UNIVERSITY OF WARSAW

On complexity of numerical problems of continuous mathematics

Leszek Plaskota
leszekp@mimuw.edu.pl

Colloquium WMIM
November 18, 2021

Introduction

Numerical analysis deals with solving numerically problems of continuous mathematics

- Creation and analysis of algorithms
- Software implementations
- Computational complexity

Problems of continuous mathematics

UNIVERSITY OF WARSAW

1. Linear equations

$$A\vec{x} = \vec{b}, \qquad A \in \mathbb{R}^{m,n}, \quad \vec{b} \in \mathbb{R}^m$$

2. Integration

$$\int_D f(\vec{x})\,\mathrm{d}\vec{x}, \qquad f \in F$$

3. Approximation

$$f \in F \hookrightarrow G$$

4. .....

What is going on?

Consider $n \times n$ system

$$A\vec{x} = \vec{b}$$

$$A = \begin{pmatrix} 1 & & & & & 1 \\ -1 & 1 & & & & 1 \\ -1 & -1 & 1 & & & 1 \\ -1 & -1 & -1 & 1 & & 1 \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ -1 & -1 & \cdots & -1 & -1 & 1 \end{pmatrix}, \qquad \vec{b} = A \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ \vdots \\ 1 \end{pmatrix}$$

MATLAB returns a solution with relative l_∞ error

$$\begin{cases} 0 & \text{for} \quad 1 \leq n \leq 54 \\ 1 & \text{for} \quad n = 55 \end{cases}$$

even though A is very well conditioned

A general framework

IBC = Information-Based Complexity

$$
\begin{aligned}
\textit{The problem:} &\quad S : F \to G \\
\textit{Information:} &\quad \mathcal{I} : F \to Y \\
\textit{Algorithm:} &\quad \varphi : Y \to \mathcal{Z} \qquad\qquad Y, \mathcal{Z} \subset \cup_{n=0}^{\infty} \mathbb{R}^n \\
\textit{Interpretation:} &\quad \mathcal{Z} : Y \to G
\end{aligned}
$$

If S a functional then $G = \mathcal{Z} = \mathbb{R}$

$$
\begin{array}{ccc}
F & \stackrel{S}{\longrightarrow} & \mathbb{R} \\
{\scriptstyle \mathcal{I}}\big\downarrow & \nearrow_{\varphi} & \\
Y & &
\end{array}
\qquad \text{or} \qquad
\begin{array}{ccc}
F & \stackrel{S}{\longrightarrow} & G \\
{\scriptstyle \mathcal{I}}\big\downarrow & & \big\uparrow{\scriptstyle \mathcal{T}} \\
Y & \stackrel{\varphi}{\longrightarrow} & Z
\end{array}
$$

Information

$$\mathcal{I}f = (y_1, y_2, \ldots, y_n) \in Y$$

Nonadaptive information $\quad (Y = \mathbb{R}^n)$:

$$y_i = L_i f, \qquad L_i \in \Lambda$$

Adaptive information $\quad$ (Y is a prefix-free set):

$$y_i = L_i(f; y_1, \ldots, y_{i-1}), \qquad L_i(\,\cdot\,; y_1, \ldots, y_{i-1}) \in \Lambda$$

Here Λ is a class of permissible functionals

Information ε-complexity

In what follows $\Phi = \mathcal{T} \circ \varphi$

$$\mathrm{comp}(\varepsilon) = \inf\left\{\mathrm{cost}(\mathcal{I}) : \ \Phi \ \text{s.t.} \ \mathrm{error}(\Phi, \mathcal{I}) \leq \varepsilon\right\}$$

Cost and error are defined depending on the setting:

- worst case setting
- average case setting
- asymptotic setting
- randomized settings
- various mixed settings
- .....

Worst case setting

Let $\mathcal{F} \subset F$ be a class problem instances

$$\begin{aligned} \text{error}(\Phi, \mathcal{I}) &= \sup \left\{ \|Sf - \Phi(\mathcal{I}f)\|_G : f \in \mathcal{F} \right\} \\ \text{cost}(\mathcal{I}) &= \sup \left\{ n = |\mathcal{I}f|, f \in \mathcal{F} \right\} \end{aligned}$$

Some general issues (for linear problems):

- linear vs. nonlinear algorithms
- adaptive vs. nonadaptive information
- worst case vs. asymptotic approach
- .....

