

Lecture 1

October 14, 2019

```
[1]: %%HTML
<style>
.rendered_html table, .rendered_html th, .rendered_html tr, .rendered_html td {
    font-size: 100%;
}
</style>
```

<IPython.core.display.HTML object>

1 Metody Numeryczne

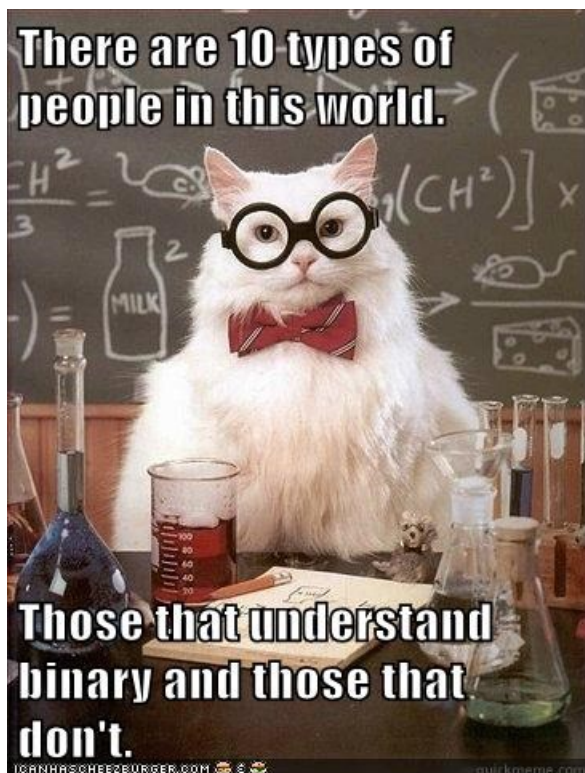
1.1 Elementy analizy numerycznej

1.1.1 dr hab. in. Jerzy Baranowski, Prof. nadzw.

1.2 Informacje ogólne

- Katedra Automatyki i Robotyki, C3, p. 214
- Konsultacje
 - Czwartki 11:00-12:00 (o ile nie ma Kolegium Wydziałowego lub seminarium)
- jb@agh.edu.pl
- wykady dostępne tutaj: https://github.com/KAIR-ISZ/public_lectures

2 Reprezentacja liczb



2.1 Kod binarny

- Zapis liczby z wykorzystaniem dwóch symboli **1** i **0**
- Podstawa współczesnego sposobu reprezentacji informacji

2.2 Zamierzcha historia

- Pingala, Chandastra i Prozodia
 - Ok. 4 wiek pne
 - Wykorzystanie zapisu w formie zer i jedynek do opisu metrum
- Chiny, hexagramy, Shao Yong, I-Ching
- Leibniz

2.3 Algebra Boole'a

$x \wedge y = xy$	Koniunkcja
$x \vee y = x + y - xy$	Alternatywa
$\neg x = 1 - x$	Negacja
$x \rightarrow y = (\neg x \vee y)$	Implikacja
$x \oplus y = (x \vee y) \wedge \neg(x \wedge y)$	EXOR
$x = y = \neg(x \oplus y)$	Równowartość

2.4 Nieco mniej zamierzcha historia

- 1937 Shannon – przekanikowa realizacja operacji binarnych i algebry Boole’a
- 1937 Stibitz – Pierwszy komputer przekanikowy (dodawanie)

2.5 Kod binarny

	0	0	1	0	1	0	1	1
2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0	

Co daje $\$ = 2^{5+2}3 + 2^{1+2}0 = 32 + 8 + 2 + 1 = 43\$$

2.6 Liczby naturalne

- Ogólnie zakres od 0 do $2^n - 1$
- 8 bit – zakres od 0 do 255
- 16 bit – zakres od 0 do 65,535 (short, int)
- 32 bit – zakres od 0 do 4,294,967,295 (long)

W Pythonie i matlabie za bardzo nie przejmujemy si typami, chyba e je wymusimy

2.7 Operacje na liczbach binarnych

- Dodawanie
 - $0+0=0$
 - $0+1=1$
 - $1+0=1$
 - $1+1=0$, przenie 1
- Jak w dodawaniu pisemnym

```
  1 1 1 1 1(cyfry przenoszone)
0 1 1 0 1(1310)
+ 1 0 1 1 1(2310)
-----
=1 0 0 1 0 0 (3610)
```

