SMOOTHNESS ESTIMATION FOR WHITTLE-MATÉRN PROCESSES ON CLOSED RIEMANNIAN MANIFOLDS

MORITZ KORTE-STAPFF[1], TONI KARVONEN[2], AND ÉRIC MOULINES[3]

ABSTRACT. The family of Matérn kernels are often used in spatial statistics, function approximation and Gaussian process methods in machine learning. One reason for their popularity is the presence of a smoothness parameter that controls, for example, optimal error bounds for kriging and posterior contraction rates in Gaussian process regression. On closed Riemannian manifolds, we show that the smoothness parameter can be consistently estimated from the maximizer(s) of the Gaussian likelihood when the underlying data are from point evaluations of a Gaussian process and, perhaps surprisingly, even when the data comprise evaluations of a non-Gaussian process. The points at which the process is observed need not have any particular spatial structure beyond quasi-uniformity. Our methods are based on results from approximation theory for the Sobolev scale of Hilbert spaces. Moreover, we generalize a well-known equivalence of measures phenomenon related to Matérn kernels to the non-Gaussian case by using Kakutani's theorem.

CONTENTS

[1]DEPARTMENT OF STATISTICS, UNIVERSITY OF MICHIGAN
[2]DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF HELSINKI
[3]CMAP, ÉCOLE POLYTECHNIQUE
E-mail addresses: [1]kortestumich.edu, [2]toni.karvonenhelsinki.fi,
[3]eric.moulinespolytechnique.edu.

1. Introduction

When modelling the covariance structure of spatial or spatio-temporal observations, one of the most common choices is to use parametric models based on Matérn kernels. In the simplest isotropic form, they give the covariance between the observations $u(x_1)$ and $u(x_2)$, indexed by sampling sites $x_1, x_2 \in \mathbb{R}^d$, as

$$(1) \quad \mathbb{C}\mathrm{ov}(u(x_1), u(x_2)) = \sigma^2 \frac{2^{1-(s-d/2)}}{\Gamma(s-d/2)} \sqrt{\tau}^{s-d/2} \|x_1 - x_2\|^{s-d/2} \mathcal{K}_{s-d/2}\left(\sqrt{\tau}\|x_1 - x_2\|\right),$$

where $s > d/2$ and $\tau, \sigma^2 > 0$ are covariance parameters, $\|\cdot\|$ denotes the Euclidean norm on $\mathbb{R}^d$ and $\mathcal{K}_s$ is the modified Bessel function of the second kind. The Matérn covariance functions are attributed to Matérn (1960) and are ubiquitously used in spatial statistics. In addition, they are popular as priors for Bayesian Gaussian process regression and as radial basis functions for scattered data interpolation. For a review of Matérn kernels and its application in a varitey of fields such as machine learning and function approximation, we refer to the recent review paper by Porcu et al. (2023). An important characterization was proved by Whittle (1963). Let $\mathcal{W}$ be spatial Gaussian white noise and $v(s, \tau)$ a suitable normalization function. Whittle showed that the stationary solution u of the stochastic partial differential equation (SPDE)

$$(2) \qquad\qquad (\tau - \Delta)^{s/2} u = \sigma \sqrt{v(s, \tau)}\, \mathcal{W},$$

on $\mathbb{R}^d$ is a Gaussian process with covariance (1). Here Δ is the Laplacian and $\mathcal{W}$ is Gaussian white noise. This characterization was used in Lindgren et al. (2011), among others, to define Whittle-Matérn processes on Riemannian manifolds by replacing the Euclidean Laplace operator in (2) by the Laplace–Beltrami operator of the respective

Riemannian manifold. Whittle-Matérn Gaussian processes on Riemannian manifolds were further investigated, for example, in Borovitskiy et al. (2020); Lang and Schwab (2015); Li et al. (2023).

Stein strongly advocates the use of Matérn covariance functions (Stein, 1999, p. 14: "Use the Matérn."), mainly because the parameter s controls the local behaviour of corresponding Gaussian processes. More precisely, a Gaussian process with the covariance structure (1) is $\lceil s - d/2 \rceil - 1$ times mean square differentiable (Stein, 1999, p. 31). This is related to the fact that Matérn covariance functions are reproducing kernels for the scale of Sobolev spaces $H^s(\mathbb{R}^d)$. In fact, for many tasks for which Matérn Kernels are used, the smoothness parameter s is the *only* parameter that matters asymptotically. For example, it is shown in van der Vaart and van Zanten (2011) that for Bayesian Gaussian process regression optimal contraction rates can only be attained if the smoothness parameter of the Gaussian process prior matches the smoothness of the unknown regression function. This is similar to results in, for example, Wang et al. (2020) and Narcowich et al. (2006), where optimal approximation and kriging error rates depend only on the smoothness of the kernel and the function to be approximated. Moreover, estimation of smoothness parameters is necessary to guarantee the reliability of uncertainty quantification provided by non-parametric methods, such as Gaussian process regression (Bull, 2012; Giné and Nickl, 2010; Hadji and Szabó, 2021; Szabó et al., 2015).

Despite the ubiquity of (Whittle-)Matérn kernels in spatial statistics, Gaussian process regression and function approximation, and the importance of the smoothness parameter s, the consistency of the maximum likelihood estimator of s has been an open problem in the *infill* (or *fixed-domain*) asymptotic framework (Zhang and Zimmerman, 2005), i.e., when the sampling sites become increasingly dense in a bounded domain over which the process is observed. This is the case even when the observations do stem from a correctly specified Gaussian process and is in contrast to the case of *increasing-domain* asymptotics, where it has long been known that maximum likelihood estimators of all parameters are consistent and asymptotically normal in the well-specified case (Mardia and Marshall, 1984). In addition to the comparative lack of positive results, it is known for Gaussian processes that consistent estimation of the parameters τ and σ^2 in (1) or (2) is *not* possible if $d < 4$. In this case only the so-called microergodic parameter $\tau^{s-d/2}\sigma^2$ can be consistently estimated (Bolin and Kirchner, 2023; Li et al., 2023; Zhang, 2004). This is due to equivalence of the involved Gaussian measures. In dimension 5 or higher, consistent estimators are given in Anderes (2010). Due to the difficulty of establishing theoretic guarantees for estimation

of the smoothness parameter, there is a large amount of literature on estimation of the microergodic parameter under the seldomly satisfied assumption that the correct smoothness is known (e.g., Bachoc, 2014; Du et al., 2009; Kaufman and Shaby, 2013; Li, 2022; Stein, 1999; Wang and Loh, 2011; Zhang, 2004).

Recently, however, progress has been made in estimating the smoothness parameter s in the context of infill asymptotics. A consistent estimator for s based on observations of a Gaussian process with covariance function (1) on $[0, 1)^d$ for $d = 1, 2, 3$ was presented in Loh et al. (2021). This work was extended in Loh and Sun (2023) to take into account, for example, noisy observations of the process. On manifolds, consistency of the maximum likelihood estimator for the smoothness parameter on the torus and the circle was investigated by Chen et al. (2021) and Petit (2023) under somewhat restrictive assumptions. The estimation of Sobolev regularity of deterministic functions observed over bounded Euclidean domains using Gaussian likelihoods was studied in Karvonen (2023).

We focus on the maximum likelihood estimation for stochastic processes with Whittle–Matérn covariance structure observed on closed Riemannian manifolds. Such manifolds are necessarily compact and therefore we work in the infill asymptotic framework. Our main contributions are as follows:

(1) THEOREM 4.8 and COROLLARY 4.10: We show that maximizers of the Gaussian likelihood for the smoothness parameter s are consistent estimators under the assumption that the points at which the process is observed are *quasi-uniform*. This continues to hold for observations from a variety of non-Gaussian processes and thus justifies the use of Gaussian maximum likelihoods for smoothness estimation even when there is no reason to assume that the underlying processes is Gaussian.

(2) THEOREM 5.1: We show that the phenomenon of equivalence of measures and consequently the *inconsistent* separation of range scale (τ) and variance (σ^2) parameters is not restricted to Gaussian processes. Specifically, we show that for a variety of other non-Gaussian stochastic processes the same phenomenon of equivalence of measures occurs as in the Gaussian case.

Let us describe Theorem 4.8 and Corollary 4.10 in more detail. Let $\{e_i\}$ be eigenfunctions of the negative Laplacian $-\Delta$ on a closed Riemannian manifold and $\{\lambda_i\}$ the corresponding eigenvalues. We observe a stochastic process u at pairwise distinct points $(x_1, \ldots, x_n) = \boldsymbol{x}$ on the manifold. The process, which has positive magnitude

and range scale parameters σ_0 and τ_0 and smoothness $s_0 > d/2$, is defined by

$$u(x) = \sigma_0\sqrt{v(s_0,\tau_0)}\sum_{i=1}^{\infty}(\tau_0 + \lambda_i)^{-s_0/2}\xi_i e_i(x),$$

where $\{\xi_i\}$ are i.i.d. zero-mean random variables with unit variance and $v(s_0,\tau_0)$ is a suitable normalization constant. Note that the random variables $\{\xi_i\}$ need not be Gaussian. We allow the likelihood to be misspecified in two ways. First, we do not assume the family of candidate covariance functions contains the true covariance of u. We consider any suitable family of kernels $\{K_s\}_{s>d/2}$ such that the reproducing kernel Hilbert space of K_s is norm-equivalent to a Sobolev space of order s. Second, in the place of the true likelihood, which may be non-Gaussian, we use the Gaussian log-likelihood

$$\ell(s; u(\boldsymbol{x})) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\left(\det(K_s(\boldsymbol{x})) - \frac{1}{2}u(\boldsymbol{x})^{\top}K_s(\boldsymbol{x})^{-1}u(\boldsymbol{x}),\right.$$

where $K_s(\boldsymbol{x})$ is the $n \times n$ covariance matrix with entries $(K_s(\boldsymbol{x}))_{ij} = K_s(x_i, x_j)$ and $u(\boldsymbol{x}) = (u(x_1), \ldots, u(x_n)) \in \mathbb{R}^n$. Theorem 4.8 and Corollary 4.10 can then be formulated somewhat informally as follows.

Theorem. *Assume that the sequence of points $\{x_i\}$ is quasi-uniform and that the family of kernels $\{K_s\}_{s>d/2}$ has a certain norm monotonicity property. Let $\{\hat{s}_n\}$ be a sequence of maximizers of the Gaussian log-likelihood $\ell(s; u(\boldsymbol{x}))$ on a bounded interval containing s_0.*

(1) If $s_0 > d$, then $\hat{s}_n \to s_0$ in probability.
(2) If $s_0 > d/2$ and $\{\xi_i\}$ are Gaussian, then $\hat{s}_n \to s_0$ almost surely.

These results are of practical importance, as we show that the smoothness parameter can be consistently estimated by using *any* family of kernels whose reproducing kernel Hilbert spaces are the Sobolev scale of Hilbert spaces. This enables the estimation of smoothness by using, for example, the generalized Wendland class of compactly supported kernels that are computationally cheaper than Matérn kernels. It also shows that the parameters σ^2 and τ don't need to be known to estimate the smoothness. Moreover, in Section 4.5 we establish an interesting connection between the parameters σ^2 and s, namely that the estimation of σ^2 compensates to some extent for misspecification of smoothness. Again, these results do not require any Gaussianity assumptions. This is similar to the results in Karvonen (2021); Karvonen et al. (2020); and Naslidnyk et al. (2023). Although our main focus is on stochastic processes on Riemannian manifolds, we provide essentially the same results for processes observed on bounded Euclidean

domains based on Dirichlet Laplacian operators (Bolin and Kirchner, 2023; Lindgren et al., 2011).

All of the above mentioned work on estimating smoothness assumes that the sampling sites are sufficiently uniform in some sense. We assume that the sampling sites are quasi-uniform, an assumption that is frequently used in the function approximation literature and therefore allows us to exploit the wealth of results obtained there. Although it is a rather strict assumption, quasi-uniformity is more general than the uniformity assumptions used in Chen et al. (2021) and Petit (2023). As far as we know, no consistency results have yet been obtained under this assumption in the context of estimating covariance parameters for stochastic processes. However, since the samples of a stochastic process with a Whittle–Matérn covariance kernel are homogeneous in the sense that their smoothness does not vary spatially, *any* sequence of sampling sites should suffice. Indeed, many results for estimating the microergodic parameter do not include assumptions on the sampling sites (e.g., Kaufman and Shaby, 2013; Li, 2022; Ying, 1991; Zhang, 2004). Refined techniques are needed to avoid uniformity assumptions when estimating smoothness.

The rest of the work is structured as follows. In Section 2 we introduce the manifold setting in which we work and give some functional analytical background information. In Section 3 we introduce our model and provide some regularity results. Our main results are presented in Section 4. Our techniques are similar to those used in Karvonen (2023) and build on known results from the literature on function approximation. Furthermore, we provide additional results for the Gaussian case, discuss the estimation with other kernels and briefly the estimation of the microergodic parameters. In Section 5 we state our results concerning the equivalence of measures. The main tool is a theorem of Kakutani (Kakutani, 1948) on the equivalence of infinite product measures. In Section 6 we present a simple translation of our results for processes on bounded Euclidean domains based on Dirichlet Laplacian operators. Finally, in Section 7 we present a simulation study assessing our theoretical results regarding the stability of smoothness estimation against model misspecification and evaluating finite sample properties.

2. Notation and Background

We equip the Euclidean space $\mathbb{R}^k$ with the Euclidean norm $\|\cdot\|$. All other norms are specifically defined. For two sequences a_n and b_n we write $a_n \lesssim b_n$ if $a_n \leq C b_n$ for a positive constant C. If $a_n \lesssim b_n$ and $b_n \lesssim a_n$, then $a_n \asymp b_n$. Finally, we generally use c and C for generic positive constants that occur in lower and upper bounds respectively. These constants may differ from one line to the next.

2.1. **Closed Manifolds and the Laplace–Beltrami Operator.** We begin with a brief introduction to our manifold setting. See Lee (2013, 2018) for more details. Let $\mathcal{M} \subset \mathbb{R}^k$ be a connected and closed (i.e. compact and without boundary) embedded d-dimensional Riemannian submanifold in the sense of (Lee, 2013, Chapter 5), where the Riemannian metric on $\mathcal{M}$ is given by the restriction $g|_{\mathcal{M}}$ of the standard Riemannian metric g on $\mathbb{R}^k$. Note that any smooth Riemannian manifold can be isometrically embedded into $\mathbb{R}^k$ for k large enough by the Nash embedding theorem. Let $d_{\mathcal{M}}$ be the distance function on $\mathcal{M}$ based on g. Then $(\mathcal{M}, d_{\mathcal{M}})$ is a complete metric space. Let ∇ denote the Levi-Civita connection. The exponential map $\exp_p^{\mathcal{M}} : \mathcal{E} \subset T_p\mathcal{M} \to \mathcal{M}$ is given by

$$\exp_p^{\mathcal{M}}(v) = \gamma_v(1),$$

where γ_v is the unique geodesic with $\gamma(0) = p$ and $\gamma'(0) = v$ and $T_p\mathcal{M}$ is the tangent space to p. Since $\mathcal{M}$ is compact, it has a finite but positive injectivity radius $r_0 > 0$, which means that whenever $r \leq r_0$ the exponential map $\exp_p^{\mathcal{M}}$ is a diffeomorphism on $B_r(\mathbf{0}) = \{v \in \mathbb{R}^d \mid \|v\| < r\}$ for all $p \in \mathcal{M}$ (where we have identified $T_p\mathcal{M}$ with $\mathbb{R}^d$) with image $\exp_p^{\mathcal{M}}(B_r(\mathbf{0})) = \Omega_p(r)$. This results in local charts $\{(\Omega_p(r), (\exp_p^{\mathcal{M}})^{-1})\}$, which give the so-called normal coordinates. Due to compactness, we can cover $\mathcal{M}$ with finitely many $\{\Omega_{p_i}(r)\}$, where the p_i are points on $\mathcal{M}$ to obtain an atlas for $\mathcal{M}$ of normal coordinates. To simplify notation, we write ϕ_i for $(\exp_{p_i}^{\mathcal{M}})^{-1}$ and Ω_i for $\Omega_{p_i}(r)$. Furthermore, let $\{\psi_i\}$ be a partition of unity subordinate to $\{(\Omega_i, \phi_i)\}$, which is to say that $\operatorname{supp}\psi_i \subset \Omega_i$. Such a partition always exists, see Proposition 7.2.1 in Triebel (1992). We require $0 < r \leq r_0/32$ for technical reasons (cf. Triebel, 1992, Remark 7.2.1/2). We let $\mathrm{d}\lambda_{\mathcal{M}}$ denote the Riemannian volume form, which allows us to define $L_2(\mathcal{M})$ as the Hilbert space of equivalence classes of square integrable functions equipped with the usual inner product $\langle \cdot, \cdot \rangle_0$ and norm $\|\cdot\|_0$.

The Laplace–Beltrami operator will be important to us. To define this operator, let $C^\infty(\mathcal{M})$ be the infinitely often differentiable functions on $\mathcal{M}$ and $\boldsymbol{G} = (g_{ij})_{ij}$ the component matrix of the Riemannian metric on $\mathcal{M}$ in local coordinates. The Laplace–Beltrami operator Δ is the operator given in local coordinates for $f \in C^\infty(\mathcal{M})$ as

$$\Delta f = \frac{1}{\sqrt{\det(\boldsymbol{G})}}\left(\sum_{i=1}^{n} \partial_i \sqrt{\det(\boldsymbol{G})} \sum_{j=1}^{n} g^{ij}\partial_j f\right),$$

where g^{ij} are the entries of $\boldsymbol{G}^{-1}$ and $\{\partial_i\}_{i=1}^d$ are the basis for the $T_p\mathcal{M}$. There is a unique, self-adjoint and negative semi-definite extension of Δ [originally defined on $C^\infty(\mathcal{M})$] to a domain $\mathcal{D}(\Delta) \subset L_2(\mathcal{M})$ (Strichartz, 1983). We identify the extension

with Δ. By Theorem 1.3 in (Craioveanu et al., 2001, Chapter 3) there is an orthonormal basis $\{e_i\}$ of $L_2(\mathcal{M})$ consisting of infinitely often differentiable eigenfunctions of $-\Delta$ such that the corresponding eigenvalues $\{\lambda_i\}$ satisfy Weyl's law. That is, $\lambda_1 = 0$ and there are $C, c > 0$ such that

$$(3) \qquad\qquad ci^{2/d} \leq \lambda_i \leq Ci^{2/d} \quad \text{for all} \quad i \geq 2.$$

2.2. **Function Spaces.** We now present several families of function spaces that are useful for studying the sample paths of stochastic processes. We also present some well-known results that establish connections between these spaces that will be important later.

2.2.1. *Sobolev Spaces.* We define Sobolev spaces on $\mathbb{R}^d$ as Bessel potential spaces and refer to (Triebel, 1992, Chapters 1 and 2) for other equivalent definitions. Define the Fourier transform $\hat{f}$ of $f \in L_2(\mathbb{R}^d)$ as $\hat{f}(\xi) = \int_{\mathbb{R}^d} f(x)e^{-ix^\top \xi}\,\mathrm{d}x$. The Sobolev space of order $s \geq 0$ is

$$H^s(\mathbb{R}^d) = \left\{ f \in L_2(\mathbb{R}^d) \ \middle| \ \|f\|^2_{H^s(\mathbb{R}^d)} := \int_{\mathbb{R}^d} (1 + \|\xi\|^2)^s |\hat{f}(\xi)|^2\,\mathrm{d}\xi < \infty \right\}.$$

On closed manifolds we define Sobolev spaces via the eigenpairs $\{(e_i, \lambda_i)\}$ of $-\Delta$. For other definitions and equivalence results, see Chapter 7 in Triebel (1992). For a real a, we write $a - \Delta$ for $aI - \Delta$, where I is the identity. The operator $1 - \Delta$ must have positive eigenvalues $\{1 + \lambda_i\}$ and from Weyl's law (3) we have $ci^{2/d} \leq 1 + \lambda_i \leq Ci^{2/d}$ for all $i \geq 1$. For $s \in \mathbb{R}$ we define the fractional powers $(1 - \Delta)^s$ via the functional calculus

$$(1 - \Delta)^s f = \sum_{i=1}^{\infty} (1 + \lambda_i)^s \langle f, e_i \rangle_0 e_i.$$

We may now define the Sobolev space of order $s \geq 0$ on $\mathcal{M}$ as

$$H^s_L = \left\{ f \in L_2(\mathcal{M}) \ \middle| \ \|f\|^2_{H^s_L} := \|(1 - \Delta)^{s/2} f\|^2_0 = \sum_{i=1}^{\infty} (1 + \lambda_i)^s \langle f, e_i \rangle^2_0 < \infty \right\}$$

$$= \left\{ f \in L_2(\mathcal{M}) \ \middle| \ (1 - \Delta)^{s/2} f \in L_2(\mathcal{M}) \right\}.$$

It is well known and easy to show that these spaces are Hilbert spaces with the inner product

$$\langle f, g \rangle_{H^s_L} = \langle (1 - \Delta)^{s/2} f, (1 - \Delta)^{s/2} g \rangle_0,$$

see Strichartz (1983). We have the following useful connection between $H^s(\mathbb{R}^d)$ and H^s_L. Here we implicitly make use of Theorem 7.4.5 in Triebel (1992). See Große and Schneider (2013) for a result for more general submanifolds.

Theorem 2.1 (Theorem 1 in Skrzypczak, 1990)**.** *Let* $s \geq 0$*. The restriction operator* $T_{\mathcal{M}} f = f|_{\mathcal{M}}$ *is a bounded linear operator* $T_{\mathcal{M}} : H^{s+(k-d)/2}(\mathbb{R}^k) \to H_L^s$*. Moreover,* $T_{\mathcal{M}}$ *has a bounded inverse operator* $E_{\mathcal{M}} : H_L^s \to H^{s+(k-d)/2}(\mathbb{R}^k)$*.*

In the definition of H_L^s, we could have used the operator $\tau - \Delta$ instead for any $\tau > 0$. Clearly, the resulting spaces are norm equivalent. In the statistical context considered in the remainder of the paper, τ has a concrete interpretation as a correlation range scale parameter. To account for τ, we introduce a modified inner product and norm on H_L^s. Let $\theta = (s, \tau)$ and

$$\langle f, g \rangle_\theta := v(\theta) \langle (\tau - \Delta)^{s/2} f, (\tau - \Delta)^{s/2} g \rangle_0 \quad \text{and} \quad \|f\|_\theta = \sqrt{\langle f, f \rangle_\theta},$$

where $v(\theta)$ is a positive "normalizing" function. It is easy to show that the norms $\|\cdot\|_{H_L^s}$ and $\|\cdot\|_\theta$ are equivalent. The function $v(\theta)$ will be discussed more extensively in Section 2.3.

The following lemma about monotonicity of the norms will be very convenient. Note that $\|f\|_{H_L^s} \leq \|f\|_{H_L^{s'}}$ and $\|f\|_{H^s(\mathbb{R}^k)} \leq \|f\|_{H^{s'}(\mathbb{R}^k)}$ whenever $s \leq s'$.

Lemma 2.2. *Let* $\theta = (s, \tau)$ *and* $\theta' = (s', \tau')$*. Suppose there exists* C *that does not depend on* s, τ *such that*

$$v(\theta')(\tau' + \lambda)^{s'} \leq C v(\theta)(\tau + \lambda)^s$$

whenever $s' \leq s$*. Let* $f \in H_L^s$*. Then* $\|f\|_{\theta'} \leq C\|f\|_\theta$ *whenever* $s' \leq s$*.*

Proof. The result follows from

$$\|f\|_{\theta'}^2 = v(\theta') \sum_{i=1}^\infty (\tau' + \lambda_i)^{s'} \langle f, e_i \rangle_0^2 \leq C v(\theta) \sum_{i=1}^\infty (\tau + \lambda_i)^s \langle f, e_i \rangle_0^2 = C\|f\|_\theta^2. \qquad \square$$

The assumption in Lemma 2.2 holds if, for example, τ is fixed and $v(\theta) \geq c\tau^{-s}$ or v is continuous and θ varies in a compact set. Finally, for $s \in \mathbb{R}$, define the scale of Hilbert spaces $\hat{H}_L^s$ and their inner product as

$$\hat{H}_L^s = \left\{ a = \{a_i\}_{i=1}^\infty \subset \mathbb{R} \; \middle| \; \sum_{i=1}^\infty a_i^2 (1 + \lambda_i)^s < \infty \right\} \quad \text{and} \quad \langle a, b \rangle_{\hat\theta} = v(\theta) \sum_{i=1}^\infty (\tau + \lambda_i)^s a_i b_i.$$

For $s \geq 0$ we have the obvious isomorphism $\iota : H_L^s \to \hat{H}_L^s$ given by $(\iota(f))_i = \langle f, e_i \rangle_0$. Observe that $\hat{H}_L^0$ equals the Hilbert space of square summable sequences.

