High-dimensional Approximation: Transforming periodic Approximation vs. Random Fourier Features

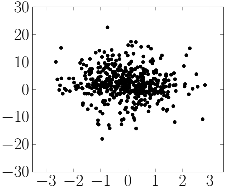
Laura Lippert
joint work with Daniel Potts, Tino Ullrich and Rachel Ward

Chemnitz University of Technology
Applied Functional Analysis

Algorithms and Complexity for Continuous Problems, Dagstuhl-Seminar 23351
01.09.2023

Given data

- sample points $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^d$ with $|\mathcal{X}| = M$, i.i.d. according to density $\mu\colon \mathbb{R}^d \to \mathbb{R}_+$
- function values $\boldsymbol{f} = (f(\boldsymbol{x}))_{\boldsymbol{x}\in\mathcal{X}}$



Aim

Approximate the function $f\colon \mathbb{R}^d \to \mathbb{C}$

Trafo approach

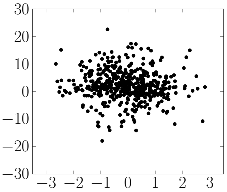
- transform the samples to the torus $\mathbb{T}^d = [-\frac{1}{2}, \frac{1}{2})^d$
- use approximation operator on $\mathbb{T}^d$

Random Fourier Features

- draw frequencies $\omega_j \in \mathbb{R}^d$ at random
- $f(\cdot) \approx \sum_j a_j \mathrm{e}^{\mathrm{i}\langle \omega_j, \cdot\rangle}$

Given data

- sample points $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^d$ with $|\mathcal{X}| = M$, i.i.d. according to density $\mu\colon \mathbb{R}^d \to \mathbb{R}_+$
- function values $\boldsymbol{f} = (f(\boldsymbol{x}))_{\boldsymbol{x}\in\mathcal{X}}$



Aim

Approximate the function $f\colon \mathbb{R}^d \to \mathbb{C}$

Trafo approach

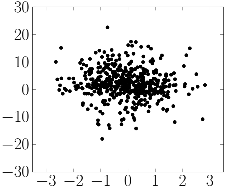
- transform the samples to the torus $\mathbb{T}^d = [-\frac{1}{2}, \frac{1}{2})^d$
- use approximation operator on $\mathbb{T}^d$

Random Fourier Features

- draw frequencies $\omega_j \in \mathbb{R}^d$ at random
- $f(\cdot) \approx \sum_j a_j \mathrm{e}^{\mathrm{i}\langle \omega_j, \cdot\rangle}$

Given data

- sample points $\boldsymbol{x} \in \mathcal{X} \subset \mathbb{R}^d$ with $|\mathcal{X}| = M$, i.i.d. according to density $\mu \colon \mathbb{R}^d \to \mathbb{R}_+$
- function values $\boldsymbol{f} = (f(\boldsymbol{x}))_{\boldsymbol{x} \in \mathcal{X}}$



Aim

Approximate the function $f \colon \mathbb{R}^d \to \mathbb{C}$

Trafo approach

- transform the samples to the torus $\mathbb{T}^d = [-\frac{1}{2}, \frac{1}{2})^d$
- use approximation operator on $\mathbb{T}^d$

Random Fourier Features

- draw frequencies $\boldsymbol{\omega}_j \in \mathbb{R}^d$ at random
- $f(\cdot) \approx \sum_j a_j \mathrm{e}^{\mathrm{i}\langle \boldsymbol{\omega}_j, \cdot \rangle}$

Outline

1. Approximation of periodic function
2. Variable Transformations
3. Approximation results on $\mathbb{R}^d$
4. Random Fourier Features
5. ANOVA decomposition

L. Lippert, D.Potts, T. Ullrich
Fast Hyperbolic Wavelet Regression meets ANOVA
Numer. Math. 154, 155-207 (2023)

L. Lippert, D.Potts
Variable Transformations in combination with Wavelets and ANOVA for high-dimensional approximation
arXiv:2108.13197, 2022

Approximation of periodic functions

given data: samples $\mathcal{X} \subset \mathbb{T}^d$, $|\mathcal{X}| = M$, function values $\boldsymbol{f} = (f(\boldsymbol{x}))_{\boldsymbol{x}\in\mathcal{X}}$

procedure:

- choose basis functions: $\psi^{\mathrm{per}}_{\boldsymbol{j},\boldsymbol{k}}(x)$ (periodized Chui-Wang wavelets of order m)
- also other basis functions are possible, e.g. $\mathrm{e}^{\mathrm{i}\langle \boldsymbol{k},\boldsymbol{x}\rangle}, \cos(\langle \boldsymbol{k},\boldsymbol{x}\rangle), \dots$
- use index-set $(\boldsymbol{j},\boldsymbol{k}) \in I_n$ with $N := |I_n| = \mathcal{O}(2^n n^{d-1})$
- choose wavelet level n according to logarithmic oversampling $M \gtrsim N \log N$
- solve minimizing problem iteratively:

$$\boldsymbol{a} = \operatorname{argmin} \|\boldsymbol{A}\boldsymbol{a} - \boldsymbol{f}\|_2\,, \quad \boldsymbol{A} = \left(\psi^{\mathrm{per}}_{\boldsymbol{j},\boldsymbol{k}}(\boldsymbol{x})\right)_{\boldsymbol{x}\in\mathcal{X},(\boldsymbol{j},\boldsymbol{k})\in I_n} \in \mathbb{R}^{M\times N}$$

- construct approximant

$$S_n^{\mathcal{X}} f(\boldsymbol{x}) := \sum_{(\boldsymbol{j},\boldsymbol{k})\in I_n} a_{\boldsymbol{j},\boldsymbol{k}} \psi^{\mathrm{per}}_{\boldsymbol{j},\boldsymbol{k}}(\boldsymbol{x})$$

Approximation of periodic functions

given data: samples $\mathcal{X} \subset \mathbb{T}^d$, $|\mathcal{X}| = M$, function values $\boldsymbol{f} = (f(\boldsymbol{x}))_{\boldsymbol{x}\in\mathcal{X}}$

procedure:

