Obliczenia Naukowe
Ćwiczenia

Jakub Kowal

1 Lista 1

1.1

$a_1 = 0,25$ $\tilde{a_1} = 0,25 * 10^0$
$a_2 = 0,0046$ $\tilde{a_2} = 0,46 * 10^{-2}$
$a_3 = 0,00079$ $\tilde{a_3} = 0,79 * 10^{-3}$
$a_4 = 0,061$ $\tilde{a_4} = 0,61 * 10^{-1}$
$((\tilde{a_1} \oplus \tilde{a_2}) \oplus \tilde{a_3}) \oplus \tilde{a_4}$
$\tilde{a_1} \oplus \tilde{a_2} = 0,25*10^0 \oplus 0,46*10^{-2} = 0,25*10^0 \oplus 0,0046*10^0 = rd(0,2546*10^0) = 0,25*10^0 = \tilde{a_{12}}$
$\tilde{a_{12}} \oplus \tilde{a_3} = 0,25 * 10^0 \oplus 0,79 * 10^{-3} = rd(0,25079 * 10^0) = 0,25 * 10^0 = \tilde{a_{123}}$
$\tilde{a_{123}} \oplus \tilde{a_4} = 0,25 * 10^0 \oplus 0,61 * 10^{-1} = rd(0,311 * 10^0) = 0,31 * 10^0$
$((\tilde{a_3} \oplus \tilde{a_2}) \oplus \tilde{a_4}) \oplus \tilde{a_1}$
$\tilde{a_3} \oplus \tilde{a_2} = 0,54 * 10^{-2}$
$\tilde{a_{32}} \oplus \tilde{a_4} = 0,66 * 10^{-1}$
$\tilde{a_{432}} \oplus \tilde{a_1} = 0,32 * 10^0$
Wynik rzeczywisty = 0,31639
$\delta = \frac{|0,31639-0,31|}{|0,31369|} = 0,0201966$
$\delta = \frac{|0,31639-0,32|}{|0,31369|} = 0,01141$

1.2

1.2.a)

$x = 0,54617$ $\tilde{x} = 0,5462$
$y = 0,54601$ $\tilde{y} = 0,5460$
$r = x - y = 0,00016$ $\tilde{r} = \tilde{x} \ominus \tilde{y} = rd(0,00016) = 0,2 * 20^{-3}$
$\epsilon = 0,5\beta^{1-t} = 0,5 * 10^{-3}$

1.2.b)

$\beta = 10$ $t = 3$
$a = 1,22$

$b = 3,34$
$c = 2,28$
$\tilde{\Delta} = \tilde{b} \odot \tilde{b} \ominus 4 \odot \tilde{a} \odot \tilde{c} = rd(3,34 * 3,34) \ominus \cdots = 11,16 \ominus 4 \odot \tilde{a} \odot \tilde{c} = 0,4 * 10^1 \odot$
$0,122 * 10^1 = (0,488 * 10^{-1})$
$(0,488 * 10^1) \odot (0,228 * 10^1) = (0,111 * 10^2)$
$\tilde{\Delta} = (0,112 * 10^[2]) \ominus (0,111 * 10^2) = (0,001 * 10^2) = 0,1$
$\Delta = 0,0292$
$\frac{|\Delta - \tilde{\Delta}|}{|\Delta|} = \frac{|0,0292 - 0,1|}{|\Delta|} = 2,42$
$\epsilon = \frac{1}{2}\beta^{1-t} = \frac{1}{2} * 10^{-2} = 0,005$

1.3

$(0,78125)_{10} = (0,11001)_2$
$z = \pm m_z 2^c m_z \in [\frac{1}{2}, 1)$
$0,11001 * 2^0$
$0,101111001 * 2^{10}$
$(754)_{10} = (1011110010)_2$

1.4

1.4.a)

$t = 23$
$n \in [1,2)$
$x = 2^{-1} + 2^{-26}$
$2^{-1} = (\frac{1}{2})_{10} = (0,1)_2$
$2^{-26} = (\frac{1}{2^{26}})_{10} = (0,\underbrace{0\ldots0}_{25}1)_2$
$x = 0,1\underbrace{0\ldots0}_{24}1$
$\tilde{x} = 1,\underbrace{0\ldots0}_{24}1 * 2^{-1}$ Odp. Nie

1.4.b)

$y = \frac{1}{3} = 0,(3)$
$\tilde{y} = 1,(01)_2 * 2^{-2}$
$0,(3) = 0,(01)_2$
$0,(3) * 2$
$0,(6) * 2$
$1,3 * 1$ Odp. Nie

1.4.c)

$z = \frac{1}{5} = 0,2$
$\tilde{z} = 1,(0011) * 2^{-3}$ Odp. Nie

1.4.d)