2.2.2. *Hölder Spaces and the Sobolev Imbedding Theorem.* For $0 < \alpha < 1$ and a function $f : \mathcal{M} \to \mathbb{R}$ such that the mth covariant derivative ∇^m exists we define the Hölder seminorm

$$|\nabla^m f|_\alpha := \sup_{x \neq y \in \mathcal{M}} \frac{\|\nabla^m f(x) - \nabla^m f(y)\|_g}{d_\mathcal{M}(x, y)^\alpha},$$

where the supremum is taken over all y in the same normal coordinate chart as x and $\nabla^m f(y)$ denotes the tensor at x after parallel transport along the geodesic between x and y; see Aubin (1998, p. 8) for the definition of parallel transport. The norm $\|\cdot\|_g$ is the natural norm induced by the Riemannian metric g, see for example Hangelbroek et al. (2010) for details.

Following Aubin (1998), for any $l \in \mathbb{N}$ and $0 < \alpha < 1$ we define the Hölder space

$$C^{l,\alpha}(\mathcal{M}) = \left\{ f : \mathcal{M} \to \mathbb{R} \;\middle|\; \|f\|_{C^{l,\alpha}(\mathcal{M})} := |\nabla^l f|_\alpha + \max_{m \leq l} \sup_{x \in \mathcal{M}} \|\nabla^m f(x)\|_g < \infty \right\}.$$

The relevance of Hölder spaces in our context arises in part from the Sobolev imbedding theorem. Specifically, the Sobolev space H_L^s for $s > d/2$ can be continuously embedded in $C^{l,\alpha}(\mathcal{M})$ for all $l + \alpha < s - d/2$; see Aubin (1998, Theorem 2.20) or Hangelbroek et al. (2010, Section 3.3). In particular, the point evaluation functionals on H_L^s are bounded and thus the spaces H_L^s are reproducing kernel Hilbert spaces (RKHS) for $s > d/2$. We identify the reproducing kernels and discuss their properties in the next section. Apart from the Sobolev embedding theorem, the Hölder spaces will appear in connection with sample path properties of Gaussian processes. The continuous functions will be deonted by $C(\mathcal{M})$.

2.3. **Kernels.** Note that the operator $\sqrt{v(\theta)}(\tau - \Delta)^{-s/2} : L_2(\mathcal{M}) \to H_L^s \subset L_2(\mathcal{M})$ is Hilbert–Schmidt for $s > d/2$ as

$$\sum_{i=1}^\infty \left\| \sqrt{v(\theta)}(\tau - \Delta)^{-s/2} e_i \right\|_0^2 = v(\theta) \sum_{i=1}^\infty (\tau + \lambda_i)^{-s} \leq C v(\theta) \sum_{i=1}^\infty i^{-2s/d} < \infty.$$

It thus follows (Reed and Simon, 1980, Theorem VI.23) that $\sqrt{v(\theta)}(\tau - \Delta)^{-s/2}$ is an integral operator with the corresponding kernel $K_\theta \in L_2(\mathcal{M} \times \mathcal{M})$ given by

$$(4) \qquad\qquad K_\theta(x, y) = v(\theta) \sum_{i=1}^\infty (\tau + \lambda_i)^{-s} e_i(x) e_i(y).$$

The kernel K_θ is the reproducing kernel of H_L^s endowed with the $\langle \cdot, \cdot \rangle_\theta$ inner product. The pointwise definition makes sense as e_i can be identified with $C^\infty(\mathcal{M})$ functions. We write $K_\theta(x)$ for $K_\theta(x, x)$. The following proposition collects some facts about K_θ.

Additional regularity properties of the kernel can be established via the results obtained in Kerkyacharian et al. (2018).

Proposition 2.3. *Let K_θ be the kernel defined in (4) with $s > d/2$. Then*

    (i) K_θ is continuous and the series in (4) converges uniformly and absolutely.
    (ii) K_θ is strictly positive definite.

Proof. Let $\alpha \in (0,1)$ be such that $\alpha < s - d/2$. We first prove (i). According to the Sobolev embedding theorem, $H_L^s \subset C^{0,\alpha}(\mathcal{M})$ and the inclusion operator is continuous. It is a well-known consequence of the Arzèla–Ascoli theorem that the embedding of $C^{0,\alpha}(\mathcal{M})$ in $C(\mathcal{M})$ is compact, and thus also the embedding of H_L^s in $C(\mathcal{M})$. The continuity of K_θ then follows from Lemma 2.2 in Ferreira and Menegatto (2013). The remaining assertions in (i) follow from the continuity of the kernel and the compactness of $\mathcal{M}$ in combination with Dini's theorem; see for example Lemma 4.6.6 in Hsing and Eubank (2015).

To prove (ii), we repeat an argument from Hangelbroek et al. (2010). Positive semi-definiteness is obvious. For $\boldsymbol{x} = (x_1, \ldots, x_n)$, define $(K_\theta(\boldsymbol{x}))_{ij} = K_\theta(x_i, x_j)$. A straight-forward calculation yields

$$\boldsymbol{c}^\top K_\theta(\boldsymbol{x})\boldsymbol{c} = \sum_{i,j=1}^n \langle c_i K_\theta(\cdot, x_i), c_j K_\theta(\cdot, x_j)\rangle_\theta^2 = \left\|\sum_{i=1}^n c_i K_\theta(\cdot, x_i)\right\|_\theta^2.$$

Assume that $x_i \neq x_j$ for $i \neq j$. If the right hand side above is 0 for a non-zero $\boldsymbol{c}$, then $\{K_\theta(\cdot, x_i)\}$ are *not* linearly independent. This means that, for every $f \in H_L^s$,

$$f(x_i) = \langle f, K_\theta(\cdot, x_i)\rangle_\theta = \sum_{j \neq i}\langle f, a_j K_\theta(\cdot, x_j)\rangle_\theta = \sum_{j \neq i} a_j f(x_j)$$

for some coefficients a_j. Because H_L^s contains bump functions such that $f(x_i) \neq 0$ and $f(x_j) = 0$ for $j \neq i$, this yields a contradiction and thus K_θ is strictly positive definite. $\qquad\square$

In addition to s and τ, we consider a magnitude parameter σ^2, which is included in the model via the scaled kernel function $(x, y) \mapsto \sigma^2 K_\theta(x, y)$.

At this point, we want to briefly discuss the normalizing function $v(\theta)$. Different choices for $v(\theta)$ have been advocated for in the literature. The function $v(\theta) = \tau^{-s+d/2}$ is used in Sanz-Alonso and Yang (2022), who provide a compelling justification in their Remark 2.1: This choice balances the average marginal variance of a process with covariance function $K_\theta(x, y)$ and thus the parameter σ^2 approximately controls the marginal variance. The function $v(\theta) = \tau^{-s+d/2}$ also appears in the spectral density of

the Euclidean Matérn covariance function given in (1). Similarly motivated is the use in Borovitskiy et al. (2020) and Li et al. (2023) of the normalizing function

$$(5) \qquad v(\theta) = \left(\frac{1}{\text{vol}(\mathcal{M})} \int_{\mathcal{M}} \sum_{i=1}^{\infty} (\tau + \lambda_i)^{-s} e_i(x)^2 \, d\lambda_{\mathcal{M}} \right)^{-1}.$$

On certain manifolds (e.g., the sphere) $K_{\theta}(x)$ does not depend on x (Borovitskiy et al., 2020; Narcowich and Ward, 2002). With $v(\theta)$ in (5), σ^2 is exactly the marginal variance. However, explicit computation of this normalization is difficult. The precise form of $v(\theta)$ will not be important as long as we can use Lemma 2.2.

3. WHITTLE-MATÉRN PROCESSES ON CLOSED MANIFOLDS

We will consider the following random elements in $L_2(\mathcal{M})$. Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space and let $\{\xi_i\}$ be a sequence of i.i.d. random variables defined on Ω such that $\mathbb{E}(\xi_i) = 0$ and $\mathbb{E}(\xi_i^2) = 1$. Set

$$(6) \qquad u = \sigma \sqrt{v(\theta)} \sum_{i=1}^{\infty} (\tau + \lambda_i)^{-s/2} \xi_i e_i.$$

For $s > d/2$ this is indeed an element of $L_2(\mathcal{M})$.

Proposition 3.1. *The random element u defined in (6) is almost surely in $H_L^{s-d/2-\epsilon}$ for all $s - d/2 > \epsilon > 0$. In particular, $u \in L_2(\mathcal{M})$ if $s > d/2$.*

Proof. Let $\epsilon > 0$ be such that $s - d/2 - \epsilon > 0$. By Weyl's law in (3),

$$\lim_{n \to \infty} \mathbb{E} \left\| \sum_{i=1}^{n} (\tau + \lambda_i)^{-s/2} \xi_i e_i \right\|_{\theta}^2 = \lim_{n \to \infty} v(\theta) \sum_{i=1}^{n} (\tau + \lambda_i)^{s-d/2-\epsilon-s} \mathbb{E} \left(\xi_i^2 \right)$$

$$\leq \lim_{n \to \infty} C \sum_{i=1}^{n} (i^{2/d})^{-d/2-\epsilon}$$

$$= \lim_{n \to \infty} C \sum_{i=1}^{n} i^{-1-2\epsilon/d} < \infty.$$

The monotone convergence theorem yields $\mathbb{E}\|u\|_{\theta}^2 < \infty$, so that $\|u\|_{\theta}^2 < \infty$ almost surely. $\qquad \square$

Note that $\mathbb{E}(u) = 0$. Furthermore, the covariance operator of u is by construction

$$\mathbb{C}\text{ov}(u) = \mathbb{E}(u \otimes u) = \sigma^2 v(\theta)(\tau - \Delta)^{-s},$$

where for two elements $v_1, v_2 \in L_2(\mathcal{M})$ $(v_1 \otimes v_2)(f) = \langle v_1, f \rangle_0 v_2$. If $s > d$, the Sobolev embedding theorem guarantees that u almost surely has continuous sample paths so

that the measurability of the pointwise evaluations is not a problem. Normally, however, the smoothness parameter is considered over the range $s > d/2$, which only guarantees that u is an element in $L_2(\mathcal{M})$. Fortunately, it follows from the results of Section 2.3 that the assumptions of Steinwart (2019, Theorem 3.3) are fulfilled for any $s > d/2$. This theorem ensures the existence of a measurable version $\tilde{u}$ of u such that $\mathbb{Cov}(\tilde{u}(x), \tilde{u}(y)) = \sigma^2 K_\theta(x, y)$ and, moreover, $\tilde{u}$ admits the Karhunen–Loève expansion

$$\tilde{u}_\omega(x) = \sum_{i=1}^{\infty} \langle \tilde{u}_\omega, e_i \rangle_0 e_i(x) = \sigma \sqrt{v(\theta)} \sum_{i=1}^{\infty} (\tau + \lambda_i)^{-s/2} \xi_i(\omega) e_i(x)$$

for $\mathbb{P}$-almost all ω. Here measurability of $\tilde{u}$ means that the map $(x, \omega) \mapsto \tilde{u}_\omega(x)$ is measurable with respect to the product σ-algebra $\mathcal{F} \otimes \mathcal{B}(\mathcal{M})$, where $\mathcal{B}(\mathcal{M})$ is the σ-algebra generated by the open subsets of $\mathcal{M}$. We may therefore assume that point evaluations of u are measurable.

An alternative way to define u is to define it as the (unique) solution to the Whittle–Matérn stochastic partial differential equation

(7)
$$(\tau - \Delta)^{s/2} u = \sigma \sqrt{v(\theta)} \mathcal{W},$$

where $\mathcal{W}$ is white noise. This approach is due to Peter Whittle (Whittle, 1963) and is the approach taken in Lindgren et al. (2011) and Bolin and Kirchner (2023). In both works more general differential operators than $(\tau - \Delta)$ are considered. We believe that our results on smoothness estimation should translate to other differential operators but for concreteness and simplicity of presentation, we restrict our attention to the standard Whittle-Matérn covariance operators. To make sense of (7) and see that this yields the same model as (6), define white noise to be any sequence of i.i.d. centered random variables $\mathcal{W} = \{\xi_i\}$ with finite second moments and then note that (i) for $\theta = (s, \tau)$ where $s < -d/2$, the norm $\|\mathcal{W}\|_{\hat{\theta}}$ is finite almost surely by the argument used in the proof of Proposition 3.1, and (ii) the operator $\iota^{-1} \circ (\tau - \hat{\Delta})^s$ defined through

$$(\iota^{-1} \circ (\tau - \hat{\Delta})^{-s})(a) = \sum_{i=1}^{\infty} (\tau + \lambda_i)^{-s} a_i e_i$$

is an isometric isomorphism from $\hat{H}_L^\beta$ to $H_L^{\beta+2s}$ for all $\beta \geq -2s$. A more standard construction that yields the same model is used in Lindgren et al. (2011) and Bolin and Kirchner (2023).

We emphasize that we do *not* assume that the ξ_i are Gaussian random variables. Unless otherwise specified, we only assume that ξ_i have finite second moments. However, we do assume that the random variables are independent. In the Gaussian case, this is not a restriction. Any mean zero Gaussian process with covariance function K_θ

admits a Karhunen–Loève decomposition (6) with independent $\{\xi_i\}$. For non-Gaussian processes, $\{\xi_i\}$ are only guaranteed to be uncorrelated.

4. GAUSSIAN MAXIMUM LIKELIHOOD ESTIMATION

We want to recover the smoothness parameter s via maximum likelihood estimation. Fix therefore the true parameter vector $\theta_0 = (s_0, \tau_0)$ with $s_0 > d/2$ and let u be defined by

$$(8) \qquad u(x) = \sigma_0 \sqrt{v(\theta_0)} \sum_{i=1}^{\infty} (\tau_0 + \lambda_i)^{-s_0/2} \xi_i e_i(x),$$

where $\{\xi_i\}$ are i.i.d. random variables with $\mathbb{E}(\xi_i) = 0$ and $\mathbb{E}(\xi_i^2) = 1$. By construction, the covariance operator of u is $\sigma_0^2 v(\theta_0)(\tau_0 - \Delta)^{-s_0}$. The corresponding covariance function is

$$\mathbb{C}\mathrm{ov}(u(x), u(y)) = \sigma_0^2 K_{\theta_0}(x, y) = \sigma_0^2 v(\theta_0) \sum_{i=1}^{\infty} (\tau_0 + \lambda_i)^{-s_0} e_i(x) e_i(y).$$

Assume that we observe the process u at pairwise distinct points $(x_1, \ldots, x_n) = \boldsymbol{x}$. Using this notation, we write $u(\boldsymbol{x})^\top = (u(x_1), \ldots, u(x_n))$. To estimate the smoothness parameter s_0 of the process u, we will use the maximum likelihood estimators. However, we allow the likelihood to be misspecified in two different ways. First, instead of requiring that the true family of covariance kernels $\sigma^2 K_\theta$ is used, we only require the use of a family of kernels $\{K_s\}_{s>d/2}$ such that for each s their RKHS $\mathcal{H}^s$ with the norm $\|\cdot\|_s$ is norm equivalent to H_L^s. This includes the correctly specified kernels $\sigma^2 K_\theta$ with $\sigma \neq \sigma_0$ or $\tau \neq \tau_0$. Moreover, it turns out that there are other well-known and useful kernels whose RKHS are norm equivalent to H_L^s. We discuss some of these kernels in Section 4.4. Second, we always use the Gaussian log-likelihood

$$(9) \qquad \ell(s; u(\boldsymbol{x})) = -\frac{n}{2} \log(2\pi) - \frac{1}{2} \log \left(\det(K_s(\boldsymbol{x})) - \frac{1}{2} u(\boldsymbol{x})^\top K_s(\boldsymbol{x})^{-1} u(\boldsymbol{x}).$$

Here, we write $K_s(\boldsymbol{x})$ for the matrix with entries $(K_s(\boldsymbol{x}))_{ij} = K_s(x_i, x_j)$. Of course, this can be a correctly specified likelihood only if u is a Gaussian process.

It is well known that the spatial sampling of the points $x_1, \ldots, x_n$ has significant impact on the quality of maximum likelihood estimators (e.g., Bachoc, 2014). To establish consistency of the maximum likelihood estimator, we will make a fairly restrictive assumption on the uniformity of sampling. However, this assumption is standard in the function approximation literature (e.g., Narcowich et al., 2006; Wendland, 2004) and thus allows us to use the wealth of results established there. Moreover, all existing results on smoothness estimation for Gaussian processes require some type of uniformity

assumption (Chen et al., 2021; Karvonen, 2023; Loh and Sun, 2023; Loh et al., 2021; Petit, 2023). Define the *fill-distance* $h_{\boldsymbol{x}}$ and the *separation radius* $q_{\boldsymbol{x}}$ as

$$(10) \qquad h_{\boldsymbol{x}} = \sup_{x \in \mathcal{M}} \min_{1 \leq i \leq n} d_{\mathcal{M}}(x, x_i) \quad \text{and} \quad q_{\boldsymbol{x}} = \frac{1}{2} \min_{i \neq j} d_{\mathcal{M}}(x_i, x_j).$$

Also define the *mesh norm* (or *mesh ratio*) $\rho_{\boldsymbol{x}} = h_{\boldsymbol{x}}/q_{\boldsymbol{x}}$. Note that by definition $q_{\boldsymbol{x}} \leq h_{\boldsymbol{x}}$ and therefore $\rho_{\boldsymbol{x}} \geq 1$. Our consistency results assume that the point sequence is quasi-uniform.

Definition 4.1. *A sequence of points $\{x_i\} \subset \mathcal{M}$ is quasi-uniform if*

$$h_{\boldsymbol{x}} \asymp q_{\boldsymbol{x}} \asymp n^{-1/d}.$$

This implies $\rho_{\boldsymbol{x}}$ is uniformly bounded from above.

Note that uniformly sampled random points are *not* quasi-uniform. In Loh et al. (2021, Section 4), an estimator for the smoothness of Gaussian random fields on bounded Euclidean domains with the Matérn covariance function is introduced, which is consistent when x_i are sampled from a density that is uniformly lower bounded on the domain. However, the estimator there effectively uses only a subset of the sampling sites that are quasi-uniform up to factors of $\log(n)$, but whose cardinality is of order n. A modification of the construction in Loh et al. (2021, Section 4) for the manifold $\mathcal{M}$ could be used to establish variants of our consistency results for random points.

Our proofs are based on connecting the quadratic form $u(\boldsymbol{x})^{\top} K_s(\boldsymbol{x})^{-1} u(\boldsymbol{x})$ and the log-determinant $\log(\det(K_s(\boldsymbol{x}))$ in the Gaussian likelihood (9) to quantities related to optimal interpolation in the RKHS $\mathcal{H}^s$. Namely, the quadratic form is the squared RKHS norm of the *minimum norm interpolant* and the log-determinant is related to *worst-case approximation error*. Let K be a positive definite reproducing kernel of an RKHS $\mathcal{H}$ and u a function that has been observed at the points $x_1, \ldots, x_n$. The unique minimum norm interpolant $m_{\mathcal{H},n}$ of u is

$$m_{\mathcal{H},n} = \operatorname*{arg\,min}_{f \in \mathcal{H}} \left\{ \|f\|_{\mathcal{H}} \mid f(x_i) = u(x_i), \ 1 \leq i \leq n \right\}.$$

We write $m_{s,n}$ for the minimum norm interpolant based on the kernel K_s. It is well known and easy to verify that the quadratic form in (9) equals the squared norm of this interpolant:

$$(11) \qquad \|m_{s,n}\|_s^2 = u(\boldsymbol{x})^{\top} K_s(\boldsymbol{x})^{-1} u(\boldsymbol{x}).$$

For a proof, see for example Theorem 3.5 in Kanagawa et al. (2018).

For the connection to worst-case approximation error, define the "conditional variances"

$$\text{(12)} \qquad \mathbb{V}_s(y|\boldsymbol{x}) = K_s(y) - K_s(y, \boldsymbol{x}) K_s(\boldsymbol{x})^{-1} K_s(\boldsymbol{x}, y).$$

Here, $K_s(y, \boldsymbol{x})$ is the $1 \times n$ vector with entries $K_s(y, \boldsymbol{x})_j = K_s(y, x_j)$. If u is a Gaussian process with covariance function given by K_s, this is indeed the conditional variance of $u(y)$ given $\{u(x_j) \mid 1 \leq j \leq n\}$. The variance $\mathbb{V}_s(\cdot|\boldsymbol{x})$ is the worst-case approximation error in that

$$\sup_{\|f\|_{\mathcal{H}^s} \leq 1} |f(y) - m_{s,n}(y)| = \sqrt{\mathbb{V}_s(y|\boldsymbol{x})}.$$

See, for example, Kanagawa et al. (2018, Corollary 3.11). Now, let $\boldsymbol{x}^m$ denote the set consisting of $x_1, \ldots, x_m$. Using the Shermann–Morrison–Woodbury identity and (12) repeatedly, we write

$$\det(K_s(\boldsymbol{x})) = \big(K_s(x_n) - K_s(x_n, \boldsymbol{x}^{n-1}) K_s(\boldsymbol{x}^{n-1}) K_s(\boldsymbol{x}^{n-1}, x_n)\big) \det(K_s(\boldsymbol{x}^{n-1}))$$

$$\text{(13)} \qquad = \mathbb{V}_s(x_n|\boldsymbol{x}^{n-1}) \det(K_s(\boldsymbol{x}^{n-1}))$$

$$= \prod_{i=1}^n \mathbb{V}_s(x_i|\boldsymbol{x}^{i-1}).$$

The identities (11) and (13) show that precise control of the $\mathbb{V}_s(y|\boldsymbol{x})$ and $\|m_{s,n}\|_s$ allows for precise control over the log-likelihood (9). We next establish bounds for these quantities. From here on, we fix a family of kernels K_s, their reproducing kernel Hilbert spaces $(\mathcal{H}^s, \|\cdot\|_s)$ required to be equal to H_L^s up to equivalent norms and the corresponding minimum norm interpolants $m_{s,n}$ of u.

4.1. **Norm and Determinant Bounds.** In this section we present a number of auxiliary results. The proofs are moved to A as they are either technical in nature or modifications of well-known results. We assume throughout that u is defined by (8). To simplify the notation, we use u for both the random element and its sample paths.

First, we set up bounds for $\|m_{s,n}\|_s$. The size of the norm depends on the smoothness of u. If u is very smooth, then $\|m_{s,n}\|_s$ is small because $m_{s,n}$ is "simple"; if u is very coarse, then the norm is large because $m_{s,n}$ must be "complicated". The following proposition restricts how smooth u can be by providing a lower bound for the decay rate of its Fourier coefficients. The proofs are given in the Appendix.

Proposition 4.2. *Let $s_0 > d/2$ and let $\{\xi_i\}$ be the sequence of i.i.d. random variables used in the definition of u in (8). Then almost surely, there is a positive constant c*

such that

$$\sum_{i=R}^{\infty}(\tau_0 + \lambda_i)^{-s_0}\xi_i^2 \geq cR^{-2s_0/d+1} \quad \text{for all} \quad R \in \mathbb{N}.$$

Next, we introduce a lower bound for $\|m_{s,n}\|_s$ under an approximation condition. That this condition holds for some useful kernels, in particular for $\sigma^2 K_\theta$, will be confirmed later.

