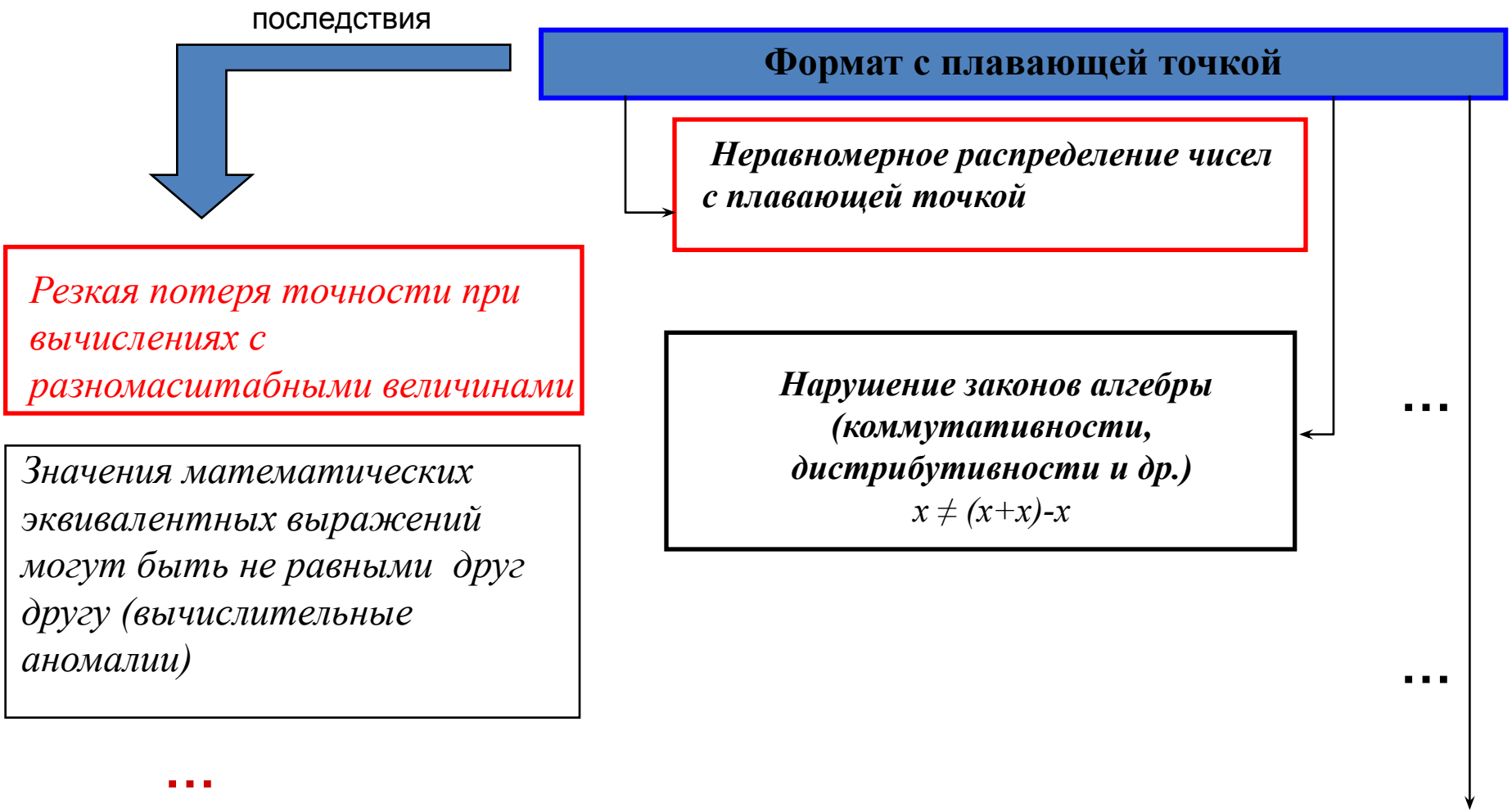
Лекция №6
по курсу
«Машинная арифметика в рациональных чисел»

