Lengthscale-Informed Sparse Grids for High Dimensional Gaussian Process Emulation

30th Biennial Numerical Analysis Conference, Strathclyde

Elliot Addy[1], Jonas Latz[2] & Aretha Teckentrup[1].

[1]School of Mathematics and Maxwell Institute for Mathematical Sciences, University of Edinburgh

[2]Department of Mathematics, University of Manchester.

arXiv:2506.07797

Outline

- Introduction to Gaussian process emulation
- Link to scattered data approximation
- Three main assumptions for high-dimensional approximation:
  1. Sobolev spaces : Matérn kernels
  2. Bounded mixed derivatives : Separable kernels & sparse grids
  3. Function anisotropy : Anisotropic hyperparameters & designs
- Lengthscale-informed sparse grids
- Numerical examples

Gaussian process emulation

- We wish to approximate a function

$$f : \Omega \subset \mathbb{R}^d \to \mathbb{R}$$

for which point evaluations are computationally expensive, for example, a linear functional on parametric PDEs

$$f(\mathbf{x}) = \mathcal{F}(\mathbf{u}(\mathbf{x}))$$

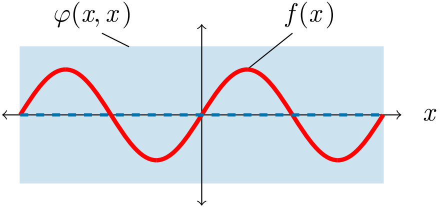
Gaussian process emulation

- We wish to approximate a function
$$f : \Omega \subset \mathbb{R}^d \to \mathbb{R}$$
for which point evaluations are computationally expensive, for example, a linear functional on parametric PDEs
$$f(\mathbf{x}) = \mathcal{F}(\mathbf{u}(\mathbf{x}))$$
- A *Gaussian process emulator* for f is a type of *surrogate model*; a relatively cheap approximation to f constructed from data $D = \{(\mathbf{x}, f(\mathbf{x}))\}_{\mathbf{x} \in \mathcal{X} \subset \Omega}$

Gaussian process emulation

- We wish to approximate a function
$$f \colon \Omega \subset \mathbb{R}^d \to \mathbb{R}$$
for which point evaluations are computationally expensive, for example, a linear functional on parametric PDEs
$$f(\mathbf{x}) = \mathcal{F}(\mathbf{u}(\mathbf{x}))$$
- A *Gaussian process emulator* for f is a type of *surrogate model*; a relatively cheap approximation to f constructed from data $D = \{(\mathbf{x}, f(\mathbf{x}))\}_{\mathbf{x} \in \mathcal{X} \subset \Omega}$

- Model f as a Gaussian process
$$f_0 \sim \mathrm{GP}(0, \varphi(\cdot, \cdot))$$



Gaussian process emulation

- We wish to approximate a function

$$f : \Omega \subset \mathbb{R}^d \to \mathbb{R}$$

for which point evaluations are computationally expensive, for example, a linear functional on parametric PDEs

$$f(\mathbf{x}) = \mathcal{F}(\mathbf{u}(\mathbf{x}))$$

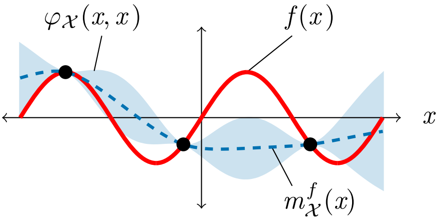
- A *Gaussian process emulator* for f is a type of *surrogate model*; a relatively cheap approximation to f constructed from data $D = \{(\mathbf{x}, f(\mathbf{x}))\}_{\mathbf{x} \in \mathcal{X} \subset \Omega}$

- Model f as a Gaussian process

$$f_0 \sim \mathrm{GP}(0, \varphi(\cdot, \cdot))$$

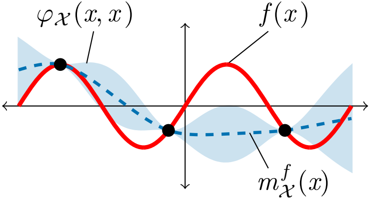
- Approximate f by the *mean function* of the posterior process, $m_{\mathcal{X}}^f$

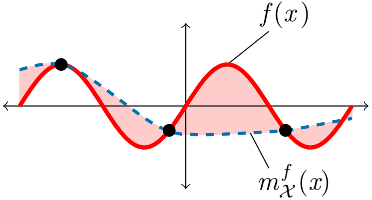
$$f_0 | D \sim \mathrm{GP}(m_{\mathcal{X}}^f(\cdot), \varphi_{\mathcal{X}}(\cdot, \cdot))$$



Link to scattered data approximation

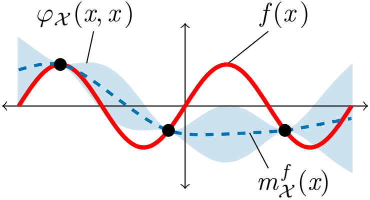
- We are interested in estimating the approximation error $\|m_{\mathcal{X}}^f - f\|$

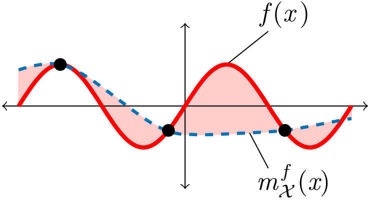




Link to scattered data approximation

- We are interested in estimating the approximation error $\|m_{\mathcal{X}}^f - f\|$





- The posterior mean function, $m_{\mathcal{X}}^f$, is exactly a *kernel interpolant* with respect to the covariance kernel $\varphi(\cdot,\cdot)$ and point-set $\mathcal{X} \subset \Omega$,

$$m_{\mathcal{X}}^f(\cdot) = s_{\mathcal{X},\varphi}(f) := \underset{\substack{g \in \mathcal{N}_\varphi(\Omega) \\ g(\mathcal{X}) = f(\mathcal{X})}}{\arg\min} \|g\|_{\mathcal{N}_\varphi(\Omega)}$$

where $\mathcal{N}_\varphi(\Omega)$ is the *Native Space* or *Reproducing Kernel Hilbert Space* (RKHS) of the kernel φ

Assumption 1: Sobolev Native spaces

- Use Matérn covariance kernels,

$$\phi_{\nu,\lambda}(\mathbf{x},\mathbf{x}') := \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{\|\mathbf{x}-\mathbf{x}'\|_2}{\lambda}\right)^{\nu} K_\nu \left(\sqrt{2\nu}\frac{\|\mathbf{x}-\mathbf{x}'\|_2}{\lambda}\right),$$

where K_ν is the Bessel function of the second kind, and ν, λ, and σ represent the *smoothness*, *lengthscale* and *standard deviation* of the kernel, respectively

Assumption 1: Sobolev Native spaces

- Use Matérn covariance kernels,

$$\phi_{\nu,\lambda}(\mathbf{x},\mathbf{x}') := \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu}\frac{\|\mathbf{x}-\mathbf{x}'\|_2}{\lambda}\right)^{\nu} K_\nu\left(\sqrt{2\nu}\frac{\|\mathbf{x}-\mathbf{x}'\|_2}{\lambda}\right),$$

  where K_ν is the Bessel function of the second kind, and ν, λ, and σ represent the *smoothness*, *lengthscale* and *standard deviation* of the kernel, respectively

