Zestaw 1 - Arytmetyka komputerowa

Piotr Karamon

Znaleźć maszynowe epsilon

Maszynowe epsilon dla systemu zmiennoprzecinkowego jest to najmniejsza dodatnia liczba ϵ którą można dodać do 1, aby otrzymać wynik różny od 1. Taka liczba musi mieć ten sam wykładnik jak liczba 1 oraz jak najmniejszą mantysę. Maszynowe epsilon zależy od podstawy systemu β i precyzji p. Wyraża się wzorem:

$$\epsilon = \beta^{1-p}$$

Maszynowe epsilon można również obliczyć poprzez dzielenie zmiennej a której początkowa wartość to 1 przez podstawę systemu. Ostatnia wartość a która spełnia warunek $a + 1 > 1$ jest maszynowym epsilon.

Problem ewaluacji funkcji $\sin(x)$

Rozważamy problem ewaluacji funkcji $\sin(x)$, m.in. propagację błędu danych wejściowych, tj. błąd wartości funkcji ze względu na zakłócenie h w argumencie x.

Błąd bezwzględny przy ewaluacji $\sin(x)$

$$\Delta \sin(x) = |\sin(x + h) - \sin(x)|$$

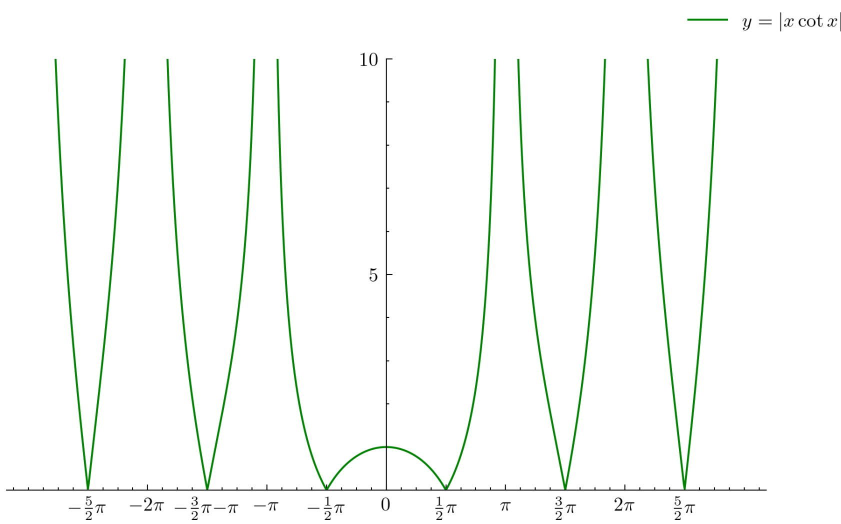
Błąd względny przy ewaluacji $\sin(x)$

$$\frac{\Delta \sin(x)}{\sin x} = \frac{|\sin(x + h) - \sin(x)|}{\sin x}$$

Uwarunkowanie

$$\mathrm{cond}(f(x)) = \left| \frac{x \cdot f'(x)}{f(x)} \right|$$

$$\mathrm{cond}(\sin(x)) = \left| \frac{x \cdot \cos(x)}{\sin(x)} \right| = |x \cot x|$$



Rysunek 1: Wykres funkcji będącej współczynnikiem uwarunkowania dla problemu obliczania funkcji $\sin x$

Ocena czułości problemu

Problem jest bardzo czuły w miejscach, gdzie $\cot x$ zbiega do ∞. Są to liczby postaci $x = k\pi$, gdzie k jest liczbą całkowitą z wyjątkiem 0. Problem jest bardzo dobrze uwarunkowany w miejscach w których $\cos x$ przyjmuje wartość 0. Są to liczby postaci $x = k\pi + \frac{\pi}{2}$, gdzie k jest liczbą całkowitą.

Wnioski

Problem ewaluacji $\sin x$ jest najgorzej uwarunkowany w otoczeniu miejsc zerowych tej funkcji. Problem jest jednocześnie najlepiej uwarunkowany w otoczeniu ekstremów lokalnych funkcji $\sin x$.

