An Asymptotic Framework for Kriging under the Matérn Covariance Model

Eva M. Furrer, Douglas W. Nychka and Cécile Piret

Institute for Mathematics Applied to Geosciences, National Center for Atmospheric Research,
P.O. Box 3000, Boulder, CO 80307

1 Introduction

A general statistical problem is to estimate a smooth function from discrete and possibly noisy observations. This problem is central to the analysis of spatial data in many fields including our particular interest in climate observations and climate change. We have found that a common spatial estimator, Kriging, can be analyzed for its asymptotic properties by identifying the estimator as a special type of spline. By extending the asymptotic results from smoothing splines to this more general case it is possible to identify an equivalent kernel estimator that approximates Kriging. This connection then gives simple forms for the bias and variance of the curve estimate as the sample size becomes large and the density of observation points increases. The equivalence between Kriging and splines is not new and there is some work on the mean squared error of Kriging estimates. However, this new approach makes it possible to understand the asymptotic properties under irregular spacings of observations and for fixed functions. Kernel smoothers for estimating curves and surfaces have broad appeal as nonparametric curve estimators and also are supported by a rich base of statistical theory. Our results have interest because they tie the Kriging method to this more common class of statistical methods. Although our direct analysis is only for one dimensional problems and for a specific class of covariance functions (Matérn) there appear to be no substantial obstacles to extend this approach to other covariance families and two or more dimensions.

A basic additive model for nonparametric regression is

$$y_i = g(x_i) + \epsilon_i, \quad i = 1, \ldots, n,$$

where g is a smooth function observed at "locations" $(x_1, \ldots, x_n)$ and with "measurement" errors ϵ_i. The goal is to estimate g based on the observations $(y_1, \ldots, y_n)$. The geostatistical interpretation of this model adds some additional stochastic elements. For this discussion we assume that g is a realization of a stochastic process with mean zero and covariance function k.

One fruitful approach to derive nonparametric curve estimators is through a variational or penalized likelihood approach. We sketch this method with the motivation that this form will include Kriging estimators and also splines. Consider a penalized least squares criterion:

$$\mathcal{L}(f) = \frac{1}{n}\sum_{i=1}^{n}\big(y_i - f(x_i)\big)^2 + \lambda[f,f], \tag{1}$$

λ is the smoothing parameter and $[\cdot,\cdot]$ is a bilinear form that is always nonnegative and defines the penalty. The estimate for g is taken to be the minimizer over an appropriate Hilbert space of functions where the penalty is finite. For the case of smoothing splines in one dimension this inner product is the integral of the squared mth derivative, i.e. penalizing roughness of the estimated curve or surface and the Hilbert space is the mth order Sobelev space. For Kriging the penalty inner product and the Hilbert space are derived in a simple way from the covariance function and more details will be given in the next section. In either case, when this expression is minimized over an appropriate space of functions under simple conditions the minimizer exists, is unique and has a simple finite dimensional form. For this discussion the resulting estimator takes the form

$$\hat{g}(x) = \frac{1}{n}\sum_{i=1}^{n} w(x,x_i)y_i \tag{2}$$

for a weighting function w that depends on the inner product, the locations and λ but not on $\mathbf{y}$. For the geostatistical method of Kriging the estimator is usually described as a weighted linear combination of the observations with the weights derived as the best linear and unbiased estimate of g under a stochastic model. Assume that g is a spatial process with mean zero and covariance function k and assume the errors are mean zero, uncorrelated and with variance σ^2. Equating $\lambda = \sigma^2/n$ the weight function described above related to the minimization of (1) is identical to the Kriging weights.

The main contribution of this paper is to rigorously adapt the techniques described in **?** for smoothing splines to the case of one-dimensional Kriging based on a Matérn covariance model. More specifically, we will identify a kernel function, $G_\lambda(x,x')$ having a simple form that approximates w in (2) in the sense that the bias, variance and mean squared error of $\hat{g}(x)$ can be approximated under some conditions using

$$\tilde{g}(x) = \frac{1}{n}\sum_{i=1}^{n} G_\lambda(x,x_i)y_i. \tag{3}$$

Using this simpler form for G_λ one can follow the standard analysis for kernel smoothers to infer the asymptotic bias and variance, in reference to this we refer to G_λ as the equivalent kernel. We end this introduction by stating the main result of this paper.

Assume g to be a mean zero stationary process with covariance given by the Matérn model (e.g., **?**)

$$k(x,x') = \frac{\phi}{\Gamma(\nu)2^{\nu-1}}(\alpha r)^{\nu}K_\nu(\alpha r),$$

where $r = |x - x'|$, $\phi, \alpha, \nu > 0$ and K_ν is a modified Bessel function of the second kind and we assume $\phi = 1$ without loss of generality. α is a scale parameter and ν indexes the differentiability of g. Specifically, if m is the integer part of ν then g has m derivatives existing in the mean square sense.

Let F_n denote the empirical distribution function for $\{x_1, ..., x_n\}$, F the uniform distribution on $[0,1]$ and $D_n = \sup |F_n - F|$.

Theorem 1.1. *Still true? Let g be a mean zero stochastic process following a Matérn covariance with parameters $\nu > 1$ and $\alpha > 0$. If $D_n n^{1/(2\nu-1)} \to 0$ then*

$$\mathcal{E}_{g,\boldsymbol{\epsilon}}\big[\big(\hat{g}(x) - g(x)\big)^2\big] = Cn^{\frac{2\nu}{2\nu+1}}\big(1 + o(1)\big)$$

as $n \to \infty$ and for x in the interior of [0,1].