- Notable special cases include the *exponential* and *Gaussian* kernels,

$$\phi_{1/2,\lambda}(\mathbf{x},\mathbf{x}') = \sigma^2 \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|_2}{\lambda}\right), \quad \phi_{\nu\to\infty,\lambda}(\mathbf{x},\mathbf{x}') = \sigma^2 \exp\left(-\frac{\|\mathbf{x}-\mathbf{x}'\|_2^2}{2\lambda^2}\right)$$

Assumption 1: Sobolev Native spaces

- Use Matérn covariance kernels,

$$\phi_{\nu,\lambda}(\mathbf{x}, \mathbf{x}') := \sigma^2 \frac{2^{1-\nu}}{\Gamma(\nu)} \left(\sqrt{2\nu} \frac{\|\mathbf{x} - \mathbf{x}'\|_2}{\lambda} \right)^{\nu} K_{\nu} \left(\sqrt{2\nu} \frac{\|\mathbf{x} - \mathbf{x}'\|_2}{\lambda} \right),$$

  where K_ν is the Bessel function of the second kind, and ν, λ, and σ represent the *smoothness*, *lengthscale* and *standard deviation* of the kernel, respectively

- Notable special cases include the *exponential* and *Gaussian* kernels,

$$\phi_{1/2,\lambda}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|_2}{\lambda} \right), \quad \phi_{\nu\to\infty,\lambda}(\mathbf{x}, \mathbf{x}') = \sigma^2 \exp\left(-\frac{\|\mathbf{x} - \mathbf{x}'\|_2^2}{2\lambda^2} \right)$$

Matérn Native spaces, *Wendland 2004*

Let $\Omega \subset \mathbb{R}^d$ be bounded. The corresponding Native spaces of Matérn kernels are isomorphic to Sobolev spaces, with equivalent norms

$$\mathcal{N}_{\phi_{\nu,\lambda}}(\Omega) = H^{\nu+d/2}(\Omega)$$

Assumption 1: Sobolev Native spaces

Sobolev error bound, *Arcangéli, Silanes, and Torrens 2012*

Let $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, be a bounded Lipshitz domain, and let $\mathcal{X} \subset \Omega$ be a discrete set. Let $\alpha, \beta \in \mathbb{N}_0$, $\beta \geq d/2$ be such that $\beta \geq \alpha$. Then

$$\| I - s_{\mathcal{X},\phi_{\beta-1/2,\lambda}} \|_{H^\beta(\Omega) \to H^\alpha(\Omega)} \leq C_1 h_{\mathcal{X},\Omega}^{\beta-\alpha},$$

where $h_{\mathcal{X},\Omega}$ is the *fill distance*, defined by

$$h_{\mathcal{X},\Omega} \coloneqq \sup_{\mathbf{x}\in\Omega} \inf_{\mathbf{x}'\in\mathcal{X}} \|\mathbf{x} - \mathbf{x}'\|_2.$$

[1] See e.g. Teckentrup 2020 and the references therein.

Assumption 1: Sobolev Native spaces

Sobolev error bound, *Arcangéli, Silanes, and Torrens 2012*

Let $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, be a bounded Lipshitz domain, and let $\mathcal{X} \subset \Omega$ be a discrete set. Let $\alpha, \beta \in \mathbb{N}_0$, $\beta \geq d/2$ be such that $\beta \geq \alpha$. Then

$$\|I - s_{\mathcal{X},\phi_{\beta-1/2,\lambda}}\|_{H^\beta(\Omega)\to H^\alpha(\Omega)} \leq C_1 h_{\mathcal{X},\Omega}^{\beta-\alpha},$$

where $h_{\mathcal{X},\Omega}$ is the *fill distance*, defined by

$$h_{\mathcal{X},\Omega} := \sup_{\mathbf{x}\in\Omega} \inf_{\mathbf{x}'\in\mathcal{X}} \|\mathbf{x} - \mathbf{x}'\|_2.$$

- **Problem:** Error is driven by the fill distance, which grows exponentially with dimension, d. In the best case [1], $h_{\mathcal{X},\Omega}$ decays like $N^{-1/d}$, with $N = |\mathcal{X}|$. To combat this, we need additional assumptions on f that allows us to

[1] See e.g. Teckentrup 2020 and the references therein.

Assumption 1: Sobolev Native spaces

Sobolev error bound, *Arcangéli, Silanes, and Torrens 2012*

Let $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, be a bounded Lipshitz domain, and let $\mathcal{X} \subset \Omega$ be a discrete set. Let $\alpha, \beta \in \mathbb{N}_0$, $\beta \geq d/2$ be such that $\beta \geq \alpha$. Then

$$\|I - s_{\mathcal{X},\phi_{\beta-1/2,\lambda}}\|_{H^\beta(\Omega)\to H^\alpha(\Omega)} \leq C_1 h_{\mathcal{X},\Omega}^{\beta-\alpha},$$

where $h_{\mathcal{X},\Omega}$ is the *fill distance*, defined by

$$h_{\mathcal{X},\Omega} \coloneqq \sup_{\mathbf{x}\in\Omega} \inf_{\mathbf{x}'\in\mathcal{X}} \|\mathbf{x} - \mathbf{x}'\|_2.$$

- **Problem:** Error is driven by the fill distance, which grows exponentially with dimension, d. In the best case [1], $h_{\mathcal{X},\Omega}$ decays like $N^{-1/d}$, with $N = |\mathcal{X}|$. To combat this, we need additional assumptions on f that allows us to
  1. Avoid the d-dimensional 2-norm in the definition of $h_{\mathcal{X},\Omega}$, and

[1] See e.g. Teckentrup 2020 and the references therein.

Assumption 1: Sobolev Native spaces

Sobolev error bound, *Arcangéli, Silanes, and Torrens 2012*

Let $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}$, be a bounded Lipshitz domain, and let $\mathcal{X} \subset \Omega$ be a discrete set. Let $\alpha, \beta \in \mathbb{N}_0$, $\beta \geq d/2$ be such that $\beta \geq \alpha$. Then

$$\|I - s_{\mathcal{X},\phi_{\beta-1/2,\lambda}}\|_{H^\beta(\Omega)\to H^\alpha(\Omega)} \leq C_1 h_{\mathcal{X},\Omega}^{\beta-\alpha},$$

where $h_{\mathcal{X},\Omega}$ is the *fill distance*, defined by

$$h_{\mathcal{X},\Omega} := \sup_{\mathbf{x}\in\Omega} \inf_{\mathbf{x}'\in\mathcal{X}} \|\mathbf{x} - \mathbf{x}'\|_2.$$

- **Problem:** Error is driven by the fill distance, which grows exponentially with dimension, d. In the best case [1], $h_{\mathcal{X},\Omega}$ decays like $N^{-1/d}$, with $N = |\mathcal{X}|$. To combat this, we need additional assumptions on f that allows us to
  1. Avoid the d-dimensional 2-norm in the definition of $h_{\mathcal{X},\Omega}$, and
  2. Prioritise more 'sensitive' directions over others

[1] See e.g. Teckentrup 2020 and the references therein.