Москва, 2020

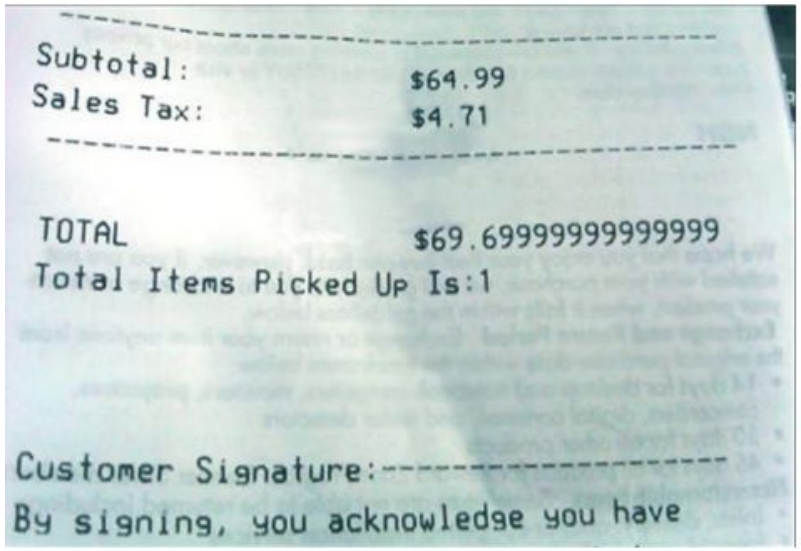
Особенности формата с плавающей точкой



Недостатки формата с плавающей точкой

1. Числа с плавающей точки дают различные результаты на различных аппаратных платформах.
2. Сложность использования численных методов (требуются экспертные знания в области Error Analyze)
3. Резкий рост времени вычислений при увеличении точности
4. В формате с плавающей точкой скрыты ошибки переполнения, исчезновения порядка (на флаги процессора никто не смотрит)

Пример ошибки при сложении чисел в формате с плавающей точкой:



Нарушение законов алгебры

- No associative property for floats
- $(a + b) + (c + d)$ (parallel) $\neq ((a + b) + c) + d$ (serial)
- Looks like a “wrong answer”

Считаете с одной точностью и с другой если результаты примерно одинаковые, то вероятно задача решена правильно.

ПРИМЕР ЗАДАЧИ, ИМЕЮЩЕЙ РЕЗКИЙ РОСТ ОШИБОК ОКРУГЛЕНИЯ

Матрица Гильберта $A = \{a_{ij}\},\ a_{ij} = \dfrac{1}{i + j - 1}$

Обращение матрицы Гильберта порядка 3

$$A = \begin{bmatrix} 1 & 1/2 & 1/3 \\ 1/2 & 1/3 & 1/4 \\ 1/3 & 1/4 & 1/5 \end{bmatrix}$$

С точностью 2 знака после запятой

$$A_{прибл}^{-1} = \begin{bmatrix} -1{,}17 & 19{,}51 & -23 \\ 19{,}51 & -112{,}94 & 112 \\ -23 & 112 & -100 \end{bmatrix}$$

Макс. относ. погрешн. более 100%.

С точностью 3 знака после запятой

$$A_{прибл}^{-1} = \begin{bmatrix} 10{,}101 & 29{,}598 & 64{,}798 \\ -41{,}039 & -192{,}78 & -202{,}4 \\ 34{,}6 & 202{,}4 & 200 \end{bmatrix}$$

Макс. относ. погрешность более 100%.

Точный результат:

$$A_{точн}^{-1} = \begin{bmatrix} 9 & -36 & 30 \\ -36 & 192 & -180 \\ 30 & -180 & 180 \end{bmatrix}$$

Матрица Гильберта

Для матрицы 12 уже неверные результаты.
Число обусловленности растёт экспоненциально!

Некоторые матрицы имеют число обусловленности пропорционально n или n^2

Модулярная арифметика

1931 – 2014

Амербаев В.М.

МОДУЛЯ́РНАЯ АРИФМЕ́ТИКА (система остаточных классов), базируется на известном в теории чисел частном виде отношения эквивалентности – понятии сравнения целого числа по натуральному модулю и возникающему при этой операции вычету. Для задания модулярной арифметики необходимо задать математические соотношения, позволяющие взаимнооднозначно отображать значение числовой величины в компоненты векторного представления и обратно, а также алгоритмы выполнения машинных операций над компонентами векторного модулярного представления.

Деление с остатком

Теорема 1.1. *Для любых целого числа a и натурального b существуют единственным образом определенные целые числа q, r, удовлетворяющие условиям*

$$a = bq + r, \qquad 0 \le r < b. \tag{1.1}$$

Доказательство. Докажем существование чисел q, r.