2.8 Operacje na liczbach binarnych

- Odejmowanie
 - $0-0=0$
 - $0-1=1$, poyczka 1
 - $1-0=1$
 - $1-1=0$,
- Analogicznie

$$\begin{array}{r}
 \\
 * & * & * & *(poyczki) \\
 1\ 1\ 0\ 1\ 1\ 1\ 0(11010) \\
 - & & 1\ 0\ 1\ 1\ 1(2310) \\
 ----- \\
 = 1\ 0\ 1\ 0\ 1\ 1\ 1(8710)
 \end{array}$$

2.9 Co z liczbami ujemnymi?

Uzupeniamy zapis o tzw. bit znaku

$$\begin{array}{cccccccc} & \overline{1} & \overline{0} & \overline{1} & \overline{0} & \overline{1} & \overline{0} & \overline{1} & \overline{1} \\ S & 2^6 & 2^5 & 2^4 & 2^3 & 2^2 & 2^1 & 2^0 & \end{array}$$

Co daje $\$ = (-1)^{1(2 \cdot 3 + 2^{1+2} \cdot 0)} = -(8 + 2 + 1) = -11\$$

Zmieniaj si zakresy: - 8 bit (-128 do 127) - 16 bit (32,768 do 32,767) - itd

2.10 Problemy

- Niepraktyczny zapis
- Trzeba przekodowywać wyniki operacji
- Potencjalnie podatniejsze na bdy

2.11 Kod uzupełnienia do 2 (U2)

$$-2^7 \quad \begin{array}{ccccccc} \hline \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{1} & \mathbf{0} & \mathbf{1} & \mathbf{1} \\ \hline 2^6 & 2^5 & 2^4 & 2^3 & 2^2 & 2^1 & 2^0 \end{array}$$

Co daje $\$ = -2^{7+2}6 + 2^{5+2}4 + 2^{3+2}1 + 2^0\$$

$$= -128 + 64 + 32 + 16 + 8 + 2 + 1 = -5$$

2.12 Bardzo atwa konwersja

- Liczby dodatnie s takie same jak byy
- Aby zamieni liczb na jej przeciwn wystarczy zanegowa wszystkie bity i do wyniku doda 1 (w obie strony)

	0	0	0	0	0	1	0	1	510	orygina
1	1	1	1	1	0	1	0			negacja
1	1	1	1	1	0	1	1	-510		dodanie 1
0	0	0	0	0	1	0	0			negacja
0	0	0	0	0	1	0	1	510		dodanie 1
-2^7	2^6	2^5	2^4	2^3	2^2	2^1	2^0			

2.13 Jaka z tego korzy?

- Odejmowanie staje się dodawaniem (prawie)

$$A - B = A + \neg B + 1$$

- Przykład $13 - 7$ (na 8 bitach)

$$\begin{array}{r}
 1\ 1\ 1\ 1 \quad 1(\text{cyfry przenoszone}) \\
 0\ 0\ 0\ 0\ 1\ 1\ 0\ 1(1310) \\
 1\ 1\ 1\ 1\ 1\ 0\ 0\ 0(\text{zanegowane } 710) \\
 + \quad 1(\text{jedynka}) \\
 \hline
 = 0\ 0\ 0\ 0\ 0\ 1\ 1\ 0(610)
 \end{array}$$

2.14 Operacje na liczbach binarnych

Mnożenie równie przypomina mnożenie pisemne

$$\begin{array}{r}
 1\ 0\ 1\ 1110 \\
 * 1\ 0\ 1\ 01010 \\
 \hline
 0\ 0\ 0\ 0 \\
 + \quad 1\ 0\ 1\ 1 \\
 + \quad 0\ 0\ 0\ 0 \\
 + \quad 1\ 0\ 1\ 1 \\
 \hline
 = 1\ 1\ 0\ 1\ 1\ 1\ 011010
 \end{array}$$