Assumption 2: Bounded mixed derivatives

- A common assumption in high-dimensional approximation to let f live in a Sobolev space with *bounded mixed smoothness*

$$\|f\|_{H^\alpha}^2 = C \sum_{\substack{\boldsymbol{\beta}\in\mathbb{N}_0^d \\ |\boldsymbol{\beta}|_1\le\alpha}} \|D^{\boldsymbol{\beta}} f\|_{L^2}^2 \qquad \|f\|_{H^{\boldsymbol{\alpha}}_{\text{mix}}} = C' \sum_{\substack{\boldsymbol{\beta}\in\mathbb{N}_0^d \\ \beta_j\le\alpha_j}} \|D^{\boldsymbol{\beta}} f\|_{L^2}^2$$

Assumption 2: Bounded mixed derivatives

- A common assumption in high-dimensional approximation to let f live in a Sobolev space with *bounded mixed smoothness*

$$\|f\|_{H^\alpha}^2 = C \sum_{\substack{\boldsymbol{\beta}\in\mathbb{N}_0^d \\ |\boldsymbol{\beta}|_1\leq\alpha}} \|D^{\boldsymbol{\beta}} f\|_{L^2}^2 \qquad \|f\|_{H^{\boldsymbol{\alpha}}_{\text{mix}}} = C' \sum_{\substack{\boldsymbol{\beta}\in\mathbb{N}_0^d \\ \beta_j\leq\alpha_j}} \|D^{\boldsymbol{\beta}} f\|_{L^2}^2$$

Sparse grid kernel interpolant

Let $\{\mathcal{X}_l\}_{l\in\mathbb{N}_0}$ be a nested sequence of discrete sets, $\mathcal{X}_l \subset \mathcal{X}_{l+1} \subset \Omega$. We define the *Matérn sparse grid interpolation operator* of level $L \in \mathbb{N}_0$ by

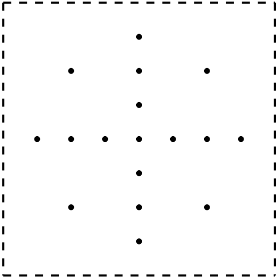
$$S_{L,\Phi_{\boldsymbol{\nu},\boldsymbol{\lambda}}} = \sum_{\substack{|\mathbf{l}|_1\leq L \\ \mathbf{l}\in\mathbb{N}_0^d}} \bigotimes_{j=1}^{d} \left(s_{\mathcal{X}_{l_j},\phi_{\nu_j,\lambda_j}} - s_{\mathcal{X}_{l_{j-1}},\phi_{\nu_j,\lambda_j}} \right),$$

where $\boldsymbol{\nu} = (\nu_1, \dots \nu_d) \in \mathbb{N}^d - 1/2$ and $\boldsymbol{\lambda} = (\lambda_1, \dots, \lambda_d) \in \mathbb{R}^d_{>0}$.

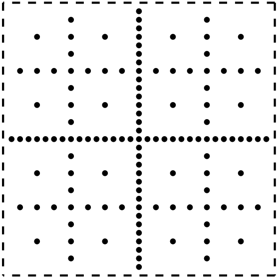
Assumption 2: Bounded mixed derivatives

- With respect to GP emulation, this is equivalent to using *separable Matérn* covariance kernels on points in a *sparse grid* design

$$\Phi_{\boldsymbol{\nu},\boldsymbol{\lambda}}(\mathbf{x},\mathbf{x}') := \prod_{j=1}^{d} \phi_{\nu_j,\lambda_j}(x_j, x_j') \qquad \mathcal{N}_{\Phi_{\boldsymbol{\nu},\boldsymbol{\lambda}}}(\Omega) = H_{\text{mix}}^{\boldsymbol{\nu}+1/2}(\Omega)$$



$L = 2$



$L = 4$

Assumption 2: Bounded mixed derivatives

Sparse grid error bound, *Nobile, Tempone, and Wolfers 2018*[a]

[a] As presented in *Teckentrup 2020*.

Let $0 \leq \alpha_j < \nu_j$ for all $1 \leq j \leq d$. For large enough $L \in \mathbb{N}_0$, there exists a constant C_2, independent of L, such that the error of a Matérn kernel interpolant with training points arranged in a sparse grid is bounded by

$$\|I - S_{L,\boldsymbol{\nu},\boldsymbol{\lambda}}\|_{H_{\mathrm{mix}}^{\boldsymbol{\nu}+1/2}(\Omega) \to H_{\mathrm{mix}}^{\boldsymbol{\alpha}+1/2}(\Omega)} \leq C_2 N^{-b_{\min}} (\log N)^{(1+b_{\min})(d-1)}, \qquad (1)$$

where $N = |\mathcal{X}_{d,L}^{\otimes}|$ and $b_{\min} = \min_{1 \leq j \leq d} \nu_j - \alpha_j$.

Assumption 2: Bounded mixed derivatives

Sparse grid error bound, *Nobile, Tempone, and Wolfers 2018*[a]

[a] As presented in *Teckentrup 2020*.

Let $0 \leq \alpha_j < \nu_j$ for all $1 \leq j \leq d$. For large enough $L \in \mathbb{N}_0$, there exists a constant C_2, independent of L, such that the error of a Matérn kernel interpolant with training points arranged in a sparse grid is bounded by

$$\|I - S_{L,\boldsymbol{\nu},\boldsymbol{\lambda}}\|_{H_{\text{mix}}^{\boldsymbol{\nu}+1/2}(\Omega) \to H_{\text{mix}}^{\boldsymbol{\alpha}+1/2}(\Omega)} \leq C_2 N^{-b_{\min}} (\log N)^{(1+b_{\min})(d-1)}, \tag{1}$$

where $N = |\mathcal{X}_{d,L}^{\otimes}|$ and $b_{\min} = \min_{1 \leq j \leq d} \nu_j - \alpha_j$.

- The *rate* of growth of points now only depends on the dimension d via a logarithmic factor

Assumption 2: Bounded mixed derivatives

Sparse grid error bound, *Nobile, Tempone, and Wolfers 2018*[a]

[a] As presented in *Teckentrup 2020*.

Let $0 \leq \alpha_j < \nu_j$ for all $1 \leq j \leq d$. For large enough $L \in \mathbb{N}_0$, there exists a constant C_2, independent of L, such that the error of a Matérn kernel interpolant with training points arranged in a sparse grid is bounded by

$$\|I - S_{L,\boldsymbol{\nu},\boldsymbol{\lambda}}\|_{H_{\mathrm{mix}}^{\boldsymbol{\nu}+1/2}(\Omega) \to H_{\mathrm{mix}}^{\boldsymbol{\alpha}+1/2}(\Omega)} \leq C_2 N^{-b_{\min}} (\log N)^{(1+b_{\min})(d-1)}, \tag{1}$$

where $N = |\mathcal{X}_{d,L}^{\otimes}|$ and $b_{\min} = \min_{1 \leq j \leq d} \nu_j - \alpha_j$.

