Politechnika Wrocławska

Metody numeryczne w fizyce

W11OPA-SM0060G

FZP002934wcL
rok akademicki 2022/23
semestr letni

Wykład 1

Karol Tarnowski

karol.tarnowski@pwr.edu.pl

L-1 p. 220

Politechnika Wrocławska

Plan wykładu (1)

- Zapis liczb w różnych układach
- Typy danych numerycznych
- Reprezentacja zmiennoprzecinkowa liczb
- Dokładność operacji na liczbach zmiennoprzecinkowych

Na podstawie:

- D. Kincaid, W. Cheney, *Analiza numeryczna*

Politechnika Wrocławska

Plan wykładu (2)

- Metody wyznaczania miejsc zerowych
  - Metoda bisekcji
  - Metoda Newtona i metoda siecznych
- Zbieżność
- Notacja *O* i *o*

Na podstawie:

- D. Kincaid, W. Cheney, *Analiza numeryczna*

Arytmetyka komputerowa

Zapis liczb w różnych układach

$814{,}72 =$

$= 8\times10^{2} + 1\times10^{1} + 4\times10^{0} + 7\times10^{-1} + 2\times10^{-2}$

$(1110{,}10100)_2 =$

$= 1\times2^{3} + 1\times2^{2} + 1\times2^{1} + 0\times2^{0} + 1\times2^{-1} +$

$0\times2^{-2} + 1\times2^{-3} + 0\times2^{-4} + 0\times2^{-5} =$

$= 8 + 4 + 2 + 0{,}5 + 0{,}125 = (14{,}625)_{10}$

$\phi = 1{,}618033988\ 7\ldots$

Politechnika Wrocławska

Arytmetyka komputerowa

Zapis liczb w różnych układach

$$\frac{1}{2} = (0,5)_{10} = (0,1)_2$$

$$\frac{1}{10} = (0,1)_{10} = (0,00011001100\ldots)_2$$

$$\frac{1}{3} = (0,33333\ldots)_{10} = (0,01010101\ldots)_2$$

Politechnika Wrocławska

Typy danych

- całkowite
- zmiennoprzecinkowe

Politechnika Wrocławska

Typy danych

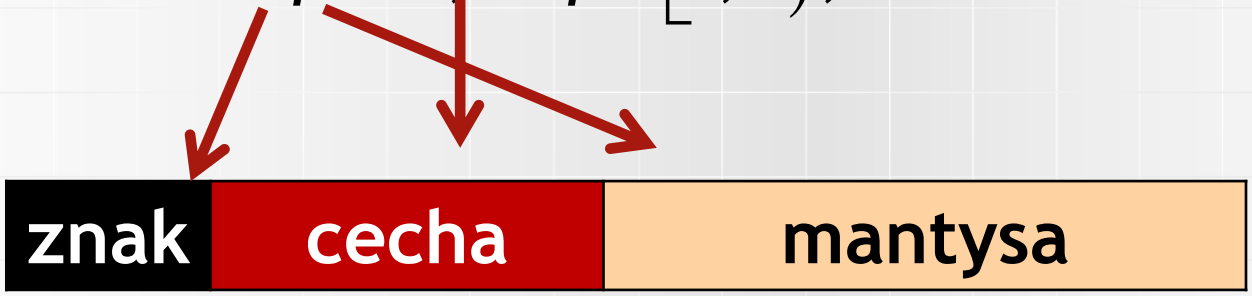
`double`	podwójna precyzja (64 bity)
`single`	pojedyncza precyzja (32 bity)
`int8`	8-bitowa liczba całkowita ze znakiem (signed integer)
`int16`	16-bitowa liczba całkowita ze znakiem
`int32`	32-bitowa liczba całkowita ze znakiem
`int64`	64-bitowa liczba całkowita ze znakiem
`uint8`	8-bitowa liczba całkowita bez znaku (unsigned integer)
`uint16`	16-bitowa liczba całkowita bez znaku
`uint32`	32-bitowa liczba całkowita bez znaku
`uint64`	64-bitowa liczba całkowita bez znaku

Typy danych

Liczby zmiennoprzecinkowe

$$x = \pm r \times 10^n, \quad r \in [1,10), \quad n \in C$$

$$x = \pm q \times 2^n, \quad q \in [1,2), \quad n \in C$$



$$x = (-1)^s \cdot (1, q_1 q_2 q_3 \ldots) \times 2^{n-b}$$

bias
przesunięcie

Politechnika Wrocławska

Typy danych

Liczby zmiennoprzecinkowe

- rozmiar i zachowanie zależy od implementacji
- standard IEEE 754 określa arytmetykę liczb pojedynczej (32 bity) oraz podwójnej (64 bity) precyzji

