Parameter estimation and uncertainty quantification for Gaussian process interpolation

Toni Karvonen

Department of Mathematics and Statistics
University of Helsinki, Finland

toni.karvonen@helsinki.fi — https://tskarvone.github.io/

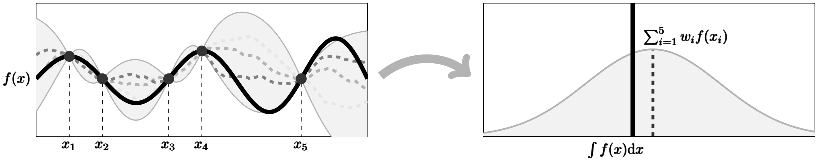
UNIVERSITY OF HELSINKI
FACULTY OF SCIENCE

MASCOT-NUM 2023
Le Croisic, France

6 April 2023

Motivation: Probabilistic numerical integration

Approximate $\int_\Omega f \, \mathrm{d}P$ by modelling f with a Gaussian process.



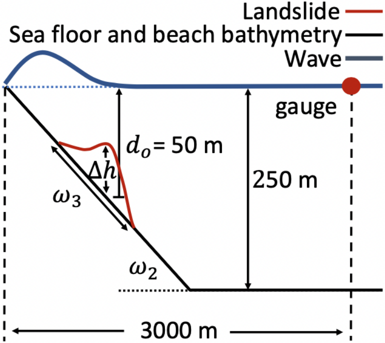
O'Hagan (1991). Bayes–Hermite quadrature. *Journal of Statistical Planning and Inference*, 29(3):245–260.

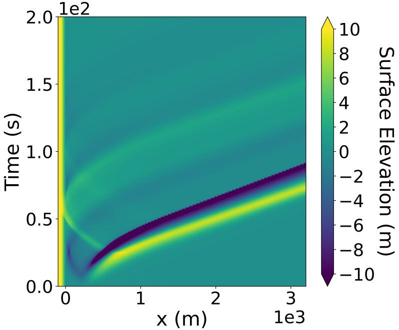
Cockayne, Oates, Sullivan & Girolami (2019). Bayesian probabilistic numerical methods. *SIAM Review*, 61(4):756–789.

Hennig, Osborne & Kersting (2022). *Probabilistic Numerics: Computation as Machine Learning*. Cambridge University Press.

Tsunami model example

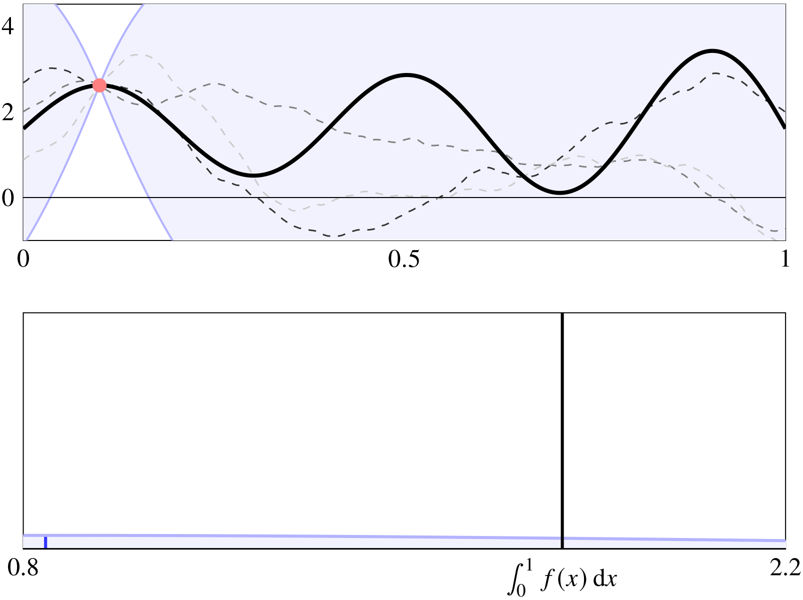
$$\int_{0.125}^{0.5} \int_{5}^{15} \int_{100}^{200} f(x_1, x_2, x_3)\, dx_1\, dx_2\, dx_3 \qquad \textbf{(1)}$$



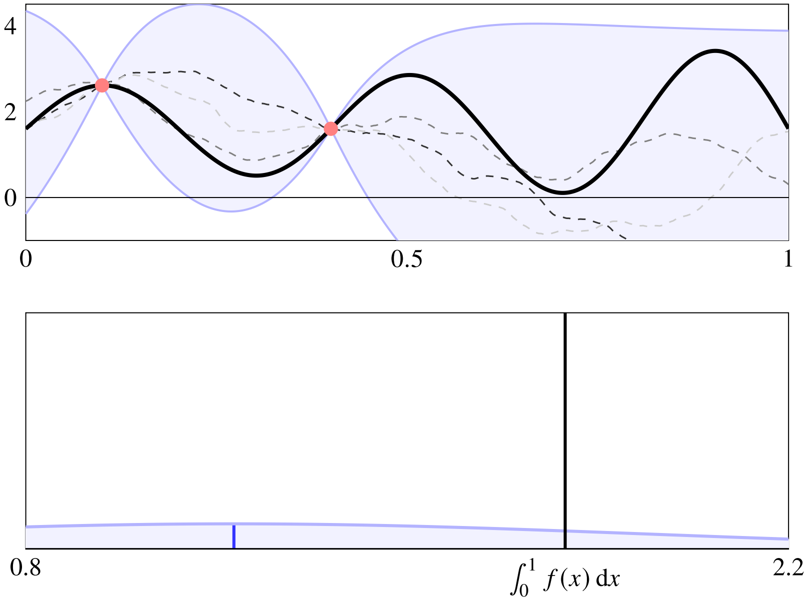


Li, Giles, Karvonen, Guillas & Briol (2023). Multilevel Bayesian quadrature. *AISTATS*. To appear.

