Jak trudne jest numeryczne całkowanie (O złożoności zadań ciągłych)

Leszek Plaskota

Uniwersytet Warszawski
Wydział Matematyki, Informatyki i Mechaniki
leszekp@mimuw.edu.pl

Horyzonty 2014
17-03-2014
Będlewo

Zadania numeryczne - przykłady

- *Rozwiązywanie układów równań liniowych:*
  dla danych $A \in \mathbb{R}^{n,n}$, $\det A \neq 0$, $\vec{b} \in \mathbb{R}^n$, obliczyć $A^{-1}\vec{b}$
- *Interpolacja wielomianowa:*
  dla danych wartości $f(t_j)$ funkcji $f : \mathbb{R} \to \mathbb{R}$, $0 \leqslant j \leqslant n$, obliczyć współczynniki $a_0, \ldots, a_n$ rozwinięcia w bazie $p_0, \ldots, p_n$ wielomianu $w = \sum_{i=0}^{n} a_j p_j$, takiego że

$$w(t_j) = f(t_j), \quad 0 \leqslant j \leqslant n$$

- *Całkowanie funkcji:*
  dla danej funkcji $f : D \to \mathbb{R}$ obliczyć (przybliżyć) całkę $\int_D f(x)\,\mathrm{d}x$ na podstawie wartości $f(t_j)$, $1 \leqslant j \leqslant n$
- $\cdots$

Zadania numeryczne - schemat ogólny

Chcemy aproksymować wartości $S(f)$ odwzorowania

$$S : F \to G \qquad \text{(zadanie)}$$

gdzie F przestrzeń liniowa (dane), G unormowana (wyniki)

Schemat aproksymacji:

$$f \;\mapsto\; N(f) \;\mapsto\; \varphi(N(f))$$

$$\begin{aligned} N &: F \to Y \qquad \text{(informacja)} \\ \varphi &: Y \to G \qquad \text{(algorytm)} \end{aligned}$$

$$S \sim \varphi \circ N$$

Błąd algorytmu (przypadek najgorszy)

Błąd algorytmu φ korzystającego z informacji N, na klasie $E \subseteq F$:

$$e^{\mathrm{wor}}(\varphi, N) \;:=\; \sup_{f \in E} \|S(f) - \varphi(Nf))\|_G$$

Algorytm φ^* jest optymalny dla N gdy

$$e^{\mathrm{wor}}(\varphi^*, N) \;=\; \inf_{\varphi} e^{\mathrm{wor}}(\varphi, N)$$

Twierdzenie (Promień informacji)

$$r^{\mathrm{wor}}(N) \;:=\; \inf_{\varphi} e^{\mathrm{wor}}(\varphi, N) \;=\; \sup_{y \in N(E)} \mathrm{rad}\big(S(E \cap N^{-1}y)\big)$$

gdzie $\mathrm{rad}(A)$ *jest promieniem Czebyszewa zbioru* A
$S(E \cap N^{-1}y) = \{S(h) : \; h \in E, Nh = y\}$

Optymalność algorytmów liniowych dla funkcjonałów

Niech zadanie $S : F \to \mathbb{R}$ będzie funkcjonałem liniowym, a $N : F \to \mathbb{R}^n$ informacją liniową

Twierdzenie (Smolak 1963)

Jeśli klasa $E \subset F$ jest wypukła i zbalansowana to dla informacji N istnieje algorytm liniowy φ^{lin}, który jest optymalny,

$$e^{\mathrm{wor}}(\varphi^{\mathrm{lin}}, N) = r^{\mathrm{wor}}(N)$$

Wniosek: kwadratury numeryczne $Q_n(f) = \sum_{i=1}^{n} a_i f(t_i)$ korzystające z wartości funkcji są optymalne o ile klasa funkcji podcałkowych jest wypukła i zbalansowana

Dowód Tw. Smolaka

Lemat

$$r^{\mathrm{wor}}(N) \;=\; \sup\{|Sh| \,:\; h \in E \cap \ker N\} \;=\; \mathrm{rad}\big(S(E \cap \ker N)\big)$$

