The dimension weighted fast multipole method for scattered data approximation

Helmut Harbrecht, Michael Multerer, Jacopo Quizi

Università della Svizzera italiana

Outline

1. Problem formulation
2. Kernels on anisotropic data sets
3. Weighted total degree polynomial approximation
4. The dimension weighted fast multipole method
5. Numerical results

Harbrecht et al. 2016

Problem formulation

- We consider the following stationary diffusion problem

$$-\Delta u(\omega) = f \text{ in } D(\omega), \quad u(\omega) = 0 \text{ on } \partial D(\omega), \quad \omega \in \Omega,$$

for an uncertain domain $D(\omega) \subset \mathbb{R}^N$.

- We denote by $D_0 \subset \mathbb{R}^D$, for $d \in \mathbb{N}$, a domain with Lipschitz continuous boundary.
- To model the shape uncertainty, we introduce a random deformation field $\boldsymbol{\chi}\colon D_0 \times \Omega \to \mathbb{R}^n$.
- We impose the uniformity condition

$$\|\boldsymbol{\chi}(\omega)\|_{C^1(\overline{D_0};\mathbb{R}^n)}, \|\boldsymbol{\chi}^{-1}(\omega)\|_{C^1(\overline{D(\omega)};\mathbb{R}^n)} \lesssim 1 \quad \text{for } \mathbb{P}\text{-a.e. } \omega \in \Omega,$$

which satisfies

$$D(\omega) = \boldsymbol{\chi}(D_0, \omega).$$

- We consider the Karhunen-Loève expansion

$$\chi(\omega) = \mathrm{Id} + \sum_{k=1}^{M} \sigma_k \chi_k Y_k(\omega), \quad \omega \in \Omega.$$

- $\{Y_k\}$ are i.i.d. random variables uniformly distributed on $[-1, 1]$.
- Parametrizing the random variable by their image, we obtain the parametric deformation field

$$\chi(\zeta) = \mathrm{Id} + \sum_{k=1}^{M} \sigma_k \chi_k \zeta_k, \quad \zeta \in [-1, 1]^M. \tag{1}$$

- We have now a parametric diffusion problem

$$-\Delta u(\zeta) = f \text{ in } D(\zeta), \quad u(\zeta) = 0 \text{ on } \partial D(\zeta), \quad \zeta \in [-1, 1]^M.$$

- The goal we consider consists of the solution-based functional reconstruction:

$$F(u)(\zeta) = \left(u \circ \boldsymbol{\chi}^{-1}(\zeta), \phi\right)_{L^2(D_0)}. \tag{2}$$

- Each ζ_k in (1) introduces additional dimensions to the problem.
- The importance of each dimension is weighted by σ_k, which decays as k increases, creating anisotropic data.
- To reconstruct the quantity of interest we use the fast anisotropic kernel method.
- **Goal**: Given functional observed values $\boldsymbol{F}$, we want to solve $\boldsymbol{K}\boldsymbol{\alpha} = \boldsymbol{F}$, where $\boldsymbol{K}$ representes the kernel matrix associated with the interpolation problem.

Kernels on anisotropic data sets

Definition. Let $\Omega \subset \mathbb{R}^d$ and $\big(\mathcal{H}, (\cdot,\cdot)_\mathcal{H}\big)$ be a Hilbert space of functions $h\colon \Omega \to \mathbb{R}$. A *reproducing kernel* $\mathcal{K}$ for $\mathcal{H}$ is a function $\mathcal{K} : \Omega \times \Omega \to \mathbb{R}$ such that

1. $\mathcal{K}(\cdot, \boldsymbol{x}) \in \mathcal{H}$ for all $\boldsymbol{x} \in \Omega$,
2. $h(\boldsymbol{x}) = \big(h, \mathcal{K}(\cdot, \boldsymbol{x})\big)_\mathcal{H}$ for all $h \in \mathcal{H}$ and all $\boldsymbol{x} \in \Omega$.

If $\mathcal{H}$ exhibits a reproducing kernel, we call it a *reproducing kernel Hilbert space (RKHS)*.

- Given a set of *data sites* $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\} \subset \mathbb{R}^d$ and *data values* $F_1, \dots, F_N \in \mathbb{R}$, we may want to solve the kernel ridge regression

$$\min_{\alpha_1,\dots,\alpha_N} \sum_{i=1^n} \big(F_i - s_X(\boldsymbol{x}_i)\big)^2 + \lambda \|s_X\|_\mathcal{H}^2 \to (\boldsymbol{K} + \lambda \boldsymbol{I})\boldsymbol{\alpha} = \boldsymbol{F},$$

where $s_X(\boldsymbol{x}) = \sum_{i=1}^N \alpha_i \, \mathcal{K}(\boldsymbol{x}, \boldsymbol{x}_i)$ is the kernel interpolant.

