Stability and error guarantees for least squares approximation given noisy samples

Felix Bartel

Summer school "Analysis and Theoretical Numerics"

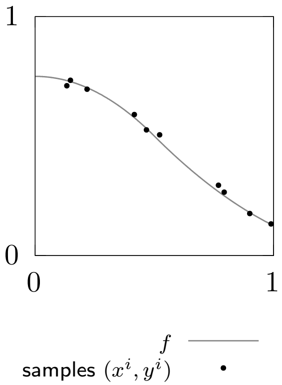
2nd of August 2022

Mathematik!
TU Chemnitz

Motivation: scattered data approximation

Given:

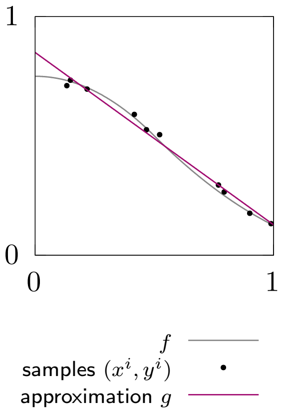
- $f\colon D \to \mathbb{C}$,
- points $x^1, \dots, x^n \in D$ distributed w.r.t. $\mathrm{d}\nu$,
- samples $\boldsymbol{y} = (\underbrace{f(x^1) + \varepsilon_1}_{y_1}, \dots, \underbrace{f(x^n) + \varepsilon_n}_{y_n})^\mathsf{T}$.



Motivation: scattered data approximation

Given:

- $f\colon D \to \mathbb{C}$,
- points $x^1, \dots, x^n \in D$ distributed w.r.t. $\mathrm{d}\nu$,
- samples $\boldsymbol{y} = (\underbrace{f(x^1) + \varepsilon_1}_{y_1}, \dots, \underbrace{f(x^n) + \varepsilon_n}_{y_n})^{\mathsf{T}}$.



Goal: find approximation $g\colon D \to \mathbb{C}$ with

- $g \in V_m = \operatorname{span}\{\eta_0, \dots, \eta_{m-1}\}$, i.e.,

$$g(x) = \sum_{k=0}^{m-1} \hat{a}_k \eta_k(x)\,,$$

- stable algorithm,
- small L_2-error, i.e., $\int_0^1 |f-g|^2 \,\mathrm{d}\mu$ small.

Motivation: domain adaptation

- source measure $\mathrm{d}\nu$ and target measure $\mathrm{d}\mu$ may differ, e.g.
  - train a self learning car
  - train a spam filter
  - diagnostic models trained on earlier diseases predicting data associated with new viruses

Motivation: domain adaptation

- source measure $\mathrm{d}\nu$ and target measure $\mathrm{d}\mu$ may differ, e.g.
  - train a self learning car
  - train a spam filter
  - diagnostic models trained on earlier diseases predicting data associated with new viruses

Radon-Nikodym theorem

Let
- μ and ν two σ-finite measures and
- ν be absolutely continuous w.r.t. μ.

Then there exists a measurable function $\varrho\colon D \to [0,\infty)$, s.t.

$$\nu(A) = \int_A \varrho \,\mathrm{d}\mu \quad \forall A \subset D\,.$$

Least squares in scattered data approximation

Ansatz: weighted least squares $g(x) = \sum_{k=0}^{m-1} \hat{g}_k \eta_k(x)$ with $m \leq n$,

$$\hat{\boldsymbol{g}} = \operatorname*{arg\,min}_{\hat{\boldsymbol{a}} \in \mathbb{C}^m} \|\boldsymbol{L}\hat{\boldsymbol{a}} - \boldsymbol{y}\|_{\boldsymbol{W}}^2 = (\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{L})^{-1}\boldsymbol{L}^*\boldsymbol{W}\boldsymbol{y}\,,$$

with

$$\boldsymbol{L} = \begin{bmatrix} \eta_0(x^1) & \dots & \eta_{m-1}(x^1) \\ \vdots & \ddots & \vdots \\ \eta_0(x^n) & \dots & \eta_{m-1}(x^n) \end{bmatrix} \in \mathbb{C}^{n\times m}, \quad \boldsymbol{W} = \begin{bmatrix} \frac{1}{\varrho(x^1)} & & \\ & \ddots & \\ & & \frac{1}{\varrho(x^n)} \end{bmatrix} \in [0,\infty)^{n\times n}$$

Q: Is weighted least squares stable?

