Metody Numeryczne

Wykład 1
Literatura
Dokładność

Wojciech Kordecki

Collegium Witelona
Wydział Nauk Technicznych i Ekonomicznych
Zakład Informatyki

Semestr letni 2023/24

Literatura podstawowa

[KS] W. Kordecki, K. Selwat.
Metody numeryczne dla informatyków.
Helion, Gliwice, 2020.

[FMW] Z. Fortuna, B. Macukow, J. Wąsowski.
Metody numeryczne.
Wydawnictwo Naukowe PWN, Warszawa, wydanie 7, 2017.

[PLSG] B. Pańczyk, E. Łukasik, J. Sikora, T. Guziak.
Metody numeryczne w przykładach.
Politechnika Lubelska, 2012.
http://bc.pollub.pl/Content/1370/metody.pdf.

Literatura uzupełniająca

[KCh] D. Kincaid, W. Cheney.
Analiza numeryczna.
Wydawnictwa Naukowo-Techniczne, 2006.

[SzWM] R. Szmurło, S. Wincenciak, T. Markiewicz.
Metody numeryczne. Wykłady na Wydziale Elektrycznym Politechniki Warszawskiej.
Oficyna Wydawnicza Politechniki Warszawskiej, 2015.

Literatura – programowanie

[BP] B. Baron, Ł. Piątek.
Metody numeryczne w C++ Builder.
Helion, 2004.

[MW] P. Mikołajczak, M. Ważny,
Metody numeryczne w C++,
UMCS w Lublinie, Lublin 2012.
http://informatyka.umcs.lublin.pl/files/mikolajczak_metody_numeryczne_w_cpp.pdf

Literatura – klasyka

[K] D. Knuth,
Sztuka programowania, t. 2, Algorytmy seminumeryczne.
Wydawnictwa Naukowo-Techniczne, 2002.

Źródła błędów

Źródła błędów

- Błędy danych wejściowych – dane zaokrąglone z wcześniejszych obliczeń, dane z pomiarów itp.

Źródła błędów

- Błędy danych wejściowych – dane zaokrąglone z wcześniejszych obliczeń, dane z pomiarów itp.
- Błędy zaokrągleń w czasie pomiarów.

Źródła błędów

- Błędy danych wejściowych – dane zaokrąglone z wcześniejszych obliczeń, dane z pomiarów itp.
- Błędy zaokrągleń w czasie pomiarów.
- Błędy obcięcia, na przykład obliczenie zbyt małej liczby wyrazów,

Źródła błędów

- Błędy danych wejściowych – dane zaokrąglone z wcześniejszych obliczeń, dane z pomiarów itp.
- Błędy zaokrągleń w czasie pomiarów.
- Błędy obcięcia, na przykład obliczenie zbyt małej liczby wyrazów,
- Zbytnie uproszczenie modelu matematycznego.

Źródła błędów

- Błędy danych wejściowych – dane zaokrąglone z wcześniejszych obliczeń, dane z pomiarów itp.
- Błędy zaokrągleń w czasie pomiarów.
- Błędy obcięcia, na przykład obliczenie zbyt małej liczby wyrazów,
- Zbytnie uproszczenie modelu matematycznego.
- Błędy programisty.

Błędy bezwzględne i względne

Niech

- wartość prawdziwa x,
- wartość przybliżona $\tilde{x}$ wartości x.

Wtedy

Błędy bezwzględne i względne

Niech

- wartość prawdziwa x,
- wartość przybliżona $\tilde{x}$ wartości x.

Wtedy

- $|\tilde{x} - x|$ – błąd bezwzględny ,

Błędy bezwzględne i względne

Niech

- wartość prawdziwa x,
- wartość przybliżona $\tilde{x}$ wartości x.

Wtedy

- $|\tilde{x} - x|$ – błąd bezwzględny ,
- $|(\tilde{x} - x)/x|$ – błąd względny.

Błąd sumy i różnicy

Niech $x = x_1 + x_2$ lub $x = x_1 - x_2$

$$\begin{aligned} d_1 &= |\tilde{x}_1 - x_1| \\ d_2 &= |\tilde{x}_2 - x_2|, \\ d &= |\tilde{x} - x|. \end{aligned}$$

Błąd sumy i różnicy

Niech $x = x_1 + x_2$ lub $x = x_1 - x_2$

$$d_1 = |\tilde{x}_1 - x_1|$$
$$d_2 = |\tilde{x}_2 - x_2|,$$
$$d = |\tilde{x} - x|.$$

Stąd

$$d \leq d_1 + d_2 \,,$$

czyli błąd bezwzględny sumy i różnicy jest niewiększy od sumy błędów bezwzględnych.