Dowód (lematu). Dla dowolnego $y \in N(E)$ mamy

$$\begin{aligned}
\mathrm{rad}(S(E \cap N^{-1}y)) &= \frac{1}{2}\mathrm{diam}(S(E \cap N^{-1}y)) \\
&= \sup\left\{\left|S\left(\frac{f_1 - f_2}{2}\right)\right| : f_1, f_2 \in E, N\left(\frac{f_1 - f_2}{2}\right) = 0\right\} \\
&\leqslant \sup\{|Sh| : h \in E, Nh = 0\} = \mathrm{rad}(S(E \cap \ker N))
\end{aligned}$$

Z drugiej strony, $r^{\mathrm{wor}}(N) \geqslant \mathrm{rad}(S(E \cap \ker N))$. $\square$

Dowód Tw. Smolaka, c.d.

Niech $r := r^{\mathrm{wor}}(N) > 0$,

$$A = \{(Nf, Sf) : f \in E\} \subset \mathbb{R}^{n+1}$$

A jest wypukły i zbalansowany. Przy (upraszczających) założeniach, że jest też pochłaniający i ograniczony, *funkcjonał Minkowskiego*

$$\|u\| := \inf\{t > 0 : u/t \in A\}$$

definiuje normę w $\mathbb{R}^{n+1}$.
Zdefiniujmy na $P = \{(0, g) : g \in \mathbb{R}\}$ funkcjonał

$$\xi_1(0, g) = g/r$$

Ponieważ $\|(0, g)\| = |g|/r$ to $|\xi_1(0, g)| \leqslant \|(0, g)\|$.

Dowód Tw. Smolaka, c.d.

ξ_1 można teraz rozszerzyć (*Tw. Hahna-Banacha!*) do funkcjonału ξ_2 na $\mathbb{R}^{n+1}$ z zachowaniem normy, tzn.

(i) $\xi_2(x) = \xi_1(x)$ dla $x = (0, g)$, $g \in \mathbb{R}$

(ii) $|\xi_2(x)| \leqslant 1$ dla $x \in \mathbb{R}^{n+1}$

Stąd

$$|\xi_2(y, g)| = |\xi_2(0, g) + \xi_2(y, 0)| = \left|\frac{g}{r} + \xi_2(y, 0)\right| \leqslant 1$$

Przyjmując $y = Nf$, $g = Sf$,

$$\varphi(y) = -r\,\xi_2(y, 0)$$

dostajemy $|Sf - \varphi(Nf)| \leqslant r$. $\square$

Adaptacja nie pomaga

Informacja *adaptacyjna* $y = N^{\mathrm{ada}}(f)$

$$\begin{aligned} y_1 &= f(x_1) \\ y_2 &= f(x_2(y_1)) \\ &\cdots \\ y_n &= f(x_n(y_1, y_2, \ldots, y_{n-1})) \end{aligned}$$

tzn. wybór x_i zależy od wartości $y_1, \ldots, y_{i-1}$ z poprzednich kroków

Dla N^{ada} definiujemy informację *nieadaptacyjną*

$$N^{\mathrm{non}}(f) = [f(x_1), f(x_2(0)), \ldots, f(x_n(\underbrace{0, \ldots, 0}_{n-1}))]$$

Adaptacja nie pomaga, c.d.

Twierdzenie (Bachwałow 1971)

$$r^{\mathrm{wor}}(N^{\mathrm{non}}) \leqslant r^{\mathrm{wor}}(N^{\mathrm{ada}})$$

Dowód.

$$\begin{aligned} r^{\mathrm{wor}}(N^{\mathrm{ada}}) &= \sup_{y \in N^{\mathrm{ada}}(E)} \mathrm{rad}(\{Sf : f \in E, N^{\mathrm{ada}}(f) = y\}) \\ &\geqslant \mathrm{rad}(\{Sh : h \in E, N^{\mathrm{ada}}(h) = 0\}) \\ &= \mathrm{rad}(\{Sh : h \in E, N^{\mathrm{non}}(h) = 0\}) \\ &= r^{\mathrm{wor}}(N^{\mathrm{non}}). \qquad \square \end{aligned}$$

Całkowanie funkcji lipschitzowskich

Zadanie:

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

$$E = \{f : [0,1]^d \to \mathbb{R} : \|f\|_F \leqslant 1\}$$

$$\|f\|_F = \max\left(\sup_{x\in[0,1]^d} |f(x)|, \sup_{x,y\in[0,1]^d} \frac{|f(x)-f(y)|}{\|x-y\|_\infty} \right)$$

Informacja:

$$N_n(f) = [f(t_1), f(t_2), \ldots, f(t_n)], \qquad t_j \in [0,1]^d$$

ε-złożoność zadania:

$$n^{\mathrm{wor}}(\varepsilon, d) = \min\{n : \exists N_n \; r^{\mathrm{wor}}(N_n) \leqslant \varepsilon\}$$

Całkowanie funkcji lipschitzowskich, c.d.

