Замена целочисленного деления на умножение и сдвиг (теория)

Материал к видеолекции «Беседы о программировании 002»
Караваев Артём Михайлович, 24.12.2015
Текст является неотъемлемой частью видеозаписи
http://zealcomputing.ru

Интуитивные соображения

Пример для десятичной системы счисления.

$$q = \left\lfloor \frac{123}{7} \right\rfloor = 17.$$

Вместо деления умножаем на обратный элемент $1/7 \approx 0{,}14$:

$$q = \lfloor 123 \cdot 0{,}14 \rfloor = \lfloor 17{,}22 \rfloor = 17.$$

Вместо обратного элемента возьмём аппроксимацию в виде дроби

$$\frac{1}{7} \approx \frac{14}{10^2} \quad \Rightarrow \quad q = \left\lfloor 123 \cdot \frac{14}{10^2} \right\rfloor = \left\lfloor \frac{1722}{10^2} \right\rfloor = 17.$$

Всегда можно подобрать аппроксимацию вида

$$\frac{1}{d} \approx \frac{\nu}{10^m},$$

вопрос лишь в точности, определяемой величиной m.

Например,

$$q = \left\lfloor \frac{1234}{7} \right\rfloor = 176 \neq \left\lfloor 1234 \cdot \frac{14}{10^2} \right\rfloor = \left\lfloor \frac{17276}{10^2} \right\rfloor = 172.$$

точность недостаточна, однако для $m = 3$ имеем

$$\frac{1}{7} \approx \frac{143}{10^3} \quad \Rightarrow \quad q = \left\lfloor 1234 \cdot \frac{143}{10^3} \right\rfloor = \left\lfloor \frac{176462}{10^3} \right\rfloor = 176,$$

что верно.

Аналогично, в двоичной системе счисления для целого $d > 0$ должна существовать дробь

$$\frac{1}{d} \approx \frac{v}{2^m},$$

с помощью которой деление можно заменить
на умножение и сдвиг:

$$\left\lfloor \frac{a}{d} \right\rfloor = \left\lfloor \frac{av}{2^m} \right\rfloor.$$

Например, для чисел a, умещающихся в 32 бита без знака

$$\left\lfloor \frac{a}{3} \right\rfloor = \left\lfloor \frac{2\,863\,311\,531 \cdot a}{2^{33}} \right\rfloor = \left\lfloor \frac{\mathtt{AAAAAAAB}_{(16)} \cdot a}{2^{33}} \right\rfloor.$$

Попробуем угадать ν

Договоримся, что $a \geqslant 0$ — числитель, $d > 0$ — знаменатель. Оба целые, и

$$a = qd + r, \quad \text{где } q = \left\lfloor \frac{a}{d} \right\rfloor \text{ и } 0 \leqslant r < d.$$

Внимание! Дальше идёт не строгое математическое рассуждение, а попытка «прикинуть» область поиска значения ν.

Нам нужно, чтобы

$$q = \left\lfloor \frac{a}{d} \right\rfloor = \left\lfloor \frac{a\nu}{2^m} \right\rfloor.$$

Это возможно, если

$$q \leqslant \frac{a\nu}{2^m} < q + 1.$$

$$\frac{q \cdot 2^m}{a} \leqslant \nu < \frac{(q+1) \cdot 2^m}{a}.$$

$$\frac{(a-r) \cdot 2^m}{ad} \leqslant \nu < \frac{(a-r+d) \cdot 2^m}{ad}.$$

$$\frac{2^m}{d} - \frac{r \cdot 2^m}{ad} \leqslant \nu < \frac{2^m}{d} - \frac{r \cdot 2^m}{ad} + \frac{2^m}{a}.$$

В частности, это должно быть верно при $r = 0$:

$$\frac{2^m}{d} \leqslant \nu < \frac{2^m}{d} + \frac{2^m}{a}.$$

$$\left\lceil \frac{2^m}{d} \right\rceil \leqslant \nu < \frac{2^m}{d} + \frac{2^m}{a}.$$