$j = \frac{1}{10} = (0,1)_{10}$
$0,1 * 2$
$\tilde{j} = 1,(00011)_2 * 2^{-4}$

1.5

33-bitowe słowo $x = sm2^c$
cecha 8 bitów ze znakiem
mantysa 24 bity $\in [\frac{1}{2}, 1)$

1.5.a)

$x_{max} = (0.11...1)_2 * 2^{127} = (1 - 2^{-24}) * 2^{127} = 1,7 * 10^{38}$
$x_{min} = (0.10...0)_2 * 2^{-127} = 2^{-1} * 2^{-127} = 2^{-128}$
$[-x_{max}, -x_{min}] \cup [x_{min}, x_{max}]$

1.5.b)

$(-x_{min}, x_{min})$

1.5.c)

$\epsilon = \frac{1}{2} * \beta^{1-t} = 2^{-1} * 2^{-23} = 2^{-24}$

Single

1.5.a)

$x_{max} = (2 * 2^{-23}) * 2^{127} \approx 3,4 * 10^{38}$
$x_{min} = 1 * 2{-}126 = 2^{-126}$
$x_{minsub} = (0.00.....01)_2 * 2^{-126} = 2^{-23} * 2^{-126} = 2^{-149}$
$[-x_{max}, -x_{min}] \cup \{0\} \cup [x_{min}, x_{max}]$

1.5.b)

$(-x_{minsub}, x_{minsub})$

1.5.c)

$\epsilon = \frac{1}{2}\beta^{1-t} = 2^{-1} * 2^{-23} = 2^{-24}$

1.6

1.6.a)

x $= 1 + 2^{-24}$
$\epsilon = 2^{-23}$
$x^{-} = 1$
$x^{+} = 1 + 2^{-23}$

1.6.b)

$x \bigoplus 1 = 1$
$x \in (0, 2^{-24} + 2^{-25}]$

1.6.c)

$x \bigoplus = x$
$x < \frac{x}{2^{23}}$

1.7

$A(a_1, ..., a_n) = (...(a_1 \oplus a_2) \oplus a_3) \oplus ...) \oplus a_n = (...(a_1 + a_2)(1+\delta_1) + a_3)(1+\delta_2) + ...) + a_n)(1+\delta_{n-1}) = a_1 + a_1 * E_1 + a_2 + a_2 * E_1 + a_3 + a_3 * E_2 ... + a_n + a_n * E_{n-1} = (a_1 + a_2 + ... + a_n) + (a_1 * E_1 + a_2 * E_1 + ... + a_n * E_{n-1}) = S + (a_1 * E_1 + a_2 * E_1 + ... + a_n * E_{n-1})$
$|\delta| \leq \epsilon$
$1 + E_k = \prod_{i=1}^{k}(1 + \delta_k) \leftarrow$ *wprowadzamy*
$(a_1 * E_1 + a_2 * E_1 + ... + a_n * E_{n-1}) = E_{max} = \prod_{i=1}^{n-1}(1 + |\delta_i|) - 1 \leq \prod_{i=1}^{n-1}(1+\epsilon) - 1$
$a_1 * E_1 + a_2 * E_1 + ... + a_n * E_{n-1} \leq (a_1 + a_2 + ... + a_n) * E_{max} \leq ((a_1 + a_2 + ... + a_n) * \prod_{i=1}^{n-1}(1 + \epsilon)) - 1 = S(1 + \epsilon)^{n-1} - 1$
$\frac{|\tilde{S} - S|}{|S|} \leq \frac{\cancel{S} + S * (1+s)}{|S|} \leftarrow$ Zmazał za szybko
$\tilde{S} \leq S + S * (1 + \epsilon)^{n-1} - 1$

1.8

P(n-1)=$\prod_{i=0}^{n-1} q_i$ — dokładny wynik
Q(n-1)=fl(P(n-1)) —wynik w fl
Q(n-1)=$P(n-1)(1 + \sum_{i=1}^{n-1} q_i), \forall i |\delta_i| \leq \epsilon$ n=1 Q(1)=$a_0 \odot a_1 = a_0 a_1 (1 + \delta_1) = P(1)(1 + \sum_{i=1}^{1} q_i)$
Zachodzi dla 1
Q(n+1)=$a_0 \odot a_1 \odot \cdots \odot a_{n+1}$=Q(n)$\odot a_{n+1} = (Q(n) a_{n+1})(1 + \delta_n + 1) = P(n)(1 + \sum_{i=1}^{n}) a_{n+1} (1 + \delta_{n+1}) = P(n+1)(1 + \sum_{i=1}^{n} \delta_1)(1 + \delta_{n+1}) \approx P(n+1)(1 + \delta_{n+1} + \sum_{i=1}^{n} \delta_i) = P(n+1)(1 + \sum_{i=1}^{n+1} \delta_i)$