Если $b|a$, т. е. $b = au$ с некоторым целым u, положим $q = u$, $r = 0$. При таком выборе условия (1.1), очевидно, выполнены.

Далее будем предполагать, что число a не делится на b. Множество M натуральных чисел, представимых в виде $a - bk$ с некоторым целым k, не пусто.

Действительно, при $k = -|a| - 1$ имеем

$$a - kb = a + (|a| + 1)b \ge a + (|a| + 1) \ge 1.$$

Деление с остатком

Обозначим буквой r наименьшее из чисел множества M. Тогда $r \geq 1$, и с некоторым целым q выполняется равенство $r = a - bq$. Поскольку $a - b(q+1) = r - b < r$, то $a - b(q+1) \notin M$, и, значит, выполнено $a - b(q+1) = r - b \leq 0$. Учитывая, что равенство $a - b(q+1) = 0$ невозможно, ведь $b \nmid a$, заключаем $r - b < 0$. Итак, $1 \leq r < b$. Существование нужных чисел q, r доказано и в этом случае.

Допустим теперь, что вместе с парой чисел (q, r), удовлетворяющих условиям (1.1), существует еще одна, отличная от нее пара чисел (u, v), для которой

$$a = bu + v, \qquad 0 \leq v < b. \tag{1.2}$$

Из равенств (1.1) и (1.2) следует $bq + r = bu + v$. Поэтому $r - v = b(u - q)$, так что $b|(r - v)$. Из неравенств (1.1) и (1.2) находим также $|r - v| < b$. Согласно последнему свойству леммы 1.1 можно утверждать теперь, что $r - v = 0$, т. е. $v = r$. Но тогда $b(u - q) = 0$ и, значит, $u = q$. Получаем, что пары (q, r) и (u, v) должны совпадать вопреки нашему предположению. Это противоречие завершает доказательство теоремы 1.1. ■

Лемма 1.2. *1. Если a — целое, b — натуральное и $b|a$, то множество общих делителей чисел a и b совпадает с множеством делителей b. В частности, $(a, b) = b$.*

2. Если

$$a = bq + r, \qquad 0 \leq r < b,$$

то $\mathcal{D}\{a, b\} = \mathcal{D}\{b, r\}$, в частности, $(a, b) = (b, r)$.

Доказательство. Множество общих делителей чисел a и b содержится среди делителей b, т. е. $\mathcal{D}\{a,b\} \subset \mathcal{D}\{b\}$. В то же время каждый делитель числа b, в силу условия $b|a$, будет делителем числа a, т. е. принадлежит $\mathcal{D}\{a,b\}$. Следовательно, множество общих делителей чисел a и b совпадает с множеством делителей b. Это доказывает первое утверждение леммы 1.2.

Для доказательства второго утверждения заметим, что по свойствам делимости каждый общий делитель чисел b и r делит число $bq + r = a$ и, значит, принадлежит множеству $\mathcal{D}\{a,b\}$. Точно так же каждый общий делитель чисел a и b делит число $a - bq = r$, так что принадлежит множеству $\mathcal{D}\{b,r\}$. Этим доказано равенство $\mathcal{D}\{a,b\} = \mathcal{D}\{b,r\}$. Отсюда следует и совпадение наибольших общих делителей пар чисел a, b и b, r. ■

Алгоритм Евклида

Рассмотрим алгоритм подробнее. Положим $r_0 = a$, $r_1 = b$ и обозначим $r_2, \ldots, r_n$ — последующие делители в алгоритме Евклида. Таким образом, получаются следующие равенства

$$\begin{aligned}
a &= r_0 = bq_1 + r_2, & 0 \le r_2 < b,\\
b &= r_1 = r_2q_2 + r_3, & 0 \le r_3 < r_2,\\
r_2 &= r_3q_3 + r_4, & 0 \le r_4 < r_3,\\
&\ldots\ldots\ldots\ldots & \ldots\ldots\ldots\ldots\\
r_{n-2} &= r_{n-1}q_{n-1} + r_n, & 0 \le r_n < r_{n-1},\\
r_{n-1} &= r_nq_n.
\end{aligned} \tag{1.3}$$

Алгоритм останавливается, когда деление произойдет без остатка (последний остаток $r_{n+1} = 0$). В соответствии с леммой 1.2 находим равенства

$$(a, b) = (b, r_2) = (r_2, r_3) = (r_3, r_4) = \cdots = (r_{n-1}, r_n) = (r_n, 0) = r_n.$$

Таким образом, наибольший общий делитель равен последнему делителю (он же последний ненулевой остаток) в алгоритме Евклида.

