$$

We use the notation $\big(\hat{\theta}_n(kt)\big)_{n \in \mathbb{N}_+}$ and $\big(\hat{\theta}_n(k\tilde{t})\big)_{n \in \mathbb{N}_+}$ for ML and truncated-ML estimators for θ_0 with respect to $\{k_\theta t_{\beta_0} \colon \theta \in \Theta\}$ and $\{(k_\theta \tilde{t}_{m,\beta_0}) \colon \theta \in \Theta\}$, respectively. In accordance with the literature about tapered-ML estimators, the estimators $\big(\hat{\theta}_n(kt)\big)_{n \in \mathbb{N}_+}$ and $\big(\hat{\theta}_n(k\tilde{t})\big)_{n \in \mathbb{N}_+}$ are then further referred to as tapered-ML estimators and truncated-tapered ML estimators, respectively. We can now state the following theorem:

Theorem 7.1. *We have that* $\mathbb{P}$ *a.s.*

$$\sup_{\theta \in \Theta} \big| d_{n,\theta} - \tilde{d}_{n,\theta} \big| \xrightarrow{n \to \infty} 0, \tag{15}$$

and as $n \to \infty$,

$$d_{n,\hat{\theta}_n(k\tilde{t})} = \inf_{\theta \in \Theta} d_{n,\theta} + \delta_n, \tag{16}$$

where $\delta_n \xrightarrow[n \to \infty]{\mathbb{P}} 0$.

Therefore, in the given scenario, truncated-tapered ML estimators asymptotically minimize the conditional Kullback-Leibler divergence of taper misspecified distributions from the true distribution (compare also to Theorem 3.3 in [5]). Thus, in terms of Kullback-Leibler divergence, truncated-tapered ML estimators and tapered-ML estimators perform asymptotically equally well.

8. Discussion and outlook

With the introduction of truncated-likelihood functions, we allow for more far-reaching forms of covariance approximations, such as linear interpolations or polynomial approximations. Our approximation approach relates directly to the presumed covariance function. Thus, combinations with existing approximation methods such as low-rank or covariance tapering approaches are well possible. We studied the quality of truncated-ML estimators from an asymptotic point of view. For compactly supported covariance functions, the conditions imposed in Sections 3 and 5 permit us to obtain truncated-ML estimators that are asymptotically well-behaving. That is, we obtain estimators that are consistent and asymptotically normal. Our proof strategies were strongly influenced by [4]. We have provided a comprehensive analysis for the family of generalized Wendland covariance functions. That is, we give precise conditions on smoothness, variance and range parameters, under which ML estimators for variance and range parameters are consistent and asymptotically normal. To our knowledge, this does

not exists in the literature so far (compare also to [9], within the infill-domain asymptotic context). Further, we gave four examples of generalized Wendland approximations, for which truncated-ML estimators preserve consistency and asymptotic normality.

We now discuss some open questions. Our results on consistency and asymptotic normality depend on the condition that correlations vanish beyond a certain distance. It would be of interest to recover the consistency and asymptotic normality results for truncated-ML estimators, where the assumption of a compact support is dropped. To this end, we recall that the imposed conditions on covariance functions and approximations resulted in the uniform asymptotic equivalence of covariance matrices and approximations. Using this, we established the existence of a positive integer N, after which covariance matrix approximations remain positive-definite. Expanding to non-compactly supported covariance function, this result remains unchanged, as long as covariance matrices and approximations are uniformly asymptotically equivalent (uniformly on the parameter and sample space). Thus, in this case, consistency and asymptotic normality can be recovered, even when presumed covariance functions are no longer compactly supported. However, as a mere condition, the asymptotic equivalence of covariance matrices and approximations is of little practical importance. Thus, the case of non-compactly supported covariance functions deserves further attention.

From a more applied point of view, our results provide a strong theoretical basis for further research. It remains to test and extend the given examples of covariance approximations. The four examples of generalized Wendland approximations and their effect on parameter estimations were discussed from a theoretical point of view. An important next step is to provide numerical implementations and practical comparisons.

In conclusion, for large datasets built upon correlated data, the present work provides an essential missing piece in the area of covariance approximations.

Appendix A: Covariance approximations for isotropic random fields

We consider families of approximations $\left\{(\tilde{\varphi}_{m,\theta})\colon \theta \in \Theta\right\}$ for $\{\varphi_\theta \colon \theta \in \Theta\}$ on $\mathbb{R}_+$ and translate (recycling the notation of Section 3) Assumptions 3.1 and 3.2 as follows:

Assumption A.1 (Regularity conditions on φ_θ).

(1) There exist real constants C, $L < \infty$, which are independent of $\theta \in \Theta$, such that $\varphi_\theta \in \mathcal{B}_\mathrm{C}(\mathbb{R}_+; S_\theta)$, with $S_\theta \subset [0, C]$ and $\|\varphi_\theta\|_\infty \leq L$.
(2) For any $t \in \mathbb{R}_+$, the first, second and third order partial derivatives of $\theta \mapsto \varphi_\theta(t)$ exist. In addition, for any $q = 1, 2, 3$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, $\frac{\partial^q \varphi_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}} \in \mathcal{B}_\mathrm{C}(\mathbb{R}_+; S_\theta(i_1, \ldots, i_q))$, where there exist constants C', $L' < \infty$, which are independent of $\theta \in \Theta$, such that $S_\theta(i_1, \ldots, i_q) \subset [0, C']$ and $\left\|\frac{\partial^q \varphi_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}\right\|_\infty \leq L'$.

(3) Fourier inversion holds, that is for any $\theta \in \Theta$,

$$\varphi_\theta(\|s\|) = \int_{\mathbb{R}^d} \hat{c}_\theta(f) \, \mathrm{e}^{\mathrm{i}\langle f, s \rangle} \, \mathrm{d}f,$$

where $\Theta \times \mathbb{R}^d \ni (\theta, f) \mapsto \hat{c}_\theta(f)$ is continuous and strictly positive.

Assumption A.2 (Regularity conditions on $\tilde{\varphi}_{m,\theta}$).

(1) For any $\theta \in \Theta$, for any $m \in \mathbb{N}_+$, the function $\tilde{\varphi}_{m,\theta} \colon (\mathbb{R}_+, \mathfrak{B}(\mathbb{R}_+)) \to (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ is measurable.
(2) For any $m \in \mathbb{N}_+$, $\tilde{\varphi}_{m,\theta}$ satisfies (1) of Assumption A.1, where respective constants $\widetilde{C}$ and $\widetilde{L}$ can be further chosen independently of $m \in \mathbb{N}_+$.
(3) $\sup_{\theta \in \Theta} \|\tilde{\varphi}_{m,\theta} - \varphi_\theta\|_\infty \xrightarrow{m \to \infty} 0$.
(4) For any $m \in \mathbb{N}_+$, $\tilde{\varphi}_{m,\theta}$ satisfies (2) of Assumption A.1, where respective constants $\widetilde{C}'$ and $\widetilde{L}'$ can be further chosen independently of $m \in \mathbb{N}_+$.
(5) For any $q = 1, 2, 3$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, we have that

$$\sup_{\theta \in \Theta} \left\| \frac{\partial^q \tilde{\varphi}_{m,\theta}}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}} - \frac{\partial^q \varphi_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}} \right\|_\infty \xrightarrow{m \to \infty} 0.$$

Note that the family $\{\varphi_\theta \colon \theta \in \Theta\}$ satisfies Assumption A.1 if and only if $\{c_\theta \colon \theta \in \Theta\}$ satisfies Assumption 3.1. Further, for any $n \in \mathbb{N}_+$ and $\theta \in \Theta$, we have that

$$\Sigma_{n,\theta} = \left[\varphi_\theta(\|S_i - S_j\|)\right]_{1 \le i, j \le n},$$

on $(\Omega, \mathcal{F}, \mathbb{P})$. Thus, a sequence of truncated-ML estimators for θ_0 based on $\{c_\theta \colon \theta \in \Theta\}$ is a sequence of truncated-ML estimators for θ_0 based on $\{\varphi_\theta \colon \theta \in \Theta\}$. If we define a sequence of truncated-ML estimators $(\hat{\theta}_n(\tilde{\varphi}))_{n \in \mathbb{N}_+}$ for θ_0 based on a given $\{(\tilde{\varphi}_{m,\theta}) \colon \theta \in \Theta\}$ upon replacing $\widetilde{\Sigma}_{n,\theta}$ in (7) with the random $n \times n$ matrix $\left[\tilde{\varphi}_{r(n),\theta}(\|S_i - S_j\|)\right]_{1 \le i, j \le n}$, we can recover the results of Sections 4 and 5:

Theorem A.1. *Let* $(\hat{\theta}_n(\tilde{\varphi}))_{n \in \mathbb{N}_+}$ *be a sequence of truncated-ML estimators for* θ_0 *based on* $\{(\tilde{\varphi}_{m,\theta}) \colon \theta \in \Theta\}$. *Assume that* $\{\varphi_\theta \colon \theta \in \Theta\}$ *satisfies Assumption A.1 (regarding (2), $q = 1$ and the continuity of first order partial derivatives is sufficient) and* $\{c_\theta \colon \theta \in \Theta\}$ *satisfies Assumption 5.1. Suppose further that* $\{(\tilde{\varphi}_{m,\theta}) \colon \theta \in \Theta\}$ *satisfies Assumption A.2 (regarding (4) and (5), $q = 1$ and the continuity of first order partial derivatives is sufficient). Then,*

$$\hat{\theta}_n(\tilde{\varphi}) \xrightarrow[n \to \infty]{\mathbb{P}} \theta_0.$$

Theorem A.2. *Let* $(\hat{\theta}_n(\tilde{\varphi}))_{n \in \mathbb{N}_+}$ *be an sequence of truncated-ML estimators for* θ_0 *based on* $\{(\tilde{\varphi}_{m,\theta}) \colon \theta \in \Theta\}$. *Suppose that* $\{\varphi_\theta \colon \theta \in \Theta\}$ *satisfies Assumption A.1 and* $\{c_\theta \colon \theta \in \Theta\}$ *satisfies Assumptions 5.1 and 5.2. Assume further that* $\{(\tilde{\varphi}_{m,\theta}) \colon \theta \in \Theta\}$ *satisfies Assumption A.2. Then, we have that*

$$n^{1/2}\left(\hat{\theta}_n(\tilde{\varphi}) - \theta_0\right) \xrightarrow[n \to \infty]{\mathrm{d}} \mathcal{N}(0, \Lambda^{-1}),$$

with Λ as in Theorem 5.4.

Appendix B: Supporting results

Let $r\colon \mathbb{N}_+ \to \mathbb{N}_+$ be such that $r(n) \to \infty$ as $n \to \infty$. For the families $\{c_\theta \colon \theta \in \Theta\}$ and $\{(\tilde{c}_{m,\theta})\colon \theta \in \Theta\}$, we introduce, for any $n \in \mathbb{N}_+$ and $\theta \in \Theta$, for an arbitrary $s_{(n)} \in \mathcal{G}_n$, for any $q = 1, 2, 3$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, the non-random $n \times n$ matrices

$$\frac{\partial^q \Sigma_\theta(s_{(n)})}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}} := \left[\frac{\partial^q c_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}} (s_i - s_j) \right]_{1 \leq i, j \leq n},$$

and

$$\frac{\partial^q \widetilde{\Sigma}_\theta(s_{(n)})}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}} := \left[\frac{\partial^q \tilde{c}_{r(n),\theta}}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}} (s_i - s_j) \right]_{1 \leq i, j \leq n},$$

whenever the above partial derivatives with respect to θ exist. Further, for Borel measurable sequences of functions $\{(\tilde{c}_{m,\theta})\colon \theta \in \Theta\}$, we introduce, on $(\Omega, \mathcal{F}, \mathbb{P})$, the $n \times n$ random matrices

$$\omega \mapsto \frac{\partial^q \Sigma_{n,\theta}}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}(\omega) := \frac{\partial^q \Sigma_\theta(S_{(n)}(\omega))}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}},$$

and

$$\omega \mapsto \frac{\partial^q \widetilde{\Sigma}_{n,\theta}}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}(\omega) := \frac{\partial^q \widetilde{\Sigma}_\theta(S_{(n)}(\omega))}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}},$$

whenever the above partial derivatives with respect to θ exist.

Lemma B.1. *Let C, $L < \infty$ be some real constants. Consider $g\colon \mathbb{R}^d \to \mathbb{R}_+$ such that $g \in \mathcal{B}_C(\mathbb{R}^d; S)$, with $S \subset \mathrm{B}[0; C]$ and $\|g\|_\infty \leq L$. Then, for any $i \in \mathbb{N}_+$, for any sequence $(s_j)_{j \in \mathbb{N}_+} \in \mathcal{G}$,*

$$\sum_{j \in \mathbb{N}_+} g(s_i - s_j) \leq LR(d, C, \tau), \tag{17}$$

where $R(d, C, \tau) := (2^{2d} d C^{d-1})/\Delta_\tau^d$, with $\Delta_\tau = 1 - 2\tau$. Further, we also have that

$$\sum_{\substack{j \in \mathbb{N}_+ \\ |v_i - v_j|_\infty \geq C+1}} g(s_i - s_j) = 0. \tag{18}$$

Remark B.1. We would like to point out that Lemma B.1 resembles Lemmas D.1 and D.3 of [4], where $f\colon \mathbb{R}^d \to \mathbb{R}_+$, which is such that $f(s) \leq 1/(1 + |s|_\infty^{d+1})$, is replaced with a compactly supported function g, defined as in Lemma B.1.

Lemma B.2. *Let $\widetilde{C}$, $\widetilde{L} < \infty$ be some real constants. Consider a sequence of functions $(g_m)_{m \in \mathbb{N}_+}$, with values in $\mathbb{R}_+$, where for any $m \in \mathbb{N}_+$, $g_m \in$*

$\mathcal{B}_C(\mathbb{R}^d; \widetilde{S}_m)$, *with* $\widetilde{S}_m \subset \mathrm{B}\big[0; \widetilde{C}\big]$ *and* $\|g_m\|_\infty \leq \widetilde{L}$. *Then, for any* $i \in \mathbb{N}_+$, *for any sequence* $(s_j)_{j\in\mathbb{N}_+} \in \mathcal{G}$,

$$\sup_{m\in\mathbb{N}_+} \sum_{j\in\mathbb{N}_+} g_m(s_i - s_j) \leq \widetilde{L}R(d, \widetilde{C}, \tau), \tag{19}$$

where $R(d, \widetilde{C}, \tau) := (2^{2d} d \widetilde{C}^{d-1})/\Delta_\tau^d$, *with* $\Delta_\tau = 1 - 2\tau$. *Further we also have that*

$$\sup_{m\in\mathbb{N}_+} \sum_{\substack{j\in\mathbb{N}_+ \\ |v_i - v_j|_\infty \geq \widetilde{C}+1}} g_m(s_i - s_j) = 0. \tag{20}$$

Lemma B.3. *Assume that* $\{c_\theta \colon \theta \in \Theta\}$ *satisfies* (1) *and* (3) *of Assumption 3.1. Consider* $\big\{(\tilde{c}_{m,\theta}) \colon \theta \in \Theta\big\}$ *that satisfies* (1), (2) *and* (3) *of Assumption 3.2. Then, we have that*

$$\sup_{n\in\mathbb{N}_+} \sup_{s_{(n)}\in\mathcal{G}_n} \sup_{\theta\in\Theta} \big\|\Sigma_\theta(s_{(n)})\big\|_2 < \infty, \quad \sup_{n\in\mathbb{N}_+} \sup_{s_{(n)}\in\mathcal{G}_n} \sup_{\theta\in\Theta} \big\|\widetilde{\Sigma}_\theta(s_{(n)})\big\|_2 < \infty, \tag{21}$$

and in particular

$$\sup_{s_{(n)}\in\mathcal{G}_n} \sup_{\theta\in\Theta} \big\|\Sigma_\theta(s_{(n)}) - \widetilde{\Sigma}_\theta(s_{(n)})\big\|_2 \xrightarrow{n\to\infty} 0. \tag{22}$$

Further, we have that

$$\inf_{n\in\mathbb{N}_+} \inf_{s_{(n)}\in\mathcal{G}_n} \inf_{\theta\in\Theta} \lambda_n\big(\Sigma_\theta(s_{(n)})\big) > 0, \tag{23}$$

and there exists $N \in \mathbb{N}_+$ *such that*

$$\inf_{n\geq N} \inf_{s_{(n)}\in\mathcal{G}_n} \inf_{\theta\in\Theta} \lambda_n\big(\widetilde{\Sigma}_\theta(s_{(n)})\big) > 0. \tag{24}$$

Corollary B.4. *Let* $\{c_\theta \colon \theta \in \Theta\}$, $\big\{(\tilde{c}_{m,\theta}) \colon \theta \in \Theta\big\}$ *and* N *be as in Lemma B.3. Then, we have that*

$$\sup_{n\in\mathbb{N}_+} \sup_{s_{(n)}\in\mathcal{G}_n} \sup_{\theta\in\Theta} \big\|\Sigma_\theta(s_{(n)})^{-1}\big\|_2 < \infty, \quad \sup_{n\geq N} \sup_{s_{(n)}\in\mathcal{G}_n} \sup_{\theta\in\Theta} \big\|\widetilde{\Sigma}_\theta(s_{(n)})^{-1}\big\|_2 < \infty.$$