Arytmetyka komputerowa

Reprezentacja zmiennoprzecinkowa liczb

Liczby pojedynczej precyzji (przesunięcie 127)

z	c	c	c	c	c	c	c	c	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m	m
1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2	3	4	5	6	7	8	9	0	1	2

Liczby podwójnej precyzji (przesunięcie 1023)

z	c	m
1	2-12	13-64

8-bitowa liczba zmiennopozycyjna (przesunięcie 3)

z	c	c	c	m	m	m	m
1	2	3	4	5	6	7	8

Arytmetyka komputerowa

Reprezentacja zmiennoprzecinkowa liczb

0 0 1 1 0 0 0 0	+1. 0 0 0 0	$\times 2^{3-3}$	= 1,
0 0 1 1 1 0 0 0	+1. 1 0 0 0	$\times 2^{3-3}$	= 1,5
0 0 1 1 0 0 0 1	+1. 0 0 0 1	$\times 2^{3-3}$	= 1,0625 $=1+2^{-4}$
0 0 1 0 0 0 0 0	+1. 0 0 0 0	$\times 2^{2-3}$	= 0,5
0 0 0 1 0 0 0 0	+1. 0 0 0 0	$\times 2^{1-3}$	= 0,25 $= 2^{-2}$
0 1 1 0 0 0 0 0	+1. 0 0 0 0	$\times 2^{6-3}$	= 8,0
0 1 1 0 1 1 1 1	+1. 1 1 1 1	$\times 2^{6-3}$	= 15,5 $= 2^3(2-2^{-4})$
1 1 0 1 0 0 1 0	-1. 0 0 1 0	$\times 2^{5-3}$	= -4,5

Arytmetyka komputerowa

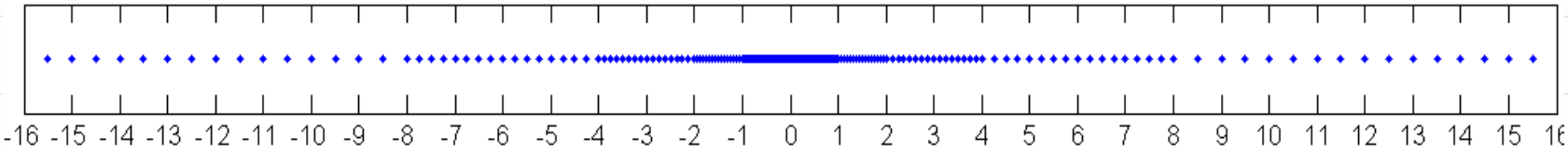
Reprezentacja zmiennoprzecinkowa liczb

0	0	0	0	0	0	0	0	= +0,
1	0	0	0	0	0	0	0	= -0,
0	1	1	1	0	0	0	0	= +Inf (infinity)
1	1	1	1	0	0	0	0	= -Inf (infinity)
0	0	0	0	1	0	0	0	$+0.1000 \times 2^{1-3}$ = 0,125
0	0	0	0	0	0	0	1	$+0.0001 \times 2^{1-3}$ = 0,015625 $=2^{-2}\times 2^{-4}$
0	1	1	1	1	0	0	0	= NaN (not a number)

Arytmetyka komputerowa

Reprezentacja zmiennoprzecinkowa liczb

Rozmieszczenie liczb zmiennopozycyjnych



- precyzja arytmetyki (epsilon maszynowy ε) $0{,}0625 = 2^{-4}$
- największa liczba zmiennopozycyjna $15{,}5 = (2-2^{-4})2^3$
- najmniejsza liczba zmiennopozycyjna
  - znormalizowana $0{,}25 = 2^{-2}$
  - zdenormalizowana $0{,}015625 = 2^{-6}$

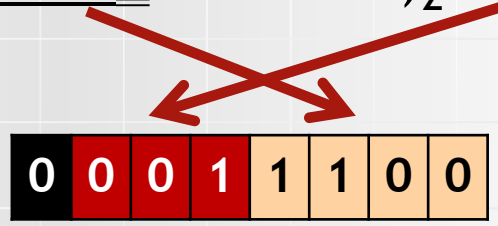
Arytmetyka komputerowa
Reprezentacja zmiennoprzecinkowa liczb

Jak wygląda reprezentacja liczby $^4/_9$ w rozważanej arytmetyce?