Таким образом, можно взять

$$v = \left\lceil \frac{2^m}{d} \right\rceil$$

С другой стороны, можно было рассуждать иначе, нам нужно

$$\frac{1}{d} \approx \frac{v}{2^m},$$

откуда

$$v \approx \frac{2^m}{d}.$$

Таким образом, допустим, что

$$v = \left\lceil \frac{2^m}{d} \right\rceil = \left\lfloor \frac{2^m + d - 1}{d} \right\rfloor = \left\lfloor \frac{2^m - 1}{d} \right\rfloor + 1.$$

Строгое доказательство

Обозначим
$\mathbb{Z}_{\geqslant 0}$ — множество целых неотрицательных чисел и
$\mathbb{Z}_{>0}$ — множество целых положительных чисел.
$\mathbb{Z}_{>1}$ — множество целых чисел, больших единицы.

Теорема

$\forall\, A \in \mathbb{Z}_{\geqslant 0},\ d \in \mathbb{Z}_{>1}, d$ нечётное $\exists\, m \in \mathbb{Z}_{\geqslant 0}$: при $\nu = \lceil 2^m/d \rceil$ выполняется

$$\left\lfloor \frac{a}{d} \right\rfloor = \left\lfloor \frac{a\nu}{2^m} \right\rfloor, \quad \forall\, 0 \leqslant a \leqslant 2^A - 1.$$

Доказательство

Мы хотим, чтобы выполнялось равенство

$$\left\lfloor \frac{a}{d} \right\rfloor = \left\lfloor \frac{a\nu}{2^m} \right\rfloor.$$

Насколько близки друг к другу числа $\frac{a}{d}$ и $\frac{a\nu}{2^m}$?

$$\frac{a\nu}{2^m} - \frac{a}{d} = ?$$

$$\nu = \left\lfloor \frac{2^m - 1}{d} \right\rfloor + 1 = \frac{(2^m - 1) - (2^m - 1) \bmod d}{d} + 1$$

$$\frac{a\nu}{2^m} - \frac{a}{d} = \frac{ad\left(\dfrac{(2^m-1)-(2^m-1) \bmod d}{d} + 1\right) - a \cdot 2^m}{d \cdot 2^m} =$$

$$= \frac{a}{d \cdot 2^m}\Big((2^m-1)-(2^m-1) \bmod d + d - 2^m\Big) =$$

$$= \frac{a}{d \cdot 2^m}\Big((d-1)-(2^m-1) \bmod d\Big) =$$

$$= \frac{a}{d \cdot 2^m}\Big(((d-1)-(2^m-1)) \bmod d\Big) =$$

$$= \frac{a}{d \cdot 2^m}\Big((-2^m) \bmod d\Big).$$

Обозначим $k \stackrel{\text{def}}{=} (-2^m) \bmod d$. (Вот почему $d > 1$ и нечётное!).

$$\boxed{\sigma \stackrel{\text{def}}{=} \frac{a\nu}{2^m} - \frac{a}{d} = \frac{k \cdot a}{d \cdot 2^m}.}$$

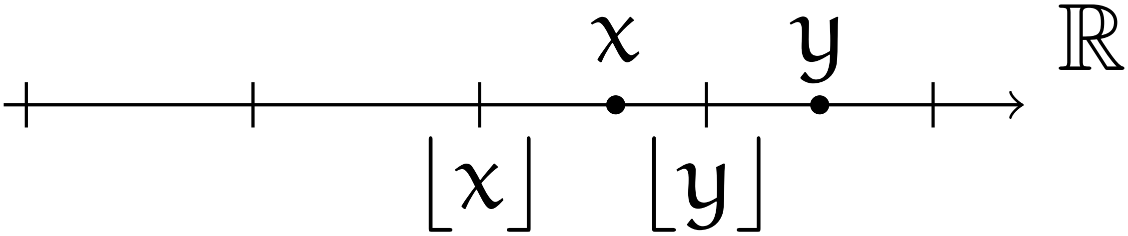
Если увеличивать m, можно сделать разницу сколь угодно близкой к нулю.