In addition we can conclude that

$$\sup_{s_{(n)}\in\mathcal{G}_n} \sup_{\theta\in\Theta} \big\|\Sigma_\theta(s_{(n)})^+ - \widetilde{\Sigma}_\theta(s_{(n)})^+\big\|_2 \xrightarrow{n\to\infty} 0.$$

In particular we have that $\mathbb{P}$ *a.s.*

$$\sup_{n\in\mathbb{N}_+} \sup_{\theta\in\Theta} \big\|\Sigma_{n,\theta}^{-1}\big\|_2, \quad \sup_{n\geq N} \sup_{\theta\in\Theta} \big\|\widetilde{\Sigma}_{n,\theta}^{-1}\big\|_2 < \infty \ \text{ and } \ \sup_{\theta\in\Theta} \big\|\Sigma_{n,\theta}^+ - \widetilde{\Sigma}_{n,\theta}^+\big\|_2 \xrightarrow{n\to\infty} 0.$$

Lemma B.5. *Suppose that $\{c_\theta \colon \theta \in \Theta\}$ satisfies (2) of Assumption 3.1. Consider $\{(\tilde{c}_{m,\theta}) \colon \theta \in \Theta\}$ that satisfies (1), (4) and (5) of Assumption 3.2. Then, for any $q = 1, 2, 3$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, we have that (21) and (22) of Lemma B.3 are satisfied with $\Sigma_\theta(s_{(n)})$ and $\widetilde{\Sigma}_\theta(s_{(n)})$ replaced with the respective partial derivatives $\frac{\partial^q \Sigma_\theta(s_{(n)})}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}$ and $\frac{\partial^q \widetilde{\Sigma}_\theta(s_{(n)})}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}$. In particular, for any $q = 1, 2, 3$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, we have that $\mathbb{P}$ a.s.*

$$\sup_{n \in \mathbb{N}_+} \sup_{\theta \in \Theta} \left\| \frac{\partial^q \Sigma_{n,\theta}}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}} \right\|_2 < \infty, \quad \sup_{n \in \mathbb{N}_+} \sup_{\theta \in \Theta} \left\| \frac{\partial^q \widetilde{\Sigma}_{n,\theta}}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}} \right\|_2 < \infty,$$

and in addition it is true that for any $q = 1, 2, 3$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, $\mathbb{P}$ a.s.

$$\sup_{\theta \in \Theta} \left\| \frac{\partial^q \Sigma_{n,\theta}}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}} - \frac{\partial^q \widetilde{\Sigma}_{n,\theta}}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}} \right\|_2 \xrightarrow{n \to \infty} 0.$$

Lemma B.6. *Let $I \in \mathbb{N}_+$ be fixed. On $(\Omega, \mathcal{F}, \mathbb{P})$, for $k = 1, \ldots, I$, we consider a sequence of $n \times n$ random symmetric matrices $(\widetilde{A}_{k,n,\theta})_{n \in \mathbb{N}_+}$, $\theta \in \Theta$, such that $\mathbb{P}$ a.s., for any $k = 1, \ldots, I$, $\sup_{n \in \mathbb{N}_+} \sup_{\theta \in \Theta} \|\widetilde{A}_{k,n,\theta}\| < \infty$. Further we assume that there exists $N \in \mathbb{N}_+$ such that $\mathbb{P}$ a.s., for $k = 1, \ldots, I$, $\inf_{n \geq N} \inf_{\theta \in \Theta} \lambda_n(\widetilde{A}_{k,n,\theta}) > 0$. Let $(A_{k,n,\theta})_{n \in \mathbb{N}_+}$, $\theta \in \Theta$, $k = 1, \ldots, I$, be another sequence of $n \times n$ random symmetric matrices, defined on the same probability space, which is such that $\mathbb{P}$ a.s., for $k = 1, \ldots, I$,*

$$\sup_{n \in \mathbb{N}_+} \sup_{\theta \in \Theta} \|A_{k,n,\theta}\|_2 < \infty \quad and \quad \inf_{n \geq N} \inf_{\theta \in \Theta} \lambda_n(A_{k,n,\theta}) > 0.$$

Finally we also assume that $\mathbb{P}$ a.s., for any $k = 1, \ldots, I$,

$$\sup_{\theta \in \Theta} \|\widetilde{A}_{k,n,\theta} - A_{k,n,\theta}\|_2 \xrightarrow{n \to \infty} 0.$$

Then, we have that $\mathbb{P}$ a.s.

$$\sup_{\theta \in \Theta} \left| \frac{1}{n} \log \left(\det_+ \left(\prod_{k=1}^{I} A_{k,n,\theta} \right) \right) - \frac{1}{n} \log \left(\det_+ \left(\prod_{k=1}^{I} \widetilde{A}_{k,n,\theta} \right) \right) \right| \xrightarrow{n \to \infty} 0.$$

Lemma B.7. *On $(\Omega, \mathcal{F}, \mathbb{P})$, consider two sequences of $n \times n$ random matrices $(A_{n,\theta})_{n \in \mathbb{N}_+}$ and $(\widetilde{A}_{n,\theta})_{n \in \mathbb{N}_+}$, $\theta \in \Theta$, such that $\mathbb{P}$ a.s.*

$$\sup_{\theta \in \Theta} \|A_{n,\theta} - \widetilde{A}_{n,\theta}\|_2 \xrightarrow{n \to \infty} 0.$$

Then, we have that

$$\sup_{\theta \in \Theta} \frac{1}{n} \left| \langle Z_{(n)}, A_{n,\theta} Z_{(n)} \rangle - \langle Z_{(n)}, \widetilde{A}_{n,\theta} Z_{(n)} \rangle \right| \xrightarrow[n \to \infty]{\mathbb{P}} 0. \tag{25}$$

Lemma B.8. *Suppose that $\{c_\theta \colon \theta \in \Theta\}$ satisfies Assumption 3.1 and 5.2 (regularity conditions for partial derivatives up to order $q = 2$ are sufficient). Suppose further that $\{(\tilde{c}_{m,\theta}) \colon \theta \in \Theta\}$ satisfies Assumption 3.2 (regularity conditions for partial derivatives up to order $q = 2$ are sufficient). Let N be as in Proposition 5.1 and define $\{(G_{n,N}(\theta))_{n \in \mathbb{N}_+} \colon \theta \in \Theta\}$ and $\{(\widetilde{G}_{n,N}(\theta))_{n \in \mathbb{N}_+} \colon \theta \in \Theta\}$ as in (10) and (11), respectively. We then have that*

$$\left\| J_{\widetilde{G}_{n,N}}(\theta_0) - J_{G_{n,N}}(\theta_0) \right\|_2 \xrightarrow[n \to \infty]{\mathbb{P}} 0. \tag{26}$$

Further, we conclude that the random $p \times p$ matrix $J_{\widetilde{G}_{n,N}}(\theta_0)$ converges in probability $\mathbb{P}$ to a non-random matrix 2Λ, where $\mathrm{S}_{p \times p} \ni \Lambda \succ 0$.

Appendix C: Proofs

C.1. Proof of results in Appendix B

Proof of Lemma B.1. Let $(s_j)_{j \in \mathbb{N}_+} \in \mathcal{G}$. For $j \in \mathbb{N}_+$ such that $|v_i - v_j|_\infty \geq C+1$ we have that $|s_i - s_j|_\infty \geq C$ and thus $\|s_i - s_j\| \geq C$ as well (since $|w|_\infty \leq \|w\|$ for any $w \in \mathbb{R}^d$). Therefore, (18) follows since we have assumed that g has compact support $S \subset \mathrm{B}[0; C]$. The proof of (17) depends on the fact that there exists a minimal spacing $\Delta_\tau > 0$ between any two distinct observation points (see (4)). This allows us to show that for some arbitrary $i \in \mathbb{N}_+$, if $N_{s_i,C}$ denotes the cardinality of the set $\{j \in \mathbb{N}_+ \colon \|s_j - s_i\| \leq C\} \subset \{j \in \mathbb{N}_+ \colon |s_j - s_i|_\infty \leq C\}$, we have that $N_{s_i,C} \leq R(d, C, \tau)$. For a complete argument one could for example consider the proof of Lemma 4 in [16]. Using this we can estimate,

$$\begin{aligned}
\sum_{j \in \mathbb{N}_+} g(s_i - s_j) &= \sum_{j \in \mathbb{N}_+} g(s_i - s_j) \, \mathbb{1}_{[0,C]}(\|s_i - s_j\|) \\
&\leq L \sum_{j \in \mathbb{N}_+} \mathbb{1}_{[0,C+1]}(\|s_i - s_j\|) \\
&\leq L R(d, C, \tau),
\end{aligned}$$

and thus also (17) is proven. $\qquad \square$

Proof of Lemma B.2. The proof is similar to the proof of Lemma B.1 and hence we consider the lemma as proven. $\qquad \square$

Proof of Lemma B.3. Let C, L and $\widetilde{C}$, $\widetilde{L}$ be defined as in (1) of Assumption 3.1 and (2) of Assumption 3.2, respectively. We use Lemma B.1 to show that there exists a real constant $M > 0$, which does not depend on $n \in \mathbb{N}_+$, $s_{(n)} \in \mathcal{G}_n$ and $\theta \in \Theta$ such that for any $n \in \mathbb{N}_+$, $s_{(n)} \in \mathcal{G}_n$,

$$\max \left\{ \left\| \Sigma_\theta(s_{(n)}) \right\|_2, \left\| \widetilde{\Sigma}_\theta(s_{(n)}) \right\|_2 \right\} \leq M. \tag{27}$$

To see this, let $C_* := \max\{C, \widetilde{C}\}$ and $L_* := \max\{L, \widetilde{L}\}$. Using (1) of Assumption 3.1, we have that for any $\theta \in \Theta$, $c_\theta \in \mathcal{B}_\mathrm{C}(\mathbb{R}^d; S_\theta)$, where now $S_\theta \subset \mathrm{B}[0; C_*]$

and $\|c_\theta\|_\infty \leq L_*$, with C_* and L_* finite constants that are independent of $n \in \mathbb{N}_+$ and $\theta \in \Theta$. Thus we can write, for any $n \in \mathbb{N}_+$, $s_{(n)} \in \mathcal{G}_n$ and $\theta \in \Theta$, by Gershgorin circle theorem,

$$\begin{aligned}
\left\|\Sigma_\theta(s_{(n)})\right\|_2 &\leq \max_{i=1,\ldots,n} \sum_{j=1}^n |c_\theta(s_i - s_j)| \\
&\leq \sup_{i \in \mathbb{N}_+} \sum_{j \in \mathbb{N}_+} |c_\theta(s_i - s_j)| \leq L_* R(d, C_*, \tau) =: M,
\end{aligned}$$

under application of Lemma B.1, with $R(d, C_*, \tau) = (2^{2d} d C_*^{d-1})/\Delta_\tau^d$, where $\Delta_\tau = 1 - 2\tau$. Note that M is independent of $n \in \mathbb{N}_+$, $s_{(n)} \in \mathcal{G}_n$ and $\theta \in \Theta$. Similarly, by (2) of Assumption 3.2 we then use Lemma B.2, together with Gershgorin circle theorem, to show that for any $n \in \mathbb{N}_+$, $s_{(n)} \in \mathcal{G}_n$ and $\theta \in \Theta$, $\|\widetilde{\Sigma}_\theta(s_{(n)})\|_2 \leq M$ as well. This shows (27). Thus, we have established that

$$\sup_{n \in \mathbb{N}_+} \sup_{s_{(n)} \in \mathcal{G}_n} \sup_{\theta \in \Theta} \left\|\widetilde{\Sigma}_\theta(s_{(n)})\right\|_2 \leq M,$$

and

$$\sup_{n \in \mathbb{N}_+} \sup_{s_{(n)} \in \mathcal{G}_n} \sup_{\theta \in \Theta} \left\|\Sigma_\theta(s_{(n)})\right\|_2 \leq M,$$

and therefore (21) of Lemma B.3 is verified. It is shown in [4] (Proposition D.4) that because of the increasing-domain setting, where there exists a minimal distance between any two observation points (see (4)), and since (3) of Assumption 3.1 is satisfied,

$$\inf_{n \in \mathbb{N}_+} \inf_{x_{(n)} \in \mathcal{Q}^n} \inf_{\theta \in \Theta} \lambda_n\left(\Sigma_\theta(s_{(n)})\right) > 0.$$

This shows (23) of Lemma B.3. Using this result, we can fix some $\delta > 0$ (small enough, independent of $n \in \mathbb{N}_+$, $s_{(n)} \in \mathcal{G}_n$ and $\theta \in \Theta$), such that for any $s_{(n)} \in \mathcal{G}_n$,

$$0 < \varepsilon := \frac{\delta}{R(d, C_*, \tau)} < \min_{\|a\|=1} \langle a, \Sigma_\theta(s_{(n)})a \rangle.$$

For the above $\delta > 0$, we can then find $N \in \mathbb{N}_+$ such that,

$$\sup_{n \geq N} \sup_{s_{(n)} \in \mathcal{G}_n} \sup_{\theta \in \Theta} \left\|\Sigma_\theta(s_{(n)}) - \widetilde{\Sigma}_\theta(s_{(n)})\right\|_2 \leq \delta. \tag{28}$$

This is valid since for the given $\varepsilon > 0$, by the uniform convergence of $(\tilde{c}_{r(n),\theta})$ to c_θ (see (3) of Assumption 3.2), we find $N \in \mathbb{N}_+$ such that for any $n \geq N$, for any $s_{(n)} \in \mathcal{G}_n$ and $1 \leq i, j \leq n$,

$$\sup_{\theta \in \Theta} \left| \left[\Sigma_\theta(s_{(n)})\right]_{i,j} - \left[\widetilde{\Sigma}_\theta(s_{(n)})\right]_{i,j} \right| < \varepsilon.$$

Then, if we define

$$\mathbb{R}^d \ni s \mapsto g_{r(n)}(s) := \sup_{\theta \in \Theta} \left| \left(c_\theta - \tilde{c}_{r(n),\theta} \right)(s) \right|, \quad n \geq N,$$

since we have assumed that the families $\{c_\theta \colon \theta \in \Theta\}$ and $\{(\tilde{c}_{m,\theta}) \colon \theta \in \Theta\}$ have compact supports, which belong to $\mathrm{B}\left[0; C_*\right]$, we have that $g_{r(n)}(s) = 0$ for $\|s\| \geq C_*$. Thus, by Gershgorin circle theorem, under application of Lemma B.2, for $n \geq N$ and $s_{(n)} \in \mathcal{G}_n$,

$$\left\| \widetilde{\Sigma}_\theta(s_{(n)}) - \Sigma_\theta(s_{(n)}) \right\|_2 \leq \max_{i=1,\dots,n} \sum_{j=1}^n g_{r(n)} \left(s_i - s_j \right)$$

$$\leq \sup_{i \in \mathbb{N}_+} \sum_{j \in \mathbb{N}_+} g_{r(n)} \left(s_i - s_j \right) \leq \varepsilon R \left(d, C_*, \tau \right).$$

Since $\varepsilon R \left(d, C_*, \tau \right)$ is independent of $n \in \mathbb{N}_+$, $s_{(n)} \in \mathcal{G}_n$ and $\theta \in \Theta$, we can conclude that (28) must be satisfied. Using (28), we have, for $n \geq N$, $s_{(n)} \in \mathcal{G}_n$ and $\theta \in \Theta$, and for vectors a such that $\|a\| = 1$, that

$$\left| \langle a, \Sigma_\theta(s_{(n)})a \rangle - \langle a, \widetilde{\Sigma}_\theta(s_{(n)})a \rangle \right| = \left| \langle a, \left(\Sigma_\theta(s_{(n)}) - \widetilde{\Sigma}_\theta(s_{(n)}) \right)a \rangle \right|$$

$$\leq \left\| \Sigma_\theta(s_{(n)}) - \widetilde{\Sigma}_\theta(s_{(n)}) \right\|_2 \leq \delta,$$

under application of the Cauchy–Schwarz inequality. In conclusion we have for vectors a such that $\|a\| = 1$, for $n \geq N$, $s_{(n)} \in \mathcal{G}_n$ and $\theta \in \Theta$,

$$\min_{\|a\|=1} \langle a, \Sigma_\theta(s_{(n)})a \rangle - \delta \leq \min_{\|a\|=1} \langle a, \widetilde{\Sigma}_\theta(s_{(n)})a \rangle.$$

But we know that $\inf_{n \geq N} \inf_{s_{(n)} \in \mathcal{G}_n} \inf_{\theta \in \Theta} \min_{\|a\|=1} \langle a, \Sigma_\theta(s_{(n)})a \rangle > 0$ and $\delta > 0$ was chosen small enough (but otherwise arbitrary). Thus, we have also proven (24) of Lemma B.3. Notice that (22) is proven with (28), hence the proof of Lemma B.3 is complete. □

Proof of Corollary B.4. This follows from Lemma B.3. □

Proof of Lemma B.5. We omit a formal argument and argue that one can proof Lemma B.5 using the same way of reasoning as in the proof of Lemma B.3. □