- **Problem:** Direct computation of all pairwise interactions for large N is costly.
- To compress $\boldsymbol{K}$, we exploit kernel expansions obtained by interpolation.
- In particular, to deal with posssibly high dimensional data, we use interpolation by anistropic total degree polynomials.

Definition. The data set $X = \{\boldsymbol{x}_1, \dots, \boldsymbol{x}_N\} \subset \Omega$ is quasi-uniform if the *fill distance*

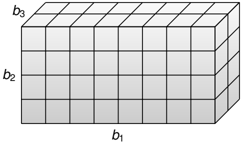
$$h_{X,\Omega} := \sup_{\boldsymbol{x}\in\Omega} \min_{\boldsymbol{x}_i \in X} \|\boldsymbol{x} - \boldsymbol{x}_i\|_2$$

is proportional to the separation radius $q_X := \min_{i\neq j} \|\boldsymbol{x}_i - \boldsymbol{x}_j\|_2$, i.e., there exists a constant $c_{X,\Omega} \in (0,1)$ such that $0 < c_{X,\Omega} \leq \frac{q_X}{h_{X,\Omega}} \leq \frac{1}{c_{X,\Omega}}$.

- We assume that the set X is *quasi-uniform* and contained in an anisotropic axis parallel cuboid

$$\mathcal{B} = [0, b_1] \times \cdots \times [0, b_d],$$

with *dimension weights* $b_1 \geq \ldots \geq b_d \geq 0$.



In this framework, the first step is to show the existence of an analytic extension of the kernel function:

- We introduce the linear transformation

$$\boldsymbol{B}\colon [0,1]^d \to \mathcal{B}, \quad \hat{\boldsymbol{x}} \mapsto \boldsymbol{x} = \boldsymbol{B}\hat{\boldsymbol{x}},$$

  which is given by the matrix $\boldsymbol{B} := \mathrm{diag}(b_1, \ldots, b_d)$.
- Thus, considering the kernel $\mathcal{K}\colon \mathcal{B} \times \mathcal{B} \to \mathbb{R}$ yields to the *transported kernel*

$$\mathcal{K}_{\boldsymbol{B}}\colon [0,1]^d \times [0,1]^d \to \mathbb{R}, \quad \mathcal{K}_{\boldsymbol{B}}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}}) := \mathcal{K}(\boldsymbol{B}\hat{\boldsymbol{x}}, \boldsymbol{B}\hat{\boldsymbol{y}}). \tag{3}$$

- The analysis of the kernel interpolation is based on the *asymptotical smoothness* property of the kernel $\mathcal{K}$, that is

$$\left|\frac{\partial^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}}{\partial \boldsymbol{x}^{\boldsymbol{\alpha}} \partial \boldsymbol{y}^{\boldsymbol{\beta}}} \mathcal{K}(\boldsymbol{x}, \boldsymbol{y})\right| \leq c_{\mathcal{K}} \frac{(|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|)!}{\rho^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}\|\boldsymbol{x}-\boldsymbol{y}\|_2^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}}, \quad c_{\mathcal{K}}, \rho > 0. \tag{4}$$

- The derivatives of the transported kernel satisfy the bound

$$\left|\frac{\partial^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}}{\partial \hat{\boldsymbol{x}}^{\boldsymbol{\alpha}} \partial \hat{\boldsymbol{y}}^{\boldsymbol{\beta}}} \mathcal{K}_{\boldsymbol{B}}(\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}})\right| \leq c_{\mathcal{K}} \left(\frac{\boldsymbol{b}}{\rho}\right)^{\boldsymbol{\alpha}+\boldsymbol{\beta}} \frac{(|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|)!}{\|\boldsymbol{B}(\hat{\boldsymbol{x}}-\hat{\boldsymbol{y}})\|_2^{|\boldsymbol{\alpha}|+|\boldsymbol{\beta}|}},$$

with $\boldsymbol{b} := [b_1 \ldots, b_d]$.

- Inserting a lower bound $\eta > 0$ for the distance of now yields the analyticity of the transported kernel for any pair of points $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}} \in [0,1]^d$ satisfying $\|\boldsymbol{B}(\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}})\|_2 \geq \eta$.