$$A \le \sigma_{\min}(\boldsymbol{W}^{1/2}\boldsymbol{L}) \le \sigma_{\max}(\boldsymbol{W}^{1/2}\boldsymbol{L}) \le B$$

Least squares in scattered data approximation

Theorem (Least squares is stable) [Nagel, Schäfer, T. Ullrich 2021][Moeller, T. Ullrich 2021][B 2022]

Let
- $x^1, \ldots, x^n$ be distributed according to $\mathrm{d}\nu = \varrho\,\mathrm{d}\mu$,
- $V_m = \mathrm{span}\{\eta_0, \ldots, \eta_{m-1}\}$ an m-dimensional function space, $t > 0$ such that

$$6 \left\| \frac{\sum_{k=0}^{m-1} |\eta_k|^2}{\varrho} \right\|_\infty (\log(m) + t) \leq n$$

- $\boldsymbol{W} = \mathrm{diag}(1/\varrho(x^1), \ldots, 1/\varrho(x^n))$.

Then we have with probability $1 - \exp(-t)$ each

$$\frac{1}{2} \leq \sigma_{\min}^2\left(\tfrac{1}{\sqrt{n}}\boldsymbol{W}^{1/2}\boldsymbol{L}\right) \qquad \text{and} \qquad \sigma_{\max}^2\left(\tfrac{1}{\sqrt{n}}\boldsymbol{W}^{1/2}\boldsymbol{L}\right) \leq \frac{3}{2}\,.$$

Thus, we have fast convergence for iterative solvers like CG or LSQR.

Q: Can we bound the error?

$$P(f, V_m) = \underset{g \in V_m}{\arg\min} \|f - g\|^2_{L_2(D,\mu)}$$

Least squares in scattered data approximation

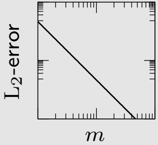
Theorem (Error bound without noise) [B 2022]

Let
- $f\colon D \to \mathbb{C}$, $x^1, \dots, x^n$, $n \in \mathbb{N}$ be points drawn according $\mathrm{d}\nu = \varrho \mathrm{d}\mu$,
- $\boldsymbol{y} = (f(x^1), \dots, f(x^n))^\mathsf{T}$ function values,
- $t \ge 0$, V_m satisfying

$$6 \left\| \frac{\sum_k |\eta_k|^2}{\varrho} \right\|_\infty (\log(m) + t) \le n \,.$$

Then, for g the weighted least squares approximation, we have with probability exceeding $1 - 2\exp(-t)$:

$$\|f - g\|_{L_2}^2 \le 8 \left(1 + \sqrt{\frac{t \| \sum_k |\eta_k|^2 \|_\infty}{n}} \right)^2 \|f - P(f, V_m)\|_{L_2}^2 \,.$$



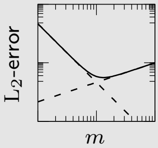
Least squares in scattered data approximation

Theorem (Error bound with noise) [B 2022]

Let
- the assumptions from the theorem before hold,
- $\boldsymbol{y} = (f(x^1) + \varepsilon_1, \dots, f(x^n) + \varepsilon_n)^\mathsf{T}$ noisy function values,
- $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)^\mathsf{T}$ mean-zero rv's with $\mathbb{E}(|\varepsilon_i|^2) \le \sigma^2$ and $\|\boldsymbol{\varepsilon}\|_\infty \le B$.

Then, for g the weighted least squares approximation, we have with probability exceeding $1 - 3\exp(-t)$:

$$\|f-g\|_{L_2}^2 \le 14 \left(1 + \sqrt{\frac{t\|\sum_k |\eta_k|^2\|_\infty}{n}}\right)^2 \|f - P(f, V_m)\|_{L_2}^2 + \|4/\varrho\|_\infty \left(\frac{m}{n}\left(14B\sqrt{t\sigma^2} + \sigma^2\right) + \frac{128B^2t}{n}\right).$$



Least squares in scattered data approximation

Theorem (continuation) [B 2022]

and

$$\|f-g\|_{L_\infty} \le \left(2+\sqrt{20\left\|\sum_k |\eta_k|^2\right\|_\infty}\right)\|f-P(f,V_m,L_\infty)\|_{L_\infty}$$
$$+\sqrt{\|2/\varrho\|_\infty\left\|\sum_k |\eta_k|^2\right\|_\infty}\sqrt{\frac{m}{n}\left(14B\sqrt{t\sigma^2}+\sigma^2\right)+\frac{128B^2t}{n}}\,.$$

Q: How do the involved quantities behave?