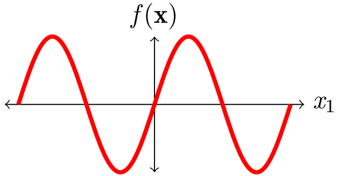
- The *rate* of growth of points now only depends on the dimension d via a logarithmic factor
- **However:** Practically, the maximum number of dimensions is limited to ~ 10

Assumption 3: Function anisotropy

- To apply these methods to functions in arbitrarily high dimensions, we need to *weight* each dimension differently according to the 'sensitivity' of f

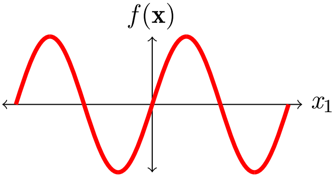
Assumption 3: Function anisotropy

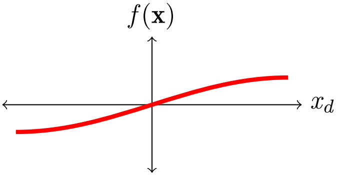
- To apply these methods to functions in arbitrarily high dimensions, we need to *weight* each dimension differently according to the 'sensitivity' of f
- By ordering dimensions from *most* to *least* sensitive, we assume the function gets progressively 'flatter'



Assumption 3: Function anisotropy

- To apply these methods to functions in arbitrarily high dimensions, we need to *weight* each dimension differently according to the 'sensitivity' of f
- By ordering dimensions from *most* to *least* sensitive, we assume the function gets progressively 'flatter'

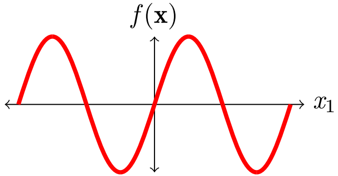




- **Idea:** Less sensitive directions require fewer points to approximate well

Assumption 3: Function anisotropy

- To apply these methods to functions in arbitrarily high dimensions, we need to *weight* each dimension differently according to the 'sensitivity' of f
- By ordering dimensions from *most* to *least* sensitive, we assume the function gets progressively 'flatter'



- **Idea:** Less sensitive directions require fewer points to approximate well
- How do we formalise *sensitivity*?

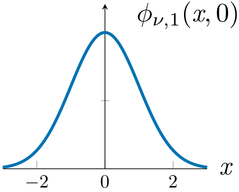
Assumption 3: Function anisotropy

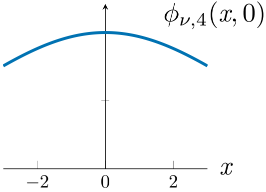
- Anisotropic methods often require additional regularity, ν, in later dimensions[2], limiting the space of functions they apply to

[2] See e.g. Nobile, Tempone, and Webster 2008.

Assumption 3: Function anisotropy

- Anisotropic methods often require additional regularity, ν, in later dimensions[2], limiting the space of functions they apply to
- Instead consider anisotropic lengthscale: Larger λ corresponds to a flatter kernel

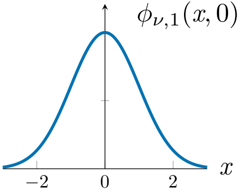


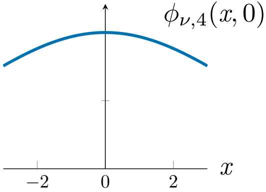


[2] See e.g. Nobile, Tempone, and Webster 2008.

Assumption 3: Function anisotropy

- Anisotropic methods often require additional regularity, ν, in later dimensions[2], limiting the space of functions they apply to
- Instead consider anisotropic lengthscale: Larger λ corresponds to a flatter kernel





- Assume f is bounded in the Native space norm of a separable Matérn kernel with *growing* lengthscale, $\boldsymbol{\lambda} \in \mathbb{R}^d_{>0}$, $\lambda_j < \lambda_{j+1}$,

$$\|f\|_{\mathcal{N}_{\Phi_{\boldsymbol{\nu},\boldsymbol{\lambda}}}(\Omega)} \leq C$$

[2]See e.g. Nobile, Tempone, and Webster 2008.

Lengthscale-informed sparse grids

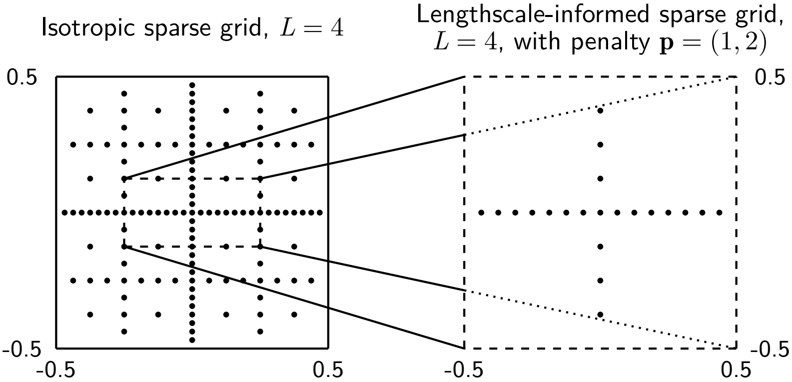
- **Idea:** Link the lengthscale and the *onset* of growth of points in each direction via a penalty parameter $\mathbf{p} \in \mathbb{N}_0$, such that

$$p_j = 2^{\lambda_j}$$

Lengthscale-informed sparse grids

- **Idea:** Link the lengthscale and the *onset* of growth of points in each direction via a penalty parameter $\mathbf{p} \in \mathbb{N}_0$, such that

$$p_j = 2^{\lambda_j}$$



Lengthscale-informed sparse grids

Native space error bound

Let $d \in \mathbb{N}$, $\mathbf{p} \in \mathbb{N}_0^d$, $L \in \mathbb{N}_0$ and $\Omega = (-1/2, 1/2)^d$. Let $\boldsymbol{\nu}, \boldsymbol{\alpha} \in \mathbb{N}_0^d$ be such that $\nu_j - \alpha_j = c \in \mathbb{N}$ for all $1 \leq j \leq d$. Then,

$$\|I - P_{L,\mathbf{p},\Phi_{\boldsymbol{\nu},2\mathbf{p}}}\|_{\mathcal{N}_{\Phi_{\boldsymbol{\nu},2\mathbf{p}}}(\Omega) \to \mathcal{N}_{\Phi_{\boldsymbol{\alpha},2\mathbf{p}}}(\Omega)} \leq C_{\boldsymbol{\nu},\boldsymbol{\alpha}} \sum_{k=0}^{d} 2^{-ck} \sum_{\mathfrak{u} \in \mathcal{P}_k^d} 2^{-c|\mathbf{p}_{\mathfrak{u}}|} \epsilon_{\boldsymbol{\nu}_{\mathfrak{u}},\boldsymbol{\alpha}_{\mathfrak{u}}}^{(k)}(L - |\mathbf{p}_{\mathfrak{u}}| - k), \tag{2}$$

