Theoretical Limits of Approximation and Optimization for Continuous Black Box Functions

Erich Novak

Friedrich-Schiller University Jena

Lorentz Center, Leiden, 29. Feb. - 4. March 2016

Plan for the talk

- Formulation of the problem
  - What is a black box function?
  - What is Information Based Complexity?
- Example: C^∞-functions
- What is tractability?
- Tractability by smoothness?
- Tractability by structure
- Tractability by randomization

What is a black box function?

A black box function

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

is an unknown function; it is assumed that we can compute function values

$$f(x_1), f(x_2), \ldots, f(x_n)$$

or values of linear functionals: Fourier coefficients, wavelet coefficients, and so on.

We want to compute something, for example

$$S(f) = \max f,$$

but we have incomplete information on f.

What is Information-Based Complexity?

Again we have a black box function $f : [0,1]^d \to \mathbb{R}$ and we want to compute something, for example $S(f) = \max f$.

How many values

$$f(x_1), f(x_2), \dots, f(x_n)$$

do we need to compute $S(f)$ up to an error $\varepsilon > 0$?

Makes only sense if we know that $f \in F$ for a given class F of "possible" functions; the class F is the input class or problem class for our algorithms.

Information complexity

$$n(\varepsilon, S, F)$$

is the smallest n needed to solve the problem up to an error ε.

A simple example

Assume that $f : [0,1] \to \mathbb{R}$ ist Lipschitz,

$$F = \{f : [0,1] \to \mathbb{R} \mid |f(x) - f(y)| \le |x - y|\},$$

we want to compute $S(f) = \max f$ using n function values. The optimal x_i are equidistant, $x_i = \frac{2i-1}{2n}$. The true value of $S(f)$ is in the interval

$$[\max_i f(x_i), \ \max_i f(x_i) + 1/(2n)].$$

An optimal algorithm is

$$A_n^*(f) = \max_i f(x_i) + \frac{1}{4n},$$

the worst case error is $\frac{1}{4n}$. Obtain $\quad n(\varepsilon, S, F) = \lceil 1/(4\,\varepsilon) \rceil$.

Curse of Dimensionality (Bellman)

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

$$F = \{f : [0,1]^d \to \mathbb{R} \mid \; |f(x) - f(y)| \le \max_i |x_i - y_i|\},$$

we want to compute $S(f) = \max f$ using n function values. For $n = m^d$, the optimal x_i form a grid. The true value of $S(f)$ is in the interval

$$[\max_i f(x_i), \; \max_i f(x_i) + 1/(2m)].$$

An optimal algorithm is

$$A_n^*(f) = \max_i f(x_i) + \frac{1}{4m},$$

the worst case error is $\frac{1}{4m}$. Obtain $\quad n(\varepsilon, S, F) = \lceil 1/(4\,\varepsilon) \rceil^d$.

Computing time increases exponentially with dimension.

Approximation or Optimization of C^k-functions

Consider $f \in C^k([0,1]^d) = F$.

Optimal methods: order of convergence of the error ε is $n^{-k/d}$, or $n(\varepsilon, d) \asymp \varepsilon^{-d/k}$.

The order is excellent if k/d is large.

Does it mean that the problem is easy?

What about $k = \infty$?

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". Let $e(n,d) = \inf_{S_n} e(S_n)$ be the (worst case) error of an optimal algorithm using n function values,

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

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

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

The problem is **polynomially tractable** iff

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

The problem is **weakly tractable** iff

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

Result for C^∞ functions

N. & Woźniakowski, 2009

For optimization (or L_∞-approximation) over F_d we have

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

or

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

The problem is intractable.

Proof

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

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

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 any information $N : V_d \to \mathbb{R}^{2^s - 1}$, there is a $f^* \in V_d$ with $\max f^* = 1$ such that $N(f^*) = 0$.

This f^* cannot be distinguished from $f_0 = 0$ ("principle of incomplete information") and hence the error is at least $1/2$ for any algorithm based on N.

An open problem

Consider the same class

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

and numerical integration,

$$S_d(f) = \int_{[0,1]^d} f(x) \, \mathrm{d}x.$$

Is this problem tractable? Curse of dimension?

For the larger classes

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

we have the curse of dimension, see

Hinrichs, N., Ullrich, Woźniakowski (2013).

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

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

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

$\mathrm{disc}_\infty(\{t_1, \ldots, t_n\})$ 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

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

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

The star-discrepancy is tractable

Heinrich, N., Wasilkowski, Woźniakowski (2001)

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

The dependence on d is optimal since

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

Hinrichs (2004).

A kind of structure: 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, Werschulz, Woźniakowski 2009

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 $[d] := \{1, 2, \ldots, d\}$ 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 $\Lambda^{\rm std}$.

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:**

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

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

  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)

Rank One Tensors

N., Rudolf 2016

Assume that $f : [0,1]^d \to \mathbb{R}$ is a rank one tensor,

$$f(x_1, x_2, \ldots, x_d) = \prod_{i=1}^{d} f_i(x_i)$$

with

$$\|f_i\|_\infty \leq 1, \quad \|f_i^{(r)}\|_\infty \leq M.$$

The problem of approximation or optimization should be easy since f is given by d univariate functions. But:

The curse of dimensionality is present if and only if $M \geq 2^r r!$.

For smaller M, we construct a randomized algorithm that, for fixed $\varepsilon > 0$, has polynomial (in d) cost.

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:

- Problems with a structure (weighted spaces)
- Randomized algorithms, mainly if you want to compute integrals