Linear vs. nonlinear algorithms

Radius of information:

$$\mathrm{rad}(\mathcal{I}) = \inf\left\{\mathrm{error}(\Phi, \mathcal{I}) :\ \Phi \text{ - arbitrary }\right\}$$

Theorem (Smolyak 1963)

Let

- *the problem* $S : F \to \mathbb{R}$ *be a linear functional*
- *information* $\mathcal{I} = (L_1, \ldots, L_n)$ *be nonadaptive*
- *the class* $\mathcal{F} \subset F$ *be convex and balanced*

Then there is a linear algorithm φ *that is optimal, i.e.,*

$$\mathrm{error}(\varphi, \mathcal{I}) = \mathrm{rad}(\mathcal{I}).$$

Adaptive vs. nonadaptive information

Theorem (Bakhvalov 1971, Taub & Woźniakowski 1980)

Let

- *the problem* $S : F \to \mathbb{R}$ *be a linear operator*
- *the class* $\mathcal{F} \subset F$ *be convex and balanced*

Then for any adaptive information $\mathcal{I}^{\mathrm{ada}}$ *there is nonadaptive information* $\mathcal{I}^{\mathrm{non}}$ *such that*

$$\begin{aligned} \mathrm{cost}(\mathcal{I}^{\mathrm{non}}) &\leq \mathrm{cost}(\mathcal{I}^{\mathrm{ada}}) \\ \mathrm{rad}(\mathcal{I}^{\mathrm{non}}) &\leq 2 * \mathrm{rad}(\mathcal{I}^{\mathrm{ada}}) \end{aligned}$$

If S *is a functional then* $\mathrm{rad}(\mathcal{I}^{\mathrm{non}}) \leq \mathrm{rad}(\mathcal{I}^{\mathrm{ada}})$.

Worst case vs. asymptotic setting

In the asymptotic approach we are interested in the behavior of

$$\|Sf - \Phi_n(\mathcal{I}_n f)\|_G \qquad \text{as} \quad n \to +\infty$$

for *each* individual $f \in F$, where $\mathcal{I}_n$ uses n evaluations

Theorem (Trojan 1980)

Let

- *F be a Banach space*
- *the problem $S : F \to G$ be a linear operator*

Then for any positive sequence $\{\delta_n\}_{n \geq 1}$ converging to zero the set

$$\left\{ f \in F : \limsup_{n \to +\infty} \frac{\|Sf - \Phi_n(\mathcal{I}_n f)\|_G}{\delta_n * \mathrm{rad}(\mathcal{I}_n)} < +\infty \right\}$$

does not contain any nontrivial ball.
(Here the radius is with respect to the unit ball of F)

Numerical integration

Formulation

Let F be a Banach space of functions $f \in C^r([a,b])$ with norm

$$\|f\| = \|f\|_\infty + \|f^{(r)}\|_\infty$$

and the class $\quad \mathcal{F} = \{f \in F : \ \|f\|_F \leq 1\}$

The problem:

$$S(f) = \int_a^b f(x)\,\mathrm{d}x, \qquad f \in F$$

Information:

$$L \in \Lambda \quad \text{iff} \quad Lf = f(x) \quad \text{for some} \quad x \in [a,b]$$

Complexity

Theorem

For the integration problem, the worst case complexity is

$$\mathrm{comp}(\varepsilon) \asymp \varepsilon^{-r}$$

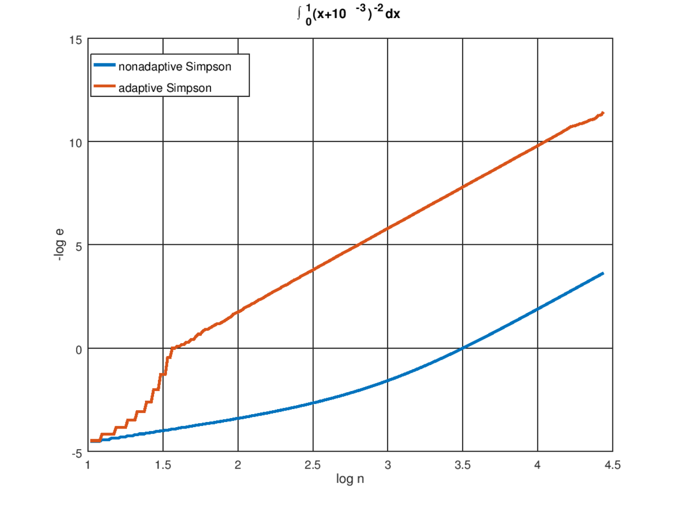
and is achieved by nonadaptive quadratures