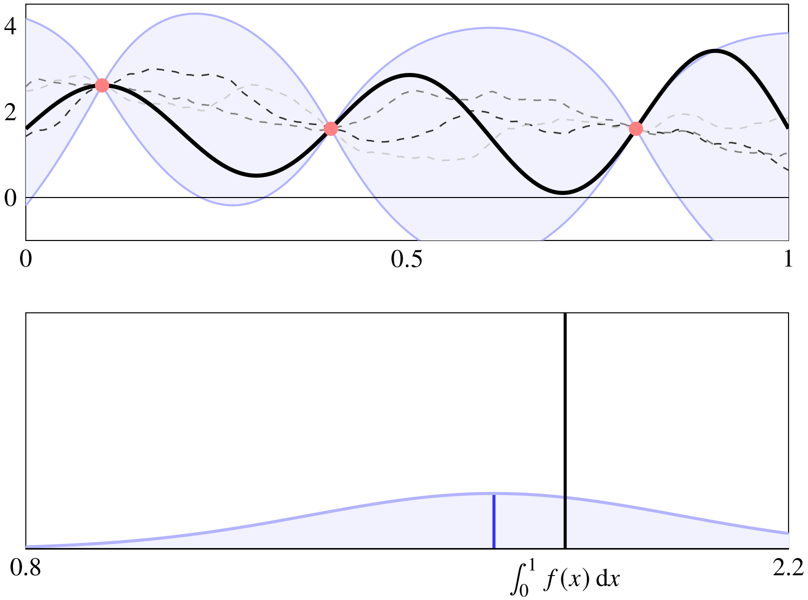
Example



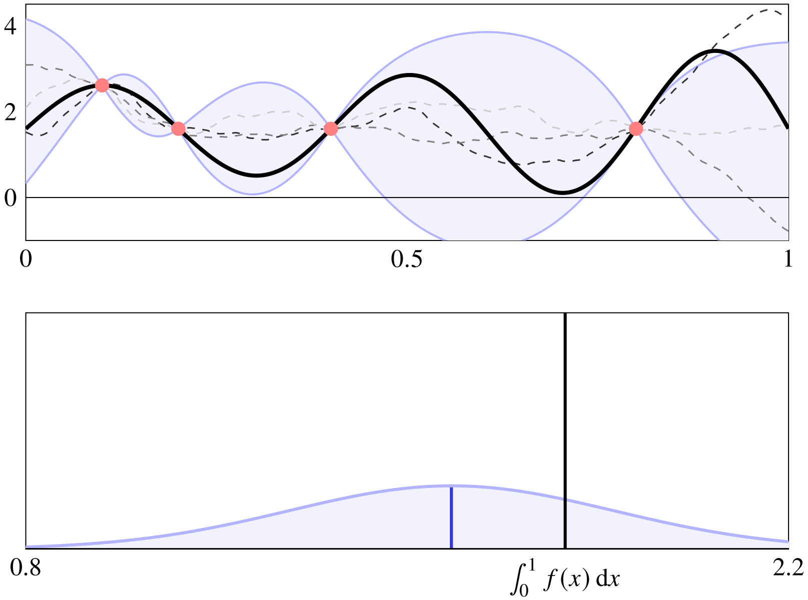
Example



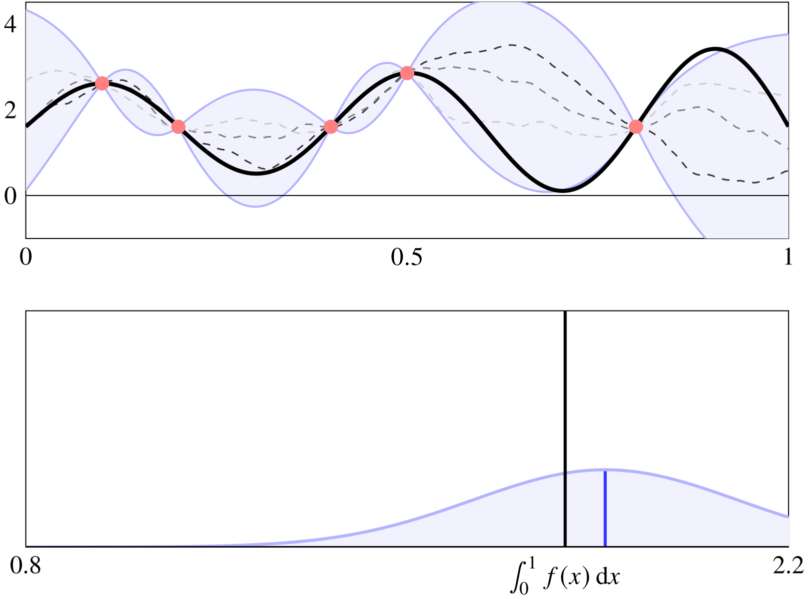
Example



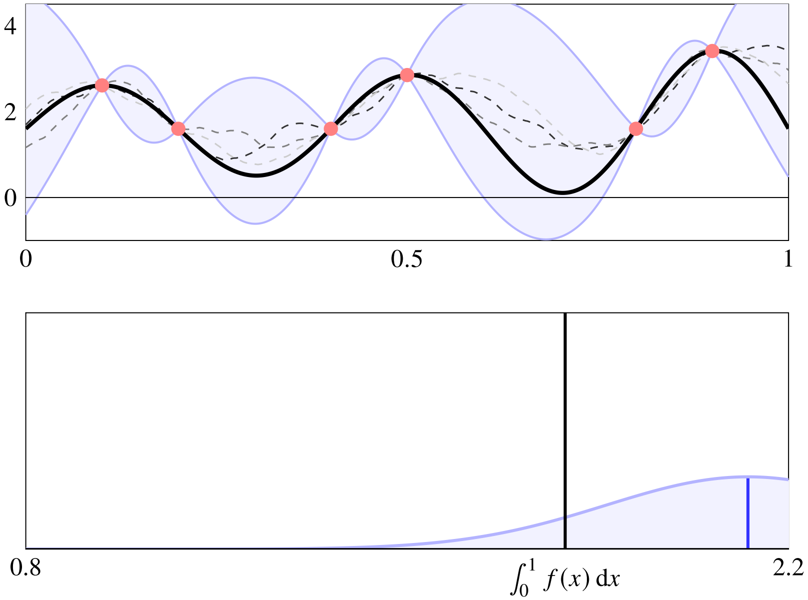
Example



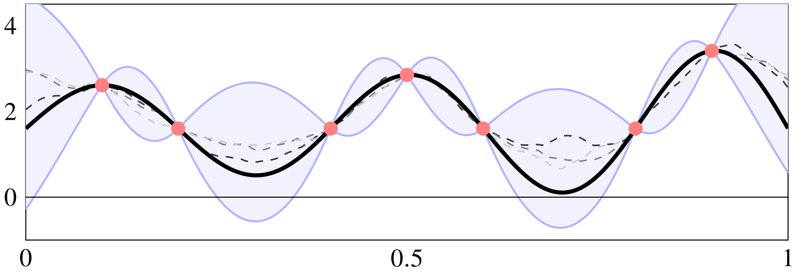
Example

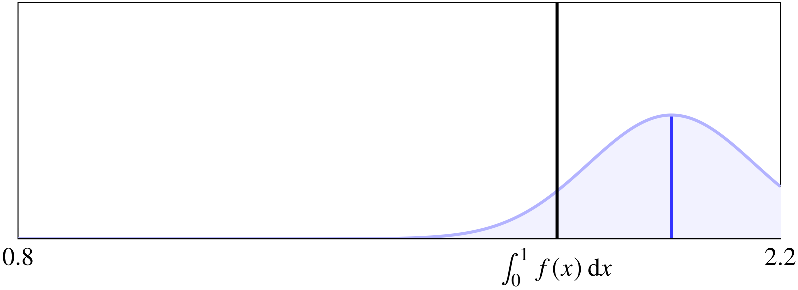


Example

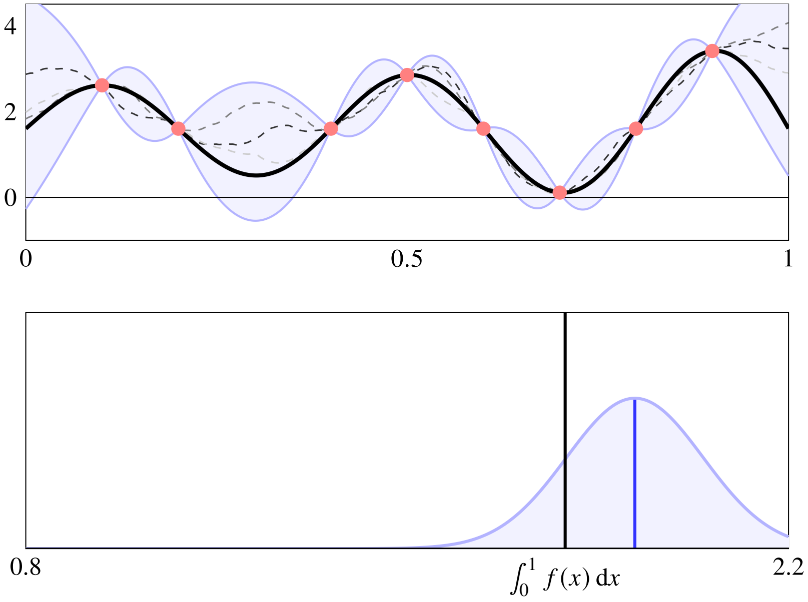


Example

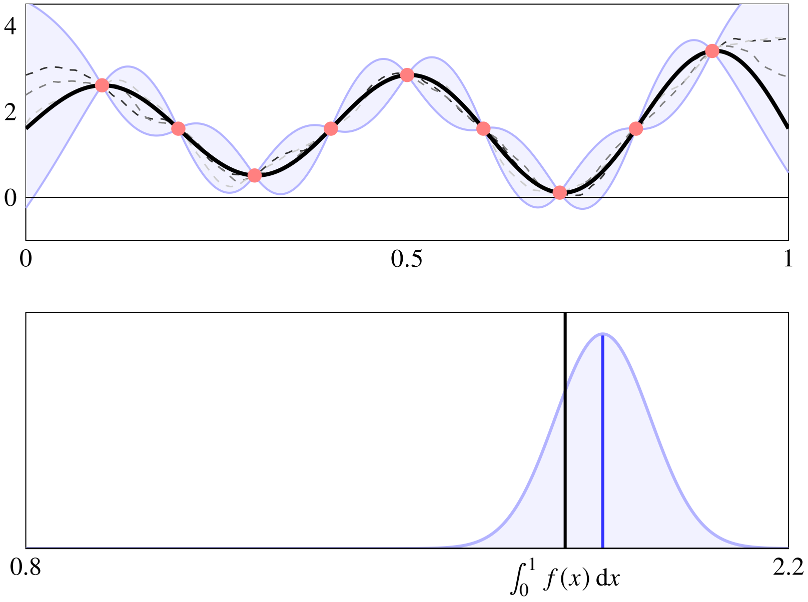




Example



Example



Setting and objectives

1. Let $f : \Omega \to \mathbb{R}$ be a data-generating function on a sufficiently regular and bounded set $\Omega \subset \mathbb{R}^d$.
2. Model f as a Gaussian process $f_{\mathrm{GP}} \sim \mathrm{GP}(0, K_\theta)$.
3. Obtain *noiseless data* $\mathcal{D}_n = \{(\mathbf{x}_1, f(\mathbf{x}_1)), \ldots, (\mathbf{x}_n, f(\mathbf{x}_n))\}$ at some pairwise distinct points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \Omega$.
4. Estimate kernel parameters θ from the data.
5. Compute the posterior $f_{\mathrm{GP}} \mid \mathcal{D}_n$.

(a) How do parameter estimates behave as $n \to \infty$?

(b) Is posterior standard deviation commensurate to the true approximation error as $n \to \infty$?

Frequentist coverage in Bayesian nonparametrics

1. **Bull (2012)**. Honest adaptive confidence bands and self-similar functions. *Electronic Journal of Statistics*, 6:1490–1516.
2. **Szabó, van der Vaart & van Zanten (2015)**. Frequentist coverage of adaptive nonparametric Bayesian credible sets. *The Annals of Statistics*, 43(4):1391–1428.