Теорема 1.4. *Количество делений, необходимое для вычисления с помощью алгоритма Евклида наибольший общий делитель двух положительных целых чисел, не превышает пятикратного количества цифр в десятичной записи меньшего из этих двух чисел.*

Доказательство. Пусть $a \geq b$ — натуральные числа и b записывается m цифрами в десятичной системе счисления. Это значит, что $10^{m-1} \leq b < 10^m$. Требуется доказать, что алгоритм Евклида, примененный к числам a, b, выполнит не более $5m$ делений с остатком.

Положим $r_0 = a$ и обозначим $r_1 = b, r_2, \ldots, r_n$ — последовательность делителей в алгоритме Евклида. Тогда

$$r_{i-1} = q_i r_i + r_{i+1}, \quad 0 \leq r_{i+1} < r_i, \quad i = 1, 2, \ldots, n-1, \quad r_n = (a, b).$$

Докажем справедливость неравенств

$$r_{n-i} \geq \lambda^i, \qquad i = 0, 1, \ldots, n-1, \tag{1.4}$$

где $\lambda = \dfrac{1+\sqrt{5}}{2} = 1,61...$ есть корень квадратного уравнения $\lambda^2 - \lambda - 1 = 0$. Для этого воспользуемся методом математической индукции. При $i = 0$ имеем $r_n \geq 1 = \lambda^0$. При $i = 1$ находим $r_{n-1} \geq r_n + 1 \geq 2 > \lambda$, так что неравенство (1.4) опять справедливо. Предположим теперь, что $1 \leq k < n - 1$ и неравенство (1.4) выполняется при всех $i \leq k$. Имеем следующую цепочку равенств и неравенств

$$r_{n-k-1} = q_{n-k} r_{n-k} + r_{n-k+1} \geq r_{n-k} + r_{n-k+1} \geq \lambda^k + \lambda^{k-1} = \lambda^{k+1}.$$

Таким образом, неравенство (1.4) выполняется и при $i = k+1$. Это доказывает справедливость (1.4) при всех $i = 0, 1, \ldots, n-1$.

Поскольку $10^m > b$ и $b = r_1 \geq \lambda^{n-1}$, видим, что

$$m > (n-1)\lg\lambda > (n-1)/5.$$

В последнем неравенстве была использована оценка $\lambda > 10^{1/5} = 1,58...$ Таким образом, $n < 5m + 1$. ■

ПРИНЦИПЫ РЕАЛИЗАЦИИ МОДУЛЯРНЫХ ВЫЧИСЛЕНИЙ

Пусть m – некоторое простое число, называемое ***модулем***, тогда по теореме о делении с остатком любое целое число a можно представить в виде:

$$a = m \cdot q + r ,$$

где q – частное, r – остаток от деления a на m.

Если два целых числа при делении на m дают один и тот же остаток, то они называются ***равноостаточными*** или ***сравнимыми по модулю*** m. Например, числа 13 и 8 сравнимы по модулю 5, т.к. 13=2*5+3 и 8=1*5+3, или числа 10 и 3 сравнимы по модулю 7, т.к. 10=1*7+3 и 3=0*7+3.

Математически свойство сравнимости по модулю m двух чисел a и b записывается так:

$$a \equiv r(\bmod m) \quad \text{или} \quad |a|_m \equiv r \qquad (1)$$

Пример. $13 \equiv 8(\bmod 5)$, $10 \equiv 3(\bmod 7)$.

ПРИНЦИПЫ РЕАЛИЗАЦИИ МОДУЛЯРНЫХ ВЫЧИСЛЕНИЙ

Свойства сравнений.

1°. ***Рефлексивность***.

Если $a = b$, то $a \equiv b (\mathrm{mod}\ m)$. А если $a < m$ и $b < m$, то из $a \equiv b (\mathrm{mod}\ m)$ следует, что $a = b$.

2°. Если $a \equiv r (\mathrm{mod}\ m)$, то и $a \pm k \cdot m \equiv r (\mathrm{mod}\ m)$.

