Представление чисел в формате с плавающей запятой

Вещественные числа (конечные и бесконечные десятичные дроби) хранятся и обрабатываются в компьютере в формате *с плавающей запятой*. В этом случае положение запятой в записи числа может изменяться.
Формат чисел *с плавающей запятой* базируется на экспоненциальной форме записи, в которой может быть представлено любой число.

Представление двоичных чисел с плавающей запятой

Число А в форме с плавающей запятой представляется в виде

$$A = m_n \cdot q^p,$$

где m_n – нормализованная мантисса числа А;
P – порядок (характеристика) числа А;
q – основание системы счисления.

Мантисса m_n представляет собой правильную дробь, удовлетворяющую условию

$$q^{-1} \leq | m_n | < 1.$$

Числа А1 и А2 представлены следующим образом:

$$A_1 = m_1 \cdot q^{P1}; \quad A_2 = m_2 \cdot q^{P2}.$$

Арифметическое сложение или вычитание мантисс двух чисел может быть выполнено только в случае равенства их порядков.

Таким образом, в нормализованных числах первая цифра после точки должна быть значащей:

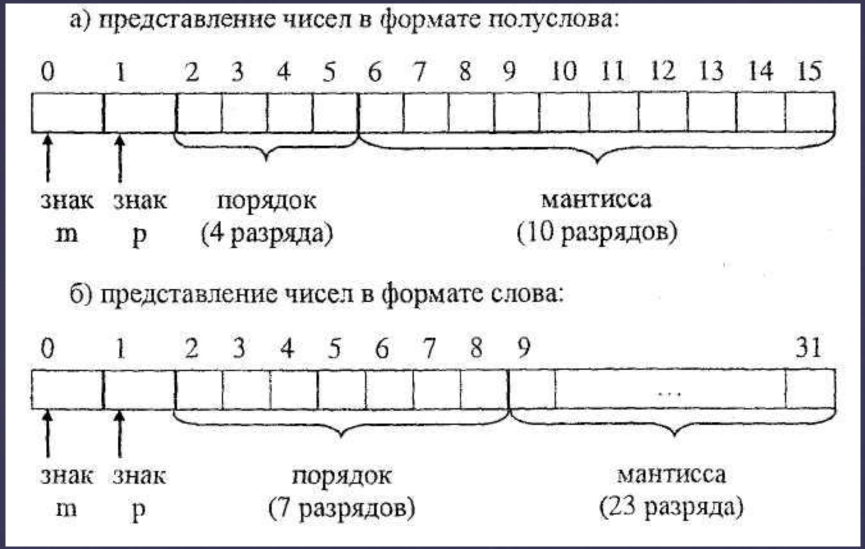
$$\underbrace{0.0832*10^3}_{\text{ненормализованное число}} = \underbrace{0.832*10^2}_{\text{нормализованное число}}$$

Пример. Преобразуйте десятичное число 888,888, записанное в естественной форме, в экспоненциальную форму с нормализованной мантиссой.

$888{,}888 = 0{,}888888 \times 10^3$

Нормализованная мантисса $m = 0{,}888888$, порядок $n = 3$.

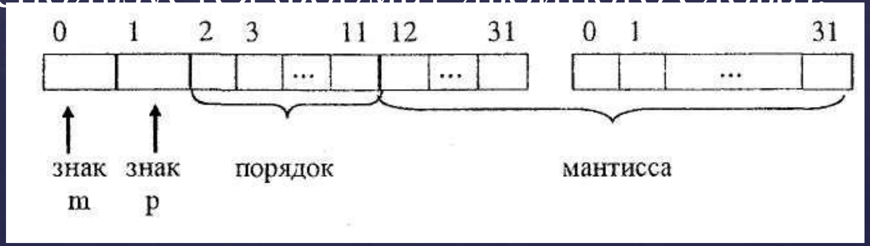
Для представления чисел в машинном слове выделяют группы разрядов для изображения мантиссы, порядка, знака числа и знака порядка:



с плавающей точкой в формате 32-разрядного слова будет иметь вид

Максимальным числом, представимым в формате слова, будет число

Числа с плавающей точкой позволяют увеличить диапазон обрабатываемых чисел, но при этом точность их изображения определяется только разрядами мантиссы и уменьшается по сравнению с числами с фиксированной точкой. При записи чисел в формате слова диапазон представимых чисел будет от - $1\text{-}2^{127}$ до 1-2 (2 =10), а точность их представления будет определяться мантиссой, состоящей из 23 разрядов. Точность может быть повышена путем увеличения количества разрядов мантиссы. Это реализуется путем представления чисел с так называемой двойной точностью (используется формат двойного слова):



Алгоритм сложения двух чисел с плавающей запятой:

1. Выравнивание порядков суммируемых чисел:

а) вычитание из порядка числа A_1 порядка числа A_2 с целью определения, порядок какого числа больше и на сколько:

$P_1 – P_2 = \Delta P$;

б) мантисса числа с меньшим порядком сдвигается вправо на величину разности порядков ΔP.

Если разность порядков положительная, то сдвигается мантисса второго числа, если отрицательная, то сдвигается мантисса первого числа. Обоим числам присваивается больший порядок.

2. Алгебраическое суммирование мантисс (аналогично суммированию дробных чисел с фиксированной запятой).

