Tractability of Multivariate Problems

Erich Novak

University of Jena

Plan for the talk

- Example: approximation of C^∞-functions
- What is tractability?
- Tractability by smoothness?
- Tractability by sparsity, finite order weights, or structure
- Tractability by randomization
  - Integral equations
  - Markov chain Monte Carlo
  - Optimal importance sampling

Example: approximation of C^∞-functions

Approximation of $f \in C^k([0,1]^d)$ by linear algorithms

$$S_n(f) = \sum_{i=1}^{n} L_i(f)\, g_i.$$

Optimal methods: order of convergence is $n^{-k/d}$, error in L_∞.

The order is excellent if k/d is large.
Does it mean that the problem is easy?
What about $k = \infty$?

We also will allow nonlinear methods

$$S_n = \phi \circ N \quad \text{with continuous} \quad N : C^k \to \mathbb{R}^n, \quad \phi : \mathbb{R}^n \to L_\infty.$$

A class of very smooth functions

$$F_d = \{f : [0,1]^d \to \mathbb{R} \mid \|D^\alpha f\|_\infty \le 1 \text{ for all } \alpha \in \mathbb{N}_0^d\}.$$

The class is “small”, error bounds should be “excellent”.

$$S_n = \phi \circ N \quad \text{with continuous} \quad N : F_d \to \mathbb{R}^n, \quad \phi : \mathbb{R}^n \to L_\infty,$$

$$e(S_n) = \sup_{f \in F_d} \|f - S_n(f)\|_\infty,$$

$$e(n,d) = \inf_{S_n} e(S_n), \quad n(\varepsilon, d) = \inf\{n \mid e(n,d) \le \varepsilon\}.$$

Well known: For any d and $r > 0$

$e(n,d) = \mathcal{O}(n^{-r}), \quad n(\varepsilon, d) = \mathcal{O}(\varepsilon^{-1/r}).$

Conventional conclusion: The problem is easy since the order of convergence is excellent.

Tractability

Information Complexity

$$n(\varepsilon, d) = \inf\{n \mid e(n,d) \le \varepsilon\}.$$

The problem is **strongly polynomially tractable** iff

$$n(\varepsilon, d) \le C\, \varepsilon^{-p} \quad \text{for all} \quad \varepsilon \in (0,1)\,, \quad d \in \mathbb{N}.$$

The problem is **polynomially tractable** iff

$$n(\varepsilon, d) \le C\, d^{\,q}\, \varepsilon^{-p} \quad \text{for all} \quad \varepsilon \in (0,1)\,, \quad d \in \mathbb{N}.$$

The problem is **weakly tractable** iff

$$\lim_{\varepsilon^{-1}+d\to\infty} \frac{\ln\, n(\varepsilon, d)}{\varepsilon^{-1} + d} \;=\; 0\,.$$

Introduced by Woźniakowski, 2 papers in 1994.

Result

N. & Woźniakowski, 2009

For L_∞-approximation over F_d we have

$$e(n,d) = 1 \quad \text{for all} \quad n \le 2^{\lfloor d/2 \rfloor} - 1$$

or

$$n(\varepsilon, d) \ge 2^{\lfloor d/2 \rfloor} \quad \text{for all } \varepsilon \in (0,1).$$

The problem is intractable.

Proof

Take $s = \lfloor d/2 \rfloor$ and consider $f : [0,1]^d \to \mathbb{R}$,

$$f(x) = \sum_{i \in \{0,1\}^s} a_i (x_1 + x_2)^{i_1} (x_3 + x_4)^{i_2} \ldots (x_{2s-1} + x_{2s})^{i_s}.$$

The space V_d of such functions has dimension 2^s and

$$\|f\|_\infty = \sup_\alpha \|D^\alpha f\|_\infty \quad \text{for all } f \in V_d.$$

For continuous $N : V_d \to \mathbb{R}^{2^s - 1}$, there is a $f \in V_d$ with $\|f\|_\infty = 1$ such that $N(f) = N(-f)$; follows from the Borsuk-Ulam Theorem.

Hence $S_n(f) = \phi(N(f)) = S_n(-f)$ and $e(S_n) \geq 1$ for $n = 2^s - 1$.

Tractability by smoothness assumptions?

Usually, we cannot obtain tractability even by strong smoothness assumptions, see the L_∞ approx. problem for C^∞ functions.

Sometimes: yes.