where $\mathcal{P}_k^d$ is the set of all k-length subsets of $\{1, \ldots, d\}$ and $\epsilon_{\boldsymbol{\nu}_{\mathfrak{u}},\boldsymbol{\alpha}_{\mathfrak{u}}}^{(k)}(L)$ is the k-dimensional isotropic error bound

$$\|I - S_{L,\Phi_{\boldsymbol{\nu}_{\mathfrak{u}},1}}\|_{H_{\text{mix}}^{\boldsymbol{\nu}_{\mathfrak{u}}+1/2}(\Omega) \to H_{\text{mix}}^{\boldsymbol{\alpha}_{\mathfrak{u}}+1/2}(\Omega)} \leq \epsilon_{\boldsymbol{\nu}_{\mathfrak{u}},\boldsymbol{\alpha}_{\mathfrak{u}}}^{(k)}(L). \tag{3}$$

Lengthscale-informed sparse grids

Native space error bound

Let $d \in \mathbb{N}$, $\mathbf{p} \in \mathbb{N}_0^d$, $L \in \mathbb{N}_0$ and $\Omega = (-1/2, 1/2)^d$. Let $\boldsymbol{\nu}, \boldsymbol{\alpha} \in \mathbb{N}_0^d$ be such that $\nu_j - \alpha_j = c \in \mathbb{N}$ for all $1 \leq j \leq d$. Then,

$$\begin{aligned} &\|I - P_{L,\mathbf{p},\Phi_{\boldsymbol{\nu},2\mathbf{p}}}\|_{\mathcal{N}_{\Phi_{\boldsymbol{\nu},2\mathbf{p}}}(\Omega) \to \mathcal{N}_{\Phi_{\boldsymbol{\alpha},2\mathbf{p}}}(\Omega)} \\ &\leq C_{\boldsymbol{\nu},\boldsymbol{\alpha}} \sum_{k=0}^{d} 2^{-ck} \sum_{\mathfrak{u} \in \mathcal{P}_k^d} 2^{-c|\mathbf{p}_\mathfrak{u}|} \epsilon^{(k)}_{\boldsymbol{\nu}_\mathfrak{u},\boldsymbol{\alpha}_\mathfrak{u}}(L - |\mathbf{p}_\mathfrak{u}| - k), \end{aligned} \tag{2}$$

where $\mathcal{P}_k^d$ is the set of all k-length subsets of $\{1, \ldots, d\}$ and $\epsilon^{(k)}_{\boldsymbol{\nu}_\mathfrak{u},\boldsymbol{\alpha}_\mathfrak{u}}(L)$ is the k-dimensional isotropic error bound

$$\|I - S_{L,\Phi_{\boldsymbol{\nu}_\mathfrak{u},1}}\|_{H_{\text{mix}}^{\boldsymbol{\nu}_\mathfrak{u}+1/2}(\Omega) \to H_{\text{mix}}^{\boldsymbol{\alpha}_\mathfrak{u}+1/2}(\Omega)} \leq \epsilon^{(k)}_{\boldsymbol{\nu}_\mathfrak{u},\boldsymbol{\alpha}_\mathfrak{u}}(L). \tag{3}$$

- **Key:** Highlighted constants become very small in high dimensions, and so error grows much more slowly for 'small' L

L^∞ and posterior marginal variance in GPs

L^∞-bound

Let $\boldsymbol{\nu}, \boldsymbol{\alpha} \in \mathbb{R}^d$ such that $1/2 \leq \alpha_j \leq \nu_j$ for all $1 \leq j \leq d$. For $f \in \mathcal{N}_{\Phi_{\boldsymbol{\nu},2^{\mathbf{p}}}}(\Omega)$, we have

$$\|f - P_{L,\boldsymbol{\nu},\mathbf{p}}(f)\|_{L^\infty(\Omega)} \leq \sigma \|f - P_{L,\boldsymbol{\nu},\mathbf{p}}(f)\|_{\mathcal{N}_{\Phi_{\boldsymbol{\alpha},2^{\mathbf{p}}}}(\Omega)}, \tag{4}$$

where $\sigma = \sigma_1 \sigma_2 \cdots \sigma_d$, with σ_j the standard deviation of the one-dimensional Matérn kernel $\phi_{\alpha_j, 2^{p_j}}$.

L^∞ and posterior marginal variance in GPs

L^∞-bound

Let $\boldsymbol{\nu}, \boldsymbol{\alpha} \in \mathbb{R}^d$ such that $1/2 \leq \alpha_j \leq \nu_j$ for all $1 \leq j \leq d$. For $f \in \mathcal{N}_{\Phi_{\boldsymbol{\nu},2^{\mathbf{p}}}}(\Omega)$, we have

$$\|f - P_{L,\boldsymbol{\nu},\mathbf{p}}(f)\|_{L^\infty(\Omega)} \leq \sigma \|f - P_{L,\boldsymbol{\nu},\mathbf{p}}(f)\|_{\mathcal{N}_{\Phi_{\boldsymbol{\alpha},2^{\mathbf{p}}}}(\Omega)}, \tag{4}$$

where $\sigma = \sigma_1 \sigma_2 \cdots \sigma_d$, with σ_j the standard deviation of the one-dimensional Matérn kernel $\phi_{\alpha_j, 2^{p_j}}$.

Posterior marginal variance

The posterior marginal variance is bounded by

$$\|\tilde{\Phi}_{L,\boldsymbol{\nu},\mathbf{p}}(\cdot,\cdot)^{1/2}\|_{L^\infty(\Omega)} \leq \sigma \|I - P_{L,\boldsymbol{\nu},\mathbf{p}}\|_{\mathcal{N}_{\Phi_{\boldsymbol{\nu},2^{\mathbf{p}}}}(\Omega) \to \mathcal{N}_{\Phi_{\boldsymbol{\alpha},2^{\mathbf{p}}}}(\Omega)}, \tag{5}$$

for all $\boldsymbol{\alpha} \in \mathbb{R}^d$ such that $1/2 \leq \alpha_j \leq \nu_j$, $1 \leq j \leq d$. Here, $\sigma := \sigma_1 \sigma_2 \cdots \sigma_d$, where σ_j is the standard deviation of the one-dimensional Matérn kernel $\phi_{\alpha_j, 2^{p_j}}$.