|P(n)-Q(n)|=$|P(n)\sum_{i=1}^{n}\delta_i|$

1.9

$A1(a,b) = a^2 - b^2$
$fl(a^2 - b^2) = fl(fl(a^2) - fl(b^2)) = (a^2(1+\delta_1) - b^2(1+\delta_2))(1+\delta_3) \approx a^2 - b^2 + (a^2 - b^2)\delta_3 + a^2\delta_1 - b^2\delta_2$
Błąd względny:
$\frac{|a^2-b^2+(a^2-b^2)\delta_3+a^2\delta_1-b^2\delta_2-(a^2-b^2)|}{|a^2-b^2|} = \frac{|(a^2-b^2)\delta_3+a^2\delta_1-b^2\delta_2|}{|a^2-b^2|} \leq |\delta_3| + \frac{|a^2\delta_1-b^2\delta_2|}{|a^2-b^2|} \leq |\delta_3| + \frac{a^2|\delta_1|-b^2|\delta_2|}{|a^2-b^2|} \leq \epsilon + \frac{a^2\epsilon+b^2\epsilon}{|a^2-b^2|} = \epsilon(1 + \frac{a^2+b^2}{|a^2-b^2|})$
$A2(a,b) = (a-b)(a+b)$
$((a-b)(1+\delta_1)(a+b)(1+\delta_2))(1+\delta_3) = (a^2-b^2)(1+\delta_1)(1+\delta_2)(1+\delta_3) \approx (a^2-b^2)(1+\delta_1+\delta_2+\delta_3)$
Błąd względny:
$\frac{|(a^2-b^2)(1+\delta_1+\delta_2+\delta_3)-(a^2-b^2)|}{|a^2-b^2|} = |\delta_1+\delta_2+\delta_3| \leq |\delta_1| + |\delta_2| + |\delta_3| \leq 3\epsilon$

2 Lista 2

2.1

$y = \sqrt{x^2+1} - 1$
TW z wykładu:
Jeśli x,y - dodatnie liczby w dwójkowej arytmetyce float, takie że:
$x > y, 2^{-q} \leq 1 - \frac{y}{x}$
to przy odejmowaniu tracimy najwyżej q bitów q = 2
$2^{-2} \leq 1 - \frac{1}{\sqrt{x^2+1}}$
$\frac{1}{4} = 1 - \frac{1}{\sqrt{x^2+1}}$
$\frac{1}{\sqrt{x^2+1}} \leq \frac{3}{4}$
$\sqrt{x^2+1} >= \frac{3}{4}$

2.2

Wskazówka:
Twierdzenie z 1 zadania: $2^{-q} <= 1 - \frac{a}{b} <= 2^{-P} \wedge b - a, b > a$
$x = \frac{1}{2}$
$\cos\frac{1}{2} = 0,8775825$
$2^{-q} <= 1 - \frac{0,8775825}{1} <= 2^{-P}$
$2^{-q} <= 0,1224175 <= 2^{-P}$
$q = 4 \wedge P = 3$

2.3

f(x)=$x^{-1}(1-\cos x)$
a) $\lim_{x\to 0}\frac{1-\cos x}{x}(De'Hospital)=\lim_{x\to 0}\frac{\sin x}{1}=0$
b) $\cos x\approx 1$
$x\approx 2\pi k, k\in Z$
$1-\cos x=2\sin^2\frac{x}{2}$
f(x)=$\frac{2\sin^2\frac{x}{2}}{x}=(t=\frac{x}{2})=\frac{\cancel{2}\sin^2 t}{\cancel{2}t}=t(\frac{\sin t}{t})^2=\frac{x}{2}(\frac{\sin\frac{x}{2}}{\frac{x}{2}})^2$
$\lim_{x\to 0}\frac{\sin t}{t}=1$ dla $x\to 0(t\to 0)$ mamy $f(t)=t1^2=t=\frac{x}{2}$

2.4

$f(x)=\sqrt{x+2}-\sqrt{x}$
Problem dla dużych x : $\sqrt{x}\approx\sqrt{x+2}$
$\sqrt{x+2}-\sqrt{x}/*\frac{\sqrt{x+2}+\sqrt{x}}{\sqrt{x+2}+\sqrt{x}}$
$\frac{x+2-x}{\sqrt{x+2}+\sqrt{x}}=\frac{2}{\sqrt{x+2}+\sqrt{x}}$
$a^2-b^2=(a-b)(a+b)$