Błąd iloczynu

Niech $x = x_1 x_2$.

$$\begin{aligned} r_1 &= \left|\left(\tilde{x}_1 - x_1\right)/x_1\right| \\ r_2 &= \left|\left(\tilde{x}_2 - x_2\right)/x_2\right|, \\ r &= \left|\left(\tilde{x} - x\right)/x\right|. \end{aligned}$$

Błąd iloczynu

Niech $x = x_1 x_2$.

$$\begin{aligned} r_1 &= \left|\left(\tilde{x}_1 - x_1\right)/x_1\right| \\ r_2 &= \left|\left(\tilde{x}_2 - x_2\right)/x_2\right|, \\ r &= \left|\left(\tilde{x} - x\right)/x\right|. \end{aligned}$$

Stąd

$$r \approx r_1 + r_2\,,$$

o ile $r_1 \ll 1$, $r_2 \ll 1$, czyli błąd względny iloczynu jest w przybliżeniu równy sumie błędów względnych.

Błąd ilorazu

Niech $x = x_1/x_2$.

$$r_1 = \left|\left(\tilde{x}_1 - x_1\right)/x_1\right|$$
$$r_2 = \left|\left(\tilde{x}_2 - x_2\right)/x_2\right|,$$
$$r = \left|\left(\tilde{x} - x\right)/x\right|.$$

Błąd ilorazu

Niech $x = x_1/x_2$.

$$
\begin{aligned}
r_1 &= \left| \left(\tilde{x}_1 - x_1\right) / x_1 \right| \\
r_2 &= \left| \left(\tilde{x}_2 - x_2\right) / x_2 \right|, \\
r &= \left| \left(\tilde{x} - x\right) / x \right|.
\end{aligned}
$$

Stąd

$$
r \approx \left| r_1 - r_2 \right|,
$$

o ile $r_2 \ll 1$, czyli błąd względny ilorazu jest w przybliżeniu równy wartości bezwzględnej różnicy błędów względnych.

Liczba π to trzy z hakiem

Liczba π to trzy z hakiem

$$\pi \approx 3.141592653589793$$

Liczba π to trzy z hakiem

$$\pi \approx 3.141592653589793$$
$$\pi \approx 3.14$$

Liczba π to trzy z hakiem

$$\pi \approx 3.141592653589793$$
$$\pi \approx 3.14$$

Z dokładnością do 16 cyfr znaczących:

$$\sin\left(3.141592653589793/2\right) = 1$$

Liczba π to trzy z hakiem

$$\pi \approx 3.141592653589793$$
$$\pi \approx 3.14$$

Z dokładnością do 16 cyfr znaczących:

$$\sin\left(3.141592653589793/2\right) = 1$$
$$\sin\left(3.14/2\right) = 0.9999996829318346$$

Liczba π to trzy z hakiem

$$\pi \approx 3.141592653589793$$
$$\pi \approx 3.14$$

Z dokładnością do 16 cyfr znaczących:

$$\sin(3.141592653589793/2) = 1$$
$$\sin(3.14/2) = 0.9999996829318346$$

$$e \approx 2.718281828459045$$

Liczba π to trzy z hakiem

$$\pi \approx 3.141592653589793$$
$$\pi \approx 3.14$$

Z dokładnością do 16 cyfr znaczących:

$$\sin(3.141592653589793/2) = 1$$
$$\sin(3.14/2) = 0.9999996829318346$$

$$e \approx 2.718281828459045$$
$$e \approx 2.71$$

Liczba π to trzy z hakiem

$$\pi \approx 3.141592653589793$$
$$\pi \approx 3.14$$

Z dokładnością do 16 cyfr znaczących:

$$\sin(3.141592653589793/2) = 1$$
$$\sin(3.14/2) = 0.9999996829318346$$

$$e \approx 2.718281828459045$$
$$e \approx 2.71$$
$$e \approx 2.72$$

Liczba π to trzy z hakiem

$$\pi \approx 3.141592653589793$$
$$\pi \approx 3.14$$

Z dokładnością do 16 cyfr znaczących:

$$\sin(3.141592653589793/2) = 1$$
$$\sin(3.14/2) = 0.9999996829318346$$

$$e \approx 2.718281828459045$$
$$e \approx 2.71$$
$$e \approx 2.72$$

Z dokładnością do 16 cyfr znaczących:

$$\ln(2.718281828459045) = 1$$

Liczba π to trzy z hakiem

$$\pi \approx 3.141592653589793$$
$$\pi \approx 3.14$$

Z dokładnością do 16 cyfr znaczących:

$$\sin(3.141592653589793/2) = 1$$
$$\sin(3.14/2) = 0.9999996829318346$$

$$e \approx 2.718281828459045$$
$$e \approx 2.71$$
$$e \approx 2.72$$

Z dokładnością do 16 cyfr znaczących:

