УДК 519.7
©Л. Мороз[1]*, А. Гринчишин*[1] *, 2015*

МОДИФІКОВАНИЙ МЕТОД ШВИДКОГО ОБЧИСЛЕННЯ ОБЕРНЕНОГО КВАДРАТНОГО КОРЕНЯ З ВИКОРИСТАННЯМ МАГІЧНОЇ КОНСТАНТИ

Запропоновано модифікований метод обчислення оберненого квадратного кореня з метою підвищення точності обчислень. Коригування сталих коефіцієнтів формул Ньютона-Рафсона дає змогу зменшити максимальне значення відносної похибки приблизно у 7 разів у всьому діапазоні значень чисел типу float .

Ключові слова: обчислення, обернений квадратний корінь, формула Ньютона-Рафсона, похибка обчислення.

This paper suggests a modifying method of calculation of inverse square root to improve the accuracy of calculations. Correction of constant coefficients of Newton-Raphson formulas enables to decrease maximum value of relative errors approximately in 7 times in all range of numbers of float type.

Keywords: calculation, inverse square root, Newton-Raphson formula, calculation error.

1. МЕТА РОБОТИ

Відомо, що метод обчислення оберненого квадратного кореня з використанням оптимальної магічної константи (“fastinversesquareroot”) [1,2]: floatxhalf = 0.5f*x;

```
int i = *(int*)&x;
i = 0x5f375a86 - (i>>1);
y= *(float*)&i;{ початкове лінійне наближення y0 }
y:=  y*(1.5f- xhalf *y*y);{перша ітерація - y1 }
y:= y*(1.5f- xhalf *y*y); {друга ітерація- y2 }
```

у деяких випадках не задовольняє вимоги обчислювальних практик в окремих областяхнауки [3,4] через відносно низьку точність (17.7 коректних біти результату після двох Ньютонівських ітерацій), тому зменшення похибок таких обчислень є актуальною задачею.

[1]Національний університет "Львівська політехніка"

2. ОПИС ВІДОМОГО МЕТОДУ

Для функції оберненого квадратного кореняпочаткове лінійне наближення y_0 у наближеному вигляді записується як [2]

$$y_{00}=(\alpha x+\beta)\cdot 2^{E_p}, \tag{1}$$

де

$$\alpha=-2^{-(E_x+1)} \tag{2}$$

$$\beta=[\frac{1}{2}(3-bias)+Q+t-\frac{1}{2}E_x-e_p], \tag{3}$$

можна подати уявному вигляді за допомогою трьох лінійних наближень y_{01}, y_{02} та y_{03}:

$$y_{01}=-x+\frac{3}{2}+t,\ x\in[0.5,x_t); \tag{4}$$

$$y_{02}=-\frac{1}{2}x+\frac{5}{4}+\frac{t}{2},\ x\in[x_t,1); \tag{5}$$

$$y_{03}=-\frac{1}{4}x+1+\frac{t}{2},\ x\in[1,2). \tag{6}$$

Тут

$$T=3627654,\ t=0.432450056076, \tag{7}$$

$$x_t=0.5+t=.932450056076,$$

$$bias=127,\ Q=190,$$

$$t=\frac{t}{N_m},\ N_m=2^{23}.$$

Абсолютна Δ_1 та відносна δ_1 похибки після проведення першої ітерації за класичною формулою Ньютона-Рафсона

$$y_1=y_0(1.5-0.5xy_0^2) \tag{8}$$

для $x\in[0.5,x_t)$, де розташовані максимуми цих похибок, матимуть вигляд

$$\Delta_1=\frac{1}{16}\cdot(3+2t-2x)\cdot(-12+9x+12xt-12x^2+4xt^2-8x^2t+4x^3)-\frac{1}{\sqrt{x}};$$

$$\delta_1=\Delta_1\cdot\sqrt{x}. \tag{9}$$

Максимуми відносної похибки δ_1 розміщені в точках :

$x_{11m}=1/2+t/3$; (10)

$x_{12m}=x_t=1/2+t$; (11)