- choose basis functions: $\psi_{\boldsymbol{j},\boldsymbol{k}}^{\mathrm{per}}(\boldsymbol{x})$ (periodized Chui-Wang wavelets of order m)
- also other basis functions are possible, e.g. $\mathrm{e}^{\mathrm{i}\langle \boldsymbol{k},\boldsymbol{x}\rangle}, \cos(\langle \boldsymbol{k},\boldsymbol{x}\rangle), \dots$
- use index-set $(\boldsymbol{j},\boldsymbol{k}) \in I_n$ with $N := |I_n| = \mathcal{O}(2^n n^{d-1})$
- choose wavelet level n according to logarithmic oversampling $M \gtrsim N \log N$
- solve minimizing problem iteratively:

$$\boldsymbol{a} = \operatorname{argmin} \|\boldsymbol{A}\boldsymbol{a} - \boldsymbol{f}\|_2\,, \qquad \boldsymbol{A} = \left(\psi_{\boldsymbol{j},\boldsymbol{k}}^{\mathrm{per}}(\boldsymbol{x})\right)_{\boldsymbol{x}\in\mathcal{X},(\boldsymbol{j},\boldsymbol{k})\in I_n} \in \mathbb{R}^{M\times N}$$

- construct approximant

$$S_n^{\mathcal{X}} f(\boldsymbol{x}) := \sum_{(\boldsymbol{j},\boldsymbol{k})\in I_n} a_{\boldsymbol{j},\boldsymbol{k}}\psi_{\boldsymbol{j},\boldsymbol{k}}^{\mathrm{per}}(\boldsymbol{x})$$

Theorem ([L., Potts, Ullrich, '23])

Let $\mathcal{X} \sim$ *i.i.d. uniformly,* $M \geq r\, N \log N$, m *the order of the wavelets, if* $\frac{1}{2} < s < m$:

$$\mathbb{P}\left(\left\|f - S_n^{\mathcal{X}} f\right\|_{L_2(\mathbb{T}^d)} \lesssim 2^{-ns} n^{(d-1)/2} \, \|f\|_{\boldsymbol{B}^s_{2,\infty}(\mathbb{T}^d)}\right) \geq 1 - 2\, M^{-r}.$$

and if $m = s$

$$\mathbb{P}\left(\left\|f - S_n^{\mathcal{X}} f\right\|_{L_2(\mathbb{T}^d)} \lesssim 2^{-ns} n^{(d-1)/2} \, \|f\|_{H^s_{\mathrm{mix}}(\mathbb{T}^d)}\right) \geq 1 - 2\, M^{-r}.$$

important assumptions:

- uniformly distributed samples with $M \gtrsim N \log N$, ($N \sim 2^n n^{d-1}$)
- regularity of the function: $f \in H^s_{\mathrm{mix}}(\mathbb{T}^d)$ or $f \in \boldsymbol{B}^s_{2,\infty}(\mathbb{T}^d)$

result: error decay $\sim N^{-s} (\log N)^{(s+1/2)(d-1)}$

Theorem ([L., Potts, Ullrich, '23])

Let $\mathcal{X} \sim$ *i.i.d. uniformly,* $M \geq r\, N \log N$, m *the order of the wavelets, if* $\frac{1}{2} < s < m$:

$$\mathbb{P}\left(\left\| f - S_n^{\mathcal{X}} f \right\|_{L_2(\mathbb{T}^d)} \lesssim 2^{-ns} n^{(d-1)/2} \left\| f \right\|_{\boldsymbol{B}^s_{2,\infty}(\mathbb{T}^d)} \right) \geq 1 - 2\, M^{-r}.$$

and if $m = s$

$$\mathbb{P}\left(\left\| f - S_n^{\mathcal{X}} f \right\|_{L_2(\mathbb{T}^d)} \lesssim 2^{-ns} n^{(d-1)/2} \left\| f \right\|_{H^s_{\mathrm{mix}}(\mathbb{T}^d)} \right) \geq 1 - 2\, M^{-r}.$$

important assumptions:

- uniformly distributed samples with $M \gtrsim N \log N$, $(N \sim 2^n n^{d-1})$
- regularity of the function: $f \in H^s_{\mathrm{mix}}(\mathbb{T}^d)$ or $f \in \boldsymbol{B}^s_{2,\infty}(\mathbb{T}^d)$

result: error decay $\sim N^{-s} (\log N)^{(s+1/2)(d-1)}$

Theorem ([L., Potts, Ullrich, '23])

Let $\mathcal{X} \sim$ *i.i.d. uniformly,* $M \geq r\, N \log N$, m *the order of the wavelets, if* $\frac{1}{2} < s < m$:

$$\mathbb{P}\left(\left\|f - S_n^{\mathcal{X}} f\right\|_{L_2(\mathbb{T}^d)} \lesssim 2^{-ns} n^{(d-1)/2} \|f\|_{\boldsymbol{B}^s_{2,\infty}(\mathbb{T}^d)}\right) \geq 1 - 2\, M^{-r}.$$

and if $m = s$

$$\mathbb{P}\left(\left\|f - S_n^{\mathcal{X}} f\right\|_{L_2(\mathbb{T}^d)} \lesssim 2^{-ns} n^{(d-1)/2} \|f\|_{H^s_{\text{mix}}(\mathbb{T}^d)}\right) \geq 1 - 2\, M^{-r}.$$

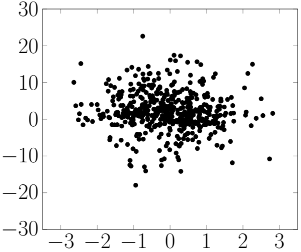
important assumptions:

- uniformly distributed samples with $M \gtrsim N \log N$, $(N \sim 2^n n^{d-1})$
- regularity of the function: $f \in H^s_{\text{mix}}(\mathbb{T}^d)$ or $f \in \boldsymbol{B}^s_{2,\infty}(\mathbb{T}^d)$

result: error decay $\sim N^{-s} (\log N)^{(s+1/2)(d-1)}$

Constructing the transformation

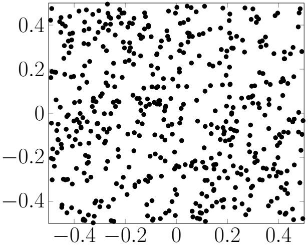
$\mathbb{R}^d$



samples $\mathcal{X}$

$\xrightarrow{\mathrm{R}}$

$\mathbb{T}^d$

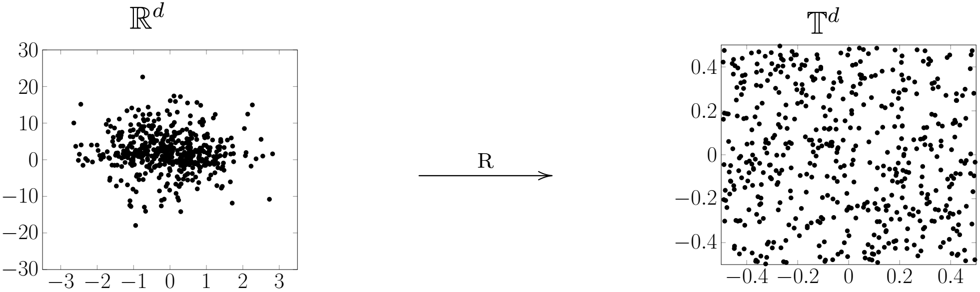


transformed samples $\mathcal{Y}$

one-dimensional case:

$$\mathrm{R}\colon \mathbb{R} \to [-\tfrac{1}{2}, \tfrac{1}{2}], \quad \mathrm{R}(x) := \int_{-\infty}^{x} \mu(t)\mathrm{d}t - \tfrac{1}{2}$$