2.15 A co z ułamkami?

Są dwa sposoby zapisu liczb niecałkowitych - Staoprzecinkowy (staopozycyjny) - Zmiennoprzecinkowy (zmiennopozycyjny)

2.16 Zapis staoprzecinkowy

$$\begin{array}{cccccccc}
 & & \overline{1} & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\
 2^1 & 2^0 & 2^{-1} & 2^{-2} & 2^{-3} & 2^{-4} & 2^{-5} & 2^{-6} & &
 \end{array}$$

$$2^1 + 2^{-1} + 2^{-2} + 2^{-3} = 2 + \frac{1}{2} + \frac{1}{4} + \frac{1}{8} = 2.875$$

2.17 Zalety zapisu staoprzecinkowego

- Nie ma różnicy w kodowaniu
- Mamy stałe określenie dokładności, którą możemy w miarę dokładności kształtować
- Stosunkowo prostota
- Małe wymagania sprzętowe

2.18 Wady zapisu staoprzecinkowego

Problemy z dokadnoci, np. nie da si dokadnie przedstawi liczby 0.1 - Na 3 bitach czci uamkowej rónica wynosi 0.025 - Na 7 bitach czci uamkowej rónica wynosi ok. 0.001

2.19 Jak wykonujemy dziaania?

- Działania wykonujemy traktując zapis liczby staoprzecinkowej jako normaln binarn
- Kod U2 dalej dziaa
- Naley pamita, e wtedy liczba jest pomnoona przez 2^n gdzie n to ilo bitów czci uamkowej
- W liczbach poddanych działaniu liczba bitów czci cakowitej i uamkowej musi by równa

2.20 Dziaania staoprzecinkowe

- Dodawanie wykonujemy identycznie
- W przypadku mnożenia wynik musimy podzielić przez $2n$
- Mnożenie liczb staoprzecinkowych przez potęg 2 polega tylko na przesuwaniu bitów (bardzo proste w realizacji)

				1	0	1	1	1	0	0	0	
0	0	1	0	1	1	1	0	0	Podzielenie przez 2^2			
2^1	2^0	2^{-1}	2^{-2}	2^{-3}	2^{-4}	2^{-5}	2^{-6}					

2.21 Format zmiennoprzecinkowy

- Bardziej zaawansowany sposób przedstawiania liczb
- Ustandaryzowany norm IEEE
- Dajcy pod pewnymi wzgladami wiksza dokadno

2.22 Format zmiennoprzecinkowy

Reprezentacja liczby

$$x = S \cdot M \cdot B^E$$

- S – znak (*sign*)
- M – mantysa (*mantissa*, *take fraction*)
- B – podstawa (*base*, zazwyczaj 2, rzadziej 10)
- E - wykładnik (*exponent*)

2.23 Mantysa

- Liczba odpowiadająca za uamkow cz zapisu
- Format staoprzecinkowy, zazwyczaj liczba z przedziału [1,2)

2.24 Podstawa i wykadnik

- Pozwalaj na określenie szerokiego zakresu

NumPy - float64, ale w zasadzie każda liczba w Pythonie i Matlabie to double, chyba że wymusimy inaczej

2.29 Wyświetlanie liczb

- Normalnie
– $3700 = 3.7 \cdot 10^3$, $0.12 = 120 \cdot 10^{-3}$
- Notacja inżynierska
– $3700=3.7E3$, $0.12=1.2E-1$

3 Błędy numeryczne

3.1 Podstawowe definicje

Warto dokładna

$$y = \tilde{y} + \varepsilon$$

- \tilde{y} - wartość przybliżona - ε - błąd

3.2 Błąd bezwzględny

Warto bezwzględna różnicy między rozwiązaniem dokładnym i przybliżonym

$$\varepsilon = |y - \tilde{y}|$$

3.3 Błąd względny

Stosunek błęd bezwzględnego do wartości bezwzględnej rozwiązania

$$\eta = \frac{|y - \tilde{y}|}{|y|} = \left| \frac{y - \tilde{y}}{y} \right| = \left| 1 - \frac{\tilde{y}}{y} \right|$$

