



An asymptotic study of the joint maximum likelihood estimation of the regularity and the amplitude parameters of a Matérn model on the circle

Sébastien Petit

► To cite this version:

Sébastien Petit. An asymptotic study of the joint maximum likelihood estimation of the regularity and the amplitude parameters of a Matérn model on the circle. 2023. hal-03778527v6

HAL Id: hal-03778527

<https://centralesupelec.hal.science/hal-03778527v6>

Preprint submitted on 22 Sep 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

An asymptotic study of the joint maximum likelihood estimation of the regularity and the amplitude parameters of a Matérn model on the circle

Sébastien J. Petit^{*1, 2}

¹Laboratoire National de Métrologie et d'Essais, 78197 Trappes Cedex, France

² Université Paris-Saclay, CNRS, CentraleSupélec,
Laboratoire des signaux et systèmes, 91190, Gif-sur-Yvette, France

September 22, 2023

Abstract

This work considers parameter estimation for Gaussian process interpolation with a periodized version of the Matérn covariance function introduced by Stein. Convergence rates are studied for the joint maximum likelihood estimation of the regularity and the amplitude parameters when the data are sampled according to the model. The mean integrated squared error is also analyzed with fixed and estimated parameters, showing that maximum likelihood estimation yields asymptotically the same error as if the ground truth was known. Finally, the case where the observed function is a fixed deterministic element of a Sobolev space of continuous functions is also considered, suggesting that a joint estimation does not select the regularity parameter as if the amplitude were fixed.

1 Introduction

Gaussian process interpolation or kriging is a common technique for inferring an unknown function from noiseless data, which has applications in geostatistics (Stein, 1999), computer experiments (Santner et al., 2003), and machine learning (Rasmussen and Williams, 2006). A covariance function fully characterizes a zero-mean Gaussian process model. The need for tailoring this function to the task at hand is widely acknowledged in the literature. The common practice consists in choosing it within a parametric family. Stein (1999) promotes using the Matérn (1986) family of stationary covariance functions. Assuming isotropy and using the parameterization from Stein (1999, p. 31), this family is defined on $\mathbb{R}^d$ by its spectrum:

$$\widehat{k}: \omega \in \mathbb{R}^d \mapsto \frac{\phi}{(\alpha^2 + \|\omega\|^2)^{\nu+d/2}}, \quad (1)$$

^{*}sebastien.petit@lne.fr

which is indexed by three parameters: the *regularity* parameter ν , what we shall call the *amplitude* parameter ϕ , and the parameter α . See (Stein, 1999) for a comprehensive description of the effect of these parameters. In short, the parameter ν is shown to be the key quantity governing the asymptotics of the prediction error. The amplitude parameter ϕ does not impact the posterior mean predictions but matters for uncertainty quantification, whereas α is less important asymptotically.

One can safely say that cross-validation and maximum likelihood estimation are the most popular techniques for selecting Gaussian process parameters from data. We shall focus on the latter for the rest of this article.

Assuming observations from a Matérn process with parameter $\theta_0 = (\nu_0, \phi_0, \alpha_0)$, a distinction is often made between increasing and fixed-domain asymptotic frameworks (see, e.g., Bachoc, 2021, for a review). While several increasing-domain asymptotic frameworks have been exhaustively studied (see, e.g., Mardia and Marshall, 1984; Bachoc, 2014), a comprehensive asymptotic analysis of maximum likelihood estimation in fixed-domain frameworks—i.e., on bounded domains—has long been an open question. Previous works mainly consider the estimation of ϕ and α for a known ν (see, e.g., Ying, 1991, 1993; van der Vaart, 1996; Zhang, 2004; Loh, 2006; Kaufman and Shaby, 2013; Li, 2020, who often use alternative parametrizations).

The asymptotics of $\hat{\nu}_n$ seem to have been little studied. Stein (1999, Section 6.7) proposes an asymptotic framework with equispaced observations on the torus and makes a conjecture about the asymptotic behavior of the joint maximum likelihood estimate $\hat{\theta}_n = (\hat{\nu}_n, \hat{\phi}_n, \hat{\alpha}_n)$ based on the Fisher information matrix (see also Stein, 1993, in the case of noisy observations). This topic has only recently regained popularity. Indeed, Chen et al. (2021) used the framework from Stein to show that $\hat{\nu}_n$ is consistent if the other parameters remain fixed (i.e., enforced to arbitrary values, which may not be ϕ_0 and α_0). Continuing with fixed ϕ and α , Karvonen (2023) has recently shown that $\liminf \hat{\nu}_n \geq \nu_0$ in the (more general) case of quasi-uniform observations on a “nice” bounded domain of $\mathbb{R}^d$.

Another long-standing open problem (see notably Putter and Young 2001 and Stein 1999, in the Preface) is that of predictions with estimated parameters: how accurate and reliable are the predictions if one selects a parameter $\hat{\theta}_n$ from data and uses it to make subsequent predictions? The critical influence of ν on the kriging error suggests that the asymptotic behavior of $\hat{\nu}_n$ is a key element in answering this question.

Another research line consists in studying parameters estimation assuming observations from a fixed deterministic function f . The definition of a ground truth θ_0 is not obvious in this setting. Instead, the aim is to study which “features” of f are used by the estimator to select a Gaussian process model and how this affects predictions. See (Karvonen et al., 2020; Karvonen and Oates, 2023) for analyses of maximum likelihood estimators of other parameters given a fixed regularity. Regarding $\hat{\nu}_n$, the tight lower bound shown by Karvonen (2023) also covers the case of a continuous function from a Sobolev space. The result shows an interesting connection with sample path properties. More precisely, define the smoothness $\nu_0(f)$ of f in a Sobolev sense so that $\nu_0(\xi) = \nu_0$ holds almost surely for any Matérn process ξ with regularity ν_0 . For fixed ϕ and α , Karvonen (2023) showed that $\liminf \hat{\nu}_n \geq \nu_0(f)$ and, under (essentially) a self-similarity

hypothesis on the spectrum of f , that $\hat{\nu}_n$ converges to $\nu_0(f)$. This means that, if the spectrum of f is well-behaved, then maximum likelihood estimation fits ν so that f and the sample paths have the same Sobolev smoothness. It echoes similar findings in Bayesian nonparametric statistics with noise-corrupted observations (see notably [Belitser and Ghosal 2003](#), [Knapik et al. 2016](#), p. 779, and [Szabó et al. 2015](#), pp. 1397 and 1404), where, with our notations, similar conditions on the truth imply that $\hat{\nu}_n \rightarrow \nu_0(f)$.

This article focuses on the one-dimensional version of the framework proposed by [Stein \(1999, Section 6.7\)](#) to analyze the joint maximum likelihood estimation of (ν, ϕ, α) . This simplified framework is very convenient for such a study, as explained in [Section 3](#). On the one hand, a $\sqrt{n}$ -rate asymptotic normality result is shown for $(\hat{\nu}_n, \hat{\phi}_n)$ when observing a Matérn process. Whether the (non-identifiable) parameter α is known or estimated does not affect the limiting distribution. Furthermore, one consequence is that the ratio between the mean squared error with estimated parameters and the one with known parameters converges to unity. On the other hand, it is shown that a joint estimation does not result in the behavior discussed in the previous paragraph. The key takeaway is that only the smaller asymptotic bound $\liminf \hat{\nu}_n \geq \nu_0(f) - 1/2$ holds. This means that the reproducing kernel Hilbert space is asymptotically too small to contain f but does not say whether the Sobolev smoothness of the sample paths exceeds or converges to $\nu_0(f)$. To give a quantitative description of the behavior above $\nu_0(f) - 1/2$, we derive the large sample limit of the (profile) likelihood on a class of functions that is small but satisfies the usual spectrum conditions ensuring that $\hat{\nu}_n \rightarrow \nu_0(f)$ for fixed ϕ and α . The minimizer of this limit has no closed-form expression (see [\(16\)](#)), but we show that a numerical approximation is not maximized by $\nu_0(f)$. A strong consistency result on sample paths shows that the set of functions f such that $\hat{\nu}_n \rightarrow \nu_0(f)$ has probability one under a Matérn process.

To summarize, the contributions of the present article are threefold. First, we prove consistency and asymptotic normality results on the maximum likelihood estimates of the parameters ν and ϕ . Then, we leverage these convergence rates to analyze the expected integrated error, showing that estimating the parameters yields the same error asymptotically as if the ground truth was known. Finally, we investigate model selection by maximum likelihood estimation on a deterministic function.

The article is organized as follows. [Section 2](#) introduces the periodic framework and our notations and [Section 3](#) discusses how this framework helps for circumventing the challenges posed by the study of the profile likelihood. Then, [Section 4](#) gives the main results. Finally, [Section 5](#) provides our results on the deterministic case.

2 Gaussian process interpolation on the circle

2.1 Framework

Let $f : [0, 1] \rightarrow \mathbb{R}$ be a continuous periodic function observed on a regular grid: $\{j/n, 0 \leq j \leq n-1\}$. Consider the periodic version of the Matérn family of stationary covariance functions [\(1\)](#) introduced by [Stein \(1999, Section 6.7\)](#) and

defined by the uniformly absolutely convergent Fourier series

$$k_\theta : x \in \mathbb{R} \mapsto \sum_{j \in \mathbb{Z}} \mathfrak{L}_j(\theta) e^{2\pi i x j}$$

with coefficients:

$$\mathfrak{L}_j(\theta) = \frac{\phi}{(\alpha^2 + j^2)^{\nu+1/2}}, \quad \text{for } j \in \mathbb{Z} \text{ and } \theta = (\nu, \phi, \alpha) \in (0, +\infty)^3. \quad (2)$$

The function k_θ is continuous and strictly positive definite (see, e.g., [Gneiting, 2011](#), Theorem 1). The description of the parameters ν , ϕ , and α from the [Introduction](#) carries to this periodic one-dimensional version. A specificity is that α is not identifiable as different values yield equivalent probability measures. However, ν and ϕ are identifiable (see, e.g., [Stein, 1999](#), Chapter 4 and Section 6.7).

Assuming a centered process, the usual task in Gaussian process interpolation is to use the model $\xi \sim \text{GP}(0, k_\theta)$ to infer the function f from the noiseless data

$$Z = (f(0), f(1/n), \dots, f(1 - 1/n))^T. \quad (3)$$

The function f is usually predicted using the posterior mean function given by the kriging equations ([Matheron, 1971](#)). This predictor can be written simply in the framework presented above.

Proposition 2.1. Let $n \geq 1$ and $f : [0, 1] \rightarrow \mathbb{R}$ be a continuous periodic function with absolutely summable Fourier coefficients $c_j(f)$. Writing $\hat{f}_n$ for the posterior mean function given Z and the parameter θ , we have:

$$\hat{f}_n(x) = \sum_{j \in \mathbb{Z}} \left(\frac{\sum_{q \in j+n\mathbb{Z}} c_q(f)}{\sum_{q \in j+n\mathbb{Z}} \mathfrak{L}_q(\theta)} \right) \mathfrak{L}_j(\theta) e^{2\pi i x j} \quad \text{for } x \in [0, 1]. \quad (4)$$

The convergence of (4) holds uniformly absolutely.

The proof is deferred to Section [A.4](#) of the [Appendix](#).

The expression (4) shows how the posterior mean function approximates f : it transforms the Fourier coefficients of k_θ into those of f using the ratio of their discrete Fourier transforms. Finally, we also define the integrated squared error:

$$\text{ISE}_n(\nu, \alpha; f) = \int_0^1 (f - \hat{f}_n)^2. \quad (5)$$

Note that it does not depend on ϕ .

2.2 Maximum likelihood estimation

Given the observations Z and $\Theta \subset (0, +\infty)^3$, a maximum likelihood estimate is defined by $\hat{\theta}_n = (\hat{\nu}_n, \hat{\phi}_n, \hat{\alpha}_n)$ minimizing (a linear transform of) the negative log-likelihood:

$$\mathbb{L}_n : \theta \in \Theta \mapsto n^{-1} (\ln(\det(K_\theta)) + Z^T K_\theta^{-1} Z), \quad (6)$$

with ties broken arbitrarily and K_θ the covariance matrix of Z according to k_θ .

The estimators $\hat{\nu}_n$ and $\hat{\alpha}_n$ are assumed bounded in this work, i.e., we take $\Theta = N \times (0, +\infty) \times A$ with N and A compact intervals. However, keeping $\hat{\phi}_n$ unbounded is key to our main results and for discussing the deterministic case in Section 5. Write $K_\theta = \phi R_{\nu,\alpha}$ for $\theta = (\nu, \phi, \alpha) \in (0, +\infty)^3$. The following proposition gives an expression for the *profile* likelihood, i.e., the infimum of $\mathbb{L}_n(\nu, \phi, \alpha)$ with respect to $\phi \in (0, +\infty)$ for fixed ν and α .

Proposition 2.2. (see, e.g., [Santner et al., 2003](#), Section 3.3.2) Let $\nu, \alpha > 0$. It holds that

$$\inf_{\phi > 0} \mathbb{L}_n(\nu, \phi, \alpha) = 1 + n^{-1} \ln(\det(R_{\nu,\alpha})) + \ln \left(\frac{Z^\top R_{\nu,\alpha}^{-1} Z}{n} \right). \quad (7)$$

Moreover, if Z is nonzero, then the infimum is uniquely reached by $\hat{\phi}_n = Z^\top R_{\nu,\alpha}^{-1} Z / n$.

(The case $Z = 0$ is covered since both sides of (7) match.)

3 Studying the profile likelihood using discrete Fourier transforms

3.1 Linking the spectra of k_θ and K_θ

As [Craven and Wahba \(1979\)](#) and [Stein \(1999, Section 6.7\)](#) point out, the framework introduced in Section 2.1 is convenient. More precisely, it provides a natural link between K_θ and k_θ using discrete Fourier transforms (see Section A.2 of the [Appendix](#) for details). In particular, it gives a closed-form identity

$$n^{-1} \phi \lambda_{m,n} = \sum_{j \in \mathbb{Z}} \underline{\mathcal{L}}_{m+nj}(\theta)$$

linking the eigenvalues $\phi \lambda_{0,n}, \dots, \phi \lambda_{n-1,n}$ of K_θ to those of k_θ given by (2). Furthermore, the matrices K_θ share the same eigenvectors. One has $n^{-1} \phi \lambda_{m,n} \rightarrow \underline{\mathcal{L}}_m(\theta)$ for a fixed m but the ratio $n^{-1} \phi \lambda_{m,n} / \underline{\mathcal{L}}_m(\theta)$ remains bounded away from one for m close to $n/2$.

3.2 The consistency of $\hat{\nu}_n$ for fixed ϕ and α ([Chen et al., 2021](#))

Assuming observations from $\xi \sim \text{GP}(0, k_{\theta_0})$ under a similar model with $\theta_0 = (\nu_0, \phi_0, \alpha_0) \in (0, +\infty)^3$, [Chen et al. \(2021\)](#) show the consistency of $\hat{\nu}_n$ for equispaced observations on the d -dimensional torus for fixed parameters ϕ and α . A sketch of their reasoning for $d = 1$ is provided in this paragraph. The spectrum of K_θ is studied by showing that

$$n^{-1} \phi \lambda_{m,n} = e^{\mathcal{O}(1)} \underline{\mathcal{L}}_m(\theta) = e^{\mathcal{O}(1)} m^{-2\nu-1} \quad (8)$$

uniformly in ν and $1 \leq m \leq n/2$.¹ This approximation yields:

$$\begin{cases} \ln(\det(K_\theta)) &= -2\nu n \ln(n) + n \ln(\phi) + n\mathcal{O}(1), \\ Z^\top K_\theta^{-1} Z &= \phi^{-1} \phi_0 \mathcal{O}_{\mathbb{P}}(\ln(n)) \text{ if } \nu \leq \nu_0 - 1/2, \\ Z^\top K_\theta^{-1} Z &= \phi^{-1} \phi_0 e^{\mathcal{O}_{\mathbb{P}}(1)} n^{1+2(\nu-\nu_0)} \text{ if } \nu > \nu_0 - 1/2, \end{cases} \quad (9)$$

¹The $\lambda_{m,n}$ satisfy $\lambda_{m,n} = \lambda_{n-m,n}$.

with uniform $\mathcal{O}_P$ -terms on some regularity ranges. The consistency for fixed parameters ϕ and α follows by observing that ν_0 is the turning point where the quadratic form starts dominating the log-determinant. The latter claims are also true if ν is estimated jointly with $\phi \in F$ for a set F bounded away from zero and infinity.

3.3 Profiling the likelihood

Consider now the case $F = (0, +\infty)$ by plugging (9) into (7), for $\nu > \nu_0 - 1/2$, to get

$$\inf_{\phi > 0} \mathbb{L}_n(\nu, \phi, \alpha) = 1 + \mathcal{O}(1) + \ln(\phi_0) + \mathcal{O}_P(1) - 2\nu_0 \ln(n) = \mathcal{O}_P(1) - 2\nu_0 \ln(n), \quad (10)$$

which is not sharp enough. Therefore, a more precise analysis of how the spectrum of K_θ fluctuates around the one of k_θ is needed to study the profile likelihood. The following section provides an ingredient for this purpose. Coordination with tools for proving uniform central limit theorems makes it possible to study convergence rates for parameter estimation and prediction error in Section 4. Developments for studying the profile likelihood are used to provide insights on model selection in the case of a fixed deterministic function from a Sobolev space in Section 5, which also discusses related works in this setting.

