

Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский университет ИТМО»

Факультет программной инженерии и компьютерной техники

Домашняя работа 6 по дисциплине «Основы дискретной математики»

СЛОЖЕНИЕ ЧИСЕЛ С ПЛАВАЮЩЕЙ ЗАПЯТОЙ

Вариант №53

Выполнил: Баукин Максим Александрович

Группа: Р3132

Принимающий: Поляков Владимир Иванович

Должность: доцент факультета ПИиКТ

Г. Санкт-Петербург 2024

| | |
|-------|-------|
| 231,2 | 260,8 |
|-------|-------|

Задача:

- Заданные числа A и B представить в форме с плавающей запятой в разрядных сетках форматов $\Phi 1$ и $\Phi 2$ с укороченной мантиссой (12 двоичных разрядов).
Примечание: общее число разрядов в формате – 20.
- Выполнить операцию сложения заданных чисел со следующими комбинациями знаков операндов: “++”, “+–”, “–+” в разрядных сетках форматов $\Phi 1$ и $\Phi 2$.
- Результаты представить в форматах операндов, перевести в десятичную систему счисления и проверить их правильность.
- Определить абсолютную и относительную погрешности результатов и обосновать их причину.
- Сравнить погрешности результатов аналогичных операций для форматов $\Phi 1$ и $\Phi 2$ и объяснить причины их сходства или различия.

Выполнение:

$$A = 231,2$$

$$B = 260,8$$

1.1 Формат $\Phi 1$

$$A = (231,2)_{10} = (E7,333333)_{16} = (0,E733333)_{16} \cdot 16^2$$

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

$$B = (260,8)_{10} = (104,CCCCCD)_{16} = (0,104CCCCCD)_{16} \cdot 16^3$$

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

$$\begin{array}{r}
 X_A = \underline{-1} \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \\
 X_B = \underline{1} \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 1 \\
 \hline
 (X_A - X_B)_{\text{доп.}} = 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1 \quad 1
 \end{array}$$

$(X_A - X_B)_{\text{доп.}} = -1; X_C = X_B = 3.$

a) Оба операнда положительные ($A > 0, B > 0$):

$$\begin{array}{r}
 2,3) \quad \overrightarrow{\mathbf{M}_A} = \overset{4}{\overbrace{. \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 0}} \quad \overbrace{0 \quad 1 \quad 1 \quad 1} \\
 \mathbf{M}_B = \overline{+ \quad . \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 1} \\
 \mathbf{M}_C = \quad . \quad 0 \quad 0 \quad 0 \quad 1 \quad 1 \quad 1 \quad 1 \quad 0 \quad 1 \quad 1 \quad 0 \quad 0
 \end{array}$$

Результат сложения нормализован.

$$M_C = .00011101100$$

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

$$C^* = M_C \cdot 16^{P_c} = (0,1EC)_{16} \cdot 16^3 = 492.$$

Определим абсолютную и относительную погрешности результата:
 $\Delta C = 492 - 492 = 0$

$$\delta C = \left| \frac{0}{492} \right| \cdot 100\% = 0\%$$

6) A>0, B<0:

$$M_A = .000011100111$$

$$M_B = \overline{.000100000101}$$

$$M_C = \overline{.111111100010}$$

Результат вычитания денормализован вправо и представлен в дополнительном коде.

$$M_C = .111000100000$$

Т.к. выполнен сдвиг мантиссы влево, характеристику результата нужно уменьшить на 1 ($X_C = X_C - 1 = 2$).

| | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

$$C^* = M_C \cdot 16^{P_c} = (-0,1E0)_{16} \cdot 16^2 = -30.$$

Определим абсолютную и относительную погрешности результата:
 $\Delta C = -29,6 - (-30) = 0,4$

$$\delta C = \left| \frac{0,4}{-29,6} \right| \cdot 100\% = 1,35135\%$$

Результат получился представленным с избыtkом. Этот факт можно объяснить потерей значащих разрядов мантиссы результата при его нормализации.

c) A<0, B>0:

$$\begin{array}{r} M_B = .000100000101 \\ - \\ M_A = .000011100111 \\ \hline M_C = .000000011110 \end{array}$$

Результат вычитания денормализован вправо.

$$M_C = .000111100000$$

Т.к. выполнен сдвиг мантиссы влево, характеристику результата нужно уменьшить на 1 ($X_C = X_C - 1 = 2$).

