Quotient estimation in multiprecision division algorithms

Saša Singer, Mladen Rogina

Math. Subj. Class. (1980) 68 B 99

ABSTRACT:
Quotient estimates in multiprecision division algorithms are considered. Estimates from above and from below, given in [1] are compared and a sharp error bound is obtained. When used with the usual normalization of operands both estimates have the same error bound of 2. A new normalization procedure is presented, which, in both cases, reduces the error bound to the optimal value of 1. The normalization is obtained using the leading digit of the divisor.

PROCJENA KVOCIJENATA U ALGORITMIMA DIJELJENJA S VIŠESTRUKOM TOČNOŠĆU
Razmatra se procjena kvocijenata u algoritmima za dijeljenje s višestrukom točnošću. Uspoređuju se procjene odozgo i odozdo, dane u [1] i daje se precizna ocjena pogreške. Uz uobičajenu normalizaciju operanda, obje procjene imaju pogrešku ograđenu s 2. Izlaže se novi postupak normalizacije koji, u oba slučaja, daje optimalnu ocjenu pogreške s iznosom 1. Normalizacija je bazirana samo na vodećoj znamenki divizora.

1 Introduction

We consider the problem of finding the integer quotient $q = \lfloor u/v \rfloor$, where u and v are integers given in positional notation in a certain base b, $b \geq 2$. The problem is of special interest when u and v are large enough to be unrepresentable in a computer. Such numbers are called multiprecision integers. We assume that u and v have standard representation in the following form:

$$u = \sum_{k=0}^{m} u_k b^k, \qquad 0 \le u_k \ < b, \qquad k + 0, \ldots, m \tag{1}$$

$$v = \sum_{k=1}^{n} v_k b^k, \quad 0 \le v_k \ < b, \quad k = 0, \ldots, n, \quad v_n > 0 \tag{2}$$

Without loss of generality we confine ourselves to the problem of finding a single–digit quotient. In general case this procedure is repeatedly applied, with reminders from earlier divisions as dividends. Therefore, we assume $m = n + 1$ in 1 and $q < b$.

If u and v are multiprecision numbers, it is impossible to calculate q exactly using machine arithmetics. So we seek a calculable estimate $\hat{q}$ of the quotient. It is assumed that b^2 is exactly representable which means that all operations involving individual digits can be performed exactly. The goal is to find $\hat{q}$ with the optimal error estimate using the smallest possible number of digits from u and v. We show that this is always possible using only the leading two digits of u and the leading digit of v.

2 Quotient estimates

Knuth [1] uses the following estimate

$$\hat{q} = \min\{\lfloor \frac{u_{n+1} b + u_n}{v_n} \rfloor, b - 1\} \tag{3}$$

and merely mentions the more natural one:

$$\hat{q}_L = \lfloor \frac{u_{n+1} b + u_n}{v_n + 1} \rfloor. \tag{4}$$

It is easily shown that:

Theorem 2.1 *$\hat{q}$ is a lower bound for q, and $\hat{q}_U$ is an upper bound for q:*

$$\hat{q}_L \le q \le \hat{q}_U. \tag{5}$$

The following result bounds the differences of these estimates:

Theorem 2.2

$$\hat{q}_U - \hat{q}_L \leq 1 = \frac{b-2}{v_n+1}. \tag{6}$$

Proof: We consider two cases depending on the value of $\hat{q}_U$. If $\lfloor (u_{n+1}b + u_n)/v_n \rfloor \geq b-1$, then from (3) $\hat{q}_U = b-1$. It follows

$$\hat{q}_U - \hat{q}_L = b - 1 - \lfloor \frac{u_{n+1}b + u_n}{v_n+1} \leq b - 1 - (\frac{u_{n+1}b + u_n}{v_n+1} - 1 + \frac{1}{v_n+1}) =$$

$$= b - \frac{u_{n+1}b + u_n + 1}{v_n+1}.$$

But $u_{n+1}b + u_n \geq (b-1)v_n$, so we obtain

$$\hat{q}_U - \hat{q}_L \leq b - \frac{(b-1)v_n + 1}{v_n+1} = 1 + \frac{b-2}{v_n+1}.$$

If $\lfloor (u_{n+1}b + u_n)/v_n \rfloor < b-1$ then $\hat{q}_U = \lfloor \frac{u_{n+1}b+u_n}{v_n} \rfloor$ and we have

$$\hat{q}_U - \hat{q}_L \leq \frac{u_{n+1}b + u_n}{v_n} - (\frac{u_{n+1}b + u_n}{v_n+1} - 1 + \frac{1}{v_n+1}) =$$

$$= \frac{u_{n+1}b + u_n}{v_n(v_n+1)} + 1 - \frac{1}{v_n+1}.$$

But now $u_{n+1}b + u_n < (b-1)v_n$, so

$$\hat{q}_U - \hat{q}_L < \frac{b-1}{v_n+1} + 1 - \frac{1}{v_n+1} = 1 + \frac{b-2}{v_n+1}. \bullet$$

3 Normalization

Theorem 2.2 shows that the error depends on the leading digit of the divisor. To make v_n as large as possible we use the so called normalization. Before division, numbers u and v are multiplied with a satiable normalization constant d dependent on the leading digit of v. It is common to use the following normalization:

Theorem 3.1 *(Knuth [1]) If we define*

$$d_1 = \lfloor \frac{b}{v_n + 1} \rfloor, \qquad \bar{v} = d_1 v \tag{7}$$

then

$$b - 1 \geq \bar{v}_n \geq \lfloor \frac{b}{2} \rfloor. \tag{8}$$

This result gives the error bounds:

Theorem 3.2

$$\hat{q}_U - \hat{q}_L \leq \begin{cases} 2 & \text{if} \quad b \geq 5 \\ 1 & \text{if} \quad b \leq 4. \end{cases} \tag{9}$$

Proof: From $v_n \geq \lfloor b/2 \rfloor$ we get $v_n + 1 \geq (b+1)/2$ and, from (6) it follows that

$$\hat{q}_U - \hat{q}_L \leq 1 + 2\frac{b-2}{b+1} = 3 - \frac{6}{b+1}.$$

The difference $\hat{q}_U - \hat{q}_L$ is an integer which completes the proof.

Corollary 3.1 *Relation (9) includes the following result [1]*

$$q \leq \hat{q}_U \leq q + 2 \tag{10}$$

as well as

$$q - 2 \leq \hat{q}_L \leq q. \tag{11}$$

It is easily shown that estimates in (9) and (10) are the best possible using normalization (7). Note that this normalization is achieved through a single-digit multiplication so it is very fast.

Using a slightly more complicated normalization better–even optimal error bounds are obtained:

Theorem 3.3 *If $v_n \geq b - 2$ then*

$$\hat{q}_U - \hat{q}_L \leq 1 \tag{12}$$

$$q - 1 \leq \hat{q}_L \leq q \leq \hat{q}_U \leq q + 1. \tag{13}$$

Proof: trivial.

There remains to describe the corresponding normalization:

Theorem 3.4 *Let v be such that $v_n \geq \lfloor b/2 \rfloor$ and define*

$$d_2(v) = \lfloor \frac{b^2}{v_n + 1} \rfloor \frac{1}{b} \tag{14}$$

$$\bar{v} = d_2(v)v, \qquad \tilde{v} = d_2(\tilde{v})\tilde{v}. \tag{15}$$

Then $\tilde{v}$ is normalized

$$b - 1 \geq \tilde{v}_n \geq b - 2. \tag{16}$$

Proof: It is obvious that for every v, $d_2(v) \geq d_2(v)v_n b^n$ so it suffices to consider $d_2(v)v_n$.

$$d_2(v)v_n = \lfloor \frac{b^2}{v_n = 1} \rfloor \frac{v_n}{b} \geq (\frac{b^2+1}{v_n+1} - 1)\frac{v_n}{b} = \frac{b^2+1}{b} - \frac{b^2+1}{b(v_n+1)} - \frac{v_n}{b}$$

and we define function f as

$$f(v_n) = \frac{b^2+1}{b} - \frac{b^2+1}{b(v_n+1)} - \frac{v_n}{b} \quad \text{for} \quad \mathrm{v_n \in [\alpha, b-1]}, \tag{17}$$

where $\alpha \geq 1$. Function f is easily seen to be monotonically increasing on the given interval. Therefore its minimal value is

$$\min_{v_n \in [\alpha, b-1]} f(v_n) = f(\alpha) \tag{18}$$

and we have

$$v_n \geq \alpha \Rightarrow d_2(v)v_n \geq f(\alpha). \tag{19}$$

From $v_n \geq \lfloor b/2 \rfloor$ it follows that

$$d_2(v)v_n \geq f(\lfloor b/2 \rfloor) \geq f(\frac{b-1}{2} \geq b - \frac{5}{2} + \frac{7b-1}{2b(b+1)} \tag{20}$$

which gives

$$\bar{v}_n \geq \begin{cases} b-3 & \text{if} \quad b \geq 6 \\ b-2 & \text{if} \quad b \leq 5. \end{cases} \tag{21}$$

For $\tilde{v}$, from (21) we have $\bar{v}_n \geq b - 3$, and from (19) it follows

$$d_2(\bar{v})\bar{v}_n \geq f(b-3) = b - 2 + \frac{2b-9}{b^2-2b} \tag{22}$$

and we conclude

$$\tilde{v}_n \geq b - 2 \qquad \text{if} \qquad \text{b} \geq 6. \tag{23}$$

If $b \leq 5$ then $\bar{v}$ is already properly normalized. •

Note that $d_2(v)$ has two digits in general. Therefore the normalized divisor $\tilde{v}$ can have two more digits than v, and the whole normalization requires more work than the previous one. But, we obtained a better estimate and that simplifies the division algorithm because it requires at most one subtraction to get the proper remainder.

We conclude by observing that the bound in (12), (13) is optimal. It is easy to construct examples in which it is impossible to estimate the quotient from any number of leading digits in v, but the whole number is required to calculate q exactly.

REFERENCES:

[1] KNUTH D.E., The Art of Computer Programming, Vol. 2 – Seminumerical Algorithms, Addison-Wesley, Reading 1969.