Czasami błąd względny wyrażamy w procentach

3.4 Przykłady

Pierwiastek kwadratowy ze 122

$$y = \sqrt{122} \approx 11.04536$$

$$\tilde{y} = 11$$

$$\varepsilon = |y - \tilde{y}| = 0.04536$$

$$\eta = \frac{|y - \tilde{y}|}{|y|} = 0.00411$$

3.5 Przykłady

Liczba obywateli Polski (stan na ostatni spis powszechny z 2011)

$$y = 38\,538\,447$$

$$\tilde{y} = 38\,500\,000$$

$$\varepsilon = |y - \tilde{y}| = 38\,447$$

$$\eta = \frac{|y - \tilde{y}|}{|y|} = 9.97627 \cdot 10^{-4} \approx 0.001$$

3.6 Przykłady

Obliczanie stałej grawitacji

$$y = 6.673841 \cdot 10^{-11}$$

$$\tilde{y} = 6.7 \cdot 10^{-11}$$

$$\varepsilon = |y - \tilde{y}| = 2.6159 \cdot 10^{-13}$$

$$\eta = \frac{|y - \tilde{y}|}{|y|} = 0.00391$$

3.7 Źródła błędów

Błędy powstające przy formułowaniu zadania - Błędy pomiaru - Błędy wynikające z przyjęcia określonych przybliżeń opisu zjawisk fizycznych

Błędy powstające przy obliczeniach - Błędy grube (pomyłki) - Błędy metody (obcicia) - Błędy zaokrąglenia

3.8 Błędy grube

- Błąd przy wpisywaniu wzoru do komputera np. $x=A/b$ zamiast $x=A \setminus b$
- Zła implementacja algorytmu
- Niewłaściwa kolejność wykonywania działań

3.9 Błędy metody (obcicia)

- Błąd obcicia jest nieodczynnym elementem obliczeń numerycznych.
- Błąd obcicia jest to błąd wynikający z tego, że do uzyskania dokładnego rozwiązania potrzebujemy wykonać nieskończenie wiele obliczeń

3.10 Przykłady błędów metody

Mona wykazać, że

$$\begin{aligned}\sin x &= x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \dots = \\ &= \sum_{n=0}^{\infty} (-1)^n \frac{x^{2n+1}}{(2n+1)!}\end{aligned}$$

Błąd obcicia będzie

$$\sin x \approx x - \frac{x^3}{3!} + \frac{x^5}{5!}$$

3.11 Przykłady bđów metody

Metoda bisekcji

```
[2]: def bisection(f,a,b,N):  
    a_n = a  
    b_n = b  
    for n in range(1,N+1):  
        m_n = (a_n + b_n)/2  
        f_m_n = f(m_n)  
        if f(a_n)*f_m_n < 0:  
            a_n = a_n  
            b_n = m_n  
        elif f(b_n)*f_m_n < 0:  
            a_n = m_n  
            b_n = b_n  
    return (a_n + b_n)/2
```

Szukamy pierwiastka wielomianu $x^2 - 2$, w przedziale $[1,2]$. Rozwiązanie to $\sqrt{2}$.

```
[3]: f = lambda x: x**2 - 2 # definicja funkcji  
bisection(f,1,2,5) # 5 kroków
```

```
[3]: 1.421875
```

```
[4]: bisection(f,1,2,10) # 10 kroków
```

```
[4]: 1.41455078125
```

```
[5]: bisection(f,1,2,15) # 15 kroków
```

```
[5]: 1.4141998291015625
```

```
[6]: import numpy as np  
np.sqrt(2)
```

```
[6]: 1.4142135623730951
```

3.12 Bđ metody - podsumowanie

- Praktycznie wszystkie metody numeryczne mają jakiegoś bđ metody
- Dobre algorytmy podają jednak jego oszacowanie, w ten sposób wiemy jak daleko jesteśmy od rozwiązania nawet jak przerwiemy obliczenia

3.13 Bdy zaokrąglenia

Kolejne nieusuwalne w pełni źródło błędów, nad którym mamy mniej kontrol niż nad bđem metody

