The effect of sparsity and relaxations thereof in certain function approximation problems

Sebastian Mayer

Institute for Numerical Simulation, University of Bonn

Linz, June 3, 2015

Institute for Numerical Simulation
Rheinische Friedrich-Wilhelms-Universität Bonn

A (de-)motivating example

Consider the class F_d of **infinitely differentiable** functions

$$f : [0,1]^d \to \mathbb{R}$$

such that

$$\|f\| := \sup_{\alpha} \|D^{\alpha} f\|_{\infty} \leq 1.$$

Minimal deterministic worst-case error:

$$\text{error}(n,d) := \inf_{A_n} \sup_{f \in F_d} \|f - A_n f\|_{\infty}$$

A_n: deterministic algorithm **adaptively** using n arbitrary **linear functionals** (e.g. Fourier coefficients, line integrals, derivatives)

A (de-)motivating example

For any $r > 0$, we have

$$\mathrm{error}(n, d) \leq C_{r,d}\; n^{-r}.$$

Information complexity:

$$n(\varepsilon, d) := \min\{n \in \mathbb{N} : \mathrm{error}(n, d) \leq \varepsilon\}.$$

Theorem ([Novak, Woźniakowski (2009)])

For all $\varepsilon \in (0, 1)$ and all $d \in \mathbb{N}$, we have

$$n(\varepsilon, d) \geq 2^{\lfloor d/2 \rfloor}.$$

The curse of dimensionality despite of smoothness

Corollary

The L_∞-approximation of infinitely differentiable functions from F_d suffers from

the curse of dimensionality*.*

Remarkable:

- **Smoothness** does not guarantee tractability in high dimensions!
- **Asymptotic decay rates** give no indication for the complexity!

How to overcome the curse?

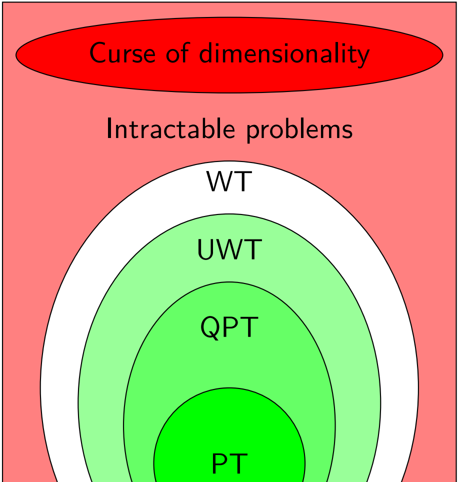
What additional **a priori** knowledge guarantees tractability?

- Sparsity?
- Structured functions?

What determines the complexity?

asymptotic decay rates vs. **preasymptotics**

Notions of tractability



Weak tractability (WT):

$$\lim_{1/\varepsilon+d\to\infty} \frac{\log n(\varepsilon, d)}{1/\varepsilon + d} = 0.$$

Uniform weak tractability (UWT):

for all $\alpha, \beta > 0$

$$\lim_{1/\varepsilon+d\to\infty} \frac{\log n(\varepsilon, d)}{1/\varepsilon^{\alpha} + d^{\beta}} = 0.$$

Quasi-polynomial tractability (QPT):

there exist constants $C, p, q > 0$ such that

$$n(\varepsilon, d) \leq C\,(1/\varepsilon)^{p(1+\log d)}\, d^{q}$$

Sobolev spaces of periodic functions

$\mathbb{T}^d = [0, 2\pi)^d$

The classical Sobolev space $H^m(\mathbb{T}^d)$ consists of all

periodic functions $f \in L_2(\mathbb{T}^d)$

such that

$$D^\gamma f \in L_2(\mathbb{T}^d) \qquad \text{for all } \gamma \in \mathbb{N}_0^d : \|\gamma\|_1 \leq m.$$

What norm to equip $H^m(\mathbb{T}^d)$ with?