Proof of Lemma B.6. For $n \geq N$ (N as in the statement) and $\theta \in \Theta$, we can write $\mathbb{P}$ a.s.

$$\mathrm{det}_+ \left(\prod_{k=1}^I \widetilde{A}_{k,n,\theta} \right) = \det \big(\underbrace{\widetilde{A}_{I,n,\theta}^{1/2} \cdots \widetilde{A}_{2,n,\theta}^{1/2} \widetilde{A}_{1,n,\theta} \widetilde{A}_{2,n,\theta}^{1/2} \cdots \widetilde{A}_{I,n,\theta}^{1/2}}_{=: \widetilde{B}_{n,\theta}} \big),$$

and

$$\mathrm{det}_+ \left(\prod_{k=1}^I A_{k,n,\theta} \right) = \det \big(\underbrace{A_{I,n,\theta}^{1/2} \cdots A_{2,n,\theta}^{1/2} A_{1,n,\theta} A_{2,n,\theta}^{1/2} \cdots A_{I,n,\theta}^{1/2}}_{=: B_{n,\theta}} \big).$$

Note that $\widetilde{B}_{n,\theta}$ and $B_{n,\theta}$ are random symmetric matrices. Further, for each of the random symmetric matrices

$$\widetilde{A}_{I,n,\theta}^{1/2}, \ldots, \widetilde{A}_{2,n,\theta}^{1/2}, \ \widetilde{A}_{1,n,\theta}, \ A_{I,n,\theta}^{1/2}, \ldots, A_{2,n,\theta}^{1/2}, \ A_{1,n,\theta},$$

the smallest eigenvalue is strictly greater that zero, $\mathbb{P}$ a.s., uniformly in $n \geq N$ and $\theta \in \Theta$ and hence we have that

$$\inf_{n \geq N} \inf_{\theta \in \Theta} \lambda_n \left(B_{n,\theta} \right) > 0 \text{ and } \inf_{n \geq N} \inf_{\theta \in \Theta} \lambda_n \left(\widetilde{B}_{n,\theta} \right) > 0, \ \ \mathbb{P} \text{ a.s.} \tag{29}$$

In addition, since $\mathbb{P}$ a.s. for $k = 1, \ldots, I$, by assumption

$$\sup_{n \geq N} \sup_{\theta \in \Theta} \left\| \widetilde{A}_{k,n,\theta} \right\|_2, \ \sup_{n \geq N} \sup_{\theta \in \Theta} \left\| \widetilde{A}_{k,n,\theta} \right\|_2 < \infty,$$

and

$$\sup_{\theta \in \Theta} \left\| \widetilde{A}_{k,n,\theta} - A_{k,n,\theta} \right\|_2 \xrightarrow{n \to \infty} 0,$$

we also have that

$$\sup_{n \geq N} \sup_{\theta \in \Theta} \left\| \widetilde{B}_{n,\theta} \right\|_2, \ \sup_{n \geq N} \sup_{\theta \in \Theta} \| B_{n,\theta} \|_2 < \infty, \tag{30}$$

and

$$\sup_{\theta \in \Theta} \left\| B_{n,\theta} - \widetilde{B}_{n,\theta} \right\|_2 \xrightarrow{n \to \infty} 0. \tag{31}$$

Using (29), (30) and (31), we pick $\delta > 0$ arbitrary and define

$$0 < \varepsilon := \frac{\delta}{\sup_{n \geq N} \sup_{\theta \in \Theta} \| B_{n,\theta} \|_2}$$

such that for some given integer $N^* \geq N$, $\mathbb{P}$ a.s.,

$$\sup_{n \geq N^*} \sup_{\theta \in \Theta} \left\| \widetilde{B}_{n,\theta}^{-1} - B_{n,\theta}^{-1} \right\|_2 < \varepsilon.$$

Now write

$$\frac{1}{n} \log \left(\frac{\det(\prod_{k=1}^I A_{k,n,\theta})}{\det(\prod_{k=1}^I \widetilde{A}_{k,n,\theta})} \right) = \frac{1}{n} \log \left(\det \left(B_{n,\theta} \widetilde{B}_{n,\theta}^{-1} \right) \right)$$

$$= \frac{1}{n} \operatorname{tr} \left(\log \left(B_{n,\theta} \widetilde{B}_{n,\theta}^{-1} \right) \right)$$

$$= \frac{1}{n} \sum_{i=1}^n \log \left(\lambda_i \left(B_{n,\theta} \widetilde{B}_{n,\theta}^{-1} \right) \right). \tag{32}$$

We can then estimate (32) from above and below as

$$\log \left(\lambda_n \left(B_{n,\theta} \widetilde{B}_{n,\theta}^{-1} \right) \right) \leq \frac{1}{n} \sum_{i=1}^n \log \left(\lambda_i \left(B_{n,\theta} \widetilde{B}_{n,\theta}^{-1} \right) \right) \leq \log \left(\lambda_1 \left(B_{n,\theta} \widetilde{B}_{n,\theta}^{-1} \right) \right).$$

But for the given $\varepsilon > 0$, for $n \geq N^*$, we have that $\mathbb{P}$ a.s.

$$
\begin{aligned}
\lambda_1 \left(B_{n,\theta} \widetilde{B}_{n,\theta}^{-1} \right) &\leq \lambda_1 \left(B_{n,\theta} \right) \lambda_1 \left(\widetilde{B}_{n,\theta}^{-1} \right) \\
&= \left\| B_{n,\theta} \right\|_2 \left\| \widetilde{B}_{n,\theta}^{-1} \right\|_2 \\
&\leq \left\| B_{n,\theta} \right\|_2 \left\| \widetilde{B}_{n,\theta}^{-1} - B_{n,\theta}^{-1} \right\|_2 + \left\| B_{n,\theta} \right\|_2 \left\| B_{n,\theta}^{-1} \right\|_2 \\
&\leq 1 + \delta.
\end{aligned}
$$

On the other hand, by (31), we also have that for $n \geq N^*$, $\mathbb{P}$ a.s.

$$
\begin{aligned}
\lambda_n \left(B_{n,\theta} \widetilde{B}_{n,\theta}^{-1} \right) &\geq \lambda_n \left(B_{n,\theta} \right) \lambda_n \left(\widetilde{B}_{n,\theta}^{-1} \right) \\
&= \left(\min_{\{a\,:\,\|a\|=1\}} \langle a, B_{n,\theta} a \rangle \right) \left(\min_{\{a\,:\,\|a\|=1\}} \left\langle a, \widetilde{B}_{n,\theta}^{-1} a \right\rangle \right) \\
&\geq \left(\min_{\{a\,:\,\|a\|=1\}} \langle a, B_{n,\theta} a \rangle \right) \left(\min_{\{a\,:\,\|a\|=1\}} \left\langle a, B_{n,\theta}^{-1} a \right\rangle - \varepsilon \right) \\
&\geq 1 - \delta.
\end{aligned}
$$

Since $\delta > 0$ was arbitrary and independent of $\theta \in \Theta$, the lemma is proven. $\square$

Proof of Lemma B.7. First, using the Cauchy–Schwarz inequality and the compatibility of the spectral norm with the Euclidean norm, we can estimate $\mathbb{P}$ a.s.

$$
\frac{1}{n} \left| \langle Z_{(n)}, A_{n,\theta} Z_{(n)} \rangle - \langle Z_{(n)}, \widetilde{A}_{n,\theta} Z_{(n)} \rangle \right| \leq \sup_{\theta \in \Theta} \left\| A_{n,\theta} - \widetilde{A}_{n,\theta} \right\|_2 \frac{\left\| Z_{(n)} \right\|^2}{n}
$$

Let us fix some arbitrary $\varepsilon > 0$ such that for n large enough we have that $\mathbb{P}$ a.s.,

$$
\sup_{\theta \in \Theta} \left\| A_{n,\theta} - \widetilde{A}_{n,\theta} \right\|_2 < \varepsilon.
$$

Then, let $\delta > 0$ be arbitrary and notice that

$$
\mathbb{P} \left(\varepsilon n^{-1} \left\| Z_{(n)} \right\|^2 > \delta \,\Big|\, S_{(n)} = s_{(n)} \right) = \mathbb{P} \left(\varepsilon n^{-1} \left\| \Sigma_{\theta_0} (s_{(n)})^{1/2} V_n \right\|^2 > \delta \right),
$$

where V_n is a Gauss vector, defined on $(\Omega, \mathcal{F}, \mathbb{P})$, with zero-mean vector and identity covariance matrix. Then, we use Markov's inequality to estimate

$$
\begin{aligned}
\mathbb{P} \left(\varepsilon n^{-1} \left\| \Sigma_{\theta_0} (s_{(n)})^{1/2} V_n \right\|^2 > \delta \right) &\leq \varepsilon n^{-1} \delta^{-1} \mathbb{E} \left[\left\| \Sigma_{\theta_0} (s_{(n)})^{1/2} V_n \right\|^2 \right] \\
&\leq \varepsilon \delta^{-1} \left\| \Sigma_{\theta_0} (s_{(n)})^{1/2} \right\|_2^2,
\end{aligned}
$$

where the latter term is bounded uniformly in $s_{(n)} \in \mathcal{G}_n$ and $n \in \mathbb{N}_+$ (see Lemma B.3). Thus we conclude that

$$
\sup_{s_{(n)} \in \mathcal{G}_n} \mathbb{P} \left(\sup_{\theta \in \theta} \left\| A_{n,\theta} - \widetilde{A}_{n,\theta} \right\|_2 \frac{\left\| Z_{(n)} \right\|^2}{n} > \delta \,\Big|\, S_{(n)} = s_{(n)} \right) \xrightarrow{n \to \infty} 0,
$$

which shows that

$$\sup_{\theta \in \theta} \left\| A_{n,\theta} - \widetilde{A}_{n,\theta} \right\|_2 \frac{\left\| Z_{(n)} \right\|^2}{n} \xrightarrow[n \to \infty]{\mathbb{P}} 0,$$

and thus the proof is complete. $\qquad \square$

Proof of Lemma B.8. For $n \in \mathbb{N}_+$, let $h(n) = n + N - 1$. Then, for $k = 1, \ldots, p$, we have that $\mathbb{P}$ a.s.

$$\frac{\partial \tilde{l}_{n,N}}{\partial \theta_k}(\theta_0) = \frac{1}{h(n)} \left(\operatorname{tr} \left(\widetilde{\Sigma}_{h(n),\theta_0}^{-1} \frac{\partial \widetilde{\Sigma}_{h(n),\theta_0}}{\partial \theta_k} \right) \right.$$
$$\left. - \langle Z_{(h(n))}, \widetilde{\Sigma}_{h(n),\theta_0}^{-1} \frac{\partial \widetilde{\Sigma}_{h(n),\theta_0}}{\partial \theta_k} \widetilde{\Sigma}_{h(n),\theta_0}^{-1} Z_{(h(n))} \rangle \right),$$

and

$$\mathbb{E} \left[\frac{\partial \tilde{l}_{n,N}}{\partial \theta_k}(\theta_0) \,\middle|\, S_{(h(n))} \right] = \frac{1}{h(n)} \left(\operatorname{tr} \left(\widetilde{\Sigma}_{h(n),\theta_0}^{-1} \frac{\partial \widetilde{\Sigma}_{h(n),\theta_0}}{\partial \theta_k} \right) \right.$$
$$\left. - \operatorname{tr} \left(\widetilde{\Sigma}_{h(n),\theta_0}^{-1} \frac{\partial \widetilde{\Sigma}_{h(n),\theta_0}}{\partial \theta_k} \widetilde{\Sigma}_{h(n),\theta_0}^{-1} \Sigma_{h(n),\theta_0} \right) \right).$$

Similar expressions can then be calculated for $l_{n,N}$ based on $\Sigma_{h(n),\theta_0}$. We can further calculate, for $n \in \mathbb{N}_+$, for $1 \le k, l \le p$, $\mathbb{P}$ a.s.,

$$\frac{\partial^2 \tilde{l}_{n,N}}{\partial \theta_k \partial \theta_l}(\theta_0) = \frac{1}{h(n)} \operatorname{tr} \left(\widetilde{A}_{1,h(n),\theta_0}^{kl} \right) + \frac{1}{h(n)} \langle Z_{(h(n))}, \widetilde{A}_{2,h(n),\theta_0}^{kl} Z_{(h(n))} \rangle,$$

where

$$\widetilde{A}_{1,h(n),\theta_0}^{kl} := -\widetilde{\Sigma}_{h(n),\theta_0}^{-1} \frac{\partial \widetilde{\Sigma}_{h(n),\theta_0}}{\partial \theta_k} \widetilde{\Sigma}_{h(n),\theta_0}^{-1} \frac{\partial \widetilde{\Sigma}_{h(n),\theta_0}}{\partial \theta_l} + \widetilde{\Sigma}_{h(n),\theta_0}^{-1} \frac{\partial^2 \widetilde{\Sigma}_{h(n),\theta_0}}{\partial \theta_k \partial \theta_l}, \quad (33)$$

and

$$\widetilde{A}_{2,h(n),\theta_0}^{kl} := 2 \widetilde{\Sigma}_{h(n),\theta_0}^{-1} \frac{\partial \widetilde{\Sigma}_{h(n),\theta_0}}{\partial \theta_k} \widetilde{\Sigma}_{h(n),\theta_0}^{-1} \frac{\partial \widetilde{\Sigma}_{h(n),\theta_0}}{\partial \theta_l} \widetilde{\Sigma}_{h(n),\theta_0}^{-1}$$
$$- \widetilde{\Sigma}_{h(n),\theta_0}^{-1} \frac{\partial^2 \widetilde{\Sigma}_{h(n),\theta_0}}{\partial \theta_k \partial \theta_l} \widetilde{\Sigma}_{h(n),\theta_0}^{-1}. \quad (34)$$

In addition, for $n \in \mathbb{N}_+$, we also have that $\mathbb{P}$ a.s.,

$$\frac{\partial \left(\mathbb{E} \left[\frac{\partial \tilde{l}_{n,N}}{\partial \theta_l}(\theta_0) \,\middle|\, S_{(h(n))} \right] \right)}{\partial \theta_k} = \frac{1}{h(n)} \left(\operatorname{tr} \left(\widetilde{A}_{1,h(n),\theta_0}^{kl} \right) + \operatorname{tr} \left(\widetilde{A}_{2,h(n),\theta_0}^{kl} \Sigma_{h(n),\theta_0} \right) \right).$$

Again, similar expressions can be obtained for $l_{n,N}$ based on $\Sigma_{h(n),\theta_0}$, where for $n \in \mathbb{N}_+$, $1 \le k, l \le p$, the respective terms $A_{1,h(n),\theta_0}^{kl}$ and $A_{2,h(n),\theta_0}^{kl}$ are defined

as in (33) and (34), respectively, but $\widetilde{\Sigma}_{h(n),\theta_0}$ is replaced with $\Sigma_{h(n),\theta_0}$. Then, we have for $n \in \mathbb{N}_+$, for $k, l = 1, \ldots, p$, $\mathbb{P}$ a.s.,

$$
\begin{aligned}
\left| \frac{\partial^2 \tilde{l}_{n,N}}{\partial \theta_k \partial \theta_l}(\theta_0) - \frac{\partial^2 l_{n,N}}{\partial \theta_k \partial \theta_l}(\theta_0) \right| &\leq \frac{1}{h(n)} \left| \operatorname{tr} \left(\widetilde{A}_{1,h(n),\theta_0}^{kl} - A_{1,h(n),\theta_0}^{kl} \right) \right| \\
&\quad + \frac{1}{h(n)} \left| \langle Z_{(h(n))}, \widetilde{A}_{2,h(n),\theta_0}^{kl} - A_{2,h(n),\theta_0}^{kl} Z_{(h(n))} \rangle \right| \\
&\leq \frac{1}{h(n)} \left| \operatorname{tr} \left(\widetilde{A}_{1,h(n),\theta_0}^{kl} - A_{1,h(n),\theta_0}^{kl} \right) \right| \\
&\quad + \left\| \widetilde{A}_{2,h(n),\theta_0}^{kl} - A_{2,h(n),\theta_0}^{kl} \right\|_2 \frac{\left\| Z_{(h(n))} \right\|^2}{h(n)}.
\end{aligned}
$$

We can apply Lemma B.7 to the sequence of random matrices $\left(\widetilde{A}_{2,h(n),\theta_0}^{kl} \right)_{n \in \mathbb{N}_+}$ and $\left(A_{2,h(n),\theta_0}^{kl} \right)_{n \in \mathbb{N}_+}$ to conclude under application of Lemma 4.1 (see also Corollary B.4 and Lemma B.5) that

$$
\left\| \widetilde{A}_{2,h(n),\theta_0}^{kl} - A_{2,h(n),\theta_0}^{kl} \right\|_2 \frac{\left\| Z_{(h(n))} \right\|^2}{h(n)} \xrightarrow[n \to \infty]{\mathbb{P}} 0.
$$

We also have $\mathbb{P}$ a.s.