$$\ln(2.718281828459045) = 1$$
$$\ln(2.71) = 0.9969486348916096$$

Liczba π to trzy z hakiem

$$\pi \approx 3.141592653589793$$
$$\pi \approx 3.14$$

Z dokładnością do 16 cyfr znaczących:

$$\sin(3.141592653589793/2) = 1$$
$$\sin(3.14/2) = 0.9999996829318346$$

$$e \approx 2.718281828459045$$
$$e \approx 2.71$$
$$e \approx 2.72$$

Z dokładnością do 16 cyfr znaczących:

$$\ln(2.718281828459045) = 1$$
$$\ln(2.71) = 0.9969486348916096$$
$$\ln(2.72) = 1.000631880307906$$

Liczba π to trzy z hakiem c.d.

Błędy bezwzględne i względne są przy obliczeniu $\log(e)$ i $\sin(\pi/2)$ takie same.

Liczba π to trzy z hakiem c.d.

Błędy bezwzględne i względne są przy obliczeniu $\log(e)$ i $\sin(\pi/2)$ takie same.

$$|1 - \ln(2.72)| = 6.318803079059521 \cdot 10^{-4},$$

Liczba π to trzy z hakiem c.d.

Błędy bezwzględne i względne są przy obliczeniu $\log(e)$ i $\sin(\pi/2)$ takie same.

$$|1 - \ln(2.72)| = 6.318803079059521 \cdot 10^{-4},$$
$$|1 - \sin(3.14/2)| = 3.170681653896779 \cdot 10^{-7}.$$

Liczba π to trzy z hakiem c.d.

Błędy bezwzględne i względne są przy obliczeniu $\log(e)$ i $\sin(\pi/2)$ takie same.

$$|1 - \ln(2.72)| = 6.318803079059521 \cdot 10^{-4},$$
$$|1 - \sin(3.14/2)| = 3.170681653896779 \cdot 10^{-7}.$$

Błąd względny obliczenia iloczynu:

$$|1 - (\ln(2.72))(\sin(3.14/2)) = 6.315630393913807 \cdot 10^{-4}$$

Suma błędów względnych:

$$6.318803079059521 \cdot 10^{-4} + 3.170681653896779 \cdot 10^{-7}$$
$$= 6.315632397405624 \cdot 10^{-4}.$$

Liczba e

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \approx 2.718281828459045$$

$$1/e = \lim_{n\to\infty}\left(1-\frac{1}{n}\right)^n \approx 0.3678794411714423$$

Liczba e – kolejne aproksymacje

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n \approx 2.718281828459045$$

Liczba e – kolejne aproksymacje

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \approx 2.718281828459045$$

n	$(1+1/n)^n$
1	2.000000000000000

Liczba e – kolejne aproksymacje

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \approx 2.718281828459045$$

n	$(1+1/n)^n$
1	2.000000000000000
10	2.593742460100002

Liczba e – kolejne aproksymacje

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \approx 2.718281828459045$$

n	$(1+1/n)^n$
1	2.000000000000000
10	2.593742460100002
100	2.704813829421528

Liczba e – kolejne aproksymacje

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \approx 2.718281828459045$$

n	$(1+1/n)^n$
1	2.000000000000000
10	2.593742460100002
100	2.704813829421528
1000	2.716923932235594

Liczba e – kolejne aproksymacje

$$e = \lim_{n\to\infty} \left(1 + \frac{1}{n}\right)^n \approx 2.718281828459045$$

n	$(1+1/n)^n$
1	2.000000000000000
10	2.593742460100002
100	2.704813829421528
1000	2.716923932235594
10000	2.718145926824926

Liczba e – kolejne aproksymacje

$$e = \lim_{n\to\infty}\left(1+\frac{1}{n}\right)^n \approx 2.718281828459045$$

n	$(1+1/n)^n$
1	2.000000000000000
10	2.593742460100002
100	2.704813829421528
1000	2.716923932235594
10000	2.718145926824926
100000	2.718268237192297

Liczba e – kolejne aproksymacje

$$e = \lim_{n \to \infty} \left(1 + \frac{1}{n}\right)^n \approx 2.718281828459045$$

n	$(1+1/n)^n$
1	2.000000000000000
10	2.593742460100002
100	2.704813829421528
1000	2.716923932235594
10000	2.718145926824926
100000	2.718268237192297
1000000	2.718280469095753

Typy całkowite

Wybrane typy.

Typy całkowite

Wybrane typy.

Ze znakiem

- shortint, 8 bitów, $-128\ldots127$,
- smallint, 16 bitów, $-32768\ldots32767$,
- longint, 32 bity, $-2147483648\ldots2147483647$.

Typy całkowite

Wybrane typy.