В частности, мы выяснили следующее:

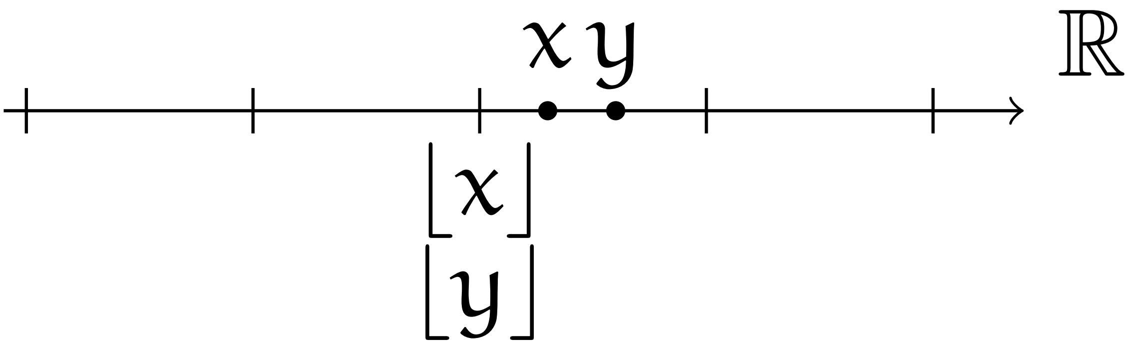
- $\dfrac{a\nu}{2^m} \geqslant \dfrac{a}{d}$.
- Если $m \geqslant A$, то $\sigma < 1$ и $\lim\limits_{m\to\infty} \sigma = 0$.

Но сколь угодно высокая степень близости двух величин, скажем x и y, не означает, что их целые части будут равны.

Так может быть:



А так надо, чтобы было:



Правило следующее: $y - \lfloor y \rfloor \geqslant y - x$. То есть:

$$\frac{av}{2^m} - \left\lfloor \frac{av}{2^m} \right\rfloor \geqslant \sigma.$$

$$\left\{ \frac{av}{2^m} \right\} \geqslant \sigma.$$

$$\frac{(av) \bmod 2^m}{2^m} \geqslant \frac{k \cdot a}{d \cdot 2^m}.$$

$$(av) \bmod 2^m \geqslant \frac{k \cdot a}{d}.$$

Во-первых, отмемим, что

$$v \bmod 2^m = \left(\left\lfloor \frac{2^m - 1}{d} \right\rfloor + 1 \right) \bmod 2^m =$$

$$= \left(\frac{(2^m - 1) - (2^m - 1) \bmod d}{d} + 1 \right) \bmod 2^m =$$

$$= \left(\frac{(d - 1) - (2^m - 1) \bmod d}{d} \right) \bmod 2^m = \frac{k}{d} \bmod 2^m.$$

Следовательно, когда $a \bmod d = 0$, или $a = qd$ имеем

$$(av) \bmod 2^m \geqslant \frac{ak}{d} \quad \Leftrightarrow \quad (qk) \bmod 2^m \geqslant qk.$$

Это выполнено, когда $qk < 2^m$, то есть всегда, т. к. $m \geqslant A$. Более того, имеем строгое равенство $(qk) \bmod 2^m = qk$.