Function approximation exemplified on $[0, 1]$

Scenario 1

- domain $D = [0, 1]$,
- source measure $\nu(x) = \frac{\mathrm{d}x}{\sqrt{1-(2x-1)^2}}$,
- target measure $\mu(x) = \frac{\mathrm{d}x}{\sqrt{1-(2x-1)^2}}$, and
- $V_m = \mathrm{span}\{1, x, \ldots, x^{m-1}\}$.

Function approximation exemplified on $[0,1]$

We have

- $\eta_k(x) = T_k(x) = \cos(k \arccos(2x-1))$ Chebyshev polynomials,
- $\varrho \equiv 1$,
- $\sum_{k=0}^{m-1} |\eta_k(x)|^2 = \frac{m}{2} + \frac{\cos((m-1)\arccos(2x-1))\sin(m\arccos(2x-1))}{2\sqrt{1-x^2}} \leq m$,

and

Theorem [Trefethen 2018]

If $f, \ldots, f^{(s-1)}$ absolute continuous and $f^{(s)}$ of bounded variation V. Then

$$\|f - P(f, V_m)\|_{L_2([0,1], \frac{\mathrm{d}x}{\sqrt{1-(2x-1)^2}})} \leq \frac{2V}{\pi\sqrt{(s-1)}(m-s-1)^{s+1/2}} \lesssim m^{-(s+1/2)}.$$

Function approximation exemplified on $[0,1]$

Corollary

[B 2022]

Let the source and target measure be the Chebyshev measure and $V_m = \mathrm{span}\{1, x, \dots, x^{m-1}\}$. Further, let the assumptions from the theorem before hold and $6m(\log(m) + t) \le n$. Then

$$\begin{aligned} \|f - g\|_{L_2}^2 &\le 20\|f - P(f, V_m)\|_{L_2}^2 \\ &+ 4\left(\frac{m}{n}\left(14B\sqrt{t\sigma^2} + \sigma^2\right) + \frac{128B^2t}{n}\right). \end{aligned}$$

Scenario 2

- domain $D = [0, 1]$,
- source measure $\nu(x) \equiv 1$,
- target measure $\mu(x) = \frac{\mathrm{d}x}{\sqrt{1-(2x-1)^2}}$, and
- $V_m = \mathrm{span}\{1, x, \ldots, x^{m-1}\}$.

Function approximation exemplified on $[0,1]$

We have

- $\eta_k(x) = T_k(x) = \cos(k \arccos(2x-1))$ Chebyshev polynomials,
- $\varrho(x) = \sqrt{1-(2x-1)^2}$,
- $\sum_{k=0}^{m-1} |\eta_k(x)|^2 = \frac{m}{2} + \frac{\cos((m-1)\arccos(2x-1))\sin(m\arccos(2x-1))}{2\sqrt{1-x^2}} \leq m$.

Thus,

$$\|1/\varrho\|_\infty = \infty \quad \text{and} \quad \left\| \frac{\sum_{k=0}^{m-1} |\eta_k|^2}{\varrho} \right\|_\infty = \infty \,.$$

Function approximation exemplified on $[0, 1]$

"Corollary" [B 2022]

Let the source measure be uniform and target measure be the Chebyshev measure and $V_m = \mathrm{span}\{1, x, \ldots, x^{m-1}\}$. Further, let the assumptions from the theorem before hold and $\infty \leq n$. Then

$$\|f - g\|_{L_2}^2 \leq 14 \left(1 + \sqrt{\frac{tm}{n}}\right)^2 \|f - P(f, V_m)\|_{L_2}^2 + \infty \,.$$

Function approximation exemplified on $[0, 1]$

Scenario 3

- domain $D = [0, 1]$,
- source measure $\nu(x) \equiv 1$,
- target measure $\mu(x) \equiv 1$, and
- $V_m = \text{span}\{1, x, \ldots, x^{m-1}\}$.