3. Результату присваивается больший порядок.

4. Проверка на нормализацию результата. При выполнении операции сложения нарушение нормализации суммы может быть влево только на один разряд ($|m_{ns}|<1$) или вправо на любое число разрядов ($q^{-1} \leq |m_{ns}|$).

Нормализация мантиссы

Мантисса считается нормализованной, если $2^{-1} \leq |M_x^{Н}| < 1$.

Положительная нормализованная мантисса в двоичном коде имеет вид:

$M_{x\ \ пр.обр.\ доп.}^{Н} = +\,0{,}1\ X_{-2}\ X_{-3}\ \ldots\ X_{-n} = M_{x\ \ обр.}^{Н} = M_{x\ \ доп.}^{Н}$

Отрицательная нормализованная мантисса:

$M_x^{Н} = -\,0{,}1\ X_{-2}\ X_{-3}\ \ldots\ X_{-n}$;

$M_{x\ \ пр.}^{Н} = 1{,}1\ X_{-2}\ X_{-3}\ \ldots\ X_{-n}$;

$M_{x\ \ обр.}^{Н} = 1{,}0\ \overline{X}_{-2}\overline{X}_{-3}\ \ldots\ \overline{X}_{-n}$;

$M_{x\ \ доп.}^{Н} = 1{,}0\ \overline{X}_{-2}\overline{X}_{-3}\ \ldots\ \overline{X}_{-n} + 2^{-n}$.

Признак нормализованной мантиссы в обратном и дополнительном кодах – различие цифр, стоящих по обе стороны от запятой;

в прямом коде – наличие единицы в старшем значащем разряде.

Нарушение нормализации мантиссы вправо

Признак нарушения нормализации вправо для дополнительных и обратных кодов является совпадение значения знакового разряда со значением старшего значащего разряда мантиссы.

Для устранения этого нарушения мантисса сдвигается влево до тех пор, пока не будут выполняться условия нормализации. Порядок P_x при этом уменьшается на число выполненных сдвигов.

В обратном коде в освобождающиеся младшие разряды при сдвиге для положительного числа записываются «0», а для отрицательного – «1».

В дополнительном коде в освобождающиеся разряды записываются «0» независимо от знака числа.

Примеры.

Для прямого, обратного и дополнительного кодов:

$M_x = 00,001011;$ $P_x = 00,101.$

$M_x^н = 00,101100;$ $P_x = 00,011.$

Для дополнительного кода:

$M_{x\,доп.} = 11,1101011;$ $P_x = 11,001.$

$M_x{}^{Н}{}_{доп.} = 11,0101100;$ $P_x = 11,011.$

Для обратного кода:

$\mathbf{M_{x\,обр.} = 11,1101010;}$ $\mathbf{P_x = 11,001.}$

$\mathbf{M_x{}^{Н}{}_{обр.} = 11,0101011}$ $\mathbf{P_x = 11,011.}$

Нарушение нормализации мантиссы влево

Признак нарушения нормализации влево для дополнительных и обратных кодов – это сочетание 01 или 10 в знаковых разрядах модифицированных кодов.

Устранение этого нарушения состоит в модифицированном сдвиге мантиссы M_x вправо на 1 разряд и увеличении порядка P_x на единицу.

Примеры для доп. и обр. кодов:

$M_x = 01{,}110111$, $P_x = 00{,}101$

После нормализации:

$M_x = 00{,}111011$ ①*, $P_x = 00{,}110$

$M_x = 10{,}10011$, $P_x = 00{,}011$

После нормализации:

$M_x = 11{,}01001$ ①*, $P_x = 00{,}100$

* – разряд, вышедший за пределы разрядной сетки, используется для округления или отбрасывается.

Пример.

1. Выполнить сложение чисел

$$X=+0.101\bullet10^{+101}$$

$$Y= -0.1011\bullet10^{+100}$$

(количество разрядов порядка Np = 4, мантиссы Nq = 8).

$$p_x - p_y=p_x + (-p_y)$$

а) Выполним выравнивание порядков чисел, для чего вычислим значение разности

p_y = 1100 - прямой код порядка числа Y;

p_y = 1011 - обратный код порядка числа Y;

p_y = 1100 - дополнительный код порядка числа Y.

Находим разность порядков:

$$p_x - p_y=(0101 + 1100)_{доп} = (0001)_{пр}$$

Так как $p_x>p_y$, необходимо сдвинуть мантиссу числа Y на один разряд вправо:

$$Y_{пр} = 1.1011000.$$

Сдвигаем:

$$Y_{пр} = 1.0101100.$$

б) Выполним сложение (вычитание) мантисс:

$$Y_{доп} = 11010011_{обр} + 00000001 = 11010100_{доп}$$

Суммируем мантиссы:

$q_z = (q_x + q_y) = (01010000_{пр} + 11010100_{доп}) = 00100100_{пр}$.

в) Порядок результата принимаем равным порядку числа X, т.е.

$$p_z = +101.$$

г) Полученную сумму нормализуем, так как

$$|q_z| < \frac{1}{S}.$$

Для нормализации необходимо p_z уменьшить на 1, а q_z сдвинуть на один разряд влево:

$$Z = 0.1001\bullet10^{+100}.$$