To give a simple statement the expectation is taken over both the measurement errors, $\boldsymbol{\epsilon}$ and also g. This is unusual for an analysis of a nonparametric regression method but is standard from a geostatistical perspective. The main theorem in Section 3 addresses more than one? the asymptotic bias and variance for fixed functions, however, whether they are realizations of a stochastic process or elements in a Hilbert space. The main contribution in this paper is being able to handle fairly arbitrary distributions of locations provided that they are sufficiently close to a limiting uniform density. This is in contrast to previous work in this area (**?**) that requires observations on regular grids or makes specific assumptions on the minimum and maximum spacing of the locations for the one dimensional case. A practical focus is on two-dimensional problems where the locations are not only irregular but can change in their local density. The US surface temperature observational network is an example of this kind of heterogenity. Although our rigorous analysis does not include this case we believe the basic mathematical tools are well suited to higher dimensions and nonuniform location densities.

2 Penalized Least Squares and Kriging

This section gives an outline of the estimators derived from minimizing a penalized least squares criterion. The main point is that the exact and approximate weight functions can be expressed as reproducing kernels. The reproducing property in turn provides the basic identity for the approximation bounds. First we give a few necessary definitions.

Definition 2.1 (Reproducing kernel)**.** *Let $\mathcal{H}$ be a Hilbert space of functions on the real line with inner product $\langle \cdot, \cdot \rangle$. In our discussion $r(\cdot, \cdot)$ is a reproducing kernel provided that $r(\cdot, x) \in \mathcal{H}$ for all $x \in \mathbb{R}$ and that for all $f \in \mathcal{H}$ and $x \in \mathbb{R}$ $\langle f, r(\cdot, x) \rangle = f(x)$. From this property it follows that $\langle r(\cdot, x), r(\cdot, x') \rangle = r(x, x')$ which in turn implies that r will be a symmetric and positive definite function.*

Besides identifying a reproducing kernel from a specific Hilbert space with inner product it is also possible to proceed in reverse direction. That is, given a positive definite kernel, r, one can formally construct an inner product and a Hilbert space such that r is the reproducing kernel. The reader is referred to **?** for details.

Definition 2.2 (Fourier transform). *The Fourier transform of a function $f \in C^{\infty}(\mathbb{R})$ and its inverse are defined as*

$$\mathcal{F}(f)(\omega) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-i\omega t} f(t)dt \quad \text{and} \quad \mathcal{F}^{-1}(\mathcal{F}(f))(t) = \int_{\mathbb{R}} e^{i\omega t}\mathcal{F}(f)(\omega)d\omega.$$

For k a stationary covariance function we will adopt the bilinear form

$$[f_1, f_2] = \int_{\mathbb{R}} \frac{\mathcal{F}(f_1)(\omega)\mathcal{F}(f_2)(\omega)}{\mathcal{F}(k)(\omega)}\,\mathrm{d}\omega,$$

as the penalty in (1). With this definition $[\cdot,\cdot]$ can also be interpreted as an inner product for a Hilbert space, $\mathcal{H}$ such that k is a reproducing kernel. $\mathcal{H}$ is also a Hilbert space under the inner product:

$$\langle f, g\rangle_{\lambda}^{n} = (1/n)\sum_{i=1}^{n} f(x_i)g(x_i) + \lambda[f, g]$$

and finally we identify w from (2) as the reproducing kernel under this inner product.

? introduced the approach of characterizing the weighting function w through the first order conditions for the minimizer of $\mathcal{L}$ in (1) and it was subsequently used by **?** for smoothing splines and **?** for Kriging. See **?** for a recent derivation that w corresponds indeed to the Kriging weights with respect to the stationary covariance function, k.

The approximating kernel G_{λ} used through this work is the reproducing for $\mathcal{H}$ but with respect to the inner product.

$$\langle f, g\rangle_{\lambda} = \int_{\infty}^{\infty} f(t)g(t)\,dt + \lambda[f_1, f_2],$$

Here the intuition is that we have replaced the discrete sum over the observations with an integral approximation. This is consistent with the assumption that the observations converge to a uniform distribution under infill asymptotics.

By using the reproducing properties of both w and G_{λ} and the fact that both inner products contain the same penalty term it follows that

$$\begin{aligned} w(u,v) - G_{\lambda}(u,v) &= \langle w(\cdot,v), G_{\lambda}(\cdot,u)\rangle_{\lambda} - \langle G_{\lambda}(.,u), w(.,v)\rangle_{\lambda}^{n} \\ &= \big(\textstyle\int_{\infty}^{\infty} w(t,v)G_{\lambda}(u,t)\,dt - \frac{1}{n}\sum_{i=1}^{n} w(x_i,v)G_{\lambda}(x_i,u)\big) \\ &= \textstyle\int_0^1 w(t,v)G_{\lambda}(t,u)\mathrm{d}\big[F(t) - F_n(t)\big] + \int_{\mathbb{R}\setminus[0,1]} w(t,v)G_{\lambda}(t,u)dt \end{aligned} \tag{4}$$

Thus, provided that the difference between the integral and sum is negligible and some decay G_{λ} will be a good approximation to w, see Theorem 3.4 and Section 4.2 for more details.