A family of equivalent norms on $H^m(\mathbb{T}^d)$

The classical norm:

$$\|f\|^2 = \sum_{\|\gamma\|_1 \leq m} \|D^\gamma f\|_2^2 \asymp_m \sum_{k \in \mathbb{Z}^d} (1 + \sum_{i=1}^{d} |k_i|^2)^m \, |\hat{f}(k)|^2$$

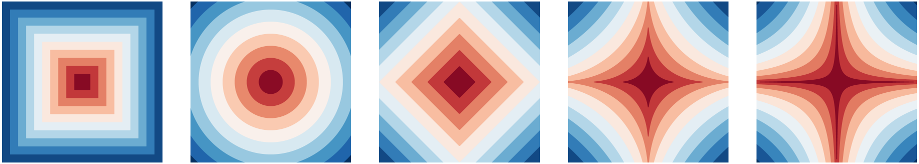
A family of equivalent norms on $H^m(\mathbb{T}^d)$

$$\|f\|_{H^{m,p}(\mathbb{T}^d)} := \left(\sum_{k \in \mathbb{Z}^d} w_{m,p}(k)^2 \, |\hat{f}(k)|^2 \right)^{1/2}$$

Weights:

$$w_{m,p}(k) := (1 + \sum_{i=1}^{d} |k_i|^p)^{m/p}$$

$$w_{m,\infty}(k) := \max\{1, |k_1|, \dots, |k_d|\}^m$$



$p = \infty$ $\quad$ $p = 2$ $\quad$ $p = 1$ $\quad$ $p = 0.5$ $\quad$ $p = 0.1$

$p < 1$: **compressibility constraint** on frequency vectors

Approximation numbers

Sobolev space of fractional smoothness $\alpha > 0$:

$$H^{\alpha,p} := \{f \in L_2(\mathbb{T}^d) : \|f\|_{H^{\alpha,p}(\mathbb{T}^d)} < \infty\}$$
$$F_d := \{f \in L_2(\mathbb{T}^d) : \|f\|_{H^{\alpha,p}(\mathbb{T}^d)} \leq 1\}$$

Minimal non-adaptive worst-case error
(aka. **approximation numbers**):

$$\begin{aligned} \mathrm{error}(n,d) &= \inf_{A_n} \sup_{\|f\|_{H^{\alpha,p}(\mathbb{T}^d)} \leq 1} \|f - A_n f\|_2 \\ &= a_{n+1}(\mathrm{Id} : H^{\alpha,p}(\mathbb{T}^d) \to L_2(\mathbb{T}^d)) \end{aligned}$$

Only non-adaptive algorithms!

Historic perspective

In 1967, J. W. Jerome proved

$$c_{\alpha,d}\ n^{-\alpha/d} \leq \mathrm{error}(n,d) \leq C_{\alpha,d}\ n^{-\alpha/d}$$

Result holds true for **any** of the norms $\|\cdot\|_{H^{\alpha,p}(\mathbb{T}^d)}$.

Where is the compressibility information gone?

All information is in the weights

$(\sigma_n)_{n\in\mathbb{N}}$: **non-increasing rearrangment** of $(1/w_{\alpha,p}(k))_{k\in\mathbb{Z}^d}$

Optimal algorithm:

- Order frequency vectors $k \in \mathbb{Z}^d$ according to $(\sigma_n)_{n\in\mathbb{N}}$:

$$k^{(1)}, k^{(2)}, k^{(3)}, \dots$$

- Take

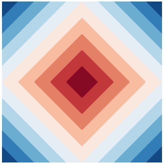
$$A_n^* f(x) = \frac{1}{(2\pi)^{d/2}} \sum_{j=1}^{n} \hat{f}(k^{(j)})\, e^{ik^{(j)}\cdot x}$$

$$\mathrm{error}(n,d) = \sup_{\|f\|_{H^{\alpha,p}(\mathbb{T}^d)}\le 1} \|f - A_n^* f\|_2 = \sigma_{n+1}$$

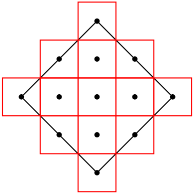
Counting via entropy

Consider, for instance, $p = 1$

$$\|f\|_{H^{\alpha,p}(\mathbb{T}^d)} = \left(\sum_{k \in \mathbb{Z}^d} (1 + \|k\|_p^p)^{2\alpha/p} \, |\hat{f}(k)|^2 \right)^{1/2}$$



Counting: $\sharp\{k \in \mathbb{Z}^d \mid \|k\|_p \le m\} \longleftrightarrow$



$1/m$-covering of B_p^d in ℓ_∞^d

Entropy numbers $\varepsilon_n(B_p^d, \ell_\infty^d)$:
minimal number of ℓ_∞^d-balls of radius ε which cover B_p^d