3.4 A symmetrized version of the Hurwitz zeta function

Stein (1999, Section 6.7) uses the function

$$\gamma: (s; x) \in (1, +\infty) \times (0, 1) \mapsto \sum_{j \in \mathbb{Z}} \frac{1}{|j + x|^s},$$

for deriving the asymptotics of the Fisher information matrix of the model presented in Section 2.1. It will also play a major role in our analysis of the likelihood criterion.

The function γ is (jointly) smooth and related to the Hurwitz zeta function ζ_H by:

$$\gamma(s; x) = \zeta_H(s; x) + \zeta_H(s; 1 - x), \quad (s, x) \in (1, +\infty) \times (0, 1). \quad (11)$$

Moreover, the function $\gamma(s; \cdot)$ is symmetric with respect to $1/2$ for $s > 1$.

4 Main results

4.1 Standing assumptions

Consider the framework presented in Section 2.1 and suppose that the function is sampled according to the (real-valued) centered Gaussian process:

$$\xi: x \in [0, 1] \mapsto \frac{1}{\sqrt{2}} \sum_{j \in \mathbb{Z}} \sqrt{\mathcal{L}_j(\theta_0)} (U_{1,|j|} + iU_{2,|j|}\text{sign}(j)) e^{2\pi i x j}, \quad (12)$$

with $\theta_0 = (\nu_0, \phi_0, \alpha_0) \in (0, +\infty)^3$ and $(U_{q,j})_{q \in \{1,2\}, j \geq 0}$ independent random variables such that $U_{2,0} = 0$, $U_{1,0} \sim \mathcal{N}(0, 2)$, and $U_{q,j} \sim \mathcal{N}(0, 1)$ for $q \in \{1, 2\}$

and $j \geq 1$. Let $\mathbf{P}$ be the measure defined on the underlying probability space (assumed to be the completion of the product space, so the $U_{q,j}$ s are coordinate projections). The convergence of the expansion (12) is meant pointwise both in $L^2(\mathbf{P})$ and almost surely. We further assume $\nu_0 > 1/2$ to avoid dealing with conditionally convergent series. It holds that $\xi \sim \text{GP}(0, k_{\theta_0})$.

Let $\hat{\theta}_n = (\hat{\nu}_n, \hat{\phi}_n, \hat{\alpha}_n)$ be a maximum likelihood estimate as defined in Section 2.2 for some $\Theta = N \times (0, +\infty) \times A$ with $A, N \subset (0, +\infty)$ compact intervals and $\nu_0 \in N$. The following sections give convergence rates for parameter estimation and prediction error.

4.2 Convergence rates of maximum likelihood estimation

The following result states the strong consistency of $\hat{\nu}_n$.

Theorem 4.1. *Let $\Theta = N \times (0, +\infty) \times A$ with N and A compact intervals and $\nu_0 \in N$. Then, the convergence $\hat{\nu}_n \rightarrow \nu_0$ holds almost surely.*

The proof is deferred to Section A.5 of the Appendix. A key step is to show that (a shift of) the profile likelihood converges almost surely to

$$\int_0^1 \ln(\gamma(2\nu + 1; \cdot)) + \ln\left(\int_0^1 \frac{\gamma(2\nu_0 + 1; \cdot)}{\gamma(2\nu + 1; \cdot)}\right), \quad (13)$$

for $\nu > \nu_0 - 1/2$. The first term is a refinement of the $\mathcal{O}(1)$ appearing in (9) for the log-determinant. The second term is a refinement of the $\mathcal{O}_{\mathbf{P}}(1)$ appearing for the quadratic form. Jensen inequality shows that (13) is minimized by taking $\nu = \nu_0$.

Furthermore, similarly to Stein (1999, Section 6.7), let us define

$$\psi_\nu: x \in (0, 1) \mapsto \frac{\sum_{j \in \mathbb{Z}} |x + j|^{-2\nu-1} \ln |x + j|}{\sum_{j \in \mathbb{Z}} |x + j|^{-2\nu-1}}, \quad \text{for } \nu > 0, \quad (14)$$

which is square integrable on $(0, 1)$. The following result proves the conjecture made by Stein (1999, p. 194) when $d = 1$ and $\hat{\nu}_n$ and $\hat{\alpha}_n$ are bounded. The proof is deferred to Section A.6 of the Appendix.

Theorem 4.2. *Let $\Theta = N \times (0, +\infty) \times A$ with N and A compact intervals and $\nu_0 \in N$. Then,*

$$\sqrt{2n} \left(\frac{\hat{\phi}_n - \phi_0}{2\phi_0} - (\ln(n) + \mathbb{E}(\psi_{\nu_0}(V))) (\hat{\nu}_n - \nu_0) \right) \rightsquigarrow \mathcal{N}(0, I_2),$$

where V is a random variable distributed uniformly on $(0, 1)$.

Observe that the asymptotic behavior of $(\hat{\nu}_n, \hat{\phi}_n)$ is not influenced by whether the parameter α is fixed, estimated, or even known.

4.3 Convergence rates of the integrated squared error

This section states our results about the expectation of (5) with fixed and estimated parameters. The proofs are deferred to Section C. We begin with the case of fixed parameters.

For $\nu, \nu_0 > 0$, define

$$\vartheta_{\nu; \nu_0} : x \in (0, 1) \mapsto \frac{\gamma(4\nu + 2; x) \gamma(2\nu_0 + 1; x)}{\gamma^2(2\nu + 1; x)} + \gamma(2\nu_0 + 1; x) - 2 \frac{\gamma(2\nu + 2\nu_0 + 2; x)}{\gamma(2\nu + 1; x)}$$

which is smooth and integrable when $\nu > (\nu_0 - 1)/2$.

The following result states the asymptotics of the prediction error with fixed parameters.

Theorem 4.3. *Let $(\nu, \alpha) \in (0, +\infty)^2$. Then,*

$$\mathbb{E}(\text{ISE}_n(\nu, \alpha; \xi)) \lesssim \frac{1}{n^{4\nu+2}}, \text{ for } \nu < (\nu_0 - 1)/2,$$

$$\mathbb{E}(\text{ISE}_n(\nu, \alpha; \xi)) \lesssim \frac{\ln(n)}{n^{2\nu_0}}, \text{ for } \nu = (\nu_0 - 1)/2,$$

and

$$n^{2\nu_0} \mathbb{E}(\text{ISE}_n(\nu, \alpha; \xi)) \rightarrow \phi_0 \int_0^1 \vartheta_{\nu; \nu_0}, \text{ otherwise.}$$

The symbol $\lesssim$ denotes an inequality up to a universal constant.

This result shows that half of the smoothness is sufficient for optimal convergence rates. However, the constant $\int_0^1 \vartheta_{\nu; \nu_0}$ is minimized by taking $\nu = \nu_0$. This is in line with the result of [Stein \(1999, Theorem 3\)](#) obtained in a different framework.

Then, our last result gives the asymptotic behavior of the prediction error with estimated parameters.

Theorem 4.4. *Let $\Theta = N \times (0, +\infty) \times A$ with N and A compact intervals and $\nu_0 \in N$. Then,*

$$n^{2\nu_0} \mathbb{E}(\text{ISE}_n(\hat{\nu}_n, \hat{\alpha}_n; \xi)) \rightarrow \phi_0 \int_0^1 \vartheta_{\nu_0; \nu_0}.$$

This last result shows that estimating the parameters yields asymptotically the same error as if the ground truth was known.

5 The deterministic case

Let $\beta > 1/2$ and define the Sobolev space

$$H^\beta[0, 1] = \left\{ g \in L^2[0, 1], \|g\|_{H^\beta[0, 1]}^2 = \sum_{j \in \mathbb{Z}} (1 + j^2)^\beta |c_j(f)|^2 < +\infty \right\}$$

of (continuous) periodic functions. This section studies maximum likelihood estimation with equispaced observations [\(3\)](#) from a fixed deterministic periodic function $f : [0, 1] \rightarrow \mathbb{R}$ lying in a Sobolev space. Define the (Sobolev) smoothness

$$\nu_0(f) = \inf \{ \beta > 1/2, f \notin H^\beta[0, 1] \}$$

of f as [Karvonen \(2023\)](#) and [Wang and Jing \(2022\)](#). We will assume that $\nu_0(f) \in (1, +\infty)$. The restriction $\nu_0(f) > 1$ is imposed for convenience as it

ensures that f has absolutely summable Fourier coefficients. Section D contains the proofs for this section.

On “nice” bounded regions of $\mathbb{R}^d$, Karvonen (2023) shows that $\liminf \hat{\nu}_n \geq \nu_0(f)$ if α and ϕ are fixed. Karvonen also shows that $\hat{\nu}_n \rightarrow \nu_0(f)$ for a class of compactly supported self-similar functions. It is not hard to check that $\nu_0(\xi) = \nu$ holds almost surely for any Matérn process with regularity parameter ν . With that in mind, one can interpret the previous results the following way. Maximum likelihood estimation chooses the parameter ν so that the sample paths are asymptotically smoother than f and, under more assumptions, so that the (Sobolev) smoothnesses match. Interestingly, the proof is close in spirit to the reasoning described in Section 3.2. A sketch is briefly provided with the notations of the framework from Section 2.1. The log-det term in (6) is not data-dependent, so the expansion from (9) applies.² Then, for $\nu > \nu_0(f) - 1/2$, the uniform inequality

$$Z^\top K_\theta^{-1} Z \lesssim \phi^{-1} n^{1+2(\nu-\nu_0(f))}$$

is (essentially³) shown. However, establishing (sufficient results slightly weaker than) the reverse inequality requires additional assumptions on f , such as membership in a class of functions with self-similar spectra (see Karvonen 2023, Definition 3.1 and also Szabó et al. 2015, p. 1398, in the context of the inverse signal-in-white-noise model). For the present purposes, it suffices to consider the “prototypical” subclass (see Karvonen, 2023, p. 14) of functions f such that

$$C_1 |j|^{-\nu_0(f)-1/2} \leq |c_j(f)| \leq C_2 |j|^{-\nu_0(f)-1/2} \quad \text{when } |j| \geq N, \quad (15)$$

for some $N \geq 0$ and $C_2 \geq C_1 > 0$. This notation is compatible with the definition of $\nu_0(f)$. As in previous works, the following property holds for this class of functions with well-behaved spectra.

Proposition 5.1. Let $\Theta = N \times F \times A$ with N , F , and A compact intervals and N containing $\nu_0(f)$. Assume that f satisfies (15). Then, the convergence $\hat{\nu}_n \rightarrow \nu_0(f)$ holds.

Having ϕ and α estimated on compact intervals jointly with ν is somewhat anecdotal, so nothing is new in this result. The details of the proof sketched in the previous paragraph are therefore omitted. However, since the proof roughly follows the lines from Section 3.2, the observation from Section 3.3 applies also in this setting. Beforehand, the following preliminary step is required.

Proposition 5.2. Let $\Theta = N \times (0, +\infty) \times A$ with N and A compact intervals and $\max N \geq \nu_0(f) - 1/2$. Then, it holds that $\liminf \hat{\nu}_n \geq \nu_0(f) - 1/2$.

Note the difference with the previous $\liminf \hat{\nu}_n \geq \nu_0(f)$ for fixed ϕ . A smoothness estimate larger than $\nu_0(f)$ means that f is rougher than the sample paths. The weaker inequality $\hat{\nu}_n \geq \nu_0(f) - 1/2$ only means that the function f is rougher than the elements of the reproducing kernel Hilbert space. A computation similar to (10) shows that the behavior above $\nu_0(f) - 1/2$ is, roughly

²The proof of Karvonen (2023) uses bounds with matching rates for the conditional variance instead of an inequality like (8).

³Provided that $f \in H^{\nu_0(f)}[0, 1]$, which does not necessarily hold with our notations. However, the bound $n^{1+2(\nu-\nu_0(f)+\epsilon)}$ given by $f \in H^{\nu_0(f)-\epsilon}[0, 1]$, for any $\epsilon > 0$, suffices.

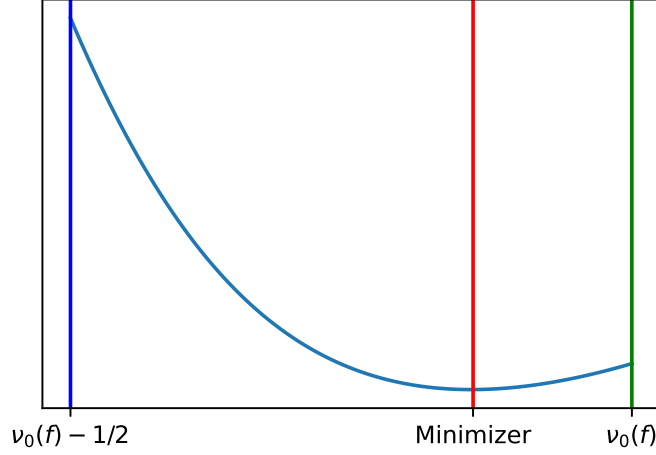


Figure 1: The function $\mathbb{M}_\infty^f$, for $\nu_0(f) = 3/2$. A numerical approximation to the minimizer is about 1.359. Note that $\mathbb{M}_\infty^f$ is approximated numerically using finite sums for γ and discretizations for the integrals.

speaking, governed by $\mathcal{O}(1)$ -terms. It is possible to give a quantitative description of what happens for a class smaller than (15). Define

$$\mathbb{M}_n^f: (\nu, \alpha) \in (\nu_0(f) - 1/2, +\infty) \times (0, +\infty) \mapsto \inf_{\phi > 0} \mathbb{L}_n(\nu, \phi, \alpha) + 2\nu_0(f) \ln(n) - 1.$$

Proposition 5.3. Suppose that $c_j(f) = \left(1 + \mathcal{O}(|j|^{-1})\right) |j|^{-\nu_0(f)-1/2}$ for nonzero j . Then, we have

$$\mathbb{M}_n^f(\nu, \alpha) \rightarrow \mathbb{M}_\infty^f(\nu) = \int_0^1 \ln(\gamma(2\nu + 1; \cdot)) + \ln\left(\int_0^1 \frac{\gamma^2(\nu_0(f) + 1/2; \cdot)}{\gamma(2\nu + 1; \cdot)}\right), \quad (16)$$

uniformly on compact subsets of $(\nu_0(f) - 1/2, +\infty) \times (0, +\infty)$.

We could not identify the minimizer(s) of the limit analytically. Figure 1 shows a numerical approximation to $\mathbb{M}_\infty^f$.

After inspection of the proof of Proposition 5.3, it does not seem obvious to exhibit a function f such that $\hat{\nu}_n \rightarrow \nu_0(f)$ holds when the amplitude parameter is jointly estimated. However, Theorem 4.1 shows that the set of such functions has probability one under a Matérn process with regularity $\nu_0 > 1/2$ belonging to N .

Acknowledgments. The author thanks Julien Bect and Emmanuel Vazquez for their patience and guidance.

A Proofs

A.1 Notations

The symbol $\lesssim$ denotes an inequality up to a universal constant. For compactness, the symbol $\approx$ is used when the two-way inequality $\lesssim$ holds.

Write $K_\theta = \phi R_{\nu,\alpha}$ and $\underline{c}_j(\theta) = \phi c_j(\nu, \alpha)$, for $\theta = (\nu, \phi, \alpha) \in (0, +\infty)^3$ and $j \in \mathbb{Z}$. All results suppose that $\Theta = N \times (0, +\infty) \times A$ with $N = [\nu_{\min}, \nu_{\max}]$, $A = [\alpha_{\min}, \alpha_{\max}]$, $0 < \nu_{\min} < \nu_0 < \nu_{\max} < +\infty$, and $0 < \alpha_{\min} \leq \alpha_{\max} < +\infty$ unless explicitly stated otherwise. Without loss of generality, suppose that $\nu_{\min} < \nu_0 - 1/2$ and define $N_\epsilon = [\nu_0 - 1/2 + \epsilon, +\infty) \cap N$ for $\epsilon > 0$. The notation $l = \lfloor (n-1)/2 \rfloor$ will often be used throughout the following.

A.2 Circulant matrices and useful facts

The framework introduced in Section 2.1 is convenient for analyzing kernel-based regression methods (see, e.g., [Craven and Wahba, 1979](#)). This section reviews the properties needed for our purposes.