3. **Hadji & Szabó (2021)**. Can we trust Bayesian uncertainty quantification from Gaussian process priors with squared exponential covariance kernel? *SIAM/ASA Journal on Uncertainty Quantification*, 9(1):185–230.

- Let $f = \sum_{i=1}^{\infty} f_i \psi_i$ for some basis $(\psi_i)_{i=1}^{\infty}$ of H and $f_i = \langle f, \psi_i \rangle_H$.
- One observes the noisy sequence $X = (X_1, X_2, \ldots)$, where

$$X_i = f_i + \frac{1}{\sqrt{n}} Z_i \quad \text{for} \quad \text{i.i.d. } Z_i \sim N(0,1). \qquad \textbf{(2)}$$

- Place a Gaussian prior on the coefficients f_i of f.
- Study the coverage of credible sets as $n \to \infty$.

Our setting: **(a)** No noise **(b)** n = number of observations $\neq$ noise level

Gaussian processes

Model f as a zero-mean Gaussian process $f_{\mathrm{GP}} \sim \mathrm{GP}(0, K)$ with a positive-definite covariance kernel $K: \Omega \times \Omega \to \mathbb{R}$.

Covariance kernel

Kernel defines covariance structure: $\mathrm{Cov}[f_{\mathrm{GP}}(\mathbf{x}), f_{\mathrm{GP}}(\mathbf{y})] = K(\mathbf{x}, \mathbf{y})$. For example,

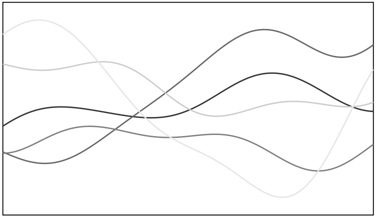
$$K_\theta(\mathbf{x}, \mathbf{y}) = K(\mathbf{x}, \mathbf{y}) = \sigma^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{y}\|^2}{2\lambda^2}\right) \quad \text{with} \quad \theta = \{\sigma, \lambda\}. \quad \textbf{(3)}$$

Under this model $[f_{\mathrm{GP}}(\mathbf{x}_1), \ldots, f_{\mathrm{GP}}(\mathbf{x}_n)] \in \mathbb{R}^n$ is a multivariate normal random variable with mean zero and covariance matrix $\mathbf{K}_n$:

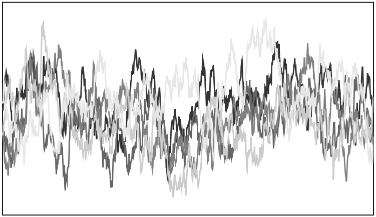
$$\begin{bmatrix} f_{\mathrm{GP}}(\mathbf{x}_1) \\ \vdots \\ f_{\mathrm{GP}}(\mathbf{x}_n) \end{bmatrix} \sim \mathrm{N}(\mathbf{0}_n, \mathbf{K}_n) = \mathrm{N}\left(\begin{bmatrix} 0 \\ \vdots \\ 0 \end{bmatrix}, \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}_1) & \cdots & K(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ K(\mathbf{x}_n, \mathbf{x}_1) & \cdots & K(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix} \right).$$

Gaussian process priors

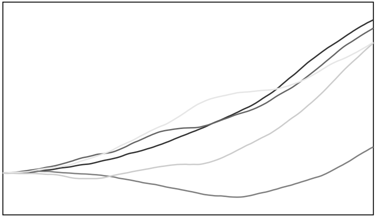
Gaussian: $K(x,y) = e^{-(x-y)^2/2}$



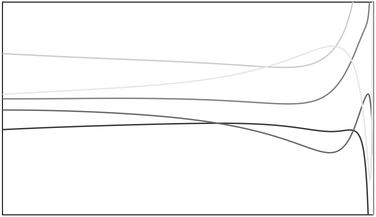
Matérn: $K(x,y) = e^{-|x-y|}$



BM: $K(x,y) = \frac{\min\{x,y\}^3}{3} + \frac{|x-y|\min\{x,y\}^2}{2}$



Hardy: $K(x,y) = \frac{1}{1-xy}$



Gaussian process interpolation

Recall that we have access to the *noiseless* data

$$\mathcal{D}_n = \left\{(\mathbf{x}_1, f(\mathbf{x}_1)), \ldots, (\mathbf{x}_n, f(\mathbf{x}_n))\right\} \qquad \textbf{(4)}$$

at some pairwise distinct points $\mathbf{x}_1, \ldots, \mathbf{x}_n \in \boldsymbol{\Omega} \subset \mathbb{R}^d$.

Conditional Gaussian process

The conditional process $f_{\mathrm{GP}} \mid \mathcal{D}_n$ is also a Gaussian process. Standard Gaussian conditioning formulae give

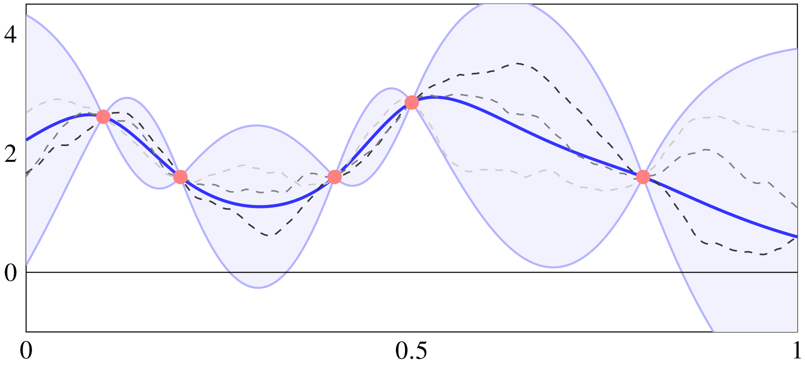
$$\mu_n(\mathbf{x}) = \mathbb{E}[f_{\mathrm{GP}}(\mathbf{x}) \mid \mathcal{D}_n] = \mathbf{k}_n(\mathbf{x})^\mathsf{T} \mathbf{K}_n^{-1} \mathbf{f}_n, \qquad \textbf{(5)}$$

$$\mathbb{V}_n(\mathbf{x}) = \mathbb{V}[f_{\mathrm{GP}}(\mathbf{x}) \mid \mathcal{D}_n] = K(\mathbf{x}, \mathbf{x}) - \mathbf{k}_n(\mathbf{x})^\mathsf{T} \mathbf{K}_n^{-1} \mathbf{k}_n(\mathbf{x}). \qquad \textbf{(6)}$$

Here

$$\mathbf{f}_n = \begin{bmatrix} f(\mathbf{x}_1) \\ \vdots \\ f(\mathbf{x}_n) \end{bmatrix}, \quad \mathbf{k}_n(\mathbf{x}) = \begin{bmatrix} K(\mathbf{x}, \mathbf{x}_1) \\ \vdots \\ K(\mathbf{x}, \mathbf{x}_n) \end{bmatrix}, \quad \mathbf{K}_n = \begin{bmatrix} K(\mathbf{x}_1, \mathbf{x}_1) & \cdots & K(\mathbf{x}_1, \mathbf{x}_n) \\ \vdots & \ddots & \vdots \\ K(\mathbf{x}_n, \mathbf{x}_1) & \cdots & K(\mathbf{x}_n, \mathbf{x}_n) \end{bmatrix}.$$

Example



Uncertainty quantification

- Suppose that the prior covariance is parametric: $f_{\mathrm{GP}} \sim \mathrm{GP}(0, K_\theta)$.
- The conditional process is $f_{\mathrm{GP}} \mid \mathcal{D}_N \sim \mathrm{GP}(\mu_{\theta,n}, \mathbb{V}_{\theta,n})$.
- For any $a \in (0,1)$ and every $\mathbf{x} \in \Omega$,

$$\mathbb{P}\left[|f_{\mathrm{GP}}(\mathbf{x}) - \mu_{\theta,n}(\mathbf{x})| \leq c(a) \sqrt{\mathbb{V}_{\theta,n}(\mathbf{x})} \;\middle|\; \mathcal{D}_N \right] = 1 - a. \tag{7}$$

- Plug in a some estimator $\theta_n = \theta(\mathcal{D}_n)$ of θ.

Objective

Understand the behaviour of **(i)** θ_n and **(ii)** the standard score

$$C_n(\mathbf{x}) = \frac{|f(\mathbf{x}) - \mu_{\theta_n,n}(\mathbf{x})|}{\sqrt{\mathbb{V}_{\theta_n,n}(\mathbf{x})}} \quad \text{as} \quad n \to \infty \tag{8}$$

for different **(a)** functions f, **(b)** kernel parametrisations $\{K_\theta\}_{\theta \in \Theta}$ and **(c)** parameter estimators.

Matérn class

We consider kernels of the isotropic Matérn class and the estimation of

scale σ and smoothness ν.