Ze znakiem

- shortint, 8 bitów, $-128 \ldots 127$,
- smallint, 16 bitów, $-32768 \ldots 32767$,
- longint, 32 bity, $-2147483648 \ldots 2147483647$.

Bez znaku

- byte, 8 bitów, $0 \ldots 255$,
- word, 16 bitów, $0 \ldots 65535$,
- longword, 32 bity, $0 \ldots 4294967295$.

Typy całkowite

Wybrane typy.

Ze znakiem

- shortint, 8 bitów, $-128\ldots127$,
- smallint, 16 bitów, $-32768\ldots32767$,
- longint, 32 bity, $-2147483648\ldots2147483647$.

Bez znaku

- byte, 8 bitów, $0\ldots255$,
- word, 16 bitów, $0\ldots65535$,
- longword, 32 bity, $0\ldots4294967295$.

Również o większym zakresie.

Typy całkowite w C++

- `short` – przynajmniej 16 bitów,
- `int` – nie mniejsze od `short`,
- `long` – przynajmniej 32 bity i nie mniejsze od `int`,
- `long long` – przynajmniej 64 bity i nie mniejsze od `long`.

Typy zmiennoprzecinkowe

Wybrane typy.

Typy zmiennoprzecinkowe

Wybrane typy.
32 bity – pojedyncza precyzja (single):

- zakres $1.5 \cdot 10^{-45} \ldots 3.4 \cdot 10^{38}$,
- liczba cyfr znaczących 7 lub 8.

Typy zmiennoprzecinkowe

Wybrane typy.
32 bity – pojedyncza precyzja (single):

- zakres $1.5 \cdot 10^{-45} \ldots 3.4 \cdot 10^{38}$,
- liczba cyfr znaczących 7 lub 8.

64 bity – podwójna precyzja (double)

- zakres $1.5 \cdot 10^{-324} \ldots 1.7 \cdot 10^{308}$,
- liczba cyfr znaczących 15 lub 16.

Typy zmiennoprzecinkowe

Wybrane typy.

32 bity – pojedyncza precyzja (single):

- zakres $1.5 \cdot 10^{-45} \ldots 3.4 \cdot 10^{38}$,
- liczba cyfr znaczących 7 lub 8.

64 bity – podwójna precyzja (double)

- zakres $1.5 \cdot 10^{-324} \ldots 1.7 \cdot 10^{308}$,
- liczba cyfr znaczących 15 lub 16.

Również o większym zakresie i dokładności.

Typy zmiennoprzecinkowe w C++

- `float` – przynajmniej 32 bity,
- `double` – przynajmniej 48 bitów i nie mniejsze niż `float`,
- `long double` – przynajmniej tyle bitów co w `double`.

Typy zmiennoprzecinkowe w C++

- `float` – przynajmniej 32 bity,
- `double` – przynajmniej 48 bitów i nie mniejsze niż `float`,
- `long double` – przynajmniej tyle bitów co w `double`.

Zwykle

- `float` – 32 bity,
- `double` – 64 bity,
- `long double` – 80, 96 lub 128 bitów.

Typy zmiennoprzecinkowe – reprezentacja (1)

Liczby zmiennoprzecinkowe są podzbiorem liczb wymiernych postaci

$$z = (-1)^s m p^w, \tag{1}$$

- p – podstawa (liczba naturalna > 1),
- m – mantysa (nieujemna liczba wymierna o skończonym rozwinięciu przy podstawie p),
- s – znak ($s = 0$ lub $s = 1$),
- w – wykładnik (liczba całkowita).

Typy zmiennoprzecinkowe – reprezentacja (2)

Jeżeli

$$1 \leq m < p, \tag{2}$$

to przedstawienie liczby $z \neq 0$ wzorem (1)est jednoznaczne.

W komputerze zawsze jest $p = 2$ (notacja binarna) natomiast do „normalnego" użytku jest $p = 10$ (notacja dziesiętna).

W notacji dziesiętnej stosuje się zapis $z = x\,Ey = x10^y$, gdzie y jest liczbą całkowitą, a x jest liczbą wymierną o skończonym rozwinięciu dziesiętnym.

Typy zmiennoprzecinkowe – reprezentacja (3)

Dla $p = 2$, jeżeli spełniony jest warunek (2) to część całkowita mantysy jest równa 1. Można wtedy przyjąć ją za domyślną i nie zapamiętywać, a zapamiętywać tylko część ułamkową. Dla takich m jest to liczba znormalizowana.
Wykładnik spełnia nierówność $w_{\text{min}} < w < w_{\text{max}}$.
Liczby

$$w_{\text{min}} + 1 \text{ oraz } w_{\text{max}} - 1$$

określają najmniejszy i największy wykładnik liczby.