Characterization by entropy numbers

$$\mathrm{error}(n,d) = \inf_{A_n} \sup_{\|f\|_{H^{\alpha,p}(\mathbb{T}^d)}\le 1} \|f - A_n f\|_2$$

Theorem ([Kühn, M., Ullrich (2015)])

For all $\alpha > 0$ and all $0 < p \le \infty$, we have

$$\mathrm{error}(n,d) \asymp \varepsilon_{n+1}(B_p^d, \ell_\infty^d)^\alpha.$$

(With constants independent of n and d)

Characterization by entropy numbers

$\varepsilon_n(B_p^d, \ell_q^d)$: behavior in n and d **completely understood**!

For $p = \infty$, we obtain

$$\mathrm{error}(n,d) \asymp_{\alpha,p} \begin{cases} 1 & , \ 1 \le n \le 2^d \\ n^{-\alpha/d} & , \ n \ge 2^d \end{cases}$$

$\longrightarrow$ *curse of dimensionality*

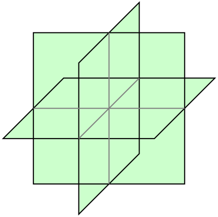
For $p < \infty$, we obtain

$$\mathrm{error}(n,d) \asymp_{\alpha,p} \begin{cases} 1 & , \ 1 \le n \le d \\ \left(\frac{\log(1+d/\log n)}{\log n}\right)^{\alpha/p} & , \ d \le n \le 2^d \\ d^{-\alpha/p} n^{-\alpha/d} & , \ n \ge 2^d \end{cases}$$

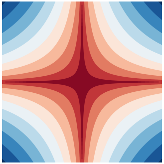
$\longrightarrow$ *weak tractability iff* $\alpha > p$ **strong compressibility** $\longrightarrow$ *almost uniform weak tractability*

Relaxations of sparsity

Sparsity

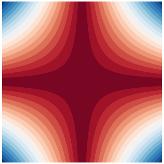


Compressibility



$w_{\alpha,p}(k) = (1 + \sum_{i=1}^{d} |k_i|^p)^{\alpha/p}$

Hyperbolic cross



$w_{\alpha,p}(k) = \prod_{i=1}^{d}(1 + |k_i|^2)^{\alpha/2}$

Sobolev spaces of dominating mixed smoothness

$$\mathrm{error}(n,d) = \inf_{A_n} \sup_{\|f\|_{H^{\alpha}_{\mathrm{mix}}(\mathbb{T}^d)}\le 1} \|f - A_n f\|_2$$

Theorem ([Kühn, Sickel, Ullrich (2015)])

For $1 \le n \le 4^d$, we have

$$\mathrm{error}(n,d) \le \left(\frac{e^2}{n}\right)^{\frac{\alpha}{2+\log d}} . \longrightarrow \textit{quasi-polyomial tractability}$$

For $n \ge 4^d$, we have

$$\mathrm{error}(n,d) \asymp n^{-\alpha}(\log n)^{\alpha(d-1)}$$

(with constants decaying super-exponentially in d).

Sparsified Sobolev spaces

$$H^{\alpha}(\mathbb{T}^d, s) := \{ f \in H^{\alpha,\infty}(\mathbb{T}^d) : \hat{f}(k) = 0 \text{ if } \operatorname{supp}(k) > s \}$$

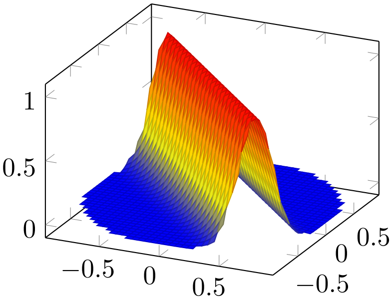
Theorem ([Kühn, M., Ullrich (2015)])

$$\operatorname{error}(n, d) \asymp \begin{cases} 1 & n \leq 2^s \binom{d}{s}, \\ d^s n^{-\alpha/s} & n > 2^s \binom{d}{s}. \end{cases}$$

$\longrightarrow$ *polynomial tractability*

(this time the **asymptotic decay** is crucial!)