$$
\frac{1}{n} \left| \operatorname{tr} \left(\widetilde{A}_{1,h(n),\theta_0}^{kl} - A_{1,h(n),\theta_0}^{kl} \right) \right| \xrightarrow{n \to \infty} 0,
$$

using the triangular inequality, von Neumann's trace inequality and Lemma 4.1 (see also Corollary B.4 and Lemma B.5). Hence, we have shown that for any $k, l = 1, \ldots, p$

$$
\left| \frac{\partial^2 \tilde{l}_{n,N}}{\partial \theta_k \partial \theta_l}(\theta_0) - \frac{\partial^2 l_{n,N}}{\partial \theta_k \partial \theta_l}(\theta_0) \right| \xrightarrow[n \to \infty]{\mathbb{P}} 0.
$$

In addition, we have that for any $k, l = 1, \ldots, p$, $\mathbb{P}$ a.s., the expression

$$
\left| \frac{\partial \left(\mathrm{E} \left[\frac{\partial \tilde{l}_{n,N}}{\partial \theta_l}(\theta_0) \mid S_{(h(n))} \right] \right)}{\partial \theta_k} - \frac{\partial \left(\mathrm{E} \left[\frac{\partial l_{n,N}}{\partial \theta_l}(\theta_0) \mid S_{(h(n))} \right] \right)}{\partial \theta_k} \right|
$$

is bounded from above by

$$
\frac{\left| \operatorname{tr} \left(\widetilde{A}_{1,h(n),\theta_0}^{kl} - A_{1,h(n),\theta_0}^{kl} \right) \right| + \left| \operatorname{tr} \left(\left(\widetilde{A}_{2,h(n),\theta_0}^{kl} - A_{2,h(n),\theta_0}^{kl} \right) \Sigma_{h(n),\theta_0} \right) \right|}{h(n)},
$$

which again, under application of the triangular inequality, von Neumann's trace inequality and Lemma 4.1 (see also Corollary B.4 and Lemma B.5), converges to zero $\mathbb{P}$ a.s. Hence, we have shown that for any $k, l = 1, \ldots, p$,

$$
\left| \left(J_{\widetilde{G}_{n,N}}(\theta_0) \right)_{kl} - \left(J_{G_{n,N}}(\theta_0) \right)_{kl} \right| \xrightarrow[n \to \infty]{\mathbb{P}} 0,
$$

which concludes the proof of (26). Now it is shown in [4] (see Propositions D.7 and D.8 and also consider the proofs of Propositions 3.2 and 3.3), under application of Lemmas B.1, B.2, 4.1, B.5 and Corollary B.4, that

$$J_{G_{n,N}}(\theta_0) \xrightarrow[n\to\infty]{\mathbb{P}} 2\Lambda,$$

where Λ is the $\mathbb{P}$ a.s. limit of a sequence of $p \times p$ matrices $\big(H_{h(n)}(\theta_0)\big)_{n\in\mathbb{N}_+}$ defined as

$$\left\{ \left[\frac{1}{2h(n)} \operatorname{tr} \left(\Sigma_{h(n),\theta_0}^{-1} \frac{\partial \Sigma_{h(n),\theta_0}}{\partial\theta_k} \Sigma_{h(n),\theta_0}^{-1} \frac{\partial \Sigma_{h(n),\theta_0}}{\partial\theta_l} \right) \right]_{1\le k,l\le p} : n \in \mathbb{N}_+ \right\}.$$

Further, by Assumption 5.2, it is concluded that the limit Λ is such that $\Lambda \succ 0$. But then, we use (26) to show that

$$J_{\widetilde{G}_{n,N}}(\theta_0) \xrightarrow[n\to\infty]{\mathbb{P}} 2\Lambda,$$

as well, which concludes the proof of Lemma B.8. $\square$

C.2. Proof of results in Section 4

Proof of Lemma 4.1. We rely on Lemma B.3 and a proof is evident. $\square$

C.3. Proof of results in Section 5

To simplify the notation, we write $\big(\tilde{\theta}_n\big)_{n\in\mathbb{N}_+} \coloneqq \big(\hat{\theta}_n(\tilde{c})\big)_{n\in\mathbb{N}_+}$.

Proof of Proposition 5.1. The statement is verified as a consequence of Lemmas 4.1, B.6 and B.7. $\square$

Proof of Theorem 5.2. Let $N \in \mathbb{N}_+$ be as in Lemma 4.1 (or Proposition 5.1) and define, for any $\omega \in \Omega$, the sequence $\big(\tilde{l}_{n,N}(\theta)\,(\omega)\,\big)_{n\in\mathbb{N}_+}$ as in (9) of Section 5.1. We note that, under the given assumptions of Theorem 5.2, the first order partial derivatives with respect to θ exist for the sequence $\big(\tilde{l}_{n,N}(\theta)\,(\omega)\,\big)_{n\in\mathbb{N}_+}$. Then, we define the sequence of estimators $\big(\tilde{\theta}_{n,N}\big)_{n\in\mathbb{N}_+} \coloneqq \big(\tilde{\theta}_{n+N-1}\big)_{n\in\mathbb{N}_+}$. Therefore $\tilde{\theta}_{n,N}$ minimizes $\tilde{l}_{n,N}(\theta)$ $\mathbb{P}$ a.s. for any $n \in \mathbb{N}_+$. To prove that

$$\tilde{\theta}_n \xrightarrow[n\to\infty]{\mathbb{P}} \theta_0,$$

it is sufficient to show that

$$\tilde{\theta}_{n,N} \xrightarrow[n\to\infty]{\mathbb{P}} \theta_0. \tag{35}$$

We consider a similar approach as given in [4]. As N is fixed, we write for $n \in \mathbb{N}_+$, $h(n) = n + N - 1$. Under the assumptions of the theorem we have that $\mathbb{P}$ a.s.

$$\operatorname{Var}\big(\tilde{l}_{n,N}(\theta) \mid S_{(h(n))}\big) \xrightarrow{n\to\infty} 0, \tag{36}$$

and

$$\max_{k=1,\dots,p} \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta_k} \tilde{l}_{n,N}(\theta) \right| = O_{\mathbb{P}}(1) \text{ as } n \to \infty. \tag{37}$$

To see it, we remark that $\mathbb{P}$ a.s. (using Proposition 5.1),

$$\text{Var}\left(\tilde{l}_{n,N}(\theta) \mid S_{(h(n))}\right) = \frac{2}{h(n)^2} \text{tr}\left(\underbrace{\widetilde{\Sigma}_{h(n),\theta}^{-1} \Sigma_{h(n),\theta_0} \widetilde{\Sigma}_{h(n),\theta}^{-1} \Sigma_{h(n),\theta_0}}_{=:\widetilde{A}_{h(n),\theta}} \right).$$

From here, we can use von Neumann's trace inequality to show that $\mathbb{P}$ a.s.

$$\left| \text{tr}\left(\widetilde{A}_{h(n),\theta}\right) \right| \leq h(n) \left\| \widetilde{\Sigma}_{h(n),\theta}^{-1} \right\|_2^2 \left\| \Sigma_{h(n),\theta_0} \right\|_2^2.$$

Now, by Lemma 4.1 (and Corollary B.4) we can conclude that there exists a real constant $M_0 > 0$, such that for any $n \in \mathbb{N}_+$, $\mathbb{P}$ a.s., $\text{Var}\left(\tilde{l}_{n,N}(\theta) \mid S_{(h(n))}\right) \leq M_0/h(n)$, which proofs (36). For (37), we first notice that by Lemma B.5, there exist constants $M_1, M_2 > 0$ (which are independent of $n \in \mathbb{N}_+$, $s_{(n)} \in \mathcal{G}_n$ and $\theta \in \Theta$) such that $\mathbb{P}$ a.s.

$$\sup_{\theta \in \Theta} \left\| \widetilde{\Sigma}_{n,\theta}^{-1} \right\|_2 < M_1 \text{ and } \max_{k=1,\dots,p} \sup_{\theta \in \Theta} \left\| \frac{\partial}{\partial \theta_k} \widetilde{\Sigma}_{n,\theta} \right\|_2 \leq M_2.$$

Using this result have that $\mathbb{P}$ a.s.

$$\begin{aligned}
\max_{k=1,\dots,p} \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta_k} \tilde{l}_{n,N}(\theta) \right| &= \max_{k=1,\dots,p} \sup_{\theta \in \Theta} \left| \frac{1}{h(n)} \text{tr}\left(\widetilde{\Sigma}_{h(n),\theta}^{-1} \frac{\partial}{\partial \theta_k} \widetilde{\Sigma}_{h(n),\theta} \right) \right. \\
&\qquad \left. - \frac{1}{h(n)} \langle Z_{(h(n))}, \widetilde{\Sigma}_{h(n),\theta}^{-1} \frac{\partial}{\partial \theta_k} \widetilde{\Sigma}_{h(n),\theta} \widetilde{\Sigma}_{h(n),\theta}^{-1} Z_{(h(n))} \rangle \right| \\
&\leq M_1 M_2 + M_1^2 M_2 \frac{\left\| Z_{(h(n))} \right\|^2}{h(n)}.
\end{aligned} \tag{38}$$

Let $V_{h(n)}$ be a Gauss vector on $(\Omega, \mathcal{F}, \mathbb{P})$, with zero-mean vector and $h(n) \times h(n)$ identity covariance matrix. Then (see also Remark 2.1), for any finite $M > 0$, we have that the probability

$$\mathbb{P}\left(M_1 M_2 + M_1^2 M_2 h(n)^{-1} \left\| Z_{(h(n))} \right\|^2 > M \mid S_{(h(n))} = s_{(h(n))}\right)$$

is bounded from above by

$$\mathbb{P}\left(M_1 M_2 \left(1 + h(n)^{-1} \| V_{h(n)} \|^2\right) > M\right).$$

Therefore, $\mathbb{P}$ a.s.,

$$\mathbb{P}\left(M_1 M_2 + M_1^2 M_2 h(n)^{-1} \left\| Z_{(h(n))} \right\|^2 > M \mid S_{(h(n))}\right)$$

is bounded from above by $\mathbb{P}\big(M_1 M_2\big(1 + h(n)^{-1}\|V_{h(n)}\|^2\big) > M\big)$ as well. Since $M_1 M_2\big(1 + h(n)^{-1}\|V_{h(n)}\|^2\big) = O_{\mathbb{P}}(1)$ as $n \to \infty$, (37) is shown.

Notice further that Θ is convex, $\theta \mapsto \tilde{l}_{n,N}(\theta)$ is continuously differentiable and by (38)

$$\sup_{n \in \mathbb{N}_+} \mathbb{E}\left[\max_{k=1,\ldots,p} \sup_{\theta \in \Theta} \left| \frac{\partial}{\partial \theta_k} \tilde{l}_{n,N}(\theta) \right| \right] < \infty.$$

Thus, under application of Corollary 2.2 of [27], with (36) and (37), we can conclude that

$$\sup_{\theta \in \Theta} \left| \tilde{l}_{n,N}(\theta) - \mathbb{E}\left[\tilde{l}_{n,N}(\theta) \mid S_{(h(n))} \right] \right| \xrightarrow[n \to \infty]{\mathbb{P}} 0. \tag{39}$$

To continue, we define the sequences of random variables

$$\big(D_{h(n),\theta,\theta_0}\big)_{n \in \mathbb{N}_+} := \big(\mathbb{E}\big[l_{n,N}(\theta) \mid S_{(h(n))}\big] - \mathbb{E}\big[l_{n,N}(\theta_0) \mid S_{(h(n))}\big]\big)_{n \in \mathbb{N}_+},$$

$$\big(\widetilde{D}_{h(n),\theta,\theta_0}\big)_{n \in \mathbb{N}_+} := \big(\mathbb{E}\big[\tilde{l}_{n,N}(\theta) \mid S_{(h(n))}\big] - \mathbb{E}\big[\tilde{l}_{n,N}(\theta_0) \mid S_{(h(n))}\big]\big)_{n \in \mathbb{N}_+}.$$

For any $n \in \mathbb{N}_+$, we have that $\mathbb{P}$ a.s.,

$$D_{h(n),\theta,\theta_0} = \frac{1}{h(n)} \log\big(\det\big(\Sigma_{h(n),\theta}\big)\big) + \frac{1}{h(n)} \operatorname{tr}\big(\Sigma_{h(n),\theta}^{-1} \Sigma_{h(n),\theta_0}\big)$$
$$- \frac{1}{h(n)} \log\big(\det\big(\Sigma_{h(n),\theta_0}\big)\big) - \frac{1}{h(n)} \operatorname{tr}\big(\Sigma_{h(n),\theta_0}^{-1} \Sigma_{h(n),\theta_0}\big).$$

Similarly, For any $n \in \mathbb{N}_+$, we have that $\mathbb{P}$ a.s.

$$\widetilde{D}_{h(n),\theta,\theta_0} = \frac{1}{h(n)} \log\Big(\det\Big(\widetilde{\Sigma}_{h(n),\theta}\Big)\Big) + \frac{1}{h(n)} \operatorname{tr}\Big(\widetilde{\Sigma}_{h(n),\theta}^{-1} \Sigma_{h(n),\theta_0}\Big)$$
$$- \frac{1}{h(n)} \log\Big(\det\Big(\widetilde{\Sigma}_{h(n),\theta_0}\Big)\Big) - \frac{1}{h(n)} \operatorname{tr}\Big(\widetilde{\Sigma}_{h(n),\theta_0}^{-1} \Sigma_{h(n),\theta_0}\Big).$$

Notice that because of (39) we have that

$$\sup_{\theta \in \Theta} \left| \big(\tilde{l}_{n,N}(\theta) - \tilde{l}_{n,N}(\theta_0)\big) - \widetilde{D}_{h(n),\theta,\theta_0} \right| \xrightarrow[n \to \infty]{\mathbb{P}} 0.$$

Further, it is shown in [4] (see the proof of Proposition 3.1) that under application of Lemma B.3 there exists some constant $B > 0$ (which does not depend on $n \in \mathbb{N}_+$) such that $\mathbb{P}$ a.s.

$$D_{h(n),\theta,\theta_0} \geq B \underbrace{\frac{1}{h(n)} \sum_{i,j=1}^{h(n)} \Big(c_\theta(S_i - S_j) - c_{\theta_0}(S_i - S_j)\Big)^2}_{=:D_{2,h(n),\theta,\theta_0}}.$$

Under application of Lemmas B.1, B.2, B.3 and Corollary B.4, it is then shown in the proof of Proposition 3.1 of [4] that either, if $\tau = 0$, $D_{2,h(n),\theta,\theta_0}$ is deterministic and we have that

$$\sup_{\theta \in \Theta} \left| D_{2,h(n),\theta,\theta_0} - D_{\infty,\theta,\theta_0} \right| \xrightarrow{n \to \infty} 0,$$

where the limit is given by $D_{\infty,\theta,\theta_0} = \sum_{z \in \mathbb{Z}^d} \left(c_\theta(z) - c_{\theta_0}(z) \right)^2$. Or $\tau > 0$ and it is concluded that

$$\sup_{\theta \in \Theta} \left| D_{2,h(n),\theta,\theta_0} - D_{\infty,\theta,\theta_0} \right| \xrightarrow[n \to \infty]{\mathbb{P}} 0,$$

where in this case

$$D_{\infty,\theta,\theta_0} = \int_{D_\tau} \left(c_\theta(s) - c_{\theta_0}(s) \right)^2 f(s) ds + \left(c_\theta(0) - c_{\theta_0}(0) \right)^2.$$