Constructing the transformation



samples $\mathcal{X}$ transformed samples $\mathcal{Y}$

one-dimensional case:

$$\mathrm{R}\colon \mathbb{R} \to [-\tfrac{1}{2}, \tfrac{1}{2}], \quad \mathrm{R}(x) := \int_{-\infty}^{x} \mu(t)\mathrm{d}t - \tfrac{1}{2}$$

multi-dimensional case

assumptions:

- independent input variables, i.e. $\mu(\boldsymbol{x}) = \prod_{i=1}^{d} \mu_i(x_i)$

$\rightarrow$ transform every dimension separately:

$$\mathrm{R}(\boldsymbol{x}) := (\mathrm{R}_1(x_1), \dots, \mathrm{R}_d(x_d))$$

useful properties of the transformation

- there exists an inverse

$$\mathrm{R}^{-1}(\boldsymbol{y}) := (\mathrm{R}_1^{-1}(y_1), \dots, \mathrm{R}_d^{-1}(y_d))$$

- $\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}} \mathrm{R}(\boldsymbol{x}) = \mu(\boldsymbol{x})$

multi-dimensional case

assumptions:

- independent input variables, i.e. $\mu(\boldsymbol{x}) = \prod_{i=1}^{d} \mu_i(x_i)$

$\rightarrow$ transform every dimension separately:

$$\mathrm{R}(\boldsymbol{x}) := (\mathrm{R}_1(x_1), \ldots, \mathrm{R}_d(x_d))$$

useful properties of the transformation

- there exists an inverse

$$\mathrm{R}^{-1}(\boldsymbol{y}) := (\mathrm{R}_1^{-1}(y_1), \ldots, \mathrm{R}_d^{-1}(y_d))$$

- $\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}} \mathrm{R}(\boldsymbol{x}) = \mu(\boldsymbol{x})$

multi-dimensional case

assumptions:

- independent input variables, i.e. $\mu(\boldsymbol{x}) = \prod_{i=1}^{d} \mu_i(x_i)$

$\rightarrow$ transform every dimension separately:

$$\mathrm{R}(\boldsymbol{x}) := (\mathrm{R}_1(x_1), \ldots, \mathrm{R}_d(x_d))$$

useful properties of the transformation

- there exists an inverse

$$\mathrm{R}^{-1}(\boldsymbol{y}) := \left(\mathrm{R}_1^{-1}(y_1), \ldots, \mathrm{R}_d^{-1}(y_d)\right)$$

- $\frac{\mathrm{d}}{\mathrm{d}\boldsymbol{x}} \mathrm{R}(\boldsymbol{x}) = \mu(\boldsymbol{x})$

our procedure:

- transform samples $\mathcal{Y} = \mathrm{R}(\mathcal{X})$
- use approximation operator $S_n^{\mathcal{Y}}$ on $\mathbb{T}^d$
- transform back to $\mathbb{R}^d$

$$
\begin{tikzcd}
\underset{\mathcal{X}}{(\sim\mu)} \arrow[rr, "\mathrm{R}"] & & \underset{\mathcal{Y}}{(\sim\mathcal{U})}
\end{tikzcd}
$$

$$
\begin{tikzcd}
\mathbb{R}^d \arrow[rr, bend left, "\mathrm{R}"] \arrow[dr, bend right, "f"'] & & \mathbb{T}^d \arrow[ll, bend left, "\mathrm{R}^{-1}"] \arrow[dl, bend left, "f\circ\mathrm{R}^{-1}"] \\
& \mathbb{C} &
\end{tikzcd}
$$

$$\left(S_n^{\mathcal{Y}}(f \circ \mathrm{R}^{-1})\right) \circ \mathrm{R}$$

preservation of L_2-norm:

$$\|f\|_{L_2(\mathbb{R}^d,\mu)}^2 := \int_{\mathbb{R}^d} |f(\boldsymbol{x})|^2 \mu(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \|f \circ \mathrm{R}^{-1}\|_{L_2(\mathbb{T}^d)}^2$$

$\rightarrow$ measure error in $\|\cdot\|_{L_2(\mathbb{R}^d,\mu)}$

idea:

introduce function spaces $H_{\mathrm{mix}}^m(\mathbb{R}^d,\mu)$ with $\|f \circ \mathrm{R}^{-1}\|_{H_{\mathrm{mix}}^m(\mathbb{T}^d)} = \|f\|_{H_{\mathrm{mix}}^m(\mathbb{R}^d,\mu)}$

$$
\begin{tikzcd}
\underset{\mathcal{X}}{(\sim\mu)} \arrow[rr, "\mathrm{R}"] & & \underset{\mathcal{Y}}{(\sim\mathcal{U})}
\end{tikzcd}
$$

$$
\begin{tikzcd}
\mathbb{R}^d \arrow[rr, bend left, "\mathrm{R}"] \arrow[dr, bend right, "f"'] & & \mathbb{T}^d \arrow[ll, bend left, "\mathrm{R}^{-1}"] \arrow[dl, bend left, "f\circ\mathrm{R}^{-1}"] \\
& \mathbb{C} &
\end{tikzcd}
$$

our procedure:

- transform samples $\mathcal{Y} = \mathrm{R}(\mathcal{X})$
- use approximation operator $S_n^{\mathcal{Y}}$ on $\mathbb{T}^d$
- transform back to $\mathbb{R}^d$

$$(S_n^{\mathcal{Y}}(f \circ \mathrm{R}^{-1})) \circ \mathrm{R}$$

preservation of L_2-norm:

$$\|f\|^2_{L_2(\mathbb{R}^d,\mu)} := \int_{\mathbb{R}^d} |f(\boldsymbol{x})|^2 \mu(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \|f \circ \mathrm{R}^{-1}\|^2_{L_2(\mathbb{T}^d)}$$