Proposition 4.3. *Let* $\theta = (s, \tau)$ *with* $s > d/2$. *Suppose that* $f \in \mathcal{H}^s$ *and, almost surely,*

(14)
$$\|u - f\|_0 \leq C_1 n^{-s_0/d+1/2+\epsilon/d} + C_2 n^{-s/d}\|f\|_s$$

for every $\epsilon > 0$, *where* $C_1, C_2 > 0$ *can depend on* $s, s_0, \epsilon, \mathcal{M}$ *and the sample path. Then almost surely*

(15)
$$\|f\|_s^2 \geq Cn^{1+2(s-s_0)/d-\epsilon'}$$

for every $\epsilon' > 0$, *where* $C > 0$ *depends on* $s, s_0, \epsilon', \mathcal{M}$ *and the sample path.*

Remark 4.4. *We would like to discuss the assumption in* (14), *which may seem unusual. We are mainly interested in applying Proposition 4.3 with* $f = m_{s,n}$. *In this scenario one commonly has the error estimate*

$$\|u - m_{s,n}\|_0 \leq Cn^{-s_0/d+1/2+\epsilon/d}\|u\|_{s_0-d/2-\epsilon},$$

which clearly implies (14), *and we will set up a similar bound in Proposition 4.5 below. Note, however, that we do not assume that* $f = m_{s,n}$. *The proposition holds for any* f *that approximates* u *sufficiently well. This additional flexibility will be important later.*

The following proposition shows that the assumption for Proposition 4.3 is generally satisfied if $s \geq s_0$. Since they apply to functions outside the RKHS of $\mathcal{H}^s$, results of this type are often referred to as "escape" results and are based on the fundamental work by Narcowich and Ward (2002) and Narcowich et al. (2006). The statement and the proof are essentially the same as in Fuselier and Wright (2012). For completeness, we provide the proof in the Appendix. The proof combines known results for the Euclidean case from Narcowich et al. (2006) with Theorem 2.1.

Proposition 4.5. *Let* $f \in H_L^{s_0}$ *with* $s_0 > d$ *and let* $m_{s,n}^f$ *be its minimum norm interpolant from the RKHS* $\mathcal{H}^s$. *Then there exists a positive constant* C *that depends on* s_0 *and* $\mathcal{M}$ *such that we have almost surely*

(16)
$$\|f - m_{s,n}^f\|_0 \leq Ch_{\boldsymbol{x}}^{s_0}\rho_{\boldsymbol{x}}^{s-s_0}\|f\|_{s_0}$$

when $h_{\boldsymbol{x}}$ is small enough. Thus, if $h_{\boldsymbol{x}} \to 0$ and C is allowed to depend on f, then (16) *holds for all n.*

See Lemma 3.8 in Hangelbroek et al. (2010) for a precise statement on how small $h_{\boldsymbol{x}}$ needs to be. Under quasi-uniformity, $h_{\boldsymbol{x}} \leq Cn^{-1/d}$ and $\rho_{\boldsymbol{x}}$ is uniformly upper bounded and for $f = u$ we thus obtain

$$
(17) \quad
\begin{aligned}
\|u - m_{s,n}\|_0 &\leq Ch_{\boldsymbol{x}}^{s_0 - d/2 - \epsilon} \rho_{\boldsymbol{x}}^{s - (s_0 - d/2 - \epsilon)} \|u\|_{s_0 - d/2 - \epsilon} \\
&\leq Cn^{-s_0/d + 1/2 + \epsilon/d} \|u\|_{s_0 - d/2 - \epsilon}
\end{aligned}
$$

for all $\epsilon > 0$, where C can depend on the sample path. Together, Propositions 4.3 and 4.5 provide lower bounds on $\|m_{s,n}\|_s$ when $s \geq s_0 > d$. Next we provide an upper bound.

Proposition 4.6. *Let u be as in* (8). *Suppose that $s_0 > d$ and $s > d/2$. Let $m_{s,n}$ be the minimum norm interpolant for u. Then almost surely*

$$
(18) \quad \|m_{s,n}\|_s^2 \leq Cn^{2(s - s_0) + 1 + \epsilon}
$$

for all $\epsilon > 0$, where C depends on s, θ_0, ϵ and the sample path. If ξ_i are Gaussian, then it is only necessary to assume $s_0 > d/2$ to obtain

$$
(19) \quad cn \leq \|m_{s_0,n}\|_{s_0}^2 \leq Cn
$$

almost surely for positive constants c and C that depend only on θ_0 and the sample path.

See the Appendix for a proof. The bound in (18) is well known in the Euclidean setting and its proof only needs to be slightly adapted. The bound in (19) follows from a simple argument based on the law of large numbers. In the Gaussian case, the bound (19) is strictly stronger than (18) if $s = s_0$.

Finally, we obtain bounds for the worst-case approximation error. These results are standard and we merely adapt known proofs for manifolds. The proofs are included in the Appendix.

Proposition 4.7. *Let $s > d/2$.*

(1) *Suppose the family of norms $\{\|\cdot\|_s\}_{s > d/2}$ is such that $\|f\|_s \leq C\|f\|_{s'}$ for every f whenever $s \leq s'$. Let $\boldsymbol{x}_i^n$ be all points in $\boldsymbol{x}$ except x_i. Then, for $s \leq S$, there is a constant $c > 0$, that only depends on S, such that*

$$
\min_{1 \leq i \leq n} \mathbb{V}_s(x_i | \boldsymbol{x}_i^n) \geq cq_{\boldsymbol{x}}^{2S - d}.
$$

(2) *There exists a positive constant C, that only depends on s and $\mathcal{M}$, such that*

$$\sup_{y \in \mathcal{M}} \mathbb{V}_s(y|\boldsymbol{x}) \leq C h_{\boldsymbol{x}}^{2s-d}.$$

4.2. **Estimation of the Smoothness Parameter.** Based on the previous auxiliary results, we may now prove the main result of this paper.

Theorem 4.8. *Let u be defined as in (8) with $s_0 > d$. Let the family of kernels $\{K_s\}_{s>d/2}$ be such that the associated norms $\{\|\cdot\|_s\}_{s>d/2}$ satisfy the monotonicity condition that, whenever $s \leq s'$,*

(20)
$$\|f\|_s \leq C\|f\|_{s'}$$

for some positive C that does not depend on s and s'. Assume that the sequence of points $\{x_i\}$ is quasi-uniform. If $\{\hat{s}_n\}$ is a sequence of maximizers of the Gaussian log-likelihood

(21)
$$\ell(s; u(\boldsymbol{x})) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(\det(K_s(\boldsymbol{x}))) - \frac{1}{2}u(\boldsymbol{x})^\top K_s(\boldsymbol{x})^{-1}u(\boldsymbol{x})$$

in $(d/2, S_{\max}]$ with $s_0 \leq S_{\max}$ for arbitrary but fixed $S_{\max}$, then

$$\hat{s}_n \to s_0 \quad \text{in probability.}$$

Proof. Any maximizer of $\ell(s; u(\boldsymbol{x}))$ also maximizes the logarithm of the likelihood ratio

$$\mathcal{L}(s) := 2(\ell(s; u(\boldsymbol{x})) - \ell(s_0; u(\boldsymbol{x})))$$
$$= \log\left(\frac{\det(K_{s_0}(\boldsymbol{x}))}{\det(K_s(\boldsymbol{x}))}\right) + u(\boldsymbol{x})^\top K_{s_0}(\boldsymbol{x})^{-1}u(\boldsymbol{x}) - u(\boldsymbol{x})^\top K_s(\boldsymbol{x})^{-1}u(\boldsymbol{x}).$$

Note that

$$\{|\hat{s}_n - s_0| > \delta\} \subset \left\{\sup_{d/2 < s \leq s_0 - \delta} \mathcal{L}(s) \geq 0\right\} \cup \left\{\sup_{s_0 - \delta \leq s \leq S_{\max}} \mathcal{L}(s) \geq 0\right\}$$

for every $\delta > 0$ for which the suprema are over non-empty sets. Thus it suffices to show that

(22)
$$\mathbb{P}\left(\sup_{d/2 < s \leq s_0 - \delta} \mathcal{L}(s) \geq 0\right) + \mathbb{P}\left(\sup_{s_0 - \delta \leq s \leq S_{\max}} \mathcal{L}(s) \geq 0\right) \to 0 \quad \text{as } n \to \infty$$

for every such $\delta > 0$. Consider first the case $s \geq s_0 + \delta$. Choose $0 < \epsilon < \min\{2\delta/d, s_0 - d\}$. It follows from combining Propositions 4.3 and 4.5 and quasi-uniformity that almost surely

$$u(\boldsymbol{x})^\top K_{s_0+\delta}(\boldsymbol{x})^{-1}u(\boldsymbol{x}) = \|m_{s_0+\delta,n}\|_{s_0+\delta}^2 \geq Cn^{1+2\delta/d-\epsilon}.$$

Moreover, for any $s > s_0 + \delta$, we have by (20) and the minimum norm property that

$$Cn^{1+2\delta/d-\epsilon} \leq \|m_{s_0+\delta,n}\|_{s_0+\delta}^2 \leq \|m_{s,n}\|_{s_0+\delta}^2 \leq C_2\|m_{s,n}\|_s^2.$$

Note that C in the above display depends on δ and s_0 but not on s and C_2 also doesn't depend on s. Next, we have by Proposition 4.6, equation (18) that almost surely

$$u(\boldsymbol{x})^\top K_{s_0}(\boldsymbol{x})^{-1}u(\boldsymbol{x}) = \|m_{s_0,n}\|_{s_0} \leq Cn^{1+\epsilon'}$$

for some $\epsilon' < 2\delta/d - \epsilon$, where C only depends on θ_0, ϵ' and the sample path. Clearly, we can find such an ϵ' for every δ. This gives

$$\text{(23)} \quad \sup_{s\geq s_0+\delta} u(\boldsymbol{x})^\top K_{s_0}(\boldsymbol{x})^{-1}u(\boldsymbol{x}) - u(\boldsymbol{x})^\top K_s(\boldsymbol{x})^{-1}u(\boldsymbol{x}) \lesssim n^{1+\epsilon'} - n^{1+2\delta-\epsilon}$$

$$\lesssim -n^{1+2\delta-\epsilon},$$

almost surely, where the hidden constants only depend on $\delta, \epsilon, \theta_0$ and the sample path. For the terms involving the log-determinants we have, as in (13),

$$-\log(\det(K_s(\boldsymbol{x}))) + \log(\det(K_{s_0}(\boldsymbol{x}))) = \sum_{i=1}^n \log\left(\frac{\mathbb{V}_{s_0}(x_i|\boldsymbol{x}^{i-1})}{\mathbb{V}_s(x_i|\boldsymbol{x}^{i-1})}\right).$$

By quasi-uniformity, the bounds in Proposition 4.7 apply to each $\mathbb{V}_s(x_i|\boldsymbol{x}^{i-1})$. Therefore for all $s > d/2$

$$\mathbb{V}_s(x_i|\boldsymbol{x}^{i-1}) \geq ci^{-(2S_{\max}-d)/d} \quad \text{and} \quad \mathbb{V}_{s_0}(x_i|\boldsymbol{x}^{i-1}) \leq Ci^{-(2s_0-d)/d}.$$

Thus, for $s \geq s_0 + \delta$,

$$\text{(24)} \quad \sum_{i=1}^n \log\left(\frac{\mathbb{V}_{s_0}(x_i|\boldsymbol{x}^{i-1})}{\mathbb{V}_s(x_i|\boldsymbol{x}^{i-1})}\right) \leq n\log(C/c) + \sum_{i=1}^n \log\left(i^{(2S_{\max}-2s_0)/d}\right)$$

$$= n\log(C/c) + \frac{2(S_{\max}-s_0)}{d}\sum_{i=1}^n \log(i) \asymp n\log(n),$$

where we have used Stirling's approximation $\sum_{i=1}^n \log(i) = \log(n!) = n\log(n) - O(n)$ and the hidden constants depend on s_0 and δ but not on s. It follows from the estimates (23) and (24) that almost surely there exist positive constants C_1 and C_2 such that

$$\lim_{n\to\infty} \sup_{s_0+\delta\leq s<S_{\max}} \mathcal{L}(s) \leq \lim_{n\to\infty} C_1 n\log(n) - C_2 n^{1+2\delta/d-\epsilon} = -\infty.$$

From this it follows that $\limsup_{n\to\infty} \hat{s}_n \leq s_0$ almost surely, which is stronger than what is required for (22).

Consider then the case $d/2 < s \le s_0 - \delta$. By repeating the steps in (24), we obtain

$$(25) \qquad \sum_{i=1}^{n} \log \left(\frac{\mathbb{V}_{s_0}(x_i|\boldsymbol{x}^{i-1})}{\mathbb{V}_s(x_i|\boldsymbol{x}^{i-1})} \right) \lesssim (s_0 - \delta - s_0)n \log(n) = -\delta n \log(n).$$

Let $m_{\theta_0,n}$ denote the minimum norm interpolant based on the correct kernel $\sigma_0^2 K_{\theta_0}$. By norm equivalence and the minimum norm property, $\|m_{s_0,n}\|_{s_0} \le \|m_{\theta_0,n}\|_{s_0} \le C\|m_{\theta_0,n}\|_{\theta_0}$. Thus

$$u(\boldsymbol{x})^\top K_{s_0}(\boldsymbol{x})^{-1}u(\boldsymbol{x}) \le \sigma_0^{-2}Cu(\boldsymbol{x})^\top K_{\theta_0}(\boldsymbol{x})^{-1}u(\boldsymbol{x})$$

$$(26) \qquad = C\tilde{\boldsymbol{\xi}}_n^\top K_{\theta_0}(\boldsymbol{x})^{1/2}K_{\theta_0}(\boldsymbol{x})^{-1}K_{\theta_0}(\boldsymbol{x})^{1/2}\tilde{\boldsymbol{\xi}}_n \le C\sum_{i=1}^{n} \tilde{\xi}_{n,i}^2,$$

where $\{\tilde{\boldsymbol{\xi}}_n = (\xi_{n,1}, \ldots, \tilde{\xi}_{n,n})\}$ forms a triangular array of pairwise uncorrelated random variables with mean 0 and variance 1. The estimates (25) and (26) and Markov's inequality yield

$$\mathbb{P}\left(\sup_{d/2 < s \le s_0 - \delta} \mathcal{L}(s) \ge 0 \right) \le \mathbb{P}\left(-C_1 \delta n \log(n) + C_2 \sum_{i=1}^{n} \tilde{\xi}_{n,i}^2 - u(\boldsymbol{x})^\top K_s(\boldsymbol{x})^{-1}u(\boldsymbol{x}) \ge 0 \right)$$

$$\le \mathbb{P}\left(\sum_{i=1}^{n} \tilde{\xi}_{n,i}^2 \ge cn \log(n) \right) \le \frac{1}{c \log(n)} \to 0.$$

This concludes the proof of (22). $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

The assumption $s_0 > d$ was only used to prove $\mathbb{P}(\limsup_{n\to\infty} \hat{s}_n \ge s_0) = 1$. Thus if we relax the assumption to $s_0 > d/2$, we still obtain $\mathbb{P}(\hat{s}_n \le s_0) \to 0$. However, if we impose additional assumptions on the $\{\xi_i\}$ we can strengthen the statement of Theorem 4.8.

4.3. **The Gaussian Case.** In this section, we will discuss the particular case where the $\{\xi_i\}$ are standard Gaussian random variables, for concreteness and because of the importance of Gaussian processes in spatial statistics and other related fields. We first establish a path regularity result, similar to Lang and Schwab (2015, Theorem 4.6), which gives the result for the sphere $\mathcal{M} = \mathbb{S}^{d-1}$. A stronger result for the sphere is given in Lan et al. (2018). For various other path regularity results, see Henderson (2023); Meerschaert et al. (2013); Scheuerer (2010); Steinwart (2019).

Proposition 4.9. *Let $\{\xi_i\}$ be an i.i.d. sequence of standard Gaussian random variables. Let $l + \alpha < s_0 - d/2$ with $l \in \mathbb{N}$ and $0 < \alpha < 1$ and let u be defined as in (8). Then there is a version of u whose sample paths are almost surely elements of $C^{l,\alpha}(\mathcal{M})$.*

Proof. As in the proof of Theorem 4.6 in Lang and Schwab (2015), it suffices to establish the claim for $d/2 < s_0 \leq d/2+1$ because for any u and $1 < l \in \mathbb{N}$ such that $s_0 - l > d/2$ we have

$$\frac{1}{\sigma_0}u = \frac{v(\theta_0)}{v(\tilde{\theta}_0)}(\tau_0 - \Delta)^{-l/2} \sum_{i=1}^{\infty} v(\tilde{\theta}_0)(\tau_0 - \lambda_i)^{-s_0/2+l/2}\xi_i e_i =: \frac{v(\theta_0)}{v(\tilde{\theta}_0)}(\tau_0 - \Delta)^{-l/2}\tilde{u}.$$

and for integer l the operator $(\tau_0 - \Delta)^{-l/2}$ is an isomorphism from $C^{0,\alpha}$ to $C^{l,\alpha}$ for $0 < \alpha < 1$ by Theorem XI.2.5 in Taylor (1981). Moreover, $\tilde{u}$ is a mean-zero Gaussian process with covariance function $K_{\tilde{\theta}_0}$ and $d/2 < \tilde{s}_0 \leq d/2+1$ where $\tilde{s}_0$ is the smoothness parameter in $\tilde{\theta}_0 = (\tilde{s}_0, \tau_0)$. The claim is now an immediate consequence of the results in Kerkyacharian et al. (2018). More concretely, note that the results there apply to the Laplace–Beltrami operator on compact manifolds as mentioned in the Section 3 and the Examples on their page 283. Then apply first their Corollary 5.2, then their Theorem 4.6 and finally their Proposition 3.2. $\qquad\square$

To simplify the notation in what follows we assume that $s_0 - d/2 = l + \alpha$ with $0 < \alpha \leq 1$. It follows from Proposition 4.9 that in the Gaussian case we have in fact

$$u \in H_L^{s_0-d/2-\epsilon}(\mathcal{M}) \cap C^{l,\alpha-\epsilon}(\mathcal{M}) = H_L^{l+\alpha-\epsilon}(\mathcal{M}) \cap C^{l,\alpha-\epsilon}(\mathcal{M})$$

for all $\epsilon > 0$ small enough. That is, u possesses Sobolev and Hölder regularity of the same order. Thus one might hope to have the following analog of Proposition 4.5:

$$(27) \qquad\qquad \|u - m_{s,n}\|_0 \leq Cn^{-s_0/d+1/2+\epsilon/d}\|u\|_{C^{l,\alpha-\epsilon}(\mathcal{M})}$$

for all $s_0 > d/2$ when the sequence of points is quasi-uniform. That this is indeed the case on closed manifolds has recently been shown in Corollary 1 of Hangelbroek et al. (2018), which is based on results in Hangelbroek et al. (2010). However, their result applies only to even $s \in \mathbb{N}$. While it appears likely that results like (27) should hold for general $s > d/2$, the result in Hangelbroek et al. (2018) suffices for our purposes.

Corollary 4.10. *Let the family of kernels $\{K_s\}_{s>d/2}$ and the sequence of points $\{x_i\}$ be as in Theorem 4.8. Assume that the sequence $\{\xi_i\}$ used to define u in (8) is a sequence of i.i.d. standard Gaussian random variables and that $s_0 > d/2$. Let $\{\hat{s}_n\}$ be a sequence of maximizers of the Gaussian log-likelihood (21) in $(d/2, S_{\max})$. Then*

$$\hat{s}_n \to s_0 \qquad \text{almost surely.}$$

Proof. The conclusion is stronger than that of Theorem 4.8 in two ways: (i) $s_0 > d/2$ instead of $s_0 > d$ and (ii) the convergence is almost sure rather than in probability. The assumption $s_0 > d$ was used in the proof of Theorem 4.8 only to show that almost

surely

$$\|m_{s,n}\|_s^2 \geq n^{1+2(s-s_0)/d-\epsilon}$$

via Proposition 4.3, which is only needed to show that $\mathbb{P}(\limsup_{n\to\infty} \hat{s}_n \leq s_0) = 1$. Thus it suffices to show that in the Gaussian case the assumption for Proposition 4.3 is fulfilled for $s_0 > d/2$. For this purpose, let b be a natural number such that $2b > \max\{S_{\max}, s_0\}$ and let $\tilde{m}_{2b,n}$ denote the minimum norm interpolant from $\mathcal{H}^{2b}$. Note that because $u(x_i) = m_{s,n}(x_i)$ for all $1 \leq i \leq n$, it follows that $\tilde{m}_{2b,n}$ is the minimum norm interpolant for both u and $m_{s,n}$. Then Proposition 4.5 and Corollary 1 in Hangelbroek et al. (2018) give

$$\|u - m_{s,n}\|_0 \leq \|u - \tilde{m}_{2b,n}\|_0 + \|\tilde{m}_{2b,n} - m_{s,n}\|_0$$
$$\leq C_1 n^{-s_0/d+1/2+\epsilon/d}\|u\|_{C^{k,\alpha-\epsilon}(\mathcal{M})} + C_2 n^{-s/d}\|m_{s,n}\|_s$$

for constants $C_1, C_2 > 0$ that do not depend on n. Thus the assumptions of Proposition 4.3 are fulfilled. Repeating the proof of Theorem 4.8 yields $\mathbb{P}(\limsup_{n\to\infty} \hat{s}_n \leq s_0) = 1$.

It remains to prove $\mathbb{P}(\liminf_{n\to\infty} \hat{s}_n \geq s_0) = 1$. For this, note that the estimate (19) in Proposition 4.6 shows that almost surely there exists a positive constant C such that

$$u(\boldsymbol{x})^\top K_{s_0}(\boldsymbol{x})^{-1} u(\boldsymbol{x}) \leq Cn.$$

Repeating the same steps as in the proof of Theorem 4.8 yields that, almost surely,

$$\sup_{d/2<s\leq s_0-\delta} \mathcal{L}(s) \leq -C_1\delta n \log(n) + C_2 n - u(\boldsymbol{x})^\top K_s(\boldsymbol{x})^{-1} u(\boldsymbol{x}) \lesssim -n\log(n) \to -\infty,$$

which implies $\mathbb{P}(\liminf_{n\to\infty} \hat{s}_n \geq s_0) = 1$ and thus completes the proof. $\qquad\square$

4.4. **Estimation Using Other Kernels.** In this section we discuss some kernels to which Theorem 4.8 applies. Of course, the theorem applies to the correctly specified kernels $\sigma^2 K_\theta$ from Section 2, as long as the corresponding norms fulfill the norm monotonicity property; see Lemma 2.2. Although these kernels have many nice theoretical properties, their evaluation requires the calculation of an infinite sum. Furthermore, the eigenvalue decomposition of the Laplace–Beltrami operator is only known in certain special cases, e.g. on the sphere or the torus. Therefore, even obtaining an approximate evaluation of K_θ can be hopeless. Another aspect of practical importance is that the matrices $K_\theta(\boldsymbol{x})$ are dense, so that the evaluation of the Gaussian likelihood for large n is computationally intensive. Therefore, it would be desirable to find other classes

$\{K_s\}_{s>d/2}$ of kernels that have good practical properties and allow a consistent estimation of s_0. The crucial property that these kernels have to satisfy is that their RKHS are equal to H_L^s up to equivalent norms.

One class of kernels we consider here are kernels from the ambient space $\mathbb{R}^k$. Such kernels were applied to function approximation on closed manifolds in Fuselier and Wright (2012) and some of the results in this section are essentially from their work. For manifolds $\mathcal{M} \subset \mathbb{R}^k$ it is quite easy to arrive at positive definite kernels by using kernels from the ambient space $\mathbb{R}^k$. Specifically, if $\tilde{K}$ is a positive definite kernel on $\mathbb{R}^k \times \mathbb{R}^k$, then its restriction is

$$K(\cdot, \cdot) = \tilde{K}(\cdot, \cdot)|_{\mathcal{M} \times \mathcal{M}}$$

is of course positive definite as well. The RKHS of a particular class of kernels $\tilde{K}$ were identified in Fuselier and Wright (2012).