а нуль цієї ж похибки – у точці :

$$x_{10\_zero}=-1/12\cdot(-36t^2-8t^3-54t+81+6(-48t^3-216t^2-$$
$$-324t+162)^{1/2})^{1/3}+3(-t^2/9-t/3-1/4)/(-36t^2-8t^3- \qquad (12)$$
$$-54t+81++6(-48t^3-216t^2-324t+162)^{1/2})^{1/3}+1+2/3t.$$

Максимуми та нуль відносної похибки δ_2 після другої ітерації за класичною формулою Ньютона-Рафсона

$$y_2=y_1(1.5-0.5xy_1^2) \qquad (13)$$

на цьому ж проміжку $x\in[0.5,x_t)$, знаходяться у тих самих точках.

3. ОПИС МОДИФІКОВАНОГО МЕТОДУ

Для підвищення точності обчислень оберненого квадратного кореняможна застосувати спосіб адитивної корекції результатів обчислень на кожній ітерації, запропонований у роботі [5].

Для цього слід провести першу ітерацію за модифікованою формулою Ньютона-Рафсона:

$$y_{1b}=y_0(1.5+k_1-0.5xy_0^2). \qquad (14)$$

Тоді отримаємо таку відносну похибку :

$$\delta_{1b}=(y_{1b}-y_t)/y_t, \qquad (15)$$

або

$$\delta_{1b}=(9/4+3/2k_1-51/16x-27/8tx+27/8x^2-9/4xt^2+9/2tx^2-9/4x^3+ =$$
$$+3/2t+tk_1-1/2xt^3+3/2t^2x^2-3/2tx^3-xk_1+1/2x^4-x^{-1/2})x^{1/2}.$$

Існують три точки максимуму відносної похибки δ_{1b} для першої ітерації, проведеної за модифікованою формулою Ньютона-Рафсона :

$x_{11mb}=x_{11m}=1/2+t/3;$

$$x_{12mb} = x_{10\_zero} = -1/12 \cdot (-36t^2 - 8t^3 - 54t + 81 + 6(-48t^3 -$$
$$216t^2 - 324t + 162)^{1/2})^{1/3} + 3(-t^2/9 - t/3 - 1/4)/(-36t^2 - 8t^3$$
$$-54t + +81 + 6(-48t^3 - 216t^2 - 324t + 162)^{1/2})^{1/3} + 1 + 2/3t;$$
$$x_{13mb} = x_{12m} = x_t = 1/2 + t\,.$$

Для знаходження значення k_1 створимо систему з двох рівнянь:

1. $x = x_{10\_zero}$;

$$\delta_{11b} = (9/4 + 3/2k_1 - 51/16x - 27/8tx + 27/8x^2 - 9/4xt^2 + 9/2tx^2 - 9/4x^3 +$$
$$+3/2t + tk_1 - 1/2xt^3 + 3/2t^2x^2 - 3/2tx^3 - xk_1 + 1/2x^4 - x^{-1/2})x^{1/2};$$

2. $x = x_t$;

$$\delta_{12b} = (9/4 + 3/2k_1 - 51/16x - 27/8tx + 27/8x^2 - 9/4xt^2 + 9/2tx^2 - 9/4x^3 +$$
$$+3/2t + tk_1 - 1/2xt^3 + 3/2t^2x^2 - 3/2tx^3 - xk_1 + 1/2x^4 - x^{-1/2})x^{1/2}.$$

Похибки при цих значеннях x мають бути рівними за модулем, але протилежними за знаками:

$$\delta_{11b} + \delta_{12b} = 0\,. \qquad (16)$$

Розв'язком цього рівняння буде значення k_1

$$k_1 = 0.00089090\,. \qquad (17)$$

Однак слід підкреслити, що дане теоретичне значення не враховує похибок заокруглень та відсікання, тому на практиці може спостерігатись відхилення у наймолодшому двійковому розряді мантиси при поданні уточненого коефіцієнта $(1.5+k_1)$ у вигляді числа типу float. З врахуванням (17) рівняння (14) набере вигляду

$$y_{1b} = y_0(1.50089090 - 0.5xy_0^2) \qquad (17)$$

Аналогічно, проведемо другу ітерацію за подібною модифікованою формулою:

$$y_{2b} = y_{1b}(1.5 + k_2 - 0.5xy_{1b}^2)\,, \qquad (18)$$

де $k_2 = 0.00000060$. Формула (18) в програмі матиме вигляд:

$$y_{2b} = y_{1b}(1.50000060 - 0.5xy_{1b}^2)\,. \qquad (19)$$

Тоді програма набере остаточного вигляду:

```
float xhalf = 0.5f*x;
int i = *(int*)&x;
i = 0x5f375a86 - (i>>1);
y= *(float*)&i;
    y:=  y*(1.50089090- xhalf *y*y);
    y:= y*(1.50000060- xhalf *y*y);
```

Графіки похибок δ_2 та δ_{2b} наведені на рис.1.

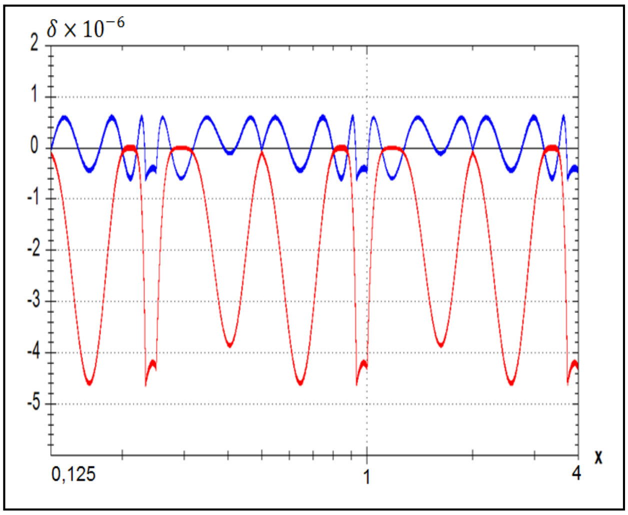

Рис.1 Графіки відносних похибок δ_2 (червона крива) та δ_{2b} (синя крива) для $x \in [0.125, 4)$

Модифікація сталих коефіцієнтів формул Ньютона-Рафсона дає змогу підвищити точність обчислень приблизно у 7 разів у всьому діапазоні значень чисел типу float (20.5 коректних біти результату після двох Ньютонівських ітерацій) :

$$\delta_{2\max} = -4.65 \cdot 10^{-6};$$

$$\delta_{2b\max} = -6.52 \cdot 10^{-7}$$

$$v = \frac{\delta_{2\max}}{\delta_{2b\max}} \approx 7.$$

Слід відмітити, що перехід від первісної магічної константи $0x5f3759df$ до оптимальної константи $0x5F375A86$ зменшив максимальну відносну похибку лише на 0.12% [1], а запропонований метод – понад 700%.

4.ВИСНОВКИ

У даній роботі запропоновано модифікований метод обчислення оберненого квадратного кореня з використанням магічної константи з метою підвищення точності обчислень. Уточнення сталих коефіцієнтів формул Ньютона-Рафсона дає змогу зменшити максимальне значення відносної похибки приблизно у 7 разів у всьому діапазоні значень чисел типу float .

1. Chris Lomont, Fast Inverse Square Root, 2003. http://www.lomont.org/Math/Papers/2003/InvSqrt.pdf 2.Л. Мороз, А. Гринчишин . Швидке обчислення оберненого квадратного кореня з використанням магічної константи - аналітичний підхід. "Комп'ютерні технології друкарства", № 32, 2014, с.38-51. 3.T.O. Hands, I. Griffiths, D.A. Marshall, G. Douglas. The fast inverse square root in scientific computing. Journal of Physics Special Topics, A2_1, Oct 20, 2011. 4.Grimm, Simon L., and Joachim G. Stadel. "The GENGA Code: Gravitational Encounters in N-body simulations with GPU Acceleration." The Astrophysical Journal 796.1 (2014): 23. 5.Мороз Л.В. Теорія та швидкодіючі апаратно-програмні засоби ітераційних методів обчислення функцій. - Автореф. д.т.н. – Національний університет "Львівська політехніка", Львів, 2013.