Matérn class

Matérn kernel of order (or smoothness) $\nu > 0$ is

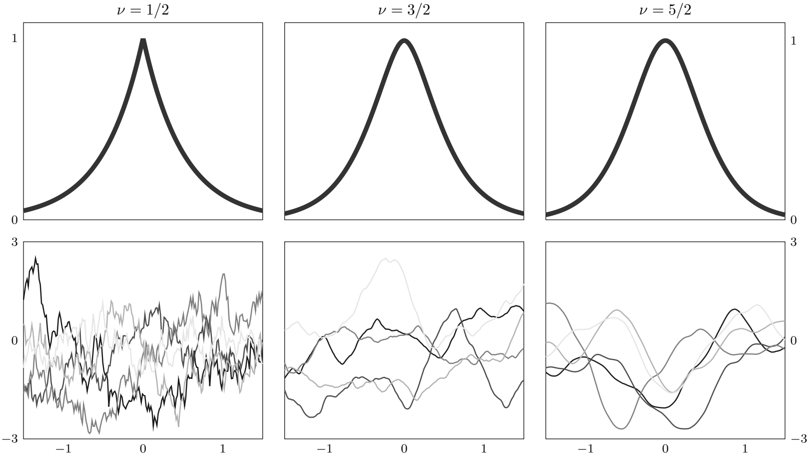
$$K(\mathbf{x},\mathbf{y}) = \Phi(\mathbf{x}-\mathbf{y}) \quad \text{for} \quad \Phi(\mathbf{z}) = \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\frac{\sqrt{2\nu}\,\|\mathbf{z}\|}{\lambda} \right)^{\nu} \mathcal{K}_\nu \left(\frac{\sqrt{2\nu}\,\|\mathbf{z}\|}{\lambda} \right).$$

Here $\mathcal{K}_\nu$ is the modified Bessel function of the second kind of order ν.

For example, $\nu = 1/2$ and $\nu = 3/2$ give

$$\Phi_{\nu=1/2}(\mathbf{z}) = \sigma^2 e^{-\frac{1}{\lambda}\|\mathbf{z}\|} \quad \text{and} \quad \Phi_{\nu=3/2}(\mathbf{z}) = \sigma^2 \left(1 + \frac{\sqrt{3}\,\|\mathbf{z}\|}{\lambda}\right) e^{-\frac{\sqrt{3}}{\lambda}\|\mathbf{z}\|}.$$

Examples of Matérns



Reproducing kernel Hilbert spaces

Reproducing kernel Hilbert space

For every positive-semidefinite kernel $K\colon \Omega \times \Omega \to \mathbb{R}$ there is a unique reproducing kernel Hilbert space (RKHS), $H(K)$:

1. Each element of $H(K)$ is a function $f\colon \Omega \to \mathbb{R}$.
2. The kernel has the reproducing property $\langle f, K(\cdot, \mathbf{x}) \rangle_{H(K)} = f(\mathbf{x})$.

Optimal approximation in RKHS

Let $B = \{ f \in H(K) \ : \ \|f\|_{H(K)} \leq 1 \}$. Posterior STD equals worst-case error:

$$\sqrt{\mathbb{V}_n(\mathbf{x})} = \sup_{f \in B} |f(\mathbf{x}) - \mu_n(\mathbf{x})| = \inf_{A_n^{\text{std}}} \sup_{f \in B} |f(\mathbf{x}) - (A_n^{\text{std}} f)(\mathbf{x})| \,. \tag{9}$$

The RKHS of a Matérn K_ν of order ν is a norm-equivalent to a Sobolev space of order $\nu + d/2$:

$$H(K_\nu) = H^{\nu + d/2}(\mathbb{R}^d) = \left\{ f \in L^2(\mathbb{R}^d) \ : \ \int_{\mathbb{R}^d} \left(1 + \|\boldsymbol{\omega}\|^2\right)^{\nu + d/2} |\widehat{f}(\boldsymbol{\omega})|^2 \, \mathrm{d}\boldsymbol{\omega} < \infty \right\}.$$

Sample path properties

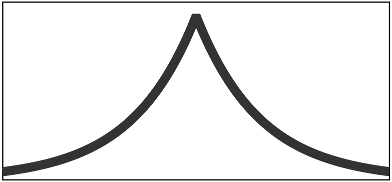
Driscoll's theorem for Matérns

If $f_{\mathrm{GP}} \sim \mathrm{GP}(0, K_\nu)$, then

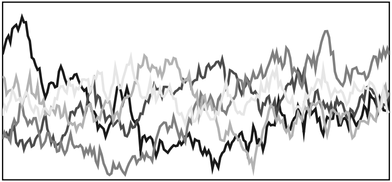
$$\mathbb{P}[f_{\mathrm{GP}} \in H^\alpha(\Omega)] = 1 \text{ if } \alpha < \nu \;\; \textbf{and} \;\; \mathbb{P}[f_{\mathrm{GP}} \in H^\alpha(\Omega)] = 0 \text{ if } \alpha \geq \nu.$$

$\implies$ Samples are "$d/2$ less smooth than the RKHS" as $H(K_\nu) = H^{\nu + d/2}(\Omega)$.
$\implies$ Samples are *not* in the $H(K_\nu) = H^{\nu + d/2}(\Omega)$ since $\alpha = \nu + d/2 \geq \nu$.

Kernel $(\nu = 1/2)$



Samples $(\nu = 1/2)$



Driscoll (1973); Lukić & Beder (2001); Scheuerer (2010); Steinwart (2019); Henderson (2022); Karvonen (2023).

Maximum likelihood estimation

Let $\theta \in \Theta$ be kernel parameter(s). The log-likelihood function is

$$L(\theta; \mathcal{D}_n) = -\frac{1}{2}\left[\mathbf{f}_n^\top \mathbf{K}_{\theta,n}^{-1}\mathbf{f}_n + \log\det \mathbf{K}_{\theta,n} + C\right], \tag{10}$$

where $\mathbf{f}_n = [f(\mathbf{x}_i)]_{i=1}^n$ and $\mathbf{K}_{\theta,n} = [K_\theta(\mathbf{x}_i, \mathbf{x}_j)]_{i,j=1}^n$.

