Some old and new results on Information-Based Complexity

part two

Erich Novak

Friedrich-Schiller University Jena

INI February 2019

Second talk: Complexity results for integration

- For a d-variate function $f : [0,1]^d \to \mathbb{R}$, approximate the integral
$$S_d(f) = \int_{[0,1]^d} f(x)\,\mathrm{d}x.$$
- **Problem instances:** functions f from a function class F_d of d-variate functions; often we have an initial error
$$\max_{f \in F_d} |S_d(f)| = 1.$$
- **Algorithms:** deterministic or randomized algorithms using n function values.

The problem

- **Information complexity:** $n(\varepsilon, F_d)$ is the minimal number of function values needed by an optimal algorithm to approximate S_d up to error $\varepsilon < 1$ for all $f \in F_d$.
- **Optimal error bounds:** Sometimes more convenient:

$$e(n, F_d) = \inf_{A_n} \sup_{f \in F_d} |S_d(f) - A_n(f)|$$

$=$ minimal error achievable with n function values.

What is known about the numbers $n(\varepsilon, F_d)$ or $e(n, F_d)$?

Plan for the talk

- Some classical results (≤ 1971)
- Randomized algorithms
- Tensor product problems, tractability, weighted spaces
- Curse of dimensionality for C^k functions

Classical results

- Optimality of linear algorithms
- Optimal order for C^k functions
- Lipschitz functions and the curse of dimensionality

Optimality of linear algorithms

Smolyak 1965 and Bakhvalov 1971

Theorem

Assume that F_d is convex and symmetric. Then non-linear or adaptive algorithms cannot be better than linear algorithms

$$A_n(f) = \sum_{i=1}^{n} a_i f(x_i)$$

and

$$e(n, F_d) = \inf_{x_1,\ldots,x_n} \sup_{f \in F_d,\, f(x_i)=0} S_d(f).$$

See Hickernell, Plaskota, Wasilkowski and others for different F_d.

Smooth functions

$$F_d = C_d^k = \{f \in C^k([0,1]^d) \;\mid\; \|D^\beta f\|_\infty \le 1 \text{ for all } \|\beta\|_1 \le k\}$$

Theorem (Bakhvalov 1959)

$$e(n, C_d^k) \asymp n^{-k/d}$$

or: for all $\varepsilon \in (0,1)$ *and* $d \in \mathbb{N}$

$$a_{d,k}\,\varepsilon^{-d/k} \le n(\varepsilon, C_d^k) \le b_{d,k}\,\varepsilon^{-d/k}.$$

Lower bound trivial for fixed $\varepsilon > 0$ and large d.

Arbitrary Lipschitz domains: N. and Triebel 2006.

Fractals: Dereich, Müller-Gronbach 2014.

Lipschitz functions

$$F_d = \{f \in C([0,1]^d) \;\mid\; |f(x) - f(y)| \le \|x - y\|_\infty\}$$

Theorem (Maung Zho Newn and Sharygin 1971)

$$e(n, F_d) = \frac{d}{2d+2} \cdot n^{-1/d}$$

for $n = m^d$ *with* $m \in \mathbb{N}$.

Observe that

$$e(2^d, F_d) = \frac{1}{2}\, e(1, F_d).$$

Curse of dimension.

Similar results: Babenko, Sukharev, Chernaya

Randomized algorithms

- Optimal algorithm for L_p, $p \geq 2$
- Optimal order for C^k functions

Optimal algorithm for L_p, $p \geq 2$

Error of A_n now

$$e(A_n) = \sup_{f \in F_d} (\mathbb{E}(S(f) - A_n(f))^2)^{1/2},$$

numbers $n^{\mathrm{ran}}(\varepsilon, F_d)$ and $e^{\mathrm{ran}}(n, F_d)$ defined as before.

Let F_d be the unit ball of $L_p([0,1]^d)$ with $2 \leq p \leq \infty$.

Theorem (Mathé 1995)

$$e^{\mathrm{ran}}(n, F_d) = \frac{1}{1+\sqrt{n}}$$

and $A_n(f) = \frac{1}{n+\sqrt{n}} \sum_{i=1}^n f(X_i)$ *with i.i.d.* X_i *is optimal.*

Smooth functions

$$F_d = C_d^k = \{f \in C^k([0,1]^d) \ \mid \ \|D^\beta f\|_\infty \le 1 \ \text{ for all } \ \|\beta\|_1 \le k\}$$

Theorem (Bakhvalov 1959)

$$e^{\mathrm{ran}}(n, C_d^k) \asymp n^{-k/d-1/2}$$

Algorithm:

1) For $n = 2m$, use m values for a good L_2-approximation f_m.