2.5

$u=\pm\sqrt{\frac{x+\sqrt{x^2+y^2}}{2}}$
$v=\frac{y}{2u}$
$liczbazespolona=x+iy$
$x\neq 0$
$pierwiastektou+iv$
$x\geq 0:$ jeśli y«x, to może dojść do "pochłonięcia" y
poza tym dokładność generalnie dobra
$x<0:$ może się zdarzyć, że $|x|\approx\sqrt{x^2+y^2}$
i nastąpi utrata cyfr znaczących w liczniku, więc dokładność wzoru jest słaba
alternatywny wzór na $k<0:$
$x+\sqrt{x^2+y^2}=\frac{(x+\sqrt{x^2+y^2})(x-\sqrt{x^2+y^2})}{(x-\sqrt{x^2+y^2})}=\frac{x^2-x^2-y^2}{x-\sqrt{x^2+y^2}}=\frac{-y^2}{x-\sqrt{x^2+y^2}}=\frac{-y^2}{-(|x|+\sqrt{x^2+y^2})}$
Wtedy $u=\sqrt{\frac{-y^2}{2(x-\sqrt{x^2+y^2})}}=\sqrt{\frac{y^2}{2(|x|+\sqrt{x^2+y^2})}}$

2.6

$4ac\to 0$
x=$\frac{-b\mp\sqrt{b^2-4ac}}{2a}$
$x_1x_2=\frac{c}{a}\ q=\frac{1}{2a}(-b-sqm?(b)\sqrt{b^2-4ac})$
$x_1=\frac{-b+\sqrt{b^2-4ac}}{2a}$
$x_2=\frac{-b-\sqrt{b^2-4ac}}{2a}=\frac{-(b+\sqrt{b^2-4ac})}{2a}$

$x_1 = \frac{c}{ax_2}$

$$b = 0ax^2 + c = 0$$
$$\Delta = -4ac$$

$b < 0$
$x_1 = \frac{|b| + \sqrt{b^2 - 4ac}}{2a}$
$x_2 = \frac{c}{2a}$

2.8

$\frac{|f(\hat{x}) - f(x)|}{|f(x)|} = \frac{|f(x + x\delta) - f(x)|}{|f(x)|} \approx \frac{|f'(x)x\delta|}{|f(x)|} = cond(x)|\delta|$
$x^\alpha (x^\alpha)' = \alpha x^{\alpha-1} \frac{|(\alpha x^{alpha-1})x\delta|}{|x^\alpha|} = \frac{|\alpha x^\alpha \delta|}{|x^\alpha|} = \alpha\delta$
$\sin x (\sin x)' = \cos x \frac{|(\cos x)x\delta|}{|\sin x|} = \frac{x\delta}{\tan x}$
$e^x (e^x)' = e^x$
$x^{-1}e^x (x^{-1}e^x)' = x^{-1}e^x + (-1)x^{-2}e^x = \frac{e^x}{x} - \frac{e^x}{x^2}$
$\arcsin x$

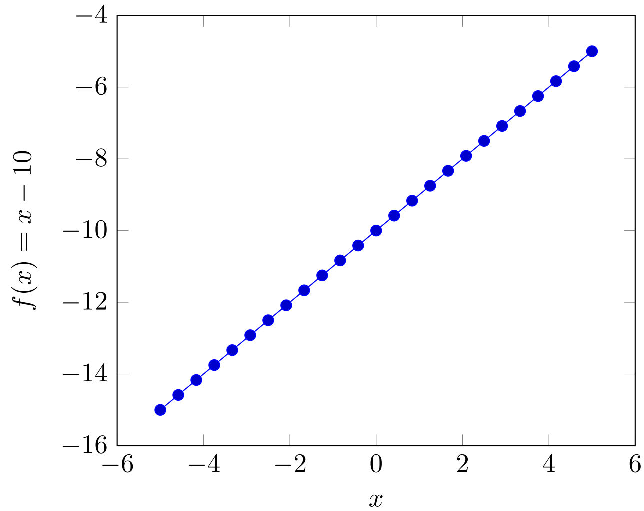
3 Lista 3

3.1

Bisekcji $[a_0, b_0]$ $c_n = \frac{a_n + b_n}{2}$
$\lim_{n\to\infty} c_n = r$
$e_n = r - c_n$

3.1.a)

$f(x) = x - 10$
Czy relacja $|e_0| \geq |e_1| \geq \dots$
$r = 10$
$f(10) = 0$
$a_0 = 0$
$b_0 = 22$
$c_0 = 11$
$e_0 = -1$
$|e_0| = 1$
$a_1 = 0$
$b_1 = 11$
$c_1 = 5.5$
$e_1 = 4.5$
$|e_0| < |e_1|$



3.1.b)