Using Parseval's theorem and properties of the Fourier transform **?** find that

$$\mathcal{F}(G_\lambda)(\omega,\omega') = \Big(1 + \lambda \frac{1}{\mathcal{F}(k)(\omega,\omega')}\Big)^{-1}, \tag{5}$$

i.e. the Fourier transform of the reproducing kernel G_λ is essentially determined by the Fourier transform of the underlying covariance function. In the case of the Matérn covariance model we have

$$\mathcal{F}(k)(\omega,\omega') = \frac{\phi\Gamma(\nu+1/2)\alpha^{2\nu}}{\Gamma(\nu)\pi^{1/2}(\alpha^2+\rho^2)^{\nu+1/2}},$$

with $\rho = |\omega - \omega'|$ and as a consequence $G_\lambda(x,x')$ analogous to k depends on its arguments only through $r = |x - x'|$. This simple characterization of G_λ under the Matérn model allows to determine its behavior for $t \to \infty$, see Theorem 3.1.

3 Theorems

Theorem 3.1 (Exponential Envelope Condition)**.** *For arbitrary Matérn range parameter $\alpha > 0$ and smoothing parameter $\lambda > 0$ tending to zero and if the Matérn smoothness parameter $\nu + 1/2 = m/\ell > 1$ with $m, \ell \in \{1, 2, \dots\}$ then the equivalent kernel G_λ satisfies*

$$|G_\lambda^{(k)}(x,x')| \le C\lambda^{-\frac{k+1}{2\nu+1}} \exp\big[-A\lambda^{-1/(2\nu+1)}|x-x'|\big], \quad \textit{for}\, x, x' \in \mathbb{R}, k = 0,1,2, \tag{6}$$

where C and A are positive constants depending on α and ν only. We refer to equation (6) as the Exponential Envelope Condition.

Proof: See Section 4.1. □

Remark 3.2. **?** *proves asymptotic forms of bias and variance only for the case of $m = 1$ due to a lacking formal proof of the Exponential Envelope Condition for higher order smoothing splines. Smoothing splines of order m correspond to $\alpha = 0$ and $\nu + 1/2 = m \in \{1, 2 \dots\}$ in our notation and the technical results can be extended easily to this case pointing to a simple way of proving the conjectures of* **?**.

Assumption 3.3 (Rate of infill asymptotics)**.** $\delta_n = 2C\Big(\frac{1}{\Delta} + \frac{1}{A-\Delta}\Big)D_n\lambda^{-1/(2\nu+1)} < 1$, *where $A > \Delta > 0$, ν is the smoothness parameter of the Matérn covariance function and C, A are the constants of the Exponential Envelope Condition (Theorem 3.1).*

Theorem 3.4 (Approximation)**.** *Let $A > \Delta > 0$. For Matérn range parameter $\alpha > 0$, smoothing parameter $\lambda > 0$ tending to zero, Matérn smoothness parameter $\nu+1/2 = m/\ell > 1$*

with $m, \ell \in \{1, 2, \dots\}$ *and under Assumption 3.3 we have*

$$\big|w(x,x')\big| < \frac{C}{1-\delta_n}\lambda^{-\frac{1}{2\nu+1}} \exp\Big(-(A-\Delta)\lambda^{-\frac{1}{2\nu+1}}|x-x'|\Big)$$

$$\big|w(x,x') - G_\lambda(x,x')\big| < \frac{\delta_n C}{1-\delta_n}\lambda^{-\frac{1}{2\nu+1}} \exp\Big(-(A-\Delta)\lambda^{-\frac{1}{2\nu+1}}|x-x'|\Big)$$

$$\Big|\frac{\partial}{\partial x}w(x,x')\Big| < \frac{C}{1-\delta_n}\lambda^{-\frac{2}{2\nu+1}} \exp\Big(-(A-\Delta)\lambda^{-\frac{1}{2\nu+1}}|x-x'|\Big)$$

uniformly over $x, x' \in [0,1]$ *for* C, A *the constants of the Exponential Envelope Condition (Theorem 3.1).*

Proof: See Section 4.2

Theorem 3.5 (Bias and Variance)**.** *Assume a Matérn smoothness parameter satisfying* $\nu + 1/2 = m/\ell > 1$ *with* $m, \ell \in \{1, 2, \dots\}$, $D_n \to 0$ *as* $n \to \infty$ *and that*

$$\lambda_n \sim D_n^{\frac{2\nu+1}{2\nu+2}} \log n \text{ as } n \to \infty. \tag{7}$$

1. *Suppose that a given path of the spatial field* g *is bounded on* $[0,1]$ *with bounded derivative and* $\mathcal{F}(g)(\omega) \sim |\omega|^{-\mu}$ *for* $\mu > 1$ *and* $|\omega| \to \infty$ *(this implies conditions on* ν*), then the bias of the Kriging estimator satisfies as* $\lambda_n \to 0$

$$\begin{aligned}\mathcal{E}\big[\hat{g}(x)|g\big] &= \int_{\mathbb{R}} G_{\lambda_n}(|x-u|)g(u)\,du - g(x) + \mathcal{O}\Big(D_n\lambda_n^{\frac{1}{2\nu+1}}\Big)\\ &= -\,\lambda_n^{\frac{\mu-1}{2\nu+1}} \int_{\mathbb{R}} \frac{(\lambda_n^{\frac{1}{\nu+1/2}}\alpha^2+\omega^2)^{\nu+1/2}}{\alpha^{2\nu}c + (\lambda_n^{\frac{1}{\nu+1/2}}\alpha^2+\omega^2)^{\nu+1/2}}|\omega|^{-\nu}\,d\omega + \mathcal{O}\Big(D_n\lambda_n^{-\frac{1}{2\nu+1}}\Big)\end{aligned}$$

