Metody numeryczne

Wykład 1 - Wprowadzenie

Janusz Szwabiński

Plan wykładu

1. Sprawy administracyjne
2. Warsztat pracy
3. Arytmetyka komputerowa
4. Błędy w obliczeniach numerycznych

Sprawy administracyjne

- Kontakt:
  `https://prac.im.pwr.edu.pl/~szwabin/`
- Materiały do kursu:
  `https://eportal.pwr.edu.pl/course/view.php?id=1820`
- Zasady zaliczenia:
  `https://eportal.pwr.edu.pl/course/view.php?id=1820`

Plan kursu

- Układy równań liniowych
- Równania nieliniowe
- Interpolacja i aproksymacja
- Całkowanie numeryczne
- Różniczkowanie numeryczne
- Równania różniczkowe zwyczajne

Bibliografia

1. Jaan Kiusalaas, *Numerical Methods in Engineering with Python*
2. G. Dahlquist, A. Björk, *Metody numeryczne*
3. pozostałe pozycje jak w karcie przedmiotu

Warsztat pracy

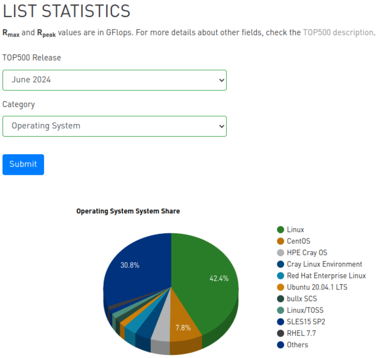
Podstawowe narzędzia:

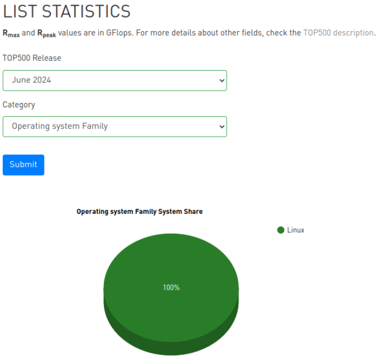
- Python 3.X
- biblioteki numeryczne: numpy, scipy
- inne biblioteki: matplotlib, sympy

Warto znać:

- pygsl
- Maxima, Yacas
- GNU Octave

Warsztat pracy





https://top500.org/statistics/list/

Arytmetyka komputerowa (i jej ograniczenia)

$\boxed{\text{Demo}}$

Reprezentacja stałopozycyjna

Rozważmy liczby w formacie 5-cyfrowym, z dwoma cyframi w części ułamkowej

			,		

Liczba 256,78 ma w tym formacie naturalną reprezentację

2	5	6	,	7	8

Reprezentacja stałopozycyjna

Najmniejsza liczba

0	0	0	,	0	0

Największa liczba

9	9	9	,	9	9

Reprezentacja stałopozycyjna

Liczba 256,786 będzie miała tylko reprezentację przybliżoną

2	5	6	,	7	8

Obcięcie

2	5	6	,	7	9

Zaokrąglenie

- w oby przypadkach błąd mniejszy od 0.01
- ogólnie przy zaokrągleniu błąd średnio dwukrotnie mniejszy niż przy obcięciu

Podstawowe błędy

Błąd bezwględny:

$$|X - X_0|$$

gdzie X_0 to wartość dokładna.

Błąd względny:

$$\frac{|X - X_0|}{X_0}$$

Podstawowe błędy

Przykład

Liczby 256,786 i 3,546 mają takie same błędy bezwzględne w naszej reprezentacji z zaokrąglaniem:

$$
\begin{aligned}
|X^{(1)} - X_0^{(1)}| &= |256.786 - 256.79| = 0,004 \\
|X^{(2)} - X_0^{(2)}| &= |3,546 - 3,55| = 0,004
\end{aligned}
$$

Podstawowe błędy

Przykład

Błędy względne są większe dla małych liczb:

$$
\begin{aligned}
\epsilon_1 &= \frac{256.786 - 256.79}{256.786} * 100\% = 0,001558\% \\
\epsilon_2 &= \frac{3,546 - 3,55}{3,546} * 100\% = 0,11280\%
\end{aligned}
$$

Arytmetyka zmiennopozycyjna

$$\mathit{ZNAK} \times \mathit{MANTYSA} \times 10^{\mathit{WYKLADNIK}}$$

Co zyskujemy?