Lengthscale-informed sparse grids

Theorem *Counting points in lengthscale-informed sparse grids*

Let $d \in \mathbb{N}$, $\mathbf{p} \in \mathbb{N}_0^d$ and $L \in \mathbb{N}_0$. Then $P_{L.d,\mathbf{p}}$ requires exactly

$$N_{d,\mathbf{p}}(L) := \sum_{k=0}^{d} 2^k \sum_{\mathfrak{u} \in \mathcal{P}_k^d} N_{k,\mathbf{0}}(L - |\mathbf{p}_\mathfrak{u}| - k) \tag{6}$$

many function evaluations, where $\mathcal{P}_k^d$ is the set of all k-length subsets of $\{1, \ldots, d\}$, and $N_{k,\mathbf{0}}$ is number points in an isotropic sparse grid of dimension k,

$$N_{k,\mathbf{0}}(L) = \begin{cases} \sum_{l=0}^{L-1} \binom{l+k-1}{k-1} 2^l & \text{if } L \geq 0, \text{and,} \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

[3] Extension of Plumlee 2014

Lengthscale-informed sparse grids

Theorem *Counting points in lengthscale-informed sparse grids*

Let $d \in \mathbb{N}$, $\mathbf{p} \in \mathbb{N}_0^d$ and $L \in \mathbb{N}_0$. Then $P_{L.d,\mathbf{p}}$ requires exactly

$$N_{d,\mathbf{p}}(L) := \sum_{k=0}^{d} 2^k \sum_{\mathfrak{u} \in \mathcal{P}_k^d} N_{k,\mathbf{0}}(L - |\mathbf{p}_{\mathfrak{u}}| - k) \tag{6}$$

many function evaluations, where $\mathcal{P}_k^d$ is the set of all k-length subsets of $\{1, \ldots, d\}$, and $N_{k,\mathbf{0}}$ is number points in an isotropic sparse grid of dimension k,

$$N_{k,\mathbf{0}}(L) = \begin{cases} \sum_{l=0}^{L-1} \binom{l+k-1}{k-1} 2^l & \text{if } L \geq 0, \text{and,} \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

- In general this is much less than in isotropic sparse grids!

[3] Extension of Plumlee 2014

Lengthscale-informed sparse grids

Theorem *Counting points in lengthscale-informed sparse grids*

Let $d \in \mathbb{N}$, $\mathbf{p} \in \mathbb{N}_0^d$ and $L \in \mathbb{N}_0$. Then $P_{L.d,\mathbf{p}}$ requires exactly

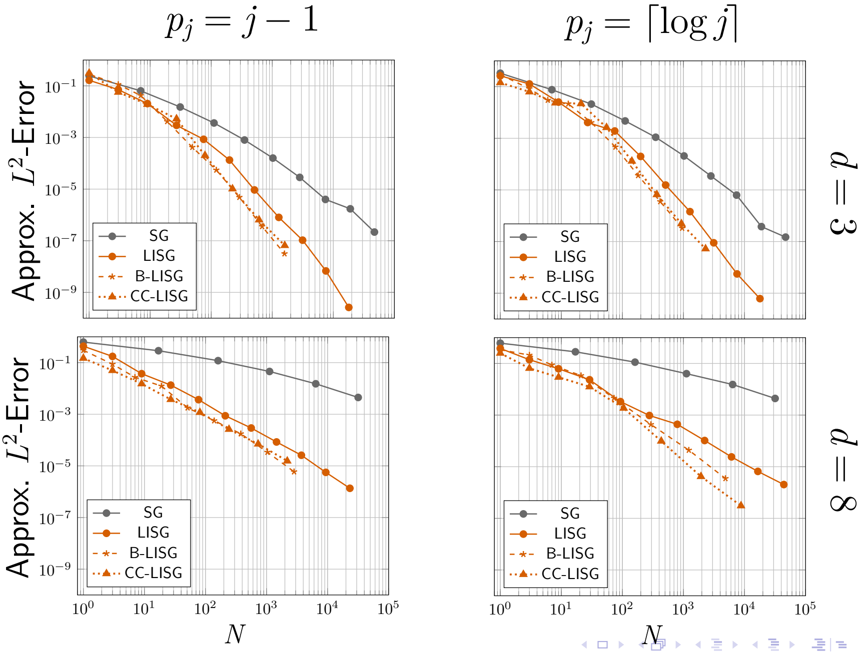
$$N_{d,\mathbf{p}}(L) := \sum_{k=0}^{d} 2^k \sum_{\mathfrak{u} \in \mathcal{P}_k^d} N_{k,\mathbf{0}}(L - |\mathbf{p}_{\mathfrak{u}}| - k) \tag{6}$$

many function evaluations, where $\mathcal{P}_k^d$ is the set of all k-length subsets of $\{1, \ldots, d\}$, and $N_{k,\mathbf{0}}$ is number points in an isotropic sparse grid of dimension k,

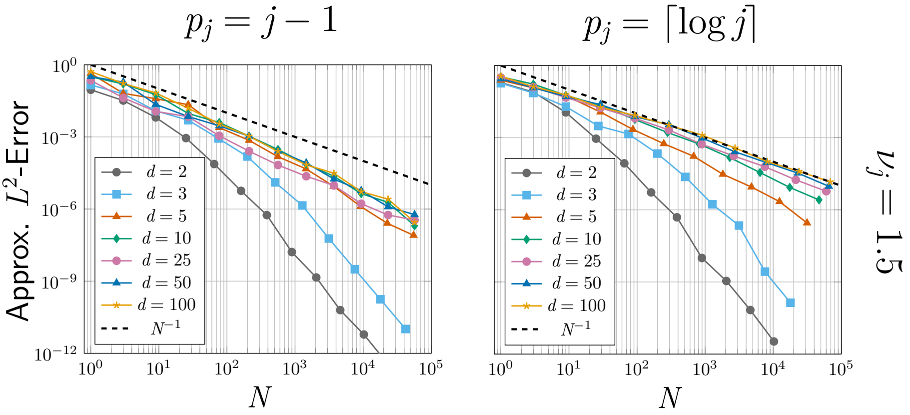
$$N_{k,\mathbf{0}}(L) = \begin{cases} \sum_{l=0}^{L-1} \binom{l+k-1}{k-1} 2^l & \text{if } L \geq 0, \text{and,} \\ 0 & \text{otherwise.} \end{cases} \tag{7}$$

- In general this is much less than in isotropic sparse grids!
- Sparse grid structure allows $\mathcal{O}(L2^L) < \mathcal{O}(N^3)$ inversions[3]

[3] Extension of Plumlee 2014



$p_j = j - 1$
$p_j = \lceil \log j \rceil$
Approx. L^2-Error
$d = 3$
$d = 8$
SG
LISG
B-LISG
CC-LISG
N



$p_j = j - 1$
$p_j = \lceil \log j \rceil$
Approx. L^2-Error
10^0
10^{-3}
10^{-6}
10^{-9}
10^{-12}
10^0 10^1 10^2 10^3 10^4 10^5
N
$d = 2$
$d = 3$
$d = 5$
$d = 10$
$d = 25$
$d = 50$
$d = 100$
N^{-1}
$\nu_j = 1.5$

Conclusions and Outlook

Summary

- With lengthscale-informed sparse grids, we are able to adapt our designs and kernels in GP emulation to match the λ-anisotropy in f
- For functions exhibiting sufficient lengthscale-anisotropy, the effect of dimension on the error is greatly reduced - allowing GP emulation in much higher dimensions without additional regularity assumptions

Conclusions and Outlook

Summary

- With lengthscale-informed sparse grids, we are able to adapt our designs and kernels in GP emulation to match the λ-anisotropy in f
- For functions exhibiting sufficient lengthscale-anisotropy, the effect of dimension on the error is greatly reduced - allowing GP emulation in much higher dimensions without additional regularity assumptions

Future Work

- Explore the natural use-cases of this lengthscale-anisotropy, e.g. in parametric PDEs
- Design adaptive inference algorithms exploiting sparse grid construction.