2. *The variance of the Kriging estimator given the spatial field* g *satisfies as* $\lambda_n \to 0$

$$\begin{aligned}Var\big[\hat{g}(x)|g\big] =& \frac{\sigma^2}{n}\Big[\int_{\mathbb{R}} G_{\lambda_n}(|x-u|)^2\,du + \mathcal{O}\Big(D_n\lambda_n^{-\frac{1}{2\nu+1}}\Big)\Big]\\ =& \frac{\sigma^2}{n}\Big[tobecalculated + \mathcal{O}\Big(D_n\lambda_n^{-\frac{1}{2\nu+1}}\Big)\Big]\end{aligned}$$

Proof: See Section 4.3.

Theorem 3.6 (MSE)**.** *Assume a Matérn smoothness parameter satisfying* $\nu+1/2 = m/\ell > 1$ *with* $m, \ell \in \{1, 2, \dots\}$, $D_n \to 0$ *as* $n \to \infty$ *and that*

$$\lambda_n \sim D_n^{\frac{2\nu+1}{2\nu+2}} \log n \text{ as } n \to \infty. \tag{8}$$

The mean squared error, where the expectation is taken over the distribution of the errors and that of g*, satisfies as* $\lambda_n \to 0$

$$\begin{aligned}\mathcal{E}_{g,\boldsymbol{\epsilon}}\big[\big(\hat{g}(x)-g(x)\big)^2\big] &= 1 - \int_{\mathbb{R}} G_{\lambda_n}(|u|)k(|u|)\,du + \mathcal{O}\Big(D_n\lambda_n^{-\frac{1}{2\nu+1}}\Big)\\ &= \lambda_n^{\frac{2\nu}{2\nu+1}} \int_{\mathbb{R}} \Big[1 + \frac{1}{\alpha^{2\nu}c}\big(\lambda_n^{\frac{1}{\nu+1/2}}\alpha^2+\omega^2\big)\Big]^{-1} d\omega + \mathcal{O}\Big(D_n\lambda_n^{-\frac{1}{2\nu+1}}\Big)\end{aligned}$$

Proof: See Section 4.3.

4 Technical Part

4.1 Proof of the Exponential Envelope Condition

We use techniques from complex analysis to prove the Exponential envelope condition for G_λ, an introduction to those techniques can be found in **?**.

Proof of Theorem 3.1:

We start by considering $\nu+1/2=m\in\mathbb{N}$ and look at the sum in Equation (11) of Lemma A.1 more closely. For $r=|x-x'|$, $f(\omega)=\big[1+\lambda/(\alpha^{2m-1}c)(\alpha^2+\omega^2)^m\big]^{-1}$ and $c=\Gamma(\nu+1/2)\pi^{-1/2}/\Gamma(\nu)$ we have

$$
\begin{aligned}
G_\lambda(x,x') &= \int_{-\infty}^{\infty} \mathrm{e}^{i\omega r} f(\omega)\,d\omega \\
&= \frac{i\pi}{m}\frac{\alpha^{2m-1}c}{\lambda}\sum_{k=0}^{m-1}\frac{\exp\big(-\lambda^{-\frac{1}{2m}}\tilde{R}_k r\sin\theta_k\big)\ \exp(iR_k r\cos\theta_k)}{\Big(\frac{\lambda}{\alpha^{2m-1}c}\Big)^{-\frac{m-1}{m}}\exp\Big(\frac{i(\pi+2k\pi)(m-1)}{m}\Big)\lambda^{-\frac{1}{2m}}\tilde{R}_k\exp(i\theta_k)} \\
&= \lambda^{-\frac{1}{2m}}\frac{i\pi\alpha^{2-\frac{1}{m}}c^{\frac{1}{m}}}{m}\sum_{k=0}^{m-1}\frac{\exp\big(-\lambda^{-\frac{1}{2m}}\tilde{R}_k r\sin\theta_k\big)\ \exp(iR_k r\cos\theta_k)}{\exp\Big(\frac{i(\pi+2k\pi)(m-1)}{m}\Big)\ \tilde{R}_k\exp(i\theta_k)}
\end{aligned}
$$

with $\tilde{R}_k=\big\{\alpha^{4-\frac{2}{m}}c^{\frac{2}{m}}-\lambda^{\frac{1}{m}}\alpha^{4-\frac{1}{m}}c^{\frac{1}{m}}\cos\big[(\pi+2k\pi)/m\big]+\lambda^{\frac{2}{m}}\alpha^4\big\}^{1/4}\sim(\alpha^{2m-1}c)^{\frac{1}{2m}}$ as $\lambda\to 0$. We derive