Let W be the $n \times n$ matrix with entries $W_{j,m} = n^{-1/2} e^{2\pi i j m / n}$, for $0 \leq j, m \leq n-1$. For every $\theta = (\nu, \phi, \alpha) \in (0, +\infty)^3$, the periodicity of k_θ implies that

$$K_\theta = \begin{pmatrix} k_\theta(0) & k_\theta(\frac{1}{n}) & \dots & k_\theta(\frac{n-1}{n}) \\ k_\theta(\frac{n-1}{n}) & k_\theta(0) & \dots & k_\theta(\frac{n-2}{n}) \\ \dots & \dots & \dots & \dots \\ k_\theta(\frac{1}{n}) & k_\theta(\frac{2}{n}) & \dots & k_\theta(0) \end{pmatrix}$$

is a circulant matrix and so is $R_{\nu,\alpha}$. Consequently (see, e.g., [Brockwell and Davis, 1987](#), p. 130), it holds that $R_{\nu,\alpha} = W \Delta_{\nu,\alpha} W^*$ with $\Delta_{\nu,\alpha} = \text{diag}(\lambda_{0,n}, \dots, \lambda_{n-1,n})$ and

$$\lambda_{m,n} = \sum_{j=0}^{n-1} e^{-2\pi i j m / n} k_{\nu,1,\alpha}(j/n) = n \sum_{j \in \mathbb{Z}} c_{m+nj}(\nu, \alpha), \quad 0 \leq m \leq n-1. \quad (17)$$

Note that $\lambda_{m,n}$ depends on ν and α but the symbols are dropped to avoid cumbersome expressions. These coefficients verify

$$\lambda_{m,n} = \lambda_{n-m,n}, \quad \text{for } 1 \leq m \leq n-1. \quad (18)$$

The eigenvalue $\lambda_{0,n}$ is simple and there are l pairs $(\lambda_{m,n}, \lambda_{n-m,n})$, for $m \in \llbracket 1, l \rrbracket$, where l is the shortcut defined in Section A.1. If n is even, then the eigenvalue $\lambda_{n/2,n}$ is also simple.

Furthermore, combining each pair of eigenvectors of W shows that $R_{\nu,\alpha} = P \Delta_{\nu,\alpha} P^\top$ for a unitary matrix P written using sines and cosines functions. Then, with $\theta_0 = (\nu_0, \phi_0, \alpha_0)$ the ground truth introduced in Section 4.1, write

$$P^\top Z = \sqrt{\phi_0} \left(\sqrt{\lambda_{0,n}^{(0)}} U_{0,n}, \dots, \sqrt{\lambda_{n-1,n}^{(0)}} U_{n-1,n} \right),$$

with $\lambda_{0,n}^{(0)}, \dots, \lambda_{n-1,n}^{(0)}$ the eigenvalues of R_{ν_0,α_0} and $U_{0,n}, \dots, U_{n-1,n}$ drawn independently from a standard Gaussian. We have

$$Z^\top R_{\nu,\alpha}^{-1} Z = \phi_0 \sum_{m=0}^{n-1} \frac{U_{m,n}^2 \lambda_{m,n}^{(0)}}{\lambda_{m,n}}.$$

Our strategy to analyze this kind of expression will often consist of: 1) studying the sum for $m \in \llbracket 1, l \rrbracket$; 2) using the equality (18); and 3) treating the remaining terms for $m = 0$ and eventually $m = n/2$ separately.

The following approximation discussed in Section 3.2 will sometimes be used.

Lemma A.1. One has $n^{-1}\lambda_{0,n} \approx c_0(\nu, \alpha) \approx 1$ and $n^{-1}\lambda_{m,n} \approx c_m(\nu, \alpha) \approx m^{-2\nu-1}$ uniformly in $\nu \in N$, $\alpha \in A$, n and $1 \leq m \leq \lfloor n/2 \rfloor$.

Proof. Let $0 \leq m \leq \lfloor n/2 \rfloor$, we have using (17)

$$c_m(\nu, \alpha) \leq \lambda_{m,n}/n \leq 2c_m(\nu, \alpha) + 2 \sum_{j=1}^{+\infty} c_{m+nj}(\nu, \alpha).$$

Moreover

$$\sum_{j=1}^{+\infty} c_{m+nj}(\nu, \alpha)/c_m(\nu, \alpha) \leq \sum_{j=1}^{+\infty} (\alpha_{\max}^2 + 1/4)^{\nu+1/2}/j^{2\nu+1} \lesssim 1,$$

uniformly using the monotonicity of the zeta function. This shows $n^{-1}\lambda_{m,n} \approx c_m(\nu, \alpha)$ and finishing the proof makes no difficulty. $\square$

Nevertheless, our results will require refined approximations, as explained in Section 3.3.

A.3 More notations and properties

For each n , it is straightforward to prove that the $\lambda_{m,n}$ s are smooth functions of $(\nu, \alpha) \in (0, +\infty)^2$ by bounding the derivatives of the c_j s uniformly on compacta (up to third-order derivatives suffice for our purposes). Using the formulas from Section A.2 then shows that $\mathbb{L}_n$ is also smooth for any realization.

Furthermore, define:

$$\mathbb{M}_n: (\nu, \alpha) \in N \times A \mapsto \inf_{\phi > 0} \mathbb{L}_n(\nu, \phi, \alpha) + 2\nu_0 \ln(n) - \ln(\phi_0) - 1,$$

with ν_0 the ground truth introduced in Section 4.1. Its expression is given by Proposition 2.2 so it is a stochastic process which is smooth on the almost sure event $Z \neq 0$. The proofs mostly consist in studying $\mathbb{M}_n$.

For a compact interval $A \subset (0, +\infty)$, define now $\mathbb{U}_n: \nu \in N \mapsto \inf_{\alpha \in A} \mathbb{M}_n(\nu, \alpha)$. The object $\mathbb{U}_n$ is a stochastic process since the infima can be replaced by countable ones. Its almost sure continuity follows from the almost sure smoothness of $\mathbb{M}_n$ and the compactity of A .

Also, write $g_\nu = \ln(\gamma(2\nu + 1; \cdot))$ for $\nu > 0$ and

$$h_{\nu; \nu_0} = \frac{\gamma(2\nu_0 + 1; \cdot)}{\gamma(2\nu + 1; \cdot)}$$

for $\nu > \nu_0 - 1/2$. These functions are smooth and integrable and we will write

$$H: \nu \in (\nu_0 - 1/2, +\infty) \mapsto \int_0^1 h_{\nu; \nu_0}, \quad G: \nu \in (0, +\infty) \mapsto \int_0^1 g_\nu,$$

and $\mathbb{U}: \nu \in (\nu_0 - 1/2, +\infty) \mapsto G(\nu) + \ln(H(\nu))$. The smoothness of these functions is ensured by dominated convergence arguments (three derivatives suffice for our purposes).

A.4 Proofs of Section 2.1

Proof of Proposition 2.1. For $x \in [0, 1]$, the kriging equations yield $\hat{f}_n(x) = k_{\theta, x}^\top K_\theta^{-1} Z$, with $k_{\theta, x} = (k_\theta(m/n - x))_{0 \leq m \leq n-1}$. The assumptions guarantee that f equals the limit of its Fourier series everywhere. Then, using the matrix W defined in Section A.2, it is straightforward to show that

$$W^* Z = \sqrt{n} \left(\sum_{j \in m+n\mathbb{Z}} c_j(f) \right)_{0 \leq m \leq n-1} \quad (19)$$

and

$$W^* k_{\theta, x} = \sqrt{n} \left(\sum_{j \in m+n\mathbb{Z}} c_j(\theta) e^{-2\pi i x j} \right)_{0 \leq m \leq n-1},$$

where the sums converge absolutely. Then, the uniform absolute-convergence of (4) follows from elementary manipulations. $\square$

A.5 Proof of Theorem 4.1

A.5.1 Proof of the theorem

Proof of Theorem 4.1. For $0 < \epsilon < 1/2$, the sequence $\mathbb{U}_n$ converges almost surely uniformly to $\mathbb{U}$ on N_ϵ by Lemma A.6. Also, the function $\mathbb{U}$ is continuous and strictly minimized by taking $\nu = \nu_0$ thanks to Jensen inequality.

The rest of the proof is dedicated to showing that $\liminf \hat{\nu}_n \geq \nu_0 - 1/2 + \epsilon$ for some $\epsilon > 0$. First for $\nu \in N$ and $\alpha \in A$, we have

$$\begin{aligned} \mathbb{M}_n(\nu, \alpha) &= G(\nu) + \mathcal{O}(\ln(n)/n) + \ln \left(\frac{\phi_0^{-1} Z^\top R_{\nu, \alpha}^{-1} Z}{n^{1+2(\nu-\nu_0)}} \right) \\ &= \mathcal{O}(1) + \ln \left(\frac{\phi_0^{-1} Z^\top R_{\nu, \alpha}^{-1} Z}{n^{1+2(\nu-\nu_0)}} \right) \end{aligned}$$

uniformly in $\nu \in N$ and $\alpha \in A$ thanks to Lemma A.5 and the continuity of G .

Now, let $0 < \epsilon < 1/4$, $\nu \in N \setminus N_\epsilon = [\nu_{\min}, \nu_0 - 1/2 + \epsilon]$ and $\alpha \in A$. It holds that:

$$\begin{aligned} \frac{\phi_0^{-1} Z^\top R_{\nu, \alpha}^{-1} Z}{n^{1+2(\nu-\nu_0)}} &\geq \frac{C}{n} \sum_{m=1}^{n-1} U_{m,n}^2 \min \left(\frac{m}{n}, 1 - \frac{m}{n} \right)^{2(\nu-\nu_0)} \quad (C > 0, \text{ by Lemma A.1 and (18)}) \\ &\geq \frac{C}{n} \sum_{m=1}^{n-1} U_{m,n}^2 \min \left(\frac{m}{n}, 1 - \frac{m}{n} \right)^{-1+2\epsilon} \quad (\nu \leq \nu_0 - 1/2 + \epsilon) \\ &= o(1) + \frac{C}{n} \sum_{m=1}^{n-1} \min \left(\frac{m}{n}, 1 - \frac{m}{n} \right)^{-1+2\epsilon} \quad (\text{a.s., using Lemma A.12}) \\ &\rightarrow \frac{C}{2^{2\epsilon}\epsilon}. \end{aligned}$$

Lemma A.6 gives $\mathbb{U}_n(\nu_0) \rightarrow \mathbb{U}(\nu_0)$ almost surely, so we have

$$\begin{aligned} \inf_{\nu \in N \setminus N_\epsilon} \mathbb{U}_n(\nu) - \mathbb{U}_n(\nu_0) &= \inf_{\nu \in N \setminus N_\epsilon, \alpha \in A} \mathbb{M}_n(\nu, \alpha) - \mathbb{U}_n(\nu_0) \\ &\geq \mathcal{O}(1) + \ln(C) - \ln(2^{2\epsilon}\epsilon) - \mathbb{U}(\nu_0) + o(1). \end{aligned}$$

Letting $\epsilon \rightarrow 0$ shows that the expression in display can be made almost surely ultimately strictly positive. $\square$

A.5.2 Approximating $\ln(\det(R_{\nu,\alpha}))$

Lemma A.2. Let $\nu \in N$, $\alpha \in A$, $1 \leq m \leq \lfloor n/2 \rfloor$, and $j \in \mathbb{Z}$. We have:

$$c_{m+nj}(\nu, \alpha) = \frac{1 + u_{n,m,j}(\nu, \alpha)}{|jn + m|^{2\nu+1}}, \quad (20)$$

with $-1 < v_m \leq u_{n,m,j}(\nu, \alpha) \leq 0$ and $v_m = \mathcal{O}(m^{-2})$.

Proof. Using (17), we have

$$c_{m+nj}(\nu, \alpha) = \frac{1}{(\alpha^2 + (jn + m)^2)^{\nu+1/2}} = \frac{1 + u_{n,m,j}(\nu, \alpha)}{|jn + m|^{2\nu+1}},$$

with $u_{n,m,j}(\nu, \alpha) = (1 + (\alpha/(jn + m))^2)^{-\nu-1/2} - 1$. Elementary operations show that

$$0 \geq u_{n,m,j}(\nu, \alpha) \geq \left(\left(\frac{\alpha_{\max}}{m} \right)^2 + 1 \right)^{-\nu_{\max}-1/2} - 1,$$

which gives the desired result thanks to the Taylor inequality. $\square$

Lemma A.3. Let $S \subset (1, +\infty)$ be a compact interval. It holds that

$$\gamma(s; x) = \frac{1}{x^s} + \frac{1}{(1-x)^s} + \mathcal{O}(1),$$

uniformly in $s \in S$ and $x \in (0, 1)$. In particular, we have $\gamma(s; x) \approx \min(x, 1-x)^{-s}$.

Proof. Let $s_{\min} = \min S$. Then, $0 \leq \gamma(s; x) - x^{-s} - (1-x)^{-s} \leq 2\zeta(s_{\min})$. $\square$

Lemma A.4. Let $S \subset (1, +\infty)$ be a compact interval. It holds that

$$\frac{\partial \gamma}{\partial s}(s; x) = -\frac{\ln(x)}{x^s} - \frac{\ln(1-x)}{(1-x)^s} + \mathcal{O}(1),$$

uniformly in $s \in S$ and $x \in (0, 1)$.

Proof. Similar to the proof of Lemma A.3. $\square$

Lemma A.5. Uniformly in $\nu \in N$ and $\alpha \in A$, we have

$$\ln(\det(R_{\nu,\alpha})) = -2\nu n \ln(n) + n \int_0^1 g_\nu + \mathcal{O}(\ln(n)).$$

Proof. Let $\nu \in N$ and $\alpha \in A$. Using (17) and Lemma A.2, we have

$$\lambda_{m,n}/n = \sum_{j \in \mathbb{Z}} c_{m+nj}(\nu, \alpha) = \sum_{j \in \mathbb{Z}} \frac{1 + u_{n,m,j}(\nu, \alpha)}{|jn + m|^{2\nu+1}}.$$

Therefore, using the notation $l = \lfloor (n-1)/2 \rfloor$, we have

$$\sum_{m=1}^l \ln(\lambda_{m,n}/n) = -(2\nu+1)l \ln(n) + a_n(\nu, \alpha) + \sum_{m=1}^l g_\nu(m/n),$$

with

$$|a_n(\nu, \alpha)| \leq \left| \sum_{m=1}^l \ln(1 + v_m) \right| = \mathcal{O}(1)$$

uniformly in $\nu \in N$ and $\alpha \in A$.

The function g_ν is symmetric with respect to $1/2$. Moreover, a direct consequence of Lemma A.3 is that

$$g_\nu(x) = -(2\nu + 1) \ln(x) + \mathcal{O}(1), \quad (21)$$

uniformly in $\nu \in N$ and $0 < x \leq 1/2$. For $\nu \in N$, the function g_ν is thus integrable on $(0, 1)$. Furthermore, verifying that it is non-increasing on $(0, 1/2]$ is straightforward using the derivative of $\gamma(2\nu + 1; \cdot)$, so we have:

$$\int_{1/n}^{(l+1)/n} g_\nu \leq \frac{1}{n} \sum_{m=1}^l g_\nu(m/n) \leq \int_0^{l/n} g_\nu.$$

Use then (21) to get $\int_0^{1/n} g_\nu = \mathcal{O}(\ln(n)/n)$, uniformly in $\nu \in N$. The remainders $\int_{l/n}^{1/2} g_\nu$ and $\int_{1/2}^{(l+1)/n} g_\nu$ are $\mathcal{O}(n^{-1})$ uniformly in $\nu \in N$ by a compacity argument using the continuity of γ .

Therefore, we have

$$\sum_{m=1}^l g_\nu(m/n) = n \int_0^{1/2} g_\nu + \mathcal{O}(\ln(n)),$$

uniformly in $\nu \in N$. Moreover, Lemma A.1 shows that $\ln(\lambda_{0,n}/n) = \mathcal{O}(1)$ and $\ln(\lambda_{n/2,n}/n) = \mathcal{O}(\ln(n))$ uniformly for n even. One can then conclude using (18). $\square$

A.5.3 Approximating $Z^\top R_{\nu,\alpha} Z$

Let us first give some definitions. For $\epsilon > 0$, Lemma A.3 can be used to show that there exists some $C > 0$ such that

$$h_{\nu;\nu_0}(x) \leq F_\epsilon(x) = C \min(x, 1-x)^{-1+2\epsilon}, \text{ for all } 0 < x < 1 \text{ and } \nu \in N_\epsilon. \quad (22)$$

The function F_ϵ will be called the envelope of the family $\mathcal{F}_\epsilon = \{h_{\nu;\nu_0}, \nu \in N_\epsilon\}$ of functions.

Lemma A.6. For $0 < \epsilon < 1/2$, the sequence $\mathbb{M}_n$ converges almost surely uniformly to $(\nu, \alpha) \mapsto \mathbb{U}(\nu)$ on $N_\epsilon \times A$.

Proof. For $\nu \in N_\epsilon$ and $\alpha \in A$, we have:

$$\begin{aligned} & \frac{\phi_0^{-1} Z^\top R_{\nu,\alpha}^{-1} Z}{n^{1+2(\nu-\nu_0)}} - \int_0^1 h_{\nu;\nu_0} \\ &= \frac{U_{0,n}^2 \lambda_{0,n}^{(0)}}{n^{1+2(\nu-\nu_0)} \lambda_{0,n}} + \frac{1}{n} \sum_{m=1}^{n-1} U_{m,n}^2 \left(\frac{\lambda_{m,n}^{(0)}}{n^{2(\nu-\nu_0)} \lambda_{m,n}} - h_{\nu;\nu_0}(m/n) \right) \\ &+ \frac{1}{n} \sum_{m=1}^{n-1} B_{m,n} h_{\nu;\nu_0}(m/n) + \left(\frac{1}{n} \sum_{m=1}^{n-1} h_{\nu;\nu_0}(m/n) - \int_0^1 h_{\nu;\nu_0} \right), \end{aligned} \quad (23)$$

with $B_{m,n} = U_{m,n}^2 - 1$. First, $\sup_{\nu \in N_\epsilon} \left| n^{-1} \sum_{m=1}^{n-1} B_{m,n} h_{\nu;\nu_0}(m/n) \right|$ converges almost surely to zero by Lemma A.11, Lemma A.13, and Arzelà-Ascoli theorem. Then, for all $\beta > 0$, a Borel-Cantelli argument shows that $U_{0,n}^2 \lesssim n^\beta$ almost surely, so the $m = 0$ -term converges almost surely uniformly to zero by Lemma A.1. Finally, Lemma A.8 and Lemma A.10 show that (23) converges almost surely uniformly. Conclude using Proposition 2.2, Lemma A.5, and the L^∞ -continuity at $H: (\nu, \alpha) \in N_\epsilon \times A \mapsto \int_0^1 h_{\nu;\nu_0}$ of the mapping ψ used in the proof of Lemma A.16. $\square$

Lemma A.7. The function $h_{\nu;\nu_0}$ is non-decreasing (resp. non-increasing) on $(0, 1/2]$ when $\nu \geq \nu_0$ (resp. $\nu \leq \nu_0$).