Formaty zmiennoprzecinkowe

Format	s	Wykładnik w_B	l	Część ułamk. m'
32 bity	31	30 23		22 0
64 bity	63	62 52		51 0
80 bitów	79	78 64	63	62 0

Daje to następujące rozmiary pól:

Format	Wykładnik	Część ułamk.
32 bity	8	23
64 bity	11	52
80 bitów	15	63

Problemy z dodawaniem (1)

Symbolicznie *Maxima*:

$$\varepsilon = 2^{-63} = \frac{1}{9223372036854775808} \approx 1.084202172485504 \cdot 10^{-19}\,.$$

Problemy z dodawaniem (1)

Symbolicznie *Maxima*:

$$\varepsilon = 2^{-63} = \frac{1}{9223372036854775808} \approx 1.084202172485504 \cdot 10^{-19}.$$

$$(1+\varepsilon)^2 = \frac{85070591730234615884290395931651604481}{85070591730234615865843651857942052864} \approx 1.0,$$

Problemy z dodawaniem (1)

Symbolicznie *Maxima*:

$$\varepsilon = 2^{-63} = \frac{1}{9223372036854775808} \approx 1.084202172485504 \cdot 10^{-19}.$$

$$(1+\varepsilon)^2 = \frac{85070591730234615884290395931651604481}{85070591730234615865843651857942052864} \approx 1.0,$$

ale

$$(1+\varepsilon)^2 - 1 = \frac{18446744073709551617}{85070591730234615865843651857942052864}$$
$$\approx 2.168404344971009 \cdot 10^{-19}.$$

Problemy z dodawaniem (2)

Tylko numerycznie – *Maxima* lub Scilab:

$$(1+\varepsilon)^2 - 1 = 1 + 2\varepsilon + \varepsilon^2 - 1 \approx 0,$$

ale (dodawanie jest przemienne!)

$$(1+\varepsilon)^2 - 1 = 1 - 1 + 2\varepsilon + \varepsilon^2 \approx 2.168404344971009 \cdot 10^{-19}.$$

Problemy z dodawaniem (2)

Tylko numerycznie – *Maxima* lub Scilab:

$$(1+\varepsilon)^2 - 1 = 1 + 2\varepsilon + \varepsilon^2 - 1 \approx 0,$$

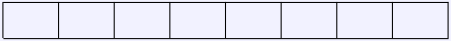
ale (dodawanie jest przemienne!)

$$(1+\varepsilon)^2 - 1 = 1 - 1 + 2\varepsilon + \varepsilon^2 \approx 2.168404344971009 \cdot 10^{-19}.$$

Problem. Porównać z wynikami otrzymanymi dla funkcji napisanych w C++ dla $\varepsilon = 2^{-w}$ i dla różnych w i dla typów `float`, `double` i .

Problemy z dodawaniem – najprostszy przykład

Kalkulator ośmiopozycyjny:



Problemy z dodawaniem – najprostszy przykład

Kalkulator ośmiopozycyjny:

Dodajemy $1 + 10^{-7} = 1 + 0.0000001$:

	1.	0	0	0	0	0	0	0	0
$+$	0.	0	0	0	0	0	0	1	0
$=$	1.	0	0	0	0	0	0	1	0

Problemy z dodawaniem – najprostszy przykład

Kalkulator ośmiopozycyjny:

Dodajemy $1 + 10^{-7} = 1 + 0.0000001$:

	1.	0	0	0	0	0	0	0	0
+	0.	0	0	0	0	0	0	1	0
=	1.	0	0	0	0	0	0	1	0

Dodajemy $1 + 10^{-8} = 1 + 0.00000001$:

	1.	0	0	0	0	0	0	0	0
+	0.	0	0	0	0	0	0	0	1
=	1.	0	0	0	0	0	0	0	1

Arytmetyka dla dociekliwych

Rozdział 4 w dziele Donalda Knutha [K] jest fascynującą choć trudną lekturą dla zainteresowanych zagadnieniami dokładności obliczeń.

Arytmetyka dla dociekliwych

Rozdział 4 w dziele Donalda Knutha [K] jest fascynującą choć trudną lekturą dla zainteresowanych zagadnieniami dokładności obliczeń.

Fragment motta do tego rozdziału (M. P. La Touche – 1878):

Jeśli na przykład dodajesz liczby od dołu do góry a potem znów od góry do dołu, to wynik zawsze będzie inny.

Arytmetyka dla dociekliwych

Rozdział 4 w dziele Donalda Knutha [K] jest fascynującą choć trudną lekturą dla zainteresowanych zagadnieniami dokładności obliczeń.

