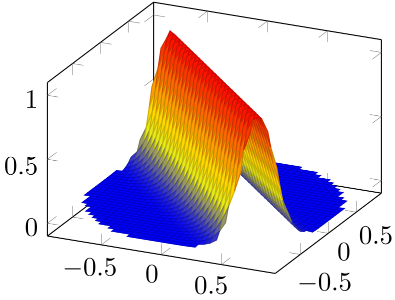
Approximation of ridge functions: tractability results

Sebastian Mayer
Universität Bonn

Joint work with T. Ullrich (U Bonn) und J. Vybíral (TU Berlin).

MCQMC 2014, Leuven, Belgium.

Ridge functions



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

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

- **ridge vector** a;
- **profile** g.

d large !

- In statistics: **single index models**, e.g. *Golubev (1992)*; *Hristache et al. (2001)*; *Gui, Li (2005)*.
- Active learning: *Cohen et al. (2012)*; *Fornasier et al. (2012)*; *Tyagi, Cevher (2012)*.

An intuitive algorithm

Given: *differentiable* ridge function $f(x) = g(a \cdot x)$ with unknown ridge vector a and profile g.

1. Find $x_0 \in \Omega$ where $g'(a \cdot x_0)$ is sufficiently large.
2. Compute **first-order differences** along all coordinates:

$$\tilde{a}_i = \frac{f(x_0 + he_i) - f(x_0)}{h}$$
$$\left(= \nabla f(x_0 + \xi_h^{(i)} e_i) \cdot e_i = g'(x_0 \cdot a + \xi_h^{(i)} a_i) a_i \right).$$

   Normalize $\hat{a} = \tilde{a} / \|\tilde{a}\|_2$.
3. Sample

$$\tilde{g} : [-1, 1] \to \mathbb{R},\ t \mapsto f(t\hat{a}) = g(ta \cdot \hat{a})$$

   using standard procedures. We obtain $\hat{g}$.
4. Put $\hat{f}(x) = \hat{g}(\hat{a} \cdot x)$.

Is Step 1 feasible?

A sampling problem: prior knowledge

Domain: the Euclidean unit ball $\Omega = \{x \in \mathbb{R}^d : \|x\|_2 \leq 1\}$.

ridge vectors: no prior knowlege, $\|a\|_2 \leq 1$.

Smoothness of profiles: Given

$$\alpha = s + \beta, \quad s \in \mathbb{N}_0, \quad 0 < \beta \leq 1.$$

For every profile $g : [-1, 1] \to \mathbb{R}$, we have

- s continuous derivatives $g^{(1)}, \ldots, g^{(s)}$,

$$\|g^{(j)}\|_\infty := \sup_{t \in [-1,1]} |g^{(j)}(t)| \leq 1.$$

- for $0 < \beta \leq 1$, the s-th derivative has **Hölder constant**

$$\left|g^{(s)}\right|_\beta \leq 1.$$

- $\|g\|_{\mathrm{Lip}(\alpha)} := \max\left\{\|g^{(0)}\|_\infty, \ldots, \|g^{(s)}\|_\infty, |g^{(s)}|_\beta\right\} \leq 1.$

A sampling problem: the setting

- **Function class:** Given $\alpha > 0$, consider

$$\mathcal{R}_d^\alpha = \left\{ g(a\cdot) : \Omega \to \mathbb{R}, \; \|g\|_{\mathrm{Lip}(\alpha)} \leq 1, \|a\|_2 \leq 1 \right\}$$

- **Algorithms:** *Adaptive*, *deterministic* approximation methods using *function values*.

***n*-th minimal worst-case error:**

$$g(n,d) := \inf \left\{ \sup_{f \in \mathcal{R}_d^\alpha} \|f - Sf\|_\infty : S \text{ using n function values } \right\}.$$

Information complexity:

$$n(\varepsilon, d) := \inf\{n \in \mathbb{N} : g(n,d) \leq \varepsilon\}$$

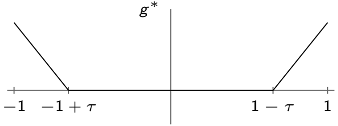
Lower bound: Construction of fooling functions

Given: Algorithm S using n function values.

Let $x_1, \dots, x_n$ the **zero function points**.

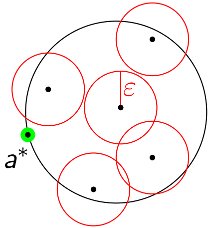
A bad profile:

For the moment: $\alpha = 1$.



$$\|g^*\|_\infty = \tau$$

A bad ridge vector:

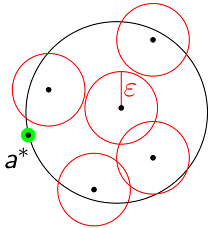


$$x_i \cdot a^* < 1 - \varepsilon^2/2$$

$$\tau = \varepsilon^2/2 \quad \Longrightarrow \quad \|g^*(a^* \cdot) - Sg^*(a^* \cdot)\|_\infty = \varepsilon^2/2.$$

$$\alpha > 0 : \quad \|g^*(a^* \cdot) - Sg^*(a^* \cdot)\|_\infty = c_\alpha \varepsilon^{2\alpha}/2.$$

Entropy numbers



What is the smallest possible ε for any S?