Proof. Suppose that $\nu \geq \nu_0$. Use (11) along with the fact that the Hurwitz Zeta function verifies

$$\frac{\partial \zeta_H}{\partial x}(s; x) = -s \zeta_H(s+1; x), \quad \text{for } x > 0, \text{ and } s > 1, \quad (24)$$

and has the representation

$$\zeta_H(s; x) = \frac{1}{\Gamma(s)} \int_0^{+\infty} \frac{t^{s-1} e^{-tx}}{1 - e^{-t}} dt, \quad \text{for } x > 0, \text{ and } s > 1,$$

where Γ is the classical Gamma function (see, e.g., Postnikov, 1988). So, for $x \in (0, 1)$, we have

$$\gamma(2\nu + 1; x) = \frac{1}{\Gamma(2\nu + 1)} \int_0^{+\infty} \frac{t^{2\nu} (e^{-tx} + e^{-t(1-x)})}{1 - e^{-t}} dt,$$

and

$$\frac{\partial \gamma}{\partial x}(2\nu + 1; x) = \frac{1}{\Gamma(2\nu + 1)} \int_0^{+\infty} \frac{t^{2\nu+1} (e^{-t(1-x)} - e^{-tx})}{1 - e^{-t}} dt.$$

Now let $x \in [1/2, 1)$, the derivative of $h_{\nu;\nu_0}$ at x has the sign of

$$\begin{aligned} & \gamma(2\nu + 1; x) \frac{\partial \gamma}{\partial x}(2\nu_0 + 1; x) - \gamma(2\nu_0 + 1; x) \frac{\partial \gamma}{\partial x}(2\nu + 1; x) \\ &= \frac{1}{\Gamma(2\nu + 1)\Gamma(2\nu_0 + 1)} \int_0^{+\infty} \int_0^{+\infty} \frac{t^{2\nu} s^{2\nu_0} (\eta(s, t; x) - \eta(t, s; x))}{\kappa(s, t)} dt ds \end{aligned}$$

with $\eta(s, t; x) = s(e^{-tx} + e^{-t(1-x)})(e^{-s(1-x)} - e^{-sx})$ and $\kappa(s, t) = (1 - e^{-t})(1 - e^{-s}) = \kappa(t, s)$ thanks to the Fubini-Lebesgue theorem. Then, one can split the integral to have:

$$\begin{aligned} & \frac{1}{\Gamma(2\nu + 1)\Gamma(2\nu_0 + 1)} \left(\int_0^{+\infty} \int_t^{+\infty} \frac{t^{2\nu} s^{2\nu_0} (\eta(s, t; x) - \eta(t, s; x))}{\kappa(s, t)} dt ds \right. \\ & \quad \left. + \int_0^{+\infty} \int_t^{+\infty} \frac{s^{2\nu} t^{2\nu_0} (\eta(t, s; x) - \eta(s, t; x))}{\kappa(t, s)} dt ds \right) \\ &= \frac{1}{\Gamma(2\nu + 1)\Gamma(2\nu_0 + 1)} \int_0^{+\infty} \int_t^{+\infty} \frac{(t^{2\nu} s^{2\nu_0} - s^{2\nu} t^{2\nu_0})(\eta(s, t; x) - \eta(t, s; x))}{\kappa(s, t)} dt ds \leq 0 \end{aligned}$$

since $t^{2\nu} s^{2\nu_0} \leq s^{2\nu} t^{2\nu_0}$ when $s \geq t$, $\kappa(s, t) \geq 0$ and $\eta(s, t; x) \geq \eta(t, s; x)$ when $s \geq t$ and $x \geq 1/2$.

So we proved that $h_{\nu; \nu_0}$ is non-increasing on $[1/2, 1)$ and the first claim is due to the symmetry with respect to $1/2$. Observe that $h_{\nu; \nu_0} = 1/h_{\nu_0; \nu}$ for the second claim. $\square$

Lemma A.8. Let $\epsilon > 0$, we have

$$\frac{1}{n} \sum_{m=1}^{n-1} h_{\nu; \nu_0}(m/n) = \int_0^1 h_{\nu; \nu_0} + \mathcal{O}\left(\frac{1}{n^{\min(1, 2\epsilon)}}\right),$$

uniformly in $\nu \in N_\epsilon$.

Proof. The proof is similar to the treatment of $n^{-1} \sum_{m=1}^{n-1} g_\nu(m/n)$ in the proof of Lemma A.5 using Lemma A.7 and (22) to get:

$$\int_0^{1/n} h_{\nu; \nu_0} \leq \int_0^{1/n} F_\epsilon = \mathcal{O}(n^{-2\epsilon}), \quad \text{uniformly in } \nu \in N_\epsilon. \quad (25)$$

$\square$

Lemma A.9. Let $1 \leq m \leq \lfloor n/2 \rfloor$, we have

$$\frac{\lambda_{m,n}^{(0)}}{n^{2(\nu-\nu_0)} \lambda_{m,n}} = (1 + \mathcal{O}(m^{-2})) h_{\nu; \nu_0}(m/n)$$

uniformly in $\nu \in N$ and $\alpha \in A$.

Proof. A direct consequence from Lemma A.2. $\square$

Lemma A.10. Let $0 < \epsilon < 1/2$ and $0 < \delta < 2\epsilon$. There exists a constant C such that

$$\limsup n^{2\epsilon-\delta} \sup_{(\nu, \alpha) \in N_\epsilon \times A} \frac{1}{n} \sum_{m=1}^{n-1} U_{m,n}^2 \left| \frac{\lambda_{m,n}^{(0)}}{n^{2(\nu-\nu_0)} \lambda_{m,n}} - h_{\nu; \nu_0}\left(\frac{m}{n}\right) \right| \leq C,$$

almost surely.

Proof. Let $\nu \in N_\epsilon$, $\alpha \in A$, and $p = 1/(1-2\epsilon+\delta)$. It holds that $n^{-1} \sum_{m=1}^{n-1} F_\epsilon^p(m/n) = \mathcal{O}(1)$. Then, Lemma A.9, the usual symmetry arguments, and Hölder inequality yield:

$$\begin{aligned} & \frac{1}{n} \sum_{m=1}^{n-1} U_{m,n}^2 \left| \frac{\lambda_{m,n}^{(0)}}{n^{2(\nu-\nu_0)} \lambda_{m,n}} - h_{\nu; \nu_0}\left(\frac{m}{n}\right) \right| \\ &= \frac{1}{n} \sum_{m=1}^{n-1} U_{m,n}^2 \mathcal{O}(m^{-2} \vee (n-m)^{-2}) h_{\nu; \nu_0}\left(\frac{m}{n}\right) \\ &\leq \frac{1}{n^{1/q}} \cdot \left(\frac{1}{n} \sum_{m=1}^{n-1} |U_{m,n}|^{2p} F_\epsilon^p(m/n) \right)^{1/p} \cdot \underbrace{\left(\sum_{m=1}^{n-1} \mathcal{O}(m^{-2q} \vee (n-m)^{-2q}) \right)^{1/q}}_{\mathcal{O}(1) \text{ uniformly}} \end{aligned}$$

with $1/q = 2\epsilon - \delta$. Conclude using Lemma A.12 and $n^{-1} \sum_{m=1}^{n-1} F_\epsilon^p(m/n) = \mathcal{O}(1)$. $\square$

For $n \geq 2$ and $1 \leq m \leq n-1$, define $B_{m,n} = U_{m,n}^2 - 1$.

Lemma A.11. Let $\nu > \nu_0 - 1/2$. Then, $n^{-1} \sum_{m=1}^{n-1} B_{m,n} h_{\nu;\nu_0}(m/n)$ converges almost surely to zero.

Proof. A direct consequence of Lemma A.12, since $0 \leq h_{\nu;\nu_0}(x) \lesssim \min(x, 1-x)^{2(\nu-\nu_0)}$. $\square$

Lemma A.12. Let $\alpha > -1$ and $g: (0, 1) \rightarrow \mathbb{R}$ such that $0 \leq g(x) \lesssim \min(x, 1-x)^\alpha$. For each n , let $D_{1,n}, \dots, D_{n-1,n}$ be i.i.d. centered variables such that $\mathbb{E}(|D_{1,2}|^q)$ is finite for all $q \geq 0$. Then, $n^{-1} \sum_{m=1}^{n-1} D_{m,n} g(m/n)$ converges almost surely to zero.

Proof. If $\alpha \geq 0$, then $g(m/n) = \mathcal{O}(1)$, so the result is given by (Taylor and Hu, 1987, Corollary 5). Otherwise if $\alpha < 0$, then let $0 < \delta < 1/2$. It holds that:

$$\begin{aligned} \left| \frac{1}{n} \sum_{m=1}^{n-1} D_{m,n} g\left(\frac{m}{n}\right) \right| &\leq \left| \frac{1}{n} \sum_{m=1}^{n-1} \underbrace{g\left(\frac{m}{n}\right) \mathbb{1}_{\lfloor \delta n \rfloor + 1 \leq m \leq n - \lfloor \delta n \rfloor - 1}}_{\lesssim \delta^\alpha} D_{m,n} \right| \\ &+ \left| \frac{1}{n} \sum_{m=1}^{n-1} (\mathbb{1}_{m \leq \lfloor \delta n \rfloor} + \mathbb{1}_{m \geq n - \lfloor \delta n \rfloor}) D_{m,n} g\left(\frac{m}{n}\right) \right|. \end{aligned}$$

The first term converges almost surely to zero by (Taylor and Hu, 1987, Corollary 5). For the second term, Hölder inequality gives (a multiple of) the bound:

$$\left(\frac{1}{n} \sum_{m=1}^{n-1} |D_{m,n}|^q \right)^{1/q} \cdot \left(\frac{2}{n} \sum_{m=1}^{\lfloor \delta n \rfloor} \left(\frac{m}{n}\right)^{p\alpha} \right)^{1/p}.$$

The first term converges almost surely to the q -norm of the $D_{m,n}$ by the previous reference and, for p close enough to one, the second is $\mathcal{O}(\delta^{\alpha+1/p})$ with $\alpha + 1/p > 0$. Take $\delta = 1/j$ and a countable intersection of almost sure events to conclude. $\square$

Lemma A.13. Let $0 < \epsilon < 1/2$ and define

$$g_n: \nu \in N_\epsilon \mapsto \frac{1}{n} \sum_{m=1}^{n-1} B_{m,n} h_{\nu;\nu_0}\left(\frac{m}{n}\right).$$

The sequence $(g_n)_{n \geq 2}$ is almost surely uniformly equicontinuous.

Proof. Lemma A.4 shows that

$$\left| \frac{\partial \gamma}{\partial s}(2\nu + 1; x) \right| \lesssim -x^{-2\nu-1} \ln(x) \lesssim x^{-2(\nu+\delta)-1} \quad (\text{with the notation } \gamma(s; x)),$$

holds uniformly in $x \in (0, 1/2]$ and $\nu \in N_\epsilon$, for any $\delta > 0$. With a slight abuse of notation, the latter fact and Lemma A.3 yield:

$$\left| \frac{\partial h_{\nu_0}}{\partial \nu}(\nu; m/n) \right| \lesssim \left(\frac{n}{m}\right)^{1-2\epsilon+2\delta}, \quad (26)$$

uniformly in n , $1 \leq m \leq \lfloor n/2 \rfloor$, and $\nu \in N_\epsilon$.

Now let $\nu_1, \nu_2 \in N_\epsilon$. If one chooses $p > 1$ and $\delta > 0$ such that $p(1-2\epsilon+2\delta) < 1$, then we have by Hölder's inequality with $1/q + 1/p = 1$

$$\begin{aligned} & |g_n(\nu_1) - g_n(\nu_2)| \\ & \leq \left(\frac{1}{n} \sum_{m=1}^{n-1} |B_{m,n}|^q \right)^{1/q} \cdot \underbrace{\left(\frac{1}{n} \sum_{m=1}^{n-1} \sup_{\nu \in N_\epsilon} \left| \frac{\partial h_{\nu_0}}{\partial \nu}(\nu; m/n) \right|^p \right)^{1/p}}_{\mathcal{O}(1) \text{ by (26)}} \cdot |\nu_1 - \nu_2|. \end{aligned}$$

Use (Taylor and Hu, 1987, Corollary 5) to conclude. $\square$

A.6 Proof of Theorem 4.2

A.6.1 An upper bound of the rate

Lemma A.14. Let $0 < \beta < 1/4$. It holds that $\hat{\nu}_n - \nu_0 = o_P(n^{-\beta})$ and $\hat{\phi}_n - \phi_0 = o_P(n^{-\beta})$.

Proof. Let $0 < \beta < 1/4$, Proposition 2.2 gives almost surely

$$\ln(\hat{\phi}_n) = \ln(\phi_0) + \ln\left(\frac{\phi_0^{-1} Z^\top R_{\hat{\nu}_n, \hat{\alpha}_n}^{-1} Z}{n^{1+2(\hat{\nu}_n - \nu_0)}}\right) + 2(\hat{\nu}_n - \nu_0) \ln(n).$$

So

$$\begin{aligned} & \frac{n^\beta}{\ln(n)} \left(\ln(\hat{\phi}_n) - \ln(\phi_0) \right) \\ & = \frac{n^\beta}{\ln(n)} \ln(H(\hat{\nu}_n)) + \frac{n^\beta}{\ln(n)} \left(\ln\left(\frac{\phi_0^{-1} Z^\top R_{\hat{\nu}_n, \hat{\alpha}_n}^{-1} Z}{n^{1+2(\hat{\nu}_n - \nu_0)}}\right) - \ln(H(\hat{\nu}_n)) \right) + 2n^\beta(\hat{\nu}_n - \nu_0). \end{aligned}$$

The latter converges to zero in probability thanks to the coordination of (28) with Slutsky's lemma in $L^\infty(N_\epsilon \times A)$ (van Der Vaart and Wellner, 1996, p. 32), Lemma A.15, and the univariate delta method since the mapping $\ln \circ H$ is smooth. This implies that $\ln(\hat{\phi}_n) - \ln(\phi_0) = o_P(n^{-\beta})$ for all $0 < \beta < 1/4$. Conclude using again the delta method. $\square$

Lemma A.15. Let $0 < \beta < 1/4$. The bound $\hat{\nu}_n - \nu_0 = o_P(n^{-\beta})$ holds in probability.

Proof. Let $1/4 < \epsilon < 1/2$ and $0 < \beta < 1/2$ and use the notations from Section A.3. Lemma A.16 implies $\sup_{\nu \in N_\epsilon} |\mathbb{U}_n(\nu) - \mathbb{U}(\nu)| = o_P(n^{-\beta})$. Moreover, the function $\mathbb{U}$ is C^3 -smooth and we have $\mathbb{U}'(\nu_0) = 0$ and, with the notation given by (14):

$$\mathbb{U}''(\nu_0) = 4 \left(\int_0^1 (\psi_{\nu_0})^2 - \left(\int_0^1 \psi_{\nu_0} \right)^2 \right) > 0,$$

thanks to Jensen inequality. Finally, Theorem 4.1 and a second-order Taylor expansion around ν_0 give the rate $n^{-\beta/2}$. $\square$

Lemma A.16. Let $1/4 < \epsilon < 1/2$. Then, the sequence

$$(\nu, \alpha) \in N_\epsilon \times A \mapsto \sqrt{n} \left(\mathbb{M}_n(\nu, \alpha) - \int_0^1 g_\nu - \ln \left(\int_0^1 h_{\nu; \nu_0} \right) \right)$$

of processes converges weakly in $L^\infty(N_\epsilon \times A)$ to

$$\text{GP} \left(0, (\nu_1, \alpha_1; \nu_2, \alpha_2) \mapsto \frac{2 \int_0^1 h_{\nu_1; \nu_0} h_{\nu_2; \nu_0}}{\int_0^1 h_{\nu_1; \nu_0} \int_0^1 h_{\nu_2; \nu_0}} \right) \quad (27)$$

which can be seen as a tight Borel probability measure. In particular, for all $\beta < 1/2$, we have

$$\sup_{\nu \in N_\epsilon, \alpha \in A} \left| \mathbb{M}_n(\nu, \alpha) - \int_0^1 g_\nu - \ln \left(\int_0^1 h_{\nu; \nu_0} \right) \right| = o_P(n^{-\beta}).$$

Proof. Use the notation $H: (\nu, \alpha) \in N_\epsilon \times A \mapsto \int_0^1 h_{\nu; \nu_0}$ for this proof.