Function approximation exemplified on $[0,1]$

We have

- $\eta_k(x) = P_k / \|P_k\|_{L_2} = \sqrt{\frac{2k+1}{2}} \frac{1}{2^k k!} \frac{\mathrm{d}^k}{\mathrm{d}x^k} (x^2-1)^k$ Legendre polynomials,
- $\varrho \equiv 1$,
- $\sum_{k=0}^{m-1} |\eta_k(0)|^2 = m^2$,

and

Theorem [Wang 2021]

If $f, \dots, f^{(s-1)}$ absolute continuous and $f^{(s)}$ of bounded variation V. Then

$$\|f - P(f, V_m)\|_{L_2([0,1],\mathrm{d}x)} \leq \frac{V}{\sqrt{\pi(s+1/2)}(m-1-s)^{s+1/2}} \lesssim m^{-(s+1/2)}.$$

Function approximation exemplified on $[0, 1]$

Corollary

[B 2022]

Let the source and target measure be uniform and $V_m = \operatorname{span}\{1, x, \dots, x^{m-1}\}$. Further, let the assumptions from the theorem before hold and $6m^2 \log(m+t) \le n$. Then

$$\|f - g\|_{L_2}^2 \le 20\|f - P(f, V_m)\|_{L_2}^2 + 4\left(\frac{m}{n}\left(14B\sqrt{t\sigma^2} + \sigma^2\right) + \frac{128B^2t}{n}\right).$$

Function approximation exemplified on $[0,1]$

Assumption: $f \in H^s$ Sobolev space with integer smoothness s, i.e.,

$$\|f\|_{H^s}^2 := \|f\|_{L_2([0,1],\mathrm{d}x)}^2 + \|f^{(s)}\|_{L_2([0,1],\mathrm{d}x)}^2 < \infty\,.$$

Function approximation exemplified on $[0,1]$

Assumption: $f \in H^s$ Sobolev space with integer smoothness s, i.e.,

$$\|f\|_{H^s}^2 := \|f\|_{L_2([0,1],\mathrm{d}x)}^2 + \|f^{(s)}\|_{L_2([0,1],\mathrm{d}x)}^2 < \infty\,.$$

Construction of V_m: For $\mathrm{Id}\colon H^s \hookrightarrow L_2([0,1],\mathrm{d}x)$ compute the eigenfunctions of $W = \mathrm{Id}^* \circ \mathrm{Id}\colon H^s \to H^s$:

$$W(f) = \sum_{k=0}^{\infty} \sigma_k \langle f, e_k\rangle_{H^s} e_k\,.$$

- e_k ONS in H^s,
- $\eta_k = \sigma_k^{-1} e_k$ ONS in $L_2([0,1],\mathrm{d}x)$,

Error estimate

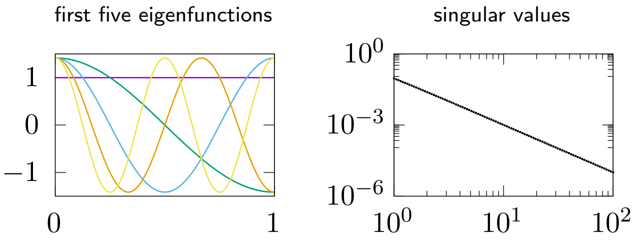
For $V_m = \mathrm{span}\{\eta_0,\ldots,\eta_{m-1}\}$ and $f = \sum_{k=0}^{\infty}\langle f, e_k\rangle_{H^s} e_k \in H^s$ we have

$$\|f - P(V_m,f)\|_{L_2([0,1],\mathrm{d}x)}^2 = \sum_{k=m}^{\infty} |\langle f, e_k\rangle_{H^s}|^2 \underbrace{\|e_k\|_{L_2}^2}_{=\sigma_k^2} \le \sup_{k\ge m} \sigma_k^2\, \|f\|_{H^s}^2\,.$$

Eigenfunctions for H^1

$W = \mathrm{Id}^* \circ \mathrm{Id} : H^1 \to H^1$ has the following singular values and eigenfunctions:

$$\eta_k(x) = \begin{cases} 1 & \text{for } k = 0 \\ \sqrt{2}\cos(\pi k x) & \text{for } k \geq 1 \end{cases} \quad \text{and} \quad \sigma_k^2 = \frac{1}{1 + \pi^2 k^2}.$$



Function approximation exemplified on $[0,1]$

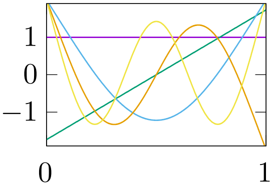
Eigenfunctions for H^2

[Iserles, Nørsett 2008][Suryanarayana, Nuyens, Cools 2016][B 2022]