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

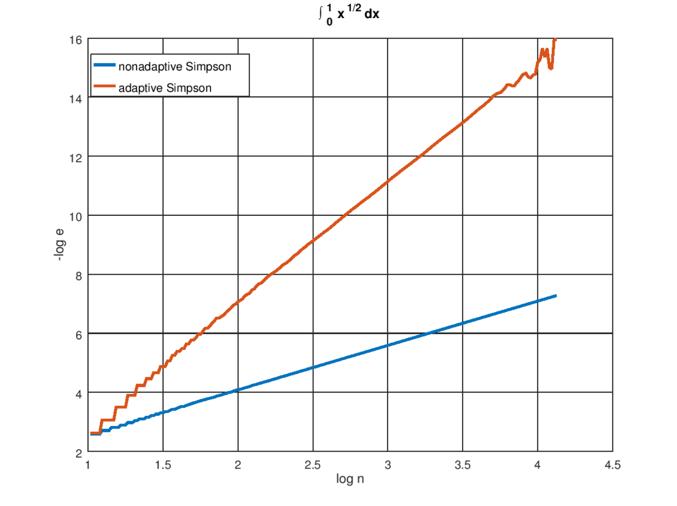
that are composed of simple quadratures of degree of exactness r. In particular, adaption does not help.

Moreover, in the asymptotic approach the worst case convergence rate n^{-r} cannot be beaten

Adaption does not help?



Adaption does not help?



L^p approximation of Hölder classes from noisy data

Formulation

Let $K = [0,1]^d,\ r \geq 0,\ 0 < \varrho \leq 1$
$H^d_{r,\varrho}$ be the space of functions $f \in C^r(K)$ s.t. all the derivatives

$$\mathcal{D}^{(\alpha)} f = \frac{\partial^{|\alpha|} f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}} \quad \text{with} \quad \alpha = (\alpha_1, \ldots, \alpha_d),\ |\alpha| = r$$

are Hölder continuous with exponent ϱ. The problem is to approximate $f \in H^d_{r,\varrho}$ in the L^p-norm with

$$1 \leq p < +\infty,$$

based on noisy observations of f.

Noisy information

Nonadaptive information $\mathbf{y} = (y_1, y_2, \ldots, y_n) \in Y$ where

$$y_j = f(\mathbf{x}_j) + e_j, \qquad e_j \sim \mathcal{N}(0, \sigma_j^2), \qquad 1 \le j \le n$$

Adaptive information

$$y_j = f\big(\mathbf{x}_j(y_1, \ldots, y_{j-1})\big) + e_j, \qquad e_j \sim \mathcal{N}\big(0, \sigma_j^2(y_1, \ldots, y_{j-1})\big)$$

The points $\mathbf{x}_j$ and precisions σ_j are subject to our choice.

Let π_f be the probability measure corresponding to the distribution of information $\mathbf{y}$ about f,

$$\Pi = \{\pi_f\}_{f \in F}$$

Error

For $f \in H^d_{r,\varrho}$ define the seminorm

$$[f]_{r,\varrho} \;=\; \max_{|\alpha|=r}\; \sup_{\mathbf{t}_1,\mathbf{t}_2 \in K} \frac{\left|\mathcal{D}^{(\alpha)} f(\mathbf{t}_1) - \mathcal{D}^{(\alpha)} f(\mathbf{t}_2)\right|}{\|\mathbf{t}_1 - \mathbf{t}_2\|_\infty^\rho} < +\infty$$

and the class

$$\mathcal{H} \;=\; \mathcal{H}^d_{r,\rho} \;=\; \left\{ f \in H^d_{r,\rho} :\; [f]_{r,\rho} \le 1 \right\}$$

The (worst case) L^p-error of Φ using information Π is

$$\mathrm{error}(\Phi, \Pi) \;=\; \sup_{f \in \mathcal{H}} \left(\int_Y \|f - \Phi(\mathbf{y})\|_{L^p}^p \, \pi_f(\mathrm{d}\mathbf{y}) \right)^{1/p}$$

where $\|f - \Phi(\mathbf{y})\|_{L^p} \;=\; \left(\int_K |(f - \Phi(\mathbf{y}))(\mathbf{x})|^p \, \mathrm{d}\mathbf{x} \right)^{1/p}$.