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

$$C^* = M_C \cdot 16^{P_c} = (0,1E0)_{16} \cdot 16^2 = 30.$$

Определим абсолютную и относительную погрешности результата:
 $\Delta C = 29,6 - 30 = -0,4$

$$\delta C = \left| \frac{-0,4}{29,6} \right| \cdot 100\% = 1,35135\%$$

Результат получился представленным с избытком. Этот факт можно объяснить потерей значащих разрядов мантиссы результата при его нормализации.

2.1 Формат Ф2

$$A = (231,2)_{10} = (E7,333333)_{16} = (0,111001110011001100110011)_2 \cdot 2^8$$

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 1 | 1 | 1 | 0 | 0 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

$$B = (260,8)_{10} = (104,CCCCCD)_{16} = (0,1000001001100110011001101)_2 \cdot 2^9$$

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

$$X_A = 10001000$$

$$X_B = 10001001$$

$$(X_A - X_B)_{\text{доп.}} = 11111111$$

$$(X_A - X_B) = -1; X_C = X_B = 9$$

a) A>0, B>0:

$$M_A = .011100111001$$

+

$$M_B = .100000100110$$

$$M_C = \overline{.111101011111}$$

Результат сложения нормализован.

$$M_C = .111101011111$$

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 1 | 1 | 1 | 0 | 1 | 0 | 1 | 1 | 1 | 1 | 1 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

$$C^* = M_C \cdot 2^{P_c} = (0,111101011111)_2 \cdot 2^9 = 491,875.$$

Определим абсолютную и относительную погрешности результата:

$$\Delta C = 492 - 491,875 = 0,125$$

$$\delta C = \left| \frac{0,125}{492} \right| \cdot 100\% = 0,02541\%$$

Результат получился представленным с избытком. Этот факт можно объяснить потерей значащих младших у первого операнда при выравнивании порядков.

б) A>0, B<0:

$$M_A = .011100111001$$

-

$$M_B = .100000100110$$

$$M_C = \overline{.111100010011}$$

Результат вычитания денормализован вправо и представлен в дополнительном коде.

$$M_C = .000100110000$$

Т.к. выполнен сдвиг мантиссы влево, характеристику результата нужно уменьшить на 4 ($X_C = X_C - 4 = 5$).

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

$$C^* = M_C \cdot 2^{P_c} = (-0,11101101)_2 \cdot 2^5 = -29,625.$$

Определим абсолютную и относительную погрешности результата:
 $\Delta C = -29,6 - (-29,625) = 0,025$

$$\delta C = \left| \frac{0,025}{-29,6} \right| \cdot 100\% = 0,08446\%$$

Результат получился представленным с избытком. Этот факт можно объяснить потерей значащих разрядов мантиссы результата при его нормализации.

c) A<0, B>0:

$$M_B = .100000100110$$

$$M_A = -.011100111001$$

$$M_C = \overline{.000011101101}$$

Результат вычитания денормализован вправо.

$$M_C = .111011010000$$

Т.к. выполнен сдвиг мантиссы влево, характеристику результата нужно уменьшить на 4 ($X_C = X_C - 4 = 5$).

| | | | | | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 0 | 1 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | 1 | 1 | 0 | 1 | 1 | 0 | 1 | 0 | 0 | 0 | 0 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|

$$C^* = M_C \cdot 2^{P_c} = (0,11101101)_2 \cdot 2^5 = 29,625.$$

Определим абсолютную и относительную погрешности результата:
 $\Delta C = 29,6 - 29,625 = -0,025$

$$\delta C = \left| \frac{-0,025}{29,6} \right| \cdot 100\% = 0,08446\%$$

Результат получился представленным с избытком. Этот факт можно объяснить потерей значащих разрядов мантиссы результата при его нормализации.

В формате Ф2 результаты получились точнее из-за того, что операнды представлены точнее и при нормализации результата сдвиг производился на один двоичный разряд, а не на четыре.