Fragment motta do tego rozdziału (M. P. La Touche – 1878):

Jeśli na przykład dodajesz liczby od dołu do góry a potem znów od góry do dołu, to wynik zawsze będzie inny.

$$\varepsilon = 2^{-63}.$$

Arytmetyka dla dociekliwych

Rozdział 4 w dziele Donalda Knutha [K] jest fascynującą choć trudną lekturą dla zainteresowanych zagadnieniami dokładności obliczeń.

Fragment motta do tego rozdziału (M. P. La Touche – 1878):

Jeśli na przykład dodajesz liczby od dołu do góry a potem znów od góry do dołu, to wynik zawsze będzie inny.

$$\varepsilon = 2^{-63}.$$

Dodawanie jest co prawda przemienne, ale ...

$$\begin{aligned}(1+\varepsilon)^2 - 1 &= 1 + 2\varepsilon + \varepsilon^2 - 1 \approx 0,\\ (1+\varepsilon)^2 - 1 &= 1 - 1 + 2\varepsilon + \varepsilon^2 \approx 2\varepsilon\\ &= 2.168404344971009 \cdot 10^{-19}.\end{aligned}$$

Obliczenia liczby e – typy word i single/double

$$e = \lim_{n\to\infty}\left(1 + \frac{1}{n}\right)^n \approx 2.71828182845905.$$

word	double		single	
n	$1 + 1/n$	$\approx e$	$1 + 1/n$	$\approx e$
10000	1.00010000000000	2.71814592682490	1.000100017	2.71859574317932
20000	1.00005000000000	2.71821387453362	1.000049949	2.71542286872864
30000	1.00003333333333	2.71823652514588	1.000033379	2.72192811965942
40000	1.00002500000000	2.71824785070851	1.000025034	2.72193431854248
50000	1.00002000000000	2.71825464612676	1.000020027	2.72193503379822
60000	1.00001666666667	2.71825917647437	1.000016689	2.72193241119385
4464	1.00022401433692	2.71797742391736	1.000223994	2.71773314476013

Obliczenia liczby e – typy longint i single/double

$$e \approx 2.71828182845905.$$

longint	double		single	
n	$1+1/n$	$\approx e$	$1+1/n$	$\approx e$
$1 \cdot 10^5$	1.00001000000000	2.71826823719229	1.000010014	2.72191071510315
$2 \cdot 10^5$	1.00000500000000	2.71827503280339	1.000005007	2.72180390357971
$3 \cdot 10^5$	1.00000333333333	2.71827729802102	1.000003338	2.72161221504211
$4 \cdot 10^5$	1.00000250000000	2.71827843051163	1.000002503	2.72104191780090
$5 \cdot 10^5$	1.00000200000000	2.71827911026038	1.000002027	2.75324344635010
$6 \cdot 10^5$	1.00000166666667	2.71827956324549	1.000001669	2.72057056427002
$7 \cdot 10^5$	1.00000142857143	2.71827988666858	1.000001431	2.71957755088806
$8 \cdot 10^5$	1.00000125000000	2.71828012943192	1.000001192	2.59271550178528
$9 \cdot 10^5$	1.00000111111111	2.71828031832189	1.000001073	2.62278890609741
$10 \cdot 10^5$	1.00000100000000	2.71828046909594	1.000000954	2.58985233306885

Niestabilność numeryczna – przykład [PLSG]

Algorytm numerycznie stabilny – algorytm, który dla nieco zaburzonych danych zwraca nieco zaburzone wyniki.

Niestabilność numeryczna – przykład [PLSG]

Algorytm numerycznie stabilny – algorytm, który dla nieco zaburzonych danych zwraca nieco zaburzone wyniki.

Obliczamy całkę

$$y_n = \int_0^1 \frac{x^n}{x+a}\,dx,$$

Niestabilność numeryczna – przykład [PLSG]

Algorytm numerycznie stabilny – algorytm, który dla nieco zaburzonych danych zwraca nieco zaburzone wyniki.

Obliczamy całkę

$$y_n = \int_0^1 \frac{x^n}{x+a}\, dx,$$

Ponieważ

$$y_n + a y_{n-1}$$

Niestabilność numeryczna – przykład [PLSG]

Algorytm numerycznie stabilny – algorytm, który dla nieco zaburzonych danych zwraca nieco zaburzone wyniki.

Obliczamy całkę

$$y_n = \int_0^1 \frac{x^n}{x+a}\, dx,$$

Ponieważ

$$y_n + a y_{n-1} = \int_0^1 \frac{x^n + a x^{n-1}}{x+a}\, dx$$

Niestabilność numeryczna – przykład [PLSG]

Algorytm numerycznie stabilny – algorytm, który dla nieco zaburzonych danych zwraca nieco zaburzone wyniki.