$\to$ measure error in $\|\cdot\|_{L_2(\mathbb{R}^d,\mu)}$

idea:

introduce function spaces $H^m_{\mathrm{mix}}(\mathbb{R}^d, \mu)$ with $\|f \circ \mathrm{R}^{-1}\|_{H^m_{\mathrm{mix}}(\mathbb{T}^d)} = \|f\|_{H^m_{\mathrm{mix}}(\mathbb{R}^d,\mu)}$

$$\overset{(\sim\mu)}{\mathcal{X}} \xrightarrow{\quad \mathrm{R} \quad} \overset{(\sim\mathcal{U})}{\mathcal{Y}}$$

$$\begin{tikzcd}[column sep=huge] \mathbb{R}^d \arrow[rr, bend left, "\mathrm{R}"] \arrow[dr, "f"'] & & \mathbb{T}^d \arrow[ll, bend left, "\mathrm{R}^{-1}"] \arrow[dl, "f \circ \mathrm{R}^{-1}"] \\ & \mathbb{C} & \end{tikzcd}$$

our procedure:

- transform samples $\mathcal{Y} = \mathrm{R}(\mathcal{X})$
- use approximation operator $S_n^{\mathcal{Y}}$ on $\mathbb{T}^d$
- transform back to $\mathbb{R}^d$

$$(S_n^{\mathcal{Y}}(f \circ \mathrm{R}^{-1})) \circ \mathrm{R}$$

preservation of L_2-norm:

$$\|f\|^2_{L_2(\mathbb{R}^d,\mu)} := \int_{\mathbb{R}^d} |f(\boldsymbol{x})|^2 \mu(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} = \|f \circ \mathrm{R}^{-1}\|^2_{L_2(\mathbb{T}^d)}$$

$\rightarrow$ measure error in $\|\cdot\|_{L_2(\mathbb{R}^d,\mu)}$

idea:

introduce function spaces $H^m_{\mathrm{mix}}(\mathbb{R}^d, \mu)$ with $\|f \circ \mathrm{R}^{-1}\|_{H^m_{\mathrm{mix}}(\mathbb{T}^d)} = \|f\|_{H^m_{\mathrm{mix}}(\mathbb{R}^d,\mu)}$

Weighted function spaces (one-dimensional)

Weighted Sobolev norms on $\mathbb{R}$:

$$\|f\|^2_{H^m(\mathbb{R},\mu)} := \sum_{k=0}^{m} \left\|\mathrm{D}^k f\right\|^2_{L_2(\mathbb{R},\Upsilon_{m,k})} \text{ with } \Upsilon_{m,k}(x) := \begin{cases} \sum_{\alpha=k}^{m} |B_{\alpha,k}(x)|^2 \mu(x) & \text{if } 1 \leq k \leq m, \\ \mu(x) & \text{if } k = 0. \end{cases}$$

Weighted function spaces (one-dimensional)

Weighted Sobolev norms on $\mathbb{R}$:

$$\|f\|^2_{H^m(\mathbb{R},\mu)} := \sum_{k=0}^{m} \left\|\mathrm{D}^k f\right\|^2_{L_2(\mathbb{R},\Upsilon_{m,k})} \text{ with } \Upsilon_{m,k}(x) := \begin{cases} \sum_{\alpha=k}^{m} |B_{\alpha,k}(x)|^2 \mu(x) & \text{if } 1 \le k \le m, \\ \mu(x) & \text{if } k = 0. \end{cases}$$

Bell polynomial

Weighted function spaces (one-dimensional)

Weighted Sobolev norms on $\mathbb{R}$:

$$\|f\|^2_{H^m(\mathbb{R},\mu)} := \sum_{k=0}^{m} \left\| \mathrm{D}^k f \right\|^2_{L_2(\mathbb{R},\Upsilon_{m,k})} \quad \text{with } \Upsilon_{m,k}(x) := \begin{cases} \sum_{\alpha=k}^{m} |B_{\alpha,k}(x)|^2 \mu(x) & \text{if } 1 \le k \le m, \\ \mu(x) & \text{if } k = 0. \end{cases}$$

Bell polynomial

Examples:

$$\|f\|^2_{H^0(\mathbb{R},\mu)} = \|f\|^2_{L_2(\mathbb{R},\mu)} = \left\|f \circ \mathrm{R}^{-1}\right\|^2_{H^0(\mathbb{T}^d)}$$

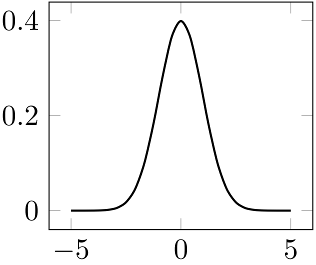
$$\|f\|^2_{H^1(\mathbb{R},\mu)} = \|f\|^2_{L_2(\mathbb{R},\mu)} + \left\|f'\right\|^2_{L_2(\mathbb{R},\frac{1}{\mu})} = \left\|f \circ \mathrm{R}^{-1}\right\|^2_{H^1(\mathbb{T}^d)}$$

$$\|f\|^2_{H^2(\mathbb{R},\mu)} = \|f\|^2_{L_2(\mathbb{R},\mu)} + \left\|f'\right\|^2_{L_2\left(\mathbb{R},\frac{1}{\mu}+\frac{(\mu')^2}{\mu^5}\right)} + \left\|f''\right\|^2_{L_2(\mathbb{R},\frac{1}{\mu^3})} = \left\|f \circ \mathrm{R}^{-1}\right\|^2_{H^2(\mathbb{T})}$$

The normal distribution

density:

$$\mu_{\mathrm{N}}(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$$



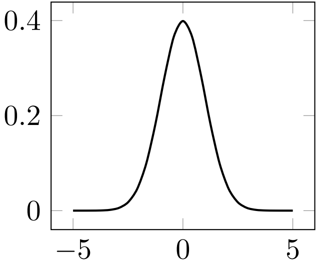
transformation:

$$\mathrm{R}(x) = \frac{1}{2} \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right), \quad \text{where} \quad \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \mathrm{e}^{-t^2} \mathrm{d}t.$$

The normal distribution

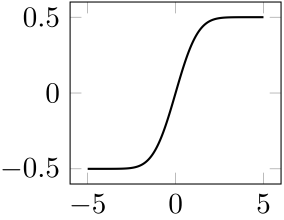
density:

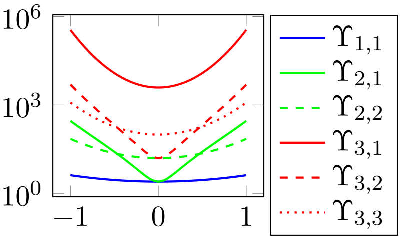
$$\mu_{\mathrm{N}}(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$$



transformation:

$$\mathrm{R}(x) = \frac{1}{2} \operatorname{erf}\left(\frac{x}{\sqrt{2}}\right), \quad \text{where} \quad \operatorname{erf}(x) = \frac{2}{\sqrt{\pi}} \int_x^{\infty} \mathrm{e}^{-t^2} \mathrm{d}t.$$





multivariate setting:

$$\|f\|^2_{H^m_{\mathrm{mix}}(\mathbb{R}^d,\mu)} = \sum_{0\le\|\boldsymbol{k}\|_\infty\le m} \left\|\mathrm{D}^{\boldsymbol{k}} f(\boldsymbol{x})\right\|^2_{L_2(\mathbb{R}^d,\Upsilon_{m,\boldsymbol{k}})} \text{ with } \Upsilon_{m,\boldsymbol{k}}(\boldsymbol{x}) := \prod_{i=1}^{d} \Upsilon_{m,k_i}(x_{k_i})$$

further weighted function spaces:

- definition for fractional smoothness via the decay of the Fourier coefficients
- transformation also for Besov regularity possible:
  define norm in $\boldsymbol{B}^s_{2,\infty}(\mathbb{R}^d,\mu)$, such that

$$\left\|f\circ \mathrm{R}^{-1}\right\|_{\boldsymbol{B}^s_{2,\infty}(\mathbb{T}^d)} \lesssim \|f\|_{\boldsymbol{B}^s_{2,\infty}(\mathbb{R}^d,\mu)},$$

  where we use characterization via Fourier coefficients of the space $\boldsymbol{B}^s_{2,\infty}(\mathbb{T}^d)$

multivariate setting:

$$\|f\|^2_{H^m_{\mathrm{mix}}(\mathbb{R}^d,\mu)} = \sum_{0\le\|\boldsymbol{k}\|_\infty\le m} \left\|\mathrm{D}^{\boldsymbol{k}} f(\boldsymbol{x})\right\|^2_{L_2(\mathbb{R}^d,\Upsilon_{m,\boldsymbol{k}})} \text{ with } \Upsilon_{m,\boldsymbol{k}}(\boldsymbol{x}) := \prod_{i=1}^d \Upsilon_{m,k_i}(x_{k_i})$$

further weighted function spaces:

- definition for fractional smoothness via the decay of the Fourier coefficients
- transformation also for Besov regularity possible:
  define norm in $\boldsymbol{B}^s_{2,\infty}(\mathbb{R}^d,\mu)$, such that

$$\left\|f \circ \mathrm{R}^{-1}\right\|_{\boldsymbol{B}^s_{2,\infty}(\mathbb{T}^d)} \lesssim \|f\|_{\boldsymbol{B}^s_{2,\infty}(\mathbb{R}^d,\mu)},$$

  where we use characterization via Fourier coefficients of the space $\boldsymbol{B}^s_{2,\infty}(\mathbb{T}^d)$

Approximation results on $\mathbb{R}^d$

Theorem ([L., Potts, '22])

Let the density μ_i be in $C^{m-1}(\mathbb{R})$ for $i \in \{1, \dots, d\}$ and let $m \in \mathbb{N}$ be the order of vanishing moments of the wavelet ψ. Let $M \gtrsim rN \log N$ ($r > 1$), $\mathcal{X} \subset \mathbb{R}^d$ be drawn i.i.d at random according to μ, $f \in C(\mathbb{R})$, $\mathcal{Y} = \mathrm{R}(\mathcal{X})$. Then

$$\mathbb{P}\left(\left\| f - (S_n^{\mathcal{Y}}(f \circ \mathrm{R}^{-1})) \circ \mathrm{R} \right\|_{L_2(\mathbb{R}^d,\mu)} \lesssim 2^{-ns} n^{(d-1)/2} \, \|f\|_{\boldsymbol{B}^s_{2,\infty}(\mathbb{R}^d,\mu)} \right) \geq 1 - 2\, M^{-r} \quad \text{if } s < m,$$

$$\mathbb{P}\left(\left\| f - (S_n^{\mathcal{Y}}(f \circ \mathrm{R}^{-1})) \circ \mathrm{R} \right\|_{L_2(\mathbb{R}^d,\mu)} \lesssim 2^{-ns} n^{(d-1)/2} \, \|f\|_{H^s_{\mathrm{mix}}(\mathbb{R}^d,\mu)} \right) \geq 1 - 2\, M^{-r} \quad \text{if } m = s.$$

Numerical example

$$\text{RMSE} = \left(\sum_{\boldsymbol{x} \in \mathcal{X}_{\text{test}}} \frac{1}{|\mathcal{X}_{\text{test}}|} |f(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x})|^2 \right)^{1/2}$$

function: $f \colon \mathbb{R}^d \to \mathbb{R}, \quad f(\boldsymbol{x}) = \mathrm{e}^{-\|\boldsymbol{x}\|_2^2}$

density:

Numerical example

$$\mathsf{RMSE} = \left(\sum_{\boldsymbol{x} \in \mathcal{X}_{\text{test}}} \frac{1}{|\mathcal{X}_{\text{test}}|} |f(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x})|^2 \right)^{1/2}$$

function: $f\colon \mathbb{R}^d \to \mathbb{R}, \quad f(\boldsymbol{x}) = \mathrm{e}^{-\|\boldsymbol{x}\|_2^2}$
density:

$$\mu_{\mathrm{N}}(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$$

Numerical example

$$\mathrm{RMSE} = \left(\sum_{\boldsymbol{x} \in \mathcal{X}_{\text{test}}} \frac{1}{|\mathcal{X}_{\text{test}}|} |f(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x})|^2 \right)^{1/2}$$

function: $f \colon \mathbb{R}^d \to \mathbb{R}, \quad f(\boldsymbol{x}) = \mathrm{e}^{-\|\boldsymbol{x}\|_2^2}$
density:

$$\mu_{\mathrm{N}}(x) = \frac{1}{\sqrt{2\pi}} \mathrm{e}^{-x^2/2}$$

- $f \in H^2(\mathbb{R}^d, \mu_{\mathrm{N}}), \quad f \notin H^3(\mathbb{R}, \mu_{\mathrm{N}})$
- $f \in \boldsymbol{B}^{5/2}_{2,\infty}(\mathbb{R}^d, \mu_N)$
- $m = 3 \to m > s$