Conclusions and Outlook

Summary

- With lengthscale-informed sparse grids, we are able to adapt our designs and kernels in GP emulation to match the λ-anisotropy in f
- For functions exhibiting sufficient lengthscale-anisotropy, the effect of dimension on the error is greatly reduced - allowing GP emulation in much higher dimensions without additional regularity assumptions

Future Work

- Explore the natural use-cases of this lengthscale-anisotropy, e.g. in parametric PDEs
- Design adaptive inference algorithms exploiting sparse grid construction.

Thank You

Selected references

Wendland, Holger (2004). *Scattered Data Approximation*. Cambridge Monographs on Applied and Computational Mathematics. Cambridge University Press. DOI: 10.1017/CBO9780511617539.

Arcangéli, Rémi, María Cruz López de Silanes, and Juan José Torrens (2012). "Extension of sampling inequalities to Sobolev semi-norms of fractional order and derivative data". In: *Numerische Mathematik* 121.3, pp. 587–608.

Teckentrup, Aretha L. (2020). "Convergence of Gaussian Process Regression with Estimated Hyper-Parameters and Applications in Bayesian Inverse Problems". In: *SIAM/ASA Journal on Uncertainty Quantification* 8.4, pp. 1310–1337. DOI: 10.1137/19M1284816. eprint: https://doi.org/10.1137/19M1284816. URL: https://doi.org/10.1137/19M1284816.

Nobile, F., R. Tempone, and S. Wolfers (2018). "Sparse approximation of multilinear problems with applications to kernel-based methods in UQ". In: *Numerische Mathematik* 139. DOI: 10.1007/s00211-017-0932-4. URL: https://doi.org/10.1007/s00211-017-0932-4.

Nobile, F., R. Tempone, and C. G. Webster (2008). "An Anisotropic Sparse Grid Stochastic Collocation Method for Partial Differential Equations with Random Input Data". In: *SIAM Journal on Numerical Analysis* 46.5, pp. 2411–2442. DOI: 10.1137/070680540. eprint: https://doi.org/10.1137/070680540. URL: https://doi.org/10.1137/070680540.

Plumlee, Matthew (2014). "Fast Prediction of Deterministic Functions Using Sparse Grid Experimental Designs". In: *Journal of the American Statistical Association* 109.508, pp. 1581–1591. ISSN: 01621459. URL: http://www.jstor.org/stable/24247392 (visited on 07/09/2024).

Preprint

Thank You

arXiv:2506.07797

Lengthscale-informed sparse grids

Theorem *Counting points in lengthscale-informed sparse grids*

Let $d \in \mathbb{N}$, $\mathbf{p} \in \mathbb{N}_0^d$ and $L \in \mathbb{N}_0$. Then $P_{L.d,\mathbf{p}}$ requires exactly

$$N_{d,\mathbf{p}}(L) := \sum_{k=0}^{d} 2^k \sum_{\mathfrak{u} \in \mathcal{P}_k^d} N_{k,\mathbf{0}}(L - |\mathbf{p}_{\mathfrak{u}}| - k) \tag{8}$$

many function evaluations, where $\mathcal{P}_k^d$ is the set of all k-length subsets of $\{1, \ldots, d\}$, and $N_{k,\mathbf{0}}$ is number points in an isotropic sparse grid of dimension k,

$$N_{k,\mathbf{0}}(L) = \begin{cases} \sum_{l=0}^{L-1} \binom{l+k-1}{k-1} 2^l & \text{if } L \geq 0, \text{and,} \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

[4] Plumlee 2014

Lengthscale-informed sparse grids

Theorem *Counting points in lengthscale-informed sparse grids*

Let $d \in \mathbb{N}$, $\mathbf{p} \in \mathbb{N}_0^d$ and $L \in \mathbb{N}_0$. Then $P_{L.d,\mathbf{p}}$ requires exactly

$$N_{d,\mathbf{p}}(L) := \sum_{k=0}^{d} 2^k \sum_{\mathfrak{u} \in \mathcal{P}_k^d} N_{k,\mathbf{0}}(L - |\mathbf{p}_{\mathfrak{u}}| - k) \tag{8}$$

many function evaluations, where $\mathcal{P}_k^d$ is the set of all k-length subsets of $\{1, \ldots, d\}$, and $N_{k,\mathbf{0}}$ is number points in an isotropic sparse grid of dimension k,

$$N_{k,\mathbf{0}}(L) = \begin{cases} \sum_{l=0}^{L-1} \binom{l+k-1}{k-1} 2^l & \text{if } L \geq 0, \text{and}, \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

- In general this is much less than in isotropic sparse grids!

[4] Plumlee 2014

Lengthscale-informed sparse grids

Theorem *Counting points in lengthscale-informed sparse grids*

Let $d \in \mathbb{N}$, $\mathbf{p} \in \mathbb{N}_0^d$ and $L \in \mathbb{N}_0$. Then $P_{L.d,\mathbf{p}}$ requires exactly

$$N_{d,\mathbf{p}}(L) := \sum_{k=0}^{d} 2^k \sum_{\mathfrak{u} \in \mathcal{P}_k^d} N_{k,\mathbf{0}}(L - |\mathbf{p}_{\mathfrak{u}}| - k) \tag{8}$$

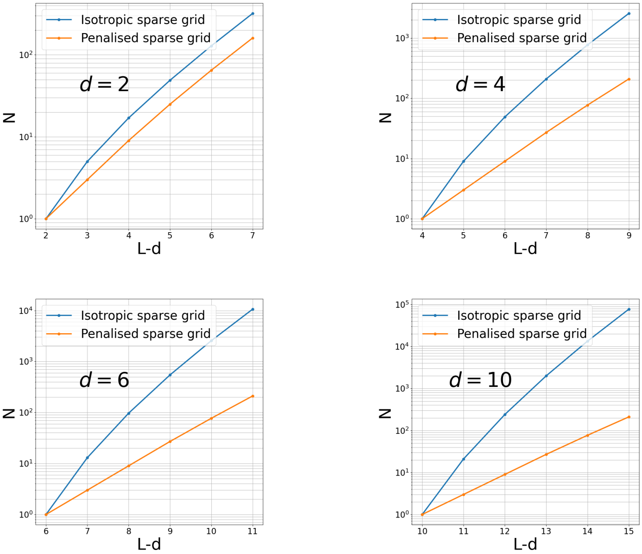
many function evaluations, where $\mathcal{P}_k^d$ is the set of all k-length subsets of $\{1, \ldots, d\}$, and $N_{k,\mathbf{0}}$ is number points in an isotropic sparse grid of dimension k,

$$N_{k,\mathbf{0}}(L) = \begin{cases} \sum_{l=0}^{L-1} \binom{l+k-1}{k-1} 2^l & \text{if } L \geq 0, \text{and,} \\ 0 & \text{otherwise.} \end{cases} \tag{9}$$

- In general this is much less than in isotropic sparse grids!
- Sparse grid structure allows $\mathcal{O}(L2^L) < \mathcal{O}(N^3)$ inversions[4]

[4]Plumlee 2014

Growth of lengthscale-informed vs isotropic sparse grids



Proposition *A*.