Notice that because of the assumption that $(X_i)_{i \in \mathbb{N}_+}$ is independent with common law that has a strictly positive probability density function, the function f is strictly positive almost everywhere with respect to the Lebesgue measure on D_τ (see the end of the proof of Proposition 3.1 in [4]). In either case, we can thus conclude that

$$\sup_{\theta \in \Theta} \left| D_{2,h(n),\theta,\theta_0} - D_{\infty,\theta,\theta_0} \right| \xrightarrow[n \to \infty]{\mathbb{P}} 0,$$

where for any $\alpha > 0$, because of Assumption 5.1, $\inf_{\theta : |\theta - \theta_0| \geq \alpha} D_{\infty,\theta,\theta_0} > 0$, and the limit $D_{\infty,\theta,\theta_0}$ is deterministic. We now want to show that there exists some $N_2 \geq N$ such that for any $n \geq N_2$, for any $\theta \in \Theta$, $\mathbb{P}$ a.s.,

$$\widetilde{D}_{h(n),\theta,\theta_0} \geq B D_{2,h(n),\theta,\theta_0}, \tag{40}$$

as well. In this case, with $D_{2,h(n),\theta,\theta_0}$ a random function on Ω and $D_{\infty,\theta,\theta_0}$ a deterministic function of $\theta \in \Theta$, we would have for any fixed $\tau \geq 0$, and for any given $\alpha > 0$,

$$\sup_{\theta \in \Theta} \left| D_{2,h(n),\theta,\theta_0} - D_{\infty,\theta,\theta_0} \right| \xrightarrow[n \to \infty]{\mathbb{P}} 0,$$

$$\inf_{\theta : |\theta - \theta_0| \geq \alpha} D_{\infty,\theta,\theta_0} > D_{\infty,\theta_0,\theta_0} = 0,$$

where the sequence of estimators $\left(\tilde{\theta}_{n,N} \right)_{n \in \mathbb{N}_+}$, is such that for $n \geq N_2$, $\mathbb{P}$ a.s.,

$$D_{2,h(n),\tilde{\theta}_{n,N},\theta_0}$$
$$= D_{2,h(n),\tilde{\theta}_{n,N},\theta_0} - \frac{1}{B}\left(\tilde{l}_{n,N}\left(\tilde{\theta}_{n,N} \right) - \tilde{l}_{n,N}\left(\theta_0 \right) \right) + \frac{1}{B}\left(\tilde{l}_{n,N}\left(\tilde{\theta}_{n,N} \right) - \tilde{l}_{n,N}\left(\theta_0 \right) \right)$$
$$\leq \frac{\widetilde{D}_{h(n),\tilde{\theta}_{n,N},\theta_0}}{B} - \frac{1}{B}\left(\tilde{l}_{n,N}\left(\tilde{\theta}_{n,N} \right) - \tilde{l}_{n,N}\left(\theta_0 \right) \right) + \frac{1}{B}\left(\tilde{l}_{n,N}\left(\tilde{\theta}_{n,N} \right) - \tilde{l}_{n,N}\left(\theta_0 \right) \right)$$

$$\leq D_{2,h(n),\theta_0,\theta_0} + \underbrace{\frac{1}{B} \sup_{\theta \in \Theta} \left| \left(\tilde{l}_{n,N}(\theta) - \tilde{l}_{n,N}(\theta_0) \right) - \widetilde{D}_{h(n),\theta,\theta_0} \right|}_{\xrightarrow[n \to \infty]{\mathbb{P}} 0},$$

and we can conclude the proof of Theorem 5.2, using Theorem 5.7 of [31]. Hence, it remains to show (40). We write $\mathbb{P}$ a.s.,

$$\left| \widetilde{D}_{h(n),\theta,\theta_0} - D_{h(n),\theta,\theta_0} \right| \leq \left| \widetilde{A}_{1,h(n),\theta,\theta_0} - A_{1,h(n),\theta,\theta_0} \right|$$
$$+ \left| \widetilde{A}_{2,h(n),\theta,\theta_0} - A_{2,h(n),\theta,\theta_0} \right|$$
$$+ \left| \widetilde{A}_{3,h(n),\theta,\theta_0} - A_{3,h(n),\theta,\theta_0} \right|,$$

where

$$\widetilde{A}_{1,h(n),\theta,\theta_0} - A_{1,h(n),\theta,\theta_0} = \frac{1}{h(n)} \log \left(\det \left(\Sigma_{h(n),\theta_0} \Sigma_{h(n),\theta}^{-1} \right) \right)$$
$$- \frac{1}{h(n)} \log \left(\det \left(\widetilde{\Sigma}_{h(n),\theta_0} \widetilde{\Sigma}_{h(n),\theta}^{-1} \right) \right),$$

$$\widetilde{A}_{2,h(n),\theta,\theta_0} - A_{2,h(n),\theta,\theta_0} = \frac{1}{h(n)} \operatorname{tr} \left(\left[\widetilde{\Sigma}_{h(n),\theta}^{-1} - \Sigma_{h(n),\theta}^{-1} \right] \Sigma_{h(n),\theta_0} \right),$$

and

$$\widetilde{A}_{3,h(n),\theta,\theta_0} - A_{3,h(n),\theta,\theta_0} = \frac{1}{h(n)} \operatorname{tr} \left(\left[\widetilde{\Sigma}_{h(n),\theta_0}^{-1} - \Sigma_{h(n),\theta_0}^{-1} \right] \Sigma_{h(n),\theta_0} \right).$$

By Lemma 4.1, Corollary B.4 and Lemma B.6, we already conclude that $\mathbb{P}$ a.s.

$$\left| \widetilde{A}_{1,h(n),\theta,\theta_0} - A_{1,h(n),\theta,\theta_0} \right|$$

converges to zero uniformly in $\theta \in \Theta$ as $n \to \infty$. Further, we can conclude that $\mathbb{P}$ a.s.

$$\left| \widetilde{A}_{2,h(n),\theta,\theta_0} - A_{2,h(n),\theta,\theta_0} \right| \leq \left\| \widetilde{\Sigma}_{h(n),\theta}^{-1} - \Sigma_{h(n),\theta}^{-1} \right\|_2 \left\| \Sigma_{h(n),\theta_0} \right\|_2,$$

and thus, since by Lemma 4.1 $\mathbb{P}$ a.s. $\|\Sigma_{h(n),\theta_0}\|_2$, $\|\widetilde{\Sigma}_{h(n),\theta}\|_2$ and $\|\Sigma_{h(n),\theta}\|_2$ are finite, uniformly in $n \in \mathbb{N}_+$ and $\theta \in \Theta$, and $\mathbb{P}$ a.s.

$$\left\| \widetilde{\Sigma}_{h(n),\theta}^{-1} - \Sigma_{h(n),\theta}^{-1} \right\|_2 \xrightarrow{n \to \infty} 0,$$

uniformly in $\theta \in \Theta$, by application of Corollary B.4, we can also see that $\mathbb{P}$ a.s.

$$\left| \widetilde{A}_{2,h(n),\theta,\theta_0} - A_{2,h(n),\theta,\theta_0} \right|$$

converges to zero as $n \to \infty$, uniformly in $\theta \in \Theta$. Using a similar argument we can also show that $\mathbb{P}$ a.s. the term $\left| \widetilde{A}_{3,h(n),\theta,\theta_0} - A_{3,h(n),\theta,\theta_0} \right|$ converges to zero as $n \to \infty$, uniformly in $\theta \in \Theta$. Hence, we have shown that $\mathbb{P}$ a.s.

$$\left| \widetilde{D}_{h(n),\theta,\theta_0} - D_{h(n),\theta,\theta_0} \right| \xrightarrow{n\to\infty} 0,$$

uniformly in $\theta \in \Theta$ and we can argue that $\mathbb{P}$ a.s. there exists some $N_2 \geq N$ such that for all $n \geq N_2$, $\widetilde{D}_{h(n),\theta,\theta_0} \geq BD_{2,h(n),\theta,\theta_0}$ on Ω, which shows (40). Therefore, we have that

$$\tilde{\theta}_{n,N} \xrightarrow[n\to\infty]{\mathbb{P}} \theta_0,$$

which concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Proof of Corollary 5.3. This follows from Theorem 5.2 when we define, for any $\theta \in \Theta$ and $m \in \mathbb{N}_+$, $\tilde{c}_{m,\theta}(s) := c_\theta(s)$, for all $s \in \mathbb{R}^d$. $\qquad\qquad\square$

Proof of Theorem 5.4. Let $N \in \mathbb{N}_+$ be as in Proposition 5.1 and define, for any $\omega \in \Omega$, the sequences of functions $\left(l_{n,N}(\theta)(\omega) \right)_{n\in\mathbb{N}_+}$ and $\left(\tilde{l}_{n,N}(\theta)(\omega) \right)_{n\in\mathbb{N}_+}$ as in (8) and (9) of Section 5.1, respectively. From the proof of Theorem 5.2 we know that sequence of estimators $(\tilde{\theta}_{n,N})_{n\in\mathbb{N}_+} := (\tilde{\theta}_{n+N-1})_{n\in\mathbb{N}_+}$, is such that

$$\tilde{\theta}_{n,N} \xrightarrow[n\to\infty]{\mathbb{P}} \theta_0.$$

Define $\left\{ (G_{n,N}(\theta))_{n\in\mathbb{N}_+} : \theta \in \Theta \right\}$ and $\left\{ (\widetilde{G}_{n,N}(\theta))_{n\in\mathbb{N}_+} : \theta \in \Theta \right\}$ as in (10) and (11) respectively. For $n \in \mathbb{N}_+$ we set $h(n) = n+N-1$. We have for $k = 1, \ldots, p$, $\mathbb{P}$ a.s., for $n \in \mathbb{N}_+$,

$$\begin{aligned}
\tilde{c}_{k,n,N}(\theta_0) &= \frac{\partial \tilde{l}_{n,N}}{\partial \theta_k}(\theta) - \mathbb{E}\left[\frac{\partial \tilde{l}_{n,N}}{\partial \theta_k}(\theta) \;\middle|\; S_{(h(n))} \right] \\
&= \frac{1}{h(n)} \operatorname{tr}\Big(\underbrace{\widetilde{\Sigma}_{h(n),\theta_0}^{-1} \frac{\partial \widetilde{\Sigma}_{h(n),\theta_0}}{\partial \theta_k} \widetilde{\Sigma}_{h(n),\theta_0}^{-1} \Sigma_{h(n),\theta_0}}_{=:\widetilde{M}_{k,h(n)}} \Big) \\
&\quad + \frac{1}{h(n)} \langle Z_{(h(n))}, \underbrace{-\widetilde{\Sigma}_{h(n),\theta_0}^{-1} \frac{\partial \widetilde{\Sigma}_{h(n),\theta_0}}{\partial \theta_k} \widetilde{\Sigma}_{h(n),\theta_0}^{-1}}_{=:\widetilde{N}_{k,h(n)}} Z_{(h(n))} \rangle,
\end{aligned}$$

where, by Lemma 4.1 (see also Corollary B.4 and Lemma B.5) $\mathbb{P}$ a.s., for $n \in \mathbb{N}_+$, $\|\widetilde{M}_{k,h(n)}\|_2$ and $\|\widetilde{N}_{k,h(n)}\|_2$ are finite, uniformly in $\theta \in \Theta$. Further, notice that $\mathbb{P}$ a.s.

$$\operatorname{tr}\left(\widetilde{M}_{k,h(n)} + \widetilde{N}_{k,h(n)}\Sigma_{h(n),\theta_0} \right) = 0 \quad \forall k \in \{1, \ldots, p\}.$$

From the proof of Lemma B.8, we already know that $\mathbb{P}$ a.s., $H_{h(n)}(\theta_0) \xrightarrow{n\to\infty} \Lambda$, where for any $k,l = 1,\ldots,p$, for any $n \in \mathbb{N}_+$,

$$\left[H_{h(n)}(\theta_0)\right]_{kl} = \frac{1}{2h(n)} \operatorname{tr}\left(\Sigma_{h(n),\theta_0}^{-1} \frac{\partial \Sigma_{h(n),\theta_0}}{\partial \theta_k} \Sigma_{h(n),\theta_0}^{-1} \frac{\partial \Sigma_{h(n),\theta_0}}{\partial \theta_l}\right).$$

Now, if we define, on $(\Omega, \mathcal{F}, \mathbb{P})$, the sequence of random $p \times p$ matrices

$$\left\{\underbrace{\frac{1}{2h(n)}\left[\operatorname{tr}\left(\widetilde{N}_{k,h(n)}\Sigma_{h(n),\theta_0}\widetilde{N}_{\{l,h(n)\}}\Sigma_{h(n),\theta_0}\right)\right]_{\{1\le k,l\le p\}}}_{=:\widetilde{H}_{h(n)}(\theta_0)} : n \in \mathbb{N}_+\right\}$$

we also have that, for any $k,l = 1,\ldots p$, $\mathbb{P}$ a.s., $\left|\left[\widetilde{H}_{h(n)}(\theta_0)\right]_{kl} - \Sigma_{kl}\right| \xrightarrow{n\to\infty} 0$. This follows from the fact that for any $k,l = 1,\ldots p$, we have that $\mathbb{P}$ a.s.

$$\left|\left[\widetilde{H}_{h(n)}(\theta_0)\right]_{kl} - \left[H_{h(n)}(\theta_0)\right]_{kl}\right| \le \frac{1}{2h(n)}\left|\operatorname{tr}\left(\widetilde{B}_{h(n)}^{kl} - B_{h(n)}^{kl}\right)\right|,$$

where

$$\widetilde{B}_{h(n)}^{kl} = \widetilde{\Sigma}_{h(n),\theta_0}^{-1} \frac{\partial \widetilde{\Sigma}_{h(n),\theta_0}}{\partial \theta_k} \widetilde{\Sigma}_{h(n),\theta_0}^{-1} \Sigma_{h(n),\theta_0} \widetilde{\Sigma}_{h(n),\theta_0}^{-1} \frac{\partial \widetilde{\Sigma}_{h(n),\theta_0}}{\partial \theta_l} \widetilde{\Sigma}_{h(n),\theta_0}^{-1} \Sigma_{h(n),\theta_0},$$

and

$$B_{h(n)}^{kl} = \Sigma_{h(n),\theta_0}^{-1} \frac{\partial \Sigma_{h(n),\theta_0}}{\partial \theta_k} \Sigma_{h(n),\theta_0}^{-1} \Sigma_{h(n),\theta_0} \Sigma_{h(n),\theta_0}^{-1} \frac{\partial \Sigma_{h(n),\theta_0}}{\partial \theta_l} \Sigma_{h(n),\theta_0}^{-1} \Sigma_{h(n),\theta_0},$$

and again, under application of the triangular inequality, von Neumann's trace inequality and Lemma 4.1 (see also Corollary B.4 and Lemma B.5) we thus have that $\mathbb{P}$ a.s. $\|\widetilde{H}_{h(n)}(\theta_0) - H_{h(n)}(\theta_0)\|_2 \xrightarrow{n\to\infty} 0$. But Λ is the $\mathbb{P}$ a.s. limit of $\left\{H_{h(n)}(\theta_0) \colon n \in \mathbb{N}_+\right\}$ and hence we conclude that Λ is also the $\mathbb{P}$ a.s. limit of $\left\{\widetilde{H}_{h(n)}(\theta_0) \colon n \in \mathbb{N}_+\right\}$. Then, we can apply Proposition D.9 of [4] to conclude that

$$h(n)^{1/2}\widetilde{G}_{n,N}(\theta_0) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}\left(0, 4\Lambda\right).$$

Notice that because the family $\left\{(\tilde{c}_{m,\theta}) \colon \theta \in \Theta\right\}$ satisfies Assumption 3.2, we have that for fixed $\omega \in \Omega$, $\theta \mapsto \widetilde{G}_{n,N}(\omega,\theta)$ is twice differentiable in θ and we can argue exactly as in the proof of Theorem 5.2 to conclude that the sequence

$$\left(\sup_{\theta \in \Theta} \max_{\{1\le k,l,m\le p\}}\left|\frac{\partial\left(\frac{\partial \widetilde{G}_{m,n,N}}{\partial \theta_l}\right)}{\partial \theta_k}(\theta)\right|\right)_{n\in\mathbb{N}_+},$$

is bounded in probability $\mathbb{P}$. In addition, by Lemma B.8, we also have that

$$\left\|J_{\widetilde{G}_{n,N}}(\theta_0) - J_{G_{n,N}}(\theta_0)\right\|_2 \xrightarrow[n\to\infty]{\mathbb{P}} 0.$$

Finally, the sequence of estimators $\big(\tilde{\theta}_{n,N}\big)_{n\in\mathbb{N}_+}$ is consistent and such that

$$
\mathbb{P}\left(\widetilde{G}_{n,N}\big(\tilde{\theta}_{n,N}\big) = 0\right) \xrightarrow{n\to\infty} 1.
$$

Thus we conclude, using for example Proposition D.10 in [4] that

$$
h(n)^{1/2}\left(\tilde{\theta}_{n,N} - \theta_0\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0, \Lambda^{-1}).
$$

Since N was fixed, we can conclude that

$$
n^{1/2}\left(\tilde{\theta}_n - \theta_0\right) \xrightarrow[n\to\infty]{\mathrm{d}} \mathcal{N}(0, \Lambda^{-1}),
$$

as well. □

Proof of Corollary 5.5. The result follows from Theorem 5.4, a proof is evident when we define the family $\big\{\big(\tilde{c}_{m,\theta}\big)\colon \theta \in \Theta\big\}$ as in the proof of Corollary 5.3. □