2) Compute integral of $f - f_m$ by simple Monte Carlo.

Advantages: Small error, even deterministic. Good error control, unbiased estimator.

Tensor product problems, tractability, weighted spaces

- Optimal order for $W_p^{k,\mathrm{mix}}([0,1]^d)$
- Smolyak algorithm
- Tractability
- Tractability for unweighted problems
- QMC for RKHS and weighted spaces
- Decomposable kernels and lower bounds

Optimal order for $W_p^{k,\mathrm{mix}}([0,1]^d)$

$1 < p < \infty$ and $k, d \in \mathbb{N}$ fixed. $\|D^\alpha f\|_p \le 1$ for $\|\alpha\|_\infty \le k$.

Theorem (upper bound: Frolov 1976, Skriganov 1994, lower bound: Roth 1954, Bykovskii 1985, Temlyakov 1990)

$$e(n, W_p^{k,\mathrm{mix}}([0,1]^d)) \asymp n^{-k} (\log n)^{(d-1)/2}.$$

First step: cubature formulas of the form

$$A_n(f) = \frac{|\det A|}{a^d} \sum_{m \in \mathbb{Z}^d} f\left(\frac{Am}{a}\right)$$

for functions with compact support; A does not depend on k.

Second step: transformation for the general case.

Smolyak algorithm / sparse grids

Theorem

Upper bounds for $W_p^{k,\mathrm{mix}}([0,1]^d)$ *and the Smolyak algorithm are almost optimal, but not quite.*

Smolyak 1963 and many others.

Dinh Dũng, Tino Ullrich 2014: $n^{-k}(\log n)^{(d-1)(k+1/2)}$ for $p = 2$.

Tractability

A problem is **strongly polynomially tractable** iff

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

A problem is **polynomially tractable** iff

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

A 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.

Tractability by smoothness assumptions?

Usually, we cannot obtain tractability even by strong smoothness assumptions.

Sometimes: yes.

- Tractability of star discrepancy

Star-discrepancy

For $x_1, \ldots, x_n \in [0,1]^d$ defined by

$$D_\infty^*(x_1, \ldots, x_n) = \sup_{t \in [0,1]^d} \left| t_1 \cdots t_d - \frac{1}{n} \sum_{i=1}^n 1_{[0,t)}(x_i) \right|$$

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

$$D_\infty^*(x_1, \ldots, x_n) = \sup_{\|f\| \le 1} |S_d(f) - Q_n(f)|,$$

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

The star-discrepancy is tractable

Theorem

- *Heinrich, N., Wasilkowski, Woźniakowski 2001:*

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

- *Hinrichs 2004:* $n(\varepsilon, F_d) \ge c\, d\, \varepsilon^{-1}$ *for* $\varepsilon \le \varepsilon_0$.
- *Aistleitner 2011:* $C = 100$.
- *Doerr 2014:*

$$\mathbb{E}(D^*_\infty(x_1, \dots, x_n)) \asymp \sqrt{\frac{d}{n}} \quad \text{for} \quad n \ge d.$$

Further results by Dick, Gnewuch, Larcher, Pillichshammer ...

Weighted Sobolev spaces and QMC

Inner product for $d = 1$ is

$$\langle f, g \rangle_{1,\gamma} = \int_0^1 f \, \mathrm{d}x \int_0^1 g \, \mathrm{d}x + \frac{1}{\gamma} \int_0^1 f'(x) g'(x) \, \mathrm{d}x,$$

tensor products for $d > 1$ and γ_i.

Theorem (Sloan, Woźniakowski 1998)

There exist points $x_1, \ldots, x_n$ for a QMC rule such that the problem is strongly polynomially tractable iff $\sum \gamma_i < \infty$.

Compute the mean of the quadratic error of QMC algorithms over all $(x_1, \ldots, x_n) \in [0,1]^{nd}$ and obtain

$$\frac{1}{n} \left(\int_{[0,1]^d} K(x,x) \, \mathrm{d}x - \int_{[0,1]^{2d}} K(x,y) \, \mathrm{d}x \, \mathrm{d}y \right).$$

Weighted Sobolev spaces and constructions

Sloan and Woźniakowski 1998 was continued in many directions.