Ridge functions



$$f: \quad \Omega \subseteq \mathbb{R}^d \to \mathbb{R},$$

$$f(x) = g(a \cdot x) = g(\sum_{i=1}^{d} a_i x_i).$$

$a \in \mathbb{R}^d$: **ridge direction**, g: univariate **profile**

Ridge functions on the Euclidean unit ball

For $\alpha > 0$ and $0 < p < 2$, ($p < 1$: **compressibility**)

$$\mathcal{R}_d^{\alpha,p}(B_2^d) = \left\{ g(a\cdot) : B_2^d \to \mathbb{R},\ \|g\|_{\mathrm{Lip}(\alpha)} \leq 1, \|a\|_p \leq 1 \right\}$$

$$\alpha = s + \beta, \quad \|g\|_{\mathrm{Lip}_\alpha} = \max\{\|g\|_\infty, \|g^{(1)}\|_\infty, \dots, \|g^{(s)}\|_\infty, |g^{(s)}|_\beta\}$$

Minimal deterministic worst-case sampling error
(aka. **sampling numbers**):

$$\mathrm{error}(n, d) = \inf_{S_n} \sup_{f \in \mathcal{R}_d^{\alpha,p}} \|f - S_n f\|_\infty$$

S_n: deterministic algorithm **adaptively** using n function values

Lipschitz spaces

Consider general **multivariate Lipschitz functions**

$$f : B_2^d \to \mathbb{R}.$$

Then,

$$\mathrm{error}(n, d) = \inf_{S_n} \sup_{\|f\|_{\mathrm{Lip}(\alpha)\leq 1}} \|f - S_n f\|_\infty \asymp \begin{cases} 1 & : n \leq 2^d \\ n^{-\alpha/d} & : n > 2^d \end{cases}$$

For **univariate Lipschitz functions**

$$g : [-1, 1] \to \mathbb{R},$$

it is well-known that

$$\mathrm{error}(n, d) = \inf_{S_n} \sup_{\|g\|_{\mathrm{Lip}(\alpha)\leq 1}} \|g - S_n g\|_\infty \asymp n^{-\alpha}$$

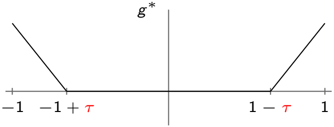
Why should recovery of ridge functions be hard?

Concentration of measure phenomenon:

$a \cdot x$ will be small with high probability for large d!

Fooling profile:

In case $\alpha = 1$, consider $f(x) = g^*(a \cdot x)$.



For n sampling points,

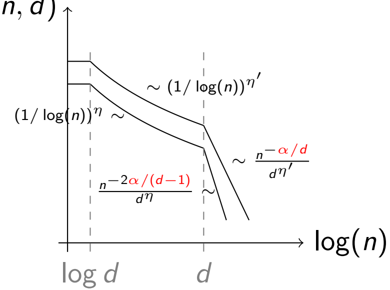
$$\tau = \varepsilon_n(\mathbb{S}_p^{d-1}, \ell_2^d)^2.$$

Characterization by entropy numbers

Theorem ([M., Ullrich, Vybíral (2014)])
For all $\alpha > 0$ and all $0 < p \leq 2$

$$\varepsilon_n(\mathbb{S}_p^{d-1}, \ell_2^d)^{2\alpha} \lesssim \mathrm{error}(n,d) \lesssim \varepsilon_{n/\binom{d+s}{s}}(B_2^d, \ell_{p'}^d)^{\alpha}$$

(p': dual index of p)



$$\eta = \alpha(1/p - 1/2)$$
$$\eta' = \alpha(1/2 - 1/p')$$

Tractability results (Euclidean ball)

Theorem ([M., Ullrich, Vybíral (2014)])

L_∞-reconstruction of ridge functions from the class $\mathcal{R}_d^{\alpha,p}(B_2^d)$ in the deterministic setting