Theorem 4.11 (Theorem 2 in Fuselier and Wright, 2012). *Let $\mathcal{M} \subset \mathbb{R}^k$ be a d-dimensional closed manifold. Let $\tilde{K}_s(x, y) = \Phi_s(\|x - y\|)$ be a positive definite kernel on $\mathbb{R}^k$ such that the Fourier transform of Φ_s satisfies*

(28) $$\hat{\Phi}_s(\xi) \asymp (1 + \|\xi\|^2)^{-(s+(k-d)/2)}.$$

Then the RKHS $\mathcal{H}^s$ with norm $\|\cdot\|_s$ of the restriction $K_s(\cdot, \cdot) = \tilde{K}_s(\cdot, \cdot)|_{\mathcal{M} \times \mathcal{M}}$ is H_L^s up to equivalent norms.

It is also known that the kernel $\tilde{K}_s$ is the kernel for an RKHS equal to $H^{s+(k-d)/2}(\mathbb{R}^k)$ up to equivalent norms (Wendland, 2004, Chapter 10). In Theorem 4.8 it is also assumed that the norms fulfil the monotonicity property (20). Fortunately, this is inherited from the norm monotonicity property of the original kernel. The following result is a direct consequence of Fuselier and Wright (2012, Lemma 4), which in turn is a reformulation of the results in Schaback (1999, Section 9), which can be found also in Wendland (2004, Theorem 10.46 and Corollary 10.47). See the Appendix, Proposition A.3 for a proof.

Proposition 4.12. *Let $\tilde{K}_s(\cdot, \cdot)$ be as in Theorem 4.11. Let $\mathcal{H}^s(\mathbb{R}^k)$ be the RKHS of $\tilde{K}_s$ and $\mathcal{H}^s$ the RKHS of $K_s(\cdot, \cdot) = \tilde{K}_s(\cdot, \cdot)|_{\mathcal{M} \times \mathcal{M}}$. Let $s \leq s'$. If $\|f\|_{\mathcal{H}^s(\mathbb{R}^k)} \leq C\|f\|_{\mathcal{H}^{s'}(\mathbb{R}^k)}$ for all $f \in \mathcal{H}^{s'}(\mathbb{R}^k)$ and some C independent of s, then $\|f\|_s \leq C\|f\|_{s'}$ for all $f \in \mathcal{H}^s$.*

The prime examples of such kernels are the isotropic Matérn kernels on $\mathbb{R}^k$ given in (1). The fact that their RKHS have monotone norms is shown for example in Karvonen (2023, Lemma 3.9).

Another class of kernels satisfying (28) are the so-called generalized Wendland functions

$$\Phi_\kappa(\|x-y\|) = \Phi_\kappa(r) = \begin{cases} \dfrac{\sigma^2 2^{1-\kappa}}{\Gamma(\kappa)} \displaystyle\int_{r/\beta}^1 u\left(u^2 - \dfrac{r^2}{\beta^2}\right)^{\kappa-1} (1-u)^\mu \, \mathrm{d}u & \text{if} \quad 0 < \dfrac{r}{\beta} < 1, \\ 0 & \text{if} \quad \dfrac{r}{\beta} \geq 1; \end{cases}$$

see Bevilacqua et al. (2019); Gneiting (2002); Hubbert and Jäger (2023). Here β and σ^2 are positive and $\mu \geq (d+1)/2+\kappa$, while $\Gamma(\cdot)$ is the gamma function. For $s = \kappa+(d+1)/2$ the RKHS of Φ_κ is equal to $H^s(\mathcal{M})$ (Bevilacqua et al., 2019, Theorem 1). Generalized Wendland functions, in a slightly different form, are studied as covariance functions for Gaussian processes in Bevilacqua et al. (2019). The difference is irrelevant for our purposes. The advantage of the generalized Wendland functions is that they are compactly supported and thus yield sparse kernel matrices. The monotonicity of the corresponding norms is proved in Proposition A.3.

If $\mathcal{M} = \mathbb{S}^{d-1}$ is the sphere, the restriction of the generalized Wendland functions to $\mathcal{M}$ was investigated by Hubbert and Jäger (2023). Further covariance functions on the sphere with explicit smoothness parameters are presented in Gneiting (2013) and Guinness and Fuentes (2016).

4.5. **Estimation of the Magnitude Parameter.** The smoothness parameter s is fixed to a largely arbitrary value in many applications. Here we consider a setting in which the model smoothness $s \geq s_0$ is fixed but the *magnitude parameter* σ is estimated via maximum likelihood. Define therefore $K_{s,\sigma} = \sigma^2 K_s$. Straightforward differentiation of the Gaussian log-likelihood (9) with respect to σ then yields the maximum likelihood estimator

$$\hat{\sigma}_n^2 = \frac{u(\boldsymbol{x})^\top K_s(\boldsymbol{x})^{-1} u(\boldsymbol{x})}{n} = \frac{\|m_{s,n}\|_s^2}{n}.$$

It is now an immediate consequence of Propositions 4.3 to 4.6 that $\hat{\sigma}_n^2$ blows up with a polynomial rate that depends on the extent, $s - s_0$, to which the model oversmooths the truth.

Corollary 4.13. *Let K_s be a kernel whose RKHS $\mathcal{H}^s$ is equal to H_L^s up to equivalent norms and let u be defined as in (8) with $s \geq s_0 > d$. Assume that the sequence of points $\{x_i\}$ is quasi-uniform. Then almost surely*

$$c\, n^{2(s-s_0)/d-\epsilon} \leq \hat{\sigma}_n^2 \leq C n^{2(s-s_0)/d+\epsilon}$$

for any $\epsilon > 0$, where the constants $c, C > 0$ depend only on s, θ_0, $\mathcal{M}$, ϵ and the sample path.

If u is a Gaussian process, an improved version of this corollary can be obtained by proceeding as in Section 4.3. According to Proposition 4.5, the minimum-norm interpolant $m_{s,n}$ with misspecified smoothness $s \geq s_0$ tends to u at essentially the same rate $n^{-s_0/d+1/2}$ as $m_{s_0,n}$. Note that the estimation of the magnitude does not affect the minimum norm of the interpolant. Corollary 4.13 is useful because it shows that a computationally simple magnitude estimate correctly compensates for the misspecification of smoothness and ensures that the conditional variance is a reliable indicator of predictive uncertainty. Proposition 4.7 and Corollary 4.13 imply that for quasi-uniform points the correct variance $\mathbb{V}_{s_0}$ and the variance $\mathbb{V}_{s,n} = \hat{\sigma}_n^2 \mathbb{V}_s$ for the kernel $K_{s,\sigma}$ with the plug-in maximum likelihood estimator $\sigma = \hat{\sigma}_n$ fulfill

$$\sup_{x \in \mathcal{M}} \mathbb{V}_{s_0}(x|\boldsymbol{x}) \asymp n^{-2s_0/d+1} \quad \text{and} \quad n^{-2s_0/d+1-\epsilon} \lesssim \sup_{x \in \mathcal{M}} \mathbb{V}_{s,n}(x|\boldsymbol{x}) \lesssim n^{-2s_0/d+1+\epsilon}.$$

This means that $\mathbb{V}_{s,n}$ decays at the essentially "correct" rate.

When smoothness is well specified (i.e. $s = s_0$), a number of consistency and asymptotic normality results for magnitude estimation in Matérn and related models on Euclidean domains have been obtained in the last 30 years (e.g., Kaufman and Shaby, 2013; Ying, 1991; Zhang, 2004). Li et al. (2023) considered essentially the same setting as in this work and established consistency of the microergodic parameter in the Gaussian case with correctly specified kernel. A corresponding asymptotic normality result of the form (assuming $d \leq 3$)

$$(29) \qquad\qquad \sqrt{n}(\hat{\sigma}_n^2 v(\hat{\theta}_n) - \sigma_0^2 v(\theta_0)) \to \mathcal{N}(0, 2\sigma_0^4 v(\theta_0)^2)$$

can be derived with known arguments, which are used for example in Bolin et al. (2023); Du et al. (2009); Kaufman and Shaby (2013); Wang and Loh (2011). A formal statement and proof can be found in the Appendix. The proof is essentially from Bolin et al. (2023, Proposition 3), where a different scenario is considered.

The results for misspecified smoothness are recent. Versions of Corollary 4.13 for the expected value of $\hat{\sigma}_n^2$ have been proved in Karvonen (2021) and Naslidnyk et al. (2023). Karvonen et al. (2020) proved the corollary for a class of self-similar functions on Euclidean domains, while Xu and Stein (2017) obtained rates for the magnitude estimator when the model has Gaussian covariance $K(x, y) = \exp(-(x - y)^2/(2\tau^2))$ for $\tau > 0$ and the observations are generated by a low-order monomial on the interval $[0, 1]$. When the observations are contaminated by noise, posterior contraction rates and the coverage of credible sets for a Brownian motion model with an estimated magnitude

parameter were investigated in Sniekers and van der Vaart (2015a,b, 2020) and Travis and Ray (2023).

5. Equivalence of Measures

In the previous section, we discussed briefly the estimation of the parameters τ and σ^2. Note that the central limit theorem in (29) is stated only for the product $\sigma^2 v(\theta)$, the so-called microergodic parameter. This is due to the well-known fact that for Gaussian processes, the parameters σ^2 and τ cannot be consistently estimated if the dimension $d \leq 3$. For bounded Euclidean domains, see for example Zhang (2004) and Bolin and Kirchner (2023); for closed manifolds, see Li et al. (2023). This is due to equivalence of the Gaussian measures describing the distribution of the processes. More concretely, in the context considered here the cited results state that if $d \leq 3$ and if θ_1 and θ_2 are such that $s_1 = s_2$ and

$$\sigma_1^2 v(\theta_1) = \sigma_2^2 v(\theta_2),$$

then the Gaussian measures $\mathcal{N}(0, \sigma_1^2 v(\theta_1)(\tau_1 - \Delta)^{s_1})$ and $\mathcal{N}(0, \sigma_2^2 v(\theta_2)(\tau_2 - \Delta)^{s_2})$ are equivalent. Otherwise they are orthogonal. The results in Bolin and Kirchner (2023) and Li et al. (2023) apply only to Gaussian measures.

However, as far as we know, there are no results for the non-Gaussian case. In this section, we will show that this continues to hold for non-Gaussian processes. To address this scenario, we will use a theorem of Kakutani (1948), for which the assumption that $\{\xi_i\}$ are independent is important. Here we briefly establish the necessary measure-theoretic framework with more details in B.

A measure m_1 is absolutely continuous with respect to a measure m_2, denoted by $m_1 \ll m_2$, if $m_2(B) = 0$ implies that $m_1(B) = 0$. Two measures m_1 and m_2 are equivalent if $m_1 \ll m_2$ and $m_2 \ll m_1$. Two probability measures are orthogonal if there is a set B for which $m_1(B) = 1$ and $m_2(B) = 0$. We start with the complete probability space $(\Omega, \mathcal{F}, \mathbb{P})$ on which the sequence $\{\xi_i\}$ is defined. To simplify the notation in this section, let

$$\lambda_i(\theta) = \sqrt{v(\theta)}(\tau + \lambda_i)^{-s/2}.$$

These are the eigenvalues of the operator $\sqrt{v(\theta)}(\tau - \Delta)^{-s/2}$.

Let $\{(\mathbb{R}, \mathcal{B}(\mathbb{R}), \mathbb{P}_{\sigma\lambda_i(\theta)})\}$ be a sequence of probability spaces with $\mathcal{B}(\mathbb{R})$ the Borel σ-algebra generated by the open sets, and $\mathbb{P}_{\sigma\lambda_i(\theta)}$ the distribution of

$$\sigma\sqrt{v(\theta)}(\tau + \lambda_i)^{-s/2}\xi_i = \sigma\lambda_i(\theta)\xi_i.$$

Consider the infinite direct product measure space $(\mathbb{R}^\infty, \mathcal{B}^\infty, \mathbb{P}_{\sigma,\theta})$, where $\mathbb{R}^\infty$ is the infinite Cartesian product, $\mathcal{B}^\infty$ is the smallest σ-algebra making all coordinate projections measurable and $\mathbb{P}_{\sigma,\theta}$ is the infinite product measure $\mathbb{P}_{\sigma,\theta} \coloneqq \bigotimes \mathbb{P}_{\sigma\lambda_i(\theta)}$.

Recall that $\hat{H}_L^0$ is the Hilbert space of square-summable sequences. One can show $\hat{H}_L^0$ is $\mathcal{B}^\infty$-measurable and that the σ-algebra $\mathcal{B}(\hat{H}_L^0)$ generated by the open sets in $\hat{H}_L^0$ in fact equals $\{A \cap \hat{H}_L^0 \mid A \in \mathcal{B}^\infty\}$, see Proposition B.2. Note that $\mathbb{P}_{\sigma,\theta}(\hat{H}_L^0) = 1$ for $s > d/2$ by Proposition 3.1. Thus we may understand each measure $\mathbb{P}_{\sigma,\theta}$ as a measure on the space $(\hat{H}_L^0, \mathcal{B}(\hat{H}_L^0))$. They induce measures on $(L_2(\mathcal{M}), \mathcal{B}(L_2(\mathcal{M})))$ via $P_{\sigma,\theta} \circ \iota^{-1}$ as ι^{-1} is an isomorphism. Of course, the measures $\mathbb{P}_{\sigma,\theta} \circ \iota^{-1}$ are the distributions of the random elements u defined in (6), see Proposition B.3.

Given two probability measures m_1 and m_2, let λ be a measure such that $m_1 \ll \lambda$ and $m_2 \ll \lambda$. Such a measure always exists (e.g., $\lambda = m_1 + m_2$). The Hellinger distance between m_1 and m_2 is

$$\rho(m_1, m_2) = \int \sqrt{\frac{\mathrm{d}m_1}{\mathrm{d}\lambda}} \sqrt{\frac{\mathrm{d}m_2}{\mathrm{d}\lambda}} \mathrm{d}\lambda,$$

where $\mathrm{d}m_1/\mathrm{d}\lambda$ and $\mathrm{d}m_2/\mathrm{d}\lambda$ are Radon–Nikodym derivatives. It is known that ρ does not depend on the choice of λ (Kakutani, 1948, Section 4). Moreover, ρ has the following properties: (i) $0 < \rho(m_1, m_2) \le 1$, (ii) $\rho(m_1, m_2) = 1 \iff m_1 = m_2$, and (iii) $\rho(m_1, m_2) = \rho(m_2, m_1)$ (Kakutani, 1948, p. 215). Kakutani's theorem (Kakutani, 1948, p. 218) links equivalence or orthogonality of the infinite product measures $\mathbb{P}_{\sigma,\theta}$ and $\mathbb{P}_{\sigma',\theta'}$ to the behaviour of $\rho(\mathbb{P}_{\sigma\lambda_i(\theta)}, \mathbb{P}_{\sigma'\lambda_i(\theta')})$ as $i \to \infty$. These Hellinger distances have a particularly convenient form in the scenario under consideration. Concretely, let $\mathbb{P}_1$ denote the distribution of the ξ_i and assume it admits the density f with respect to the Lebesgue measure. Then a simple change of variables yields

$$\rho\left(\mathbb{P}_{\sigma\lambda_i(\theta)}, \mathbb{P}_{\sigma'\lambda_i(\theta')}\right) = \int_{\mathbb{R}} \frac{1}{\sqrt{\sigma\lambda_i(\theta)\sigma'\lambda_i(\theta')}} \sqrt{f\left(\frac{x}{\sigma\lambda_i(\theta)}\right) f\left(\frac{x}{\sigma'\lambda_i(\theta')}\right)} \, \mathrm{d}x$$

$$= \int_{\mathbb{R}} \frac{\sqrt{\sigma'\lambda_i(\theta')}}{\sqrt{\sigma\lambda_i(\theta)}} \sqrt{f(x) f\left(x \frac{\sigma'\lambda_i(\theta')}{\sigma\lambda_i(\theta)}\right)} \, \mathrm{d}x$$

$$= \rho\left(\mathbb{P}_{\sigma\lambda_i(\theta)/(\sigma'\lambda_i(\theta'))}, \mathbb{P}_1\right).$$

Thus, the Hellinger distance only depends on the ratio $\sigma\lambda_i(\theta)/(\sigma'\lambda_i(\theta'))$. This motivates the function $\varphi(a) = \rho(\mathbb{P}_a, \mathbb{P}_1)$, which has a global maximum $\varphi(1) = 1$. Sufficient regularity of the function φ implies $\varphi'(1) = 0$ and $\varphi''(1) < 0$ and so a Taylor expansion

shows that

$$0 \leq -\log(\varphi(a)) = -\frac{\varphi'(1)}{\varphi(1)}(a-1) - \left[\frac{\varphi''(a^*)}{\varphi(a^*)} - \left(\frac{\varphi'(a^*)}{\varphi(a^*)}\right)^2\right](a-1)^2$$

$$= \left[-\frac{\varphi''(a^*)}{\varphi(a^*)} + \left(\frac{\varphi'(a^*)}{\varphi(a^*)}\right)^2\right](a-1)^2$$

for some a^* such that $|a^* - 1| \leq |a - 1|$. Hence $-\log(\varphi(a))$ behaves like $(a-1)^2$ in a neighborhood of 1 if φ is sufficiently regular. Under this condition, we can provide a criterion for equivalence and orthogonality for two infinite product measures $\mathbb{P}_{\sigma_1^2, \theta_1}$ and $\mathbb{P}_{\sigma_2^2, \theta_2}$.

Theorem 5.1. *Let $\{\xi_i\}$ be a sequence of i.i.d. random variables such that $\mathbb{E}(\xi_1) = 0$ and $E(\xi_1^2) < \infty$. Assume that, for all $a, b > 0$ and each i, the distributions of $a\xi_i$ and $b\xi_i$ are equivalent measures and the distribution of ξ_i admits a density with respect to the Lebesgue measure. Moreover, assume that the function $\varphi(a) = \rho(\mathbb{P}_a, \mathbb{P}_1)$ satisfies*

(30) $$c(a-1)^2 \leq -\log(\varphi(a)) \leq C(a-1)^2$$

in a neighborhood of 1. If the parameters s_1, τ_1, σ_1^2 and s_2, τ_2, σ_2^2 are such that

(31) $$s_1 = s_2 \quad and \quad \sigma_1^2 v(\theta_1) = \sigma_2^2 v(\theta_2),$$

then the measures $\mathbb{P}_{\sigma_1^2, \theta_1}$ and $\mathbb{P}_{\sigma_2^2, \theta_2}$ are equivalent if $d \leq 3$ and orthogonal if $d > 3$. If (31) fails and if φ is uniformly bounded away from 1 outside of $(1 - \epsilon, 1 + \epsilon)$ for every $\epsilon > 0$, the measures $\mathbb{P}_{\sigma_1^2, \theta_1}$ and $\mathbb{P}_{\sigma_2^2, \theta_2}$ are orthogonal.

Proof. We start with a simple observation that, for a in a neighborhood of 1,

(32) $$c(a-1)^2 \leq \log(a)^2 \leq C(a-1)^2.$$

Kakutani's theorem states that $\mathbb{P}_{\sigma_1, \theta_1}$ and $\mathbb{P}_{\sigma_2, \theta_2}$ are equivalent if

(33) $$\sum_{i=1}^{\infty} -\log\left[\rho(\mathbb{P}_{\sigma_1 \lambda_i(\theta_1)}, \mathbb{P}_{\sigma_2 \lambda_i(\theta_2)})\right] < \infty$$

and orthogonal otherwise. Let

$$a_i = \frac{\sigma_1 \sqrt{v(\theta_1)}(\tau_1 + \lambda_i)^{-s_1/2}}{\sigma_2 \sqrt{v(\theta_2)}(\tau_2 + \lambda_i)^{-s_2/2}}.$$

Under the assumptions in (31),

(34) $$a_i = \left(\frac{\tau_1 + \lambda_i}{\tau_2 + \lambda_i}\right)^{-s_1/2} \xrightarrow{i \to \infty} 1.$$

Thus, combining (30), (32) and (34), we obtain

$$-\log\left[\rho(\mathbb{P}_{\sigma_1\lambda_i(\theta_1)}, \mathbb{P}_{\sigma_2\lambda_i(\theta_2)})\right] = -\log(\varphi(a_i)) \asymp \log\left(\frac{\tau_2 + \lambda_i}{\tau_1 + \lambda_i}\right)^2 \asymp \left(\frac{\tau_2 - \tau_1}{\tau_1 + \lambda_i}\right)^2 \asymp i^{-4/d},$$

where the last $\asymp$ follows from Weyl's law (3). Thus (33) holds if and only if $\sum_{i=1}^{\infty} i^{-4/d} < \infty$. This sum is finite if and only if $d < 4$. To show that φ being uniformly bounded away from 1 outside of every neighborhood of 1 implies orthogonality of the resulting measures, note that

$$a_i \to 0 \quad \text{if } s_1 > s_2, \quad a_i \to \infty \quad \text{if } s_1 < s_2 \quad \text{and} \quad s_i \to \sigma_1^2 v(\theta_1)/\sigma_2^2 v(\theta_2) \quad \text{else.}$$

Thus, if (31) does not hold, there exist δ and N such that $\sum_{i=N}^{\infty} -\log(\varphi(a_i)) \geq \sum_{i=N}^{\infty} \delta.$
$\square$

We have assumed that the distributions of $a\xi_i$ and $b\xi_i$ are equivalent measures for each positive a and b. If this is not the case, one can easily see that the resulting infinite product measures cannot be equivalent, see Kakutani (1948, p. 218). Note that this restricts the distributions that can lead to equivalent measures. For example, if the distribution ξ_i has a Lebesgue density that vanishes on a set with positive Lebesgue measure, then the distributions $a\xi_i$ and $b\xi_i$ need not be equivalent measures for all a and b. However, this does not imply that the resulting infinite product measures are necessarily orthogonal. This means that a consistent estimate may still not be possible. For example, let $\mathbb{P}_{\sigma_1,\theta_1}$ and $\mathbb{P}_{\sigma_2,\theta_2}$ be non-orthogonal measures, even though $\sigma_1 \neq \sigma_2$. Argue as in Zhang (2004) and assume that $\{\hat{\sigma}_n\}$ is a sequence of estimators that converge to σ_1 in probability under $\mathbb{P}_{\sigma_1,\theta_1}$. Then there exists a subsequence $\{\hat{\sigma}_{k_n}\}$ such that

$$\mathbb{P}_{\sigma_1,\theta_1}\left(\lim_{k_n \to \infty} \hat{\sigma}_{k_n} = \sigma_1\right) = 1.$$

Because $\mathbb{P}_{\sigma_2,\theta_2}$ is not orthogonal to $\mathbb{P}_{\sigma_1,\theta_1}$, we have

$$\mathbb{P}_{\sigma_2,\theta_2}\left(\lim_{k_n \to \infty} \hat{\sigma}_{k_n} = \sigma_1\right) > 0$$

and so under $\mathbb{P}_{\sigma_2,\theta_2}$ consistent estimation of σ_2 may still not occur.

To show that two measures m_1 and m_2 are not orthogonal, it suffices to show that either $m_1 \ll m_2$ or $m_1 \ll m_2$. A simple deduction from the proofs in Kakutani (1948) gives a condition for $\mathbb{P}_{\sigma_1,\theta_1} \ll \mathbb{P}_{\sigma_2,\theta_2}$. This is essentially part of Kakutani's theorem.

Corollary 5.2. *Consider the setup of Theorem 5.1 but assume only that the distribution of $a\xi_i$ is absolutely continuous with respect to $b\xi_i$ if $a \leq b$. Assume (31) holds and $d \leq 3$. Moreover, let θ_1 and θ_2 be such that $1 \leq \tau_1 \leq \tau_2$. Then $\mathbb{P}_{\sigma_1,\theta_1} \ll \mathbb{P}_{\sigma_2,\theta_2}$.*

Proof. This is a consequence of Lemma 5 in Kakutani (1948) and the proof of Theorem 5.1. □

Examples 5.3. *We give three examples for the distributions of ξ_i:*

(1) *In the standard Gaussian case we recover the well-known results because a simple computation gives $\rho(\mathbb{P}_a, \mathbb{P}_1) = \sqrt{2a/(1 + a^2)}$ and it is straightforward to see that (30) holds.*

(2) *More generally, it suffices to show that φ is twice continuously differentiable to obtain (30), which can be shown to be the case if ξ_i has t-distribution with degrees of freedom larger than 2 in order to guarantee finite second moments.*

(3) *An example that pertains to Corollary 5.2 is $\xi_i + 1 \sim \mathrm{Exp}(1)$, where $\mathrm{Exp}(1)$ is the exponential distribution with scale 1. A simple computation shows that $\varphi(a) = 2\sqrt{a}/(1 + a)$ and so (30) can be shown to hold.*

We note here that while the independence assumption is crucial for the use of Kakutani's theorem, the assumption that $\{\xi_i\}$ are identically distributed may be relaxed. The proofs remain essentially unchanged if additional technical assumptions are imposed.