- General weights, different Hilbert spaces.
- Lower bounds for arbitrary algorithms: decomposable kernels (N. and Woźniakowski 2001, NW2010 Ch. 11).
- Construction of good QMC methods: ...

Books and survey papers: Dick, Kuo, Sloan 2013, Dick and Pillichshammer 2010, 2014, N. and Woźniakowski 2008, 2010, 2012.

Curse of dimensionality for C^k functions. With Aicke Hinrichs, Mario Ullrich and Henryk Woźniakowski 2017

$$F_d = C_d^k = \{f \in C^k([0,1]^d) \;\mid\; \|D^\beta f\|_\infty \le 1 \text{ for all } \|\beta\|_1 \le k\}$$

Theorem

For all k there exist c_k and $\tilde{c}_k$ such that for all $n, d \in \mathbb{N}$ with $n = m^d$ for some $m \in \mathbb{N}$ we have

$$\min(1/2,\, c_k d n^{-k/d}) \le e(n, C_d^k) \le \min(1,\, \tilde{c}_k d n^{-k/d}).$$

The complexity is roughly $(d/\varepsilon)^{d/k}$, super-exponential in d.

Summary

Numerical integration is intractable in the worst case setting for classical function spaces, like $C^k([0,1]^d)$.

Remedies:

- Weighted spaces, problems with a structure
- Randomized algorithms

More on the integration problem

- For a d-variate function $f : [0,1]^d \to \mathbb{R}$, approximate the integral
$$S_d(f) = \int_{[0,1]^d} f(x) \, \mathrm{d}x.$$
- **Problem instances:** functions f from a function class F_d of d-variate functions; often we have an initial error
$$\max_{f \in F_d} |S_d(f)| = 1.$$
- **Algorithms:** deterministic or randomized algorithms using n function values.

Curse of dimensionality for C^k functions. With Aicke Hinrichs, Mario Ullrich and Henryk Woźniakowski 2017

$$F_d = C_d^k = \{f \in C^k([0,1]^d) \mid \|D^\beta f\|_\infty \le 1 \text{ for all } \|\beta\|_1 \le k\}$$

Theorem

For all k there exist c_k and $\tilde{c}_k$ such that for all $n, d \in \mathbb{N}$ with $n = m^d$ for some $m \in \mathbb{N}$ we have

$$\min(1/2,\, c_k d n^{-k/d}) \le e(n, C_d^k) \le \min(1,\, \tilde{c}_k d n^{-k/d}).$$

The complexity is roughly $(d/\varepsilon)^{d/k}$, super-exponential in d.

Periodic functions

To simplify some statements, we consider periodic functions:

$$\mathcal{C}_d^k = \left\{ f \in C^k(\mathbb{R}^d) \colon f \text{ is 1-periodic}, \|f\|_{\mathcal{C}_d^k} \le 1 \right\}$$

with the norm $\|f\|_{\mathcal{C}_d^k} := \max_{\beta \colon |\beta|_1 \le k} \|D^\beta f\|_\infty$.

Upper bound, product rule

Haber 1970: For the product rule

$$Q_m^d(f) = \sum_{i_1=0}^{m-1} \cdots \sum_{i_d=0}^{m-1} a_{i_1} \cdots a_{i_d} \cdot f(t_{i_1}, \dots, t_{i_d}),$$

based on the one-dimensional quadrature rule

$$Q_m(f) = \sum_{i=0}^{m-1} a_i \, f(t_i)$$

we have

$$e(Q_m^d, \mathcal{C}_d^k) \le \left(\sum_{\ell=0}^{d-1} A^\ell \right) \cdot e(Q_m, \mathcal{C}_1^k),$$

where $A = \sum_{i=1}^m |a_i|$.

Proof of the lower bound

For an arbitrary set $\mathcal{P} \subset [0,1]^d$ with $\#\mathcal{P} = n$ we will define a *fooling function* f, i.e., a function with $f(x) = 0$ for $x \in \mathcal{P}$, and large integral.

For $r > 0$ we first define the initial function

$$h_r(x) = \begin{cases} 0, & \text{if} \quad \min_{y \in \mathcal{P}_n} \|x - y\|_1 \leq r, \\ 1, & \text{otherwise.} \end{cases}$$

Clearly,

$$\int_{[0,1]^d} h_r(x) \mathrm{d}x = 1 - \mathrm{vol}_d(\mathcal{P}_r) \geq 1 - n \, \mathrm{vol}_d(rB_1^d) = 1 - n \, \frac{(2r)^d}{d!}.$$

Smoothing the function

The initial function h_r is not in $\mathcal{C}_d^k$. We use convolution to make it differentiable.