$$
\begin{aligned}
\big|G_\lambda(|x-x'|)\big| &\sim \Bigg|\lambda^{-\frac{1}{2m}}\frac{i\pi(\alpha^{2m-1}c)^{\frac{1}{2m}}}{m}\sum_{k=0}^{m-1}\frac{\exp\big[-\lambda^{-\frac{1}{2m}}(\alpha^{2m-1}c)^{\frac{1}{2m}}\sin\big(\frac{\pi+2k\pi}{2m}\big)r\big]\ \exp(iR_k r\cos\theta_k)}{(\alpha^{2m-1}c)^{\frac{1}{2m}}\exp\Big(\frac{i(\pi+2k\pi)(m-1)}{m}\Big)\ \exp(i\theta_k)}\Bigg| \\
&\le \lambda^{-\frac{1}{2m}}\frac{\pi(\alpha^{2m-1}c)^{\frac{1}{2m}}}{m}\sum_{k=0}^{m-1}\exp\big[-\lambda^{-\frac{1}{2m}}(\alpha^{2m-1}c)^{\frac{1}{2m}}\sin\big(\frac{\pi+2k\pi}{2m}\big)r\big] \\
&\le \lambda^{-\frac{1}{2m}}\pi(\alpha^{2m-1}c)^{\frac{1}{2m}}\exp\big[-\lambda^{-\frac{1}{2m}}(\alpha^{2m-1}c)^{\frac{1}{2m}}\sin\big(\pi/2m\big)r\big],
\end{aligned}
$$

as $\lambda\to 0$. Therefore, choosing $C_1=\pi(\alpha^{2m-1}c)^{1/(2m)}$ and $A=(\alpha^{2m-1}c)^{1/(2m)}\sin\big(\pi/2m\big)$ we obtain an inequality of the type of Equation (6) for $k=0$.

For the derivatives of G_λ we use Equations (12) and (13). We observe under a slight abuse of notation that for $r=|x-x'|$ $\partial/\partial r G_\lambda(x,x')=\int_{-\infty}^{\infty}\exp(i\omega r)i\omega f(\omega)\,d\omega$ and $\partial^2/\partial r^2 G_\lambda(x,x')=\int_{-\infty}^{\infty}\exp(i\omega r)(i\omega)^2 f(\omega)\,d\omega$, using the same type of arguments we obtain $C_2=\pi(\alpha^{2m-1}c)^{2/(2m)}$ for the first and $C_3=\pi(\alpha^{2m-1}c)^{3/(2m)}$ for the second derivative with the same constant A in the exponential. Setting $C=\min(C_1,C_2,C_3)$ proves the claim for $\nu+1/2=m\in\mathbb{N}$, where $m>1/2$ is required for the function itself and its first derivative and $m>1$ for the second

derivative, see Lemma A.1.
For Matérn smoothness parameter $\nu + 1/2 = m/\ell$ with $m, \ell \in \{1, 2, \dots\}$ the proof of the Exponential Envelope Condition is analoguous once an analoguous result to Lemma A.1 is established rigorously. □

4.2 Proof of the Approximation Theorem 3.4

Analogous to ?, our main tool to prove that G_λ is a good approximation for the weighting function w is a series expansion of w in terms of G_λ, which will be used below to prove the Approximation Theorem 3.4.

Define the integral operator $\mathcal{R}_{n\lambda} : \mathcal{H} \to \mathcal{H}$ as

$$\mathcal{R}_{n\lambda}(h)(x) = \int_0^1 h(t)G_\lambda(x,t)d\big[t - F_n(t)\big] + \int_{\mathbb{R}\setminus[0,1]} h(t)G_\lambda(x,t)dt$$

and take $\mathcal{R}^l_{n\lambda}$ as the lth power of this operator.

Lemma 4.1. *If* $\big|\mathcal{R}^l_{n\lambda}[G_\lambda(\cdot, x_j)](x)\big| < q^l$ *for some* $q < 1$*, then*

$$w(x, x_j) = \sum_{k=0}^{\infty} \mathcal{R}^k_{n\lambda}[G_\lambda(\cdot, x_j)](x). \tag{9}$$

Proof:
From Equation (4) we deduce that w satisfies the equation

$$w(x, x_j) - G_\lambda(x, x_j) = \mathcal{R}^k_{n\lambda}[w(\cdot, x_j)](x). \tag{10}$$

We define $\psi_j(x) = \sum_{k=0}^{\infty} \mathcal{R}^k_{n\lambda}[G_\lambda(\cdot, x_j)](x)$, since by the assumption of the lemma the series converges uniformly, ψ_j is a well-defined function in $\mathcal{H}$ satisfying $\psi_j(x) = G_\lambda(x, x_j) + \mathcal{R}_{n\lambda}(\psi_j)(x)$. Therefore, ψ_j satisfies Equation (10) and if we can show that any function satisfying Equation (10) is necessarily equal to w we have proven the claim.
We denote $\mu_n(t) = t - F_n(t)$ if $t \in [0,1]$ and $\mu_n(t) = t$ for $t \in \mathbb{R}\setminus[0,1]$. Suppose $f \in \mathcal{H}$ satisfies $f(x) = G_\lambda(x, x_j) + \mathcal{R}_{n\lambda}(f)(x)$ then we have by the reproducing properties of G_λ that for any $h \in \mathcal{H}$