6. RESULTS FOR THE DIRICHLET LAPLACIAN

Some of our results can be directly applied to Laplacian operators on bounded Euclidean domains augmented with boundary conditions. They are frequently studied in connection with Markov random field approximations of Matérn type fields (Lindgren et al., 2011). As a concrete example, we take the Dirichlet Laplacian operator and present some of the results from the Sections 2 to 5. Most of the proofs are analogous, so most of the time we only provide sketches or point out where previously presented results need to be adapted.

In the following we consider an open and connected $M \subset \mathbb{R}^d$ with Lipschitz boundary ∂M (Stein, 1970, p. 189), which satisfies a *interior cone condition* (Wendland, 2004, Definition 3.6). Let $\bar{M} = M \cup \partial M$ denote the closure of M. For such a set M we define the Sobolev spaces $H^s(M)$ as

$$H^s(M) := \left\{ f \in L_2(M) \mid \text{ there is } g \in H^s(\mathbb{R}^d) \text{ such that } g|_M = f \right\},$$

which is equipped with the norm $\|f\|_{H^s(M)} = \inf\{\|g\|_{H^s(M)} \mid g \in H^s(\mathbb{R}^d) \text{ s.t. } g|_M = f\}$. Note that $H^{s'}(M) \subset H^s(M) \subset L_2(M)$ if $s' \geq s \geq 0$. The Hilbert space $L_2(M)$ and the Hölder space $C^{k,\alpha}(\bar{M})$ are defined analogously to the spaces in Section 2.

For M we consider the usual Laplacian operator

$$-\Delta = -\sum_{i=1}^{d} \frac{\partial^2}{\partial^2 x_i}.$$

However, to obtain an eigendecomposition to define (8) or to ensure that the SPDE in (7) has a unique solution, we need to augment the Laplacian with boundary conditions. We consider Dirichlet boundary conditions and denote the resulting operator by Δ_D. It is known that there is a basis for $L_2(M)$ consisting of eigenfunctions e_i of $-\Delta_D$ with eigenvalues λ_i. The functions e_i can be chosen so that they are infinitely differentiable on M and so that $e_i|_{\partial M} \equiv 0$. Moreover, Weyl's law (3) holds, that is $ci^{2/d} \le \lambda_i \le Ci^{2/d}$. See Section 6 in Davies (1995), in particular Corollary 6.2.2 and Theorem 6.2.3. We can thus define the spaces

$$H_{L_D}^s = \left\{ f \in L_2(M) \mid (1 - \Delta_D)^{s/2} f \in L_2(M) \right\},$$

$$\hat{H}_{L_D}^s = \left\{ \{a_i\} \subset \mathbb{R} \ \bigg| \ \sum_{i=1}^{\infty} a_i^2 (1 + \lambda_i)^s < \infty \right\},$$

with corresponding norms $\|\cdot\|_\theta$ and $\|\cdot\|_{\hat{\theta}}$ for some $\tau > 0$. These spaces are analogous to the spaces H_L^s and $\hat{H}_L^s$ that we defined in Section 2.2.1.

The spaces $H_{L_D}^s$ have been characterized in Lemma 4.3 of Bolin and Kirchner (2023) for wide range of values of s. Define $\mathfrak{C} = \{s \in \mathbb{R} \mid s \ne 2k + 1/2 \text{ for some } k \in \mathbb{N}\}$. For any $s \notin \mathfrak{C}$ we have

$$(35) \qquad H_{L_D}^s = \left\{ f \in H^s(M) \ \bigg| \ (1 - \Delta)^j f = 0 \text{ in } L^2(\partial M) \text{ for all } j \le \left\lfloor \frac{2s - 1}{4} \right\rfloor \right\},$$

and the norm $\|\cdot\|_\theta$ is equivalent to $\|\cdot\|_{H^s(M)}$ on $H_{L_D}^s$. This presents the key difference to the manifold case: rather than equal to the Sobolev spaces, the spaces $H_{L_D}^s$ are now proper subspaces. This makes the analysis more difficult. However, we do retain norm equivalence.

Denote the resulting kernel of $H_{L_D}^s$ by K_θ^D. Our assumptions on M guarantee that the Sobolev imbedding theorem holds (Aubin, 1998, Theorem 2.33) and thus $H^s(M) \subset C^{0,\alpha}(\bar{M})$ for all $\alpha > s - d/2$ and the imbedding is compact. Thus, Proposition 2.3 is essentially unchanged, we only need $x_i \notin \partial M$ to guarantee strict positive definiteness. The results in Section 5 apply without changes. Let u now be defined as

$$(36) \qquad u(x) = \sigma_0 \sqrt{v(\theta_0)} \sum_{i=1}^{\infty} (\tau_0 + \lambda_i)^{-s_0/2} \xi_i e_i(x),$$

where $\{(\lambda_i, e_i)\}$ are the eigenpairs of Δ_D.

Unfortunately, we cannot show that all results in Section 4 hold in the same generality. First, we need to slightly adapt our sampling assumption. Define

$$\tilde{q}_{\boldsymbol{x}} = \min\left\{\frac{1}{2}\mathrm{dist}(\boldsymbol{x}, \partial M), q_{\boldsymbol{x}}\right\} \qquad \text{and} \qquad h_{\boldsymbol{x}} = \sup_{x \in M} \min_{1 \le i \le n} \|x_i - x\|,$$

where $q_{\boldsymbol{x}} = \min_{i \ne j}\|x_i - x_j\|/2$ and $h_{\boldsymbol{x}}$ are analogous to the quantities in (10). The difference is that we now have to work with $\tilde{q}_{\boldsymbol{x}}$ to make sure that the points are far enough away from the boundary ∂M. Set $\tilde{\rho}_{\boldsymbol{x}} = \tilde{q}_{\boldsymbol{x}}/h_{\boldsymbol{x}}$. We then assume that $\tilde{q}_{\boldsymbol{x}} \asymp h_{\boldsymbol{x}} \asymp n^{-1/d}$ and call points that fulfil this condition quasi-uniform.

Then the results of Propositions 4.2 and 4.3 apply unchanged. However, the result of the Proposition 4.5 must be changed. Firstly, we need $s \notin \mathfrak{C}$. Also, at the moment we cannot show the analogue of (17),

$$\|u - m_{s,n}\|_0 \le h_{\boldsymbol{x}}^{s_0 - d/2 - \epsilon} \tilde{\rho}_{\boldsymbol{x}}^{s - s_0 + d/2 + \epsilon} \|u\|_{H_{L_D}^{s_0 - d/2 - \epsilon}}$$

for the minimum norm interpolant $m_{s,n}$ from $H_{L_D}^s$. However, we can show the following.

Proposition 6.1. *Let $m_{s,n}$ be the minimum norm interpolant from an RKHS $\mathcal{H}_D^s$ that is equal to $H_{L_D}^s$ up to equivalent norms. Let u be defined as in (36) with $s_0 > d$ and $s \ge s_0$, $s \notin \mathfrak{C}$. Then almost surely for every $0 < \epsilon < s_0 - d/2$ such that $s_0 - d/2 - \epsilon \notin \mathfrak{C}$ we have*

$$\|u - m_{s,n}\|_0 \le C h_{\boldsymbol{x}}^{s_0 - d/2 - \epsilon} \tilde{\rho}_{\boldsymbol{x}}^{s - s_0 + d/2 + \epsilon} \|u\|_{s_0 - d/2 - \epsilon} + C h_{\boldsymbol{x}}^s \|m_{s,n}\|_s.$$

Proof. Let $\tilde{m}_{s,n}$ be the minimum norm interpolant from $H^s(M)$. Since $m_{s,n} \in H^s(M)$, $\tilde{m}_{s,n}$ is also the minimum norm interpolant for $m_{s,n}$. Using the escape result from Narcowich et al. (2006, Theorem 4.2) and norm equivalence, we have

$$\begin{aligned}
\|u - m_{s,n}\|_0 &\le \|u - \tilde{m}_{s,n}\|_0 + \|\tilde{m}_{s,n} - m_{s,n}\|_0 \\
&\le C h_{\boldsymbol{x}}^{s_0 - d/2 - \epsilon} \tilde{\rho}_{\boldsymbol{x}}^{s - s_0 + d/2 + \epsilon} \|u\|_{H^{s_0 - d/2 - \epsilon}(M)} + C h_{\boldsymbol{x}}^s \|m_{s,n}\|_{H^s(M)} \\
&\le C h_{\boldsymbol{x}}^{s_0 - d/2 - \epsilon} \tilde{\rho}_{\boldsymbol{x}}^{s - s_0 + d/2 + \epsilon} \|u\|_{s_0 - d/2 - \epsilon} + C h_{\boldsymbol{x}}^s \|m_{s,n}\|_s. \qquad \square
\end{aligned}$$

As long as $s \notin \mathfrak{C}$, the result of Proposition 4.6 holds without change if we replace H_L^s by $H_{L_D}^s$. Finally, the results of Proposition 4.7 are well known in the Euclidean case (e.g., Karvonen, 2023, Proposition 3.6 and 3.7). We only need $s \notin \mathfrak{C}$ to translate these results to our setting.

Theorem 6.2. *Let u be defined as in (36) with $s_0 > d$. Let $\{K_s\}_{s > d/2}$ be any family of kernels such that the RKHS of K_s is equal to $H_{L_D}^s$ up to equivalent norms and such that the corresponding norms are monotone. Assume that the sequence of points $\{x_i\}$*

is quasi-uniform. If $\{\hat{s}_n\}$ is a sequence of maximizers of the Gaussian log-likelihood

$$\ell(s; u(\boldsymbol{x})) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log(\det(K_s(\boldsymbol{x}))) - \frac{1}{2}u(\boldsymbol{x})^\top K_s(\boldsymbol{x})^{-1}u(\boldsymbol{x})$$

in $(d/2, S_{\max})$ with $s_0 \leq S_{\max}$, then

$$\hat{s}_n \to s_0 \quad \text{in probability.}$$

Proof. The proof is the same as of that of Theorem 4.8, the only change being that Proposition 6.1 is used in the place of Proposition 4.5. Note that the restriction $s \notin \mathfrak{C}$ in (35) is not a problem. Because $u \in H_L^{s_0 - d/2 - \epsilon}$ for all $\epsilon > 0$, we can choose δ and ϵ appearing in the proof of Theorem 4.8 in such a way that we avoid any problems. $\square$

While the path regularity result in the Gaussian case can be proved in the same way as in Proposition 4.9, extending Corollary 6.2 to the case $s_0 > d/2$ is less straightforward as the results in Hangelbroek et al. (2018) are proved only for closed manifolds. However, almost sure convergence can be shown in the Gaussian case.

One should be able to establish similar results for von Neumann boundary conditions. While we are not currently aware of results completely characterizing the corresponding spaces H_L^s for $s \geq 0$, there are results in Kim (2020) which can probably be straightforwardly extended with a proof similar to that of Lemma 4.3 in Bolin and Kirchner (2023).

7. SIMULATION STUDY

We conduct simulation studies to assess finite sample properties and stability against model misspecification for smoothness estimation. As underlying manifold, we use the sphere $\mathbb{S}^2$. In this case, the spectral decomposition of the covariance operators is given as

$$(\tau - \Delta)^{-s} = \sum_{l=0}^{\infty}(\tau + \lambda_l)^{-s}\sum_{m=-l}^{l}e_{lm}\otimes e_{lm},$$

where $\lambda_l = l(l+1)$ and the functions e_{lm} are the so-called real spherical harmonics for which explicit expressions are available (Stein and Weiss, 1971, Chapter IV). Even though in this case the eigenfunctions and eigenvalues are known, in practice we must truncate the infinite sums in order to compute the kernel K_θ of $(\tau - \Delta)^s$ and sample the corresponding random fields. With the addition theorem for spherical harmonics (Stein and Weiss, 1971, Chapter IV, Corollary 2.9), it is straightforward to show that the error incurred on the diagonal of the kernel when truncating at l is of order $l^{2(1-s)}$. An application of the Cauchy–Schwarz inequality then shows that the same bounds holds

for the off-diagonal elements as well. Let K_θ stand for the truncated kernel. Because the addition theorem implies that the truncated kernel is constant as a function of x for any truncation level l, we may normalize the kernel such that $K_\theta(x) = K_\theta(x, x) \equiv 1$. In our simulation studies we chose $l = 100$. That is, we use the first $\sum_{i=0}^{100}(2i+1) = 10\,201$ eigenfunctions and eigenvalues. With s chosen as in the simulation setup below, this results in a uniform pointwise error for the kernel of order 10^{-16}.

We consider the following three cases of model misspecification: (i) we use Gaussian observations and the correct (truncated) kernel but with incorrect τ (range scale) parameters, (ii) we use non-Gaussian observations, and (iii) we estimate the smoothness using Eucliean Matérn kernels and Euclidean generalized Wendland kernels restricted to the sphere even though the observations are generated using the (truncated) kernel K_{θ_0}. For scenario (ii), we always standardize the Karhunen–Loève coefficients to have mean 0 and variance 1. In the last scenario (iii), the other kernel parameters have to be specified. In an attempt to isolate the effect of the kernel, we chose the other parameters before optimization by minimizing the difference to the true covariance. This works very well for Euclidean Matérn kernels but slightly less well for the generalized Wendland kernels.

To assess the effect of sample size, we consider sample sizes $n = 500, 1000, 2000$. For each sample size, we generate 100 samples of a Gaussian process with true parameters $\theta_0 = (s_0, \tau_0) = (5, 20)$ and 100 samples of non-Gaussian processes for scenario (ii) in which we use the true range scale parameter during smoothness estimation. The range over which we optimize is fixed to $[1 + 10^{-7}, 30]$. As our theoretical results require quasi-uniformity, we use nearly equidistant points on the sphere generated by the regular placement algorithm given in Deserno (2004). To maximize the Gaussian log-likelihood, we use the implementation of the SHGO global optimization algorithm from the Python library `scipy`. When using simpler (local) optimizers, we found that they could on occasion get stuck in local minima, especially for larger sample sizes, perhaps due to numerical instability of the likelihood evaluation. A summary of the resulting estimates for the different scenarios can be found in the violin plots in Figures 1–3. Overall, we find that smoothness estimation appears to work well even for misspecified models.

First, from Figure 1, we note that estimation appears to be stable against distribution misspecification of the Karhunen–Loève coefficients $\{\xi_i\}$. When the coefficients are Bernoulli distributed, estimation appears to work even better than in the case of a correctly specified model (compare with Figure 3b). For exponentially and t distributed

coefficients, the estimation of s_0 appears to work similarly well. In all cases, the bias appears to be very small and decreases with increasing sample size.

Next, from Figure 2, we note that misspecifying the range scale parameter results in biased smoothness estimates for finite sample sizes. Moreover, there is clear positive correlation between range scale parameter and estimated smoothness parameter. In our opinion, the bias is not very large relative to range scale misspecification. Concretely, a normalized kernel with parameter $\theta = (s, \tau) = (5, 1)$ is somewhat pathological as the correlation is never smaller than 0.975 across the entire sphere while with $\tau_0 = 20$ the correlation is essentially 0 at large distances. However, even when τ is not grossly misspecified, the bias appears to decrease only slowly as the sample size increases.

Finally, smoothness estimation also works well with misspecified kernels, see Figure 3. In this setting the Euclidean Matérn kernel appears to be superior to the generalized Wendland kernel, perhaps due to the fact that the true kernel is more closely related to the Matérn kernel. It is not clear whether the biases present in the estimation results are inherent to the kernel used or whether they are the result of the specification of the remaining kernel parameters.
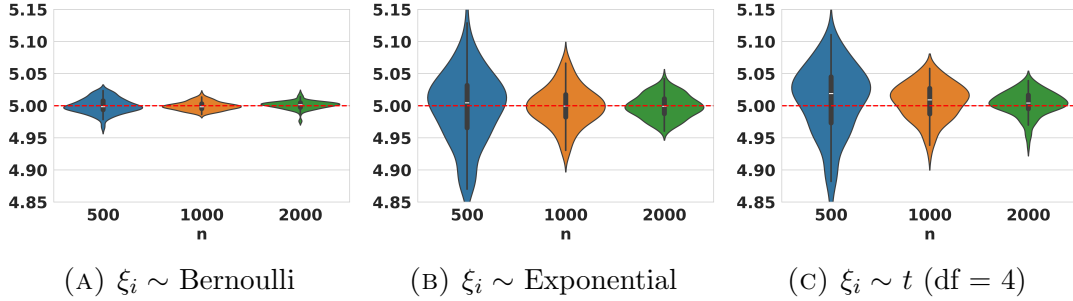

(A) $\xi_i \sim$ Bernoulli   (B) $\xi_i \sim$ Exponential   (C) $\xi_i \sim t$ (df $= 4$)

FIGURE 1. Robustness against distributional misspecification. Red dashed line corresponds to $s_0 = 5$.

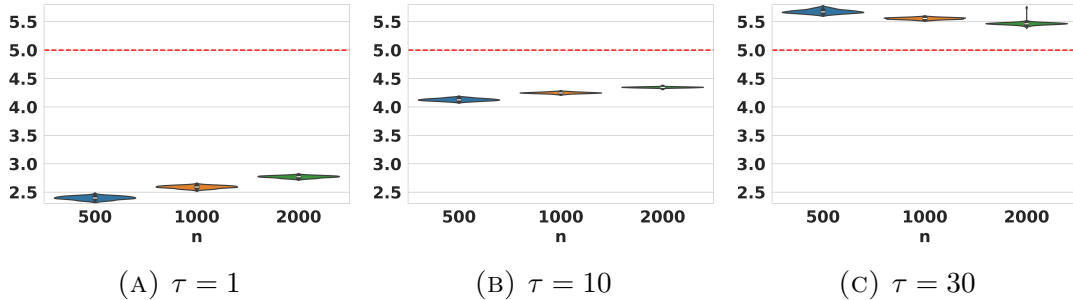(A) $\tau = 1$ (B) $\tau = 10$ (C) $\tau = 30$

FIGURE 2. Robustness against misspecification of the range scale parameter. True parameter is $\tau_0 = 20$. Red dashed line corresponds to $s_0 = 5$.
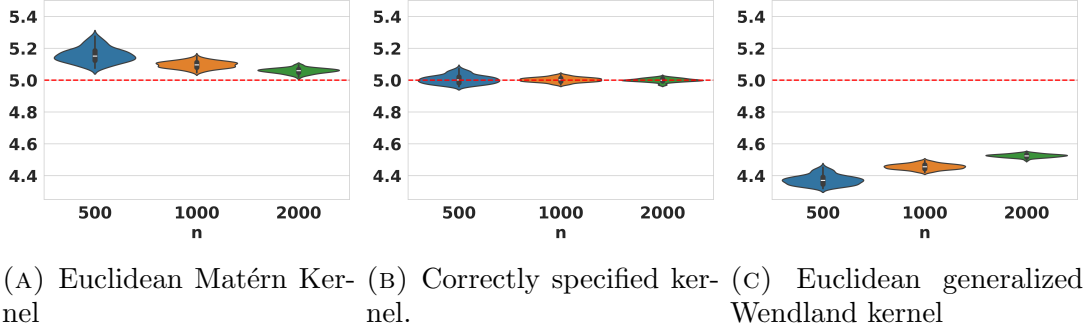
(A) Euclidean Matérn Kernel

(B) Correctly specified kernel.

(C) Euclidean generalized Wendland kernel

FIGURE 3. Robustness against misspecification of the kernel. Red dashed line corresponds to $s_0 = 5$.

APPENDIX A. PROOFS FOR SECTION 4

In this section we provide proofs for some of the results in Section 4.

A.1. **Proofs of Propositions 4.2 and 4.3.** Before we can provide the proof, we need tail bounds for sums of the form

$$\sum_{i=R}^{\infty} a_i(\xi_i - \mathbb{E}(\xi_i)),$$

where $a_i \geq 0$ and ξ_i is a sub-Gaussian random variable. These are standard results, but we have not been able to find a reference for infinite sums. Since the proof is short, we include it here. Following Definition 2.2 in Wainwright (2019), a random variable X is sub-Gaussian with sub-Gaussian parameter γ if

$$\mathbb{E}\big(e^{t(X-\mathbb{E}(X))}\big) \leq e^{\gamma^2 t^2/2}$$

for all $t \in \mathbb{R}$.