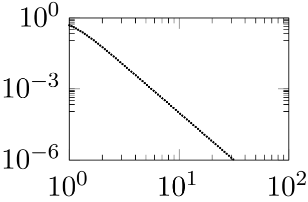
$W = \mathrm{Id}^* \circ \mathrm{Id} : H^2 \to H^2$ has the following singular values and eigenfunctions:
$\sigma_k^2 = \frac{1}{1+t_k^4}$ with $t_k > 0$ the solutions of $\cosh(t_k)\cos(t_k) = 1$ $(t_k \approx \frac{2k-1}{2}\pi)$

$$\eta_k(x) = \cosh(t_k x) + \cos(t_k x) - \frac{\cosh(t_k) - \cos(t_k)}{\sinh(t_k) - \sin(t_k)}(\sinh(t_k x) + \sin(t_k x))\,.$$

first five eigenfunctions



singular values



Function approximation exemplified on $[0,1]$

Eigenfunctions for H^2

[Iserles, Nørsett 2008][Suryanarayana, Nuyens, Cools 2016][B 2022]

$W = \mathrm{Id}^* \circ \mathrm{Id} : H^2 \to H^2$ has the following singular values and eigenfunctions:
$\sigma_k^2 = \frac{1}{1+t_k^4}$ with $t_k > 0$ the solutions of $\cosh(t_k)\cos(t_k) = 1$ $(t_k \approx \frac{2k-1}{2}\pi)$

$$\eta_k(x) = \cosh(t_k x) + \cos(t_k x) - \frac{\cosh(t_k) - \cos(t_k)}{\sinh(t_k) - \sin(t_k)}(\sinh(t_k x) + \sin(t_k x))\,.$$

Theorem: Numerically stable approximation

[B 2022]

For $k \geq 2$ let $\tilde{t}_k = \pi(2k-1)/2$ and

$$\tilde{\eta}_k(x) = \sqrt{2}\cos\left(\tilde{t}_k x + \frac{\pi}{4}\right) + \mathbb{1}_{[0,\frac{1}{2}]}(x)\exp\left(-\tilde{t}_k x\right) + \mathbb{1}_{[\frac{1}{2},1]}(x)(-1)^k \exp\left(-\tilde{t}_k(1-x)\right)\,.$$

Then $|\eta_k(x) - \tilde{\eta}_k(x)| \leq \varepsilon$ for $k \geq \frac{2}{\pi}\log(16/\varepsilon) + 1$.

Function approximation exemplified on $[0,1]$

Corollary

[B 2022]

Let the source and target measure be uniform and V_m the first eigenfunctions of W. Further, let the assumptions from the theorem before hold and $36m\log(m+t) \leq n$. Then

$$\|f-g\|_{L_2}^2 \leq 56\|f-P(f,V_m)\|_{L_2}^2 + 4\left(\frac{m}{n}\left(14B\sqrt{t\sigma^2}+\sigma^2\right)+\frac{128B^2t}{n}\right).$$

Function approximation exemplified on $[0,1]$

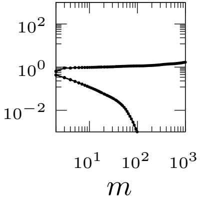
basis	assumptions	$\|f - P(f, V_m)\|_{L_2}$
Legendre	$f, \dots, f^{(s-1)}$ absolute continuous, $f^{(s)}$ bounded variation	$\mathcal{O}(m^{-(s+1/2)})$
Chebyshev		$\mathcal{O}(m^{-(s+1/2)})$
H^1-basis	$f \in H^1$	$\mathcal{O}(m^{-1})$
H^2-basis	$f \in H^2$	$\mathcal{O}(m^{-2})$

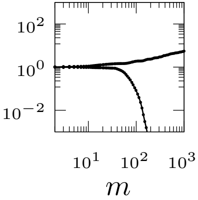
Note, for $f, \dots, f^{(s-1)}$ absolute continuous, $f^{(s)}$ bounded variation we have $f \in H^{s+1/2-\varepsilon}$.