$$\begin{aligned}
\langle f, h\rangle_\lambda &= \int_0^1 f(t)h(t)d\,t + \lambda\langle f, h\rangle \\
&= \int_{-\infty}^{\infty} G_\lambda(t, x_j)h(t)dt + \lambda\big\langle G_\lambda(\cdot, x_j), h\big\rangle + \int_{-\infty}^{\infty} \mathcal{R}_{n\lambda}(f)(t)h(t)dt + \lambda\big\langle \mathcal{R}_{n\lambda}(f), h\big\rangle \\
&= h(x_j) + \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(s)G_\lambda(t,s)d\mu_n(s)h(t)dt + \lambda\Big\langle \int_{-\infty}^{\infty} f(s)G_\lambda(\cdot, s)d\mu_n(s), h\Big\rangle \\
&= h(x_j) + \int_{-\infty}^{\infty} f(s)\Big(\int_{-\infty}^{\infty} G_\lambda(t,s)h(t)dt + \lambda\big\langle G_\lambda(\cdot - s), h\big\rangle\Big)d\mu_n(s) \\
&= h(x_j) + \int_{-\infty}^{\infty} f(s)h(s)d\mu_n(s),
\end{aligned}$$

where the second last equality, i.e. exchanging integration and penalty inner product, is shown below. Simplifying and rearranging we have shown that $f = \psi_j$ satisfies

$$\int_0^1 \psi_j(t)h(t)dF_n(t) + \lambda\langle\psi_j, h\rangle = h(x_j),$$

i.e. ψ_j is the reproducing kernel with respect to the inner product $\langle\cdot,\cdot\rangle_\lambda^n$ and hence is equal to $w(\cdot - x_j)$. Using the convolution and modulation properties of the Fourier transform and the symmetry of G_λ we show the remaining property as follows

$$\begin{aligned}\Big\langle \int_{-\infty}^{\infty} f(u)G_\lambda(\cdot,s)ds, h\Big\rangle =& \langle f * G_\lambda, h\rangle = 2\pi\int_{-\infty}^{\infty}\frac{\mathcal{F}(f)(\omega)\mathcal{F}(G_\lambda^*)(\omega)\mathcal{F}(h)(\omega)}{\mathcal{F}(k)(\omega)}d\omega\\ =& \int_{-\infty}^{\infty} f(s)\int_{-\infty}^{\infty}\frac{\mathcal{F}(G_\lambda^*)(\omega)\mathcal{F}(h)(\omega)}{\mathcal{F}(k)(\omega)}e^{-i\omega s}d\omega\, ds\\ =& \int_{-\infty}^{\infty} f(s)\big\langle G_\lambda(\cdot,s), h\big\rangle ds.\end{aligned}$$

Lemma 4.2. *Let $A > \Delta > 0$. For Matérn range parameter $\alpha > 0$, smoothing parameter $\lambda > 0$ tending to zero, Matérn smoothness parameter $\nu + 1/2 = m/\ell > 1$ with $m, \ell \in \{1, 2, \dots\}$ and under Assumption 3.3 we have*

$$\Big|\frac{\partial^k}{\partial x^k}\mathcal{R}_{n\lambda}^l[G_\lambda(\cdot,x')](x)\Big| < \delta_n^l C\lambda^{-\frac{k+1}{2\nu+1}}\exp\big[-(A-\Delta)\lambda^{-\frac{k+1}{2\nu+1}}|x-x'|\big] + \textit{something},$$

where C, A are the constants of the Exponential Envelope Condition (Theorem 3.1).

Proof:

□

Proof of Theorem 3.4:
The proof is analogous to the proof of Theorem 2.1 in **?**.

□

4.3 Proof of the Asymptotics Theorems 3.5 and 3.6

Proof of Theorem 3.5:
The proof is analogous to the proof of Theorem 2.2 of **?**.

1. For the bias we use Equation (2) to obtain

$$\begin{aligned}\mathcal{E}\big[\hat{g}(x) - g(x)|g\big] =& \int_{\mathbb{R}} G_{\lambda_n}(x,t)g(t)dt - g(x) + \int_0^1 \big[w(x,t) - G_{\lambda_n}(x,t)\big]g(t)dt\\ &- \int_{\mathbb{R}\setminus[0,1]} G_{\lambda_n}(x,t)g(t)dt + \int_0^1 w(x,t)g(t)\, d[F_n(t) - t]\end{aligned}$$

We refer to the four terms on the right hand side as $b_{\lambda_n}(x)$, R_1, R_2 and R_3 respectively.

Bound R_1, R_2 and R_3, but need the extra term in the approximation theorem first

The calculation of the leading term is different than in **?** since we need to use the reproducing property of G_{λ_n} in order to deal with the term $g(x)$. By stationarity we can consider $x = 0$ without losing generality. We have

$$\begin{aligned}
b_\lambda(0) =& \int_{\mathbb{R}} G_{\lambda_n}(0,t)g(t)dt - g(x) \\
=& \int_{\mathbb{R}} G_{\lambda_n}(0,t)g(t)dt - \int_{\mathbb{R}} G_{\lambda_n}(0,t)g(t)dt - \langle G_{\lambda_n}(0,\cdot), g\rangle_{\lambda_n} \\
=& -\lambda_n \int_{\mathbb{R}} \frac{\mathcal{F}(G_{\lambda_n})(\omega)\mathcal{F}(g)(\omega)}{\mathcal{F}(k)(\omega)} d\omega = -\lambda_n \int_{\mathbb{R}} \frac{(\alpha^2+\omega^2)^{\nu+1/2}}{\alpha^{2\nu}c + \lambda_n(\alpha^2+\omega^2)^{\nu+1/2}} g(\omega)d\omega \\
=&
\end{aligned}$$