Obliczamy całkę

$$y_n = \int_0^1 \frac{x^n}{x+a}\,dx,$$

Ponieważ

$$y_n + a y_{n-1} = \int_0^1 \frac{x^n + a x^{n-1}}{x+a}\,dx$$

$$= \int_0^1 \frac{x^{n-1}(x+a)}{x+a}\,dx$$

Niestabilność numeryczna – przykład [PLSG]

Algorytm numerycznie stabilny – algorytm, który dla nieco zaburzonych danych zwraca nieco zaburzone wyniki.

Obliczamy całkę

$$y_n = \int_0^1 \frac{x^n}{x+a}\, dx,$$

Ponieważ

$$y_n + a y_{n-1} = \int_0^1 \frac{x^n + a x^{n-1}}{x+a}\, dx$$

$$= \int_0^1 \frac{x^{n-1}(x+a)}{x+a} dx = \int_0^1 x^{n-1}\, dx$$

Niestabilność numeryczna – przykład [PLSG]

Algorytm numerycznie stabilny – algorytm, który dla nieco zaburzonych danych zwraca nieco zaburzone wyniki.
Obliczamy całkę

$$y_n = \int_0^1 \frac{x^n}{x+a}\,dx,$$

Ponieważ

$$y_n + a y_{n-1} = \int_0^1 \frac{x^n + a x^{n-1}}{x+a}\,dx$$

$$= \int_0^1 \frac{x^{n-1}(x+a)}{x+a}\,dx = \int_0^1 x^{n-1}\,dx = \frac{1}{n}\,x^n\Big|_0^1$$

Niestabilność numeryczna – przykład [PLSG]

Algorytm numerycznie stabilny – algorytm, który dla nieco zaburzonych danych zwraca nieco zaburzone wyniki.

Obliczamy całkę

$$y_n = \int_0^1 \frac{x^n}{x+a}\, dx,$$

Ponieważ

$$y_n + a y_{n-1} = \int_0^1 \frac{x^n + a x^{n-1}}{x+a}\, dx$$

$$= \int_0^1 \frac{x^{n-1}(x+a)}{x+a} dx = \int_0^1 x^{n-1}\, dx = \frac{1}{n} x^n \Big|_0^1 = \frac{1}{n},$$

Niestabilność numeryczna – przykład [PLSG]

Algorytm numerycznie stabilny – algorytm, który dla nieco zaburzonych danych zwraca nieco zaburzone wyniki.
Obliczamy całkę

$$y_n = \int_0^1 \frac{x^n}{x+a}\,dx,$$

Ponieważ

$$y_n + a y_{n-1} = \int_0^1 \frac{x^n + a x^{n-1}}{x+a}\,dx$$

$$= \int_0^1 \frac{x^{n-1}(x+a)}{x+a}\,dx = \int_0^1 x^{n-1}\,dx = \frac{1}{n} x^n \Big|_0^1 = \frac{1}{n},$$

to

$$y_n = \frac{1}{n} - a y_{n-1}.$$

Obliczenie całki – algorytm 1

Mamy więc wzór rekurencyjny na y_n:

$$y_n = \frac{1}{n} - a y_{n-1} \,.$$

Obliczenie całki – algorytm 1

Mamy więc wzór rekurencyjny na y_n:

$$y_n = \frac{1}{n} - a y_{n-1} \, .$$

Potrzebne jest jeszcze y_0:

Obliczenie całki – algorytm 1

Mamy więc wzór rekurencyjny na y_n:

$$y_n = \frac{1}{n} - a y_{n-1}\,.$$

Potrzebne jest jeszcze y_0:

$$y_0 = \int_0^1 \frac{dx}{x+a} = \ln(x+a)\Big|_0^1 = \ln(1+a) - \ln a$$
$$= \ln\frac{1+a}{a}\,.$$

Wyniki dla pojedynczej precyzji, $a = 5$

n	y_n
0	0.18232156
1	0.08839220
2	0.05803901
3	0.04313828
4	0.03430858
5	0.02845709
6	0.02438123
7	0.02095100
8	0.02024502
9	0.00988603
10	0.05056984
11	-0.16194010

Dla $n = 11$ wynik na pewno błędny, bo zawsze $y_n > 0$.

Wyniki dla pojedynczej precyzji, $a = 5$

n	y_n
0	0.18232156
1	0.08839220
2	0.05803901
3	0.04313828
4	0.03430858
5	0.02845709
6	0.02438123
7	0.02095100
8	0.02024502
9	0.00988603
10	0.05056984
11	-0.16194010

Dla $n = 11$ wynik na pewno błędny, bo zawsze $y_n > 0$.
W każdym kroku błąd jest mnożony przez a.