Corollary. Let $\hat{\boldsymbol{x}}, \hat{\boldsymbol{y}} \in [0,1]^d$ be such that $\|\boldsymbol{B}(\hat{\boldsymbol{x}} - \hat{\boldsymbol{y}})\|_2 \geq \eta$ for some $\eta > 0$. Then, each component of the kernel function $\mathcal{K}_{\boldsymbol{B}}$ from (3) admits an analytic extension into $\boldsymbol{\Sigma}(\boldsymbol{\tau})$ for $\tau_k \in [0, c_k \eta \rho / b_k]$, where $c_k > 0$ and $\gamma := \sum_{k=1}^{d} c_k < 1$, where $\boldsymbol{\Sigma}(\boldsymbol{\tau}) := \Sigma(\tau_1) \times \cdots \times \Sigma(\tau_d)$ with $\Sigma(\tau) := \{z \in \mathbb{C} : \mathrm{dist}(z, [0,1]) \leq \tau\}$.

DeVore and Lorentz 1993

Weighted total degree polynomial approximation

- Exploiting the available anisotropy for approximation, we derive corresponding estimates of the error for polynomial best approximation and extend them to the interpolation case.

Lemma. Let $\Pi_q := \operatorname{span}\{1, \ldots, x^q\}$ denote the space of all univariate polynomials up to degree q. Given a function $f \in C([0,1])$ which admits an analytic extension $\tilde{f}$ into the region $\Sigma(\tau)$ for some $\tau > 0$, there holds, for $1 < \rho := 2\tau + \sqrt{1 + 4\tau^2}$, that

$$\min_{w \in \Pi_q} \|f - w\|_{C([0,1])} \leq \frac{2\rho}{\rho - 1} e^{-(q+1)\log\rho} \|\tilde{f}\|_{C(\Sigma(\tau))}.$$

- We therefore define the projection operator

$$U_q\colon C([0,1]) \to \Pi_q, \quad U_q f := \operatorname{argmin}_{w\in\Pi_q} \|f - w\|_{C([0,1])}.$$

- By observing $(U_0 f)(x) = \frac{1}{2}\bigg(\min_{y\in[0,1]} f(y) + \max_{y\in[0,1]} f(y)\bigg)$, we infer the stability estimate

$$\|U_0 f\|_{C([0,1])} \leq \|f\|_{C([0,1])}.$$

- To define the weighted total degree approximation, we introduce the *weighted total degree index sets*

$$\Lambda_{\boldsymbol{\omega},q,d} := \bigg\{\mathbf{0} \leq \boldsymbol{\alpha} \in \mathbb{N}^d : \sum_{\ell=1}^{d} \omega_\ell \alpha_\ell \leq q\bigg\}, \boldsymbol{\omega} \in [0,\infty)^d.$$

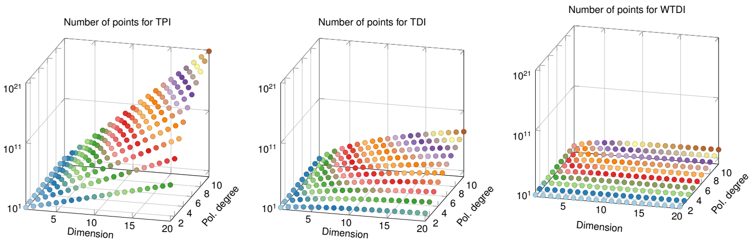


Figure: Number of interpolation points required for TPI, TDI, and WTDI based on a quadratic increase of the sequence $\{\tau_k\}$ of the convergence radii.

- Given the index set $\Lambda_{\boldsymbol{\omega},q,d}$, we define

$$\mathcal{P}_{\boldsymbol{\omega},q} := \operatorname{span}\{\boldsymbol{x}^{\boldsymbol{\alpha}} : \boldsymbol{\alpha} \in \Lambda_{\boldsymbol{\omega},q,d}\}.$$

- With respect to $\mathcal{P}_{\boldsymbol{\omega},q}$, we introduce the *weighted total degree approximation operator of degree* $q \in \mathbb{N}$ given by

$$P_{\boldsymbol{\omega},q} \colon C([0,1]^d) \to \mathcal{P}_{\boldsymbol{\omega},q}, \quad P_{\boldsymbol{\omega},q} f := \sum_{\boldsymbol{\alpha} \in \Lambda_{\boldsymbol{\omega},q,d}} \Delta_{\boldsymbol{\alpha}} f.$$