Let

$$g_{r,k}(x) = \frac{1}{\mathrm{vol}_d(\varrho_r B_1^d)} \begin{cases} 1, & \text{if } \|x\|_1 \le \frac{r}{k+1}, \\ 0, & \text{otherwise,} \end{cases}$$

and define the fooling function

$$f_r := h_r * \underbrace{g_{r,k} * \cdots * g_{r,k}}_{(k+1)\text{-fold}}.$$

We have $f_r(x) = 0$ for $x \in \mathcal{P}$ and $\int f_r \mathrm{d}x = \int h_r \mathrm{d}x$.

Bounding the derivatives

For a continuous function f it is easy to prove that

$$\|D^{e_i}[f * g_{r,k}]\|_\infty \le \frac{d(k+1)}{r}\,\|f\|_\infty$$

Inductively, we obtain

$$\|f_r\|_{\mathcal{C}_d^k} \le \max\Big\{1,\ r^{-k}\big(d(k+1)\big)^{k-1}\Big\}$$

Finishing the proof

We define

$$f_r^* \;=\; \frac{f_r}{\|f_r\|_{\mathcal{C}_d^k}} \;\in\; \mathcal{C}_d^k.$$

Using

$$\int_{[0,1]^d} f_r(x)\mathrm{d}x \;=\; 1 - n\,\frac{(2r)^d}{d!} \;>\; 1 - n\left(\frac{4er}{d}\right)^d$$

we obtain that $\int_{[0,1]^d} f_r^*(x)\mathrm{d}x \le \varepsilon$ implies that

$$n \;\geq\; \left(1 - \varepsilon \cdot \|f_r\|_{\mathcal{C}_d^k}\right) \left(\frac{d}{4er}\right)^d.$$

Finishing the proof

$$n \geq \left(1 - \varepsilon \cdot \|f_r\|_{\mathcal{C}_d^k}\right) \left(\frac{d}{4er}\right)^d .$$

Now we choose

$$r = (2\varepsilon)^{1/r} \big(d(k+1)\big)^{1-1/k}$$

to obtain

$$\|f_r\|_{\mathcal{C}_d^k} \leq \frac{1}{2\varepsilon}$$

and

$$n(\varepsilon, \mathcal{C}_d^k) \geq \frac{1}{2} \left(\frac{c_k\, d}{\varepsilon}\right)^{d/r}$$

with $c_k = 1/((4e)^k\,(k+1)^{k-1})$.

Universal algorithms   Krieg, N., Ullrich 2017

Compute the integral $I_d(f) = \int_{[0,1]^d} f(x)\,\mathrm{d}x$ by an "optimal" algorithm

$$A_n(f) = \sum_{i=1}^{n} a_i f(x_i).$$

We consider the spaces $H^s([0,1]^d)$ and $H^{r,\mathrm{mix}}([0,1]^d)$.

$H^s([0,1]^d)$ contains all f such that $D^\alpha f \in L_2$ for $\|\alpha\|_1 \le s$.

$H^{r,\mathrm{mix}}([0,1]^d)$ contains all f such that $D^\alpha f \in L_2$ for $\|\alpha\|_\infty \le r$.

Results of Bakhvalov (1959) and Frolov (1976)

Bakhvalov 1959: For $H^s([0,1]^d)$ with $s \in \mathbb{N}$ and $s > d/2$ the optimal rate is

$$|I_d(f) - A_n(f)| \leq c\, n^{-s/d} \|f\|_{H^s}.$$

Frolov 1976: For $H^{r,\mathrm{mix}}([0,1]^d)$ with $r \in \mathbb{N}$ the optimal rate is

$$|I_d(f) - A_n(f)| \leq c\, n^{-r} (\log n)^{(d-1)/2} \|f\|_{H^{r,\mathrm{mix}}}.$$

Questions:

- randomized unbiased algorithms?
- universality, same algorithm for all function spaces?

Property P1 of a dream algorithm: unbiased and positive

The algorithm A_n should be an unbiased randomized algorithm,

$$\mathbb{E}(A_n(f)) = I_d(f)$$

for $f \in L_1$.

It is also good if $a_i \geq 0$ for all i (stability).