C.4. *Proof of results in Appendix A*

Proof of Theorem A.1. The proof is similar to the proof of Theorem 5.2. □

Proof of Theorem A.2. The proof is similar to the proof of Theorem 5.4. □

C.5. *Proof of results in Section 6*

Since ν and κ are assumed to be known, we put $c_{\nu,\kappa} := \mathrm{B}(2\kappa, \nu+1)$. We define the function $f_{\nu,\kappa}(r, u) := u(u^2 - r^2)^{\kappa-1}(1-u)^\nu$, $(r, u) \in [0, 1] \times [0, 1]$. We recall that for $r = 0$,

$$
c_{\nu,\kappa} = \int_0^1 f_{\nu,\kappa}(0, u)du.
$$

Proof of Proposition 6.1. We have already seen that for any $\theta \in \Theta$, given known $\kappa > 0$ and $\nu \geq (d+1)/2 + \kappa$, ϕ_θ is continuous on $\mathbb{R}_+$. Further, for any $\theta \in \Theta$, ϕ_θ has compact support $S_\theta = [0, \beta] \subset [0, \beta_{\max}]$ and since $\kappa > 4$, we can also see that for any $\delta > 0$, for any $t \in \mathbb{R}_+$, $\phi_\theta(t + \delta) \leq \phi_\theta(t)$ and thus $\phi_\theta(t) \leq \phi_\theta(0) = \sigma^2$, which implies that $\|\phi_\theta\|_\infty \leq \sigma_{\max}^2$. Hence, with $C := \beta_{\max}$ and $L := \sigma_{\max}^2$, C and L are independent of $\theta \in \Theta$ and hence we can conclude that (1) of Assumption A.1 is satisfied with $\mathcal{B}_{\mathrm{C}}(\mathbb{R}_+; S_\theta)$ replaced with $\mathcal{C}_{\mathrm{C}}(\mathbb{R}_+; S_\theta)$. It is now sufficient to show that for any $\theta \in \Theta$, for any $q = 1, 2, 3$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, there exist constants $C_\theta(i_1, \ldots, i_q)$, $L_\theta(i_1, \ldots, i_q) < \infty$, such that

$$
\frac{\partial^q \phi_\theta}{\partial\theta_{i_1}\cdots\partial\theta_{i_q}} \in \mathcal{C}_{\mathrm{C}}(\mathbb{R}_+; S_\theta(i_1, \ldots, i_q)), \tag{41}
$$

where

$$
S_\theta(i_1, \ldots, i_q) \subset [0, C_\theta(i_1, \ldots, i_q)] \subset [0, C(i_1, \ldots, i_q)],
$$
$$
\left\| \frac{\partial^q \phi_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}} \right\|_\infty \leq L_\theta(i_1, \ldots, i_q) \leq L(i_1, \ldots, i_q), \tag{42}
$$

with $C(i_1, \ldots, i_q)$, $L(i_1, \ldots, i_q) < \infty$, independent of $\theta \in \Theta$. This means that in general we need to check the above condition for $2 + 2^2 + 2^3 = 14$ partial derivatives. Let us first focus on the partial derivatives with respect to the range parameter $\beta \in [\beta_{\min}, \beta_{\max}]$. For $r \in [0, 1]$ we write

$$
\phi_{\nu, \kappa}(r) = c_{\nu, \kappa}^{-1} \int_{a(r)}^{b(r)} f_{\nu, \kappa}(r, u) du,
$$

where $[0, 1] \ni r \mapsto a(r) = r$ and $[0, 1] \ni r \mapsto b(r) \equiv 1$ are continuously differentiable on $[0, 1]$. To simplify the notation we will put $f_{\nu, \kappa} := f$. Then, since $f \colon [0, 1] \times [0, 1] \to \mathbb{R}$ is continuous and for any $u \in [0, 1]$, since $\kappa > 2$,

$$
\frac{\partial f}{\partial r}(r, u) = -2r(\kappa - 1)u \left(u^2 - r^2 \right)^{\kappa - 2} (1 - u)^\nu,
$$

exists and is continuous on the rectangle $[0, 1] \times [0, 1]$, we can conclude, using the general Leibniz integral rule, that $[0, 1] \ni r \mapsto \frac{d\phi_{\nu, \kappa}}{dr}(r)$ is continuous and given by

$$
\frac{d\phi_{\nu, \kappa}}{dr}(r) = \underbrace{f(r, b(r)) \frac{db}{dr}(r)}_{=0} - \underbrace{f(r, a(r)) \frac{da}{dr}(r)}_{=0}
$$
$$
- 2r(\kappa - 1) c_{\nu, \kappa}^{-1} \underbrace{\int_{a(r)}^{b(r)} u \left(u^2 - r^2 \right)^{\kappa - 2} (1 - u)^\nu du}_{=c_{\nu, \kappa - 1} \phi_{\nu, \kappa - 1}(r)}.
$$

Hence, for $t \in [0, \beta]$

$$
\frac{\partial \phi_\theta}{\partial \beta}(t) = -\sigma^2 \frac{t}{\beta^2} \frac{d\phi_{\nu, \kappa}}{dr}\left(\frac{t}{\beta} \right) = \frac{2t^2(\kappa - 1)}{\beta^3} \frac{c_{\nu, \kappa - 1}}{c_{\nu, \kappa}} \sigma^2 \phi_{\nu, \kappa - 1}\left(\frac{t}{\beta} \right), \tag{43}
$$

exists and is continuous as a function of t. But clearly, as for $r \in [1, \infty)$ $\phi_{\nu, \kappa}$ is zero, we have that for $t \in [\beta, \infty)$, $\frac{\partial \phi_\theta}{\partial \beta}(t)$ exists as well and is continuous as a function of t. Hence, for any $t \in \mathbb{R}_+$, $\frac{\partial \phi_\theta}{\partial \beta}(t)$ exists, is given by (43) and is continuous as a function of t. Thus, by monotonicity of $\phi_{\nu, \kappa - 1}$ we define

$$
L_\theta(2) := \frac{2\beta^2(\kappa - 1)}{\beta^3} \frac{c_{\nu, \kappa - 1}}{c_{\nu, \kappa}} \sigma^2 = \frac{2(\kappa - 1)}{\beta} \frac{c_{\nu, \kappa - 1}}{c_{\nu, \kappa}} \sigma^2 \text{ and } C_\theta(2) := \beta,
$$

and have that

$$
\frac{\partial \phi_\theta}{\partial \beta} \in \mathcal{C}_{\mathrm{C}}(\mathbb{R}_+; [0, C_\theta(2)]), \left\| \frac{\partial \phi_\theta}{\partial \beta} \right\|_\infty \leq L_\theta(2).
$$

Further, we find

$$\sup_{\theta \in \Theta} L_\theta(2) \leq \underbrace{\frac{2(\kappa - 1)}{\beta_{\min}} \frac{c_{\nu,\kappa-1}}{c_{\nu,\kappa}} \sigma_{\sup}^2}_{=:L(2)} \text{ and } \sup_{\theta \in \Theta} C_\theta(2) \leq \underbrace{\beta_{\max}}_{=:C(2)},$$

where $L(2)$ and $C(2)$ do not depend on $\theta \in \Theta$. Since we have assumed that $\kappa > 4$, we can now repeat the arguments, which led to (43), for another two times, and conclude that for any $t \in \mathbb{R}_+$, $\frac{\partial^2 \phi_\theta}{\partial \beta^2}(t)$, and $\frac{\partial^3 \phi_\theta}{\partial \beta^3}(t)$ exits as well, are given by

$$\frac{\partial^2 \phi_\theta}{\partial \beta^2}(t) = q_1(t), \tag{44}$$

with

$$q_1(t) = \frac{4t^4(\kappa - 1)(\kappa - 2)}{\beta^6} \frac{c_{\nu,\kappa-2}}{c_{\nu,\kappa}} \sigma^2 \phi_{\nu,\kappa-2}\left(\frac{t}{\beta}\right)$$
$$- \frac{6t^2(\kappa - 1)}{\beta^4} \frac{c_{\nu,\kappa-1}}{c_{\nu,\kappa}} \sigma^2 \phi_{\nu,\kappa-1}\left(\frac{t}{\beta}\right)$$

and $\frac{\partial^3 \phi_\theta}{\partial \beta^3}(t) = q_2(t)$, with

$$q_2(t) = \frac{8t^6(\kappa - 1)(\kappa - 2)(\kappa - 3)}{\beta^9} \frac{c_{\nu,\kappa-3}}{c_{\nu,\kappa}} \sigma^2 \phi_{\nu,\kappa-3}\left(\frac{t}{\beta}\right)$$
$$+ \frac{24t^2(\kappa - 1)}{\beta^5} \frac{c_{\nu,\kappa-1}}{c_{\nu,\kappa}} \sigma^2 \phi_{\nu,\kappa-1}\left(\frac{t}{\beta}\right)$$
$$- \frac{36t^4(\kappa - 1)(\kappa - 2)}{\beta^7} \frac{c_{\nu,\kappa-2}}{c_{\nu,\kappa}} \sigma^2 \phi_{\nu,\kappa-2}\left(\frac{t}{\beta}\right),$$

and are both continuous as a function of $t \in \mathbb{R}_+$. Therefore, since $\phi_{\nu,\kappa-1}$, $\phi_{\nu,\kappa-2}$ and $\phi_{\nu,\kappa-3}$ are non-negative and monotonously decreasing, we can define

$$L_\theta(2,2) := \frac{4\beta^4(\kappa - 1)(\kappa - 2)}{\beta^6} \frac{c_{\nu,\kappa-2}}{c_{\nu,\kappa}} \sigma^2,$$

and

$$L_\theta(2,2,2) := \frac{8\beta^6(\kappa - 1)(\kappa - 2)(\kappa - 3)}{\beta^9} \frac{c_{\nu,\kappa-3}}{c_{\nu,\kappa}} \sigma^2 + \frac{24\beta^2(\kappa - 1)}{\beta^5} \frac{c_{\nu,\kappa-1}}{c_{\nu,\kappa}} \sigma^2,$$

as well as $C_\theta(2,2) = C_\theta(2,2,2) := \beta$, and have that

$$\frac{\partial^2 \phi_\theta}{\partial \beta^2} \in \mathcal{C}_{\mathrm{C}}(\mathbb{R}_+; [0, C_\theta(2,2)]), \quad \left\|\frac{\partial^2 \phi_\theta}{\partial \beta^2}\right\|_\infty \leq L_\theta(2,2),$$
$$\frac{\partial^3 \phi_\theta}{\partial \beta^3} \in \mathcal{C}_{\mathrm{C}}(\mathbb{R}_+; [0, C_\theta(2,2,2)]), \quad \left\|\frac{\partial^3 \phi_\theta}{\partial \beta^3}\right\|_\infty \leq L_\theta(2,2,2).$$

Then, we also find

$$\sup_{\theta \in \Theta} L_\theta(2,2) \le \frac{4(\kappa-1)(\kappa-2)}{\beta_{\min}^2} \frac{c_{\nu,\kappa-2}}{c_{\nu,\kappa}} \sigma_{\sup}^2 =: L(2,2),$$

as well as

$$\sup_{\theta \in \Theta} L_\theta(2,2,2) \le \frac{8(\kappa-1)}{\beta_{\min}^3} \frac{\sigma_{\sup}^2}{c_{\nu,\kappa}} ((\kappa-2)(\kappa-3)c_{\nu,\kappa-3} + 3c_{\nu,\kappa-1}) =: L(2,2,2),$$

and $C(2,2) = C(2,2,2) := \beta_{\max} = \sup_{\theta \in \Theta} C_\theta(2,2) = \sup_{\theta \in \Theta} C_\theta(2,2,2)$, where $L(2,2)$, $L(2,2,2)$ and $C(2,2)$ and $C(2,2,2)$ do not depend on $\theta \in \Theta$. This then shows that the partial derivatives of ϕ_θ with respect to the range parameter β exist up to order three and are continuous on $\mathbb{R}_+$ with uniform bounds that do not depend on $\theta \in \Theta$ and compact supports that are subsets of $[0, \beta_{\max}]$. Let us now focus on the partial derivatives with respect to σ^2. We can readily see that for $t \in \mathbb{R}_+$,

$$\frac{\partial \phi_\theta}{\partial \sigma^2}(t) = \phi_{\nu,\kappa}\left(\frac{t}{\beta}\right), \tag{45}$$

and thus with $S_\theta(1) = [0, \beta]$ and $L_\theta(1) = 1$ we can choose $L(1) = 1$ and $C(1) = \beta_{\max}$ such that (41) and (42) are satisfied. Notice that for any $t \in \mathbb{R}_+$, both $\frac{\partial^2 \phi_\theta}{\partial(\sigma^2)^2}(t)$ and $\frac{\partial^3 \phi_\theta}{\partial(\sigma^2)^3}(t)$ are zero. Thus, the existence of the desired constants $L_\theta(2,2)$, $C_\theta(2,2)$ and $L(2,2)$, $C(2,2)$ for $\frac{\partial^2 \phi_\theta}{\partial(\sigma^2)^2}(t)$ and $L_\theta(2,2,2)$, $C_\theta(2,2,2)$ and $L(2,2,2)$, $C(2,2,2)$ for $\frac{\partial^3 \phi_\theta}{\partial(\sigma^2)^3}$, such that (41) and (42) is satisfied, is clear. Let us now consider the mixed partial derivatives. Using (43) and (45), we have

$$\frac{\partial^2 \phi_\theta}{\partial \sigma^2 \partial \beta}(t) = \frac{\partial^2 \phi_\theta}{\partial \beta \partial \sigma^2}(t) = \frac{2t^2(\kappa-1)}{\beta^3} \frac{c_{\nu,\kappa-1}}{c_{\nu,\kappa}} \phi_{\nu,\kappa-1}\left(\frac{t}{\beta}\right),$$

and thus the existence of constants $L_\theta(1,2) = L_\theta(2,1)$, $C_\theta(1,2) = C_\theta(2,1)$ and $L(1,2) = L(2,1)$, $C(1,2) = C(2,1)$ for $\frac{\partial^2 \phi_\theta}{\partial \sigma^2 \partial \beta}(t)$ and $\frac{\partial^2 \phi_\theta}{\partial \beta \partial \sigma^2}(t)$ such that (41) and (42) is satisfied follows with

$$L_\theta(1,2) = \frac{2(\kappa-1)}{\beta} \frac{c_{\nu,\kappa-1}}{c_{\nu,\kappa}}, \quad C_\theta(1,2) = \beta,$$

and

$$L(1,2) = \frac{2(\kappa-1)}{\beta_{\min}} \frac{c_{\nu,\kappa-1}}{c_{\nu,\kappa}}, \quad C(1,2) = \beta_{\max}.$$

Using (43), (44) and (45) we further have that

$$\frac{\partial^3 \phi_\theta}{\partial \sigma^2 \partial \beta^2}(t) = \frac{\partial^3 \phi_\theta}{\partial \beta \partial \sigma^2 \partial \beta}(t) = \frac{\partial^3 \phi_\theta}{\partial \sigma^2 \partial \beta \partial \beta}(t) = \frac{1}{\sigma^2} q_1(t),$$

and thus

$$\frac{\partial^3 \phi_\theta}{\partial \sigma^2 \partial \beta^2} \in \mathcal{C}_{\mathrm{C}}(\mathbb{R}_+; S_\theta(1, 2, 2)), \left\| \frac{\partial^3 \phi_\theta}{\partial \sigma^2 \partial \beta^2} \right\|_\infty \le L_\theta(1, 2, 2),$$

with $S_\theta(1, 2, 2) = S_\theta(2, 2, 1) = S_\theta(2, 1, 2) = [0, \beta]$, and

$$L_\theta(1, 2, 2) = L_\theta(2, 2, 1) = L_\theta(2, 1, 2) = \frac{4(\kappa - 1)(\kappa - 2)}{\beta^2} \frac{c_{\nu, \kappa - 2}}{c_{\nu, \kappa}}.$$

Further, (42) is satisfied with $C(1, 2, 2) = C(2, 2, 1) = C(2, 1, 2) = \beta_{\max}$, and

$$L(1, 2, 2) = L(2, 2, 1) = L(2, 1, 2) = \frac{4(\kappa - 1)(\kappa - 2)}{\beta_{\min}^2} \frac{c_{\nu, \kappa - 2}}{c_{\nu, \kappa}}.$$