Tractability of star discrepancy

Can we compute

$$I_d(f) = \int_{[0,1]^d} f(x)\, dx$$

for $f : [0,1]^d \to \mathbb{R}$ from F_d in polynomial time, i.e.,

$$\mathrm{cost}(\varepsilon, F_d) \le C \cdot \varepsilon^{-\alpha} \cdot d^\beta \ ?$$

Star-discrepancy

$\mathrm{disc}_\infty(\{t_1,\dots,t_n\})$ of $t_i \in [0,1]^d$:

$$\sup_{x\in[0,1]^d} \left| x_1 \cdots x_d - \frac{1}{n}\sum_{i=1}^{n} 1_{[0,x)}(t_i) \right|$$

Sobolev space (or functions with bounded variation)

$F_1 = \{f : [0,1] \to \mathbb{R} \mid f(1) = 0,\ f' \in L_1\}$,

$\|f\| = \|f'\|_{L_1}$ and $F_d = F_1 \otimes \cdots \otimes F_1$.

Hlawka-Zaremba-equality yields

$$\mathrm{disc}_\infty(\{t_1,\dots,t_n\}) = \sup_{\|f\|\le 1} |I_d(f) - Q_n(f)|,$$

where $Q_n(f) = \frac{1}{n}\sum_{i=1}^{n} f(t_i)$.

The star-discrepancy is tractable

Heinrich, N., Wasilkowski, Woźniakowski (2001)

$$n(\varepsilon, F_d) \le C\, d\, \varepsilon^{-2}.$$

The dependence on d is optimal since

$$n(\varepsilon, F_d) \ge c\, d\, \log(\varepsilon^{-1}).$$

Improved lower bound

$$n(\varepsilon, F_d) \ge c\, d\, \varepsilon^{-1},$$

Hinrichs (2004).

Sparsity or partially separable functions

A function $f : [0,1]^d \to \mathbb{R}$ of many variables (d large) may be a sum of functions, that only depend on k variables (k small):

$$f(x_1, x_2, \ldots, x_d) = \sum_{\ell} g_\ell(x_{i_1}, x_{i_2}, \ldots, x_{i_k}).$$

In optimization such functions are called "partially separable".

See, e.g., N. & Ritter (1997), Dick, Sloan, Wang, Woźniakowski (2006).

Important for applications, Coulomb potential ...

As a rule:

Problems are tractable for such functions (with k fixed and $d \to \infty$), even if the g_ℓ are not very smooth.

Weighted Sobolev Spaces

Unit ball of the space $H_{d,\gamma}$ given by all $f : [0,1]^d \to \mathbb{R}$ with

$$\|f\|^2 = \sum_{\mathfrak{u} \subseteq [d]} \gamma_{d,\mathfrak{u}}^{-1} \int_{[0,1]^d} \left(\frac{\partial^{|\mathfrak{u}|}}{\partial x_{\mathfrak{u}}} f(x) \right)^2 dx \le 1 \qquad \frac{0}{0} = 0,$$

where $[d] := \{1, 2, \dots, d\}$ and $\gamma = \{\gamma_{d,\mathfrak{u}}\}$ are non-negative weights. Results for L_2 approximation for linear (or continuous) information $\Lambda^{\rm all}$ and for function values $\Lambda^{\rm std}$:

- For equal weights $\gamma_{d,\mathfrak{u}} = 1$ the problem is weakly tractable for $\Lambda^{\rm all}$ and not weakly tractable for $\Lambda^{\rm std}$.
- For bounded finite order weights ($\gamma_{d,\mathfrak{u}} = 0$ if $|\mathfrak{u}| > k$) the problem is always polynomially tractable, even for $\Lambda^{\rm std}$.

Werschulz & Woźniakowski (2009)

Various Weights

- **Product weights:** $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$**.** Then

$$H(K_{d,\gamma}) = H(K_{1,\gamma_{d,1}}) \otimes \cdots \otimes H(K_{1,\gamma_{d,d}})$$

  and $\gamma_{d,j}$ moderates the influence of x_j

- **Finite-order weights:**

$\gamma_{d,\mathfrak{u}} = 0$ **for all** $|\mathfrak{u}| > k$**.** Then

$$f = \sum_{\mathfrak{u} \subseteq [d], |\mathfrak{u}| \le k} f_{\mathfrak{u}}$$

  is a sum of functions depending on at most k variables.

We can model various properties of f by suitable weights.