Let $\mathbb{D}_\psi \subset L^\infty(N_\epsilon \times A)$ be the subset of positive functions bounded away from zero. One has $H \in \mathbb{D}_\psi$ and $(\nu, \alpha) \in N_\epsilon \times A \mapsto n^{-1-2(\nu-\nu_0)} \phi_0^{-1} Z^\top R_{\nu, \alpha}^{-1} Z$ lying also in $\mathbb{D}_\psi$ almost surely by continuity on the compact $N_\epsilon \times A$.

Furthermore, the mapping $\psi: g \in \mathbb{D}_\psi \subset L^\infty(N_\epsilon \times A) \mapsto \ln \circ g \in L^\infty(N_\epsilon \times A)$ is Fréchet-differentiable at H with $\psi'(H): g \in L^\infty(N_\epsilon \times A) \mapsto g/H \in L^\infty(N_\epsilon \times A)$. The weak limit given by Lemma A.17 is tight and hence separable, so we can use Theorem 3.9.4 from [van Der Vaart and Wellner \(1996\)](#) to show that

$$\sqrt{n} \left(\ln \left(\frac{\phi_0^{-1} Z^\top R_{\nu, \alpha}^{-1} Z}{n^{1+2(\nu-\nu_0)}} \right) - \ln \left(\int_0^1 h_{\nu; \nu_0} \right) \right) \quad (28)$$

converges weakly to (27) in $L^\infty(N_\epsilon \times A)$. The tightness of the limit follows from the continuity of $\psi'(H)$. Conclude with Proposition 2.2, Lemma A.5, and Slutsky's lemma. $\square$

Lemma A.17. Let $1/4 < \epsilon < 1/2$. The sequence

$$(\nu, \alpha) \in N_\epsilon \times A \mapsto \sqrt{n} \left(\frac{\phi_0^{-1} Z^\top R_{\nu, \alpha}^{-1} Z}{n^{1+2(\nu-\nu_0)}} - \int_0^1 h_{\nu; \nu_0} \right)$$

of processes converges weakly in $L^\infty(N_\epsilon \times A)$ to

$$\text{GP} \left(0, (\nu_1, \alpha_1; \nu_2, \alpha_2) \mapsto 2 \int_0^1 h_{\nu_1; \nu_0} h_{\nu_2; \nu_0} \right),$$

which can be seen as a tight Borel probability measure.

Proof. Using the continuous mapping theorem for the linear isometry $\rho: L^\infty(N_\epsilon) \rightarrow L^\infty(N_\epsilon \times A)$ mapping $g \in L^\infty(N_\epsilon)$ to the function $(\nu, \alpha) \in N_\epsilon \times A \mapsto g(\nu)$ makes it possible to rephrase the convergence given by Lemma A.20 in $L^\infty(N_\epsilon \times A)$. (The limit (29) is a tight and hence separable measure.) The rest of the proof is similar to the analysis of (23) in the proof of Lemma A.6, but using $\epsilon > 1/4$. $\square$

Lemma A.18. Let $1/4 < \epsilon < 1/2$. The family $\mathcal{F}_\epsilon$ of functions equipped with the envelope F_ϵ defined by (22) verifies the uniform entropy condition ([van Der Vaart and Wellner, 1996](#), Section 2.5.1).

Proof. For $x \in (0, 1)$ and $\nu \in N$, write $\gamma(2\nu + 1; x) = \gamma_\uparrow(2\nu + 1; x) + \gamma_\downarrow(2\nu + 1; x)$, with

$$\gamma_\downarrow(2\nu + 1; x) = \sum_{j=1}^{+\infty} (j+x)^{-2\nu-1} + \sum_{j=1}^{+\infty} (j+1-x)^{-2\nu-1},$$

and $\gamma_\uparrow(2\nu + 1; x) = x^{-2\nu-1} + (1-x)^{-2\nu-1}$. Let $h_\uparrow(\nu; x) = \gamma(2\nu_0 + 1, x) / \gamma_\downarrow(2\nu + 1, x)$ and $h_\downarrow(\nu; x) = \gamma(2\nu_0 + 1, x) / \gamma_\uparrow(2\nu + 1, x)$.⁴ The families $\mathcal{F}_\epsilon^\downarrow = \{h_\downarrow(\nu; \cdot), \nu \in N_\epsilon\}$ and $\mathcal{F}_\epsilon^\uparrow = \{1/h_\uparrow(\nu; \cdot), \nu \in N_\epsilon\}$ of functions are non-increasing with respect to the parameter ν so they are VC-subgraph classes. Indeed, let $(x_1, y_1), (x_2, y_2) \in (0, 1) \times \mathbb{R}$, there cannot be two functions f and g in one of these families such that $f(x_1) < y_1$, $f(x_2) \geq y_2$, $g(x_1) \geq y_1$, and $g(x_2) < y_2$, since we have either $g \leq f$ or $f \leq g$.

Equip $\mathcal{F}_\epsilon^\downarrow$ and $\mathcal{F}_\epsilon^\uparrow$ respectively with the envelopes F_ϵ (by increasing eventually the constant C in (22)) and $F_\epsilon^\uparrow: x \in (0, 1) \mapsto C_2 \min(x, 1-x)^{1+2\nu_0}$, for some constant $C_2 > 0$. Theorem 2.6.7 from (van Der Vaart and Wellner, 1996) shows that these families satisfy the uniform entropy condition.

Consider $\varsigma: x, y \in (0, +\infty) \mapsto (x^{-1} + y)^{-1}$. It holds that $|\frac{\partial \varsigma}{\partial x}(x, y)| \leq 1$ and $|\frac{\partial \varsigma}{\partial y}(x, y)| = \varsigma^2(x, y)$. Observe that $\varsigma(h_\downarrow(\nu_1; \cdot), 1/h_\uparrow(\nu_2; \cdot)) \lesssim F_\epsilon$, for $\nu_1, \nu_2 \in N_\epsilon$. Consequently, for $\nu_1, \nu_2, \nu_3, \nu_4 \in N_\epsilon$ and $x \in (0, 1)$, we have:

$$\begin{aligned} & (\varsigma(h_\downarrow(\nu_1; x), 1/h_\uparrow(\nu_3; x)) - \varsigma(h_\downarrow(\nu_2; x), 1/h_\uparrow(\nu_4; x)))^2 \\ & \lesssim (h_\downarrow(\nu_1; x) - h_\downarrow(\nu_2; x))^2 + F_\epsilon^4(x) \left(\frac{1}{h_\uparrow(\nu_3; x)} - \frac{1}{h_\uparrow(\nu_4; x)} \right)^2. \end{aligned}$$

Observe that $\varsigma(h_\downarrow(\nu; \cdot), 1/h_\uparrow(\nu; \cdot)) = h_{\nu; \nu_0}$ and use Theorem 2.10.20 from (van Der Vaart and Wellner, 1996) to conclude that the family

$$\mathcal{F}_\epsilon^{(\pi)} = \{\varsigma(h_\downarrow(\nu_1; \cdot), 1/h_\uparrow(\nu_2; \cdot)) - 1, \nu_1, \nu_2 \in N_\epsilon\} \quad (\text{note that } h_{\nu_0; \nu_0} = 1)$$

with envelope $F_\epsilon^{(\pi)} = 2\sqrt{F_\epsilon^2 + F_\epsilon^4(F_\epsilon^\uparrow)^2}$ satisfy the uniform entropy condition. Concluding the proof is straightforward since $\mathcal{F}_\epsilon \subset \mathcal{F}_\epsilon^{(\pi)} + 1$ and $F_\epsilon^{(\pi)} \lesssim F_\epsilon$. $\square$

Lemma A.19. For all $\epsilon > 1/4$, we have

$$\frac{1}{n} \sum_{m=1}^{n-1} (h_{\nu_1; \nu_0}(m/n) - h_{\nu_2; \nu_0}(m/n))^2 \rightarrow \int_0^1 (h_{\nu_1; \nu_0} - h_{\nu_2; \nu_0})^2,$$

uniformly in $\nu_1, \nu_2 \in N_\epsilon$.

Proof. Let $\delta > 0$, there exists $\alpha > 0$ such that:

$$\int_0^\alpha (h_{\nu_1; \nu_0} - h_{\nu_2; \nu_0})^2 \leq 4 \int_0^\alpha F_\epsilon^2 \leq \delta/5$$

and

$$\frac{1}{n} \sum_{m=1}^{\lfloor \alpha n \rfloor} (h_{\nu_1; \nu_0}(m/n) - h_{\nu_2; \nu_0}(m/n))^2 \leq \frac{4}{n} \sum_{m=1}^{\lfloor \alpha n \rfloor} F_\epsilon^2(m/n) \leq \delta/5,$$

⁴The symbols $\downarrow$ and $\uparrow$ account for the monotonicity with respect to ν for fixed x .

uniformly in $\nu_1, \nu_2 \in N_\epsilon$. The same bounds also hold by symmetry for similar quantities related to $[1 - \alpha, 1]$. Furthermore, a compacity argument using the smoothness of γ shows that the mapping $x \in (0, 1) \mapsto (h_{\nu_1; \nu_0}(x) - h_{\nu_2; \nu_0}(x))^2$ and its derivative are bounded on $[\alpha, 1 - \alpha]$ uniformly in $\nu_1, \nu_2 \in N_\epsilon$. Consequently, the standard technique for bounding approximation errors of Riemann sums gives

$$\left| \frac{1}{n} \sum_{m=\lfloor \alpha n \rfloor + 1}^{\lceil (1-\alpha)n \rceil - 1} (h_{\nu_1; \nu_0}(m/n) - h_{\nu_2; \nu_0}(m/n))^2 - \int_\alpha^{1-\alpha} (h_{\nu_1; \nu_0} - h_{\nu_2; \nu_0})^2 \right| \leq \delta/5,$$

uniformly in $\nu_1, \nu_2 \in N_\epsilon$, for sufficiently large n . $\square$

For $n \geq 2$ and $1 \leq m \leq n-1$, define $B_{m,n} = U_{m,n}^2 - 1$.

Lemma A.20. Let $1/4 < \epsilon < 1/2$. Then, the sequence

$$\nu \in N_\epsilon \mapsto \frac{1}{\sqrt{n}} \sum_{m=1}^{n-1} B_{m,n} h_{\nu; \nu_0} \left(\frac{m}{n} \right)$$

of processes converges weakly in $L^\infty(N_\epsilon)$ to

$$\text{GP} \left(0, (\nu_1, \nu_2) \mapsto 2 \int_0^1 h_{\nu_1; \nu_0} h_{\nu_2; \nu_0} \right), \quad (29)$$

which can be seen as a tight Borel probability measure.

Proof. Let $2 < \alpha < 1/(1 - 2\epsilon)$. It holds that $F_\epsilon \in L^\alpha(0, 1) \subset L^2(0, 1)$. Moreover, Lemma A.18 shows that $\mathcal{F}_\epsilon$ satisfies the uniform entropy condition (van Der Vaart and Wellner, 1996, Section 2.5.1).

Let us show that $(\mathcal{F}_\epsilon, \|\cdot\|_{L^2(0,1)})$ is totally bounded. Use the shortcut $Q_n = n^{-1}\delta_{1/2} + n^{-1} \sum_{m=1}^{n-1} \delta_{m/n}$. Since $\epsilon > 1/4$, then $\int F_\epsilon^2 dQ_n$ is bounded uniformly in n by, say, M^2 . The uniform entropy condition implies that $\mathcal{F}_\epsilon$ is totally bounded for the $L^2(Q_n)$ -norm for any n . Let $\mathcal{G}_n$ be an $(M\delta)$ -internal covering, for $\delta > 0$. Lemma A.19 makes it possible to choose n such that

$$\sup_{g_1, g_2 \in \mathcal{F}_\epsilon} \left| \int (g_1 - g_2)^2 dQ_n - \int_0^1 (g_1 - g_2)^2 \right| \leq \delta^2.$$

Therefore, $\mathcal{G}_n$ is a $(\delta\sqrt{M^2 + 1})$ -covering of $(\mathcal{F}_\epsilon, \|\cdot\|_{L^2(0,1)})$.

With $Y_{m,n}: g \in (\mathcal{F}_\epsilon, \|\cdot\|_{L^2(0,1)}) \mapsto n^{-1/2} B_{m,n} g(m/n)$, the usual measurability conditions (see van Der Vaart and Wellner, 1996, p. 205) are met since the suprema can be replaced by ones on countable sets. Indeed, using the surjection $\varrho: \nu \in N_\epsilon \mapsto h_{\nu; \nu_0} \in \mathcal{F}_\epsilon$, the suprema on subsets of $\mathcal{F}_\epsilon \times \mathcal{F}_\epsilon$ are suprema on subsets of $(N_\epsilon \times N_\epsilon, \|\cdot\|_2)$, with $\|\cdot\|_2$ standing for the euclidean norm. A subset of a separable metric space is separable. The sample path continuity of $\nu \in N_\epsilon \mapsto Y_{m,n}(\varrho(\nu))$ is inherited from the continuity of $\nu \in N_\epsilon \mapsto h_{\nu; \nu_0}(x)$, for $0 < x < 1$.

Since $2 < \alpha < 1/(1 - 2\epsilon)$, we have $n^{-1} \sum_{m=1}^{n-1} F_\epsilon^\alpha(m/n) = \mathcal{O}(1)$ so the Lyapunov condition on suprema holds:

$$\sum_{m=1}^{n-1} \mathbb{E} \left(\sup_{g \in \mathcal{F}_\epsilon} |Y_{m,n}(g)|^\alpha \right) \leq \frac{\mathbb{E}(|B_{1,2}|^\alpha)}{n^{\alpha/2}} \sum_{m=1}^{n-1} F_\epsilon^\alpha(m/n) = o(1).$$

Furthermore, for $\delta_n \rightarrow 0$, we have

$$\begin{aligned}
& \sup_{\|g_1 - g_2\|_{L^2(0,1)} < \delta_n} \sum_{m=1}^{n-1} \mathbb{E} \left((Y_{m,n}(g_1) - Y_{m,n}(g_2))^2 \right) \quad (\text{with } g_1, g_2 \in \mathcal{F}_\epsilon) \\
&= \mathbb{E} (B_{1,2}^2) \sup_{\|g_1 - g_2\|_{L^2(0,1)} < \delta_n} \frac{1}{n} \sum_{m=1}^{n-1} (g_1(m/n) - g_2(m/n))^2 \\
&= o(1) + \mathcal{O}(\delta_n^2) \rightarrow 0
\end{aligned}$$

thanks to Lemma A.19.

Now, let us show the pointwise convergence of the sequence of covariance functions. For a fixed $\nu \in N_\epsilon$, the convergence $n^{-1} \sum_{m=1}^{n-1} h_{\nu; \nu_0}^2(m/n) \rightarrow \int_0^1 h_{\nu; \nu_0}^2$ is ensured using Lemma A.7 and the same reasoning as in the proof of Lemma A.8. This fact and Lemma A.19 shows that

$$\text{Cov} \left(\sum_{m=1}^{n-1} Y_{m,n}(g_1), \sum_{m=1}^{n-1} Y_{m,n}(g_2) \right) \rightarrow 2 \int_0^1 g_1 g_2,$$

for fixed $g_1, g_2 \in \mathcal{F}_\epsilon$.

Finally, with $\mu_{n,m} = n^{-1} B_{m,n}^2 \delta_{m/n}$, one has $0 < \mu_{n,m} F_\epsilon^2 < +\infty$ almost surely and $\sum_{m=1}^{n-1} \mu_{n,m} F_\epsilon^2 = \mathcal{O}_P(1)$ using Markov's inequality.

We can then conclude using Lemma 2.11.6 and Theorem 2.11.1 from [van Der Vaart and Wellner \(1996\)](#), which also imply the tightness of the limit (see [van Der Vaart and Wellner, 1996](#), Lemma 1.3.8 and Theorem 1.5.7). The reformulation from $L^\infty(\mathcal{F}_\epsilon)$ to $L^\infty(N_\epsilon)$ is an application of the continuous mapping theorem. $\square$

A.6.2 A Taylor expansion

The proof of Theorem 4.2 is finished using a standard third-order Taylor expansion around $(\nu_0, \phi_0, \hat{\alpha}_n)$. The following technical lemmata are required. Their proofs mostly consist in reproducing the technique used by [Stein \(1999, Section 6.7\)](#) to derive the asymptotics of the Fisher information matrix. Some details are provided in Section B.

Lemma A.21. We have

$$\frac{\sqrt{n}}{2\sqrt{2}} A_n^\top \nabla \mathbb{L}_n(\nu_0, \phi_0, \hat{\alpha}_n) \rightsquigarrow \mathcal{N}(0, I_2),$$

with $\nabla \mathbb{L}_n$ the gradient with respect to (ν, ϕ) only and

$$A_n = \frac{2\phi_0}{\sqrt{\text{Var}(\psi_{\nu_0}(V))}} \begin{pmatrix} 2^{-1}\phi_0^{-1} & 0 \\ \ln(n) + \mathbb{E}(\psi_{\nu_0}(V)) & \sqrt{\text{Var}(\psi_{\nu_0}(V))} \end{pmatrix}, \quad (30)$$

where V is a random variable distributed uniformly on $(0, 1)$.