Harbrecht et al. 2024; Haji-Ali et al. 2018

Theorem. Let $v \in C([0,1]^d)$ admit an analytic extension $\tilde{v}$ into the region $\boldsymbol{\Sigma}(\boldsymbol{\tau})$ for a monotonously increasing sequence $0 < \tau_1 \leq \ldots \leq \tau_d \leq \ldots$ and $\tau_k \geq ck^r$ for some $c, r > 1$. Define $1 < \rho_k := 2\tau_k + \sqrt{1 + 4\tau_k^2}$ for $k = 1, \ldots, d$ and introduce the weights $\omega_k := \log \rho_k$. Then, there holds

$$\left\|(\mathrm{Id}^{(d)} - P_{\boldsymbol{\omega},q})v\right\|_{C([0,1]^d)} \leq c_r 2^{d+1} e^{-q(1-\frac{1}{r}\log\log d)} \|\tilde{v}\|_{C(\boldsymbol{\Sigma}(\boldsymbol{\tau}))}$$

with the constant c_r such that $|\Lambda_{\boldsymbol{\omega},q,d}| \leq c_r \log(d)^{\frac{q}{r}}$.

Corollary. Under the same hypotheses, there holds for the interpolation $I_{\boldsymbol{\omega},q}\colon C([0,1]^d) \to \mathcal{P}_{\boldsymbol{\omega},q}$ with Lebesgue constant $L_{\boldsymbol{\omega},q}$ that

$$\left\|(\mathrm{Id}^{(d)} - I_{\boldsymbol{\omega},q})v\right\|_{C([0,1]^d)} \leq (1 + L_{\boldsymbol{\omega},q}) c_r 2^{d+1} e^{-q(1-\frac{1}{r}\log\log d)} \|\tilde{v}\|_{C(\boldsymbol{\Sigma}(\boldsymbol{\tau}))}.$$

The dimension weighted fast multipole method

- We provide a cardinality balanced clustering for set of data sites X which yields a balanced binary tree $\mathcal{T}$.
- The clusters $\nu, \nu' \in \mathcal{T}$ with $j_\nu = j_{\nu'}$ are called *admissible* iff

$$\operatorname{dist}(B_\nu, B'_\nu) \geq \eta \max \left\{ \operatorname{diam} B_\nu, \operatorname{diam} B'_\nu \right\}$$

  holds for some $\eta > 0$.
- We denote by

$$\mathcal{T} \boxtimes \mathcal{T} := \{\nu \times \nu' \colon \nu, \nu' \in \mathcal{T},\ j_\nu = j'_\nu\}$$

  the level-wise Cartesian product of the cluster tree $\mathcal{T}$,

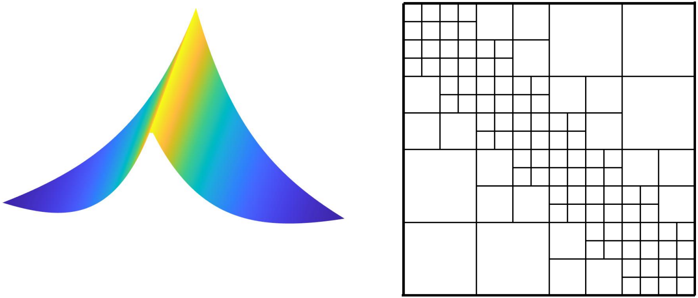


Figure: $\mathcal{H}^2$-*matrices* formalise a certain class of matrices, which exhibit a low-rank structure with respect to a given block-clustering.

- Given an admissible cluster $\nu \times \nu'$, we now approximate

$$\mathcal{K}(\boldsymbol{x}, \boldsymbol{y}) \approx \widetilde{\mathcal{K}}(\boldsymbol{x}, \boldsymbol{y}) := \sum_{\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in X_{\boldsymbol{\omega}, q, d}} c_{\boldsymbol{\alpha}, \boldsymbol{\alpha}'}^{\nu, \nu'} p_{\boldsymbol{\alpha}}^{\nu}(\boldsymbol{x}) p_{\boldsymbol{\alpha}'}^{\nu'}(\boldsymbol{y})$$

with coefficients $c_{\boldsymbol{\alpha}, \boldsymbol{\alpha}'}^{\nu, \nu'} \in \mathbb{R}$ and the transported polynomials

$$p_{\boldsymbol{\alpha}}^{\nu} := \hat{p}_{\boldsymbol{\alpha}} \circ a_{\nu}^{-1}.$$