- zwiększa się zakres liczb możliwych do przedstawienia
- błędy względne małych i dużych liczb są porównywalne

Arytmetyka zmiennopozycyjna

Przykład

Liczba 576329,78 zapisana na 5 miejscach

5	7	6	3	2

$$|X - X_0| = 29,78$$

$$\epsilon_t = 0,0051672\%$$

Liczba 256,78 zapisana na 5 miejscach

2	5	6	8	-1

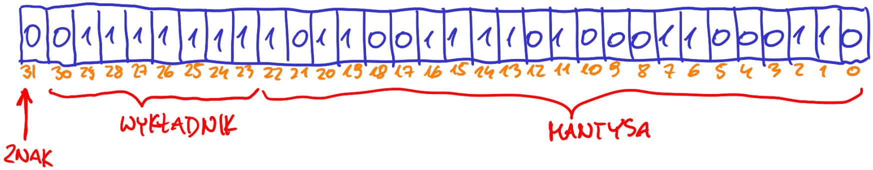
$$|X - X_0| = 0,02$$

$$\epsilon_t = 0,0077888\%$$

Standard IEEE 754

Dla liczb 32-bitowych mamy:

$$(-1)^{b_{31}} \times 2^{(b_{30}b_{29}\ldots b_{23})_2-127} \times (1.b_{22}\ldots b_0)_2$$



Standard IEEE 754

$$(-1)^{ZNAK} \times 2^{E-127} \times \left(1 + \sum_{i=1}^{23} b_{23-i} 2^{i}\right)$$

Jednostka maszynowa

Definicja

Jednostka maszynowa to najmniejsza liczba ϵ taka, że spełniony jest dla niej na komputerze warunek

$$1 + \epsilon \neq 1$$

$\boxed{\text{Demo}}$

Liczby rzeczywiste i maszynowe

Niech $X > 0$ oraz

$$X = q \times 2^m, \quad 1 \leqslant q < 2$$

Wówczas

$$X = (1.a_1 a_2 \ldots)_2 \times 2^m, \quad a_i \in \{0, 1\}$$

Liczby rzeczywiste i maszynowe

Jeśli mantysa liczb maszynowych ma t bitów po kropce:

$\Rightarrow$ liczba maszynowa bliska X powstanie przez odrzucenie zbędnych bitów $a_{t+1}, a_{t+1}, \ldots$

- obcięcie:

$$X_- = (1.a_1 a_2 \ldots a_t)_2 \times 2^m, \quad X_- \leqslant X$$

- zaokrąglenie: odrzucamy zbędne bity i jednoczesnie dodajemy jedynkę na ostatniej pozycji

$$X_+ = \left[(1.a_1 a_2 \ldots a_t)_2 + 2^{-t}\right] \times 2^m, \quad X_+ - X_- = 2^{m-t}$$

Liczby rzeczywiste i maszynowe

Definicja

Najbliższa względem X liczba maszynowa $fl(X)$ to bliższa z liczb X_+ i X_-.

Jeśli $fl(X) = X_-$, wówczas

$$|X - fl(X)| \leqslant \frac{1}{2}|X_+ - X_-| = 2^{m-t-1}$$

Podobnie, jeśli $fl(X) = X_+$, to

$$|X - fl(X)| \leqslant \frac{1}{2}|X_+ - X_-| = 2^{m-t-1}$$

Liczby rzeczywiste i maszynowe

Błąd względny:

$$\frac{|X - fl(X)|}{|X|} \leqslant \frac{2^{m-t-1}}{q2^m} = \frac{1}{q}2^{-t-1} \leqslant 2^{-t-1}$$

Niech

$$\delta = \frac{fl(X) - X}{X}$$

Wówczas

$$fl(X) = X(1 + \delta), \quad |\delta| \leqslant \epsilon$$

Liczby rzeczywiste i maszynowe

Przykład

Niech $X = \frac{2}{3}$:

- Jaka jest postać dwójkowa X dla $t = 23$?
- Ile wynoszą liczby X_- i X_+?
- Która z tych liczb będzie $fl(X)$?
- Jaki będzie błąd przybliżenia?