3.14 Zaokrąglenie i cyfry znaczące

Liczba $\tilde{y} = \text{rd}(y)$ jest poprawnie zaokrąglona do d miejsc po przecinku, jeżeli

$$\varepsilon = |y - \tilde{y}| \leq \frac{1}{2} \cdot 10^{-d}$$

k -t cyfr dziesiętn liczby \tilde{y} nazwiemy znacząc gdy

$$|y - \tilde{y}| \leq \frac{1}{2} \cdot 10^{-k}$$

oraz

$$|\tilde{y}| \geq 10^{-k}$$

3.15 Rzeczywiste obliczenia zmiennoprzecinkowe

$$\text{fl}(x + y) = \text{rd}(x + y)$$

$$\text{fl}(x - y) = \text{rd}(x - y)$$

$$\text{fl}(x \cdot y) = \text{rd}(x \cdot y)$$

$$\text{fl}(x/y) = \text{rd}(x/y)$$

3.16 Liczby maszynowe

- Liczba maszynowa, to taka liczba jak można przedstawić w komputerze. Zbiór tych liczb oznaczamy A
- Dokładność maszynowa (epsilon maszynowy) – ϵ_m , definiujemy:

$$\epsilon_m = \min\{x \in A : \text{fl}(1 + x) > 1, x > 0\}$$

Innymi słowami, jest to najmniejsza liczba, którą możemy dodać do 1, aby uzyskać coś większego od 1.

3.17 Epsilon maszynowy w różnych formatach

Zależy on od liczby bitów na część ułamkową - Single precision $\epsilon_m = 2^{-24} \approx 5.96 \cdot 10^{-8}$ - Double precision $\epsilon_m = 2^{-52} \approx 1.11 \cdot 10^{-16}$

Przykład

```
[7]: a=10**(-15)
     b=10**(-17)
     1+a>1, 1+b>1
```

```
[7]: (True, False)
```

3.18 Maksymalny błąd reprezentacji

Dla każdej liczby rzeczywistej x istnieje taka liczba ϵ , taka że $|\epsilon| < \epsilon_m$, że $\text{fl}(x) = x(1 + \epsilon)$

Oznacza to, że błąd względny między liczbą rzeczywistą, a jej najbliższą reprezentacją zmiennoprzecinkową jest zawsze mniejszy od ϵ_m

3.19 Lemat Wilkinsona

Błędy zaokrąglenia powstają podczas wykonywania działań zmiennoprzecinkowych są równoważone zastąpieniu zaburzeniu liczb, na których wykonujemy działania

$$\begin{aligned}\text{fl}(x + y) &= (x + y)(1 + \varepsilon_1) \\ \text{fl}(x - y) &= (x - y)(1 + \varepsilon_2) \\ \text{fl}(x \cdot y) &= (x \cdot y)(1 + \varepsilon_3) \\ \text{fl}(x/y) &= (x/y)(1 + \varepsilon_4) \\ |\varepsilon_i| &< \varepsilon_m\end{aligned}$$

(dla każdej pary liczb x, y zaburzenia zastępcze ε_i s inne)

3.20 Konsekwencja lematu Wilkinsona

Prawa czności i rozdzielności operacji matematycznych s ogólnie nieprawdziwe dla oblicze zmiennoprzecinkowych

3.20.1 Przykad

```
[8]: a=np.float32(0.23371258*10**(-4))
     b=np.float32(0.33678429*10**(2))
     c=np.float32(-0.33677811*10**(2))
     print([a,b,c])
```

```
[2.3371258e-05, 33.67843, -33.67781]
```