2. For the variance Equation (2) leads to

$$\begin{aligned}
\mathrm{Var}\big[\hat{g}(x)|g\big] = \frac{\sigma^2}{n}\Big\{ &\int_{\mathbb{R}} G_{\lambda_n}(x,t)^2 dt + \int_0^1 \big[w(x,t)^2 - G_{\lambda_n}(x,t)^2\big]dt \\
&- \int_{\mathbb{R}\setminus[0,1]} G_{\lambda_n}(x,t)^2 dt + \int_0^1 w(x,t)^2 d[F_n(t) - t]\Big\}
\end{aligned}$$

We refer to the four terms on the right hand side as $v_{\lambda_n}(x)$, R_4, R_5 and R_6 respectively.

Bound R_4, R_5 and R_6, but need the extra term in the approximation theorem first

By Parseval's theorem the leading term is

$$\begin{aligned}
v_{\lambda_n}(0) =& \int_{\mathbb{R}} G_{\lambda_n}(0,t)^2 dt = \int_{\mathbb{R}} \mathcal{F}(G_{\lambda_n})(\omega)^2 \mathrm{d}\omega = \int_{\mathbb{R}} \Big[1 + \frac{\lambda_n}{\alpha^{2\nu}c}(\alpha^2+\omega^2)^{\nu+1/2}\Big]^{-2} d\omega \\
=& \lambda_n^{-\frac{1}{2\nu+1}} \int_{\mathbb{R}} \Big[1 + \frac{1}{\alpha^{2\nu}c}(\lambda_n^{\frac{1}{\nu+1/2}}\alpha^2 + \omega^2)^{\nu+1/2}\Big]^{-2} d\omega \\
=&.
\end{aligned}$$

Proof of Theorem 3.6:
We have that $\mathcal{E}_{g,\boldsymbol{\epsilon}}\big[\big(\hat{g}(x) - g(x)\big)^2\big] = k(x,x) - \frac{1}{n}\sum_{i=1}^n w(x,x_i)k(x,x_i)$ and derive

$$\begin{aligned}
\mathcal{E}_{g,\boldsymbol{\epsilon}}\big[\big(\hat{g}(0) - g(0)\big)^2\big] = 1 &- \int_{\mathbb{R}} G_{\lambda_n}(0,t)k(0,t)dt - \int_{\mathbb{R}\setminus[0,1]} G_{\lambda_n}(0,t)k(0,t)dt \\
&- \int_0^1 \big[w(0,t) - G_{\lambda_n}(0,t)\big]k(0,t)dt - \int_0^1 w(0,t)k(0,t)d[F_n(t) - t]
\end{aligned}$$

We refer to the four terms on the right hand side as $m_{\lambda_n}(x)$, R_7, R_8 and R_9 respectively. Bound R_7, R_8 and R_9, but need the extra term in the approximation theorem first
The leading term is simplified using Parseval's Theorem, the fact that $\mathcal{F}(k)$ integrates to one and a substitution $u = \lambda^{1/(2\nu+1)}\omega$. □

5 Extensions/Discussion

- Nearest neighbor Kriging vs. ordinary Kriging
- Kernel estimators

A Complex Analysis

Lemma A.1. *Let $f(z)$ for $z \in \mathbb{C}$ be the analytic continuation in the complex plane of the function $f(\omega) = \mathcal{F}(G_\lambda)(\omega_1, \omega_2) = \big[1 + \lambda/(\alpha^{2m-1}c)(\alpha^2 + \omega^2)^m\big]^{-1}$, where $\alpha > 0$ is the range parameter of the Matérn covariance function, $m = \nu + 1/2 \in \{1, 2, 3, \dots\}$ with ν the smoothness parameter of the Matérn covariance function, $c = \Gamma(\nu + 1/2)\pi^{-1/2}/\Gamma(\nu)$ and $\lambda > 0$ is the smoothing parameter and $\omega = |\omega_1 - \omega_2|$.*

1. *The function f has $2m$ poles $z_0, \dots, z_{2m-1}$ given in polar coordinates as*

$$R_k = \lambda^{-\frac{1}{2m}}\Big\{\alpha^{4-\frac{2}{m}}c^{\frac{2}{m}} - \lambda^{\frac{1}{m}}\alpha^{4-\frac{1}{m}}c^{\frac{1}{m}}\cos\big[(\pi + 2k\pi)/m\big] + \lambda^{\frac{2}{m}}\alpha^4\Big\}^{1/4}, \quad k = 0, \dots, m-1,$$
$$R_k = R_{k-m}, \quad k = m, \dots, 2m-1$$
$$\theta_k = \frac{1}{2}\arctan\Big(\frac{\sin\big[(\pi + 2k\pi)/m\big]}{\cos\big[(\pi + 2k\pi)/m\big] - (\lambda\alpha/c)^{1/m}}\Big) \quad k = 0, \dots, m-1,$$
$$\theta_k = \theta_{k-m} + \pi, \quad k = m, \dots, 2m-1,$$

   where the poles $z_0, \dots, z_{m-1}$ are in the upper half plane and the poles $z_m, \dots, z_{2m-1}$ in the lower half plane.