Cost

The cost of a single observation with variance σ^2 equals $c(\sigma)$,

$$c : [0, +\infty) \to (0, +\infty]$$

is a non-trivial and non-increasing *cost function*. Then

$$\mathrm{cost}(\Pi) = \sup_{f \in \mathcal{H}} \int_Y \sum_{i=1}^{n(\mathbf{y})} c\big(\sigma_i(y_1, \ldots, y_{i-1})\big)\, \pi_f(\mathrm{d}\mathbf{y})$$

If

$$c(\sigma) = \begin{cases} +\infty, & 0 \leq \sigma < \sigma_0, \\ 1, & \sigma_0 \leq \sigma, \end{cases}$$

then we observe with fixed variance $\sigma_0^2 \geq 0$ at cost 1.

Theorem

For exact information $\mathrm{comp}(\varepsilon) \asymp \left(\frac{1}{\varepsilon}\right)^{\frac{d}{r+\varrho}}$.

Otherwise, letting $\hat{c}(x) = c(x^{-1/2})$, *we have*

(i) *If* $\hat{c}$ *is concave then* $\mathrm{comp}(\varepsilon) \asymp c(\varepsilon)\left(\frac{1}{\varepsilon}\right)^{\frac{d}{r+\varrho}}$

(ii) *If* $\hat{c}$ *is convex and* $\hat{c}(0) > 0$ *or* $\hat{c}'(0) > 0$ *then* $\mathrm{comp}(\varepsilon) \asymp \left(\frac{1}{\varepsilon}\right)^{\frac{d}{r+\varrho}+2}$

(iii) *If* $\hat{c}$ *is convex and* $\hat{c}(0) = \hat{c}'(0) = 0$ *then* $\mathrm{comp}(\varepsilon) = 0$

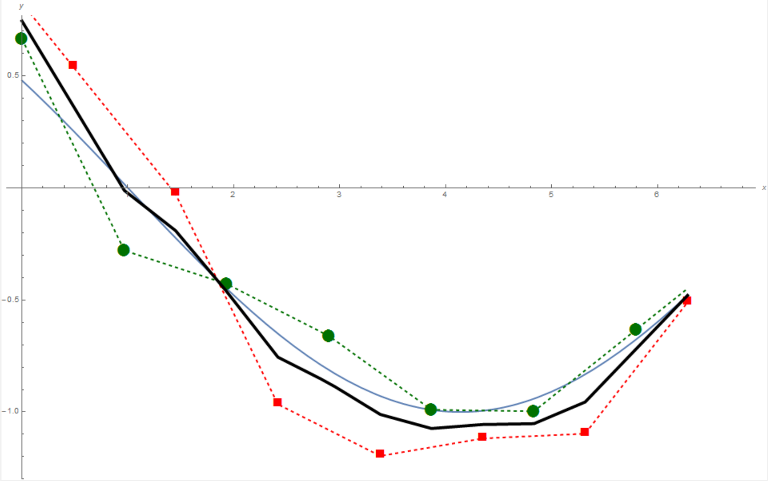
For instance, for $c(\sigma) = 1 + \sigma^{-2t}$, $t \geq 0$, we have

$$\mathrm{comp}(\varepsilon) \asymp \left(\frac{1}{\varepsilon}\right)^{\frac{d}{r+\varrho}+2\min(t,1)}$$

Remarks

1. Importance of the one-dimensional problem: approximate $f \in [-\tau, \tau]$ from its noisy observations with Gaussian noise
2. Adaption may help by a constant only
3. Nonlinear algorithms are better than linear once by a constant only
4. To achieve complexity bounds it is enough to use:
   - observations on regular grid with variance $\sigma^2 = \varepsilon^2$ and piecewise polynomial interpolation on exact data (in case of concave $\hat{c}$), or
   - observations on regular grid with variance σ_0^2 independent of ε and piecewise polynomial interpolation on smoothed data (in case of convex $\hat{c}$).

Smoothing ($d = 1, r = 1$)



References

- Traub J.F., Wasilkowski G.W., Woźniakowski H., *Information-Based Complexity*. Academic Press, New York, 1988
- Plaskota L., *Noisy Information and Computational Complexity*, Cambridge Univ. Press, Cambridge, 1996
- Novak E., Woźniakowski H., *Tractability of Multivariate Problems*, Tracts in Mathematics **6**, EMS, Vol. I (2008), Vol. II (2010), Vol. III (2012)
- Plaskota L., Automatic integration using asymptotically optimal adaptive Simpson quadrature, *Numerische Mathematik*, **131** (2015)
- Morkisz P., Plaskota L., Complexity of approximating Hölder classes from information with varying Gaussian noise, *J. Complexity* **60** (2020) 101497
- Plaskota L., Samoraj P., Automatic approximation using asyptotically optimal adaptive interpolation, *Numerical Algorithms* (2021)

Thank you