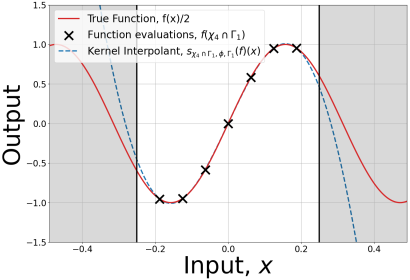
Define $\Omega_p \coloneqq (-1/2^{p+1}, 1/2^{p+1})$ and let $f \in \mathcal{N}_{\phi_{\mathrm{Mat};\ \nu,1,\sigma}}(\Omega_0)$ for some $0 \leq \alpha \leq \nu + 1/2$. Then, for $p \in \mathbb{Z}_{\geq 0}$,

$$\left\| s_{\chi_{l+p} \cap \Omega_p, \phi_{\nu,1,\sigma}, \Omega_p}(f/2^p) \right\|_{\mathcal{N}_{\phi_{\mathrm{Mat};\nu,1,\sigma}}(\Omega_p)} = \left\| s_{\chi_{l+p} \cap \Omega_p, \phi_{\nu,1,\sigma}, \Omega_0}(f/2^p) \right\|_{\mathcal{N}_{\phi_{\mathrm{Mat};\nu,1,\sigma}}(\Omega_0)}$$

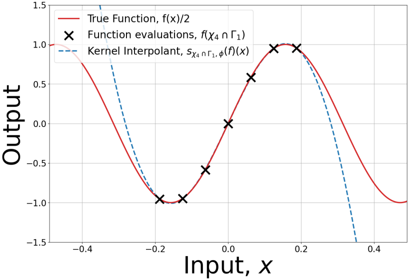
Proposition *A*.

Define $\Omega_p := (-1/2^{p+1}, 1/2^{p+1})$ and let $f \in \mathcal{N}_{\phi_{\mathrm{Mat};\ \nu,1,\sigma}}(\Omega_0)$ for some $0 \leq \alpha \leq \nu + 1/2$. Then, for $p \in \mathbb{Z}_{\geq 0}$,

$$\left\| s_{\chi_{l+p} \cap \Omega_p, \phi_{\nu,1,\sigma}, \Omega_p}(f/2^p) \right\|_{\mathcal{N}_{\phi_{\mathrm{Mat};\nu,1,\sigma}}(\Omega_p)} = \left\| s_{\chi_{l+p} \cap \Omega_p, \phi_{\nu,1,\sigma}, \Omega_0}(f/2^p) \right\|_{\mathcal{N}_{\phi_{\mathrm{Mat};\nu,1,\sigma}}(\Omega_0)}$$



$=$



Implementation and test function

- Due to induced Kronecker structure, fast covariance matrix inversions are possible; in general $\mathcal{O}(L^v 2^L) \ll \mathcal{O}(N^3)$, where v is independent of d for 'growing' lengthscales.

Implementation and test function

- Due to induced Kronecker structure, fast covariance matrix inversions are possible; in general $\mathcal{O}(L^v 2^L) \ll \mathcal{O}(N^3)$, where v is independent of d for 'growing' lengthscales.
- We consider a test functions of the form

$$f(\mathbf{x}) = \sum_{i=1}^{M} \xi_i \Phi_{\boldsymbol{\nu},\boldsymbol{\lambda}}(\mathbf{y}_i, \mathbf{x}), \tag{10}$$

with $\xi_i \sim \mathcal{N}(0, 5)$ and $\mathbf{y}_i \sim U(\Omega)$, i.i.d.

Implementation and test function

- Due to induced Kronecker structure, fast covariance matrix inversions are possible; in general $\mathcal{O}(L^v 2^L) \ll \mathcal{O}(N^3)$, where v is independent of d for 'growing' lengthscales.
- We consider a test functions of the form

$$f(\mathbf{x}) = \sum_{i=1}^{M} \xi_i \Phi_{\boldsymbol{\nu},\boldsymbol{\lambda}}(\mathbf{y}_i, \mathbf{x}), \tag{10}$$

  with $\xi_i \sim \mathcal{N}(0,5)$ and $\mathbf{y}_i \sim U(\Omega)$, i.i.d.
- Such functions have Native space norms bounded independently of d.

Implementation and test function

- Due to induced Kronecker structure, fast covariance matrix inversions are possible; in general $\mathcal{O}(L^v 2^L) \ll \mathcal{O}(N^3)$, where v is independent of d for 'growing' lengthscales.
- We consider a test functions of the form

$$f(\mathbf{x}) = \sum_{i=1}^{M} \xi_i \Phi_{\boldsymbol{\nu},\boldsymbol{\lambda}}(\mathbf{y}_i, \mathbf{x}), \tag{10}$$

  with $\xi_i \sim \mathcal{N}(0, 5)$ and $\mathbf{y}_i \sim U(\Omega)$, i.i.d.
- Such functions have Native space norms bounded independently of d.
- We consider two types of anisotropy, $\lambda = 2^{\mathbf{p}}$,

$$\mathbf{p}_{\mathrm{lin},j} = j - 1 \qquad \text{and} \qquad \mathbf{p}_{\log,j} = \lceil \log j \rceil. \tag{11}$$

Fast inference algorithm[5].

- For a given multi-index $\mathbf{a} \in \mathcal{A}^d_{\mathbf{p},L} \subset \mathbb{N}^d_0$, denote the sub-vector of $\mathbf{w}$ corresponding to the design points in the component grid $\mathcal{X}_{a_1} \times \cdots \times \mathcal{X}_{a_d} \subset \mathcal{X}^{\otimes}_{\mathbf{p},L}$.

Initialise $\mathbf{w} = \mathbf{0} \in \mathbb{R}^{N_{d,\mathbf{p}}(L)}$
for $\mathbf{a} \in \mathcal{A}^d_{\mathbf{p},L}$ **do**

$$\mathbf{w}_{\mathbf{a}} = \mathbf{w}_{\mathbf{a}} + u_{\mathbf{p},L}(\mathbf{a}) \left[\bigotimes_{j=1}^{d} \phi_{\nu_j,2^{p_j}}(\mathcal{X}_{a_j}, \mathcal{X}_{a_j})^{-1} \right] f(\mathcal{X}_{a_1} \times \cdots \times \mathcal{X}_{a_d})$$

end for

- For 'growing' lengthscales, $|\mathcal{A}^d_{\mathbf{p},L}|$ is bounded independently of d.

[5] Adapted from Algorithm 1, Plumlee 2014