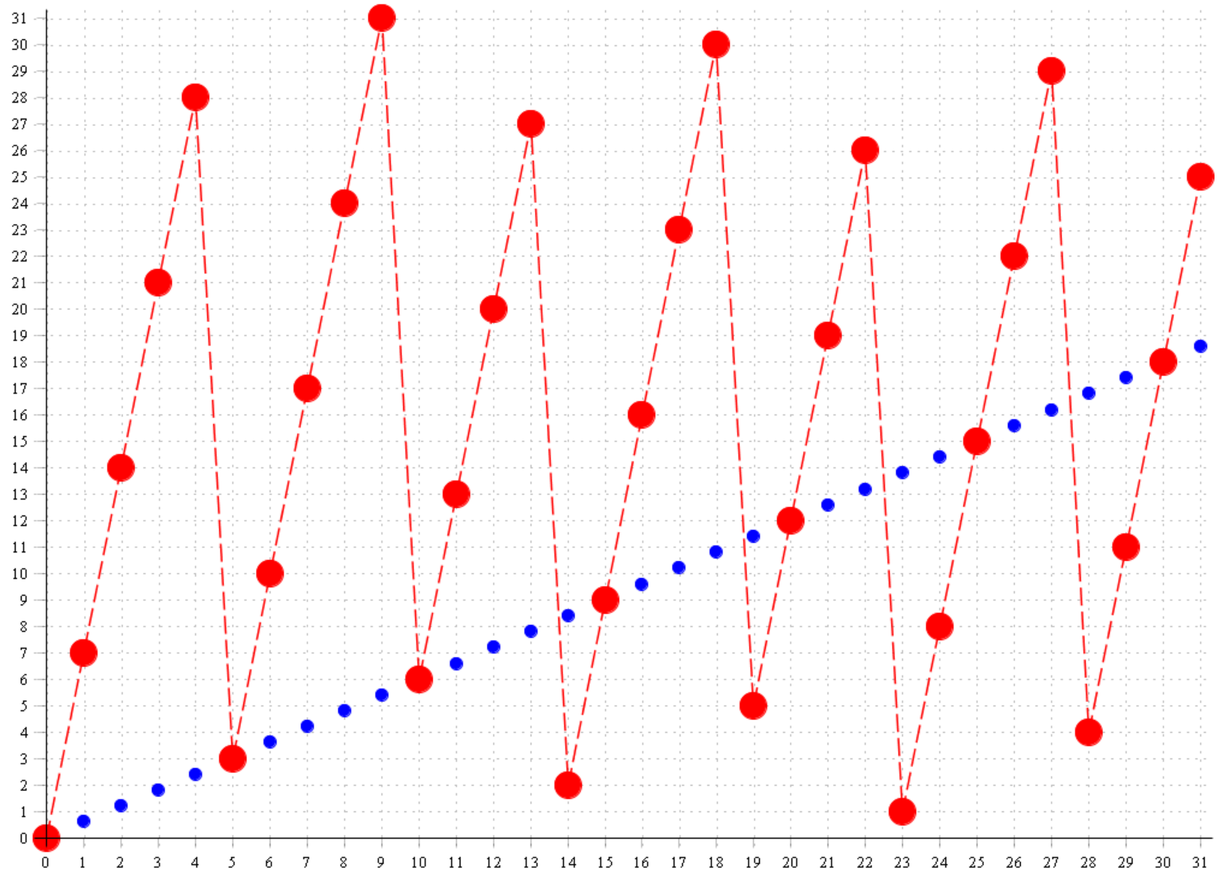
Во-вторых, ответим на вопрос, при каких ещё a

$$(av) \bmod 2^m \geqslant \frac{ak}{d}\,?$$

Рассмотрим для примера $d = 5$, $A = 5$ $(0 \leqslant a \leqslant 31)$. Предположим, что $m = 5$ и $v = \lceil 32/5 \rceil = 7$, $k = (-32) \bmod 5 = 3$.

Рассмотрим неравенство

$$(7a) \bmod 32 \geqslant \frac{3a}{5}.$$



«Редукция» av с номером s по модулю 2^m происходит для чисел a, определяемых из неравенства

$$av \geqslant s \cdot 2^m, \quad s \in \mathbb{Z}_{>0}.$$

$$a \geqslant \frac{s \cdot 2^m}{v} \quad \Leftrightarrow \quad a \geqslant \left\lceil \frac{s \cdot 2^m}{v} \right\rceil.$$

$$\left\lceil \frac{s \cdot 2^m}{v} \right\rceil = \left\lceil \frac{s \cdot 2^m}{\lceil 2^m/d \rceil} \right\rceil =$$

Пояснение

$$\left\lceil \frac{x}{y} \right\rceil = \frac{x + (-x) \bmod y}{y}$$

$$= \left\lceil \frac{sd \cdot 2^m}{2^m + k} \right\rceil = \left\lceil \frac{sd\left(2^m + k\right)}{2^m + k} - \frac{sdk}{2^m + k} \right\rceil =$$
$$= \left\lceil sd - \frac{sdk}{2^m + k} \right\rceil = sd - \left\lfloor \frac{sdk}{2^m + k} \right\rfloor.$$