Liczby rzeczywiste i maszynowe

$$\frac{2}{3} = (0.a_1a_2\ldots)_2 \quad / \cdot 2$$

$$\frac{4}{3} = (a_1.a_2a_3\ldots)_2$$

Część całkowita obu stron jest równa $1 = a_1$. Po jej odjęciu mamy

$$\frac{1}{3} = (0.a_2a_3\ldots)_2$$

Powtarzając obustronne mnożenie przez 2 i ewentualne odejmowanie części całkowitej otrzymamy

$$X = \frac{2}{3} = (0.101010\ldots)_2 = (1.010101\ldots)_2 \times 2^{-1}$$

Liczby rzeczywiste i maszynowe

Dwie bliskie liczby maszynowe:

$$
\begin{aligned}
X_- &= (1.\underbrace{010101\ldots010}_{t=23\text{ bitów}})_2 \times 2^{-1} \\
X_+ &= (1.\underbrace{010101\ldots011}_{t\text{ bitów}})_2 \times 2^{-1}
\end{aligned}
$$

Obliczmy różnice

$$
\begin{aligned}
X - X_- &= (0.101010\ldots)_2 \times 2^{-24} = \frac{2}{3} \times 2^{-24} \\
X_+ - X &= X_+ - X_- - (X - X_-) = 2^{-24} - \frac{2}{3} \times 2^{-24} = \frac{1}{3} \times 2^{-24}
\end{aligned}
$$

Liczby rzeczywiste i maszynowe

Czyli

$$fl(X) = X_+$$

Błąd bezwględny

$$|fl(X) - X| = \frac{1}{3} \times 2^{-24}$$

Błąd względny

$$\frac{|fl(X) - X|}{|X|} = \frac{\frac{1}{3} \times 2^{-24}}{\frac{2}{3}} = 2^{-23}$$

Działania arytmetyczne

Chcemy obliczyć

$$X \odot Y,$$

gdzie $\odot$ oznacza jedno z działań arytmetycznych

Niech:

- X i Y będą liczbami maszynowymi
- mantysa wyniku jest normalizowana i zaokrąglana, tzn. wynikiem działania jest $fl(X \odot Y)$

Działania arytmetyczne

Przykład

Komputer działający na liczbach z mantysą pięciocyfrową z przedziału $[0.1, 1)$. Niech:

$$\begin{aligned} X &= 0.31426_{10}\,3 \\ Y &= 0.92577_{10}\,5 \end{aligned}$$

Działania arytmetyczne

Załóżmy, że surowe wyniki umieszczane są w akumulatorze podwójnej długości (typowe w nowoczesnych komputerach):

$$
\begin{aligned}
X + Y &= 0.9289126000_{10}\,5 \\
X - Y &= -0.9226274000_{10}\,5 \\
X \times Y &= 0.2909324802_{10}\,8 \\
X/Y &= 0.3394579647_{10}\,-2
\end{aligned}
$$

Działania arytmetyczne

Po zaokrągleniu mantys:

$$\left.\begin{array}{rl} X+Y = 0.92891_{10}\,5 & \delta = 2.8_{10}\ -6 \\ X-Y = -0.92263_{10}\,5 & \delta = 2.8_{10}\ -6 \\ X\times Y = 0.29093_{10}\,8 & \delta = 8.5_{10}\ -6 \\ X/Y = 0.33946_{10}\ -2 & \delta = 6.0_{10}\ -6 \end{array}\right\}\ \delta_{\odot}\ \text{mniejsze od } 10^{-5}$$

W dobrze zaprojektowanym komputerze:

$$fl(X\odot Y) = (X\odot Y)(1+\delta),\quad |\delta| \leqslant \epsilon \quad \forall\odot$$

Działania arytmetyczne

Jeśli X i Y nie są liczbami maszynowymi:

$$fl\,(fl(X) \odot fl(Y)) = [X(1+\delta_1) \odot Y(1+\delta_2)]\,(1+\delta_3), \quad |\delta_i| \leqslant \epsilon$$

Błąd względny wyrażeń arytmetycznych

Niech X, Y i Z będą liczbami maszynowymi:

$$
\begin{aligned}
fl(X(Y+Z)) &= [X * fl(Y+Z)]\,(1+\delta_1) = [X * (Y+Z)(1+\delta_2)]\,(1+\delta_1) \\
&= X(Y+Z)(1+\delta_1+\delta_2+\delta_1\delta_2) \simeq X(Y+Z)(1+\delta_1+\delta_2) \\
&= X(Y+Z)(1+\delta_3)
\end{aligned}
$$

$$
|\delta_1| \leqslant \epsilon, \quad |\delta_2| \leqslant \epsilon, \quad |\delta_3| \leqslant 2\epsilon
$$

Błąd względny wyrażeń arytmetycznych

Twierdzenie

Jeśli $X_0, X_1, \ldots, X_n$ są dodatnimi liczbami maszynowymi, to błąd względny sumy $\sum_{i=0}^{n} X_i$ jest równy co najwyżej

$$(1+\epsilon)^n - 1 \simeq n\epsilon.$$

Utrata cyfr znaczących

Przykład

Niech

$$X = 0.3721478693, \quad Y = 0.3720230772$$

$$X - Y = 0.0001248121$$

Utrata cyfr znaczących

Przykład

W obliczeniach z 5-cyfrowymi mantysami:

$$fl(X) = 0.37215, \quad fl(Y) = 0.37202$$

$$fl(X) - fl(Y) = 0.00013$$

Różnica ma mniej cyfr znaczących w porównaniu z odjemną i odjemnikiem $\Rightarrow$ duży błąd względny!