$^4/_9 = (0{,}0111000(111000)\ldots)_2$

po normalizacji

$^4/_9 = (1{,}\underline{11000}111000\ldots)_2 \times 2^{-2} = (1{,}\underline{11000}111000\ldots)_2 \times 2^{1-3}$



$fl(^4/_9) = 1{,}1100 \times 2^{-2} = (1+0{,}5+0{,}25)\times 0{,}25 = 0{,}4375$

błąd względny $\left|\delta\right| \le \frac{\varepsilon}{2}$

Arytmetyka komputerowa

Działania na liczbach zmiennoprzecinkowych

Wynikiem operacji matematycznych na liczbach maszynowych zwykle nie jest liczba maszynowa. Przyjmujemy, że po wykonaniu działania mantysa jest normalizowana, a cecha odpowiednio korygowana.

W celu ilustracji rozpatrzymy arytmetykę liczb dziesiętnych z mantysą pięciocyfrową.
Niech $x = 9{,}7541\times10^{2}$, $y = 2{,}7849\times10^{4}$, wtedy

$x + y = 2{,}882441000\times10^{4}$, $fl(x + y) = 2{,}8824\times10^{4}$, $\delta = 1{,}43\times10^{-5}$

$x - y = -2{,}687359000\times10^{4}$, $fl(x - y) = -2{,}6874\times10^{4}$, $\delta = 1{,}53\times10^{-5}$

$x \times y = 2{,}716419309\times10^{7}$, $fl(x \times y) = 2{,}7164\times10^{7}$, $\delta = 7{,}1\times10^{-6}$

$x / y = 3{,}502495601\times10^{-2}$, $fl(x / y) = 3{,}5025\times10^{-2}$, $\delta = 1{,}3\times10^{-6}$

Arytmetyka komputerowa

Działania na liczbach zmiennoprzecinkowych

Przykładem sytuacji, w której mogą pojawić się duże błędy względne jest odejmowanie bliskich sobie liczb

$$
\begin{aligned}
x &= 8{,}14786922317801\,5,\\
fl(x) &= 8{,}14787,\\
y &= 8{,}14723586393179\,0,\\
fl(y) &= 8{,}14724,\\
x - y &= 0{,}00063359246225,\\
fl(x) - fl(y) &= 0{,}00063,\\
fl(fl(x) - fl(y)) &= 6{,}3000\times 10^{-4}
\end{aligned}
$$

$$\left|\frac{(x-y)-fl\left[fl(x)-fl(y)\right]}{x-y}\right| = \left|\frac{0{,}00063359246225-0{,}00063}{0{,}00063359246225}\right| \approx 0{,}0053$$