Numerical example

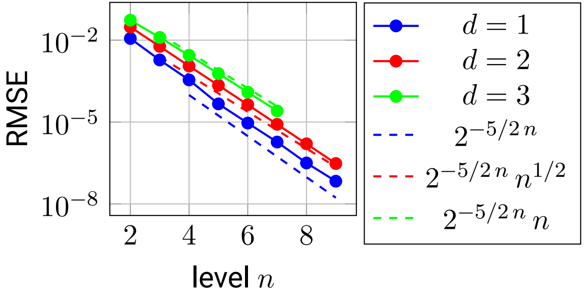
$$\text{RMSE} = \left(\sum_{\boldsymbol{x}\in\mathcal{X}_{\text{test}}} \frac{1}{|\mathcal{X}_{\text{test}}|} |f(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x})|^2 \right)^{1/2}$$

function: $f\colon \mathbb{R}^d \to \mathbb{R}, \quad f(\boldsymbol{x}) = \mathrm{e}^{-\|\boldsymbol{x}\|_2^2}$

density:

$$\mu_{\mathrm{N}}(x) = \frac{1}{\sqrt{2\pi}}\mathrm{e}^{-x^2/2}$$

- $f \in H^2(\mathbb{R}^d, \mu_{\mathrm{N}}), \quad f \notin H^3(\mathbb{R}, \mu_{\mathrm{N}})$
- $f \in \boldsymbol{B}^{5/2}_{2,\infty}(\mathbb{R}^d, \mu_N)$
- $m = 3 \to m > s$



proven error rate:

$$2^{-5/2\,n} n^{(d-1)/2}$$

Numerical example

$$\text{RMSE} = \left(\sum_{\boldsymbol{x} \in \mathcal{X}_{\text{test}}} \frac{1}{|\mathcal{X}_{\text{test}}|} |f(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x})|^2 \right)^{1/2}$$

function: $f \colon \mathbb{R}^d \to \mathbb{R}, \quad f(\boldsymbol{x}) = \mathrm{e}^{-\|\boldsymbol{x}\|_2^2}$
density:

$$\mu_{\mathrm{L}}(x) = \frac{1}{8} \mathrm{e}^{-\frac{|x-2|}{4}}$$

- $f \in H^m(\mathbb{R}^d, \mu_{\mathrm{L}})$ for all $m \in \mathbb{N}$
- $m = 3 \to m = s$

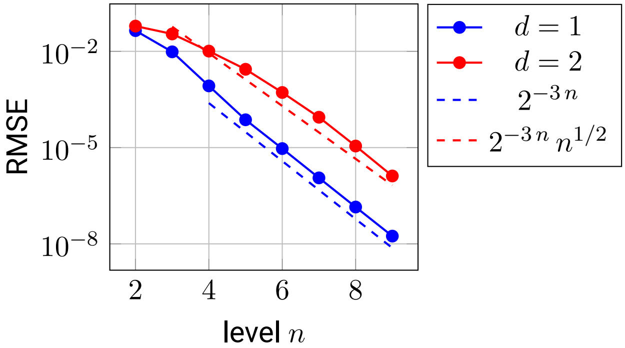
Numerical example

$$\text{RMSE} = \left(\sum_{\boldsymbol{x} \in \mathcal{X}_{\text{test}}} \frac{1}{|\mathcal{X}_{\text{test}}|} |f(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x})|^2 \right)^{1/2}$$

function: $f\colon \mathbb{R}^d \to \mathbb{R}, \quad f(\boldsymbol{x}) = \mathrm{e}^{-\|\boldsymbol{x}\|_2^2}$
density:

$$\mu_{\mathrm{L}}(x) = \frac{1}{8} \mathrm{e}^{-\frac{|x-2|}{4}}$$

- $f \in H^m(\mathbb{R}^d, \mu_{\mathrm{L}})$ for all $m \in \mathbb{N}$
- $m = 3 \to m = s$



proven error rate:

$$2^{-3n} n^{(d-1)/2}$$

Random Fourier Features

$$f(\boldsymbol{x}) \approx \sum_{j=1}^{N} a_j \mathrm{e}^{\mathrm{i}\langle \boldsymbol{\omega}_j, \boldsymbol{x} \rangle}$$

- $\boldsymbol{\omega}_j$: draw at random and keep fixed
- a_j: learn from data
- over-parametrized setting $N \gg M$
- background: approximation of a kernel κ by

$$\kappa(\boldsymbol{x}_k, \boldsymbol{x}_\ell) \approx \sum_{j=1}^{N} \mathrm{e}^{\mathrm{i}\langle \boldsymbol{\omega}_j, \boldsymbol{x}_k \rangle} \mathrm{e}^{\mathrm{i}\langle \boldsymbol{\omega}_j, \boldsymbol{x}_\ell \rangle}$$

- different algorithms so far: SRFE, SHRIMP, HARFE, . . .

A. Hashemi, H. Schaeffer, R. Shi, U. Topcu, G. Tran, R. Ward
Generalization Bounds for Sparse Random Feature Expansions
Appl. Comput. Harmon. Anal. 62, 310-330 (2023)

Y. Xie, R. Shi, H. Schaeffer, R. Ward
SHRIMP: Sparser Random Feature Models via Iterative Magnitude Pruning
Proc. Math. Sci, 190, 303–318 (2022)

E. Saha, H. Schaeffer, G. Tran
HARFE: hard-ridge random feature expansion
Sampl. Theory Signal Process. Data Anal. 21, 27 (2023)

A. Rahimi, B. Recht
Random Features for Large-Scale Kernel Machines
Adv. Neural Inf. Process. 20, (2007)

Algorithm

- draw frequencies $(\boldsymbol{\omega}_j)_{j=1}^N \subset \mathbb{R}^d$ according to density $\varrho\colon \mathbb{R}^d \to \mathbb{R}$
- construct random feature matrix $\boldsymbol{A} = \left(\mathrm{e}^{\mathrm{i}\langle \boldsymbol{\omega}_j, \boldsymbol{x}\rangle}\right)_{j=1,\boldsymbol{x}\in\mathcal{X}}^N$
- $$\boldsymbol{a}^\# = \operatorname*{argmin}_{\boldsymbol{a}} \|\boldsymbol{a}\| \text{ s.t. } \|\boldsymbol{A}\boldsymbol{a} - \boldsymbol{f}\|_2 \le \lambda$$
- optional: prune the index-set $\{1, \ldots, N\}$ iteratively
- construct approximation $f^\#(\boldsymbol{x}) = \sum_{j=1}^N a_j^\# \mathrm{e}^{\mathrm{i}\,\langle \boldsymbol{\omega}_j, \boldsymbol{x}\rangle}$