Lemma A.1. *Let $\{a_i\}$ be a square-summable sequence with $a_i \geq 0$ and $\{\xi_i\}$ an i.i.d. sequence of sub-Gaussian random variables with mean μ and sub-Gaussian parameter γ. Then for $t \geq 0$, we have*

$$\mathbb{P}\bigg(\bigg|\sum_{i=R}^{\infty} a_i(\xi_i - \mu)\bigg| \geq t\bigg) \leq 2\exp\bigg(-\frac{t^2}{2\gamma^2 \sum_{i=R}^{\infty} a_i^2}\bigg).$$

Proof. Note first that by definition, we have

$$\lim_{M\to\infty} \log\left[\mathbb{E}\left(\exp\left(\lambda\sum_{i=R}^{M} a_i(\xi_i-\mu)\right)\right)\right] \leq \frac{(\gamma\lambda)^2}{2}\sum_{i=R}^{\infty} a_i^2 < \infty.$$

Because log is continuous and monotonically increasing, Fatou's lemma implies that

$$\log\left[\mathbb{E}\left(\exp\left(\lambda\sum_{i=R}^{\infty} a_i(\xi_i-\mu)\right)\right)\right] \leq \liminf_{M\to\infty} \log\left[\mathbb{E}\left(\exp\left(\lambda\sum_{i=R}^{M} a_i(\xi_i-\mu)\right)\right)\right]$$

$$\leq \frac{(\gamma\lambda)^2}{2}\sum_{i=R}^{\infty} a_i^2.$$

Thus $\sum_{i=R}^{\infty} a_i\xi_i$ is sub-Gaussian with sub-Gaussian parameter $\gamma^2\sum_{i=R}^{\infty} a_i^2$ so that Proposition 2.5 in Wainwright (2019) yields, for $t > 0$,

$$\mathbb{P}\left(\sum_{i=R}^{\infty} a_i(\xi_i^2-1) \geq t\right) \leq \exp\left(-\frac{t^2}{2\gamma^2\sum_{i=R}^{\infty} a_i^2}\right). \qquad \square$$

PROPOSITION 4.2. *Let $s_0 > d/2$ and let $\{\xi_i\}$ be the sequence of i.i.d. random variables used in the definition of u in* (8). *Then almost surely, there is a positive constant c such that*

$$\sum_{i=R}^{\infty} (\tau_0+\lambda_i)^{-s_0}\xi_i^2 \geq cR^{-2s_0/d+1} \quad \text{for all} \quad R \in \mathbb{N}.$$

Proof. If ξ_i satisfies $0 < \mathbb{E}(\xi_i^2) < \infty$, we may select b be such that

$$\tilde{\xi}_i^2 := \min\{\xi_i^2, b\}$$

has positive mean. Then $\tilde{\xi}_i^2$ is bounded and thus sub-Gaussian. Then, because

$$\sum_{i=R}^{\infty} (\tau_0+\lambda_i)^{-s_0}\xi_i^2 \geq \sum_{i=R}^{\infty} (\tau_0+\lambda_i)^{-s_0}\tilde{\xi}_i^2,$$

it suffices to prove the claim for ξ_i such that ξ_i^2 is sub-Gaussian. First, note that Weyl's law (3) implies that

$$(37) \qquad \sum_{i=R}^{\infty} (\tau_0+\lambda_i)^{-s_0} \geq c\sum_{i=R}^{\infty} i^{-2s_0/d} \geq \int_R^{\infty} x^{-2s_0/d}\,\mathrm{d}x = cR^{-2s_0/d+1}$$

and

$$(38) \qquad \sum_{i=R}^{\infty} (\tau_0+\lambda_i)^{-2s_0} \leq C\sum_{i=R}^{\infty} i^{-4s_0/d} \leq \int_{R+1}^{\infty} x^{-4s_0/d}\,\mathrm{d}x = C(R+1)^{-4s_0/d+1}.$$

Next, we can write

$$\sum_{i=R}^{\infty}(\tau_0 + \lambda_i)^{-s_0}\xi_i^2 = \sum_{i=R}^{\infty}(\tau_0 + \lambda_i)^{-s_0}(\xi_i^2 - 1) + \sum_{i=R}^{\infty}(\tau_0 + \lambda_i)^{-s_0}.$$

Dividing through by $R^{-\alpha s_0 + 1}$ combined with (37) yields

$$R^{2s_0/d-1}\sum_{i=R}^{\infty}(\tau_0 + \lambda_i)^{-s_0}\xi_i^2 = R^{2s_0/d-1}\bigg(\sum_{i=R}^{\infty}(\tau_0 + \lambda_i)^{-s_0}(\xi_i^2 - 1) + \sum_{i=R}^{\infty}(\tau_0 + \lambda_i)^{-s_0}\bigg)$$

$$(39) \qquad\qquad \geq c + R^{2s_0/d-1}\sum_{i=R}^{\infty}(\tau_0 + \lambda_i)^{-s_0}(\xi_i^2 - 1).$$

Lemma A.1 and (38) give

$$\mathbb{P}\bigg(\bigg|R^{2s_0/d-1}\sum_{i=R}^{\infty}(\tau_0 + \lambda_i)^{-s_0}(\xi_i^2 - 1)\bigg| \geq \frac{c}{2}\bigg) \leq 2\exp\bigg(-\frac{c^2 R^{-4s_0/d+2}}{8\gamma^2\sum_{i=R}^{\infty}(\tau_0 + \lambda_i)^{-2s_0}}\bigg)$$

$$\leq 2\exp\bigg(-\frac{c^2 R^{-4s_0/d+2}}{8\gamma^2 C(R+1)^{-4s_0/d+1}}\bigg)$$

$$\leq 2\exp\left(-CR\right).$$

For the sets

$$A_R = \bigg\{\bigg|R^{2s_0/d-1}\sum_{i=R}^{\infty}(\tau_0 + \lambda_i)^{-s_0}(\xi_i^2 - 1)\bigg| \geq \frac{c}{2}\bigg\},$$

we thus have

$$\sum_{R=1}^{\infty}\mathbb{P}(A_R) \leq \sum_{R=1}^{\infty}2\exp(-CR) < \infty.$$

An application of the Borel–Cantelli lemma yields that almost surely there exists an $R_0 < \infty$ such that

$$\bigg|R^{2s_0/d-1}\sum_{i=R}^{\infty}(\tau_0 + \lambda_i)^{-s_0}(\xi_i^2 - 1)\bigg| \leq \frac{c}{2}$$

for all $R \geq R_0$. Plugging this back into (39), we obtain

$$R^{2s_0/d-1}\sum_{i=R}^{\infty}(\tau_0 + \lambda_i)^{-s_0}\xi_i^2 \geq c + R^{2s_0/d-1}\sum_{i=R}^{\infty}(\tau_0 + \lambda_i)^{-s_0}(\xi_i^2 - 1) \geq \frac{c}{2}$$

for $R \geq R_0$. This is equivalent to

$$\sum_{i=R}^{\infty}(\tau_0 + \lambda_i)^{-s_0}\xi_i^2 \geq \frac{c}{2}R^{-2s_0/d+1}.$$

Finally, note that for any R_0, the sum of the first $R_0 - 1$ is strictly positive almost surely and so adjusting the constant c yields the claim for any $R \in \mathbb{N}$. □

PROPOSITION 4.3. *Let $\theta = (s, \tau)$ with $s > d/2$. Suppose that $f \in \mathcal{H}^s$ and, almost surely,*

(40) $$\|u - f\|_0 \leq C_1 n^{-s_0/d+1/2+\epsilon/d} + C_2 n^{-s/d}\|f\|_s$$

for every $\epsilon > 0$, where $C_1, C_2 > 0$ can depend on $s, s_0, \epsilon, \mathcal{M}$ and the sample path. Then almost surely

(41) $$\|f\|_s^2 \geq C n^{1+2(s-s_0)/d-\epsilon'}$$

for every $\epsilon' > 0$, where $C > 0$ depends on $s, s_0, \epsilon', \mathcal{M}$ and the sample path.

Proof. The proof is similar to that of Theorem 8 in van der Vaart and van Zanten (2011). Recall the spaces $\hat{H}_L^s$ and the isomorphism $\iota : H_L^s \to \hat{H}_L^s$ given by $\iota(f)_i = \langle f, e_i \rangle_0$ from Section 2.2.1. To simplify notation, we let $\hat{f} = \iota(f)$. For two sequences $a, b \in \hat{H}_L^s$ we define $(ab)_i = a_i b_i$. Because $f \in H_L^s$, we have $\hat{f} \in \hat{H}_L^s$. Define $\mathbf{1}_R$ as the sequence such that $(\mathbf{1}_R)_i = 0$ if $i < R$ and $(\mathbf{1}_R)_i = 1$ if $i \geq R$. By norm equivalence,

(42)
$$\|\hat{f}\mathbf{1}_R\|_0^2 = \sum_{i=R}^{\infty}(1+\lambda_i)^{s-s}\hat{f}_i^2 \leq (1+\lambda_R)^{-s}\sum_{i=R}^{\infty}(1+\lambda_i)^s\hat{f}_i^2$$
$$\leq C(1+\lambda_R)^{-s}\|f\|_s^2.$$

By Proposition 4.2, almost surely there exists $0 < c < \infty$ such that, for all $R \in \mathbb{N}$,

(43) $$\|\hat{u}\mathbf{1}_R\|_0^2 = \sum_{i=R}^{\infty}\hat{u}_i^2 = v(\theta_0)\sum_{i=R}^{\infty}(\tau_0+\lambda_i)^{-s_0}\xi_i^2 \geq cv(\theta_0)R^{-2s_0/d+1}.$$

Via the reverse triangle inequality, we obtain

$$\|\hat{f}\mathbf{1}_R\|_{\hat{0}} \geq \|\hat{u}\mathbf{1}_R\|_{\hat{0}} - \|(\hat{f}-\hat{u})\mathbf{1}_R\|_{\hat{0}} \geq \|\hat{u}\mathbf{1}_R\|_{\hat{0}} - \|f-u\|_0.$$

Now set

$$R = \left\lceil \left(\frac{2C_1 n^{-s_0/d+1/2+\epsilon/d}}{\sqrt{cv(\theta_0)}}\right)^{1/(-s_0/d+1/2)} \right\rceil.$$

Using (40) and (43) and the above choice of R we get

$$\|\hat{f}\mathbf{1}_R\|_{\hat{0}} \geq \|\hat{u}\mathbf{1}_R\|_{\hat{0}} - \|f-u\|_0$$
$$\geq \sqrt{cv(\theta_0)}R^{-s_0/d+1/2} - C_1 n^{-s_0/d+1/2+\epsilon/d} - C_2 n^{-s/d}\|f\|_s$$
$$\geq C_1 n^{-s_0/d+1/2+\epsilon/d} - C_2 n^{-s/d}\|f\|_s.$$

Then (42) and Weyl's law in (3) give

$$
\begin{aligned}
\|f\|_s &\geq c(1+\lambda_R)^{s/2}\|\hat{f}\mathbf{1}_R\|_{\hat{0}} \\
&\geq cR^{s/d}\big(C_1 n^{-s_0/d+1/2+\epsilon/d} - C_2 n^{-s/d}\|f\|_s\big) \\
&\geq cn^{s/d-2\epsilon s/(d(2s_0-d))}\big(C_1 n^{-s_0/d+1/2+\epsilon/d} - C_2 n^{-s/d}\|f\|_s\big) \\
&= c\big(C_1 n^{1/2+(s-s_0)/d+\epsilon/d-2\epsilon s/(d(2s_0-d))} - C_2 n^{-2\epsilon s/(d(2s_0-d))}\|f\|_s\big) \\
&\geq c\big(C_1 n^{1/2+(s-s_0)/d+\epsilon/d-2\epsilon s/(d(2s_0-d))} - C_2\|f\|_s\big),
\end{aligned}
$$

where on the last line we used the fact that $2s_0 > d$. Rearranging then yields

$$
\|f\|_s \geq cn^{1/2-(s-s_0)/d+\epsilon/d-2\epsilon s/(d(2s_0-d))}.
$$

Since $\epsilon/d - 2\epsilon s/(d(2s_0 - d)) \to 0$ as $\epsilon \to 0$, relabelling completes the proof. $\square$

A.2. **Function Approximation in RKHS on Manifolds.** To prepare the remaining proofs, the following lemma will be convenient.

Lemma A.2 (Lemma 1 in Krieg and Sonnleitner, 2022). *The exponential map is bi-Lipschitz. That is,*

$$
c_j\|x - y\| \leq d_{\mathcal{M}}(\phi_j^{-1}(x), \phi_j^{-1}(y)) \leq C_j\|x - y\|
$$

for some $c_j, C_j > 0$ and all $x, y \in \phi_j(\Omega_j)$. Consequently, choosing $c = \min_{1 \leq j \leq m}$ and $C = \max_{1 \leq j \leq m} C_j$ we have

$$
c\|x - y\| \leq d_{\mathcal{M}}(\phi_j^{-1}(x), \phi_j^{-1}(y)) \leq C\|x - y\| \quad \text{for all } x, y \in \Omega_j \text{ and } 1 \leq j \leq m.
$$

PROPOSITION 4.5. *Let $f \in H_L^{s_0}$ with $s_0 > d$ and let $m_{s,n}^f$ be its minimum norm interpolant from the RKHS $\mathcal{H}^s$. Then there exists a positive constant C that depends on s_0 and $\mathcal{M}$ such that we have almost surely*

$$
(44) \qquad \qquad \|f - m_{s,n}^f\|_0 \leq Ch_{\boldsymbol{x}}^{s_0}\rho_{\boldsymbol{x}}^{s-s_0}\|f\|_{s_0}
$$

when $h_{\boldsymbol{x}}$ is small enough. Thus, if $h_{\boldsymbol{x}} \to 0$ and C is allowed to depend on f, then (16) holds for all n.

Proof. We make extensive use of Theorem 2.1. To that end, recall that $T_{\mathcal{M}}f = f|_{\mathcal{M}}$ is a bounded operator $T_{\mathcal{M}}: H^{s+(k-d)/2}(\mathbb{R}^k) \to H_L^s$. Its inverse $E_{\mathcal{M}}$ on H_L^s is bounded as well. Furthermore, note that from Lemma A.2 it follows that the point set $\{\phi_j(x_i) \mid x_i \in \Omega_j\}$ has separation radius $q_{\boldsymbol{x}}^j$ such that

$$
cq_{\boldsymbol{x}} \leq q_{\boldsymbol{x}}^j \leq Cq_{\boldsymbol{x}}.
$$

Throughout the remainder of the proof C denotes universal constants only depending on $s_0, s, \tau, \mathcal{M}$ that may be different from one occurence to the next. Note that as in Lemma 10 in Fuselier and Wright (2012), we have for $h_{\boldsymbol{x}}$ small enough

(45) $$\|f - m_{s,n}^f\|_0 \leq C h_{\boldsymbol{x}}^{s_0} \|f - m_{s,n}^f\|_{s_0}.$$

Now let $E_{\mathcal{M}}f$ be the extension of f. It follows from Theorem 3.4 and Corollary 3.5 in Narcowich et al. (2006) that there exists a sequence $\eta = \eta(n)$ such that $\eta \asymp q_{\boldsymbol{x}}^{-1}$ and corresponding functions $f_\eta : \mathbb{R}^k \to \mathbb{R}$ satisfying the following:

(1) For the Fourier transform $\hat{f}_\eta$, we have $\operatorname{supp} \hat{f}_\eta \subset B_\eta(\mathbf{0})$. That is, f_η is *bandlimited*.

(2) The function f_η interpolates u, that is

$$f_\eta(x_i) = f(x_i), \ 1 \leq i \leq n.$$

(3) We have $\|E_{\mathcal{M}}f - f_\eta\|_{H^{s_0+(k-d)/2}(\mathbb{R}^k)} \leq 5\|E_{\mathcal{M}}f\|_{H^{s_0+(k-d)/2}(\mathbb{R}^k)}$, which, together with the boundedness of the restriction operator, implies

$$\|f - T_{\mathcal{M}}f_\eta\|_{s_0} \leq C\|f\|_{s_0}.$$

(4) We have $\|f_\eta\|_{H^{s_0+(k-d)/2}(\mathbb{R}^k)} \leq C\|E_{\mathcal{M}}f\|_{H^{s_0+(k-d)/2}(\mathbb{R}^k)}$, which, together with norm equivalence, implies

$$\|f_\eta\|_{H^{s_0+(k-d)/2}(\mathbb{R}^k)} \leq C\|f\|_{s_0}.$$

Moreover, the functions f_η satisfy the Bernstein inequality

$$\|f_\eta\|_{H^s(\mathbb{R}^k)} \leq C\eta^{s-s'}\|f_\eta\|_{H^{s'}(\mathbb{R}^k)}.$$

Because u_η is bandlimited, it follows that $f_\eta \in H^s(\mathbb{R}^k)$ for all s. Then, norm equivalence and standard Hilbert space theory imply

(46) $$\|T_{\mathcal{M}}f_\eta - m_{s,n}^f\|_{H^s(\mathcal{M})} \leq C\|T_{\mathcal{M}}f_\eta - m_{s,n}^f\|_s \leq C\|T_{\mathcal{M}}f_\eta\|_s.$$

Furthermore, norm equivalence, the Bernstein inequality of bandlimited functions, property 4 above and boundedness of the extension and restriction operators imply

(47) $$C\|T_{\mathcal{M}}f_\eta\|_s \leq C\eta^{s-s_0}\|f_\eta\|_{H^{s_0+(k-d)/2}(\mathbb{R}^k)} \leq C\eta^{s-s_0}\|f\|_{s_0}.$$

Combining (46) and (47) with property 3 and using once more Lemma 10 in Fuselier and Wright (2012) gives

$$
\begin{aligned}
(48) \qquad \|f - m_{s,n}^f\|_{s_0} &\leq \|f - T_{\mathcal{M}} f_\eta\|_{s_0} + \|T_{\mathcal{M}} f_\eta - m_{s,n}^f\|_{s_0} \\
&\leq C\|f\|_{s_0} + C h_{\boldsymbol{x}}^{s-s_0} \|T_{\mathcal{M}} f_\eta - m_{s,n}^f\|_s \\
&\leq C\|f\|_{s_0} + C h_{\boldsymbol{x}}^{s-s_0} \|T_{\mathcal{M}} f_\eta\|_s \\
&\leq C\|f\|_{s_0} + C \left(\frac{h_{\boldsymbol{x}}}{q_{\boldsymbol{x}}}\right)^{s-s_0} C\|f\|_{s_0} \\
&\leq (C + C \rho_{\boldsymbol{x}}^{s-s_0})\|f\|_{s_0}.
\end{aligned}
$$

Because $\rho_{\boldsymbol{x}} \geq 1$, it follows that $(C + C\rho_{\boldsymbol{x}}^{s-s_0}) \leq C\rho_{\boldsymbol{x}}^{s-s_0}$. Thus combining (45) and (48) we obtain for $h_{\boldsymbol{x}}$ small enough

$$
\|f - m_{s,n}^f\|_0 \leq C h_{\boldsymbol{x}}^{s_0} \rho_{\boldsymbol{x}}^{s-s_0} \|f\|_{s_0}. \qquad \square
$$

PROPOSITION 4.6. *Let u be as in (8). Suppose that $s_0 > d$ and $s > d/2$. Let $m_{s,n}$ be the minimum norm interpolant for u. Then almost surely*

$$
(49) \qquad \|m_{s,n}\|_s^2 \leq C n^{2(s-s_0)+1+\epsilon}
$$

for all $\epsilon > 0$, where C depends on s, θ_0, ϵ and the sample path. If ξ_i are Gaussian, then it is only necessary to assume $s_0 > d/2$ to obtain

$$
(50) \qquad cn \leq \|m_{s_0,n}\|_{s_0}^2 \leq Cn
$$

almost surely for positive constants c and C that depend only on θ_0 and the sample path.

Proof. Let $T_{\mathcal{M}}$ and $E_{\mathcal{M}}$ be the bounded restriction and extension operators from Theorem 2.1. Let u_η be a sequence of bandlimited function as in the proof of Proposition 4.5 such that $u(x_i) = E_{\mathcal{M}} u(x_i) = u_\eta(x_i)$. Then

$$
\|u_\eta\|_{H^{s_0+(k-d)/2}(\mathbb{R}^k)} \leq C\|E_{\mathcal{M}} u\|_{H^{s_0+(k-d)/2}(\mathbb{R}^k)}.
$$

Theorem 2.1, norm equivalence, the minimum norm property and the Bernstein inequality for bandlimited functions yield

$$
\begin{aligned}
\|m_{s,n}\|_s \leq \|T_{\mathcal{M}} u_\eta\|_s &\leq C\|u_\eta\|_{H^{s+(k-d)/2}(\mathbb{R}^k)} \leq C\eta^{s_0-s-d/2-\epsilon}\|u_\eta\|_{H^{s_0+k/2-d-\epsilon}(\mathbb{R}^k)} \\
&\leq C q_{\boldsymbol{x}}^{s_0-s-d/2-\epsilon}\|u\|_{s_0-d/2-\epsilon}.
\end{aligned}
$$

This proves (49).

To prove (50), let $m_{\theta_0,n}$ be the minimum norm interpolant from $H_L^{s_0}$ equipped with the norm $\|\cdot\|_{\theta_0}$. Since $\mathbb{V}(u(\boldsymbol{x})) = \sigma_0^2 K_{\theta_0}(\boldsymbol{x})$, we can write

$$\|m_{\theta_0,n}\|_{\theta_0}^2 = u(\boldsymbol{x})^\top K_{\theta_0}(\boldsymbol{x})^{-1} u(\boldsymbol{x}) = \sigma_0^2 \tilde{\boldsymbol{\xi}}_n^\top K_{\theta_0}(\boldsymbol{x})^{1/2} K_{\theta_0}(\boldsymbol{x})^{-1} K_{\theta_0}(\boldsymbol{x})^{1/2} \tilde{\boldsymbol{\xi}}_n = \sigma_0^2 \sum_{i=1}^n \tilde{\xi}_{n,i}^2,$$

where $\{\tilde{\boldsymbol{\xi}}_n\} = \{(\tilde{\xi}_{n,1}, \ldots, \tilde{\xi}_{n,n})\}$ forms a triangular array of independent chi-squared random variables. By the strong law of large numbers,

$$\frac{1}{n} \sum_{i=1}^n \tilde{\xi}_{n,i}^2 \xrightarrow{\text{a.s.}} 1,$$

which yields the claim for $\|m_{\theta_0,n}\|_{\theta_0}$. For any other kernel the claim then follows from the minimum norm property and norm equivalence . $\square$

In the proof of Proposition 4.7 it will be convenient to use the following equivalent definition of Sobolev spaces on $\mathcal{M}$. Let $\{(\Omega_i, \phi_i)\}_{i=1}^m$ denote the atlas of normal coordinates with subordinate partition of unity $\{\psi_i\}$. We define the Sobolev spaces

$$(51) \qquad H^s(\mathcal{M}) = \left\{ f \in L_2(\mathcal{M}) \;\middle|\; \|f\|_{H^s(\mathcal{M})}^2 := \sum_{i=1}^m \|\pi_i(f)\|_{H^s(\mathbb{R}^d)}^2 < \infty \right\},$$

where

$$(52) \qquad\qquad \pi_i(f)(x) = \begin{cases} \psi_i f \circ \phi_i^{-1}(x) & \text{if } x \in \phi_i(\Omega_i) \\ 0 & \text{else.} \end{cases}$$

As in Section 2.1, we have identified $T_p\mathcal{M}$ with $\mathbb{R}^d$. It follows from Triebel (1992, Theorem 7.2.3) that this definition is independent of the choice of normal coordinates and the partition of unity up to equivalent norms and it follows from Triebel (1992, Theorem 7.4.5) that $H^s(\mathcal{M})$ is norm equivalent to H_L^s. Define balls in a bounded Euclidean domain M and on a manifold $\mathcal{M}$ as

$$B_r(x) = \{y \in M \mid \|x - y\| \le r\} \quad \text{and} \quad B_r^{\mathcal{M}}(x) = \{y \in \mathcal{M} \mid d_{\mathcal{M}}(x, y) \le r\}.$$

PROPOSITION 4.7. *Let $s > d/2$.*

(1) *Suppose the family of norms $\{\|\cdot\|_s\}_{s>d/2}$ is such that $\|f\|_s \le C\|f\|_{s'}$ for every f whenever $s \le s'$. Let $\boldsymbol{x}_i^n$ be all points in $\boldsymbol{x}$ except x_i. Then, for $s \le S$, there is a constant $c > 0$, that only depends on S, such that*

$$\min_{1 \le i \le n} \mathbb{V}_s(x_i | \boldsymbol{x}_i^n) \ge c q_{\boldsymbol{x}}^{2S-d}.$$

(2) *There exists a positive constant C, that only depends on s and $\mathcal{M}$, such that*

$$\sup_{y \in \mathcal{M}} \mathbb{V}_s(y|\boldsymbol{x}) \leq Ch_{\boldsymbol{x}}^{2s-d}.$$

Proof. We start by proving (1) and follow the same strategy as in the proof of Proposition 3.5 in Karvonen (2023). We have

$$\mathbb{V}_s(x_i|\boldsymbol{x}^{i-1}) = \sup_{\|f\|_{\mathcal{H}^s} \leq 1} |f(x_i) - m_{s,n}^f(x_i|\boldsymbol{x}^{i-1})|^2,$$

where $m_{s,n}^f(\cdot|\boldsymbol{x}_i^n)$ is the minimum norm interpolant based on the observations $\{f(x_j)|j \neq i\}$. The goal is to construct a function f such that $\|f\|_{\mathcal{H}^s} \leq 1$ and $f(x_j) = 0$ for all $j \neq i$. In this case, it is clear that $m_{s,n}^f(\cdot|\boldsymbol{x}_i^n) \equiv 0$ and so $f(x_i) - m_{s,n}^f(x_i|\boldsymbol{x}_i^n) = f(x_i)$. This is a well-known strategy and usually so-called bump functions are used. Our construction here is due to Krieg and Sonnleitner (2022, Section 2.2). Recall the atlas $\{(\Omega_j, \phi_j)\}_{j=1}^m$ and let $\{\psi_j\}_{j=1}^m$ be a subordinate partition of unity. We now generate a finite set of new atlasses based on geodesic coordinates that are particularly well suited for our purposes. First, we construct a new covering of $\mathcal{M}$ via $\{(4\Omega_j, \phi_j)\}$, where

$$4\Omega_j = \{x \in \mathcal{M} \mid d_{\mathcal{M}}(x, p_j) < 4r\}$$

and p_j is the center of Ω_j, that is, $\boldsymbol{0}$ in local coordinates of (Ω_j, ϕ_j) and the ϕ_i are now to be understood as inverses of the exponential maps on $B_{4r}(\boldsymbol{0})$. This is possible because $r < r_0/32$ and so the extensions are still diffeomorphisms and the radii of the $4\Omega_j$ are still less than $r_0/8$ as required by Chapter 7 in Triebel (1992). We construct new atlasses $\mathcal{A}_k$ and partitions of unity Ψ_k in the following way. First, let $I_k = \{j \mid \Omega_j \cap 1.5\Omega_k = \emptyset\}$. Then define the atlas $\mathcal{A}_k = \{(4\Omega_j, \phi_j)\}_{j \in I_k} \cup \{(4\Omega_k, \phi_k)\}$. Note that $4\Omega_k \cup \bigcup_{j \in I_k} 4\Omega_j$ still covers $\mathcal{M}$. Finally, define $\tilde{\psi}_k$ as

$$\tilde{\psi}_k = \sum_{j \leq m, \, j \notin J_k} \psi_j$$

and $\Psi_k = \{\psi_j\}_{j \in J_k} \cup \tilde{\psi}_k$. This is still a partition of unity such that $\operatorname{supp} \psi_j \subset 4\Omega_j$ and $\operatorname{supp} \tilde{\psi}_k \subset 4\Omega_k$ with the convenient property that by construction, $\tilde{\psi}_k \equiv 1$ on $1.5\Omega_k$. Finally, it follows from Theorem 7.2.3 in Triebel (1992) that each atlas $\mathcal{A}_k$ together with partition of unity Ψ_k generates a space that is norm equivalent to $H^s(\mathcal{M})$.