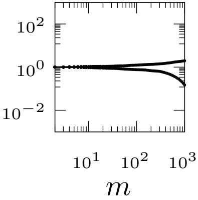
Least squares in scattered data approximation

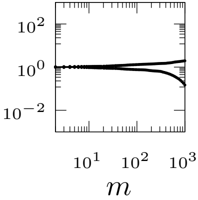
Extremal singular values of $\boldsymbol{W}^{1/2}\boldsymbol{L}$ for uniformly distributed points, i.e., $\mathrm{d}\nu = \mathrm{d}x$

Chebyshev	**Legendre**	**H^1-basis**	**H^2-basis**
$\varrho = \sqrt{1-(2x-1)^2}$	$\varrho \equiv 1$	$\varrho \equiv 1$	$\varrho \equiv 1$
$\left\lVert \frac{\sum_k \lvert\eta_k\rvert^2}{\varrho} \right\rVert_\infty \log(m)$	$\left\lVert \frac{\sum_k \lvert\eta_k\rvert^2}{\varrho} \right\rVert_\infty \log(m)$	$\left\lVert \frac{\sum_k \lvert\eta_k\rvert^2}{\varrho} \right\rVert_\infty \log(m)$	$\left\lVert \frac{\sum_k \lvert\eta_k\rvert^2}{\varrho} \right\rVert_\infty \log(m)$
$= \infty$	$= m^2 \log(m) \overset{!}{\lesssim} n$	$= 2m \log(m) \overset{!}{\lesssim} n$	$= 6m \log(m) \overset{!}{\lesssim} n$





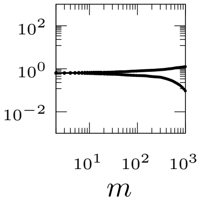
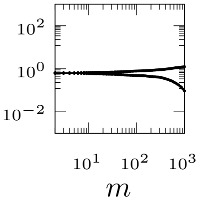
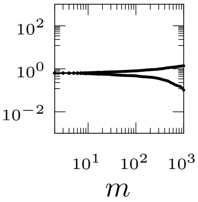
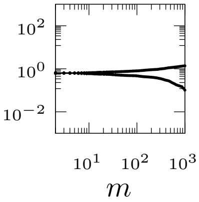




Least squares in scattered data approximation

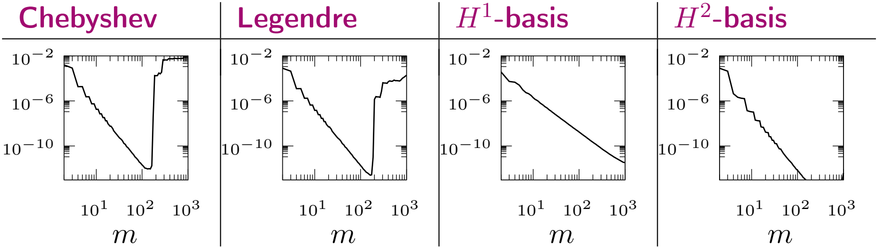
Extremal singular values of $\boldsymbol{W}^{1/2}\boldsymbol{L}$ for Chebyshev distributed points, i.e.,

$$\mathrm{d}\nu = \frac{\mathrm{d}x}{\sqrt{1-(2x-1)^2}}$$

Chebyshev	**Legendre**	**H^1-basis**	**H^2-basis**
$\varrho \equiv 1$	$\varrho = \frac{1}{\sqrt{1-(2x-1)^2}}$	$\varrho = \frac{1}{\sqrt{1-(2x-1)^2}}$	$\varrho = \frac{1}{\sqrt{1-(2x-1)^2}}$
$\left\Vert \frac{\sum_k \lvert\eta_k\rvert^2}{\varrho} \right\Vert_\infty \log(m)$	$\left\Vert \frac{\sum_k \lvert\eta_k\rvert^2}{\varrho} \right\Vert_\infty \log(m)$	$\left\Vert \frac{\sum_k \lvert\eta_k\rvert^2}{\varrho} \right\Vert_\infty \log(m)$	$\left\Vert \frac{\sum_k \lvert\eta_k\rvert^2}{\varrho} \right\Vert_\infty \log(m)$
$= 2m\log(m) \overset{!}{\lesssim} n$	$\overset{\text{[Rauhut, Ward 2012]}}{=} 2m\log(m) \overset{!}{\lesssim} n$	$= 2m\log(m) \overset{!}{\lesssim} n$	$= 6m\log(m) \overset{!}{\lesssim} n$
			

Numerical experiment

- $f \in H^{5/2-\varepsilon}$ quadratic B-spline
- $x^1, \ldots, x^n$ $10\,000$ uniformly distributed points