The distribution ϱ and the smoothness

$$\mathcal{F}(\varrho) := \left\{ f(\boldsymbol{x}) = \int_{\mathbb{R}^d} \hat{f}(\boldsymbol{\omega}) \mathrm{e}^{\mathrm{i}\langle \boldsymbol{\omega}, \boldsymbol{x} \rangle} \mathrm{d}\boldsymbol{\omega} \mid \|f\|_{\mathcal{F}(\varrho)} := \sup_{\boldsymbol{\omega} \in \mathbb{R}^d} \left| \frac{\hat{f}(\boldsymbol{\omega})}{\varrho(\boldsymbol{\omega})} \right| < \infty \right\}$$

- literature: Gaussian random features: $\varrho_N(\boldsymbol{\omega}) \sim \mathrm{e}^{-\frac{\|\boldsymbol{\omega}\|^2}{2\sigma^2}}$

$\rightarrow$ strong decay condition on the Fourier transform $\hat{f}$

$\rightarrow$ strong smoothness assumption on the function f

idea: using density

$$\varrho_{\Pi}^s \sim \prod_{i \in [d]} \frac{1}{\sigma \left(1 + \omega_i^2 / \sigma^2\right)^s}$$

work in progress:

$$H_{\text{mix}}^r(\mathbb{R}^d) \subseteq \mathcal{F}(\varrho_{\Pi}^s) \text{ if } s > \tfrac{1}{2}, r < 2s - \tfrac{1}{2},$$

$$\lim_{r \to 2s - 1/2} H_{\text{mix}}^r(\mathbb{R}^d) = \mathcal{F}(\varrho_{\Pi}^s).$$

In the literature so far:

- finite second moment of feature density ϱ is needed $\to$ can be relaxed

Work in progess:

- Matrix $\boldsymbol{A} = \left(\mathrm{e}^{\mathrm{i}\langle \boldsymbol{\omega}, \boldsymbol{x}\rangle}\right)_{j=1, \boldsymbol{x}\in\mathcal{X}}^{N}$ is well-conditioned w.h.p. if
  - $N \geq 2M$
  - $\sigma\gamma\sqrt{d} \gtrsim \log M$ $\qquad (\gamma, \sigma :$ scaling parameters of $\mu, \varrho)$
  - μ fulfills small ball property: $\mathbb{P}(\|\boldsymbol{x} - \boldsymbol{x}'\| > \delta) > \varepsilon$
  - $|\hat{\varrho}(\boldsymbol{x})| \lesssim \mathrm{e}^{-\|\boldsymbol{x}\|_2}$

The interpolation case

$$\boldsymbol{a}^{\#} = \arg \min_{\boldsymbol{f}=\boldsymbol{A}\boldsymbol{a}} \|\boldsymbol{a}\|_2 = \boldsymbol{A}^*(\boldsymbol{A}\boldsymbol{A}^*)^{-1}\boldsymbol{f}, \quad N \geq \frac{4}{\epsilon^2}\left(1 + \sqrt{\frac{1}{2}\log\left(\frac{1}{\delta}\right)}\right)^2,$$

$$\mathbb{P}\left(\left\|f - f^{\#}\right\|_{L_2(\mathbb{R}^d,\mu)} \lesssim \left(1 + \log\left(\frac{1}{\delta}\right)\right)^{1/4} M^{1/4}\right) \epsilon \|f\|_{\mathcal{F}(\varrho)}\Bigg) \geq 1 - \delta$$

Low-dimensional structures in high-dimensions

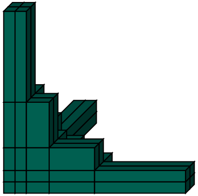
ANOVA (Analysis of variance) decomposition :

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{u} \subseteq \{1,\ldots,d\}} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}})$$

properties:

- $f_\varnothing = \int_{\mathbb{R}} f(\boldsymbol{x})\mu(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$
- $\int_{\mathbb{R}} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}}) f_{\boldsymbol{v}}(\boldsymbol{x}_{\boldsymbol{v}})\mu(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = 0, \quad \boldsymbol{u} \neq \boldsymbol{v}$

truncation: $f(\boldsymbol{x}) \approx \sum_{|\boldsymbol{u}| \leq q} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}})$



hyperbolic index-set I_3

Low-dimensional structures in high-dimensions

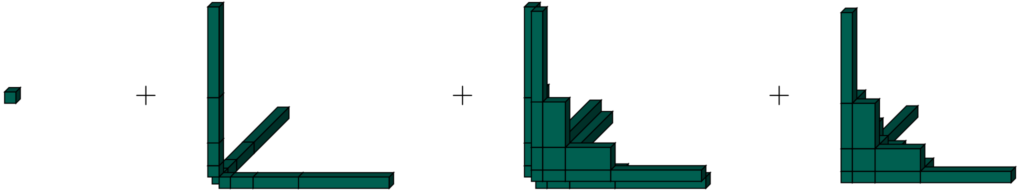
ANOVA (Analysis of variance) decomposition :

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{u} \subseteq \{1,\dots,d\}} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}})$$

properties:

- $f_{\varnothing} = \int_{\mathbb{R}} f(\boldsymbol{x})\mu(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$
- $\int_{\mathbb{R}} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}}) f_{\boldsymbol{v}}(\boldsymbol{x}_{\boldsymbol{v}})\mu(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = 0, \quad \boldsymbol{u} \neq \boldsymbol{v}$

truncation: $f(\boldsymbol{x}) \approx \sum_{|\boldsymbol{u}| \leq q} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}})$



$$I_3^{\varnothing} \quad \cup \quad I_3^{\{1\}} \cup I_3^{\{2\}} \cup I_3^{\{3\}} \quad \cup \quad I_3^{\{1,2\}} \cup I_3^{\{1,3\}} \cup I_3^{\{2,3\}} \quad \cup \quad I_3^{\{1,2,3\}}$$

Low-dimensional structures in high-dimensions

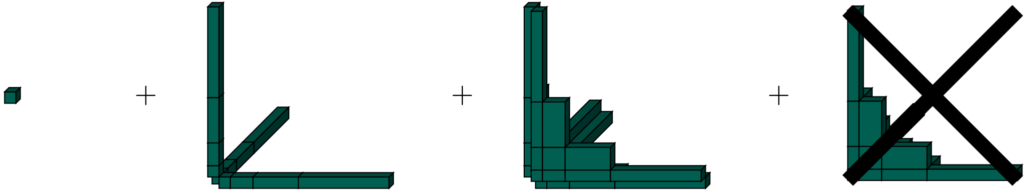
ANOVA (Analysis of variance) decomposition :

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{u}\subseteq\{1,\dots,d\}} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}})$$

properties:

- $f_\varnothing = \int_{\mathbb{R}} f(\boldsymbol{x})\mu(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$
- $\int_{\mathbb{R}} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}}) f_{\boldsymbol{v}}(\boldsymbol{x}_{\boldsymbol{v}})\mu(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = 0, \quad \boldsymbol{u} \neq \boldsymbol{v}$

truncation: $f(\boldsymbol{x}) \approx \sum_{|\boldsymbol{u}|\leq q} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}})$



$I_3^{\varnothing} \;\cup\; I_3^{\{1\}} \cup I_3^{\{2\}} \cup I_3^{\{3\}} \;\cup\; I_3^{\{1,2\}} \cup I_3^{\{1,3\}} \cup I_3^{\{2,3\}} \;\cup\; I_3^{\{1,2,3\}}$

Low-dimensional structures in high-dimensions

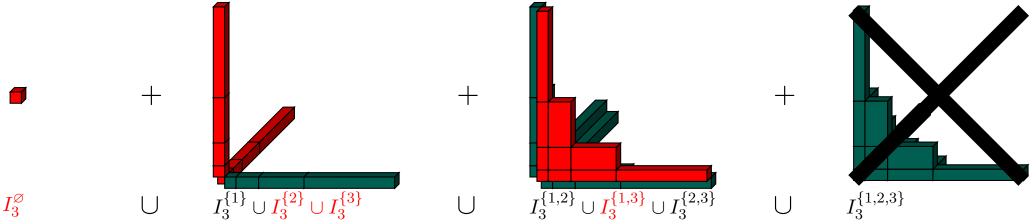
ANOVA (Analysis of variance) decomposition :

$$f(\boldsymbol{x}) = \sum_{\boldsymbol{u} \subseteq \{1,\ldots,d\}} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}})$$

properties:

- $f_\varnothing = \int_{\mathbb{R}} f(\boldsymbol{x})\mu(\boldsymbol{x})\mathrm{d}\boldsymbol{x}$
- $\int_{\mathbb{R}} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}}) f_{\boldsymbol{v}}(\boldsymbol{x}_{\boldsymbol{v}})\mu(\boldsymbol{x})\mathrm{d}\boldsymbol{x} = 0, \quad \boldsymbol{u} \neq \boldsymbol{v}$

truncation: $f(\boldsymbol{x}) \approx \sum_{|\boldsymbol{u}| \leq q} f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}})$



ANOVAapprox

$$f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}}) \approx \sum_{(\boldsymbol{j},\boldsymbol{k})\in I^{\boldsymbol{u}}} a_{\boldsymbol{j},\boldsymbol{k}} \psi^{\mathrm{per}}_{\boldsymbol{j},\boldsymbol{k}}(\mathrm{R}_{\boldsymbol{u}}^{-1}(\boldsymbol{x}_{\boldsymbol{u}}))$$

$\rightarrow$ choosing index-set with $|\operatorname{supp} \boldsymbol{j}| \leq q$

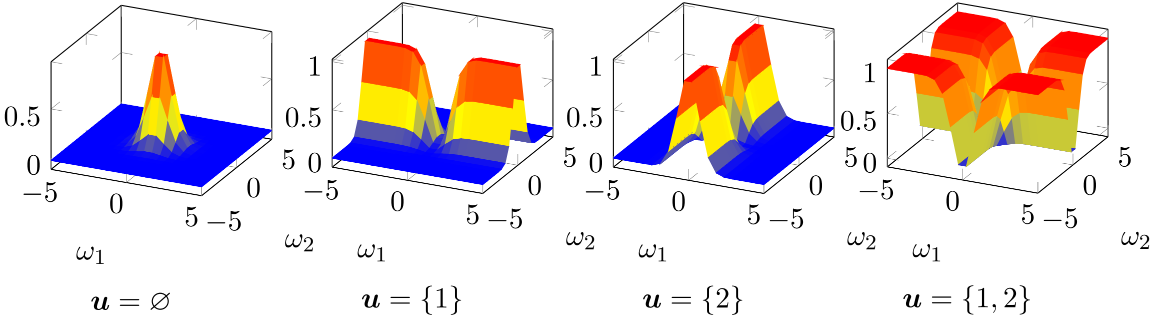
- approximate variances $\sigma^2(f_{\boldsymbol{u}})$ by $\sigma^2((S_n^{\mathcal{Y}}(f \circ \mathrm{R}^{-1}))_{\boldsymbol{u}})$ from coefficients $a_{\boldsymbol{j},\boldsymbol{k}}$
- second approximation: using only important ANOVA terms, increase accuracy for important ANOVA terms

M. Schmischke, L. Lippert, F.Nestler
NFFT/ANOVAapprox.jl: v1.1.7 (v1.1.7)
Zenodo. https://doi.org/10.5281/zenodo.7070795

L. Lippert, D. Krumm, D. Potts, S. Odenwald
Estimating vertical ground reaction forces from plantar pressure using interpretable high-dimensional approximation
submitted to Sports Eng.

Sparse Random Fourier Features

- similar idea: draw q-sparse frequencies random:
  For each $\boldsymbol{u} \subset \{1, \dots, d\}$ with $|\boldsymbol{u}| = q$ draw random $\boldsymbol{\omega}_{\boldsymbol{u}}$ and $\boldsymbol{\omega}_{\boldsymbol{u}^c} = \mathbf{0}$
- $\mathrm{e}^{\langle \boldsymbol{\omega}_j, \cdot \rangle}$ is no orthonormal system
- $f_{\boldsymbol{u}}(\boldsymbol{x}_{\boldsymbol{u}}) = \frac{1}{(2\pi)^d} \int_{\mathbb{R}^d} \hat{f}(\boldsymbol{\omega}) E(\boldsymbol{x}, \boldsymbol{\omega}, \mu, \boldsymbol{u}) \mathrm{d}\boldsymbol{\omega}$



Trafo approach

- fast multiplication with matrix $\boldsymbol{A}$ available $\to$ fast algorithm for big number of samples M
- direct connection between ANOVA terms and coefficients $a_{\boldsymbol{j},\boldsymbol{k}}$
- function spaces $H^s_{\mathrm{mix}}(\mathbb{R}^d, \mu)$

$\to$ approximation rates are transferred from $\mathbb{T}^d$ to $\mathbb{R}^d$

Random Fourier Features

- no fast algorithm available, but more parameters possible for low number of samples M (compared to dimension d)
- can be interpreted as a neural network with two layers
- function spaces $\mathcal{F}(\varrho)$
- theoretical error estimates possible, with complicated assumptions
- sensitive to the parameter choice

Thank you for your attention