Доказательство:

$a = q_1 \cdot m + r, \quad r_1 < m$

$a \pm k \cdot m = (q_1 + k) \cdot m + r$

3°. ***Сравнения можно почленно складывать***

Из $a_1 \equiv r_1 (\mathrm{mod}\ m)$ и $a_2 \equiv r_2 (\mathrm{mod}\ m)$ следует, что $a_1 + a_2 \equiv r_1 + r_2 (\mathrm{mod}\ m)$.

Доказательство:

$a_1 = q_1 \cdot m + r_1, \quad r_1 < m$

$a_2 = q_2 \cdot m + r_2, \quad r_2 < m$

$a_1 + a_2 = (q_1 + q_2) \cdot m + r_1 + r_2$

Если $r_1 + r_2 < m$, то $a_1 + a_2 = (q_1 + q_2) \cdot m + r_1 + r_2$, отсюда следует, что

$a_1 + a_2 \equiv r_1 + r_2 (\mathrm{mod}\ m)$,

Если $r_1 + r_2 \geq m$, то $a_1 + a_2 = (q_1 + q_2 + 1) \cdot m + (r_1 + r_2 - m)$

$a_1 + a_2 \equiv (r_1 + r_2 - m) (\mathrm{mod}\ m) \equiv (r_1 + r_2) (\mathrm{mod}\ m)$

ПРИНЦИПЫ РЕАЛИЗАЦИИ МОДУЛЯРНЫХ ВЫЧИСЛЕНИЙ

4°. ***Сравнения можно почленно перемножать, т.е.***

Из $a_1 \equiv b_1 (\bmod m)$ и $a_2 \equiv b_2 (\bmod m)$ следует, что $a_1 \cdot a_2 \equiv b_1 \cdot b_2 (\bmod m)$.

Из свойства 3° следует, что сравнения можно почленно возводить в степень, т.е.

Из $a \equiv b (\bmod m)$, следует, что $a^n \equiv b^n (\bmod m)$, где n – целое число.

Для представления ***отрицательных чисел*** используется формула:

$-a \equiv m - a (\bmod m)$,

Например,

$-5 \equiv 8 (\bmod 13)$,

$-15 \equiv -2 (\bmod 13) \equiv 13\text{-}2 (\bmod 13) \equiv 11 (\bmod 13)$,

т.к. $-15 = (-1)*13 - 2$

Рассмотрим ***деление сравнений***.

$$\frac{a}{b} (\bmod m) \equiv a \cdot (b)^{-1} (\bmod m),$$

где b^{-1} – число, ***обратное*** к b по модулю m, т.е. число x удовлетворяющее сравнению $x \cdot b (\bmod m) \equiv 1$.

Модулярная арифметика с дробями

Обратное число к a это такое число x, для которого выполняется соотношение:

$a \cdot x \equiv 1 (\mathrm{mod}\, m)$.

Пусть модуль $m = 7$, $a = 2$

$2 \cdot x \equiv 1 (\mathrm{mod}\, 7)$

$x = 4,\ 2 \cdot 4 \equiv 1 (\mathrm{mod}\, 7)$

Т.е. $\frac{1}{2} \mathrm{mod}\ 7 \equiv 4,$

Таблица соответствия по модулю $m = 7$.

Дробь	***Модулярное представление***
1/2	4
1/3	5
1/6	6
43/6, 1/6, 6 … ? (нет единственности)	**6**

Вычисления с дробями Фарея в модулярной арифметике

Дроби Фарея – это несократимые дроби, числители и знаменатели которых по абсолютной величине меньше чем некоторая константа, называемая порядком дробей Фарея, т.е.

$$F_N = \left(\frac{a}{b} \in Q, \quad |a| \le N, |b| \le N \right),$$

где F_N – множество дробей Фарея,

Q – множество рациональных чисел,

N – порядок дробей Фарея.

Например, дроби Фарея третьего порядка:

$$\frac{0}{1}, 1, 2, 3, -1, -2, -3, \frac{1}{2}, \frac{1}{3}, \frac{2}{3}, \frac{3}{2}, -\frac{1}{2}, -\frac{1}{3}, -\frac{2}{3}, -\frac{3}{2}.$$

Справедлива теорема о том, что если

$$2 \cdot N^2 + 1 \le m,$$

где m – модуль, то каждая дробь Фарея порядка N имеет единственное модулярное представление.