Maximum likelihood estimation

The maximum likelihood estimate (MLE) of θ is

$$\theta_n = \theta(\mathcal{D}_n) = \arg\max_{\theta \in \Theta} L(\theta; \mathcal{D}_n). \tag{11}$$

We are interested in Matérns and the parameter estimates

$$\sigma_n = \sqrt{\frac{\mathbf{f}_n^\top \mathbf{K}_{\nu,n}^{-1}\mathbf{f}_n}{n}} \quad \text{and} \quad \nu_n = \arg\min_{\nu > 0} \left[\mathbf{f}_n^\top \mathbf{K}_{\nu,n}^{-1}\mathbf{f}_n + \log\det \mathbf{K}_{\nu,n}\right]. \tag{12}$$

An identification and a decomposition

Scale MLE

$$\sigma_n = \sqrt{\frac{\mathbf{f}_n^\top \mathbf{K}_{\nu,n}^{-1} \mathbf{f}_n}{n}} = \frac{1}{\sqrt{n}} \|\mu_{\nu,n}\|_{H(K_\nu)} \tag{13}$$

Smoothness MLE

$$\nu_n = \arg\min_{\nu>0} \left[\mathbf{f}_n^\top \mathbf{K}_{\nu,n}^{-1} \mathbf{f}_n + \log\det \mathbf{K}_{\nu,n} \right] \tag{14}$$

$$= \arg\min_{\nu>0} \sum_{i=1}^{n} \left[\frac{(f(\mathbf{x}_i) - \mu_{\nu,i-1}(\mathbf{x}_i))^2}{\mathbb{V}_{\nu,i-1}(\mathbf{x}_i)} + \log \mathbb{V}_{\nu,i-1}(\mathbf{x}_i) \right] \tag{15}$$

Self-similar Sobolev functions

Self-similar Sobolev functions

We say that $f : \Omega \to \mathbb{R}$ is ν_0-self-similar, written $f \in H^{\beta}_{\mathrm{ss}}(\Omega)$, if

(a) the support of f is contained in Ω and

there is $f_e \in L^2(\mathbb{R}^d)$ such that

(b) $f_e|_\Omega = f$;

(c) $\sup_{\boldsymbol{\omega} \in \mathbb{R}^d} \|\boldsymbol{\omega}\|^{2\beta+d} |\widehat{f_e}(\boldsymbol{\omega})|^2 < \infty$;

(d) $\int_{\|\boldsymbol{\omega}\| \geq R} |\widehat{f_e}(\boldsymbol{\omega})|^2 \, \mathrm{d}\boldsymbol{\omega} \geq C R^{-2\beta}$ for all $R \geq R_0$.

Example: Function s.t. $\mathrm{supp}(f) \subset \Omega$ and $|\widehat{f_e}(\boldsymbol{\omega})| \asymp \|\boldsymbol{\omega}\|^{-(\beta+d/2)}$.

Lemma

Let $f \in H^{\beta}_{\mathrm{ss}}(\Omega)$. Then $f \in H^{\alpha}(\Omega)$ if $\alpha < \beta$ and $f \notin H^{\alpha}(\Omega)$ if $\alpha > \beta$.

Scale estimation (ν and λ fixed)

- We assume that $\{\mathbf{x}_i\}_{i=1}^{\infty}$ are quasi-uniform on $\Omega \subset \mathbb{R}^d$.
- ν = Matérn smoothness — ν_0 = "true" smoothness ($\lfloor \nu_0 + d/2 \rfloor > d/2$)

Theorem 1

(1a) If $f \in H^{\nu_0 + d/2}(\Omega)$ for $\nu_0 \leq \nu$, then $\sigma_n \lesssim n^{(\nu - \nu_0)/d - 1/2}$.

(1b) If $f \in H_{\text{ss}}^{\nu_0 + d/2}(\Omega)$ for $\nu_0 \leq \nu$, then $\sigma_n \gtrsim n^{(\nu - \nu_0)/d - 1/2 - \varepsilon}$ for any $\varepsilon > 0$.

Theorem 2

(2a) If $f \in H_{\text{ss}}^{\nu_0 + d/2}(\Omega)$ for $\nu_0 \leq \nu$, then (for any $\varepsilon > 0$ and $\mathbf{x} \in \Omega$)

$$C_n(\mathbf{x}) = \frac{|f(\mathbf{x}) - \mu_{\nu,n}(\mathbf{x})|}{\sigma_n \sqrt{\mathbb{V}_{\nu,n}(\mathbf{x})}} \lesssim \frac{n^{-\nu_0/d}\sqrt{\log n}}{n^{(\nu-\nu_0)/d - 1/2 - \varepsilon} \times n^{-\nu/d}} = n^{1/2+\varepsilon}\sqrt{\log n}. \quad \textbf{(16)}$$

(2b) If $f \in H^{\nu + d/2}(\Omega) = H(K_\nu)$, then $C_n(\mathbf{x}) \lesssim n^{1/2}$ for every $\mathbf{x} \in \Omega$.

$\Longrightarrow$ Overconfidence can occur at most "slowly".

Karvonen, Wynne, Tronarp, Oates & Särkkä (2020). Maximum likelihood estimation and uncertainty quantification for Gaussian process approximation of deterministic functions. *SIAM/ASA J. Uncertainty Quantif.*, 8(3):926–958.

Smoothness estimation (σ and λ fixed)

Suppose that $\nu_n \in [\nu_{\min}, \nu_{\max}]$ for $\nu_{\min} > 0$ and $\nu_{\max} < \infty$.

Theorem 3

(3a) If $f \in H^{\nu_0+d/2}(\Omega)$, then $\liminf_{n\to\infty} \nu_n \geq \nu_0 + d/2$.

(3b) If $f \in H_{\mathrm{ss}}^{\nu_0+d/2}(\Omega)$, then $\lim_{n\to\infty} \nu_n = \nu_0 + d/2$.

(3c) If $\nu_n \in [\nu_{\min}, \infty)$ and $f \in \bigcap_{\nu_0>0} H^{\nu_0+d/2}(\Omega)$, then $\lim_{n\to\infty} \nu_n = \infty$.

Theorem 4

If $f \in H_{\mathrm{ss}}^{\nu_0+d/2}(\Omega)$, then (for any $\varepsilon > 0$ and $\mathbf{x} \in \Omega$)

$$C_n(\mathbf{x}) = \frac{|f(\mathbf{x}) - \mu_{\nu_n,n}(\mathbf{x})|}{\sqrt{\mathbb{V}_{\nu_n,n}(\mathbf{x})}} \lesssim \frac{n^{-\nu_0/d}\sqrt{\log n}}{n^{-(\nu_0+d/2)/d-\varepsilon}} = n^{1/2+\varepsilon}\sqrt{\log n}\,. \qquad \textbf{(17)}$$

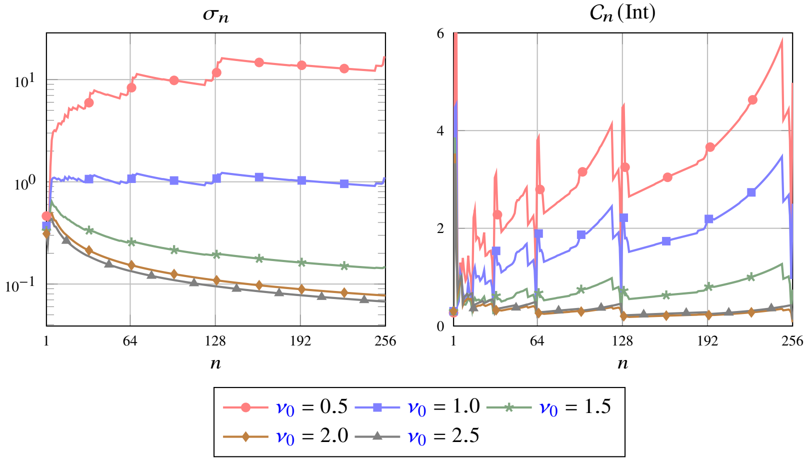
$\Longrightarrow$ Again, overconfidence can occur at most “slowly”.