Finally we can notice that

$$\frac{\partial^3 \phi_\theta}{\partial \beta \partial (\sigma^2)^2}(t) = \frac{\partial^3 \phi_\theta}{\partial \sigma^2 \partial \beta \partial \sigma^2}(t) = \frac{\partial^3 \phi_\theta}{\partial \sigma^2 \partial \sigma^2 \partial \beta}(t) = 0,$$

and hence we can verify the existence of constants

$$L_\theta(1, 1, 2) = L_\theta(1, 2, 1) = L_\theta(2, 1, 1), \quad C_\theta(1, 1, 2) = C_\theta(1, 2, 1) = C_\theta(2, 1, 1),$$

and

$$L(1, 1, 2) = L(1, 2, 1) = L(2, 1, 1), \quad C(1, 1, 2) = C(1, 2, 1) = C(2, 1, 1),$$

for $\frac{\partial^3 \phi_\theta}{\partial \beta \partial (\sigma^2)^2}$, $\frac{\partial^3 \phi_\theta}{\partial \sigma^2 \partial \beta \partial \sigma^2}$, and $\frac{\partial^3 \phi_\theta}{\partial \sigma^2 \partial \sigma^2 \partial \beta}$, such that (41) and (42) is satisfied. Thus, we have shown that for $\kappa > 4$, $\{\phi_\theta \colon \theta \in \Theta\}$ satisfies (2) of Assumption A.1, where for any $q = 1, 2, 3$, $i_1, \dots, i_q \in \{1, \dots, p\}$, $\mathcal{B}_{\mathrm{C}}(\mathbb{R}_+; S_\theta(i_1, \dots, i_q))$ can be replaced with $\mathcal{C}_{\mathrm{C}}(\mathbb{R}_+; S_\theta(i_1, \dots, i_q))$. It now remains to show that (3) of Assumption A.1 is satisfied. We already know, since $\phi_\theta \in \Phi_d$, that w_θ is continuous and non-negative definite on $\mathbb{R}^d$. We write $L_1(\mathbb{R}_+)$ and $L_1(\mathbb{R}^d)$ for the spaces of Lebesgue integrable functions on $\mathbb{R}_+$ and $\mathbb{R}^d$, respectively. Since $t \mapsto t^{d-1} \phi_\theta(t) \in L_1(\mathbb{R}_+)$ we have that $w_\theta \in L_1(\mathbb{R}^d)$. Thus we can conclude, using for example Theorems 5.26 and 6.18 in [33], that for any $s \in \mathbb{R}^d$, $\widehat{w}_\theta(s) = \mathcal{F}_d \phi_\theta(\|s\|) > 0$, where

$$\mathcal{F}_d \phi_\theta(t) = t^{1 - (d/2)} \int_0^\infty \phi_\theta(u) u^{d/2} J_{(d/2) - 1}(tu) du, \quad t \in \mathbb{R}_+,$$

with $J_{(d/2) - 1}$ the Bessel function of order $(d/2) - 1$. This also shows $s \mapsto \widehat{w}_\theta(s)$ is uniformly continuous on $\mathbb{R}^d$, a member of $L_1(\mathbb{R}^d)$ and Fourier inversion holds (see for example Theorem 1.1 and Corollary 1.26 in [29]). It remains to check that $\Theta \times \mathbb{R}^d \ni (\theta, s) \mapsto \widehat{w}_\theta(s)$ is continuous. In the present case, where $\kappa > 0$ and $\nu \ge (d + 1)/2 + \kappa$, one has actually already established a closed form representation of $\widehat{w}_\theta(s)$. We can refer to Theorem 2.1 in [11] (see also Theorem 1 in [9] for a nice summary and further results) and write for $s \in \mathbb{R}^d \setminus \{0\}$,

$$\frac{\widehat{w}_\theta(s)}{(2\pi)^d \sigma^2 L_\zeta \beta^d} = {}_1F_2\left(\frac{d + 1}{2} + \kappa; \frac{d + 1}{2} + \kappa + \frac{\nu}{2}, \frac{d + 1}{2} + \kappa + \frac{\nu}{2} + \frac{1}{2}; -\frac{(\|s\| \beta)^2}{4} \right),$$

where with $\zeta := (\nu, \kappa, d)$, $L_\zeta = K^\zeta \Gamma(\kappa)/2^{1-\kappa} \mathrm{B}(2\kappa, \nu+1)$, with

$$K^\zeta = \frac{2^{-\kappa-d+1}\pi^{-\frac{d}{2}}\Gamma(\nu+1)\Gamma(2\kappa+d)}{\Gamma\left(\kappa+\frac{d}{2}\right)\Gamma\left(\nu+2\left(\frac{d+1}{2}+\kappa\right)\right)},$$

and for any $z \in \mathbb{R}$,

$${}_1F_2(a; b, c; z) = \sum_{k=0}^{\infty} \frac{(a)_k \, z^k}{(b)_k \, (c)_k \, k!},$$

a special case of the generalized hypergeometric functions ${}_1F_2$ (see also [1]), where for $k \in \mathbb{N}_+$, $(q)_k = \Gamma(q+k)/\Gamma(q)$ denotes the Pochhammer symbol. Note that for $z \in \mathbb{R}$, $|z| \geq 1$ ($z \neq 1$), ${}_1F_2(a; b, c; z)$ is defined via its analytic continuation. Since we know that $s \mapsto \widehat{w}_\theta(s)$ is continuous on the entire $\mathbb{R}^d$ and

$$z \mapsto {}_1F_2\left(\frac{d+1}{2}+\kappa; \frac{d+1}{2}+\kappa+\frac{\nu}{2}, \frac{d+1}{2}+\kappa+\frac{\nu}{2}+\frac{1}{2}; z\right),$$

is continuous in 0, we can further note that

$$\widehat{w}_\theta(0) = (2\pi)^d \sigma^2 L_\zeta \beta^d {}_1F_2\left(\frac{d+1}{2}+\kappa; \frac{d+1}{2}+\kappa+\frac{\nu}{2}, \frac{d+1}{2}+\kappa+\frac{\nu}{2}+\frac{1}{2}; 0\right)$$
$$= (2\pi)^d \sigma^2 L_\zeta \beta^d.$$

This then shows that $\Theta \times \mathbb{R}^d \ni (\theta, s) \mapsto \widehat{w}_\theta(s)$ is continuous as a composition of continuous functions and hence the proposition is proven. $\qquad\square$

Proof of Proposition 6.2. We first show that Assumption 5.1 is satisfied. We write $\theta_1 = (\sigma_1^2, \beta_1)$ and $\theta_2 = (\sigma_2^2, \beta_2)$ and show that $\theta_1 \neq \theta_2$ implies that $\phi_{\theta_1}(\|h\|) \neq \phi_{\theta_2}(\|h\|)$ for all $h \in \mathrm{B}(0; \min\{\beta_1, \beta_2\}) \setminus \{0\}$. Suppose first that $\beta_1 = \beta_2$ but $\sigma_1^2 \neq \sigma_2^2$ we then have that $\phi_{\theta_1}(\|h\|) \neq \phi_{\theta_2}(\|h\|)$ for all $h \in \mathrm{B}(0; \min\{\beta_1, \beta_2\})$, since for any $h \in \mathrm{B}(0; \min\{\beta_1, \beta_2\})$, $\|h\|/\beta_1 = \|h\|/\beta_2 < 1$ and thus $\phi_{\nu,\kappa}(\|h\|/\beta_1) = \phi_{\nu,\kappa}(\|h\|/\beta_2) > 0$. Suppose now that either $\sigma_1^2 \neq \sigma_2^2$, with $\sigma_2^2 < \sigma_1^2$ but $\beta_1 \neq \beta_2$, or $\sigma_1^2 = \sigma_2^2 = \sigma^2$ but $\beta_1 \neq \beta_2$. Then, let us assume that $\min\{\beta_1, \beta_2\} = \beta_2$. We have with $\|h\|/\beta_1 < \|h\|/\beta_2$, by monotonicity of $r \mapsto \phi_{\nu,\kappa}(r)$, that either

$$\phi_{\theta_1}(\|h\|) - \phi_{\theta_2}(\|h\|) = \left(\phi_{\nu,\kappa}\left(\frac{\|h\|}{\beta_1}\right) - \frac{\sigma_2^2}{\sigma_1^2}\phi_{\nu,\kappa}\left(\frac{\|h\|}{\beta_2}\right)\right) > 0$$

for all $h \in \mathrm{B}(0; \min\{\beta_1, \beta_2\}) \setminus \{0\}$ or

$$\phi_{\theta_1}(\|h\|) - \phi_{\theta_2}(\|h\|) = \sigma^2\left(\phi_{\nu,\kappa}\left(\frac{\|h\|}{\beta_1}\right) - \phi_{\nu,\kappa}\left(\frac{\|h\|}{\beta_2}\right)\right) > 0$$

for all $h \in \mathrm{B}(0; \min\{\beta_1, \beta_2\}) \setminus \{0\}$. When $\min\{\beta_1, \beta_2\} = \beta_1$, we will in either of the above cases have $\phi_{\theta_1}(\|h\|) - \phi_{\theta_2}(\|h\|) < 0$ for all $h \in \mathrm{B}(0; \min\{\beta_1, \beta_2\}) \setminus \{0\}$. Further, we can also use a similar argument for the case where either $\sigma_1^2 \neq \sigma_2^2$,

with $\sigma_2^2 > \sigma_1^2$ but $\beta_1 > \beta_2$ or $\beta_1 < \beta_2$, or $\sigma_1^2 = \sigma_2^2 = \sigma^2$ but $\beta_1 > \beta_2$ or $\beta_1 < \beta_2$. Thus we have shown that $\theta_1 \neq \theta_2$ implies that $\phi_{\theta_1}(\|h\|) \neq \phi_{\theta_2}(\|h\|)$ for all $h \in \mathrm{B}\left(0; \min\{\beta_1, \beta_2\}\right) \setminus \{0\}$. Then, for $\tau = 0$, since $\min\{\beta_1, \beta_2\} > 1$, $\mathrm{B}\left(0; \min\{\beta_1, \beta_2\}\right) \setminus \{0\}$ at least contains integers $z \in \left\{p \in \mathbb{Z}^d \colon \|p\| = 1\right\}$. Therefore $\theta_1 \neq \theta_2$ implies $\phi_{\theta_1}(z) \neq \phi_{\theta_2}(z)$ on $\left\{p \in \mathbb{Z}^d \colon \|p\| = 1\right\}$. If $\tau \in (0, 1/2)$, since $\min\{\beta_1, \beta_2\} > 0$, $\mathrm{B}\left(0; \min\{\beta_1, \beta_2\}\right) \cap D_\tau$ has non zero Lebesgue measure. We have thus shown that Assumption 5.1 is satisfied. Let us now show that $\{w_\theta \colon \theta \in \Theta\}$ also satisfies Assumptions 5.2. To do so, fix some interval $I = (0, b] \subset \mathbb{R}_+$, where $1 - 2\tau < b < \beta$. We will show that for any $\theta \in \Theta$, there exists $t_0 \in I$ such that

$$
W\left(\frac{\partial \phi_\theta}{\partial \sigma^2}, \frac{\partial \phi_\theta}{\partial \beta}\right)(t_0) = \det\left(\begin{pmatrix} \frac{\partial \phi_\theta}{\partial \sigma^2}(t_0) & \frac{\partial \phi_\theta}{\partial \beta}(t_0) \\ \frac{d\frac{\partial \phi_\theta}{\partial \sigma^2}}{dt}(t_0) & \frac{d\frac{\partial \phi_\theta}{\partial \beta}}{dt}(t_0) \end{pmatrix}\right) \neq 0, \qquad (46)
$$

where $W\left(\frac{\partial \phi_\theta}{\partial \sigma^2}, \frac{\partial \phi_\theta}{\partial \beta}\right)(t_0)$ is called the Wronskian of $t \mapsto \frac{\partial \phi_\theta}{\partial \sigma^2}(t)$ and $t \mapsto \frac{\partial \phi_\theta}{\partial \beta}(t)$ at $t_0 \in I$. This then shows that the functions $t \mapsto \frac{\partial \phi_\theta}{\partial \sigma^2}(t)$ and $t \mapsto \frac{\partial \phi_\theta}{\partial \beta}(t)$ are linearly independent on the entire interval I, more explicitly, for any $t \in I$,

$$
\alpha_1 \frac{\partial \phi_\theta}{\partial \sigma^2}(t) + \alpha_2 \frac{\partial \phi_\theta}{\partial \beta}(t) = 0
$$

will imply that $\alpha_1 = \alpha_2 = 0$. This then shows that there does not exist $(\alpha_1, \alpha_2) \in \mathbb{R}^2 \setminus \{0\}$, such that for any $\theta \in \Theta$,

$$
h \mapsto \alpha_1 \frac{\partial \phi_\theta}{\partial \sigma^2}(\|h\|) + \alpha_2 \frac{\partial \phi_\theta}{\partial \beta}(\|h\|) = 0,
$$

a.e. with respect to the Lebesgue measure on on $\mathrm{B}\left[0; b\right] \setminus \{0\}$. This then justifies, for both cases, either $\tau = 0$, or $\tau > 0$, that also Assumption 5.2 must be satisfied. Hence, let us show (46). We can calculate, using arguments from the proof of Proposition 6.1, that for $t \in I$,

$$
\frac{\partial \phi_\theta}{\partial \sigma^2}(t) = \phi_{\nu,\kappa}\left(\frac{t}{\beta}\right),
$$
$$
\frac{\partial \phi_\theta}{\partial \beta}(t) = \frac{2t^2(\kappa - 1)}{\beta^3} \frac{c_{\nu,\kappa-1}}{c_{\nu,\kappa}} \sigma^2 \phi_{\nu,\kappa-1}\left(\frac{t}{\beta}\right),
$$
$$
\frac{d\frac{\partial \phi_\theta}{\partial \sigma^2}}{dt}(t) = -\frac{2t(\kappa - 1)}{\beta} \frac{c_{\nu,\kappa-1}}{c_{\nu,\kappa}} \phi_{\nu,\kappa-1}\left(\frac{t}{\beta}\right),
$$

and

$$
\frac{d\frac{\partial \phi_\theta}{\partial \beta}}{dt}(t) = \frac{4t(\kappa - 1)}{\beta^3} \frac{\sigma^2}{c_{\nu,\kappa}}\left(c_{\nu,\kappa-1}\phi_{\nu,\kappa-1}\left(\frac{t}{\beta}\right) - \frac{t^2(\kappa - 2)}{\beta} c_{\nu,\kappa-2}\phi_{\nu,\kappa-2}\left(\frac{t}{\beta}\right)\right).
$$

Therefore we have that for $t \in I$, $W\left(\frac{\partial \phi_\theta}{\partial \sigma^2}, \frac{\partial \phi_\theta}{\partial \beta}\right)(t)$ is given by

$$
\phi_{\nu,\kappa}\left(\frac{t}{\beta}\right) \frac{4t(\kappa-1)}{\beta^3} \frac{\sigma^2}{c_{\nu,\kappa}} \left(c_{\nu,\kappa-1}\phi_{\nu,\kappa-1}\left(\frac{t}{\beta}\right) - \frac{t^2(\kappa-2)}{\beta}c_{\nu,\kappa-2}\phi_{\nu,\kappa-2}\left(\frac{t}{\beta}\right)\right) \\
+ \frac{4t^3(\kappa-1)^2}{\beta^4}\sigma^2\left(\frac{c_{\nu,\kappa-1}}{c_{\nu,\kappa}}\phi_{\nu,\kappa-1}\left(\frac{t}{\beta}\right)\right)^2.
$$

But the latter expression is not equal to zero on the entire I. To see it, assume by contradiction that indeed $W\left(\frac{\partial \phi_\theta}{\partial \sigma^2}, \frac{\partial \phi_\theta}{\partial \beta}\right)(t) = 0$ for all $t \in I$. Using standard algebraic manipulations one can show that this is equivalent to assume that the function

$$
g(t) := \frac{f_1(t)}{f_2(t)},
$$

with

$$
f_1(t) := (\kappa-2)c_{\nu,\kappa}\phi_{\nu,\kappa}\left(\frac{t}{\beta}\right)c_{\nu,\kappa-2}\phi_{\nu,\kappa-2}\left(\frac{t}{\beta}\right) - (\kappa-1)\left(c_{\nu,\kappa-1}\phi_{\nu,\kappa-1}\left(\frac{t}{\beta}\right)\right)^2
$$

and

$$
f_2(t) := c_{\nu,\kappa}c_{\nu,\kappa-1}\phi_{\nu,\kappa}\left(\frac{t}{\beta}\right)\phi_{\nu,\kappa-1}\left(\frac{t}{\beta}\right),
$$

is constant equal to β on I. But this makes no sense and thus we arrive at a contradiction. Hence, there exists $t_0 \in I$ such that (46) is satisfied, which shows that Assumption 5.2 is satisfied and thus concludes the proof of Proposition 6.2. □

Proof of Proposition 6.4. The goal is to check that $\big\{(\mathfrak{T}_{m,\theta}) : \theta \in \Theta\big\}$ satisfies Assumption A.2, then we conclude using Propositions 6.1 and 6.2, as well as Theorems A.1 and A.2. We first notice that for any $\theta \in \Theta$, $m \in \mathbb{N}_+$ and any $q = 1, 2, 3$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, $\mathfrak{T}_{m,\theta}$ and $\frac{\partial^q \mathfrak{T}_{m,\theta}}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}$ are Borel measurable functions on $\mathbb{R}_+$. In addition, for any $\theta \in \Theta$ and $m \in \mathbb{N}_+$, $\mathfrak{T}_{m,\theta}$ has support $[0, U_{\theta,m}]$, with $U_{\theta,m} = \min\{C_m, \beta\}$ that satisfies $\sup_{m\in\mathbb{N}_+}\sup_{\theta\in\Theta} U_{\theta,m} = \beta_{\max}$. Further, one can verify that the family $\big\{(\mathfrak{T}_{m,\theta}) : \theta \in \Theta\big\}$ is also uniformly bounded by $\sigma^2_{\max}$ on $\mathbb{R}_+$ and it converges uniformly to ϕ_θ on $\mathbb{R}_+$, independent of $\theta \in \Theta$, that is $\sup_{\theta\in\Theta}\|\mathfrak{T}_{m,\theta} - \phi_\theta\|_\infty \xrightarrow{m\to\infty} 0$. Thus (1), (2) and (3) of Assumption A.2 are satisfied. To verify the remaining assumptions, we view $\mathfrak{T}_{m,\theta}(t)$ as the result of a truncation operator $g \mapsto \mathfrak{T}_m(g) = g\mathbb{1}_{[0,C_m]}$ evaluated at t. That is, $\mathfrak{T}_{m,\theta}(t) = \mathfrak{T}_m(\phi_\theta)(t)$. Then, we remark that for any $q = 1, 2, 3$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, for any $\theta \in \Theta$,

$$
\frac{\partial^q \mathfrak{T}_{m,\theta}}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}(t) = \mathfrak{T}_m\left(\frac{\partial^q \phi_\theta}{\partial \theta_{i_1} \cdots \partial \theta_{i_q}}\right)(t).
$$