Lemma A.22. It holds in probability that:

$$A_n^\top \nabla^2 \mathbb{L}_n(\nu_0, \phi_0, \hat{\alpha}_n) A_n \rightarrow 4I_2,$$

with A_n given by (30) and ∇ operating only on (ν, ϕ) .

Proof of Theorem 4.2. Lemmata A.21 and A.22 give the asymptotics of the score and the Hessian matrix, respectively. We are now left to bound the third derivatives uniformly locally around (ν_0, ϕ_0) . Cumbersome expressions are provided in Section B. For $\epsilon > 0$ small enough, bounding the terms individually with Lemma A.1 and Lemma B.1 makes it straightforward to show that

$$\mathbb{E} \left(\sup_{p \in \{0, 1, 2, 3\}, |\nu - \nu_0| \leq \epsilon, |\phi - \phi_0| \leq \epsilon, \alpha \in A} \left| \frac{\partial^3 \mathbb{L}_n}{(\partial \nu)^p (\partial \phi)^{3-p}} (\nu, \phi, \alpha) \right| \right) = \mathcal{O}(n^{5\epsilon}). \quad (31)$$

Lemma A.14 shows that $(\hat{\nu}_n, \hat{\phi}_n) \in [\nu_0 - \epsilon, \nu_0 + \epsilon] \times [\phi_0 - \epsilon, \phi_0 + \epsilon]$ with high probability. Write ∇ for taking derivatives with respect to (ν, ϕ) only. On this event, we have:

$$0 = \nabla \mathbb{L}_n(\nu_0, \phi_0, \hat{\alpha}_n) + \nabla^2 \mathbb{L}_n(\nu_0, \phi_0, \hat{\alpha}_n) \begin{pmatrix} \hat{\nu}_n - \nu_0 \\ \hat{\phi}_n - \phi_0 \end{pmatrix} + \mathcal{O}_{\mathbb{P}} \left(n^{5\epsilon} \left\| \begin{pmatrix} \hat{\nu}_n - \nu_0 \\ \hat{\phi}_n - \phi_0 \end{pmatrix} \right\|^2 \right),$$

thanks to (31). Multiplying by A_n^T (see (30)) and using Lemma A.14 again leads to

$$0 = A_n^T \nabla \mathbb{L}_n(\nu_0, \phi_0, \hat{\alpha}_n) + (A_n^T \nabla^2 \mathbb{L}_n(\nu_0, \phi_0, \hat{\alpha}_n) A_n + o_{\mathbb{P}}(1)) A_n^{-1} \begin{pmatrix} \hat{\nu}_n - \nu_0 \\ \hat{\phi}_n - \phi_0 \end{pmatrix},$$

where the preceding $\mathcal{O}_{\mathbb{P}}$ -term has been reformulated using a few algebraic manipulations. (Use the fact that $\|A_n\| \lesssim \ln(n)$.) Multiply by $\sqrt{2n}$ and use Slutsky's lemma to conclude. $\square$

B Proofs of technical lemmas for Theorem 4.2

Remember (see Section A.2 and Section A.3) that the $\lambda_{m,n}$ s depend smoothly on ν and α . Thus, the function $\mathbb{L}_n$ is smooth for any realization and can be written as:

$$\mathbb{L}_n(\nu, \phi, \alpha) = \ln(\phi) + \frac{1}{n} \sum_{m=0}^{n-1} \ln(\lambda_{m,n}) + \frac{\phi_0}{n\phi} \sum_{m=0}^{n-1} \frac{\lambda_{m,n}^{(0)} U_{m,n}^2}{\lambda_{m,n}}.$$

Expressions for some derivatives are given in the following. These expressions are cumbersome, but rough approximations will suffice: we only need to ensure the $\partial^p \lambda_{m,n} / \partial \nu^p$ s do not grow too fast compared to $\lambda_{m,n}$. The first-order derivative with respect to ν writes:

$$\frac{\partial \mathbb{L}_n}{\partial \nu}(\nu, \phi, \alpha) = \frac{1}{n} \sum_{m=0}^{n-1} \frac{\partial \lambda_{m,n} / \partial \nu}{\lambda_{m,n}} - \frac{\phi_0}{n\phi} \sum_{m=0}^{n-1} \frac{U_{m,n}^2 \lambda_{m,n}^{(0)} \partial \lambda_{m,n} / \partial \nu}{\lambda_{m,n}^2}.$$

Then, the second-order derivative with respect to ν writes:

$$\begin{aligned} \frac{\partial^2 \mathbb{L}_n}{\partial \nu^2}(\nu, \phi, \alpha) &= \frac{1}{n} \sum_{m=0}^{n-1} \frac{\lambda_{m,n} \partial^2 \lambda_{m,n} / \partial \nu^2 - (\partial \lambda_{m,n} / \partial \nu)^2}{\lambda_{m,n}^2} \\ &\quad - \frac{\phi_0}{n\phi} \sum_{m=0}^{n-1} \frac{U_{m,n}^2 \lambda_{m,n}^{(0)} \left(\partial^2 \lambda_{m,n} / \partial \nu^2 \lambda_{m,n} - 2 (\partial \lambda_{m,n} / \partial \nu)^2 \right)}{\lambda_{m,n}^3}. \end{aligned}$$

Finally, the third-order derivative with respect to ν writes:

$$\begin{aligned} \frac{\partial^3 \mathbb{L}_n}{\partial \nu^3}(\nu, \phi, \alpha) &= \frac{1}{n} \sum_{m=0}^{n-1} \lambda_{m,n}^{-3} \left(\frac{\partial^3 \lambda_{m,n}}{\partial \nu^3} \lambda_{m,n}^2 - 3 \frac{\partial^2 \lambda_{m,n}}{\partial \nu^2} \frac{\partial \lambda_{m,n}}{\partial \nu} \lambda_{m,n} + 2 \left(\frac{\partial \lambda_{m,n}}{\partial \nu} \right)^3 \right) \\ &\quad - \frac{\phi_0}{n\phi} \sum_{m=0}^{n-1} \lambda_{m,n}^{-4} \lambda_{m,n}^{(0)} \left(\frac{\partial^3 \lambda_{m,n}}{\partial \nu^3} \lambda_{m,n}^2 - 4 \frac{\partial^2 \lambda_{m,n}}{\partial \nu^2} \frac{\partial \lambda_{m,n}}{\partial \nu} \lambda_{m,n} + 6 \left(\frac{\partial \lambda_{m,n}}{\partial \nu} \right)^3 \right) U_{m,n}^2. \end{aligned}$$

Bounding all terms independently will suffice for our purposes. The necessary approximations are given by Lemma A.1 and the following. Exceptionally, the arguments of the $\lambda_{m,n}$ s are not dropped.

Lemma B.1. Let $0 < \delta < 2\nu_{\min}$, $0 \leq m \leq \lfloor n/2 \rfloor$, $\nu \in N$, $\alpha \in A$ and $p \in \{1, 2, 3\}$. We have:

$$\frac{1}{n} \left| \frac{\partial^p \lambda_{m,n}}{\partial \nu^p}(\nu, \alpha) \right| \lesssim \frac{1}{m^{2\nu+1-\delta}}, \quad \text{if } 1 \leq m \leq \lfloor n/2 \rfloor$$

and

$$\frac{1}{n} \left| \frac{\partial^p \lambda_{0,n}}{\partial \nu^p}(\nu, \alpha) \right| \lesssim 1,$$

uniformly in m , ν , and α .

Proof. We have

$$\frac{1}{n} \left| \frac{\partial^p \lambda_{m,n}}{\partial \nu^p}(\nu, \alpha) \right| \leq \sum_{j \in \mathbb{Z}} \frac{|\ln^p(\alpha^2 + (m+jn)^2)|}{(\alpha^2 + (m+jn)^2)^{\nu+1/2}} \lesssim \sum_{j \in \mathbb{Z}} \frac{1}{(\alpha^2 + (m+jn)^2)^{\nu+1/2-\delta/2}},$$

which equals $n^{-1} \lambda_{m,n}(\nu - \delta/2, \alpha)$, so Lemma A.1 gives the result. (Adjust the lower bound of N if needed.) $\square$

Lemma B.2. Let $A \subset (0, +\infty)$ be a compact interval and $\nu_0 > 0$. It holds that

$$\frac{\partial \lambda_{m,n} / \partial \nu(\nu_0, \alpha)}{\lambda_{m,n}(\nu_0, \alpha)} = -2 \ln(n) - 2\psi_{\nu_0}(m/n) + \mathcal{O}(m^{-2} \ln(n)),$$

uniformly in $\alpha \in A$ and $1 \leq m \leq \lfloor n/2 \rfloor$, with ψ_{ν_0} given by (14).

Proof. We have:

$$\begin{aligned} &n^{-1} \frac{\partial \lambda_{m,n}}{\partial \nu}(\nu_0, \alpha) \\ &= - \sum_{j \in \mathbb{Z}} \frac{\ln(\alpha^2 + (m+jn)^2)}{(\alpha^2 + (m+jn)^2)^{\nu_0+1/2}} \\ &= - \sum_{j \in \mathbb{Z}} \frac{2 \ln|m+jn| + \ln\left(\left(\frac{\alpha}{m+jn}\right)^2 + 1\right)}{(\alpha^2 + (m+jn)^2)^{\nu_0+1/2}} \\ &= - \sum_{j \in \mathbb{Z}} \frac{2 \ln|m+jn| + \mathcal{O}(m^{-2})}{(\alpha^2 + (m+jn)^2)^{\nu_0+1/2}} \quad (\text{uniformly, since } m \leq n/2 \Rightarrow m \leq |m+jn|) \\ &= - (1 + \mathcal{O}(m^{-2})) \sum_{j \in \mathbb{Z}} \frac{2 \ln(n) + 2 \ln|m/n+j| + \mathcal{O}(m^{-2})}{|m+jn|^{2\nu_0+1}} \quad (\text{Lemma A.2}) \\ &= - (1 + \mathcal{O}(m^{-2})) n^{-2\nu_0-1} \gamma(2\nu_0+1; m/n) (2 \ln(n) + 2\psi_{\nu_0}(m/n) + \mathcal{O}(m^{-2})). \end{aligned}$$

Thus, using Lemma A.2 again yields

$$\begin{aligned}\frac{\frac{\partial \lambda_{m,n}}{\partial \nu}(\nu_0, \alpha)}{\lambda_{m,n}(\nu_0, \alpha)} &= \frac{-(1 + \mathcal{O}(m^{-2}))(2 \ln(n) + 2\psi_{\nu_0}(m/n) + \mathcal{O}(m^{-2}))}{1 + \mathcal{O}(m^{-2})} \\ &= -(1 + \mathcal{O}(m^{-2}))(2 \ln(n) + 2\psi_{\nu_0}(m/n)) + \mathcal{O}(m^{-2}).\end{aligned}$$

Lemmata A.3 and A.4 show that $|\psi_{\nu_0}(m/n)| \lesssim \ln(n)$. $\square$

Proof of Lemma A.21. Note that the $\lambda_{m,n}$ s are random since they depend on $(\nu_0, \hat{\alpha}_n)$. First, we have:

$$\begin{aligned}\frac{\partial \mathbb{L}_n}{\partial \phi}(\nu_0, \phi_0, \hat{\alpha}_n) &= \frac{1}{\phi_0 n} \sum_{m=0}^{n-1} 1 - \frac{\lambda_{m,n}^{(0)} U_{m,n}^2}{\lambda_{m,n}} \\ &= \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n}\right) + \frac{1}{\phi_0 n} \sum_{m=1}^{n-1} 1 - \frac{\lambda_{m,n}^{(0)} U_{m,n}^2}{\lambda_{m,n}} \quad (\text{Lemma A.1}) \\ &= \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n^\beta}\right) + \frac{1}{\phi_0 n} \sum_{m=1}^{n-1} 1 - U_{m,n}^2 \quad (\text{for some } \beta > 1/2 \text{ by Lemma A.10}).\end{aligned}$$

Furthermore, one has:

$$\begin{aligned}\frac{\partial \mathbb{L}_n}{\partial \nu}(\nu_0, \phi_0, \hat{\alpha}_n) &= \frac{1}{n} \sum_{m=0}^{n-1} \frac{\partial \lambda_{m,n} / \partial \nu}{\lambda_{m,n}} \left(1 - \frac{U_{m,n}^2 \lambda_{m,n}^{(0)}}{\lambda_{m,n}}\right) \\ &= \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n}\right) + \frac{1}{n} \sum_{m=1}^{n-1} \frac{\partial \lambda_{m,n} / \partial \nu}{\lambda_{m,n}} \left(1 - \frac{U_{m,n}^2 \lambda_{m,n}^{(0)}}{\lambda_{m,n}}\right) \quad (\text{Lemmata A.1 and B.1}) \\ &= \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n}\right) + \frac{1}{n} \sum_{m=1}^{n-1} \frac{\partial \lambda_{m,n} / \partial \nu}{\lambda_{m,n}} (1 - U_{m,n}^2) \\ &\quad + \frac{1}{n} \sum_{m=1}^{n-1} \frac{\partial \lambda_{m,n} / \partial \nu}{\lambda_{m,n}} U_{m,n}^2 \mathcal{O}(m^{-2} \vee (n-m)^{-2}) \quad (\text{essentially, by Lemma A.9}) \\ &= \mathcal{O}_{\mathbb{P}}\left(\frac{1}{n}\right) + \frac{1}{n} \sum_{m=1}^{n-1} \frac{\partial \lambda_{m,n} / \partial \nu}{\lambda_{m,n}} (1 - U_{m,n}^2)\end{aligned}$$

since $\partial \lambda_{m,n} / \partial \nu(\nu_0, \hat{\alpha}_n) \lesssim m^\delta \wedge (n-m)^\delta \lambda_{m,n}(\nu_0, \hat{\alpha}_n)$ holds essentially, thanks to Lemmata A.1 and B.1. (By “essentially”, we mean that the constant does not depend on the sample path.) Then, using Lemma B.2 leads to:

$$\frac{\partial \mathbb{L}_n}{\partial \nu}(\nu_0, \phi_0, \hat{\alpha}_n) = -\frac{2 \ln(n)}{n} \sum_{m=1}^{n-1} (1 - U_{m,n}^2) - \frac{2}{n} \sum_{m=1}^{n-1} \psi_{\nu_0}(m/n) (1 - U_{m,n}^2) + \mathcal{O}_{\mathbb{P}}\left(\frac{\ln(n)}{n}\right),$$

and subsequent calculations show that $A_n^T \nabla \mathbb{L}_n(\nu_0, \phi_0, \hat{\alpha}_n)$ equals:

$$\mathcal{O}_{\mathbb{P}}\left(\frac{1}{\sqrt{n}}\right) + \frac{2}{n \sqrt{\text{Var}(\psi_{\nu_0}(V))}} \sum_{m=1}^{n-1} (1 - U_{m,n}^2) \left(\frac{\mathbb{E}(\psi_{\nu_0}(V)) - \psi_{\nu_0}(m/n)}{\sqrt{\text{Var}(\psi_{\nu_0}(V))}} \right).$$

Conclude using a standard Lindeberg-Feller argument. (Lemmata A.3 and A.4 give (a multiple of) the envelope $x \mapsto -\ln(x)$ near zero for ψ_{ν_0} . Proceed as for Lemma A.19 to show that $n^{-1} \sum_{m=1}^{n-1} \psi_{\nu_0}^p(m/n) \rightarrow \int_0^1 \psi_{\nu_0}^p \cdot$) $\square$

Proof of Lemma A.22. Observe that $A_n^\top C_n A_n = 2I_2$, with:

$$C_n = \begin{pmatrix} 2\ln^2(n) + 4\ln(n)\mathbb{E}(\psi_{\nu_0}(V)) + 2\mathbb{E}(\psi_{\nu_0}^2(V)) & -\ln(n)\phi_0^{-1} - \mathbb{E}(\psi_{\nu_0}(V))\phi_0^{-1} \\ -\ln(n)\phi_0^{-1} - \mathbb{E}(\psi_{\nu_0}(V))\phi_0^{-1} & 2^{-1}\phi_0^{-2} \end{pmatrix}.$$

The proof is left to the reader. It consists in showing that $\nabla^2 \mathbb{L}_n(\nu_0, \phi_0, \hat{\alpha}_n)/2 = C_n + \mathcal{O}_{\mathbf{P}}(n^{-\epsilon})$ for some $\epsilon > 0$ by proceeding as for the proof of Lemma A.21. $\square$

C Proofs of Theorem 4.3 and Theorem 4.4

The posterior mean does not depend on ϕ , so all derivations will be written with $\phi = 1$. Furthermore, we will use the notation $c_j(\nu, \alpha)$ defined in Section A.1. Also, we assume that $\phi_0 = 1$ without loss of generality.

We avoid dealing with conditionally convergent series since we assume that $\nu_0 > 1/2$. In this case, the coefficients of the expansion (12) are almost surely absolutely summable and so the hypotheses of Proposition 2.1 are fulfilled. The proofs will rely on using Parseval's identity.