Chcemy obliczy $a+b+c$

3.21 Obliczenia

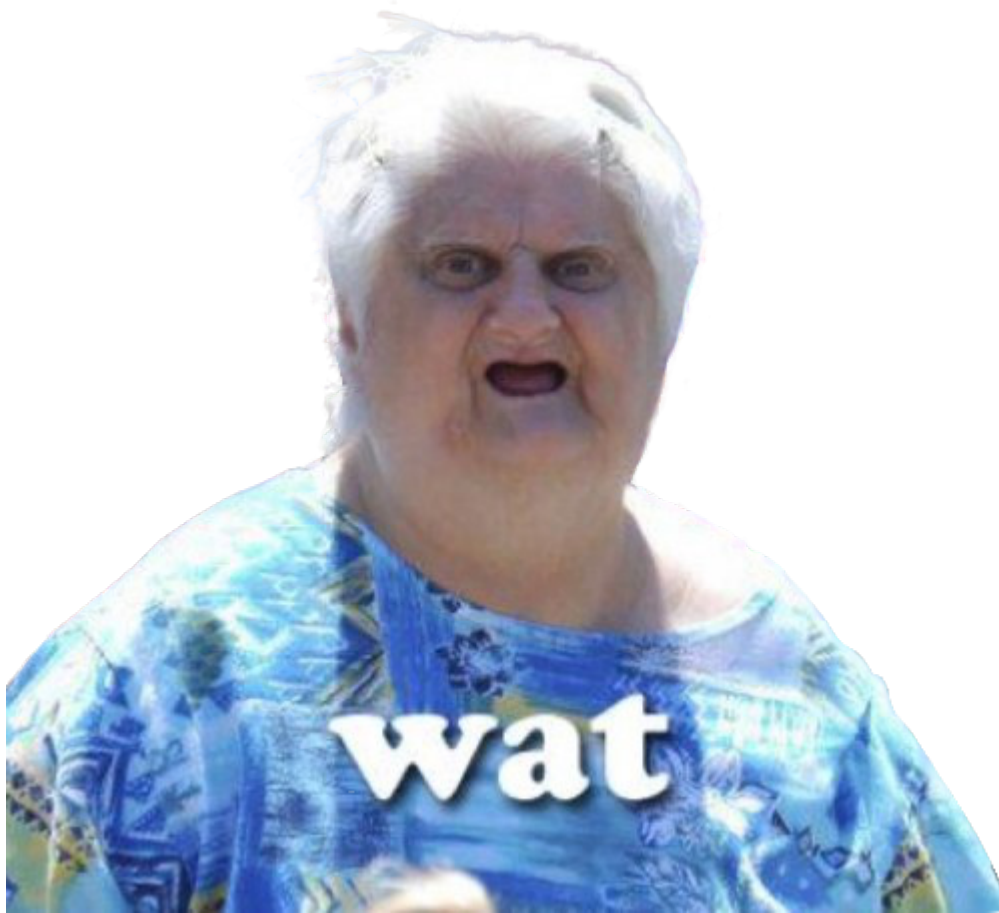
```
[9]: ## Podejcie 1
     d=b+c
     wynik_1=a+d
     print(wynik_1)
```

```
0.0006413522
```

```
[10]: ## Podejcie 2
      e=a+b
      wynik_2=e+c
      print(wynik_2)
```

```
0.00064086914
```

3.22 Co tu si porobio?



3.23 Konsekwencje obliczen zmiennoprzecinkowych

```
[11]: m_a, e_a = np.frexp(a)
      print(m_a, e_a)
      m_b, e_b = np.frexp(b)
      print(m_b, e_b)
      m_c, e_c = np.frexp(c)
      print(m_c, e_c)
```

```
0.7658294 -15
0.52622545 6
-0.5262158 6
```

Wykadnik a od wykadników b i c różni się o 21. Oznacza to, że z 23 bitów mantysy liczby a po sprowadzeniu do wspólnego wykadnika z b zostaną nam tylko 2 najbardziej znaczące.

3.24 Konsekwencje cd..

Jeeli dodajemy ma liczb do duej, zawsze musimy si liczy z zaokrgleniem i to normalne. W tym przypadku jednak dwie due liczby b i c s przeciwnych znaków i bliskie co do wartoci bezwzgldnej. Wynik tego dziaania:

```
[12]: m_d,e_d = np.frexp(d)
      print(m_d,e_d)
      print(wynik_2)
```

```
0.6328125 -10
0.00064086914
```

W konsekwencji dodajc a do d na zaokrgleniu stracimy jedynie 5 bitów mantysy a.