- *suffers from the **curse of dimensionality** iff*

$$p = 2,$$

- *is **intractable** if $p < 2$ and*

$$\alpha \leq \frac{1}{1/p - 1/2},$$

- *is **weakly tractable** if $p < 2$ and*

$$\alpha > \frac{1}{1/\max\{1, p\} - 1/2},$$

Ridge functions on the cube

For $\alpha > 0$ and $0 < p \leq 1$, ($p < 1$: **compressibility**)

$$\mathcal{R}_d^{\alpha,p}([-1,1]^d) = \Big\{ g(a\cdot) : [-1,1]^d \to \mathbb{R},\ \|g\|_{\mathrm{Lip}(\alpha)} \leq 1, \|a\|_p \leq 1 \Big\}$$

Probabilistc worst-case sampling error:

$$\text{prob-error}(n,d) := \inf_{S_n} \sup_{f \in \mathcal{R}_d^\alpha} \inf\{\varepsilon > 0 : \mathbb{P}(\|f - S_n f\|_\infty \leq \varepsilon) \geq 1/2\}$$

Theorem ([Doerr, M., Rudolf])

Let $\alpha > 1$ *and* $0 < p < 1$.

1. *For* $d < n \leq \exp(d/8)/2$, *we have*

$$\left[\frac{1}{\log(n)}\right]^{\alpha(1/p-1)} \lesssim \text{prob-error}(n,d) \lesssim \left[\frac{1}{\log(n-d)}\right]^{\alpha(1/p-1)}$$

2. *For* $n \geq 2^d + d$, *we have*

$$n^{-\alpha} \lesssim \text{prob-error}(n,d) \lesssim 2^{\alpha d}(n-d)^{-\alpha}.$$

Tractability results (cube)

Theorem ([Doerr, M., Rudolf])

L_∞-reconstruction of ridge functions from the class $\mathcal{R}_d^{\alpha,p}([-1,1]^d)$ in the probabilistic setting

1. *suffers from the* ***curse of dimensionality*** *iff*

$$p = 1,$$

2. *is* ***intractable*** *iff $0 < p < 1$ and*

$$\alpha \leq \frac{1}{1/p - 1},$$

3. *is* ***weakly tractable*** *iff $0 < p < 1$ and*

$$\alpha > \frac{1}{1/p - 1}.$$

Summary

	Sobolev	ridge(ball)	ridge(cube)
CURSE	$p = \infty$	$p = 2$	$p = 1$
WT	$\alpha > p$	$\alpha > (\frac{1}{\max\{1,p\}} - \frac{1}{2})^{-1}$	$\alpha > (\frac{1}{p} - 1)^{-1}$
(almost) UWT	strong compressibility ($p \ll 1$)		
QPT	hyperbolic cross	?	
PT	sparsity	$\lvert g'(0)\rvert > \kappa$	sparsity, $\lvert g'(0)\rvert > \kappa$, $a \geq 0$

Conclusions

For high-dimensional approximation problems:

- Asymptotic rates can have a very low significance
- Complexity is often determined by **preasymptotics**
- Smoothness can be enough
- Structure does not always help

References I

A. Cohen, I. Daubechies, R. DeVore, G. Keryacharian, D. Picard
Capturing ridge functions in high dimensions from point queries
Constructive Approximation (2012) 35:225–243.

B. Doerr, S. Mayer, D. Rudolf
Reconstructing ridge functions on the cube: tractability results
Work in progress

A. Kolleck, J. Vybíral
On some aspects of approximation of ridge functions
J. Approximation Theory (2015) 194:35–61.

T. Kühn, S. Mayer, T. Ullrich
Counting via entropy: new preasymptotics for the approximation numbers of Sobolev embeddings
ArXiv e-prints arXiv:1505.00631 [math.NA], 2015

References II

T. Kühn, W. Sickel, T. Ullrich
Approximation of mixed order Sobolev functions on the d-torus—Asymptotics, preasymptotics and d-dependence
Constructive Approximation, to appear

S. Mayer, T. Ullrich, J. Vybíral
Entropy and sampling numbers of classes of ridge functions
Constructive Approximation (accepted), DOI 10.1007/s00365-014-9267-x

E. Novak, H. Woźniakowsi
Approximation of infinitely differentiable multivariate functions is intractable
J. Complexity 25:398–404, 2009.

P. Siedlecki
Uniform weak tractability
J. Complexity 29:438–453, 2012

P. Siedlecki, M. Weimar
Notes on (s,t)-weak tractability: a refined classification of problems with (sub)exponential information complexity
ArXiv e-prints arXiv:1411.3466 [math.NA], 2014

References III