$$\frac{|X - Y - [fl(X) - fl(Y)]|}{x - y} \simeq 0.04$$

Utrata cyfr znaczących

Twierdzenie

Jeśli liczby maszynowe X i Y są takie, że $X > Y > 0$ oraz

$$2^{-q} \leqslant 1 - \frac{Y}{X} \leqslant 2^{-p}$$

(p i q są całkowite), to liczba bitów znaczących straconych przy odejmowaniu $X - Y$ jest równa co najmniej p i co najwyżej q.

- utraty cyfr znaczących można uniknąć odpowiednio planując obliczenia!
- w szczególności należy unikać odejmowania bliskich sobie liczb

Utrata cyfr znaczących

$\boxed{\text{Demo}}$

Niestabilność numeryczna

Definicja

Algorytm numeryczny jest niestabilny, jeżeli małe błędy popełnione na jakimś etapie obliczeń rosną w następnych etapach.

Niestabilność numeryczna

Przykład

Rozważmy ciąg

$$x_0 = 1, \quad x_1 = \frac{1}{3}, \quad x_{n+1} = \frac{13}{3}x_n - \frac{4}{3}x_{n-1}, \quad n > 1$$

Niestabilność numeryczna

Jeśli będziemy liczyć wyrazy ciągu w arytmetyce z 24-bitowymi mantysami:

$X_0 = 1.0000000$	
$X_1 = 0.3333333$	(7 cyfr znaczących)
$X_2 = 0.1111112$	(6 cyfr znaczących)
$X_3 = 0.0370373$	(5 cyfr znaczących)
$X_4 = 0.0123466$	(4 cyfr znaczących)
$X_5 = 0.0041187$	(3 cyfr znaczących)
$X_6 = 0.0013857$	(2 cyfr znaczących)
$X_7 = 0.0005131$	(1 cyfra znacząca)
$X_8 = 0.0003757$	(brak cyfr znaczących)
$\vdots$	
$X_{15} = 3.657493$	(błąd względny 10^8)

Niestabilność numeryczna

Można pokazać, że rozważany ciąg równoważny jest ciągowi o wyrazach

$$x_n = \left(\frac{1}{3}\right)^n, \quad n \geqslant 0$$

Niestabilność numeryczna

$\boxed{\text{Demo}}$

Uwarunkowanie

Uwarunkowanie to wrażliwość rozwiązania zadania na małe zmiany danych początkowych:

$$\begin{aligned} a &\rightarrow a + \delta a \\ W(a) &\rightarrow W(a + \delta a) \end{aligned}$$

Niech w będzie wektorem wyników oraz

$$\delta w = \underbrace{WN(a, \epsilon)}_{\text{wynik numeryczny}} - W(a)$$

Uwarunkowanie

Wskaźnik uwarunkowania $B(a)$:

$$\frac{\|\delta w\|}{\|w\|} \leqslant B(a) \frac{\|\delta a\|}{\|a\|}$$

Uwarunkowanie

Jeśli naszym zadaniem jest policzenie funkcji:

$$f(\underbrace{x+h}_{\text{zaburzenie}}) = \underbrace{f'(\xi)h}_{\text{tw. o wart. śr.}} \simeq f'(x)h$$

Wówczas

$$\frac{f(x+h)-f(x)}{f(x)} \simeq \frac{hf'(x)}{f(x)} = \underbrace{\frac{xf'(x)}{f(x)}}_{B(x)}\left(\frac{h}{x}\right)$$

Źródła błędów w obliczeniach numerycznych

- błędy wejściowe
- błędy obcięcia
- błędy zaokrągleń

Błędy wejściowe

- dane wejściowe są wynikiem pomiarów
- skończona długość słów binarnych
- wstępne zaokrąglanie liczb niewymiernych

$$\pi = 3.14\ldots$$

Warto wiedzieć

$$\pi = 4 * arctg(1.0)$$

Błędy obcięcia

Spowodowane użyciem przybliżonej formuły zamiast pełnej operacji matematycznej

- przejścia graniczne (np. pochodne i całki oznaczone)
- sumy nieskończone szeregów

Błędy zaokrągleń

- skończona długość słów binarnych