

# Some old and new results on Information-Based Complexity

Erich Novak

Friedrich-Schiller University Jena

INI February 2019

# Plan for the first talk

- What is Information Based Complexity?
- Example:  $C^\infty$ -functions
- What is tractability?
- Tractability by smoothness?
- Tractability by structure
- Tractability by randomization

## Black box or oracle

We assume that an input function  $f : [0, 1]^d \rightarrow \mathbb{R}$  is unknown, like a black box. We can compute function values

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

or values of linear functionals like Fourier coefficients.

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 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. The Information Complexity

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

is the smallest  $n$  needed to solve the problem up to an error  $\varepsilon$ .

## A simple example

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

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

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  $\frac{1}{4n}$ . We obtain  $n(\varepsilon, S, F) = \lceil 1/(4\varepsilon) \rceil$ .

## Curse of Dimensionality (Bellman)

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

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

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  $\frac{1}{4m}$ . We obtain  $n(\varepsilon, S, F) = \lceil 1/(4\varepsilon) \rceil^d$ .

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  $\varepsilon$  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

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

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,

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

**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.



## Information Complexity

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

The problem is **strongly polynomially tractable** iff

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

The problem is **polynomially tractable** iff

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

The problem is **weakly tractable** iff

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

## Result for $C^\infty$ functions

N. & Woźniakowski, 2009

For optimization (or  $L_\infty$ -approximation) over  $F_d$  we have

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

or

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

The problem is intractable, curse of dimensionality.

## Proof

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

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

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  $N : V_d \rightarrow \mathbb{R}^{2^s-1}$ , 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

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

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

Is this problem tractable? Curse of dimension?

For the larger classes

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

we have the curse of dimension, see

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

## A kind of structure: partially separable functions

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

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

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  $d \rightarrow \infty$ ), even if the  $g_{\ell}$  are not very smooth.

## Structure: Weighted Spaces, Werschulz, Woźniakowski 2009

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

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

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

- Equal weights  $\gamma_{d,\mathbf{u}} = 1$ : the problem is weakly tractable for  $\Lambda^{\text{all}}$ ; for  $\Lambda^{\text{std}}$  we have the curse of dimension even for integration.
- For bounded finite order weights ( $\gamma_{d,\mathbf{u}} = 0$  if  $|\mathbf{u}| > k$ ) the problem is always polynomially tractable, even for  $\Lambda^{\text{std}}$ .

## Various Weights

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

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

$$f = \sum_{u \subseteq [d], |u| \leq k} f_u$$

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:**  $\gamma_{d,u} = \prod_{j \in u} \gamma_{d,j}$

- 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:**  $\gamma_{d,u} = 0$  for all  $|u| > k$

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

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



## Yet another structure: Rank One Tensors

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

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

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  $M$ , there is a det. algorithm that, for fixed  $\varepsilon > 0$ , has polynomial (in  $d$ ) cost. N., Rudolf (2016), Krieg, Rudolf (2019)

# Summary

Many problems for functions  $f : [0, 1]^d \rightarrow \mathbb{R}$  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 and more)
- Randomized algorithms, mainly if you want to compute integrals And also for  $L_\infty$  approximation. But we did not discuss randomization in this lecture.