Twierdzenie (Sucharew 1979)

Jeśli $\varepsilon = 1/((2+2/d)m)$ *dla pewnej* m *to*

$$n^{\mathrm{wor}}(\varepsilon, d) = \frac{1}{(1+1/d)^d}\left(\frac{1}{2\varepsilon}\right)^d = \frac{1+o(1)}{\mathrm{e}}\left(\frac{1}{2\varepsilon}\right)^d$$

Optymalną kwadraturą jest złożona kwadratura trapezów oparta na siatce jednorodnej.

Jeśli, np. $\varepsilon = 0.05$ ale $d = 20$ to $n^{\mathrm{wor}}(\varepsilon, d) \asymp 10^{20}$

$\Longrightarrow$ przekleństwo wymiaru

Podatność i przekleństwo wymiaru

$$S_d : F_d \to G_d, \qquad d = 1, 2, 3, \ldots$$

Definicja

Zadanie S_d jest podatne wielomianowo *gdy istnieją C, p, q, że*

$$n^{\mathrm{wor}}(\varepsilon, d) \leqslant \; C\, d^q\, \varepsilon^{-p} \qquad \forall d \quad \forall \varepsilon \in (0, 1)$$

Definicja

Zadanie S_d podlega przekleństwu wymiaru *gdy istnieje $\alpha > 0$, że*

$$n^{\mathrm{wor}}(\varepsilon, d) \geqslant \; (1 + \alpha)^d \quad \textit{dla prawie wszystkich } d$$

Wniosek:

całkowanie funkcji lipschitzowskich podlega przekleństwu wymiaru

Całkowanie funkcji r-gładkich

$$F = F_{d,r} = \left\{ f : [0,1]^d \to \mathbb{R} : \ \|f\|_{d,r} := \max_{|\alpha| \leqslant r} \max_{x \in [0,1]^d} |D^\alpha f(x)| \leqslant 1 \right\}$$

Twierdzenie (Bachwałow 1959)

$$n^{\mathrm{wor}}(\varepsilon, d) \ = \ c_{r,d} \left(\frac{1}{\varepsilon} \right)^{d/r}$$

Wniosek: Całkowanie funkcji r-gładkich nie jest podatne wielomianowo

Czy występuje przekleństwo wymiaru?

TAK! $\Longrightarrow$ Hinrichs, Novak, Ullrich, Woźniakowski (2013)

Algorytmy randomizacyjne (Monte Carlo)

Informacja i algorytm:

$$(N,\varphi) = \{(N_\omega, \varphi_\omega)\} \qquad \omega \text{ - zmienna losowa}$$

Błąd algorytmu randomizacyjnego:

$$e^{\mathrm{rand}}(N,\varphi) \;=\; \sup_{f\in E} \sqrt{\mathbb{E}_\omega |S_d(f) - \varphi_\omega(N_\omega f)|^2}$$

Przykład: (klasyczne Monte Carlo)

$$\mathrm{MC}_n(f) \;=\; \frac{1}{n}\sum_{j=1}^{n} f(t_j)$$

t_j - niezależne zmienne losowe z rozkładu jednostajnego na $[0,1]^d$

Algorytmy randomizacyjne (Monte Carlo), c.d.