- The coefficients $c_{\boldsymbol{\alpha}, \boldsymbol{\alpha}'}^{\nu, \nu'}$ in matrix notation according to

$$\boldsymbol{C}_{\nu, \nu'} := [c_{\boldsymbol{\alpha}, \boldsymbol{\alpha}'}^{\nu, \nu'}]_{\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \Lambda_{\boldsymbol{\omega}, q, d}} = \boldsymbol{V}^{-1} \boldsymbol{S}_{\nu, \nu'} \boldsymbol{V}^{-\intercal}.$$

- The coupling matrix is

$$\boldsymbol{S}_{\nu, \nu'} := \left[\mathcal{K}\big(a_{\nu}(\hat{\boldsymbol{\xi}}_{\boldsymbol{\alpha}}), a_{\nu'}(\hat{\boldsymbol{\xi}}_{\boldsymbol{\alpha}'})\big)\right]_{\boldsymbol{\alpha}, \boldsymbol{\alpha}' \in \Lambda_{\boldsymbol{\omega}, q, d}}.$$

Börm 2010

- We approximate the matrix block $\boldsymbol{K}_{\nu,\nu'}$ as:

$$\boldsymbol{K}_{\nu,\nu'} \approx \widetilde{\boldsymbol{K}}_{\nu,\nu'} = \boldsymbol{P}_{\nu} \boldsymbol{C}_{\nu,\nu'} \boldsymbol{P}_{\nu'}^{\mathsf{T}} = \boldsymbol{P}_{\nu} \boldsymbol{V}^{-1} \boldsymbol{S}_{\nu,\nu'} \boldsymbol{V}^{-\mathsf{T}} \boldsymbol{P}_{\nu'}^{\mathsf{T}}.$$

- By fixing the index set $\Lambda_{\boldsymbol{\omega},q,d}$ for all clusters,

$$\boldsymbol{T}_{\nu_{\text{child}}} := \boldsymbol{V}^{-1} \left[p_{\boldsymbol{\alpha}'}^{\nu} \left(a_{\nu_{\text{child}}}(\hat{\boldsymbol{\xi}}_{\boldsymbol{\alpha}}) \right) \right]_{\boldsymbol{\alpha},\boldsymbol{\alpha}' \in \Lambda_{\boldsymbol{\omega},q,d}},$$

  there holds the two-scale relation

$$\boldsymbol{P}_{\nu} = \begin{bmatrix} \boldsymbol{P}_{\nu_{\text{child}_1}} \boldsymbol{T}_{\nu_{\text{child}_1}} \\ \boldsymbol{P}_{\nu_{\text{child}_2}} \boldsymbol{T}_{\nu_{\text{child}_2}} \end{bmatrix}.$$

Lemma. Let the assumptions of the previous Theorem be satisfied for $\tau_k \in [0, c_k\eta\rho/b_k]$. In case of quasi-uniform points, the matrix $\widetilde{\boldsymbol{K}}$ satisfies the following error estimate

$$\frac{\big\|\boldsymbol{K} - \widetilde{\boldsymbol{K}}\big\|_F}{\big\|\boldsymbol{K}\big\|_F} \lesssim (1 + L_{\boldsymbol{\omega},q})c_r 2^d e^{-q(1-\frac{1}{r}\log\log d)} \frac{c_{\mathcal{K}}}{1-\gamma}.$$

Numerical results

- The reference shape D_0 is given by the kite domain shown in this figure with bounding box $[-0.26, 0.39] \times [-0.39, 0.39]$. It is discretized by 327 680 piecewise linear parametric finite elements.

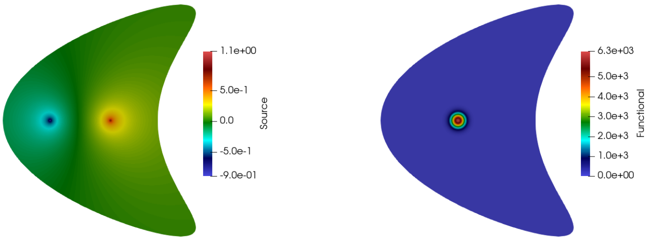


Figure: Visualization of the source term (left) and the functional ϕ (right).