Let $(\nu, \alpha) \in (0, +\infty)^2$ and $j \in \mathbb{Z}$, we have

$$\begin{aligned} 2 \left| c_j(\xi - \hat{\xi}_n) \right|^2 &= \left(\frac{c_j(\nu, \alpha) \sum_{j_1 \in \mathbb{Z} \setminus \{0\}} \sqrt{c_{j+j_1 n}(\nu_0, \alpha_0)} U_{1, |j+j_1 n|}}{\sum_{j_1 \in \mathbb{Z}} c_{j+nj_1}(\nu, \alpha)} \right. \\ &\quad \left. - \frac{\sqrt{c_j(\nu_0, \alpha_0)} U_{1, |j|} \sum_{j_1 \in \mathbb{Z} \setminus \{0\}} c_{j+nj_1}(\nu, \alpha)}{\sum_{j_1 \in \mathbb{Z}} c_{j+nj_1}(\nu, \alpha)} \right)^2 \\ &+ \left(\frac{c_j(\nu, \alpha) \sum_{j_1 \in \mathbb{Z} \setminus \{0\}} \sqrt{c_{j+nj_1}(\nu_0, \alpha_0)} U_{2, |j+nj_1|} \text{sign}(j + nj_1)}{\sum_{j_1 \in \mathbb{Z}} c_{j+nj_1}(\nu, \alpha)} \right. \\ &\quad \left. - \frac{\sqrt{c_j(\nu_0, \alpha_0)} U_{2, |j|} \text{sign}(j) \sum_{j_1 \in \mathbb{Z} \setminus \{0\}} c_{j+nj_1}(\nu, \alpha)}{\sum_{j_1 \in \mathbb{Z}} c_{j+nj_1}(\nu, \alpha)} \right)^2 \end{aligned} \quad (32)$$

after a few algebraic manipulations. The expression (32) is a sum of two independent terms. Let $m \in \llbracket 0, n-1 \rrbracket$. If $j \in m + n\mathbb{Z}$ with $m \notin \{0, n/2\}$, then the two terms are identically distributed and involve independent Gaussian variables. Thus, there exists χ_2^2 distributed variables $A_{m,j,n}$ such that

$$\left| c_{m+jn}(\xi - \hat{\xi}_n) \right|^2 = a_{m,j,n}(\nu, \alpha) A_{m,j,n}/2 \quad (33)$$

with

$$\begin{aligned} a_{m,j,n}(\nu, \alpha) &= c_{m+jn}^2(\nu, \alpha) \frac{\sum_{j_1 \in \mathbb{Z}} c_{m+nj_1}(\nu_0, \alpha_0) - c_{m+jn}(\nu_0, \alpha_0)}{\left(\sum_{j_1 \in \mathbb{Z}} c_{m+nj_1}(\nu, \alpha) \right)^2} \\ &\quad + c_{m+jn}(\nu_0, \alpha_0) \left(1 - \frac{c_{m+jn}(\nu, \alpha)}{\sum_{j_1 \in \mathbb{Z}} c_{m+nj_1}(\nu, \alpha)} \right)^2. \end{aligned} \quad (34)$$

Lemma A.1 and Lemma C.2 make it straightforward to prove the following Lemma.

Lemma C.1. Let $A, N \subset (0, +\infty)$ be compact intervals. It holds that

$$a_{m,j,n}(\nu) \lesssim (|j|n)^{-4\nu-2} m^{4\nu-2\nu_0+1} + (|j|n)^{-2\nu_0-1}, \quad \text{for } j \neq 0,$$

and

$$a_{m,0,n}(\nu) \lesssim n^{-2\nu_0-1} + m^{4\nu-2\nu_0+1} n^{-4\nu-2},$$

uniformly in $\nu \in N$, $\alpha \in A$, $j \in \mathbb{Z}$, and $1 \leq m \leq \lfloor (n-1)/2 \rfloor$.

Lemma C.2. Let $\nu, \alpha > 0$, $0 \leq m \leq \lfloor n/2 \rfloor$, and $j \neq 0$. We have:

$$c_{m+nj}(\nu, \alpha) \leq 2^{2\nu+1} (n|j|)^{-2\nu-1}.$$

Proof. Using the fact that $m \leq n/2$ leads to:

$$c_{m+nj}(\nu, \alpha) \leq (n(|j| - 1/2))^{-2\nu-1} \leq 2^{2\nu+1} (n|j|)^{-2\nu-1}.$$

□

For $m \in \{0, n/2\}$ and $j \in \mathbb{Z}$, the two terms in (32) are not identically distributed. Moreover, for $q \in \{1, 2\}$ and $m \in \{0, n/2\}$, there are duplicates among the variables $\{U_{q,|m+nj|}, j \in \mathbb{Z}\}$. Nevertheless, the two terms are sums of independent Gaussian variables, so expressions like (33) hold. However, the presence of duplicates makes the expressions more complex than (34). The upper bounds given by assuming full redundancy among the variables appearing in the two terms of (32) suffice for our purposes. The following lemmata are adaptations of Lemma C.1. The statements are made uniform with respect to regularity ranges to be used in the proof of Theorem 4.4.

Lemma C.3. Let $N, A \subset (0, +\infty)$ be compact intervals, and write $\nu_{\min} = \min N$. Then:

$$\mathbb{E} \left(\sup_{\nu \in N, \alpha \in A} \sum_{j \in \mathbb{Z}} |c_{jn}(\xi - \hat{\xi}_n)|^2 \right) \lesssim n^{-2\nu_0-1} + n^{-4\nu_{\min}-2}.$$

Lemma C.4. Let $n \geq 2$ be even and $N, A \subset (0, +\infty)$ be compact intervals. Then:

$$\mathbb{E} \left(\sup_{\nu \in N, \alpha \in A} \sum_{j \in \mathbb{Z}} |c_{n/2+jn}(\xi - \hat{\xi}_n)|^2 \right) \lesssim n^{-2\nu_0-1}.$$

Proof of Theorem 4.3. We prove the (more general) result with $\hat{\alpha}_n \in A$, for a compact interval A . This will be useful for proving Theorem 4.4.

Let $m \in \llbracket 0, n-1 \rrbracket$ such that $m \notin \{0, n/2\}$ and consider indexes $m+nj$, with $j \in \mathbb{Z}$. Lemma C.1 and (33) yields:

$$\sum_{j \in \mathbb{Z}} \mathbb{E} \left(|c_{m+jn}(\xi - \hat{\xi}_n)|^2 \right) \lesssim n^{-2\nu_0-1} + n^{-4\nu-2} m^{4\nu-2\nu_0+1}. \quad (35)$$

The first two statements then follow from Lemmata C.3 and C.4, the identity

$$\sum_{j \in \mathbb{Z}} |c_{m+jn}(\xi - \hat{\xi}_n)|^2 = \sum_{j \in \mathbb{Z}} |c_{n-m+jn}(\xi - \hat{\xi}_n)|^2, \quad (36)$$

for every $0 \leq m \leq n-1$, the Fubini-Tonelli theorem, and Parseval's identity.

For the last statement, let $\nu > (\nu_0 - 1)/2$ and $1 \leq m \leq l$ with $l = \lfloor (n-1)/2 \rfloor$. Lemma A.2 gives

$$a_{m,j,n}(\nu, \hat{\alpha}_n) = (1 + \mathcal{O}(m^{-2})) \left(|m + jn|^{-4\nu-2} \frac{\sum_{j_1 \in \mathbb{Z} \setminus \{j\}} |m + j_1 n|^{-2\nu_0-1}}{\left(\sum_{j_1 \in \mathbb{Z}} |m + j_1 n|^{-2\nu-1} \right)^2} \right. \\ \left. + |m + jn|^{-2\nu_0-1} \left(\frac{\sum_{j_1 \in \mathbb{Z} \setminus \{j\}} |m + j_1 n|^{-2\nu-1}}{\sum_{j_1 \in \mathbb{Z}} c_{m+nj_1} |m + j_1 n|^{-2\nu-1}} \right)^2 \right),$$

for every $j \in \mathbb{Z}$, essentially. Consequently, it holds that:

$$\sum_{j \in \mathbb{Z}} \mathbb{E} \left(\left| c_{m+jn}(\xi - \hat{\xi}_n) \right|^2 \right) = \frac{(1 + \mathcal{O}(m^{-2}))}{n^{2\nu_0+1}} \vartheta_{\nu;\nu_0}(m/n)$$

after a few algebraic manipulations. Using the definition of γ , it is straightforward to show that

$$\vartheta_{\nu;\nu_0}(x) \sim C_1 x^{4\nu-2\nu_0+1} + C_2 \quad (37)$$

for some nonzero constants C_1, C_2 , when $x \rightarrow 0$. Therefore, the function $\vartheta_{\nu;\nu_0}$ is integrable if $\nu > (\nu_0 - 1)/2$ and⁵

$$\frac{1}{n} \sum_{m=1}^l \vartheta_{\nu;\nu_0}(m/n) \rightarrow \int_0^{1/2} \vartheta_{\nu;\nu_0}. \quad (38)$$

Then, Lemma C.4, Lemma C.3, the identity (36), the Fubini-Tonelli theorem, and Parseval's identity gives

$$n^{2\nu_0} \mathbb{E}(\text{ISE}_n(\nu, \hat{\alpha}_n; \xi)) = o(1) + \frac{2}{n} \sum_{m=1}^l (1 + \mathcal{O}(m^{-2})) \vartheta_{\nu;\nu_0}(m/n) \rightarrow \int_0^1 \vartheta_{\nu;\nu_0},$$

killing the $\mathcal{O}(m^{-2})$ -term using Hölder inequality and (37) as in the proof of Lemma A.10. $\square$

The following lemma bounds the rate at which ν falls within the range $[\nu_0 - 1/2, \nu_{\max}]$ of values giving reproducing kernel Hilbert spaces almost surely not containing ξ . It will be useful for proving Theorem 4.4.

Lemma C.5. Let $\epsilon > 0$. With the notations of Theorem 4.4, we have:

$$\mathbb{P}(\widehat{\nu}_n \leq \nu_0 - 1/2 - \epsilon) \lesssim e^{-C\sqrt{n}},$$

for some $C > 0$.

Proof. Let α_1 be any element of A . We proceed by bounding

$$\mathbb{P} \left(\inf_{\nu_{\min} \leq \nu \leq \nu_0 - 1/2 - \epsilon, \alpha \in A} \mathbb{M}_n(\nu, \alpha) - \mathbb{M}_n(\nu_0, \alpha_1) \leq 0 \right).$$

⁵Proceed as for Lemma A.19, using (37), if $(\nu_0 - 1)/2 < \nu < (\nu_0 - 1/2)/2$.

Then, let $\alpha \in A$ and $\nu_{\min} \leq \nu \leq \nu_0 - 1/2 - \epsilon$, we have:

$$\begin{aligned}
\mathbb{M}_n(\nu, \alpha) &= \mathcal{O}(1) + \ln \left(\frac{Z^\top R_{\nu, \alpha}^{-1} Z}{n^{1+2(\nu-\nu_0)}} \right) \quad (\text{Lemma A.5}) \\
&\geq \mathcal{O}(1) + \ln \left(\frac{\sum_{m=1}^{\lfloor \sqrt{n} \rfloor} U_{m,n}^2 \lambda_{m,n}^{(0)} / \lambda_{m,n}}{n^{1+2(\nu-\nu_0)}} \right) \\
&= \mathcal{O}(1) + \ln \left(\frac{\sum_{m=1}^{\lfloor \sqrt{n} \rfloor} U_{m,n}^2 m^{2(\nu-\nu_0)}}{n^{1+2(\nu-\nu_0)}} \right) \quad (\text{Lemma A.1}) \\
&= \mathcal{O}(1) + \ln \left(\frac{1}{n} \sum_{m=1}^{\lfloor \sqrt{n} \rfloor} U_{m,n}^2 \left(\frac{m}{n} \right)^{2(\nu-\nu_0)} \right) \\
&\geq \mathcal{O}(1) + \ln \left(\frac{1}{n} \sum_{m=1}^{\lfloor \sqrt{n} \rfloor} U_{m,n}^2 \left(\frac{m}{n} \right)^{-1-2\epsilon} \right) \\
&= \mathcal{O}(1) + 2\epsilon \ln(n) + \ln \left(\sum_{m=1}^{\lfloor \sqrt{n} \rfloor} U_{m,n}^2 m^{-1-2\epsilon} \right) \\
&\geq \mathcal{O}(1) + 2\epsilon \ln(n) + \ln \left(\sum_{m=1}^{\lfloor \sqrt{n} \rfloor} U_{m,n}^2 \lfloor \sqrt{n} \rfloor^{-1-2\epsilon} \right) \\
&\geq \mathcal{O}(1) + \epsilon \ln(n) + \ln \left(\frac{1}{\lfloor \sqrt{n} \rfloor} \sum_{m=1}^{\lfloor \sqrt{n} \rfloor} U_{m,n}^2 \right) \\
&\geq \mathcal{O}(1) + \epsilon \ln(n) + \frac{1}{\lfloor \sqrt{n} \rfloor} \sum_{m=1}^{\lfloor \sqrt{n} \rfloor} \ln(U_{m,n}^2) \quad (\text{Jensen inequality})
\end{aligned}$$

with a uniform big- $\mathcal{O}$. Let $\delta > 0$ and $t > 0$, we have

$$\begin{aligned}
\mathbb{P} \left(-\frac{1}{\lfloor \sqrt{n} \rfloor} \sum_{m=1}^{\lfloor \sqrt{n} \rfloor} \ln(U_{m,n}^2) \geq \delta \right) &= \mathbb{P} \left(e^{-\frac{t}{\lfloor \sqrt{n} \rfloor} \sum_{m=1}^{\lfloor \sqrt{n} \rfloor} \ln(U_{m,n}^2)} \geq e^{t\delta} \right) \\
&\leq e^{-\delta \lfloor \sqrt{n} \rfloor / 4} \mathbb{E} \left(|U_{1,1}|^{-1/2} \right)^{\lfloor \sqrt{n} \rfloor},
\end{aligned}$$

with $t = 1/4$ and $\mathbb{E} \left(|U_{1,1}|^{-1/2} \right) < +\infty$. This gives the desired convergence rate if δ is high enough.

Furthermore, we have

$$\mathbb{M}_n(\nu_0, \alpha_1) = \mathcal{O}(1) + \ln \left(n^{-1} \sum_{m=0}^{n-1} U_{m,n}^2 \right),$$

and

$$\mathbb{P} \left(\ln \left(n^{-1} \sum_{m=0}^{n-1} U_{m,n}^2 \right) \geq \delta \right) \leq e^{-C_2 n},$$

for some $C_2 > 0$ if $\delta > 0$ is high enough, using also a Chernoff bound argument. Now, putting all the pieces together yields:

$$\begin{aligned} & \inf_{\nu_{\min} \leq \nu \leq \nu_0 - 1/2 - \epsilon, \alpha \in A} \mathbb{M}_n(\nu, \alpha) - \mathbb{M}_n(\nu_0, \alpha_1) \\ & \geq \mathcal{O}(1) + \epsilon \ln(n) + \frac{1}{\lfloor \sqrt{n} \rfloor} \sum_{m=1}^{\lfloor \sqrt{n} \rfloor} \ln(U_{m,n}^2) - \ln \left(n^{-1} \sum_{m=0}^{n-1} U_{m,n}^2 \right) \end{aligned}$$

giving the result thanks to the pigeonhole principle. $\square$

Proof of Theorem 4.4 . The proof of Theorem 4.3 already deals with estimated $\hat{\alpha}_n \in A$. It is extended to estimated $\hat{\nu}_n \in N$ by bounding derivatives and using Lemma C.5.