Wyniki dla podwójnej precyzji, $a = 5$

n	y_n
0	0.18232156
1	0.08839222
2	0.05803892
3	0.04313873
4	0.03430633
	....
10	0.01536755
11	0.01407134
	....
19	0.00745741
20	0.01271297
21	−0.01594578

Dla $n = 21$ wynik na pewno błędny, bo zawsze $y_n > 0$.

Wyniki *Maxima* nierekurencyjnie, $a = 5$

n	y_n
0	0.1823215567939547
1	0.08839221603022729
2	0.05803891984886889
3	0.04313873408901259
4	0.03430632955496549
5	0.02846835222499067
6	0.0243249055420165
7	0.02123261515225749
8	0.01883692422416061
9	0.01692648977041245
10	0.0153675489127636
	....
20	0.03125
21	0.0

Obliczenie całki – algorytm 2

Zamiast

$$y_n = \frac{1}{n} - a y_{n-1}$$

korzystamy ze wzoru

$$y_{n-1} = \frac{1}{an} - \frac{1}{a} y_n \,.$$

Dla dużych n jest $y_n = y_{n-1}$. Przyjmijmy

$$y_{20} = \frac{1}{a \cdot 21} - \frac{1}{a} y_{21} \,,$$

czyli

$$y_{20} \approx \frac{1}{a \cdot 21} + \frac{1}{a} y_{20} \,,$$

skąd

$$y_{20} \approx \frac{1}{a \cdot 21\,(1 + 1/a)} \,.$$

Obliczenie całki – algorytm 2

Dla $a = 5$:

$$y_{20} \approx \frac{1}{a \cdot 21\,(1 + 1/a)} = \frac{1}{5 \cdot 21\,(1 + 1/5)} = \frac{1}{126} = 0.0079365079365079365$$

oraz

$$y_{n-1} = \frac{1}{5n} - \frac{1}{5}\,y_n\,.$$

Wyniki dla podwójnej precyzji, $a = 5$

n	y_n
0	0.18232156
1	0.08839222
2	0.05803892
3	0.04313873
4	0.03430633
5	0.02846835
6	0.02432491
7	0.02123262
8	0.01883692
9	0.01692649
10	0.01536755

n	y_n
11	0.01407134
12	0.01297664
13	0.01203988
14	0.01122918
15	0.01052075
16	0.00989623
17	0.00934236
18	0.00884378
19	0.00841270
20	0.00793651

Dokładne wartości numerycznie

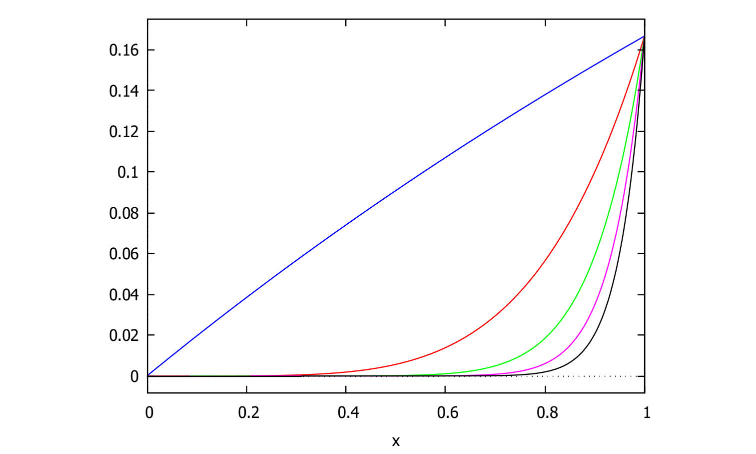
n	y_n
15	0.0105207353428904
16	0.0098963323285548
17	0.00934186776899051
18	0.00884621671060297
19	0.00840049539435356
20	0.00799752302823216
25	0.00645030526686521
30	0.00540463296514068
31	0.00523489969042563

Według https://pl.numberempire.com/definiteintegralcalculator.php

Wykresy

$$\frac{x^n}{x + a}$$

dla $a = 5,\ n = 1, 5, 10, 15, 20.$



Symbolicznie

$$\int_0^1 \frac{x^n}{x+a}\,dx$$

dla $a > 0$.

Symbolicznie

$$\int_0^1 \frac{x^n}{x+a}\,dx$$

dla $a > 0$.

n	całka
1	$a\left(\ln(a) - \ln(a+1)\right) + 1$
2	$a^2\left(\ln(a+1) - \ln(a)\right) - a + 1/2$
3	$a^3\left(\ln(a) - \ln(a+1)\right) + a^2 - a/2 + 1/3$
4	$a^4\left(\ln(a+1) - \ln(a)\right) - a^3 - a^2/2 + a/3 + 1/4$
	$\ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots$
n	$a^n(-1)^n\left(\ln(a+1) - \ln(a)\right) - \sum_{k=1}^{n}(-1)^{n-k}\frac{a^{n-k}}{k}$