Karvonen (2023). Asymptotic bounds for smoothness parameter estimates in Gaussian process interpolation. *arXiv*:2203.05400v3.

Also: Chen, Owhadi & Stuart (2021), *Math. Comput*; Petit (2023), *arXiv*:2209.07791v5.

Some numerical results for scale estimation

$$\nu = \frac{3}{2} \quad \text{and} \quad \text{“}f \in H_{\text{ss}}^{\nu_0+1/2}([0,1])\text{”} \quad \text{for} \quad \nu_0 = \frac{1}{2}, \frac{3}{2}, \ldots, \frac{5}{2}.$$



Lengthscale estimation — tangential yet interesting

Here we consider MLE of the lengthscale parameter λ:

$$K_\lambda(\mathbf{x}, \mathbf{y}) = K\left(\frac{\mathbf{x}}{\lambda}, \frac{\mathbf{y}}{\lambda}\right) = \Phi\left(\frac{\mathbf{x} - \mathbf{y}}{\lambda}\right). \tag{18}$$

Data are constant if $\mathbf{f}_n = [f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)] = [c, \ldots, c]$ for some $c \in \mathbb{R}$.

Theorem 5

Let $n \geq 2$ be **fixed** and K a Matérn kernel of any smoothness. Then

$$\lambda_n = \lambda(\mathcal{D}_n) = \infty \quad \text{and} \quad \lim_{\lambda \to \infty} L(\lambda; \mathcal{D}_n) = \infty \tag{19}$$

if and only if the data are constant. Moreover, for every $\mathbf{x} \in \mathbb{R}^d$,

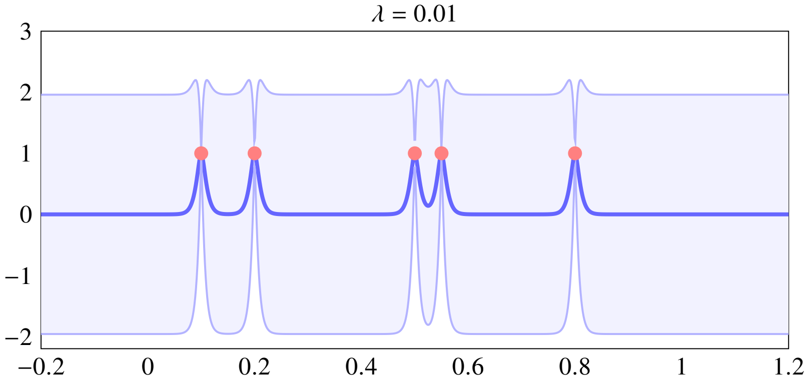
$$\lim_{\lambda \to \infty} \mu_{\lambda,n}(\mathbf{x}) = c \quad \text{and} \quad \lim_{\lambda \to \infty} \mathbb{V}_{\lambda,n}(\mathbf{x}) = 0. \tag{20}$$

$\Longrightarrow$ If the data are constant, the posterior becomes degenerate.

Karvonen & Oates (2023). Maximum likelihood estimation in Gaussian process regression is ill-posed. *Journal of Machine Learning Research*. To appear.

Lengthscale example

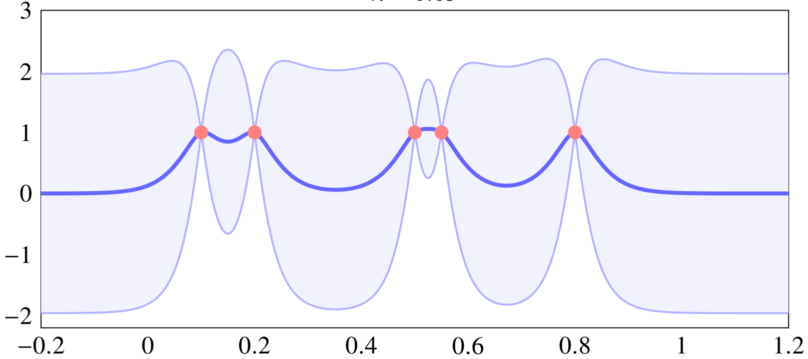
$$K_{\lambda}(x, y) = \left(1 + \frac{\sqrt{3}\,|x - y|}{\lambda}\right) \exp\left(-\frac{\sqrt{3}\,|x - y|}{\lambda}\right)$$



Lengthscale example

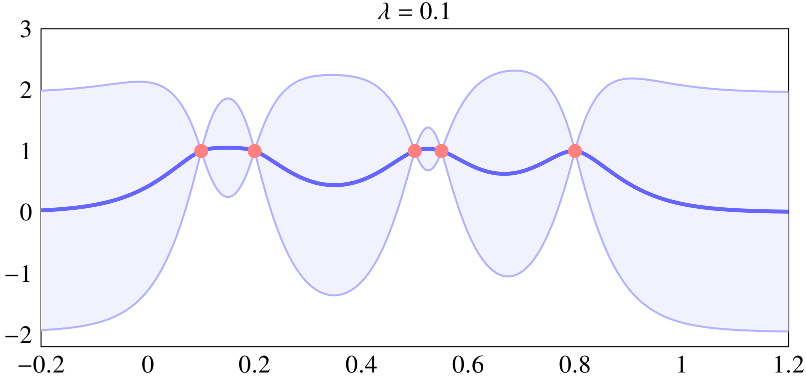
$$K_{\lambda}(x, y) = \left(1 + \frac{\sqrt{3}\,|x - y|}{\lambda}\right) \exp\left(-\frac{\sqrt{3}\,|x - y|}{\lambda}\right)$$

$\lambda = 0.05$



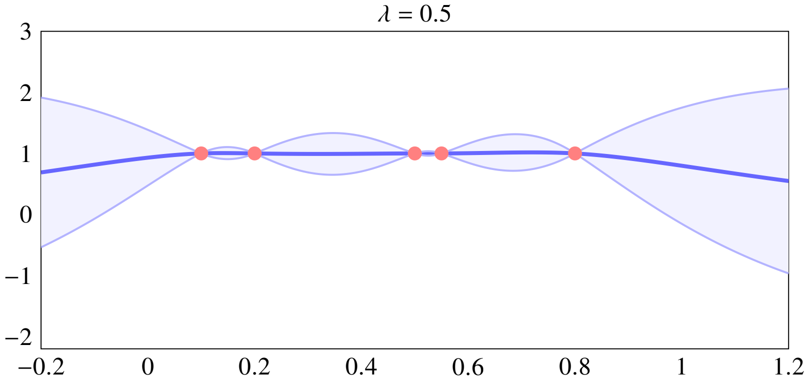
Lengthscale example

$$K_{\lambda}(x, y) = \left(1 + \frac{\sqrt{3}\,|x - y|}{\lambda}\right) \exp\left(-\frac{\sqrt{3}\,|x - y|}{\lambda}\right)$$