Таким образом, s-й момент убывания функции $(a\nu) \bmod 2^m$ возникает в точках

$$a = sd - \left\lfloor \frac{sdk}{2^m + k} \right\rfloor.$$

Если

$$\left\lfloor \frac{sdk}{2^m + k} \right\rfloor = 0,$$

то точки убываний совпадают с моментами, когда $a \bmod d = 0$, а мы доказали, что в этих точках

$$(a\nu) \bmod 2^m = \frac{k \cdot a}{d}, \quad m \geqslant A.$$

Если же

$$\left\lfloor \frac{sdk}{2^m + k} \right\rfloor \geqslant 1,$$

то момент убываний предшествует моменту, в котором $a \bmod d = 0$, то есть нарушается условие

$$(a\nu) \bmod 2^m \geqslant \frac{k \cdot a}{d}, \quad m \geqslant A.$$

Когда же это происходит?

$$\left\lfloor \frac{sdk}{2^m + k} \right\rfloor \geqslant 1.$$

$$sdk \geqslant 2^m + k.$$

$$s \geqslant \frac{2^m + k}{dk}.$$

$$s \geqslant \left\lceil \frac{2^m + k}{dk} \right\rceil.$$

$$\boxed{s = \left\lceil \frac{2^m + k}{dk} \right\rceil.}$$

И максимальное значение a, при котором ЕЩЁ НЕ произойдёт нарушение условия теоремы

$$\boxed{a_{max} = sd - \left\lfloor \frac{sdk}{2^m + k} \right\rfloor - 1.}$$

Упражнение. Покажите, что найдётся $m \geqslant A$, гарантированно подходящее под условие теоремы.

Возможно, что $m \leqslant A + D$, *где* D *— число бит в* d.

Таким образом, среди чисел из списка $[A, A + 1, \ldots]$ найдется такое m, для которого теорема верна, то есть будет выполнено условие

$$a_{max} \geqslant 2^A - 1.$$

□

Итоговые формулы

Нужно вычислить $\lfloor a/d \rfloor$,
при этом $0 \leqslant a < 2^A$, $A \in \mathbb{N}$ и $d > 0$ нечётное целое.

Важно! Если d чётное, то нужно выбрать максимальную степень двойки, на которую делится d, то есть $d = d' \cdot 2^n$ выполнить деление на d' с последующим сдвигом на n бит вправо, который можно совместить со сдвигом на m бит после умножения на ν.

Перебираем m от A до ∞, и выбираем первое, которое подходит под условие ниже.

$$k = (-2^m) \bmod d.$$

$$s = \left\lceil \frac{2^m + k}{dk} \right\rceil.$$

$$a_{max} = sd - \left\lfloor \frac{sdk}{2^m + k} \right\rfloor - 1.$$

$$? \quad a_{max} \geqslant 2^A - 1 \qquad \leftarrow \text{Вот это условие!}$$

$$\nu = \left\lceil \frac{2^m}{d} \right\rceil = \left\lfloor \frac{2^m - 1}{d} \right\rfloor + 1.$$

Тогда

$$\left\lfloor \frac{a}{d} \right\rfloor = \left\lfloor \frac{a\nu}{2^m} \right\rfloor.$$

Пример 1

$d = 3$, $A = 4$, то есть $0 \leqslant a < 16$.
Организуем перебор m от 4.
Проверяем $m = 4$:

$$k = (-2^m) \bmod d = (-16) \bmod 3 = 2.$$

$$s = \left\lceil \frac{2^m + k}{dk} \right\rceil = \left\lceil \frac{16 + 2}{3 \cdot 2} \right\rceil = 3.$$

$$a_{\max} = sd - \left\lfloor \frac{sdk}{2^m + k} \right\rfloor - 1 = 3 \cdot 3 - \left\lfloor \frac{3 \cdot 3 \cdot 2}{16 + 2} \right\rfloor - 1 = 7.$$