Results for Integration

For product weights: $\gamma_{d,\mathfrak{u}} = \prod_{j \in \mathfrak{u}} \gamma_{d,j}$

- Strong Pol. Tract. iff $\limsup_d \sum_{j=1}^d \gamma_{d,j} < \infty$
- Pol. Tract. iff $\limsup_d \frac{\sum_{j=1}^d \gamma_{d,j}}{\ln\, d} < \infty$
- Weak Tract. iff $\lim_d \frac{\sum_{j=1}^d \gamma_{d,j}}{d} = 0$

For finite-order weights: $\gamma_{d,\mathfrak{u}} = 0$ **for all** $|\mathfrak{u}| > k$

- always polynomially tractable
- for $k \geq 1$ and $\gamma_{d,\mathfrak{u}} = 1$ for $|\mathfrak{u}| \leq k$: not strongly polynomially tractable.

N. & Woźniakowski (2001, 2010), Sloan & Woźniakowski (1998, 2002), Gnewuch & Woźniakowski (2008)

Tractability by randomization

- Integral equations
- Markov chain Monte Carlo
- Optimal importance sampling

Solving Integral Equations with Random Bits

Compute $u(s)$, integral equation

$$u(x) - \int_{[0,1]^d} k(x,y)u(y)\,\mathrm{d}y = f(x)$$

on $[0,1]^d$ with Lipschitz kernel k, $\|k\|_\infty < \alpha < 1$ and right hand side. Optimal order with MC (Heinrich & Mathé 1993) $e_n \asymp n^{-1/2-1/(2d)}$.

N. & Pfeiffer (2004): With a discretized version of classical MC and results for summation we get the upper bound

$$\mathrm{cost} \le \varepsilon^{-2} + d\,(\log \varepsilon^{-1})^2,$$

only $d\,(\log \varepsilon^{-1})^2$ random bits are needed.

Problem is intractable for deterministic algorithms.

Markov chain Monte Carlo

Computation of $E_\pi(f)$ (expectation with respect to π) with a Markov chain Monte Carlo method and burn in n_0,

$$A_{n,n_0}(f) = \frac{1}{n} \sum_{k=1}^{n} f(X_{k+n_0}),$$

when it is not possible to simulate π directly.

We assume that the Markov chain (X_k) is reversible and has an L_2-spectral gap,

$$\beta = \|P - E_\pi\|_{L_2 \to L_2} < 1.$$

Then it is known (ergodic theorem) that $A_{n,n_0}(f) \to E_\pi(f)$.

Error bounds? How should we choose the burn in?

Result of Rudolf 2009

For $f \in L_p(\pi)$ with $p \geq 4$ and

$$n_0 \geq (1-\beta)^{-1} \log \left(\left\| \frac{\mathrm{d}\mu}{\mathrm{d}\pi} - 1 \right\|_\infty \right)$$

the error is bounded by

$$\sup_{\|f\|_p \leq 1} e_\mu(A_{n,n_0}, f)^2 \leq \frac{2}{n(1-\beta)} + \frac{46}{n^2(1-\beta)^2}.$$

Here μ is the initial distribution, i.e., the distribution of X_1.

The cost bound does not depend on d, but $\beta = \beta(d)$ might depend on d. One obtains different tractability results depending on the behavior of $\beta = \beta(d)$.

Explicit error bounds and a recipe for the choice of n_0.

Optimal importance sampling

$I(f) = \int_{\mathbb{R}^d} f(x)\varrho(x)\,\mathrm{d}x$ for $f \in H$, H a RKHS with

$$\|I\|^2 = \int_{\mathbb{R}^d}\int_{\mathbb{R}^d} K(x,y)\varrho(x)\varrho(y)\,\mathrm{d}x\,\mathrm{d}y < \infty.$$

Randomized error $e(A_n) = \sup_{\|f\|_H \le 1} (E(I(f) - A_n(f))^2)^{1/2}$.

Hinrichs 2010: If $K(x,y) \ge 0$ then with importance sampling

$$e(A_n) \le \left(\frac{\pi}{2}\right)^{1/2} n^{-1/2}\,\|I\|.$$

Hence such problems are strongly polynomially tractable.

N. and Woźniakowski (2010): under some additional assumptions, the algorithm of Hinrichs is optimal.

Summary

Many problems for functions $f : [0,1]^d \to \mathbb{R}$ are intractable, if considered in the worst case setting for classical function spaces, like $C^k([0,1]^d)$.

Remedies:

- Weighted spaces, problems with a structure
- Randomized algorithms