W klasie $F_{d,r}$ mamy

$$e^{\mathrm{rand}}(\mathrm{MC}_n) \;=\; n^{-1/2}$$

Rzeczywiście, oznaczając $X = \mathrm{MC}_n(f)$ mamy $\mathbb{E}(X) = S_d(f)$,

$$(e^{\mathrm{rand}}(\mathrm{MC}_n))^2 = \mathbb{E}(X - \mathbb{E}(X))^2 = \mathbb{E}(X^2) - \mathbb{E}^2(X)$$

$$\begin{aligned}\mathbb{E}\left(\frac{1}{n}\sum_{i=1}^{n} f(t_i)\right)^2 &= \frac{1}{n^2}\mathbb{E}\left(\sum_{i=1}^{n} f^2(t_i) + 2\sum_{i<j} f(t_i)f(t_j)\right) \\ &= \frac{1}{n}S_d(f^2) + \frac{n-1}{n}(S_d(f))^2\end{aligned}$$

$$\mathbb{E}(X^2) - \mathbb{E}^2(X) = \frac{S_d(f^2) - (S_d(f))^2}{n} \leqslant \frac{1}{n}\int_{[0,1]^d} f^2(t)\,\mathrm{d}t \leqslant \frac{1}{n}$$

Algorytmy randomizacyjne (Monte Carlo), c.d.

Powyższe implikuje silną wielomianową podatność

$$n^{\mathrm{rand}}(\varepsilon, d) \leqslant \varepsilon^{-2}$$

przy czym wykładnik 2 jest optymalny

Twierdzenie (Bachwałow 1959)

$$n^{\mathrm{rand}}(\varepsilon, d) = c_{r,d} \left(\frac{1}{\varepsilon}\right)^{2/(1+2r/d)}$$

Algorytm:

$$\widetilde{\mathrm{MC}}_n(f) = S_d(A_n f) + \mathrm{MC}_n(f - A_n f)$$

gdzie $A_n f$ jest kawałkami wielomianową aproksymacją funkcji f

Dyskrepancja i quasi-Monte Carlo

Definicja (Dyskrepancja)

$$\mathrm{disc}(t_1, \ldots, t_n) = \sup_{x \in [0,1]^d} \left| \mathrm{Vol}([0,x)) - \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}_{[0,x)}(t_i) \right|$$

Rozpatrujemy algorytmy postaci

$$\mathrm{QMC}_n(f) = \frac{1}{n} \sum_{j=1}^{n} f(t_j)$$

gdzie $t_1, \ldots, t_n$ są wybrane deterministycznie, dla klasy

$$E = \{ f : [0,1]^d \to \mathbb{R} : \ f(\ldots, 1, \ldots) = 0, \ \|f\|_{d,*} \leqslant 1 \}$$

$$\|f\|_{d,*} = \sup_{x \in [0,1]^d} \left| \frac{\partial^d f}{\partial x_1 \ldots \partial x_d}(x) \right|$$

Dyskrepancja i quasi-Monte Carlo, c.d.

Twierdzenie (Zaremba-Koksma-Hlavka)

$$e^{\mathrm{wor}}(\mathrm{QMC}_n) = \mathrm{disc}(t_1, \ldots, t_n)$$

Najlepsze konstrukcje punktów t_i dają

$$\mathrm{disc}(t_1, \ldots, t_n) \leqslant c_d \, \frac{\ln^{d-1} n}{n}$$

Twierdzenie (Heinrich, Novak, Wasilkowski, Woźniakowski 2001)

Istnieją punkty t_i^*, takie że

$$\mathrm{disc}(t_1^*, \ldots, t_n^*) \leqslant C \, d^{1/2} n^{-1/2}$$

To daje podatność wielomianową
Dowód jest niekonstruktywny(!)

Bibliografia (IBC)

1. A. Hinrichs, E. Novak, M. Ullrich, H. Woźniakowski, The curse of dimensionality for numerical integration of smooth functions, przyjęta do *Math. Comp.* 2013
2. E. Novak, H. Woźniakowski, *Tractability of Multivariate Problems*, *Vol. I, II, III*, EMS Tracts in Mathematics **6** 2008, **12** 2010, **18** 2012
3. L. Plaskota, *Noisy Information and Computational Complexity*, Cambridge Univ. Press, Cambridge, 1996
4. K. Ritter, *Average Case Analysis of Numerical Problems*, Lecture Notes in Math. **1733**, Springer-Verlag, Berlin, 2000
5. J.F. Traub, G.W. Wasilkowski, H. Woźniakowski, *Information-Based Complexity*, Academic Press, New York, 1988