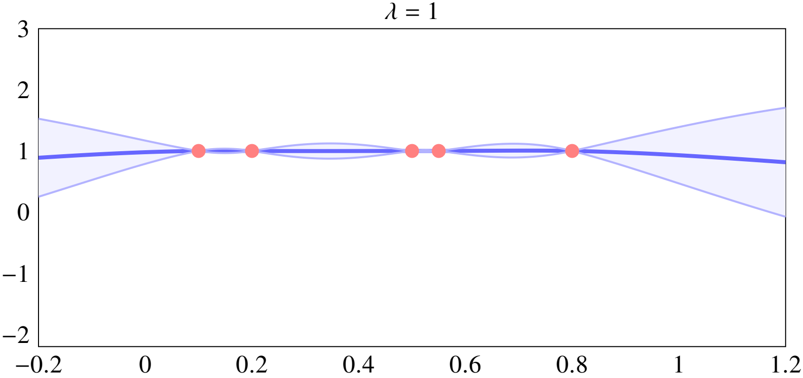
Lengthscale example

$$K_{\lambda}(x, y) = \left(1 + \frac{\sqrt{3}\,|x - y|}{\lambda}\right) \exp\left(-\frac{\sqrt{3}\,|x - y|}{\lambda}\right)$$



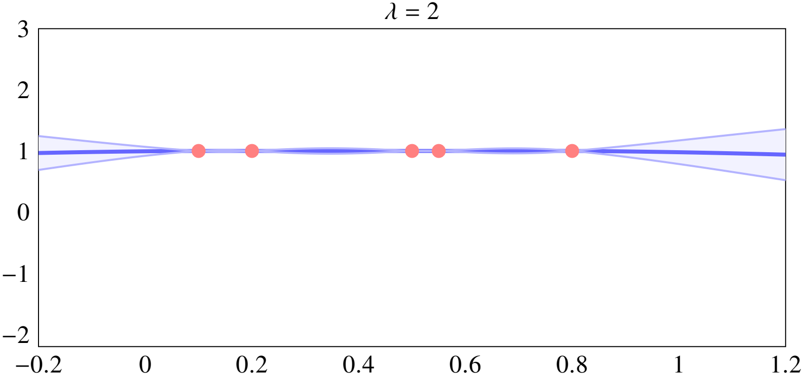
Lengthscale example

$$K_{\lambda}(x, y) = \left(1 + \frac{\sqrt{3}\,|x - y|}{\lambda}\right) \exp\left(-\frac{\sqrt{3}\,|x - y|}{\lambda}\right)$$



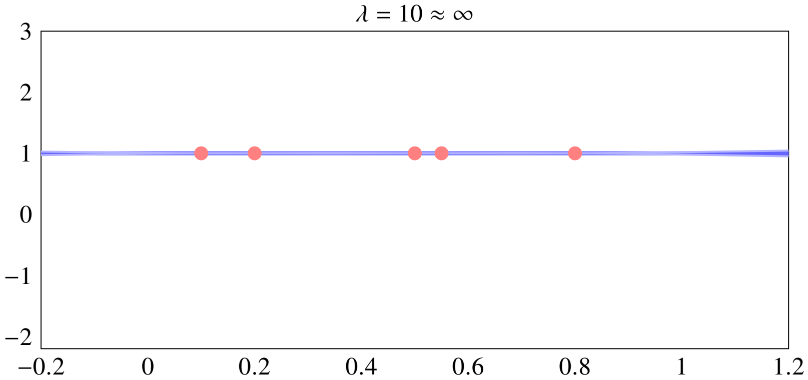
Lengthscale example

$$K_{\lambda}(x, y) = \left(1 + \frac{\sqrt{3}\,|x-y|}{\lambda}\right) \exp\left(-\frac{\sqrt{3}\,|x-y|}{\lambda}\right)$$



Lengthscale example

$$K_{\lambda}(x, y) = \left(1 + \frac{\sqrt{3}\,|x-y|}{\lambda}\right) \exp\left(-\frac{\sqrt{3}\,|x-y|}{\lambda}\right)$$



Orthonormal data?

Simplify the problem:

- Suppose f has expansion $f = \sum_{i=1}^{\infty} f_i \psi_i$ w.r.t. basis functions ψ_i.
- Instead of observing $f(\mathbf{x}_1), \ldots, f(\mathbf{x}_n)$ we observe $f_1, \ldots, f_n$.

For example, in this setting the MLE of the parameter α of the Sobolev space

$$H^\alpha = \left\{ f = \sum_{i=1}^{\infty} f_i \psi_i \;:\; \sum_{i=1}^{\infty} i^{2\alpha} f_i^2 < \infty \right\} \tag{21}$$

is simply

$$\alpha_n = \underset{\alpha \geq 0}{\arg\min} \left[\sum_{i=1}^{n} i^{2\alpha} f_i^2 - 2\alpha \log n! \right]. \tag{22}$$

$\Longrightarrow$ Life is much simpler.

Further references

1. **Driscoll (1973)**. The reproducing kernel Hilbert space structure of the sample paths of a Gaussian process. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*.
2. **Loh (2005)**. Fixed-domain asymptotics for a subclass of Matérn-type Gaussian random fields. *The Annals of Statistics*.
3. **Bachoc (2013)**. Cross validation and maximum likelihood estimations of hyper-parameters of Gaussian processes [...]. *Computational Statistics & Data Analysis*.
4. **Xu & Stein (2017)**. Maximum likelihood estimation for a smooth Gaussian random field model. *SIAM/ASA Journal on Uncertainty Quantification*.
5. **Steinwart (2019)**. Convergence types and rates in generic Karhunen-Loève expansions with applications to sample path properties. *Potential Analysis*.
6. **Teckentrup (2020)**. Convergence of Gaussian process regression with estimated hyper-parameters [...]. *SIAM/ASA Journal on Uncertainty Quantification*.
7. **Wang (2020)**. On the inference of applying Gaussian process modeling to a deterministic function. *Electronic Journal of Statistics*.
8. **Chen, Owhadi & Stuart (2021)**. Consistency of empirical Bayes and kernel flow for hierarchical parameter estimation. *Mathematics of Computation*.
9. **Karvonen (2023)**. Small sample spaces for Gaussian processes. *Bernoulli*.
10. **Petit (2023)**. Maximum likelihood estimation and prediction error for a Matérn model on the circle. *arXiv*:2209.07791v5.

Thank you for your attention!