2. *The residues of the functions $e^{izr}f(z)$, $e^{izr}izf(z)$, $e^{izr}(iz)^2f(z)$ at the poles z_k are given by*

$$Res\big(e^{izr}f(z), z_k\big) = \frac{e^{iz_kr}}{2m\lambda/(\alpha^{2m-1}c)(\alpha^2 + z_k^2)^{m-1}z_k}$$
$$Res\big(e^{izr}izf(z), z_k\big) = \frac{i\,e^{iz_kr}}{2m\lambda/(\alpha^{2m-1}c)(\alpha^2 + z_k^2)^{m-1}}$$
$$Res\big(e^{izr}(iz)^2f(z), z_k\big) = -\frac{z_k\,e^{iz_kr}}{2m\lambda/(\alpha^{2m-1}c)(\alpha^2 + z_k^2)^{m-1}}.$$

3. *On the real line we denote $z = \omega$ and derive for $r = |x - x'|$*

$$G_\lambda(x,x') = \int_{-\infty}^{\infty} e^{i\omega r} f(\omega)\, d\omega \quad = \quad \frac{\pi i \alpha^{2m-1} c}{m\lambda} \sum_{k=0}^{m-1} \frac{e^{iz_k r}}{(\alpha^2 + z_k^2)^{m-1} z_k}, \quad \textit{if } m > 1/2 \tag{11}$$

$$\frac{\partial}{\partial r} G_\lambda(x,x') = \int_{-\infty}^{\infty} e^{i\omega r} i\omega f(\omega)\, d\omega \quad = -\frac{\pi \alpha^{2m-1} c}{m\lambda} \sum_{k=0}^{m-1} \frac{e^{iz_k r}}{(\alpha^2 + z_k^2)^{m-1}}, \quad \textit{if } m > 1/2 \tag{12}$$

$$\frac{\partial^2}{\partial r^2} G_\lambda(x,x') = \int_{-\infty}^{\infty} e^{i\omega r} (i\omega)^2 f(\omega)\, d\omega = -\frac{\pi i \alpha^{2m-1} c}{m\lambda} \sum_{k=0}^{m-1} \frac{z_k e^{iz_k r}}{(\alpha^2 + z_k^2)^{m-1}}, \quad \textit{if } m > 1 \tag{13}$$

Proof:

1. The poles are obtained by straightforward calculations using that -1 can be represented as $\exp(i\pi + 2ki\pi)$ for $k \in \mathbb{Z}$. After determining the polar coordinates for z_k^2, $k = 0, \ldots, m-1$ the step to z_k is by taking the square root of the radius and halfing the angles (i.e. obtaining roots in the upper half plane). The remaining roots reflect that $\pm z_k$ lead to z_k^2 and are obtained by adding π to the angles for $k = 0, \ldots, m-1$.

2. The residues are calculated using Formula (4.1.10) of **?**.

3. The integrals are calculated using the Cauchy Residue Theorem (Theorem 4.1.1 of **?**) for the simple closed contour C composed by the section of the real line from $-b$ to b and a half circle of radius b, denoted by C_b, in the upper part of the complex plane connecting the endpoints of this section. We choose $b > R_0$ (i.e. the largest radius) then the contour C contains all singularities of f in the upper half plane. The theorem states that

$$\oint_C \exp(izr) f(z)\, dz = 2\pi i \sum_{k=0}^{m-1} \operatorname{Res}\big(\mathrm{e}^{izr} f(z), z_k\big).$$

   On the other hand, the contour integral can be decomposed into an integral over the real line and one over C_b. For $r > 0$ Jordan's Lemma (Lemma 4.2.2 of **?**) guarantees that the integral over C_b tends to zero for $b \to \infty$ since $f(z)$ tends to zero uniformly on C_b as $b \to \infty$. For $r = 0$ (Theorem 4.2.1 of **?**) implies that the integral over C_b tends to zero for $b \to \infty$ since $m > 1/2$. Therefore, letting b tend to infinity results in the desired expression of Equation (11) for the integral over the entire real line.
   Equations (12) and (13) are shown analogously while bearing in mind that for Jordan's Lemma to apply the functions $izf(z)$ and $(iz)^2 f(z)$ have to tend to zero uniformly on C_b, which is true for $m > 1/2$ and $m > 1$ respectively. □

Remark A.2. *For the case of rational exponents $\nu + 1/2 = m/\ell$ with $m, \ell \in 1, 2, \ldots$ the situation is more complicated since the function $f(z)$ is multi-valued as a complex function. The Cauchy Residue Theorem and Jordan's Lemma are still the key to calculate integrals as in Equation (11) but the contour over which to integrate is more complicated since the branch points of the multi-valued function need to be taken into consideration and one obtains only bounds not exact values. A very similar approach than the one presented in* **?** *leads to the following approximation*

$$\int_{-\infty}^{\infty} e^{i\omega t} f(\omega)\, d\omega \approx \frac{\pi i \ell \alpha^{2m-1} c}{m\lambda} \sum_{k=0}^{m-1} \frac{e^{i z_k r}}{(\alpha^2 + z_k^2)^{\frac{m-\ell}{\ell}} z_k}$$

Similar arguments lead again to corresponding results for Equations (12) and (13). A full proof of these results is beyond the scope of this paper.