Arytmetyka komputerowa

Działania na liczbach zmiennoprzecinkowych

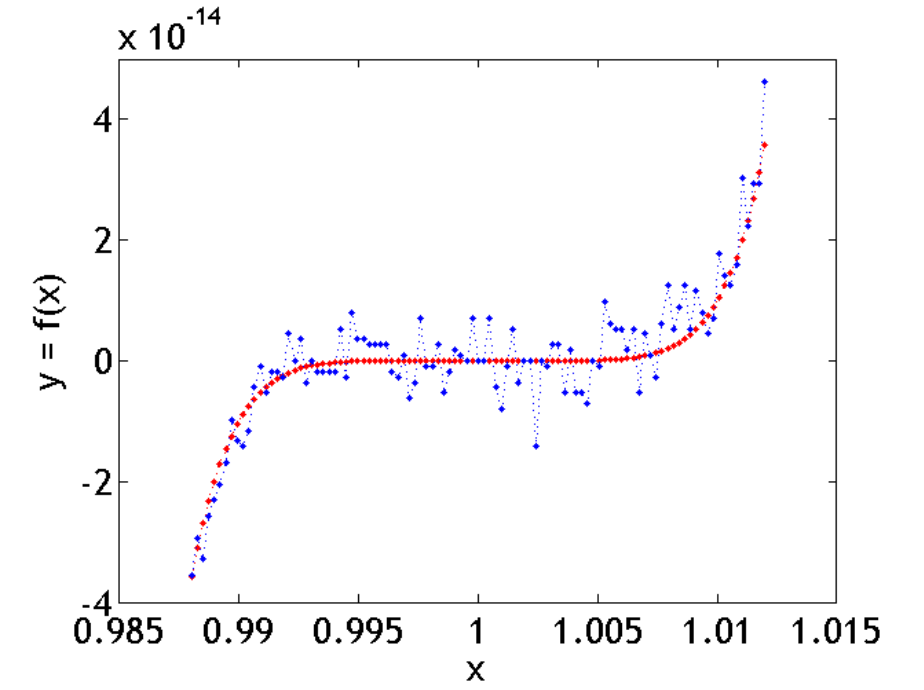
$$y \leftarrow \sqrt{x^2+1}-1$$

$$\sqrt{x^2+1}-1=\left(\sqrt{x^2+1}-1\right)\frac{\sqrt{x^2+1}+1}{\sqrt{x^2+1}+1}=\frac{x^2+1-1}{\sqrt{x^2+1}+1}=\frac{x^2}{\sqrt{x^2+1}+1}$$

$$y \leftarrow \frac{x^2}{\sqrt{x^2+1}+1}$$

Arytmetyka komputerowa

Działania na liczbach zmiennoprzecinkowych



$$f(x) = (x-1)^7 = x^7-7x^6+21x^5-35x^4+35x^3-21x^2+7x-1$$

Politechnika Wrocławska

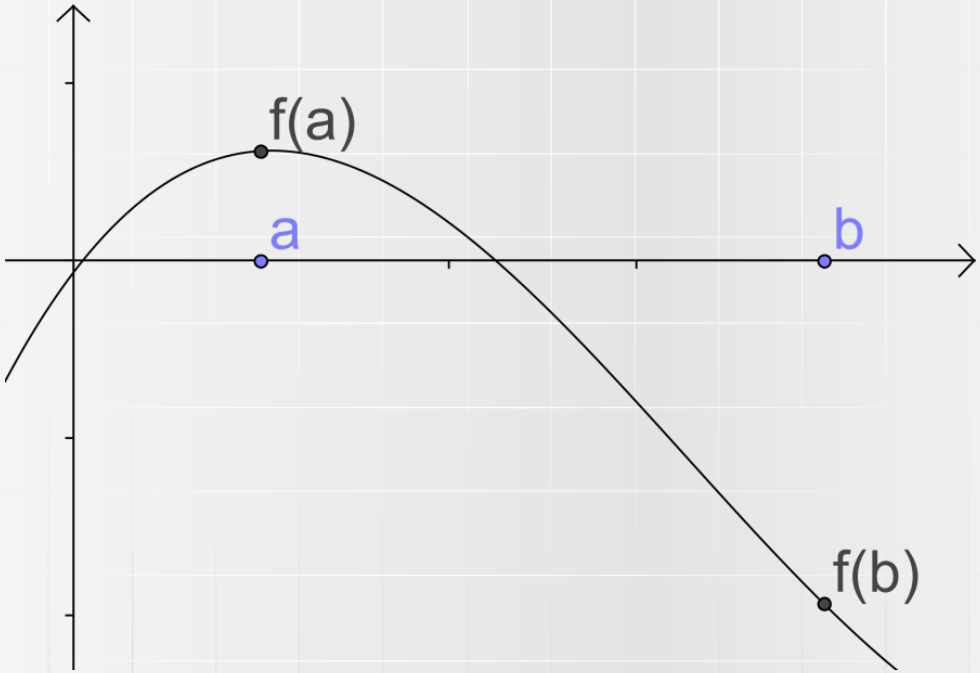
Podsumowanie (1)

- Zapis liczb w różnych układach
- Typy danych numerycznych
- Reprezentacja zmiennoprzecinkowa liczb
- Dokładność operacji na liczbach zmiennoprzecinkowych