Now suppose $q_{\boldsymbol{x}} < r/2$. Then $B_{q_{\boldsymbol{x}}}^{\mathcal{M}}(x_i) \subset 1.5\Omega_k$ whenever $x_i \in \Omega_k$ and $x_j \notin B_{q_{\boldsymbol{x}}}^{\mathcal{M}}(x_i)$ for $j \neq i$. Lemma A.2 implies that

$$\phi_k^{-1}(B_{cq_{\boldsymbol{x}}}(\phi_k(x_i))) := \phi_k^{-1}\left(\{y \in \phi_k(\Omega_k) \mid \|y - \phi_k(x_i)\| \leq cq_{\boldsymbol{x}}\}\right) \subset B_{q_{\boldsymbol{x}}}^{\mathcal{M}}(x_i) \subset 1.5\Omega_k.$$

Now, define $f : \mathcal{M} \to \mathbb{R}$ such that $f(x) = 0$ if $x \notin \Omega_k$ and, for $x \in \phi_k^{-1}(\Omega_k)$,

$$f \circ \phi_k^{-1}(x) = \begin{cases} \exp\left(-\dfrac{1}{1 - \left\| \frac{x - \phi_k(x_i)}{cq_{\boldsymbol{x}}} \right\|^2} \right) & \text{if} \quad \|x - \phi_k(x_i)\| \le cq_{\boldsymbol{x}}, \\ 0 & \text{else.} \end{cases}$$

That is, $f|_{\Omega_k} = g \circ \phi_k$ for $g : \mathbb{R}^d \to \mathbb{R}$ such that

$$g(x) = \begin{cases} \exp\left(-\dfrac{1}{1 - \left\| \frac{x - \phi_i(x_i)}{cq_{\boldsymbol{x}}} \right\|^2} \right) & \text{if} \quad \|x - \phi_k(x_i)\| \le cq_{\boldsymbol{x}}, \\ 0 & \text{else.} \end{cases}$$

Note that $\tilde{\psi}_k f = f$ and $\psi_j f \equiv 0$ for $j \ne k$. Following the exact same steps as in the proof of Proposition 3.5 in Karvonen (2023), it follows that there exists a constant C such that

$$\|\psi_k f \circ \phi_k^{-1}\|_{H^S(\mathbb{R}^d)} = \|g\|_{H^S(\mathbb{R}^d)} \le C q_{\boldsymbol{x}}^{d/2 - S}.$$

Thus, by norm equivalence, norm monotonicity and the fact that each atlas $\mathcal{A}_k$ yields a norm equivalent space and there are only finitely many such atlasses, we have that there exists an f such that $f(x_i) = 1$ and

$$\|f\|_s \le C\|f\|_{H^S(\mathcal{M})} = C\left(\sum_{j=1}^m \|\pi_j^k(f)\|_{H^S(\mathbb{R}^d)}^2 \right)^{1/2} = C\|\pi_k^k(f)\|_{H_2^S(\mathbb{R}^d)} = C\|g\|_{H^S(\mathbb{R}^d)}$$

$$\le C q_{\boldsymbol{x}}^{d/2 - S},$$

where π_j^k is the analog of the function defined in (52) based on $\mathcal{A}_k$ and Ψ_k. The constant C depends on S and $\mathcal{H}^s$. Appropriately rescaling f to f_{sc} such that $\|f_{\mathrm{sc}}\|_{\mathcal{H}^s} \le 1$ for all $s \le S$ yields $f_{\mathrm{sc}}(x_i) \ge C q_{\boldsymbol{x}}^{S - d/2}$. In the case $q_{\boldsymbol{x}} \ge r/2$ we repeat the procedure with $q_{\boldsymbol{x}}$ replaced by $r/2$. From this it follows that

$$\mathbb{V}_s(u(x_i)|\boldsymbol{x}_i^n) = \sup_{\|f\|_{\mathcal{H}^s} \le 1} |f(x_i) - m_{s,n}^f(x_i|\boldsymbol{x}_i^n)|^2$$

$$\ge f_{\mathrm{sc}}(x_i)^2$$

$$\ge \min\{C(r/2)^{2S - d}, C q_{\boldsymbol{x}}^{2S - d}\}$$

as claimed.

Next we prove (2). It follows from Lemma 10 in Fuselier and Wright (2012) that, for every $f \in H_L^s$,

$$\|f - m_{s,n}^f\|_\infty \le Ch_{\boldsymbol{x}}^{s-d/2}\|f - m_{s,n}^f\|_{\mathcal{H}^s} \le Ch_{\boldsymbol{x}}^{s-d/2}\|f\|_{\mathcal{H}^s}$$

for $h_{\boldsymbol{x}}$ small enough, see the comment below Proposition 4.5. Then

$$\sup_{x\in\mathcal{M}} \mathbb{V}_s(x|\boldsymbol{x}) = \sup_{x\in\mathcal{M}} \sup_{\|f\|_{\mathcal{H}^s}\le 1} |f(x) - m_{s,n}^f(x)|^2 = \sup_{\|f\|_{\mathcal{H}^s}\le 1} \|f - m_{s,n}^f\|_\infty^2 \le Ch_{\boldsymbol{x}}^{2s-d}\|f\|_{\mathcal{H}^s}^2$$
$$\le Ch_{\boldsymbol{x}}^{2s-d}$$

for $h_{\boldsymbol{x}}$ small enough, where C does not depend on f. Adjusting the constant C, the bound holds for all n. $\qquad\square$

A.3. **Proofs for Restricted Kernels.** Here we give proofs for some results in Section 4.4.

PROPOSITION 4.12. *Let $\tilde{K}_s(\cdot,\cdot)$ be as in Theorem 4.11. Let $\mathcal{H}^s(\mathbb{R}^k)$ be the RKHS of $\tilde{K}_s$ and $\mathcal{H}^s$ the RKHS of $K_s(\cdot,\cdot) = \tilde{K}_s(\cdot,\cdot)|_{\mathcal{M}\times\mathcal{M}}$. Let $s \le s'$. If $\|f\|_{\mathcal{H}^s(\mathbb{R}^k)} \le C\|f\|_{\mathcal{H}^{s'}(\mathbb{R}^k)}$ for all $f \in \mathcal{H}^{s'}(\mathbb{R}^k)$ and some C independent of s, then $\|f\|_s \le C\|f\|_{s'}$ for all $f \in \mathcal{H}^s$.*

Proof. Lemma 4 in Fuselier and Wright (2012) shows that for $f \in \mathcal{H}^{s'}$ there exists an extension operator E such that $\|f\|_{s'} = \|Ef\|_{\mathcal{H}^{s'}(\mathbb{R}^k)}$. Moreover, for every s the restriction operator $T_{\mathcal{M}}$ has norm bounded by 1 as operator $T_{\mathcal{M}} : \mathcal{H}^s(\mathbb{R}^k) \to \mathcal{H}_s$. It follows that $f = T_{\mathcal{M}}Ef$ and so

$$\|f\|_s = \|T_{\mathcal{M}}Ef\|_s \le \|Ef\|_{\mathcal{H}^s(\mathbb{R}^k)} \le C\|Ef\|_{\mathcal{H}^{s'}(\mathbb{R}^k)} = C\|f\|_{s'}. \qquad\square$$

Proposition A.3. *Set $\mu = \kappa+(d+1)/2$. Let $\kappa \le \kappa'$ and let $\mathcal{H}^\kappa(\mathbb{R}^k)$ and $\mathcal{H}^{\kappa'}(\mathbb{R}^k)$ be the RKHS of Φ_κ and $\Phi_{\kappa'}$ with norms $\|\cdot\|_\kappa$ and $\|\cdot\|_{\kappa'}$, respectively. Then $\mathcal{H}^{\kappa'}(\mathbb{R}^k) \subset \mathcal{H}^\kappa(\mathbb{R}^k)$ and*

$$\|f\|_\kappa \le \|f\|_{\kappa'}.$$

for all $f \in \mathcal{H}_{\kappa'}$.

Proof. By Theorem 10.12 in Wendland (2004), for any integrable and continuous radial basis function Φ on $\mathbb{R}^k$ the corresponding RKHS norm can be expressed via

$$\|f\|_\Phi^2 = \int_{\mathbb{R}^k} |\hat{f}(\xi)|^2 \hat{\Phi}(\|\xi\|)^{-1} \,\mathrm{d}\xi.$$

The generalized Wendland functions Φ_κ are such radial basis functions (Hubbert and Jäger, 2023). This implies $\mathcal{H}_{\kappa'} \subset \mathcal{H}_\kappa$ It thus suffices to show that $\hat{\Phi}_{\kappa'}(z) \le \hat{\Phi}_\kappa(z)$ whenever $\kappa \le \kappa'$. Equation 3.17 in Hubbert and Jäger (2023) provides a closed form

for the Fourier transform of Φ_κ. Set $\lambda = (d+1)/2 + \kappa$ and note that $\lambda > 1$. Then

$$\hat{\Phi}_\kappa(z) = \frac{2^\lambda \Gamma(\lambda)\Gamma(\lambda+1)}{\Gamma(3\lambda)} \; {}_1F_2\left(\lambda; \frac{3}{2}\lambda; \frac{3\lambda+1}{2}; -\left(\frac{z\beta}{2}\right)^2\right),$$

where ${}_1F_2$ is the hypergeometric function (Abramowitz and Stegun, 1964, Chapter 15). Note that the fourth argument is negative. Display 15.3.1 in Abramowitz and Stegun (1964) shows that the hypergeometric function has the integral representation

(53)
$$
\begin{aligned}
&{}_1F_2\left(\lambda; \frac{3}{2}\lambda; \frac{3\lambda+1}{2}; -\left(\frac{z\beta}{2}\right)^2\right) \\
&= \frac{\Gamma(3/2\lambda+1/2)}{\Gamma(3/2\lambda)\Gamma(1/2)} \int_0^1 t^{3/2\lambda-1}(1-t)^{-1/2}\left(1+t\left(\frac{z\beta}{2}\right)^2\right)^{-\lambda} \mathrm{d}t.
\end{aligned}
$$

Now consider

(54)
$$\vartheta(\lambda) = \frac{\Gamma(3/2\lambda+1/2)}{\Gamma(3/2\lambda)\Gamma(1/2)} \cdot \frac{2^\lambda \Gamma(\lambda)\Gamma(\lambda+1)}{\Gamma(3\lambda)}$$

and let $\psi^{(0)}$ be the digamma function, that is $\psi^{(0)}(z)\Gamma(z) = \Gamma'(z)$ (Abramowitz and Stegun, 1964, Chapter 6). It is an easy consequence of display 6.4.1 in Abramowitz and Stegun (1964) that $\psi^{(0)}$ is monotonically increasing. Straightforward calculation yields

$$\frac{\vartheta'(\lambda)}{\vartheta(\lambda)} = \frac{3}{2}\psi^{(0)}\left(\frac{3}{2}\lambda + \frac{1}{2}\right) + \log(2) + \psi^{(0)}(\lambda) + \psi^{(0)}(\lambda+1) - \frac{3}{2}\psi^{(0)}\left(\frac{3}{2}\lambda\right) - 3\psi^{(0)}(3\lambda).$$

Because $\vartheta(\lambda) > 0$, the derivative ϑ' is negative if and only if the right hand side is negative. Due to the monotonicity of of $\psi^{(0)}$, for $\lambda > 1/2$ we have

(55)
$$\frac{3}{2}\psi^{(0)}\left(\frac{3}{2}\lambda + \frac{1}{2}\right) + \psi^{(0)}(\lambda+1) - \frac{5}{2}\psi^{(0)}(3\lambda) < 0.$$

Moreover, for $\lambda = 1$ we have by displays 6.3.2 and 6.3.5 in Abramowitz and Stegun (1964) that

$$-\frac{1}{2}\left(\psi^{(0)}\left(\frac{3}{2}\lambda\right) + \psi^{(0)}(3\lambda)\right) - \psi^{(0)}\left(\frac{3}{2}\lambda\right) + \log(2) + \psi^{(0)}(\lambda) \leq 0.$$

Again by monotonicity, for all $\lambda \geq 1$ we then have

(56)
$$-\frac{1}{2}\left(\psi^{(0)}\left(\frac{3}{2}\lambda\right) + \psi^{(0)}(3\lambda)\right) - \psi^{(0)}\left(\frac{3}{2}\lambda\right) + \log(2) + \psi^{(0)}(\lambda) \leq 0.$$

Combining (55) and (56) yields $\vartheta'(\lambda) < 0$ for all $\lambda > 1$.

To see that the integral in (53) is decreasing in λ, note that, for $t \in (0, 1)$,

$$\frac{\partial}{\partial \lambda} t^{3/2\lambda-1} \left(1 + t\frac{(\beta z)^2}{4}\right)^{-\lambda} = t^{3/2\lambda-1} \left(1 + t\frac{(\beta z)^2}{4}\right)^{-\lambda} \left[\frac{3}{2}\log(t) - \log\left(1 + t\frac{(\beta z)^2}{4}\right)\right].$$

Because $0 < t < 1$, this implies that the non-negative function inside the integral in (53) is pointwise (in t) decreasing in λ, and thus the integral must be decreasing in λ. Combining (53) and (54), we thus see that

$$\Phi_\kappa(x) = \vartheta(\lambda) \int_0^1 t^{3/2\lambda-1}(1-t)^{-1/2} \left(1 + t\frac{(\beta z)^2}{4}\right)^{-\lambda} dt$$

is the product of two decreasing functions in λ. Because λ is increasing in κ, it follows that Φ_κ is decrasing in κ and thus for all $z > 0$ and for $\kappa \leq \kappa'$ we have

$$\Phi_{\kappa'}(z) \leq \Phi_\kappa(z),$$

which proves the claim. $\qquad\square$

A.4. **An Asymptotic Normality Result.** Here we formally state and proof an asymptotic normality result from Section 4.5.

Proposition A.4. *Let u be as in (8) and assume that $\{\xi_i\}$ are i.i.d. standard Gaussian. Assume that the true smoothness parameter is known and fixed at $s_0 > d/2$. Let $\hat{\sigma}_n^2$ and $\hat{\tau}_n$ be the maximizers of the log-likelihood*

$$\ell(\tau, \sigma^2; u(\boldsymbol{x})) = -\frac{n}{2}\log(2\pi) - \frac{1}{2}\log\left(\det(\sigma^2 K_\theta(\boldsymbol{x})\right) - \frac{1}{2\sigma^2}u(\boldsymbol{x})^\top K_\theta(\boldsymbol{x})^{-1}u(\boldsymbol{x}),$$

where $\theta = (s_0, \tau)$ and $\hat{\tau}_n$ is constrained to be in $[\tau_L, \tau_U]$ for some $1 \leq \tau_L < \tau_U < \infty$. Let $\hat{\theta}_n = (s_0, \hat{\tau}_n)$. If $d \leq 3$, then

$$\sqrt{n}(\hat{\sigma}_n^2 v(\hat{\theta}_n) - \sigma_0^2 v(\theta_0)) \to \mathcal{N}(0, 2\sigma_0^4 v(\theta_0)^2).$$

Note that the restriction $\tau_L \geq 1$ is not really a restriction. We can always rescale the underlying distances to obtain $\tau_L \geq 1$. The assumption $d \leq 3$ is due to the fact that both σ_0^2 and τ_0 can be consistently estimated if $d > 3$ (Anderes, 2010).

Proof. We have

(57) $$\sqrt{n}\left[\frac{u(\boldsymbol{x})^\top K_{\theta_0}(\boldsymbol{x})^{-1}u(\boldsymbol{x})}{n\sigma_0^2} - 1\right] = \frac{1}{\sqrt{n}}\sum_{i=1}^n(\tilde{\xi}_{n,i}^2 - 1) \to \mathcal{N}(0, 2),$$

for example by an application of the Lindeberg–Feller theorem (Kallenberg, 2002, Theorem 6.13). Now fix an arbitrary τ_1 and let $\theta_1 = (s_0, \tau_1)$. Let furthermore σ_1^2 be such that $\sigma_1^2 v(\theta_1) = \sigma_0^2 v(\theta_0)$ and let $\hat{\sigma}_{\tau_1,n}^2$ be the maximum likelihood estimator when using $\tau = \tau_1$

and $s = s_0$. Because $\sigma_1^2 v(\theta_1) = \sigma_0^2 v(\theta_0)$, it suffices to show $\sqrt{n}(\hat{\sigma}_{n,\tau_1}^2/\sigma_1^2 - 1) \to \mathcal{N}(0,2)$ in order to show that $\sqrt{n}(\hat{\sigma}_{\tau_1,n}^2 v(\theta_1) - \sigma_0^2 v(\theta_0)) \to \mathcal{N}(0, \sigma_0^4 v(\theta_0)^2)$. Now

$$\sqrt{n}\left(\frac{\hat{\sigma}_{\tau_1,n}^2}{\sigma_1^2} - 1\right) = \frac{1}{\sqrt{n}}\left[\frac{1}{\sigma_1^2}u(\boldsymbol{x})^\top K_{\theta_1}(\boldsymbol{x})^{-1}u(\boldsymbol{x}) - \frac{1}{\sigma_0^2}u(\boldsymbol{x})^\top K_{\theta_0}(\boldsymbol{x})^{-1}u(\boldsymbol{x})\right]$$
$$+ \sqrt{n}\left[\frac{u(\boldsymbol{x})^\top K_{\theta_0}(\boldsymbol{x})^{-1}u(\boldsymbol{x})}{n\sigma_0^2} - 1\right].$$

Asymptotic normality of the final summand follows from (57). The choice of σ_1 yields equivalence of the measures $\mathcal{N}(0, \sigma_1^2 v(\theta_1)(\tau_1-\Delta)^{s_0})$ and $\mathcal{N}(0, \sigma_0^2 v(\theta_0)(\tau_0-\Delta)^{s_0})$. Therefore the other summand tends to 0 by an application of Chebychev's inequality as the variances of

$$\frac{1}{\sigma_1^2}u(\boldsymbol{x})^\top K_{\theta_1}(\boldsymbol{x})^{-1}u(\boldsymbol{x}) - \frac{1}{\sigma_0^2}u(\boldsymbol{x})^\top K_{\theta_0}(\boldsymbol{x})^{-1}u(\boldsymbol{x})$$

are uniformly bounded; see for example Theorem 1 combined with display (2.9) in Ibragimov and Rozanov (1978). This yields the claim for arbitrary τ_1, and in particular for τ_L and τ_U. Let $\theta_L = (s_0, \tau_L)$ and $\theta_U = (s_0, \tau_U)$. Now, by construction $\hat{\tau}_n$ is a bounded sequence in $[\tau_L, \tau_U]$ and so the claim follows if we can show that $\hat{\sigma}_{\tau_L,n}v(\theta_L) \le \hat{\sigma}_{\hat{\tau}_n,n}v(\hat{\tau}_n) \le \hat{\sigma}_{\tau_U,n}v(\theta_U)$. For this purpose let $\tau_1 \le \tau_2$. Then

$$n\left[\hat{\sigma}_{\tau_2,n}^2 v(\theta_2) - \hat{\sigma}_{\tau_1,n}^2 v(\theta_1)\right] = u(\boldsymbol{x})^\top\left[v(\theta_2)K_{\theta_2}^{-1}(\boldsymbol{x}) - v(\theta_1)K_{\theta_1}^{-1}(\boldsymbol{x})\right]u(\boldsymbol{x}).$$

Thus it suffices to show that $v(\theta_2)K_{\theta_2}^{-1}(\boldsymbol{x}) - v(\theta_1)K_{\theta_1}^{-1}(\boldsymbol{x})$ is positive semi-definite. We have

$$\frac{K_{\theta_2}(\boldsymbol{x})}{v(\theta_2)} = \sum_{i=1}^\infty (\tau_2 + \lambda_i)^{-s_0}e_i(\boldsymbol{x})e_i(\boldsymbol{x})^\top \preceq \sum_{i=1}^\infty (\tau_1 + \lambda_i)^{-s_0}e_i(\boldsymbol{x})e_i(\boldsymbol{x})^\top = \frac{K_{\theta_1}(\boldsymbol{x})}{v(\theta_1)},$$

where for two matrices A, B we write $A \preceq B$ if $B - A$ is positive semi-definite. This yields the claim as $A \succeq B \iff A^{-1} \preceq B^{-1}$ for positive definite matrices A, B. $\qquad\square$

Appendix B. Infinite Product Measures and Kakutani's Theorem

In this section, we essentially repeat the content of Section 5 in Kakutani (1948) in order to state Kakutani's theorem and then provide some additional results pertaining to the setting of the present paper.

Let $\{(\Omega_i, \mathcal{F}_i, \mathbb{P}_i)\}$ be a sequence of measure spaces. Define Ω^∞ to be the set of sequences

$$\Omega^\infty = \{\{\omega_i\}_{i=1}^\infty \mid \omega_i \in \Omega_i\}$$

A subset R of Ω^∞ is called *rectangular* if it is of the form

$$R = \left(\prod_{i=1}^n B_i \right) \times \left(\prod_{i=n+1}^\infty \Omega_i \right)$$

with $B_i \in \mathcal{F}_i$, where the products $\prod$ and $\times$ are the Cartesian products. An *elementary* set is a subset of Ω^∞ that is the finite, disjoint union of rectangular sets. The family of elementary sets is a field (Kakutani, 1948, p. 217). Let $\mathcal{F}^\infty$ be the smallest σ-algebra containing all elementary sets. To define the infinite product measure $\mathbb{P} \coloneqq \bigotimes_{i=1}^\infty \mathbb{P}_i$, for a rectangular set R and an elementary set $E = \bigcup_{i=1}^n R_i$ we set

$$\tilde{\mathbb{P}}(R) = \prod_{i=1}^n \mathbb{P}_i(B_i) \quad \text{and} \quad \tilde{\mathbb{P}}(E) = \sum_{i=1}^n \tilde{\mathbb{P}}(R_i).$$

The measure $\mathbb{P}$ is then the extension (Kakutani, 1948, p. 217) of $\tilde{\mathbb{P}}$ to $\mathcal{F}^\infty$. Clearly, Ω^∞ is a rectangular set and thus if the $\mathbb{P}_i$ are probability measures, then so is $\mathbb{P}$.