$$a_{\max} = 7 \ngeqslant 15, \quad \text{не подходит.}$$

Действительно, если взять

$$v = \left\lceil \frac{16}{3} \right\rceil = 6,$$

получим, что для $a = 8$: $\lfloor 8/3 \rfloor = 2 \neq 3 = \lfloor 6 \cdot 8/16 \rfloor$.
Проверяем $m = 5$:

$$k = (-32) \bmod 3 = 1.$$

$$s = \left\lceil \frac{32 + 1}{3 \cdot 1} \right\rceil = 11.$$

$$a_{\max} = 11 \cdot 3 - \left\lfloor \frac{11 \cdot 3 \cdot 1}{32 + 1} \right\rfloor - 1 = 31.$$

$$a_{\max} = 31 \geqslant 15, \quad \text{подходит.}$$

$$v = \left\lceil \frac{32}{3} \right\rceil = 11.$$

Таким образом, для всех $0 \leqslant a \leqslant 31$

$$\left\lfloor \frac{a}{3} \right\rfloor = \left\lfloor \frac{11a}{2^5} \right\rfloor.$$

Но для $a = 32$ равенство уже не выполняется.

Пример 2

$d = 5$, $A = 5$, то есть $0 \leqslant a < 32$.
Организуем перебор m от 5.
Проверяем $m = 5$:

$$k = (-32) \bmod 5 = 3.$$

$$s = \left\lceil \frac{32+3}{5 \cdot 3} \right\rceil = 3.$$

$$a_{\max} = 3 \cdot 5 - \left\lfloor \frac{3 \cdot 5 \cdot 3}{32+3} \right\rfloor - 1 = 13.$$

$$a_{\max} = 13 \ngeqslant 31, \quad \text{не подходит.}$$

Действительно, возьмите $a = 14$ ($\nu = \lceil 32/5 \rceil = 7$):

$$\lfloor 14/5 \rfloor = 2 \neq 3 = \lfloor 14 \cdot 7/32 \rfloor.$$

Проверяем $m = 6$:

$$k = (-64) \bmod 5 = 1.$$

$$s = \left\lceil \frac{64+1}{5 \cdot 1} \right\rceil = 13.$$

$$a_{\max} = 13 \cdot 5 - \left\lfloor \frac{13 \cdot 5 \cdot 1}{64+1} \right\rfloor - 1 = 63.$$

$$a_{\max} = 63 \geqslant 31, \quad \text{подходит.}$$

$$\nu = \left\lceil \frac{64}{5} \right\rceil = 13.$$

Таким образом, для всех $0 \leqslant a \leqslant 63$

$$\left\lfloor \frac{a}{5} \right\rfloor = \left\lfloor \frac{13a}{2^6} \right\rfloor.$$

Но для $a = 64$ равенство уже не выполняется.

Пример 3

$d = 7$, $A = 32$, то есть $0 \leqslant a < 2^{32}$.
Организуем перебор m от 32.
Сразу проверяем $m = 35$ (зная ответ заранее):

$$k = 3.$$

$$s = 1\,636\,178\,018.$$

$$a_{\max} = 11\,453\,246\,124.$$

$$\nu = 4\,908\,534\,053 = 1\,24\,92\,49\,25_{(16)}$$

Число ν содержит 33 бита!

Примеры правильных значений v и m для некоторых d при $0 \leqslant a < 2^{32}$.

d	v	m
3	2 863 311 531	33
5	3 435 973 837	34
7	4 908 534 053	35
127	4 328 785 937	39
255	2 155 905 153	39
1 234 567	1 823 959 181	51
987 654 321	2 334 666 047	61
$2^{32} - 1$	$2^{31} + 1$	63
$2^{32} + 1$	1	32

!!! Не забудьте, что для чётных d нужно сначала преобразовать их к нечётным.

Вопрос: мы пытались избежать деления, но в результате получили, что для вычисления v оно всё-таки нужно.
Ответ: не страшно, так как замена деления на умножение требуется как минимум в двух случаях:

- деление на константу, известную заранее;
- деление на переменную в цикле, когда v и m рассчитываются один раз, а используются много раз.

Напомню, что это была ТЕОРИЯ, практические вопросы реализации заслуживают отдельной беседы.