Politechnika Wrocławska

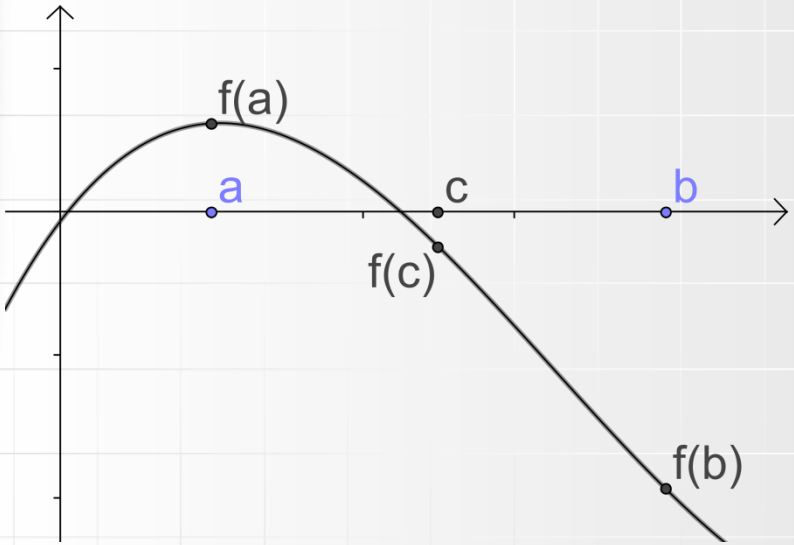
Metoda bisekcji

Jeśli f jest funkcją ciągła w przedziale $[a,b]$ i jeśli $f(a)f(b)<0$, to funkcja ta musi mieć zero w (a,b).