3.25 O ile si pomyliliśmy (w stosunku do dokadniejszych oblicze)

```
[13]: a_dbl=(0.23371258*10**(-4))
      b_dbl=(0.33678429*10**(2))
      c_dbl=(-0.33677811*10**(2))
      d_dbl=b_dbl+c_dbl
      wynik_dbl=a_dbl+d_dbl
      epsilon_1=np.abs((wynik_1)-wynik_dbl)
      eta_1=epsilon_1/np.abs(wynik_dbl)
      print("Metoda 1: Bd bezwzgldny %10.2e, Bd wzgldny %10.2e"%(epsilon_1,eta_1))
      epsilon_2=np.abs((wynik_2)-wynik_dbl)
      eta_2=epsilon_2/np.abs(wynik_dbl)
      print("Metoda 2: Bd bezwzgldny %10.2e, Bd wzgldny %10.2e"%(epsilon_2,eta_2))
```

```
Metoda 1: Bd bezwzgldny    1.91e-08, Bd wzgldny    2.97e-05
Metoda 2: Bd bezwzgldny    5.02e-07, Bd wzgldny    7.83e-04
```

4 Analiza algorytmów numerycznych

4.1 Notacja O due

- Mówimy, e dla wielkoci zalenej od parametru np. $F(n)$ zachodzi

$$F(n) = O(G(n))$$

jeeli istnieje taka staa C , e przy n zmierzajcym do nieskoczonoci (odpowiednio duym), mamy

$$F(n) \leq C G(n)$$

- Jeeli interesuje nas $O(c)$, gdzie c jest sta, zaleno ta ma zachodzi niezalenie od wielkoci parametru.
- Mówimy potocznie, gdy bd jest równy $O(n^2)$, e bd jest rzdu n^2

4.2 Ocena algorytmu

- Naszym celem jest obliczenie pewnej wielkości $f(x)$, zależnej od danych wejściowych x
- W przypadku obliczeń komputerowych zawsze mamy do czynienia z obliczaniem przybliżonym. Standardowy algorytm obliczania $f(x)$ będziemy oznaczać jako $f^*(x)$
- Dane w komputerze również są reprezentowane w sposób zaokrąglony, więc będziemy je oznaczać jako x^*

4.3 Uwarunkowanie problemu

- Mówimy, że problem $f(x)$ jest dobrze uwarunkowany, jeżeli mała zmiana x powoduje małe zmiany w $f(x)$
- Problem jest źle uwarunkowany, jeżeli mała zmiana x powoduje duże zmiany w $f(x)$
- Miara uwarunkowania jest stała κ (kappa), która (nieformalnie) określa największy iloraz zaburzeń $f(x)$ wywołanych przez najmniejsze zaburzenia x .
- Stała κ można wyliczyć tylko w niektórych przypadkach

4.4 Dokładność algorytmu

- Algorytm jest dokładny, jeżeli

$$\frac{\|f^*(x) - f(x)\|}{\|f(x)\|} = O(\varepsilon_m)$$

- Zagwarantowanie, że algorytm jest dokładny wg tej definicji jest niezwykle trudne, zwłaszcza dla źle uwarunkowanych problemów

4.5 Stabilność algorytmu

Mówimy, że algorytm jest stabilny, gdy dla każdego x , zachodzi

$$\frac{\|f^*(x) - f(x^*)\|}{\|f(x^*)\|} = O(\varepsilon_m)$$

dla takich x^* , że

$$\frac{\|x - x^*\|}{\|x\|} = O(\varepsilon_m)$$

Innymi słowami **Stabilny algorytm daje prawie dobrą odpowiedź na prawie dobre pytanie**

4.6 Stabilność wsteczna algorytmu

Algorytm jest stabilny wstecznie, jeżeli dla każdego x , zachodzi

$$f^*(x) = f(x^*)$$

dla takich x^* , że

$$\frac{\|x - x^*\|}{\|x\|} = O(\varepsilon_m)$$

Innymi słowami **Stabilny wstecznie algorytm daje prawidłową odpowiedź na prawie dobre pytanie**

4.7 Dokładność algorytmów stabilnych wstecznie przy złym uwarunkowaniu

Jeśli algorytm jest stabilny wstecznie, to jego błąd względny pogarsza się proporcjonalnie do stałej uwarunkowania tj. $O(\kappa \varepsilon_m)$