Property P2 of a dream algorithm: small worst case error

The worst case error $\sup_\omega(|I_d(f) - A_n^\omega(f)|)$ should be small for "many" classes F of functions. For all $r \in \mathbb{N}$

$$\sup_\omega(|I_d(f) - A_n^\omega(f)|) \le c\, n^{-r}(\log n)^{(d-1)/2}\|f\|_{H^{r,\mathrm{mix}}},$$

for all $s \in \mathbb{N}$ with $s > d/2$

$$\sup_\omega(|I_d(f) - A_n^\omega(f)|) \le c\, n^{-s/d}\|f\|_{H^s}.$$

Property P3 of a dream algorithm: small randomized error

The randomized error $\mathbb{E}(|I_d(f) - A_n(f)|)$ should be small for "many" classes F of functions. For all $r \in \mathbb{N}$

$$\mathbb{E}(|I_d(f) - A_n(f)|) \le c\, n^{-r-1/2} \|f\|_{H^{r,\mathrm{mix}}},$$

for all $s \in \mathbb{N}$ with $s > d/2$

$$\mathbb{E}(|I_d(f) - A_n(f)|) \le c\, n^{-s/d-1/2} \|f\|_{H^s}.$$

P4 tractability properties; P5 easy to implement

Two more properties that we do not discuss here:

P4: The algorithm should have good tractability properties.

P5: The algorithm should be easy to implement.

Construction of the algorithm for $C_c(\mathbb{R}^d)$ Part 1

Let $B \in \mathbb{R}^{d \cdot d}$ be an invertible matrix and $v \in \mathbb{R}^d$. We define

$$Q_{B,v}(f) = \frac{1}{|\det B|} \sum_{m \in \mathbb{Z}^d} f(B^{-t}(m+v)).$$

Then, for $f \in C_c(\mathbb{R}^d)$,

$$|I_d(f) - Q_{B,v}(f)| \le \sum_{m \in \mathbb{Z}^d \setminus \{0\}} |\hat{f}(Bm)|.$$

Important assumption for the following ("Frolov property"):

$$\left| \prod_{j=1}^{d} (Bm)_j \right| \ge 1$$

for all $m \in \mathbb{Z}^d \setminus \{0\}$.

Construction of the algorithm for $C_c(\mathbb{R}^d)$ Part 2

For a Frolov matrix B and $a > 0$ the randomized Frolov algorithm $M_{a,B}$ is the algorithm $Q_{aUB,v}$ as before with independent random vectors u and v, uniformly distributed in $[1, 2^{1/d}]^d$ and $[0,1]^d$, respectively and $U = \mathrm{diag}(u_1, \ldots, u_d)$.

Then $M_{a,B}$ is unbiased on $L_1(\mathbb{R}^d)$ with $n \asymp a^d$.

Main error bound:

$$\mathbb{E}|I_d(f) - M_{a,B}(f)| \leq c\, a^{-d} \int_{D_a} |\hat{f}(x)| \,\mathrm{d}x,$$

for $f \in C_c(\mathbb{R}^d)$, where D_a is the set of all $x \in \mathbb{R}^d$ with $\prod_{j=1}^d |x_j| \geq a^d$.

Final Algorithm

The algorithm $M_{a,B}$ fulfills all the error bounds from P2 and P3 for functions with compact support in $(0,1)^d$.

Let $\psi : \mathbb{R} \to \mathbb{R}$ be a monotone C^∞-function with $\psi(x) = 0$ for $x \le 0$ and $\psi(x) = 1$ for $x \ge 1$ and ψ is a diffeomorphism on $(0,1)$. Furthermore, $\Psi(x) = (\psi(x_1), \ldots, \psi(x_d))$.

For $v \in \mathbb{R}^d$ and an invertible $B \in \mathbb{R}^{d \cdot d}$ let

$$\widetilde{Q}_{B,v}(f) = \frac{1}{|\det B|} \sum_{m \in \mathbb{Z}^d} f(\Psi(B^{-t}(m+v))) \cdot |\det D\Psi(B^{-t}(m+v))|.$$

Final Algorithm for $f : [0,1]^d \to \mathbb{R}$: For a Frolov matrix B and $a > 0$ consider $\widetilde{Q}_{aUB,v}$ with independent u and v, uniformly distributed on $[1, 2^{1/d}]^d$ and $[0,1]^d$, respectively.

Main Result

This “final algorithm” is unbiased, the weights are positive, and all error bounds from P2 and P3 hold.