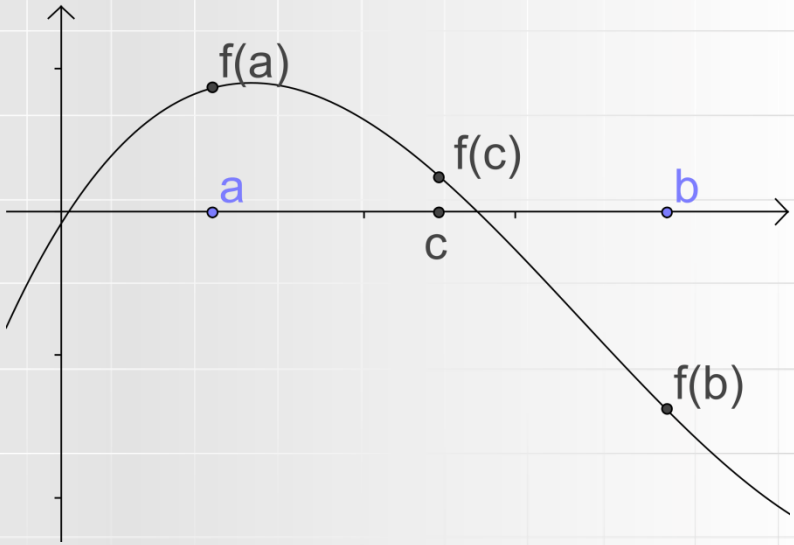


Metoda bisekcji

Wyznaczamy punkt $c = \frac{1}{2}(a+b)$ oraz wartość funkcji $f(c)$

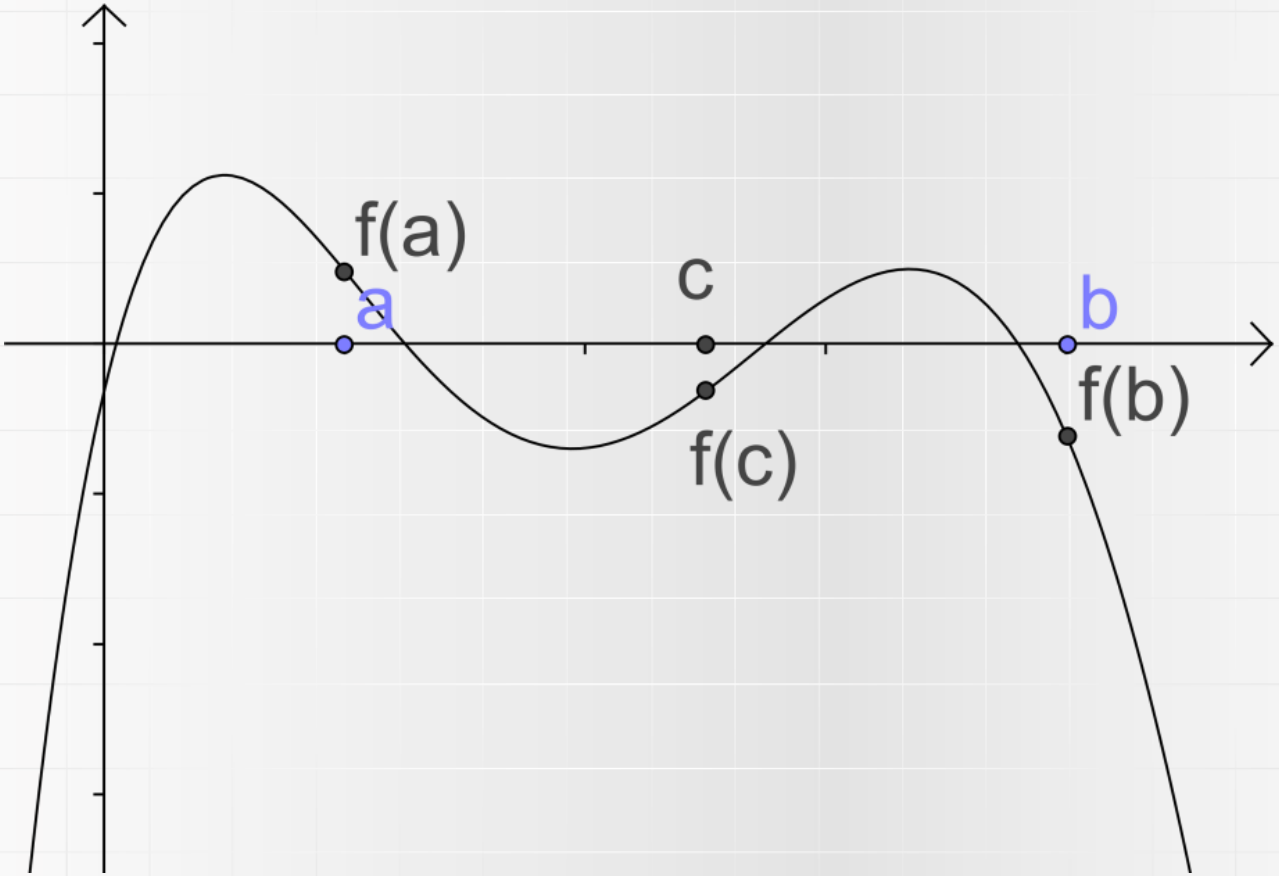


jeśli $f(a)f(c)<0$ to $b = c$



jeśli $f(b)f(c)<0$ to $a = c$

Metoda bisekcji



jeśli $f(a)f(c)<0$ to $b = c$

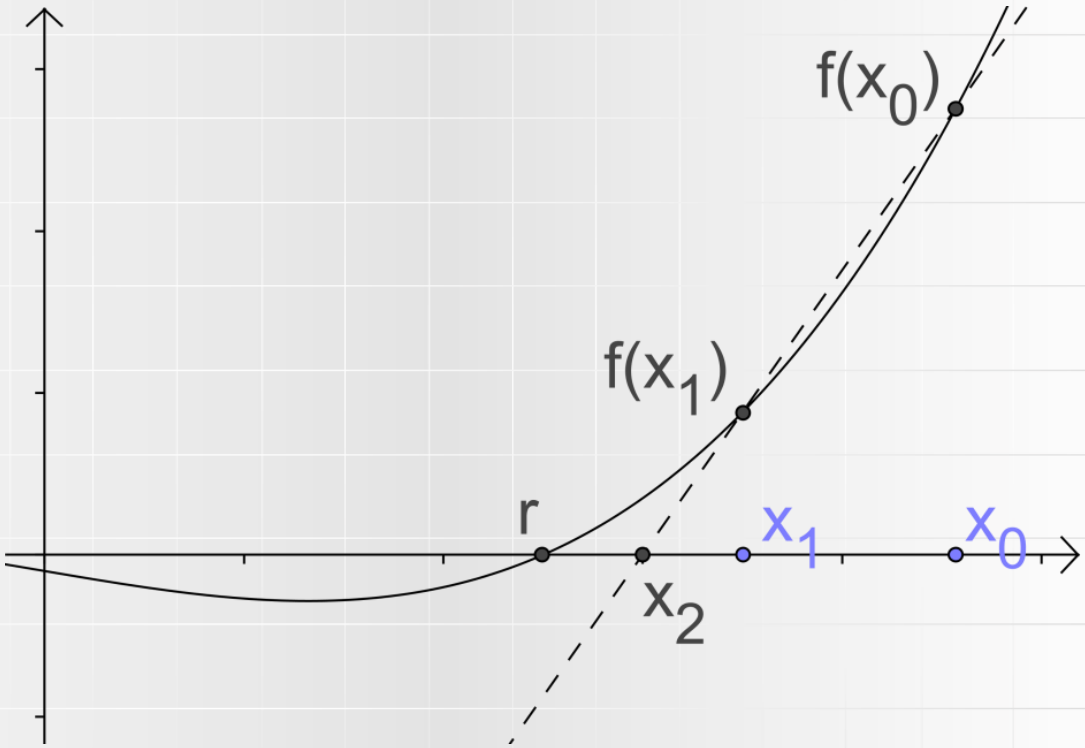
Politechnika Wrocławska

Metoda bisekcji

Kryteria zakończenia:

- przekroczenie maksymalnej liczby kroków,
- zadowalająco mały błąd,
- zadowalająco mała wartość funkcji.