Różne rodzaje błędów podczas obliczania $\sin x$

Funkcja $\sin x$ zadana jest nieskończonym ciągiem:

$$\sin(x) = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \ldots$$

$$\begin{aligned} \text{Błąd progresywny} \quad & \Delta y = |\hat{y} - y| \\ \text{Błąd wsteczny} \quad & \Delta x = |\hat{x} - x|, \quad \text{gdzie} \quad f(\hat{x}) = \hat{y} \end{aligned}$$

Błąd progresywny i wsteczny przy przybliżeniu $\sin x \approx x$

$$\begin{aligned} \text{Błąd progresywny} \quad & \Delta y = |x - \sin x| \\ \text{Błąd wsteczny} \quad & \Delta x = |\arcsin \hat{y} - y| = |\arcsin x - \sin x| \end{aligned}$$

Tablica 1: Wartości błędu progresywnego i wstecznego jeśli przybliżamy-funkcję $\sin x \approx x$

x	$y = \sin(x)$	$\hat{y} = x$	$\Delta y = \lvert\hat{y} - y\rvert$	$\hat{x} = \arcsin \hat{y}$	$\Delta x = \lvert\hat{x} - x\rvert$
0.10000000	0.09983342	0.10000000	0.00016658	0.10016742	0.00016742
0.50000000	0.47942554	0.50000000	0.02057446	0.52359878	0.02359878
1.00000000	0.84147098	1.00000000	0.15852902	1.57079633	0.57079633

Błąd progresywny i wsteczny przy przybliżeniu $\sin x \approx x - \frac{x^3}{6}$

$$\begin{aligned} \text{Błąd progresywny} \quad & \Delta y = |x - \frac{x^3}{6} - \sin x| \\ \text{Błąd wsteczny} \quad & \Delta x = \left| \arcsin\left(x - \frac{x^3}{6}\right) - \sin x \right| \end{aligned}$$

Tablica 2: Wartości błędu progresywnego i wstecznego jeśli przybliżamy-funkcję $\sin x \approx x - \frac{x^3}{6}$

x	$y = \sin(x)$	$\hat{y} = x - \frac{x^3}{6}$	$\Delta y = \lvert\hat{y} - y\rvert$	$\hat{x} = \arcsin \hat{y}$	$\Delta x = \lvert\hat{x} - x\rvert$
0.10000000	0.09983342	0.09983333	0.00000008	0.09999992	0.00000008
0.50000000	0.47942554	0.47916667	0.00025887	0.49970504	0.00029496
1.00000000	0.84147098	0.83333333	0.00813765	0.98511078	0.01488922

Wnioski

Przybliżenie z większą ilością wyrazów szeregu powoduje znaczące zmniejszenie błędu zarówno progresywnego jak i wstecznego. Możemy się spodziewać, że im więcej wyrazów weźmiemy tym dokładniejszy wynik uzyskamy.

Znormalizowany system zmiennoprzecinkowy

Zakładamy że mamy znormalizowany system zmiennoprzecinkowy z $\beta = 10$, $p = 3$, $L = -98$.

Wartość poziomu UFL(underflow) dla takiego systemu.

Wartość poziomu UFL to najmniejsza liczba dodatnia, jaka może zostać zapisana w danym systemie. System jest znormalizowany zatem mantysa takiej liczby będzie równa 1. Najmniejszy wykładnik jaki możemy uzyskać w naszym systemie jest równy $L = -98$. Zatem

$$\mathrm{UFL} = \beta^L = 10^{-98}$$

Wynik $x - y$

$x = 6.87 \cdot 10^{-97}$ $y = 6.81 \cdot 10^{-97}$

$$x - y = 6 \cdot 10^{-99} < \mathrm{UFL} \implies x - y = 0$$

Wynik operacji $x - y$ w takim systemie wyniesie 0.

Wnioski

UFL jest miarą dokładności systemu zmiennoprzecinkowego. Aby system mógł sobie radzić z małymi liczbami parametr L takiego systemu powinien być możliwe jak najmniejszy.

Bibliografia

- Prof. Michael T. Heath: *Scientific Computing: An Introductory Survey Chapter 1 – Scientific Computing*
- `https://pl.wikipedia.org/wiki/IEEE_754`
- `https://en.wikipedia.org/wiki/Machine_epsilon`