Entropy number:

$$\varepsilon_n(\mathbb{S}_2^{d-1}, \ell_2^d) = \inf\{\varepsilon > 0 : \exists \varepsilon\text{-balls } B_1, \ldots, B_n : \mathbb{S}_2^{d-1} \subseteq \bigcup B_i\}$$

Schütt (1984), Triebel (1997), Kühn (2001).

Upper bounds: Approximate Taylor polynomials

Given: ridge function $f = g(a\cdot) \in \mathcal{R}_d^{\alpha}$, $\alpha = s + \beta$, budget of n function evaluations

$T_{\bar{x},s} f$ **Taylor polynomial** of order s in the point $\bar{x} \in \Omega$.

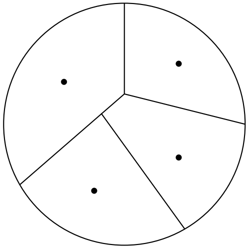
- Approximate $T_{\bar{x},s} f$ by **finite-order differences**. Takes $\binom{d+s}{s}$ function values.
- If $x \in \bar{B}_{\varepsilon}(\bar{x})$, then

$$|f(x) - T_{\bar{x},s} f(x)| \leq \|g\|_{\mathrm{Lip}(\alpha)} |a\cdot x - a\cdot\bar{x}| \leq \|x - \bar{x}\|_2^{\alpha} \leq \varepsilon^{\alpha}.$$

- Can place Taylor polynomials in

$$m = \lfloor n / \binom{d+s}{s} \rfloor$$

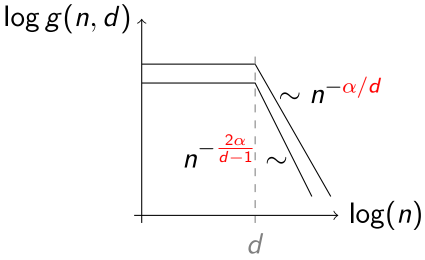
points.
- Best $\varepsilon = \varepsilon_m(\bar{B}_2^d, \ell_2^d)$.



The curse of dimensionality for ridge functions

Theorem (M., Ullrich, Vybíral '13)

$$\varepsilon_n(\mathbb{S}_2^{d-1}, \ell_2^d)^{2\alpha} \lesssim_\alpha g(n,d) \le \varepsilon_m(\bar{B}_2^d, \ell_2^d)^{\alpha}.$$



Corollary

Sampling of ridge functions from $\mathcal{R}_d^\alpha$ suffers from the curse of dimensionality.

Overcoming the curse: "sparse" ridge vectors

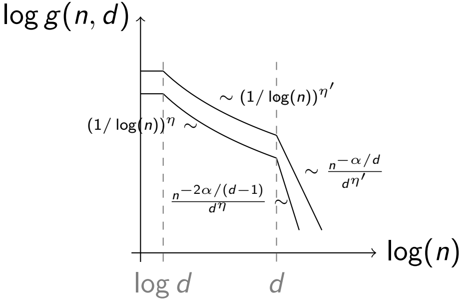
For $0 < p < 2$, consider

$$\mathcal{R}_d^{\alpha,p} := \{f = g(a\cdot) : \Omega \to \mathbb{R}, \|g\|_{\mathrm{Lip}(\alpha)}, \|a\|_p \le 1\},$$

$$\|a\|_p = \Big(\sum_{i=1}^d |a_i|^p\Big)^{1/p}.$$

Theorem (M., Ullrich, Vybíral '13)

$$\varepsilon_n(\mathbb{S}_p^{d-1}, \ell_2^d)^{2\alpha} \lesssim_{p,\alpha} g(n,d) \le \varepsilon_m(\bar{B}_2^d, \ell_{p'}^d)^{\alpha}.$$



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

Overcoming the curse: "sparse" ridge vectors

For $0 < p < 2$, consider

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

Corollary

Sampling ridge functions from $\mathcal{R}_d^{\alpha,p}$

- *never suffers from the curse of dimensionality,*
- *is intractable as long as* $\alpha \leq \frac{1}{1/p-1/2}$,
- *is weakly tractable once* $\alpha > \frac{1}{1/\max\{1,p\}-1/2}$.

Outlook: Ridge functions on the unit cube

For $\alpha > 0$ and $0 < p \leq 1$, consider

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

- **Open problem:** No lower bound so far!
- Polynomial tractability: What additional prior knowledge is necessary? *DeVore et al. (2012)*

Thank you for your attention.