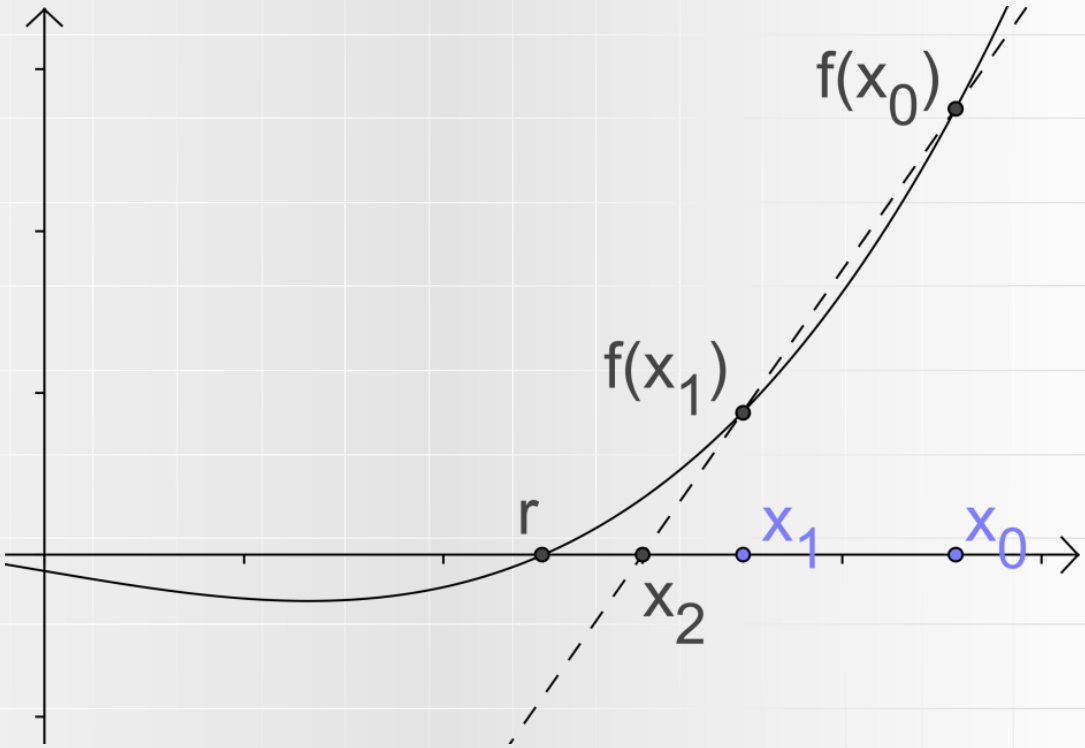
Metoda siecznych



$$x_{n+1} = x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}$$

Metoda Newtona

$$x_{n+1} = x_n - \frac{f(x_n)}{f'(x_n)}$$



$$x_{n+1} = x_n - f(x_n)\frac{x_n - x_{n-1}}{f(x_n) - f(x_{n-1})}$$

Zbieżność

W wielu przypadkach program komputerowy generuje ciąg przybliżeń rozwiązania. Zbieżność określa jak szybko uzyskamy dostatecznie dobre przybliżenie.

$$x_n = \left(1 + \frac{1}{n}\right)^n$$

$$\lim_{n\to\infty} x_n = e \approx 2{,}718281828$$

$$x_1 = 2{,}00000\ 0$$

$$x_2 = 2{,}25000\ 0$$

$$x_5 = 2{,}48832\ 0$$

$$x_{10} = 2{,}59374\ 2$$

$$x_{100} = 2{,}70481\ 4$$

$$x_{1000} = 2{,}71692\ 4$$

Zbieżność

$$x_{n+1} = \frac{x_{n-1}x_n + 1}{x_{n-1} + x_n}, x_0 = 0, x_1 = 2$$

$x_0 = 0$

$x_1 = 2$

$x_2 = 0{,}5$

$x_3 = 0{,}8$

$x_4 = 1{,}076923077$

$x_5 = 0{,}991803278$

$x_6 = 0{,}999695214873514$

$$x_{n+1} = \frac{1}{2}x_n + \frac{1}{x_n}, x_1 = 2$$

$x_1 = 2$

$x_2 = 1{,}5$

$x_3 = 1{,}416667$

$x_4 = 1{,}414216$

$\sqrt{2} = 1{,}414213$

Rząd zbieżności

Niech $\{x_n\}$ będzie ciągiem zbieżnym do x^*.