Thus, by Proposition 6.1, also (4) and (5) of Assumption A.2 are satisfied. □

Proof of Proposition 6.5. For any $m \in \mathbb{N}_+$, for any $\theta \in \Theta$, $B_{m,\theta}(t; b_m)$ is continuous on $[0, M]$ and it is also continuous as a function of $\theta \in \Theta$ (see also the proof of Proposition 6.1). Further, it converges uniformly to ϕ_θ on $[0, M]$, independent of $\theta \in \Theta$. That is

$$\sup_{\theta \in \Theta} \sup_{t \in [0,M]} |\mathfrak{P}_{m,\theta}(t) - \phi_\theta(t)| \xrightarrow{m \to \infty} 0.$$

To see this we can rely, for example, on the proof of Theorem 2.3.1 in [24]. There, it is shown that for any $t \in [0, M]$ and $\theta \in \Theta$, for any $\varepsilon > 0$, there exists $\delta(t) > 0$ such that

$$|B_{m,\theta}(t; b_m) - \phi_\theta(t)| \leq \varepsilon + 2\sigma_{\max}^2 \frac{b_m t}{m\delta(t)^2},$$

for m large enough. Since $[0, M]$ is compact and ϕ_θ is continuous, we can choose $\delta^* \equiv \delta(t)$, independent of $t \in [0, M]$ and $\theta \in \Theta$, such that the above inequality is satisfied for arbitrary $\varepsilon > 0$, with $\delta(t)$ replaced with δ^*. Then, we conclude by taking the supremum on the left and right over $[0, M]$ and $\theta \in \Theta$. For any $m \in \mathbb{N}_+$, for any $\theta \in \Theta$, we can write

$$|\mathfrak{P}_{m,\theta}(t) - \phi_\theta(t)| = |\mathfrak{P}_{m,\theta}(t) - \phi_\theta(t)| \, \mathbb{1}_{[0,M]}(t) + |\mathfrak{P}_{m,\theta}(t) - \phi_\theta(t)| \, \mathbb{1}_{(M,\infty)}(t).$$

Notice that because $M \geq \beta_{\max}$, the latter term is actually zero independent of $\theta \in \Theta$ and thus we have that $(\mathfrak{P}_{m,\theta})_{m \in \mathbb{N}_+}$ converges uniformly to ϕ_θ on the entire $\mathbb{R}_+$, independent of $\theta \in \Theta$. Thus, we have that $\sup_{\theta \in \Theta} \|\mathfrak{P}_{m,\theta} - \phi_\theta\|_\infty \xrightarrow{m \to \infty} 0$. Note also that the convergence (in the uniform norm) of $\mathfrak{P}_{m,\theta}$ to ϕ_θ in particular implies that the sequence of functions $(\mathfrak{P}_{m,\theta})_{m \in \mathbb{N}_+}$ is bounded on $\mathbb{R}_+$ for any $\theta \in \Theta$. Therefore we can use that

$$\sup_{m \in \mathbb{N}_+} \sup_{t \in \mathbb{R}_+} \sup_{\theta \in \Theta} \mathfrak{P}_{m,\theta}(t) = \sup_{m \in \mathbb{N}_+} \sup_{t \in [0,M]} \sup_{\theta \in \Theta} \mathfrak{P}_{m,\theta}(t),$$

to find $\widetilde{C} := M$ and $\widetilde{L} := \sup_{m \in \mathbb{N}_+} \sup_{t \in [0,M]} \sup_{\theta \in \Theta} \mathfrak{P}_{m,\theta}(t)$, two constants, which are independent of $m \in \mathbb{N}_+$ and $\theta \in \Theta$ (recall that Θ is compact), such that (2) of Assumption A.2 is satisfied. Clearly, for any $\theta \in \Theta$ and for any $m \in \mathbb{N}_+$, the function $\mathfrak{P}_{m,\theta} \colon (\mathbb{R}_+, \mathfrak{B}(\mathbb{R}_+)) \to (\mathbb{R}, \mathfrak{B}(\mathbb{R}))$ is measurable. In conclusion we have shown that (1), (2) and (3) of Assumption A.2 are satisfied. In the proof of Proposition 6.1 we have shown that for any $\theta \in \Theta$, for any $q = 1, 2, 3$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, there exist constants $C_\theta(i_1, \ldots, i_q)$, $L_\theta(i_1, \ldots, i_q) < \infty$, such that

$$\frac{\partial^q \phi_\theta}{\partial\theta_{i_1} \cdots \partial\theta_{i_q}} \in \mathcal{C}_{\mathrm{C}}(\mathbb{R}_+; [0, C_\theta(i_1, \ldots, i_q)]), \quad \left\|\frac{\partial^q \phi_\theta}{\partial\theta_{i_1} \cdots \partial\theta_{i_q}}\right\|_\infty \leq L_\theta(i_1, \ldots, i_q),$$

where for any $q = 1, 2, 3$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, $\sup_{\theta \in \Theta} C_\theta(i_1, \ldots, i_q) \leq \beta_{\max}$. In addition, we notice that for any $q = 1, 2, 3$, $i_1, \ldots, i_q \in \{1, \ldots, p\}$, for any $\theta \in \Theta$,

$$\frac{\partial^q \mathfrak{P}_{m,\theta}}{\partial\theta_{i_1} \cdots \partial\theta_{i_q}}(t) = \mathfrak{P}_m\left(\frac{\partial^q \phi_\theta}{\partial\theta_{i_1} \cdots \partial\theta_{i_q}}\right)(t),$$

where $g \mapsto \mathfrak{P}_m(g)$ is the Bernstein polynomial operator for a function g with support included $[0, M]$:

$$\mathfrak{P}_m(g)(t) = \sum_{k=0}^{m} g\left(b_m \frac{k}{m}\right)\binom{m}{k}\left(\frac{t}{b_m}\right)^k\left(1 - \frac{t}{b_m}\right)^{m-k},$$

for $t \leq M$ and zero otherwise. Therefore, we can rely on the same arguments that we have used to show that (2) and (3) of Assumption A.2 are satisfied, to show that also (4) and (5) of Assumption A.2 must be satisfied. This then concludes the proof of Proposition 6.5. $\qquad\square$

Proof of Proposition 6.7. The proof follows the same reasoning as the proof of Proposition 6.5. $\qquad\square$

Proof of Proposition 6.9. Since $(\delta(m))_{m \in \mathbb{N}_+}$ does not depend on $\theta \in \Theta$ and $t \in \mathbb{R}_+$, and is such that $\delta(m) \to 0$, as $m \to \infty$ we can see that $\{(\mathfrak{S}_{m,\theta}) : \theta \in \Theta\}$ satisfies Assumption A.2. Thus, using Propositions 6.1 and 6.2, under application of Theorems A.1 and A.2, the proposition is proven. $\qquad\square$

C.6. *Proof of results in Section 7*

Proof of Theorem 7.1. Given a collection $S_{(n)}$ of S, let $K_{n,\theta_{i,j}} = k_\theta(S_i - S_j)$, $1 \leq i, j \leq n$, denote the $n \times n$ covariance matrix based on the family $\{k_\theta : \theta \in \Theta\}$. We first note that under the given assumptions on the family $\{k_\theta : \theta \in \Theta\}$, we have that

$$\sup_{n \in \mathbb{N}_+} \sup_{\theta \in \Theta} \left\|K_{n,\theta}\right\|_2 < \infty \quad \text{and} \quad \inf_{n \in \mathbb{N}_+} \inf_{\theta \in \Theta} \lambda_n(K_{n,\theta}) > 0,$$

with $\mathbb{P}$ probability one. This can be seen from Proposition D.4 and Lemma D.5 in [4]. Using this, the proof of (15) is immediate, it follows from Lemmas 4.1 and B.6.

If we proof

$$d_{n,\hat{\theta}_n(kt)} = \inf_{\theta \in \Theta} d_{n,\theta} + \delta_n', \quad \text{as } n \to \infty, \tag{47}$$

where $\delta_n' \xrightarrow[n \to \infty]{\mathbb{P}} 0$, (16) follows from (15), and we are done. We note that (47) is established if we prove

$$\sup_{\theta \in \Theta} \left|l_{n,\text{t-ML}}(\theta) - \mathbb{E}\left[l_{n,\text{t-ML}}(\theta) \mid S_{(n)}\right]\right| \xrightarrow[n \to \infty]{\mathbb{P}} 0, \tag{48}$$

where

$$l_{n,\text{t-ML}}(\theta) := \frac{1}{n}\log\left(\det\left(R_{n,\theta}\right)\right) + \frac{1}{n}\langle Z_{(n)}, R_{n,\theta}^{-1} Z_{(n)}\rangle,$$

the random version of the modified log-likelihood function based on the tapered covariance function. This is seen from the proof of Theorem 3.3 in [5]. But under the given assumptions, the family $\{k_\theta t_{\beta_0} : \theta \in \Theta\}$ satisfies Assumption 3.1 (regarding (2), up to $q = 1$ and the continuity of first order partial derivatives). Thus (48) can be shown as it was shown (see (39)) in the proof of Theorem 5.2. $\qquad\square$

Acknowledgments

Funding

References

[1] ABRAMOWITZ, M. and STEGUN, I. A., eds. (1965). *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables. National Bureau of Standards Applied Mathematics Series, No. 55.* U.S. Government Printing Office, Washington, D.C. Superintendent of Documents. MR0177136

[2] ANDERES, E. (2010). On the consistent separation of scale and variance for Gaussian random fields. *The Annals of Statistics* **38** 870–893. https://doi.org/10.1214/09-AOS725. MR2604700

[3] ASKEY, R. (1973). Radial characteristic functions. Technical report, Research Center, Univ. Wisconsin-Madison, Madison, WI.

[4] BACHOC, F. (2014). Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes. *Journal of Multivariate Analysis* **125** 1–35. https://doi.org/10.1016/j.jmva.2013.11.015. MR3163828

[5] BACHOC, F. (2018). Asymptotic analysis of covariance parameter estimation for Gaussian processes in the misspecified case. *Bernoulli* **24** 1531–1575. https://doi.org/10.3150/16-BEJ906. MR3706801

[6] BACHOC, F. (2021). Asymptotic analysis of maximum likelihood estimation of covariance parameters for Gaussian processes: an introduction with proofs. In *Advances in Contemporary Statistics and Econometrics* (A. Daouia and A. Ruiz-Gazen, eds.) 283–303. Springer, Cham. MR4299285

[7] BACHOC, F., BETANCOURT, J., FURRER, R. and KLEIN, T. (2020). Asymptotic properties of the maximum likelihood and cross validation estimators for transformed Gaussian processes. *Electronic Journal of Statistics* **14** 1962–2008. https://doi.org/10.1214/20-EJS1712. MR4091860

[8] BACHOC, F. and FURRER, R. (2016). On the smallest eigenvalues of covariance matrices of multivariate spatial processes. *Stat* **5** 102–107. https://doi.org/10.1002/sta4.107. MR3478800

[9] BEVILACQUA, M., FAOUZI, T., FURRER, R. and PORCU, E. (2019). Estimation and prediction using generalized Wendland covariance functions under fixed domain asymptotics. *The Annals of Statistics* **47** 828–856. https://doi.org/10.1214/17-AOS1652. MR3909952

[10] Blasi, F., Caamaño Carrillo, C., Bevilacqua, M. and Furrer, R. (2022). A selective view of climatological data and likelihood estimation. *Spatial Statistics* **50** Paper No. 100596. https://doi.org/10.1016/j.spasta.2022.100596. MR4439327

[11] Chernih, A. and Hubbert, S. (2014). Closed form representations and properties of the generalised Wendland functions. *Journal of Approximation Theory* **177** 17–33. https://doi.org/10.1016/j.jat.2013.09.005. MR3128401

[12] Chlodovsky, I. (1937). Sur le développement des fonctions définies dans un intervalle infini en séries de polynomes de M. S. Bernstein. *Compositio Mathematica* **4** 380–393. MR1556982

[13] Cressie, N. A. C. (1993). *Statistics for Spatial Data*. John Wiley & Sons, Inc., New York. Reprint, A Wiley-Interscience Publication. https://doi.org/10.1002/9781119115151. MR1239641

[14] Du, J., Zhang, H. and Mandrekar, V. S. (2009). Fixed-domain asymptotic properties of tapered maximum likelihood estimators. *The Annals of Statistics* **37** 3330–3361. https://doi.org/10.1214/08-AOS676. MR2549562

[15] Flury, R. and Furrer, R. (2021). Discussion on competition for Spatial Statistics for large datasets. *Journal of Agricultural, Biological, and Environmental Statistics* **26** 599–603. MR4330397

[16] Furrer, R., Bachoc, F. and Du, J. (2016). Asymptotic properties of multivariate tapering for estimation and prediction. *Journal of Multivariate Analysis* **149** 177–191. https://doi.org/10.1016/j.jmva.2016.04.006. MR3507322

[17] Furrer, R., Genton, M. G. and Nychka, D. (2006). Covariance tapering for interpolation of large spatial datasets. *Journal of Computational and Graphical Statistics* **15** 502–523. https://doi.org/10.1198/106186006X132178. MR2291261

[18] Gaspari, G. and Cohn, S. E. (1999). Construction of correlation functions in two and three dimensions. *Quarterly Journal of the Royal Meteorological Society* **125** 723-757. https://doi.org/10.1002/qj.49712555417

[19] Gerber, F., Mösinger, K. and Furrer, R. (2017). Extending R packages to support 64-bit compiled code: An illustration with spam64 and GIMMS NDVI$_{3g}$ data. *Computers & Geosciences* **104** 109–119. https://doi.org/10.1016/j.cageo.2016.11.015

[20] Gneiting, T. (1999). Correlation functions for atmospheric data analysis. *Quarterly Journal of the Royal Meteorological Society* **125** 2449-2464. https://doi.org/10.1002/qj.49712555906

[21] Gneiting, T. (2002). Compactly supported correlation functions. *Journal of Multivariate Analysis* **83** 493–508. https://doi.org/10.1006/jmva.2001.2056. MR1945966

[22] Heaton, M. J., Datta, A., Finley, A. O. and et al. (2019). A case study competition among methods for analyzing large spatial data. *Journal of Agricultural, Biological, and Environmental Statistics* **24** 398–425. https://doi.org/10.1007/s13253-018-00348-w. MR3996451

[23] KAUFMAN, C. G., SCHERVISH, M. J. and NYCHKA, D. W. (2008). Covariance tapering for likelihood-based estimation in large spatial data sets. *Journal of the American Statistical Association* **103** 1545–1555. https://doi.org/10.1198/016214508000000959. MR2504203

[24] LORENTZ, G. G. (1953). *Bernstein Polynomials. Mathematical Expositions, No. 8.* University of Toronto Press, Toronto. MR0057370

[25] MARDIA, K. V. and MARSHALL, R. J. (1984). Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika* **71** 135–146. https://doi.org/10.1093/biomet/71.1.135. MR738334

[26] MATÉRN, B. (1960). *Spatial Variation: Stochastic Models and their Application to some Problems in Forest Surveys and other Sampling Investigations.* Meddelanden Fran Statens Skogsforskningsinstitut, Band 49, Nr. 5, Stockholm. MR0169346

[27] NEWEY, W. K. (1991). Uniform convergence in probability and stochastic equicontinuity. *Econometrica. Journal of the Econometric Society* **59** 1161–1167. https://doi.org/10.2307/2938179. MR1113551

[28] SCHABACK, R. (2011). The missing Wendland functions. *Advances in Computational Mathematics* **34** 67–81. https://doi.org/10.1007/s10444-009-9142-7. MR2783302

[29] STEIN, E. M. and WEISS, G. (1971). *Introduction to Fourier Analysis on Euclidean Spaces. Princeton Mathematical Series, No. 32.* Princeton University Press, Princeton, N.J. MR0304972

[30] STEIN, M. L. (2013). Statistical properties of covariance tapers. *Journal of Computational and Graphical Statistics* **22** 866–885. https://doi.org/10.1080/10618600.2012.719844. MR3173747

[31] VAN DER VAART, A. W. (1998). *Asymptotic Statistics. Cambridge Series in Statistical and Probabilistic Mathematics* **3**. Cambridge University Press, Cambridge. https://doi.org/10.1017/CBO9780511802256. MR1652247

[32] WENDLAND, H. (1995). Piecewise polynomial, positive definite and compactly supported radial functions of minimal degree. *Advances in Computational Mathematics* **4** 389–396. https://doi.org/10.1007/BF02123482. MR1366510

[33] WENDLAND, H. (2005). *Scattered Data Approximation. Cambridge Monographs on Applied and Computational Mathematics* **17**. Cambridge University Press, Cambridge. MR2131724

[34] YING, Z. (1991). Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *Journal of Multivariate Analysis* **36** 280–296. https://doi.org/10.1016/0047-259X(91)90062-7. MR1096671

[35] ZASTAVNYI, V. P. (2006). On some properties of the Buhmann functions. *Ukrainian Mathematical Journal* **58** 1045–1067. https://doi.org/10.1007/s11253-006-0128-z. MR2345078

[36] ZHANG, H. (2004). Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association* **99** 250–261. https://doi.org/10.1198/016214504000000241. MR2054303

[37] Zhang, H. and Zimmerman, D. L. (2005). Towards reconciling two asymptotic frameworks in spatial statistics. *Biometrika* **92** 921–936. https://doi.org/10.1093/biomet/92.4.921. MR2234195