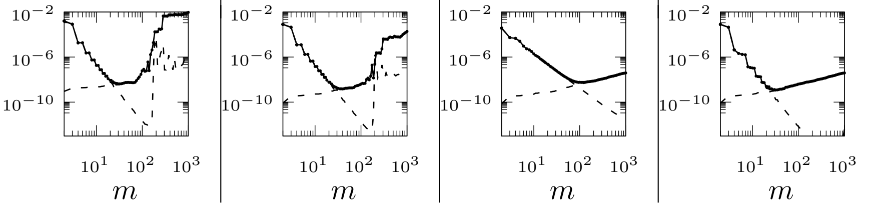


Least squares in scattered data approximation

Numerical experiment

- $f \in H^{5/2-\varepsilon}$ quadratic B-spline
- $x^1, \dots, x^n$ $10\,000$ uniformly distributed points
- 0.1% Gaussian noise in ε

Chebyshev	Legendre	H^1-basis	H^2-basis



Extension to higher dimensions

Definition

Sobolev spaces with dominating mixed smoothness on the unit cube $H^s_{\mathrm{mix}}(0,1)^d = H^s(0,1) \otimes \cdots \otimes H^s(0,1)$ with scalar product

$$\langle f, g \rangle_{H^s_{\mathrm{mix}}(0,1)^d} = \sum_{\boldsymbol{j} \in \{0,s\}^d} \langle D^{(\boldsymbol{j})} f, D^{(\boldsymbol{j})} g \rangle_{L_2} .$$

Extension to higher dimensions

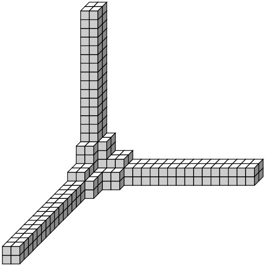
Definition

Sobolev spaces with dominating mixed smoothness on the unit cube $H^s_{\mathrm{mix}}(0,1)^d = H^s(0,1) \otimes \cdots \otimes H^s(0,1)$ with scalar product

$$\langle f, g \rangle_{H^s_{\mathrm{mix}}(0,1)^d} = \sum_{\boldsymbol{j} \in \{0,s\}^d} \langle D^{(\boldsymbol{j})} f, D^{(\boldsymbol{j})} g \rangle_{L_2}\,.$$

The singular values and eigenfunctions generalize naturally:

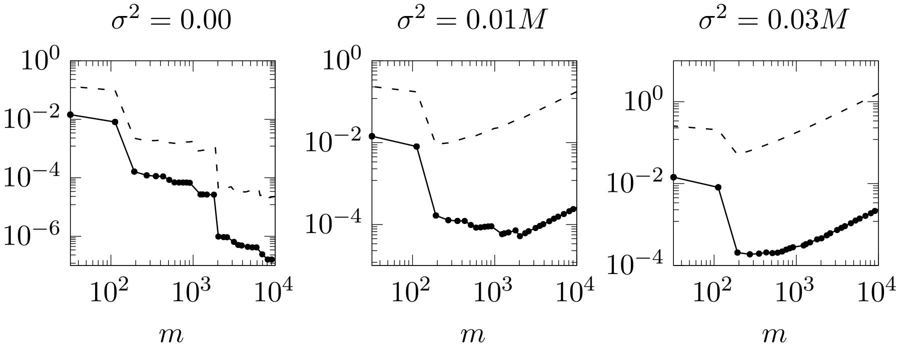
$$\sigma_{\boldsymbol{k}}^2 = \prod_{j=1}^{d} \sigma_{k_j}^2 \quad \text{and} \quad \eta_{\boldsymbol{k}}(\boldsymbol{x}) = \prod_{j=1}^{d} \eta_{k_j}(x_j)\,.$$



hyperbolic cross for $d = 3$

Numerical experiment with the H^2 basis

- $d = 5,\ f \in H_{\text{mix}}^{5/2-\varepsilon}(0,1)^5$ quadratic B-spline
- $x^1, \dots, x^n$ $1\,000\,000$ uniformly distributed points



solid: $\|f-g\|_{L_2}^2$; dashed: theoretical bound.

Conclusion and Outlook

1. Function approximation via weighted least squares for domain adaptation
   - Stability guarantees
   - Error bounds with exact function values
   - Error bounds for noisy function values
2. Exemplified the bounds on $[0, 1]$ in various settings
   - found a stable approximation for the H^2 basis

Q: How to choose the polynomial degree m when noise is involved?