Zbieżność jest co najmniej *liniowa*, jeśli istnieją stała $c < 1$ i liczba całkowita N takie, że

$$|x_{n+1} - x^*| \leq c|x_n - x^*| \ \ (n \geq N)$$

Zbieżność jest co najmniej *nadliniowa*, jeśli istnieje ciąg zbieżny do 0 $\{\varepsilon_n\}$ i liczba całkowita N takie, że

$$|x_{n+1} - x^*| \leq \varepsilon_n|x_n - x^*| \ \ (n \geq N)$$

Politechnika Wrocławska

Rząd zbieżności

Niech $\{x_n\}$ będzie ciągiem zbieżnym do x^*.

Zbieżność jest co najmniej *kwadratowa*, jeśli istnieją stała dodatnia *C* i liczba całkowita *N* takie, że

$$|x_{n+1} - x^*| \leq C|x_n - x^*|^2 \quad (n \geq N)$$

Zbieżność jest co najmniej *rzędu α,* jeśli istnieją stała dodatnia C, stała α>1 i liczba całkowita *N* takie, że

$$|x_{n+1} - x^*| \leq C|x_n - x^*|^\alpha \quad (n \geq N)$$

Notacja *O* i *o*

Niech $\{x_n\}$ i $\{\alpha_n\}$ będą dwoma różnymi ciągami.

$$x_n = O(\alpha_n)$$

jeśli istnieją takie stałe *C* i n_0, że $|x_n| \leq C|\alpha_n|$ dla każdego $n \geq n_0$.

$$x_n = o(\alpha_n)$$

jeśli $\lim_{n\to\infty}\left(\frac{x_n}{\alpha_n}\right) = 0$. Istnieje ciąg liczb nieujemnych zbieżny do 0 taki, że $|x_n| \leq \varepsilon_n|\alpha_n|$.

Notacja *O* i *o*

Jeśli $x_n \to 0, \alpha_n \to 0$ oraz $x_n = O(\alpha_n)$ to ciąg $\{x_n\}$ dąży do 0 *co najmniej tak szybko* jak $\{\alpha_n\}$.

Jeśli $x_n \to 0, \alpha_n \to 0$ oraz $x_n = o(\alpha_n)$ to ciąg $\{x_n\}$ dąży do 0 szybciej niż $\{\alpha_n\}$.

$$\frac{n+1}{n^2} = O\left(\frac{1}{n}\right) \qquad \frac{5}{n} + e^{-n} = O\left(\frac{1}{n}\right)$$

$$\frac{1}{n \ln n} = o\left(\frac{1}{n}\right) \qquad \frac{1}{n} = o\left(\frac{1}{\ln n}\right) \qquad e^{-n} = o\left(\frac{1}{n^2}\right)$$

Notacja *O* i *o*

Notacji tej używa się nie tylko dla ciągów.

$$\sin x = x - \frac{x^3}{6} + O\left(x^5\right) \quad (x \to 0)$$

Istnieją otoczenie punktu 0 i stała C takie, że w tym otoczeniu

$$\left|\sin x - x + \frac{x^3}{6}\right| \leq C\left|x^5\right|.$$

$$f(x) = O\big(g(x)\big) \quad (x \to \infty).$$

Istnieją takie stałe r i C, że $|f(x)| \leq C|g(x)|$ dla każdego $x \geq r$.

$$\sqrt{x^2 + 1} = O(x) \quad (x \to \infty)$$

Politechnika Wrocławska

Notacja *O* i *o*

$$f(x) = O\big(g(x)\big) \ \ (x \to x^*)$$

jeśli istnieją takie stałe C i otoczenie punktu x^* takie, że w tym otoczeniu

$$|f(x)| \leq C|g(x)|.$$

Podobnie

$$f(x) = o\big(g(x)\big) \ \ (x \to x^*)$$

jeśli $\lim_{x \to x^*} \left(\frac{f(x)}{g(x)} \right) = 0\,.$

Politechnika Wrocławska

Podsumowanie (1)

- Zapis liczb w różnych układach
- Typy danych numerycznych
- Reprezentacja zmiennoprzecinkowa liczb
- Dokładność operacji na liczbach zmiennoprzecinkowych

Politechnika Wrocławska

Podsumowanie (2)

- Metody wyznaczania miejsc zerowych
  - Metoda bisekcji
  - Metoda Newtona i metoda siecznych
- Zbieżność
- Notacja *O* i *o*