- In order to define the deformation field, we compute the first 20 eigenfunctions $\{\chi_k\}$ of the matrix-valued covariance kernel

$$\mathcal{C}\colon D_0 \times D_0 \to \mathbb{R}^{2\times 2}, \quad \mathcal{C}(\boldsymbol{x}, \boldsymbol{y}) := \begin{bmatrix} 10^2 e^{-5\|\boldsymbol{x}-\boldsymbol{y}\|_2^2} & e^{-\|\boldsymbol{x}-\boldsymbol{y}\|_2^2} \\ e^{-\|\boldsymbol{x}-\boldsymbol{y}\|_2^2} & 10^2 e^{-5\|\boldsymbol{x}-\boldsymbol{y}\|_2^2}, \end{bmatrix}$$

  discretized on the aforementioned finite element mesh.

- The singular values are set to be $\sigma_k := 0.25k^{-3}$.

- We consider the source term

$$f(\boldsymbol{x}) = \frac{1}{2\pi}\big(\log\|\boldsymbol{x}+\boldsymbol{x}_0\|_2 - \log\|\boldsymbol{x}-\boldsymbol{x}_0\|_2\big), \quad \boldsymbol{x} \in \mathbb{R}^2,$$

  where $\boldsymbol{x}_0 = [0.1, 0]^\intercal \in D_0$.

- The particular quantity of interest (2) involves the localized Gaussian

$$\phi(\boldsymbol{x}) = \frac{5\pi}{10^4} e^{-\frac{5}{10^4}\|\boldsymbol{x}\|_2^2}.$$

- We consider $N = 849\,375$ data points for the kernel interpolation, obtained by rescaling each component of the Monte Carlo samples $\boldsymbol{\zeta}_1, \ldots, \boldsymbol{\zeta}_N$ by the respective singular value, i.e., $x_{i,k} = \sigma_k \zeta_{i,k}$, resulting in points $\boldsymbol{x}_1, \ldots \boldsymbol{x}_N \in \times_{k=1}^{M} [-\sigma_k, \sigma_k]$.
- These points are then shifted by the choice $b_k = 2\sigma_k$ for all $k = 1, \ldots, N$ such that $X \in \mathcal{B}$.
- We perform a dimension truncation with a relative error of 10^{-3}, resulting in $d = 20$ dimensions.
- We apply the exponential kernel $\mathcal{K}(\boldsymbol{x}, \boldsymbol{y}) = \exp(\|\boldsymbol{x} - \boldsymbol{y}\|_2/\sigma)$.

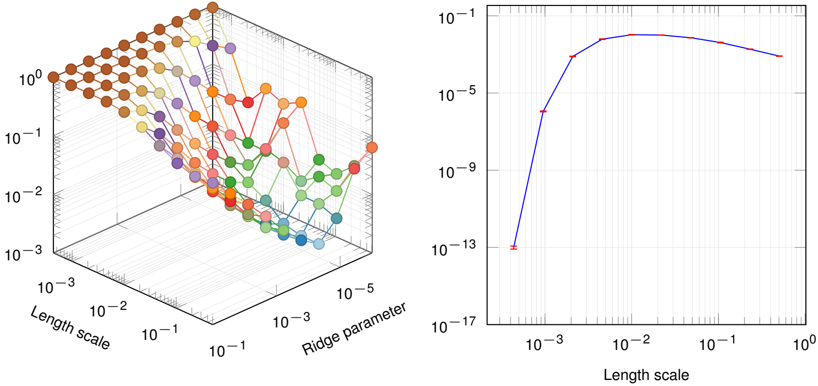


Figure: Average prediction error (left) and compression error in blue with its standard variation in red (right).

Thank you

Börm, S. (2010). *Efficient Numerical Methods for Non-Local Operators. $\mathcal{H}^2$-Matrix Compression, Algorithms and Analysis*, volume 14 of *EMS Tracts Math.* European Mathematical Society (EMS), Zürich.

DeVore, R. and Lorentz, G. (1993). *Constructive approximation*. Springer, Berlin-Heidelberg.

Haji-Ali, A.-L., Harbrecht, H., Peters, M., and Siebenmorgen, M. (2018). Novel results for the anisotropic sparse grid quadrature. *J. Complexity*, 47:62–85.

Harbrecht, H., Multerer, M., and Quizi, J. (2024). The dimension weighted fast multipole method for scattered data approximation. *arXiv preprint arXiv:2402.09531*.

Harbrecht, H., Peters, M., and Siebenmorgen, M. (2016). Analysis of the domain mapping method for elliptic diffusion problems on random domains. *Numer. Math.*, 134(4):823–856.