To state Kakutani's theorem, consider the two infinite direct product measure spaces $(\Omega^\infty, \mathcal{F}^\infty, \mathbb{P}_1)$ and $(\Omega^\infty, \mathcal{F}^\infty, \mathbb{P}_2)$ constructed from the sequences of measures spaces $\{(\Omega_i, \mathcal{F}_i, \mathbb{P}_{1,i})\}$ and $\{(\Omega_i, \mathcal{F}_i, \mathbb{P}_{2,i})\}$. Recall that the Hellinger distance is defined as

$$\rho(\mathbb{P}_{1,i}, \mathbb{P}_{2,i}) = \int_{\Omega_i} \sqrt{\frac{\mathrm{d}\mathbb{P}_{1,i}}{\mathrm{d}\lambda}} \sqrt{\frac{\mathrm{d}\mathbb{P}_{2,i}}{\mathrm{d}\lambda}} \mathrm{d}\lambda.$$

Theorem B.1 (Kakutani's theorem from Kakutani, 1948). *Assume that the measures $\mathbb{P}_{1,i}$ and $\mathbb{P}_{2,i}$ are equivalent for all i. Then the infinite product measures $\mathbb{P}_1$ and $\mathbb{P}_2$ are either equivalent or orthogonal. They are orthogonal if and only if*

$$\prod_{i=1}^\infty \rho(\mathbb{P}_{1,i}, \mathbb{P}_{2,i}) = 0 \quad \text{or equivalently} \quad \sum_{i=1}^\infty -\log\left[\rho(\mathbb{P}_{1,i}, \mathbb{P}_{2,i})\right] < \infty.$$

Moreover,

$$\rho(\mathbb{P}_1, \mathbb{P}_2) = \prod_{i=1}^\infty \rho(\mathbb{P}_{1,i}, \mathbb{P}_{2,i}).$$

Next, we consider the special case where $\Omega_i = \mathbb{R}$, and $\mathcal{F}_i = \mathcal{B}(\mathbb{R})$. Let l_2 denote the set of square-summable sequences and $\|\cdot\|_{\hat{0}}$ the corresponding norm. Then $l_2 \subset \mathbb{R}^\infty$. By definition of $\mathcal{F}^\infty$, $a \mapsto \|a\|_{\hat{0}}^2$ is the countable sum of positive, measurable functions and thus measurable on $(\mathbb{R}^\infty, \mathcal{F}^\infty)$, see Lemmas 1.10 and 1.13 in Kallenberg (2002). Thus $l_2 = (\|\cdot\|_{\hat{0}})^{-1}([0, \infty))$ is measurable.

Proposition B.2. *Let $\mathcal{B}(l_2)$ denote the σ-algebra generated by the open sets with respect to the norm topology. Then $\mathcal{B}(l_2) = \mathcal{F}^\infty|_{l_2} \coloneqq \{A \cap l_2 \mid A \in \mathcal{F}^\infty\}$.*

Proof. To show $\mathcal{B}(l_2) \subset \mathcal{F}^\infty|_{l_2}$, reasoning as above, we note that for $b \in l_2$, the function $g(a) = \|a - b\|_{\hat{0}}$ is measurable and $B_\delta(b) = g^{-1}([0, \delta))$, where $B_\delta(b)$ is the ball of radius δ around $b \in l_2$. The claim now follows from separability of l_2. For the converse, note that the sets $l_2 \cap ((a_i, b_i) \times \prod_{j \neq i} \Omega_j)$ generate $\mathcal{F}^\infty|_{l_2}$ and are open in l_2. $\qquad \square$

Next, let

$$u = \sigma \sqrt{v(\theta)} \sum_{i=1}^\infty (\tau + \lambda_i)^{-s/2} \xi_i e_i,$$

where $\{\xi_i\}$ are i.i.d. random variables with finite second moment, and let $\mathbb{P}_{\sigma, \theta}$ be as in Section 5.

Proposition B.3. *For any $B \in \mathcal{B}(L_2(\mathcal{M}))$ we have*

$$\mathbb{P}(u \in B) = \mathbb{P}_{\sigma, \theta}(\iota(B)).$$

Proof. Because ι^{-1} is an isomorphism, $\iota(B)$ is measurable. By definition, $\mathbb{P}(u \in B) = \mathbb{P}(\hat{u} \in \iota(B))$. Let $R = \prod_{i=1}^n B_i \times \prod_{i=n+1}^\infty \mathbb{R}$ be a rectangular set. Then

$$\mathbb{P}(\hat{u} \in R) = \mathbb{P}\left(\bigcap_{i=1}^n \left\{\sigma \sqrt{v(\theta)}(\tau + \lambda_i)^{s/2} \xi_i \in B_i\right\}\right) = \prod_{i=1}^n \mathbb{P}_{\sigma \lambda_i(\theta)}(B_i) = \mathbb{P}_{\sigma, \theta}(R).$$

From this the claim follows for elementary sets. By definition they generate $\mathcal{F}^\infty$ and so by the uniqueness of the extension

$$\mathbb{P}(\hat{u} \in A) = \mathbb{P}_{\sigma, \theta}(A)$$

for any measurable set A, in particular for $A = \iota(B)$ where $B \in \mathcal{B}(L_2(\mathcal{M}))$. $\qquad \square$

References

Abramowitz, M., Stegun, I.A., 1964. Handbook of mathematical functions with formulas, graphs, and mathematical tables. volume No. 55 of *National Bureau of Standards Applied Mathematics Series*. U. S. Government Printing Office, Washington, DC.

Anderes, E., 2010. On the consistent separation of scale and variance for Gaussian random fields. Ann. Stat. 38, 870 – 893. URL: `https://doi.org/10.1214/09-AOS725`, doi:`10.1214/09-AOS725`.

Aubin, T., 1998. Some nonlinear problems in Riemannian geometry. Springer Monographs in Mathematics, Springer-Verlag, Berlin. URL: `https://doi.org/10.1007/978-3-662-13006-3`, doi:`10.1007/978-3-662-13006-3`.

Bachoc, F., 2014. Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of gaussian processes. J. Multivariate Anal.

125, 1–35. URL: https://www.sciencedirect.com/science/article/pii/S0047259X13002571, doi:https://doi.org/10.1016/j.jmva.2013.11.015.

Bevilacqua, M., Faouzi, T., Furrer, R., Porcu, E., 2019. Estimation and prediction using generalized Wendland covariance functions under fixed domain asymptotics. Ann. Stat. 47, 828–856. URL: https://doi.org/10.1214/17-AOS1652, doi:10.1214/17-AOS1652.

Bolin, D., Kirchner, K., 2023. Equivalence of measures and asymptotically optimal linear prediction for Gaussian random fields with fractional-order covariance operators. Bernoulli 29, 1476 – 1504. URL: https://doi.org/10.3150/22-BEJ1507, doi:10.3150/22-BEJ1507.

Bolin, D., Simas, A., Wallin, J., 2023. Statistical inference for gaussian whittle–matérn fields on metric graphs. arXiv:2304.10372v1 arXiv:2304.10372.

Borovitskiy, V., Terenin, A., Mostowsky, P., Deisenroth (he/him), M., 2020. Matérn Gaussian processes on Riemannian manifolds, in: Larochelle, H., Ranzato, M., Hadsell, R., Balcan, M., Lin, H. (Eds.), Adv. Neur. In., Curran Associates, Inc.. pp. 12426–12437. URL: https://proceedings.neurips.cc/paper_files/paper/2020/file/92bf5e6240737e0326ea59846a83e076-Paper.pdf.

Bull, A.D., 2012. Honest adaptive confidence bands and self-similar functions. Electron. J. Stat. 6, 1490–1516. URL: https://doi.org/10.1214/12-EJS720, doi:10.1214/12-EJS720.

Chen, Y., Owhadi, H., Stuart, A.M., 2021. Consistency of empirical Bayes and kernel flow for hierarchical parameter estimation. Math. Comput. 90, 2527–2578.

Craioveanu, M., Puta, M., Rassias, T.M., 2001. Old and new aspects in spectral geometry. volume 534 of *Mathematics and its Applications*. Kluwer Academic Publishers, Dordrecht. URL: https://doi.org/10.1007/978-94-017-2475-3, doi:10.1007/978-94-017-2475-3.

Davies, E.B., 1995. Spectral theory and differential operators. volume 42 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge. URL: https://doi.org/10.1017/CBO9780511623721, doi:10.1017/CBO9780511623721.

Deserno, M., 2004. How to generate equidistributed points on the surface of a sphere. https://www.cmu.edu/biolphys/deserno/pdf/sphere_equi.pdf. Accessed: 2024-03-14.

Du, J., Zhang, H., Mandrekar, V.S., 2009. Fixed-domain asymptotic properties of tapered maximum likelihood estimators. Ann. Stat. 37, 3330–3361. URL: https://doi.org/10.1214/08-AOS676, doi:10.1214/08-AOS676.

Ferreira, J.C., Menegatto, V.A., 2013. Positive definiteness, reproducing kernel Hilbert spaces and beyond. Ann. Funct. Anal. 4, 64–88. URL: `https://doi.org/10.15352/afa/1399899838`, doi:`10.15352/afa/1399899838`.

Fuselier, E., Wright, G.B., 2012. Scattered data interpolation on embedded submanifolds with restricted positive definite kernels: Sobolev error estimates. SIAM J. Numer. Anal. 50, 1753–1776. URL: `https://doi.org/10.1137/110821846`, doi:`10.1137/110821846`, arXiv:`https://doi.org/10.1137/110821846`.

Giné, E., Nickl, R., 2010. Confidence bands in density estimation. Ann. Stat. 38, 1122–1170. URL: `https://doi.org/10.1214/09-AOS738`, doi:`10.1214/09-AOS738`.

Gneiting, T., 2002. Compactly supported correlation functions. J. Multivariate Anal. 83, 493–508. URL: `https://doi.org/10.1006/jmva.2001.2056`, doi:`10.1006/jmva.2001.2056`.

Gneiting, T., 2013. Strictly and non-strictly positive definite functions on spheres. Bernoulli 19, 1327–1349. URL: `https://doi.org/10.3150/12-BEJSP06`, doi:`10.3150/12-BEJSP06`.

Große, N., Schneider, C., 2013. Sobolev spaces on Riemannian manifolds with bounded geometry: general coordinates and traces. Math. Nachr. 286, 1586–1613. URL: `https://doi.org/10.1002/mana.201300007`, doi:`10.1002/mana.201300007`.

Guinness, J., Fuentes, M., 2016. Isotropic covariance functions on spheres: some properties and modeling considerations. J. Multivariate Anal. 143, 143–152. URL: `https://doi.org/10.1016/j.jmva.2015.08.018`, doi:`10.1016/j.jmva.2015.08.018`.

Hadji, A., Szabó, B., 2021. Can we trust Bayesian uncertainty quantification from Gaussian process priors with squared exponential covariance kernel? SIAM/ASA J. Uncertain. Quantif. 9, 185–230. URL: `https://doi.org/10.1137/19M1253010`, doi:`10.1137/19M1253010`.

Hangelbroek, T., Narcowich, F.J., Rieger, C., Ward, J.D., 2018. Direct and inverse results on bounded domains for meshless methods via localized bases on manifolds, in: Contemporary computational mathematics—a celebration of the 80th birthday of Ian Sloan. Vol. 1, 2. Springer, Cham, pp. 517–543.

Hangelbroek, T., Narcowich, F.J., Ward, J.D., 2010. Kernel approximation on manifolds I: bounding the Lebesgue constant. SIAM J. Math. Anal. 42, 1732–1760. URL: `https://doi.org/10.1137/090769570`, doi:`10.1137/090769570`.

Henderson, I., 2023. Sobolev regularity of Gaussian random fields. J. Funct. Anal. 286, 110241.

Hsing, T., Eubank, R., 2015. Theoretical foundations of functional data analysis, with an introduction to linear operators. Wiley Series in Probability and Statistics, John

Wiley & Sons, Ltd., Chichester. URL: https://doi.org/10.1002/9781118762547, doi:10.1002/9781118762547.

Hubbert, S., Jäger, J., 2023. Generalised Wendland functions for the sphere. Adv. Comput. Math. 49, Paper No. 3, 21. URL: https://doi.org/10.1007/s10444-022-10005-z, doi:10.1007/s10444-022-10005-z.

Ibragimov, I.d.A., Rozanov, Y.A., 1978. Gaussian random processes. volume 9 of *Applications of Mathematics*. Springer-Verlag, New York-Berlin. Translated from the Russian by A. B. Aries.

Kakutani, S., 1948. On equivalence of infinite product measures. Ann. Math. 49, 214–224. URL: https://doi.org/10.2307/1969123, doi:10.2307/1969123.

Kallenberg, O., 2002. Foundations of modern probability. Probability and its Applications (New York). second ed., Springer-Verlag, New York. URL: https://doi.org/10.1007/978-1-4757-4015-8, doi:10.1007/978-1-4757-4015-8.

Kanagawa, M., Hennig, P., Sejdinovic, D., Sriperumbudur, B.K., 2018. Gaussian processes and kernel methods: a review on connections and equivalences. arXiv:1805.08845v1 .

Karvonen, T., 2021. Estimation of the scale parameter for a misspecified Gaussian process model. arXiv:2110.02810v2 .

Karvonen, T., 2023. Asymptotic bounds for smoothness parameter estimates in Gaussian process interpolation. SIAM/ASA Journal on Uncertainty Quantification 11, 1225–1257. doi:10.1137/22M149288X.

Karvonen, T., Wynne, G., Tronarp, F., Oates, C., Särkkä, S., 2020. Maximum likelihood estimation and uncertainty quantification for Gaussian process approximation of deterministic functions. SIAM/ASA J. Uncertain. Quantif. 8, 926–958. URL: https://doi.org/10.1137/20M1315968, doi:10.1137/20M1315968.

Kaufman, C.G., Shaby, B.A., 2013. The role of the range parameter for estimation and prediction in geostatistics. Biometrika 100, 473–484. URL: https://doi.org/10.1093/biomet/ass079, doi:10.1093/biomet/ass079.

Kerkyacharian, G., Ogawa, S., Petrushev, P., Picard, D., 2018. Regularity of Gaussian processes on Dirichlet spaces. Constr. Approx. 47, 277–320. URL: https://doi.org/10.1007/s00365-018-9416-8, doi:10.1007/s00365-018-9416-8.

Kim, S., 2020. Fractional order Sobolev spaces for the Neumann Laplacian and the vector Laplacian. J. Korean Math. Soc. 57, 721–745. URL: https://doi.org/10.4134/JKMS.j190351, doi:10.4134/JKMS.j190351.

Krieg, D., Sonnleitner, M., 2022. Function recovery on manifolds using scattered data. arXiv:2109.04106v2 .

Lan, X., Marinucci, D., Xiao, Y., 2018. Strong local nondeterminism and exact modulus of continuity for spherical Gaussian fields. Stochastic Process. Appl. 128, 1294–1315. URL: `https://doi.org/10.1016/j.spa.2017.07.008`, doi:`10.1016/j.spa.2017.07.008`.

Lang, A., Schwab, C., 2015. Isotropic Gaussian random fields on the sphere: regularity, fast simulation and stochastic partial differential equations. Ann. Appl. Probab. 25, 3047–3094. URL: `https://doi.org/10.1214/14-AAP1067`, doi:`10.1214/14-AAP1067`.

Lee, J.M., 2013. Introduction to smooth manifolds. volume 218 of *Graduate Texts in Mathematics*. Second ed., Springer, New York.

Lee, J.M., 2018. Introduction to Riemannian manifolds. volume 176 of *Graduate Texts in Mathematics*. Second ed., Springer, Cham.

Li, C., 2022. Bayesian fixed-domain asymptotics for covariance parameters in a Gaussian process model. Ann. Stat. 50, 3334–3363. URL: `https://doi.org/10.1214/22-aos2230`, doi:`10.1214/22-aos2230`.

Li, D., Tang, W., Banerjee, S., 2023. Inference for Gaussian processes with Matérn covariogram on compact Riemannian manifolds. J. Mach. Learn. Res. 24, Paper No. [101], 26.

Lindgren, F., Rue, H., Lindström, J., 2011. An explicit link between gaussian fields and gaussian markov random fields: the stochastic partial differential equation approach. J. Roy. Stat. Soc. B 73, 423–498. URL: `https://rss.onlinelibrary.wiley.com/doi/abs/10.1111/j.1467-9868.2011.00777.x`, doi:`https://doi.org/10.1111/j.1467-9868.2011.00777.x`, arXiv:`https://rss.onlinelibrary.wiley.com/doi/pdf/10.1111/j.1467-9868.2011.00777.x`.

Loh, W.L., Sun, S., 2023. Estimating the parameters of some common Gaussian random fields with nugget under fixed-domain asymptotics. Bernoulli 29, 2519–2543. URL: `https://doi.org/10.3150/22-bej1551`, doi:`10.3150/22-bej1551`.

Loh, W.L., Sun, S., Wen, J., 2021. On fixed-domain asymptotics, parameter estimation and isotropic Gaussian random fields with Matérn covariance functions. Ann. Stat. 49, 3127–3152. URL: `https://doi.org/10.1214/21-aos2077`, doi:`10.1214/21-aos2077`.

Mardia, K.V., Marshall, R.J., 1984. Maximum likelihood estimation of models for residual covariance in spatial regression. Biometrika 71, 135–146. URL: `https://doi.org/10.1093/biomet/71.1.135`, doi:`10.1093/biomet/71.1.135`.

Matérn, B., 1960. Spatial variation: Stochastic models and their application to some problems in forest surveys and other sampling investigations. Statens Skogsforskningsinstitut, Stockholm. Meddelanden Fran Statens Skogsforskningsinstitut, Band 49, Nr. 5.

Meerschaert, M.M., Wang, W., Xiao, Y., 2013. Fernique-type inequalities and moduli of continuity for anisotropic Gaussian random fields. Trans. Amer. Math. Soc. 365, 1081–1107. URL: `https://doi.org/10.1090/S0002-9947-2012-05678-9`, doi:`10.1090/S0002-9947-2012-05678-9`.

Narcowich, F.J., Ward, J.D., 2002. Scattered data interpolation on spheres: error estimates and locally supported basis functions. SIAM J. Math. Anal. 33, 1393–1410. URL: `https://doi.org/10.1137/S0036141001395054`, doi:`10.1137/S0036141001395054`.

Narcowich, F.J., Ward, J.D., Wendland, H., 2006. Sobolev error estimates and a Bernstein inequality for scattered data interpolation via radial basis functions. Constr. Approx. 24, 175–186. URL: `https://doi.org/10.1007/s00365-005-0624-7`, doi:`10.1007/s00365-005-0624-7`.

Naslidnyk, M., Kanagawa, M., Karvonen, T., Mahsereci, M., 2023. Comparing scale parameter estimators for Gaussian process regression: cross validation and maximum likelihood. arXiv:2307.07466v1 .

Petit, S., 2023. Maximum likelihood estimation and prediction error for a Matérn model on the circle. arXiv:2209.07791 .

Porcu, E., Bevilacqua, M., Schaback, R., Oates, C., 2023. The matérn model: A journey through statistics, numerical analysis and machine learning. arXiv.2303.02759 .

Reed, M., Simon, B., 1980. Methods of modern mathematical physics. I. Second ed., Academic Press, Inc. [Harcourt Brace Jovanovich, Publishers], New York. Functional analysis.

Sanz-Alonso, D., Yang, R., 2022. The SPDE approach to Matérn fields: graph representations. Statist. Sci. 37, 519–540. URL: `https://doi.org/10.1214/21-sts838`, doi:`10.1214/21-sts838`.

Schaback, R., 1999. Native Hilbert spaces for radial basis functions. I, in: New developments in approximation theory (Dortmund, 1998). Birkhäuser, Basel. volume 132 of *Internat. Ser. Numer. Math.*, pp. 255–282.

Scheuerer, M., 2010. Regularity of the sample paths of a general second order random field. Stochastic Process. Appl. 120, 1879–1897. URL: `https://doi.org/10.1016/j.spa.2010.05.009`, doi:`10.1016/j.spa.2010.05.009`.

Skrzypczak, L., 1990. Traces of function spaces of $F_{p,q}^s - B_{p,q}^s$ type on submanifolds. Math. Nachr. 146, 137–147. URL: https://doi.org/10.1002/mana.19901460707, doi:10.1002/mana.19901460707.

Sniekers, S., van der Vaart, A., 2015a. Adaptive Bayesian credible sets in regression with a Gaussian process prior. Electron. J. Stat. 9, 2475–2527. URL: https://doi.org/10.1214/15-EJS1078, doi:10.1214/15-EJS1078.

Sniekers, S., van der Vaart, A., 2015b. Credible sets in the fixed design model with Brownian motion prior. J. Statist. Plann. Inference 166, 78–86. URL: https://doi.org/10.1016/j.jspi.2014.07.008, doi:10.1016/j.jspi.2014.07.008.

Sniekers, S., van der Vaart, A., 2020. Adaptive Bayesian credible bands in regression with a Gaussian process prior. Sankhya A 82, 386–425. URL: https://doi.org/10.1007/s13171-019-00185-0, doi:10.1007/s13171-019-00185-0.

Stein, E.M., 1970. Singular integrals and differentiability properties of functions. Princeton University Press, Princeton, N.J.,.

Stein, E.M., Weiss, G., 1971. Introduction to Fourier analysis on Euclidean spaces. volume No. 32 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ.

Stein, M.L., 1999. Interpolation of spatial data. Springer Series in Statistics, Springer-Verlag, New York. URL: https://doi.org/10.1007/978-1-4612-1494-6, doi:10.1007/978-1-4612-1494-6. some theory for Kriging.

Steinwart, I., 2019. Convergence types and rates in generic Karhunen-Loève expansions with applications to sample path properties. Potential Anal. 51, 361–395. URL: https://doi.org/10.1007/s11118-018-9715-5, doi:10.1007/s11118-018-9715-5.

Strichartz, R.S., 1983. Analysis of the Laplacian on the complete Riemannian manifold. J. Funct. Anal. 52, 48–79. URL: https://doi.org/10.1016/0022-1236(83)90090-3, doi:10.1016/0022-1236(83)90090-3.

Szabó, B., van der Vaart, A.W., van Zanten, J.H., 2015. Frequentist coverage of adaptive nonparametric Bayesian credible sets. Ann. Stat. 43, 1391–1428. URL: https://doi.org/10.1214/14-AOS1270, doi:10.1214/14-AOS1270.

Taylor, M.E., 1981. Pseudodifferential operators. volume No. 34 of *Princeton Mathematical Series*. Princeton University Press, Princeton, NJ.

Travis, L., Ray, K., 2023. Pointwise uncertainty quantification for sparse variational Gaussian process regression with a Brownian motion prior, in: Adv. Neur. In. To appear.

Triebel, H., 1992. Theory of function spaces. II. volume 84 of *Monographs in Mathematics*. Birkhäuser Verlag, Basel. URL: `https://doi.org/10.1007/978-3-0346-0419-2`, doi:`10.1007/978-3-0346-0419-2`.

van der Vaart, A., van Zanten, H., 2011. Information rates of nonparametric Gaussian process methods. J. Mach. Learn. Res. 12, 2095–2119.

Wainwright, M.J., 2019. High-dimensional statistics: a non-asymptotic viewpoint. Cambridge Series in Statistical and Probabilistic Mathematics, Cambridge University Press. doi:`10.1017/9781108627771`.

Wang, D., Loh, W.L., 2011. On fixed-domain asymptotics and covariance tapering in Gaussian random field models. Electron. J. Stat. 5, 238–269. URL: `https://doi.org/10.1214/11-EJS607`, doi:`10.1214/11-EJS607`.

Wang, W., Tuo, R., Wu, C.F.J., 2020. On prediction properties of Kriging: uniform error bounds and robustness. J. Amer. Statist. Assoc. 115, 920–930. URL: `https://doi.org/10.1080/01621459.2019.1598868`, doi:`10.1080/01621459.2019.1598868`.

Wendland, H., 2004. Scattered data approximation. Cambridge Monographs on Applied and Computational Mathematics, Cambridge University Press. doi:`10.1017/CBO9780511617539`.

Whittle, P., 1963. Stochastic processes in several dimensions. Bull. Inst. Internat. Statist. 40, 974–994.

Xu, W., Stein, M.L., 2017. Maximum likelihood estimation for a smooth Gaussian random field model. SIAM/ASA J. Uncertain. Quantif. 5, 138–175. URL: `https://doi.org/10.1137/15M105358X`, doi:`10.1137/15M105358X`.

Ying, Z., 1991. Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. J. Multivariate Anal. 36, 280–296. URL: `https://doi.org/10.1016/0047-259X(91)90062-7`, doi:`10.1016/0047-259X(91)90062-7`.

Zhang, H., 2004. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. J. Amer. Statist. Assoc. 99, 250–261. URL: `https://doi.org/10.1198/016214504000000241`, doi:`10.1198/016214504000000241`.

Zhang, H., Zimmerman, D.L., 2005. Towards reconciling two asymptotic frameworks in spatial statistics. Biometrika 92, 921–936. URL: `https://doi.org/10.1093/biomet/92.4.921`, doi:`10.1093/biomet/92.4.921`.