Let $\epsilon > 0$ and $1 \leq m \leq l = \lfloor (n-1)/2 \rfloor$ and use the notation (34). The functions $a_{m,j,n}$ are smooth. For any (fixed) $0 < \delta < \nu_{\min}$, it holds that

$$\left| \frac{\partial c_j}{\partial \nu}(\nu, \alpha) \right| \lesssim c_j(\nu - \delta, \alpha),$$

uniformly in $\nu \in [\nu_0 - 1/2 - \epsilon, \nu_{\max}]$, $\alpha \in A$, and $j \in \mathbb{Z}$. Coordination with Lemmata A.1 and C.2 makes it possible to show that

$$\begin{aligned} \left| \frac{\partial a_{m,0,n}}{\partial \nu}(\nu, \alpha) \right| & \lesssim n^{-2\nu_0-1} m^{2\delta} + \frac{m^{-2\nu_0+1}}{n^{2-2\delta}} \left(\frac{m}{n} \right)^{4\nu} \\ & \leq n^{-2\nu_0-1} m^{2\delta} + \frac{m^{-2\nu_0+1}}{n^{2-2\delta}} \left(\frac{m}{n} \right)^{4\nu_0-2-4\epsilon} \\ & = n^{-2\nu_0-1} m^{2\delta} + \frac{m^{2\nu_0-1-4\epsilon}}{n^{4\nu_0-2\delta-4\epsilon}} \\ & \leq n^{-2\nu_0-1} m^{2\delta} + \frac{m^{-1-4\epsilon}}{n^{2\nu_0-2\delta-4\epsilon}} \end{aligned}$$

and, for $j \neq 0$, that

$$\begin{aligned} \left| \frac{\partial a_{m,j,n}}{\partial \nu}(\nu, \alpha) \right| & \lesssim (|j|n)^{-4\nu-2+2\delta} m^{4\nu-2\nu_0+1+2\delta} + (|j|n)^{-2\nu_0-1+2\delta} \\ & = |j|^{-4\nu-2+2\delta} n^{-2+2\delta} \left(\frac{m}{n} \right)^{4\nu} m^{-2\nu_0+1+2\delta} + (|j|n)^{-2\nu_0-1+2\delta} \\ & \leq |j|^{-2+2\delta} n^{-2+2\delta} \left(\frac{m}{n} \right)^{4\nu_0-2+4\epsilon} m^{-2\nu_0+1+2\delta} + (|j|n)^{-2\nu_0-1+2\delta} \\ & = |j|^{-2+2\delta} n^{-4\nu_0+2\delta-4\epsilon} m^{2\nu_0-1+4\epsilon+2\delta} + (|j|n)^{-2\nu_0-1+2\delta} \end{aligned}$$

uniformly in $1 \leq m \leq l$, $j \neq 0$, $\alpha \in A$, and $\nu \in [\nu_0 - 1/2 - \epsilon, \nu_{\max}]$. Then,

$$\sum_{m=1}^l \sum_{j \in \mathbb{Z}} \mathbb{E} \left(A_{m,j,n} |a_{m,j,n}(\hat{\nu}_n, \hat{\alpha}_n) - a_{m,j,n}(\nu_0, \hat{\alpha}_n)| \mathbf{1}_{\hat{\nu}_n \geq \nu_0 - 1/2 - \epsilon} \right)$$

$$\begin{aligned}
&\leq \sum_{m=1}^l \sum_{j \in \mathbb{Z}} \mathbb{E} \left(A_{m,j,n} |\widehat{\nu}_n - \nu_0| \sup_{\nu_0 - 1/2 - \epsilon \leq \nu \leq \nu_{\max}, \alpha \in A} \left| \frac{\partial a_{m,j,n}}{\partial \nu}(\nu, \alpha) \right| \right) \\
&= \sqrt{\mathbb{E}(A_{1,0,1}^2)} \sqrt{\mathbb{E}((\widehat{\nu}_n - \nu_0)^2)} \sum_{m=1}^l \sum_{j \in \mathbb{Z}} \sup_{\nu_0 - 1/2 - \epsilon \leq \nu \leq \nu_{\max}, \alpha \in A} \left| \frac{\partial a_{m,j,n}}{\partial \nu}(\nu, \alpha) \right| = o(n^{-2\nu_0}),
\end{aligned}$$

for δ and ϵ small enough and using the above inequalities and Theorem 4.2. Therefore, Lemmata C.3 and C.4, the identity (36), and the Fubini-Tonelli theorem shows that

$$\mathbb{E}(|\text{ISE}_n(\widehat{\nu}_n, \widehat{\alpha}_n; \xi) - \text{ISE}_n(\nu_0, \widehat{\alpha}_n; \xi)| \mathbb{1}_{\widehat{\nu}_n \geq \nu_0 - 1/2 - \epsilon}) = o(n^{-2\nu_0}).$$

Furthermore, using again the Fubini-Tonelli theorem yields

$$\begin{aligned}
&\mathbb{E} \left(\sum_{m=1}^l \sum_{j \in \mathbb{Z}} |c_{m+jn}(\xi - \widehat{\xi}_n)|^2 \mathbb{1}_{\widehat{\nu}_n \leq \nu_0 - 1/2 - \epsilon} \right) \\
&= \sum_{m=1}^l \sum_{j \in \mathbb{Z}} \mathbb{E}(a_{m,j,n}(\widehat{\nu}_n, \widehat{\alpha}_n) A_{m,j,n} \mathbb{1}_{\widehat{\nu}_n \leq \nu_0 - 1/2 - \epsilon}) \\
&\leq \sum_{m=1}^l \sum_{j \in \mathbb{Z}} \sup_{\nu_{\min} \leq \nu \leq \nu_0 - 1/2 - \epsilon, \alpha \in A} a_{m,j,n}(\nu, \alpha) \mathbb{E}(A_{m,j,n} \mathbb{1}_{\widehat{\nu}_n \leq \nu_0 - 1/2 - \epsilon}) \\
&\leq \sqrt{\mathbb{E}(A_{1,0,1}^2)} \sqrt{\mathbb{E}(\mathbb{1}_{\widehat{\nu}_n \leq \nu_0 - 1/2 - \epsilon})} \sum_{m=1}^l \sum_{j \in \mathbb{Z}} \sup_{\nu_{\min} \leq \nu \leq \nu_0 - 1/2 - \epsilon, \alpha \in A} a_{m,j,n}(\nu, \alpha) \\
&\leq \sqrt{\mathbb{E}(A_{1,0,1}^2)} \sqrt{\mathbb{E}(\mathbb{1}_{\widehat{\nu}_n \leq \nu_0 - 1/2 - \epsilon})} n^\beta \quad \text{for some } \beta \text{ given by Lemma C.1} \\
&= o(n^{-2\nu_0}),
\end{aligned}$$

using Lemma C.5. Then, the sum for $j \equiv 0 \pmod{n}$ can be bounded similarly using Lemma C.3 and the sum for $j \equiv n/2 \pmod{n}$ is controlled by Lemma C.4 for n even.

Finally, the previous reasoning is easily applied to bound

$$\mathbb{E}(\text{ISE}_n(\nu_0, \widehat{\alpha}_n; \xi) \mathbb{1}_{\widehat{\nu}_n \leq \nu_0 - 1/2 - \epsilon})$$

and the desired result follows. $\square$

D Proofs of Section 5

Note that the finiteness of $\nu_0(f)$ is assumed so that f is necessarily nonzero. Consequently, the data Z is ultimately nonzero under the observation model (3) since f is continuous. Furthermore, we assume that $\nu_0(f) > 1$, so $f \in H^\beta[0, 1]$ for some $\beta > 1$. Consequently, the Sobolev embedding theorem implies that f has Hölder regularity strictly greater than $1/2$. Hence, f has absolutely summable Fourier coefficients.

The proofs are based on the observation that

$$Z^\top R_{\nu, \alpha}^{-1} Z = \sum_{m=0}^{n-1} \frac{\left| \sum_{j \in m+n\mathbb{Z}} c_j(f) \right|^2}{\sum_{j \in m+n\mathbb{Z}} c_j(\nu, \alpha)},$$

using (19) and elements from Section A.2.

Proof of Proposition 5.2 . We give a full proof only for the third assumption. The proof for the second assumption is similar and the first is a particular case of the second.

Let $\epsilon > 0$, $\nu_{\min} \leq \nu \leq \nu_0(f) - 1/2 - \epsilon$, $\alpha \in A$, and $p \in \mathbb{Z}$ such that $c_p(f) \neq 0$. Then, Proposition 2.2 and Lemma A.5 gives

$$\begin{aligned} \mathbb{M}_n^f(\nu, \alpha) &= 2(\nu_0(f) - \nu - 1/2) \ln(n) + \mathcal{O}(1) + \ln \left(\sum_{m=0}^{n-1} \frac{\left| \sum_{j \in m+n\mathbb{Z}} c_j(f) \right|^2}{\sum_{j \in m+n\mathbb{Z}} c_j(\nu, \alpha)} \right) \\ &\geq 2\epsilon \ln(n) + \mathcal{O}(1) + \ln \left(\frac{\left| \sum_{j \in p+n\mathbb{Z}} c_j(f) \right|^2}{\sum_{j \in p+n\mathbb{Z}} c_j(\nu, \alpha)} \right) \quad (\text{consider } m = p \pmod n) \\ &= 2\epsilon \ln(n) + \mathcal{O}(1) \end{aligned}$$

uniformly since $\sum_{j \in p+n\mathbb{Z}} c_j(f) \rightarrow c_p(f)$ and by Lemma A.1. (If $p < 0$, then use the symmetry of the $c_j(\nu, \alpha)$ s.)

Moreover, for $\nu = \nu_0(f) - 1/2 - \epsilon/2$ and any fixed $\alpha \in A$, we have:

$$\begin{aligned} \mathbb{M}_n^f(\nu_0(f) - 1/2 - \epsilon/2, \alpha) &= \epsilon \ln(n) + \mathcal{O}(1) + \ln(Z^\top R_{\nu, \alpha}^{-1} Z) \\ &\leq \epsilon \ln(n) + \mathcal{O}(1), \end{aligned}$$

since $f \in H^\beta[0, 1]$ for $\beta = \nu_0(f) - \epsilon/2$. Indeed, this Sobolev space is norm-equivalent to the reproducing kernel Hilbert space attached to the covariance function for $\nu = \nu_0(f) - 1/2 - \epsilon/2$ and the corresponding quadratic term $Z^\top R_{\nu, \alpha}^{-1} Z$ is the squared norm of a projection of f (see, e.g., Wendland, 2004, Theorem 13.1). This completes the proof. $\square$

Proof of Proposition 5.3 . Without loss of generality, consider a compact subset $N \times A$ with $A = [\alpha_{\min}, \alpha_{\max}]$ and $N = [\nu_0(f) - 1/2 + \epsilon, \nu_{\max}]$, for some $\epsilon > 0$. Then, Proposition 2.2 and Lemma A.5 yield:

$$\mathbb{M}_n^f(\nu, \alpha) = \int_0^1 g_\nu + \mathcal{O}\left(\frac{\ln(n)}{n}\right) + \ln \left(n^{2(\nu_0(f) - \nu - 1/2)} \sum_{m=0}^{n-1} \frac{\left| \sum_{j \in m+n\mathbb{Z}} c_j(f) \right|^2}{\sum_{j \in m+n\mathbb{Z}} c_j(\nu, \alpha)} \right),$$

with a uniform big- $\mathcal{O}$. Focus now on the term inside the logarithm. For $1 \leq m \leq n-1$, Lemma A.3 shows that $\gamma(\nu_0(f) + 3/2; m/n) \approx n(m^{-1} \vee (n-m)^{-1}) \gamma(\nu_0(f) + 1/2; m/n)$. Thus, using the hypothesis on the $c_j(f)$ we have:

$$\begin{aligned} \sum_{j \in \mathbb{Z}} c_{jn+m}(f) &= \sum_{j \in \mathbb{Z}} |jn+m|^{-\nu_0(f)-1/2} + \mathcal{O}\left(\sum_{j \in \mathbb{Z}} |jn+m|^{-\nu_0(f)-3/2}\right) \\ &= n^{-\nu_0(f)-1/2} \gamma(\nu_0(f) + 1/2; m/n) + \mathcal{O}\left(n^{-\nu_0(f)-3/2} \gamma(\nu_0(f) + 3/2; m/n)\right) \\ &= n^{-\nu_0(f)-1/2} \gamma(\nu_0(f) + 1/2; m/n) (1 + \mathcal{O}(m^{-1} \vee (n-m)^{-1})). \end{aligned}$$

(It holds that $\sum_{j \in n\mathbb{Z}} c_j(f) \rightarrow c_0(f)$, so the term for $m = 0$ is a uniform big- $\mathcal{O}$ thanks to Lemma A.1.) Then, use Lemma A.2 to get:

$$\begin{aligned}
& n^{2(\nu_0(f) - \nu - 1/2)} \sum_{m=0}^{n-1} \frac{\left| \sum_{j \in m+n\mathbb{Z}} c_j(f) \right|^2}{\sum_{j \in m+n\mathbb{Z}} c_j(\nu, \alpha)} \\
&= \mathcal{O}(n^{-2\epsilon}) + \frac{1}{n} \sum_{m=1}^{n-1} \frac{(1 + \mathcal{O}(m^{-1} \vee (n-m)^{-1})) \gamma^2(\nu_0(f) + 1/2; m/n)}{\gamma(2\nu + 1; m/n)} \\
&= \mathcal{O}(n^{-2\epsilon}) + \mathcal{O}(n^{-\epsilon}) + \frac{1}{n} \sum_{m=1}^{n-1} \frac{\gamma^2(\nu_0(f) + 1/2; m/n)}{\gamma(2\nu + 1; m/n)}
\end{aligned}$$

using Hölder inequality with $1/p = 1 - \epsilon$, similarly to the proof of Lemma A.10. The uniform convergence of the Riemann sum can be proved similarly to Lemma A.19, using (a multiple of) the envelope $x \mapsto x^{2\epsilon} - 1$. $\square$

References

- M. L. Stein. *Interpolation of spatial data*. Springer Series in Statistics. Springer-Verlag, New York, 1999. Some theory for Kriging.
- Thomas Santner, Brian Williams, and William Notz. *The Design and Analysis Computer Experiments*. 01 2003.
- C. E. Rasmussen and C. K. I. Williams. *Gaussian Processes for Machine Learning*. Adaptive Computation and Machine Learning. MIT Press, Cambridge, MA, USA, 2006.
- B. Matérn. *Spatial Variation*. Springer-Verlag New York, 1986.
- F. Bachoc. Asymptotic analysis of maximum likelihood estimation of covariance parameters for Gaussian processes: an introduction with proofs. In *Advances in Contemporary Statistics and Econometrics*, pages 283–303. Springer, 2021.
- K. V. Mardia and R. J. Marshall. Maximum likelihood estimation of models for residual covariance in spatial regression. *Biometrika*, 71(1):135–146, 1984.
- F. Bachoc. Asymptotic analysis of the role of spatial sampling for covariance parameter estimation of Gaussian processes. *Journal of Multivariate Analysis*, 125:1–35, 2014.
- Z. Ying. Asymptotic properties of a maximum likelihood estimator with data from a Gaussian process. *Journal of Multivariate analysis*, 36(2):280–296, 1991.
- Z. Ying. Maximum likelihood estimation of parameters under a spatial sampling scheme. *The Annals of Statistics*, pages 1567–1590, 1993.
- A. W. van der Vaart. Maximum likelihood estimation under a spatial sampling scheme. *The Annals of Statistics*, pages 2049 – 2057, 1996.

- H. Zhang. Inconsistent estimation and asymptotically equal interpolations in model-based geostatistics. *Journal of the American Statistical Association*, 99(465):250–261, 2004.
- W.-L. Loh. Fixed-domain asymptotics for a subclass of matérn-type Gaussian random fields. *The Annals of Statistics*, 33, 2006.
- C. G. Kaufman and B. Adam. Shaby. The role of the range parameter for estimation and prediction in geostatistics. *Biometrika*, 100(2):473–484, 2013.
- C. Li. Bayesian fixed-domain asymptotics: Bernstein-von Mises theorem for covariance parameters in a Gaussian process model. *arXiv preprint arXiv:2010.02126*, 2020.
- Michael L. Stein. Spline smoothing with an estimated order parameter. *The Annals of Statistics*, 21(3):1522–1544, 1993.
- Y. Chen, H. Owhadi, and A. Stuart. Consistency of empirical bayes and kernel flow for hierarchical parameter estimation. *Mathematics of Computation*, 90(332):2527–2578, 2021.
- T. Karvonen. Asymptotic bounds for smoothness parameter estimates in gaussian process interpolation. *arXiv preprint arXiv:2203.05400v4*, 2023.
- H. Putter and G. A. Young. On the effect of covariance function estimation on the accuracy of kriging predictors. *Bernoulli*, 7(3):421 – 438, 2001.
- T. Karvonen, G. Wynne, F. Tronarp, C. Oates, and S. Sarkka. Maximum likelihood estimation and uncertainty quantification for Gaussian process approximation of deterministic functions. *SIAM/ASA Journal on Uncertainty Quantification*, 8(3):926–958, 2020.
- T. Karvonen and C. J. Oates. Maximum likelihood estimation in gaussian process regression is ill-posed. *Journal of Machine Learning Research*, 24(120):1–47, 2023.
- E. Belitser and S. Ghosal. Adaptive Bayesian inference on the mean of an infinite-dimensional normal distribution. *The Annals of Statistics*, 31(2):536 – 559, 2003.
- B. T. Knapik, B. T. Szabó, A. W. van Der Vaart, and J. H. van Zanten. Bayes procedures for adaptive inference in inverse problems for the white noise model. *Probability Theory and Related Fields*, 164(3):771–813, 2016.
- B. Szabó, A. W. van Der Vaart, and J. H. van Zanten. Frequentist coverage of adaptive nonparametric Bayesian credible sets. *The Annals of Statistics*, 43(4):1391–1428, 2015.
- T. Gneiting. Strictly and non-strictly positive definite functions on spheres. *Bernoulli*, 19:1327–1349, 2011.
- G. Matheron. The theory of regionalized variables and its applications. Technical Report Les cahiers du CMM de Fontainebleau, Fasc. 5, Ecole des Mines de Paris, 1971.

- P. Craven and G. Wahba. Smoothing noisy data with spline functions. Estimating the correct degree of smoothing by the method of generalized cross-validation. *Numerische Mathematik*, 1979.
- W. Wang and B.-Y. Jing. Gaussian process regression: Optimality, robustness, and relationship with kernel ridge regression. *Journal of Machine Learning Research*, 23(193):1–67, 2022.
- P. J. Brockwell and R. A. Davis. *Time series: theory and methods*. Springer science & business media, 1987.
- A. G. Postnikov. *Introduction to analytic number theory*, volume 68. American Mathematical Soc., 1988.
- R. L. Taylor and T.-C. Hu. Strong laws of large numbers for arrays of rowwise independent random elements. *International Journal of Mathematics and Mathematical Sciences*, 10:805–814, 1987.
- A. W. van Der Vaart and J. A. Wellner. *Weak convergence and empirical processes: with applications to statistics*. Springer Science & Business Media, 1996.
- H. Wendland. *Scattered data approximation*, volume 17. Cambridge university press, 2004.